Published in final edited form as:

J Pharmacokinet Pharmacodyn. 2017 April ; 44(2): 95–111. doi:10.1007/s10928-016-9498-5.

Experiment Design for Nonparametric Models Based On Minimizing Bayes Risk: Application to Voriconazole¹

David S. Bayard² and Michael Neely³

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

Abstract

An experimental design approach is presented for individualized therapy in the special case where the prior information is specified by a nonparametric (NP) population model. Here, a nonparametric model refers to a discrete probability model characterized by a finite set of support points and their associated weights. An important question arises as to how to best design experiments for this type of model. Many experimental design methods are based on Fisher Information or other approaches originally developed for parametric models. While such approaches have been used with some success across various applications, it is interesting to note that they largely fail to address the fundamentally discrete nature of the nonparametric model. Specifically, the problem of identifying an individual from a nonparametric prior is more naturally treated as a problem of classification, i.e., to find a support point that best matches the patient's behavior. This paper studies the discrete nature of the NP experiment design problem from a classification point of view. Several new insights are provided including the use of Bayes Risk as an information measure, and new alternative methods for experiment design. One particular method, denoted as MMopt (Multiple-Model Optimal), will be examined in detail and shown to require minimal computation while having distinct advantages compared to existing approaches. Several simulated examples, including a case study involving oral voriconazole in children, are given to demonstrate the usefulness of MMopt in pharmacokinetics applications.

Keywords

Bayes risk; experiment design; nonparametric; population model; voriconazole

1 Introduction

Nonparametric (NP) population models are specified by a discrete parameter probability distribution. The richness of the nonparametric approach is in its ability to not simply estimate central tendency and dispersion, but rather to estimate the entire population parameter joint density. NP models arise naturally when applying maximum likelihood (ML) estimation to population modeling problems due to a remarkable fact that the optimal ML

¹Supported by NIH grants GM 068968 and HD 070886

²Consultant, Laboratory of Applied Pharmacokinetics and Bioinformatics, Children's Hospital Los Angeles

³Director, Laboratory of Applied Pharmacokinetics and Bioinformatics, Children's Hospital of Los Angeles

estimate is known to be discrete (Mallet [35], Lindsay [33]). NP models also arise in pharmacokinetic (PK) applications when computations are simplified by making discrete approximations to continuous distributions (Katz and D'Argenio [27]). Software programs for generating non-parametric PK population models are readily available, e.g., NPEM (Schumitzky [45][15]), NPAG (Leary [31], Baek [3]), and USC*PACK (Jelliffe [25]), which are now all available within the Pmetrics package for R [57].

An important question arises as to how to best design experiments with respect to non-parametric population model priors. Ideally, the experiment should be designed to extract the most information using the least number of samples. A widely used method for experiment design is to optimize the D-optimality criterion introduced by Box and Lucas [11], which consists of maximizing the determinant of the Fisher information matrix M (Silvey [49], Fedorov [18], D'Argenio [13]). One criticism of D-Optimal design is that it only uses the mean of the population model and ignores other distributional information. An improved class of approaches incorporating the full prior distribution are the so-called "robust" sampling designs, which more generally optimize the expectation of a specified scalar function of M (Pronzato and Walter [40], Tod and Rocchisani [50], Walter and Pronzato [56], D'Argenio [14]). Choices for this scalar function include the determinant of M (giving rise to ED-optimal design), the determinant of the matrix inverse of M (giving rise to EID-optimal design), and the log of the determinant of M (giving rise to ELD design). The ELD design is sometimes called the AIP design. Algorithms and software for computing the ED, EID and ELD designs are described in Pronzato and Walter [40], Tod and Rocchisani [50] [51] and Walter and Pronzato [56].

While such D-optimal approaches have been used with success across various applications, it is interesting to note that they largely fail to address the fundamentally discrete nature of the nonparametric problem. Specifically, the problem of identifying an individual from a nonparametric prior is more naturally treated as a problem of classification, i.e., of finding the support point that best matches the patient's behavior. Looking to the classification literature for inspiration, a classifier's performance is most often scored in terms of how well it minimizes the Bayes Risk. Bayes Risk is defined as the probability that a maximum a-posteriori (MAP) estimator will misclassify a subject (Duda et al. [17]). Hence, an experiment design minimizing Bayes Risk will ensure that the MAP estimate (which is one of the support points), is representative of the true patient with high probability. This is especially important for PK applications in light of the work of Sheiner et al. [46][47][48] and others (Vozech et al. [54][55], Rodman [43], Hu et al. [23]), who have demonstrated the value of MAP Bayesian control, i.e., where the MAP estimate is used as the basis for determining subsequent dosing regimens.

Interestingly, the authors are not aware of a single instance where Bayes Risk has been used as an information measure for experiment design purposes in a PK context. It is expected that this is because its computation is generally unwieldy. As pointed out in Fedorov (cf., [18], page 233), in the context of evaluating Bayes risk for model discrimination "... even for comparatively simple response surfaces, it is necessary to deal with serious computation difficulties, which make construction of the optimal design practically impossible". The current paper makes several strides in this direction. Specifically, use is made of a theoretical

upper bound on the Bayes Risk that has recently appeared in the literature in Blackmore, Rajamanoharan and Williams [10]. The upper bound is important because it can be computed completely algebraically without requiring multidimensional integration or stochastic analysis. Because of its computational advantages, this paper defines MMopt as the experiment design approach that minimizes this Bayes Risk upper bound. It will be shown that MMopt provides a flexible and effective method for designing experiments for NP models across a wide range of PK applications.

Background on experiment design is given in Section 2, with discussion of D-optimality, the ED, EID, and ELD robust designs, and the globally optimal Bayes optimal classifier (Bopt). In Section 3, the Bayes Risk upper bound of [10] is introduced. MMopt is defined as an experiment design approach that minimizes this upper bound. In Section 4, a simple two-support-point example is used to compare performance between the various designs. Sampling designs are applied in Section 5 to a more complicated four-support-point example. The MMopt experiment design approach is demonstrated on a simulated pharmacokinetic example in Section 6, and its usefulness in a voriconazole case study is demonstrated in Section 7. Section 8 summarizes the properties of MMopt relative to Bopt, ED, EID, and ELD. Appendices and a Supplemental File are included that provide detailed analysis on certain aspects of the various designs. Conclusions are postponed until Section 9.

This paper focuses on sampling times, but MMopt can be easily adapted to optimize over any other aspects of experiment design (i.e., dose timing, dose amounts, etc.)

2 Background

2.1 Dynamic Model and Measurements

A nonlinear dynamic model is defined of the form,

$$\dot{x}(t) = f(x(t), d(t), \theta) \quad (1)$$

where x is the state, d is the input, and $\theta \in \mathcal{R}^p$ is the parameter vector, and f is a function of these three quantities. The system output at time t_k is given as,

$$\mu_k = h(x(t_k), \theta) \quad (2)$$

where the observed function h depends on both the state x and parameter vector θ . A noisy measurement of μ_k is taken of the form

$$y_k = \mu_k + \sigma_k n_k \quad (3)$$

where $n_k \sim \mathcal{N}(0, 1)$ is the measurement noise, and σ_k scales n_k to its desired level. The experiment design problem is formulated to find the n optimal sampling times

$$U = \{t_1, \dots, t_n\} \quad (4)$$

that act to best identify the parameter vector θ .

The parameter vector θ can have either continuous or discrete values. When applied to NP problems, θ is assumed to belong to a discrete set $\theta \in \Theta \triangleq \{\theta_1, \dots, \theta_m\}$, having Bayesian prior probabilities $p(\theta_i)$, $i = 1, \dots, m$. The true value of θ is denoted as θ^* . A slight generalization of (3) is to allow σ_k to be a function of μ_k according to an “assay polynomial” of the form,

$$\sigma_k(\mu_k) = c_0 + c_1\mu_k + c_2\mu_k^2 + c_3\mu_k^3 \quad (5)$$

Here the coefficients $c_0, \dots, c_3$ are typically determined from a separate laboratory assay error calibration.

2.2 D-Optimal Design

The D-Optimal experiment design (denoted here as Dopt), is defined by maximizing the determinant of the Fisher Information Matrix [49][18][13],

D-Optimal Design—

$$U_D \triangleq \arg \max_U |M| \quad (6)$$

where $|(.)|$ denotes the matrix determinant, and the Fisher Information Matrix M is given by,

$$M(\theta, U) = \sum_{k=1}^n \frac{1}{\sigma_k^2} \left[\frac{\partial \mu_k}{\partial \theta} \frac{\partial \mu_k}{\partial \theta^T} \right] \bigg|_{\theta=\theta^*} \quad (7)$$

D-optimal designs have the property that they minimize the volume of confidence ellipsoids for the parameters. D-Optimal design has become a useful tool for optimizing experiments over a wide range of applications [19][29].

The partial of μ in (7) is evaluated on the true parameter value θ^* . This is cause for some concern, since the true parameter value θ^* is not known, and in fact is the main reason for doing experiment design in the first place. However, for problems where $\mu_k = h(x(t_k), \theta)$ is linear in θ ,

$$\mu_k = h(x(t_k))^T \theta \quad (8)$$

where h is a column vector, the partial becomes,

$$\frac{\partial \mu_k}{\partial \theta} = h(x(t_k)) \quad (9)$$

In this case the dependence on θ vanishes, and the problem goes away. Hence, Dopt is best suited to problems for which μ_k has the form (8) which is a linear function of θ .

Furthermore, when the Fisher Matrix M is singular its determinant is zero. This implies that D-optimal sampling design is only defined when at least p samples are taken, where p equals the number of parameters.

2.3 Robust D-Optimal Designs

A drawback to using D-Optimal designs with nonlinear models of the form (1)–(3), is that μ is nonlinear in θ . This means that the D-optimal design becomes a function of the unknown true parameter vector θ^* , which presents a problem since it is not known beforehand. A more principled approach is to use robust D-optimal designs which optimize the expectation of a cost function, where the expectation is taken over the random variable θ , and where the cost function is specified as a scalar function of the D-optimal cost. The most common robust D-optimal designs are based on the ED, EID, ELD experiment design criteria defined as (cf., Chaloner and Verdinelli [12], Pronzato and Walter [40][56], Tod [50], D'Argenio [14]),

ED Optimal Design

$$\arg \max_{\xi} E_{\theta} (|M|) \quad (10)$$

EID Optimal Design

$$\arg \min_{\xi} E_{\theta} \left(\frac{1}{|M|} \right) \quad (11)$$

ELD Optimal Design

$$\arg \max_{\xi} E_{\theta} (\log |M|) \quad (12)$$

Again, since the Fisher matrix determinant is only non-zero when the Fisher Matrix M is nonsingular, these robust designs have the same restriction as D-optimal designs in that they must contain at least p samples, where p equals the number of parameters.

Other robust D-optimal approaches are possible. One variation is to add the precision matrix Ω^{-1} to M in any of the above expressions, where Ω is the prior covariance matrix of θ . Such modifications are discussed in [12]. Another variation is to reverse the order of the determinant and expectation in the ED design (10) to give $|E_{\theta}(M)|$. Combining both of these latter two modifications gives the design criterion $|E_{\theta}(M) + \Omega^{-1}|$ which has been applied to PK problems in Hennig et al. [22] and Merle and Mentre [36].

In this paper, MMopt experiment designs will be compared to ED, EID, and ELD optimal robust designs under the assumption of an NP prior. This comparison is not intended to be exhaustive as other experiment design methods are possible. Comparisons with methods such as D_s optimal design (Goodwin and Payne [21]), T-optimal design (Atkinson et al. [2] Vajjah and Duffull [53]), alternative robust D-optimal designs [22][36], and Bayesian design (Chaloner and Verdinelli [12], Ryan et al. [42]), are left for future studies.

2.4 Bayes Optimal Classifier

Given a vector of observations $Y \in \mathcal{R}$ and a vector of control inputs $U \in \mathcal{U}$, the posterior probability $p(H_i|Y, U)$ of the i th hypothesis H_i in an m -category classification problem is calculated using Bayes rule as,

$$p(H_i|Y, U) = \frac{p(Y|H_i, U)p(H_i)}{p(Y|U)} \quad (13)$$

where $H_i, i = 1, \dots, m$ represent the m individual hypotheses, and $p(H_i), i = 1, \dots, m$ represents their corresponding prior probabilities.

REMARK 2.1: When applied to discrete priors, H_i is understood to be the hypothesis that the true value of θ corresponds to θ_i . Bayesian prior support point probabilities are denoted as $p_i = p(\theta_i), i = 1, \dots, m$.

After computing the posterior probabilities from (13), the most probable hypothesis is defined by H_i where,

$$H_i = \arg \max_j p(H_j|Y, U) \quad (14)$$

This selection rule is denoted as the **Bayes Optimal Classifier**. Since the Bayes optimal classifier maximizes the posterior probability, it is also a Maximum A-Posterior (MAP) estimator.

Since (14) is only an estimate, it will sometimes be correct and sometimes be incorrect. The probability of it being incorrect is denoted as the **Bayes Risk**. For a given U , equation (14) can be thought of as a prescription for dividing up the domain $\mathcal{R}$ into at most m regions defined by the following relation,

$$\mathcal{R}_i \triangleq \{Y | p(H_i|Y, U) > p(H_\ell|Y, U) \text{ for all } \ell \neq i\} \quad (15)$$

This definition allows one to state the Bayes optimal classifier (14) simply as an algorithm that selects model H_i as the MAP estimate if and only if Y falls into region $\mathcal{R}_i$.

Multiplying both sides of the inequality in (15) by $p(Y|U) > 0$ and rearranging allows region $\mathcal{R}_i$ to be equivalently defined as,

$$\mathcal{R}_i \triangleq \{Y | p(H_i, Y|U) > p(H_\ell, Y|U) \text{ for all } \ell \neq i\} \quad (16)$$

The Bayes Risk is calculated by summing over all possible ways that a given classifier can make a classification error, i.e., by summing over the probabilities of selecting H_j when in fact H_i is true,

$$p(error) = \sum_{i=1}^m \sum_{j \neq i} p(Y \in \mathcal{R}_j, H_i|U) \quad (17)$$

$$= \sum_{i=1}^m \sum_{j \neq i} p(Y \in \mathcal{R}_j | H_i, U) p(H_i) \quad (18)$$

$$= \sum_{i=1}^m \sum_{j \neq i} \int_{\mathcal{R}_j} p(Y | H_i, U) p(H_i) dY \quad (19)$$

The experiment design that minimizes the Bayes Risk (19) is defined as the Bayes Optimal Design, or Bopt for short,

Bayes Optimal Design (Bopt)

$$U_B \triangleq \arg \min_U p(error) \quad (20)$$

3 Multiple Model Optimal (MMopt) Design

3.1 Bayes Risk Overbound

Computing the Bayes Risk is generally a very difficult problem. However, a recent overbound on the Bayes Risk has appeared in the literature that is relevant to the special case where the likelihood functions are Gaussian, i.e.,

$$p(Y|H_i, U) = N(\mu(i), \Sigma(i)), \quad i=1, \dots, m \quad (1)$$

The overbound is given in the following result.

THEOREM 3.1 (Blackmore, Rajamanoharan and Williams [10]): *When performing hypothesis selection between m hypotheses, for Gaussian observation distributions such that $p(Y|H_i) = N(\mu(i), \Sigma(i))$, $i = 1, \dots, m$, the Bayes Risk is upper bounded as follows,*

$$p(\text{error}) \leq \sum_{i=1}^m \sum_{j>i} p(H_i)^{\frac{1}{2}} p(H_j)^{\frac{1}{2}} e^{-\kappa(i,j)} \quad (2)$$

where,

$$\kappa(i, j) = \frac{1}{4} (\mu(j) - \mu(i))^T [\Sigma(i) + \Sigma(j)]^{-1} (\mu(j) - \mu(i)) + \frac{1}{2} \ln \frac{|\frac{1}{2}\Sigma(i) + \frac{1}{2}\Sigma(j)|}{|\Sigma(i)|^{\frac{1}{2}} |\Sigma(j)|^{\frac{1}{2}}} \quad (3)$$

Proof: See Blackmore et al. [10].

3.2 MMopt Design

Theorem 3.1 is important because the expression (2) overbounds the Bayes Risk and hence represents a form of pre-posterior analysis (Raiffa and Schlaifer [41]). However, in contrast to many of the pre-posterior analysis expressions found in modern Bayesian optimal design (cf., Ryan et al. [42]), expressions (2)(3) can be calculated analytically without requiring multidimensional integration or stochastic analysis. For the purpose of this paper, a design that minimizes the Bayes Risk overbound (2) is denoted as the Multiple Model optimal experiment design (MMopt). Specifically,

MM Optimal Design

$$U_{MM} = \arg \min_U \sum_{i=1}^m \sum_{j>i} p(H_i)^{\frac{1}{2}} p(H_j)^{\frac{1}{2}} e^{-\kappa(i,j)} \quad (4)$$

MMopt designs can be computed for many problems of practical interest. The overbound was first developed and used for active estimation in Blackmore et al. [10]. These ideas were later extended by the present authors in a series of workshops on optimal experiment design (Bayard, Jelliffe and Neely [7][8]). Since Bayes Risk is always non-negative, the minimization of an overbound on the Bayes Risk will intuitively act to drive it toward lower values, as desired.

The minimization of Bayes Risk acts to make the MAP estimate a better predictor (i.e., classifier), of the true patient. For example, if the Bayes Risk is reduced to 0.1 by proper experiment design, it means that the MAP estimate (which itself is always a support point in the discrete model) will correspond to the true patient with probability 0.9. This interpretation can be very useful in PK applications where the MAP estimate is used to guide subsequent dosing and/or therapeutic recommendations. Currently, a large number of successful dosing and therapeutic drug monitoring approaches are based on MAP Bayesian control (cf., Sheiner et al. [46][47][48], and others [54][55][43]). The efficacy of such methods is directly based on the accuracy of the MAP estimate. Since MMopt acts to minimize the Bayes Risk, it improves the discriminating power of the MAP Bayesian estimate, which in turn, has the potential advantage of improving control performance across a wide range of PK applications.

4 Two-Support-Point Example

A simple two-support-point example is given to demonstrate the difference between treating an experiment design problem as a parameter estimation problem (i.e., using D,ED,EID,ELD designs) versus treating it as a classification problem (i.e., using MMopt).

4.1 System Model

A two-support-point example involving an exponential measurement model is defined as

$$y_k = \mu(t_k, a) + \sigma n_k \quad (5)$$

$$\mu(t, a) = e^{-at} \quad (6)$$

The Fisher Information Matrix M for a single sample time is computed as,

$$\frac{\partial \mu}{\partial a} = -te^{-at} \quad (7)$$

$$M(a) = \frac{1}{\sigma^2} \left(\frac{\partial \mu}{\partial a} \right)^2 = \frac{1}{\sigma^2} (te^{-at})^2 \quad (8)$$

and the D-optimal cost follows as

$$|M(a)| = t^2 e^{-2at} / \sigma^2 \quad (9)$$

Values for parameter a are given as

$$\text{Model 1: } a_1 = 1.5 \text{ (fast support point)} \quad (10)$$

$$\text{Model 2: } a_2 = 0.25 \text{ (slow support point)} \quad (11)$$

with $\sigma = 0.3$, and with the prior specified as $p_1 = .5$ and $p_2 = .5$. The two support point responses are plotted in Figures 1–9. For this example, all optimizations are performed over a discrete time grid spanning 0 to 10 hours, with a $t = 0.0001$ hour spacing.

4.2 Designs for Two-Support Point Example

It is known that the Dopt design for an exponential model of the form (5)(6) is given by $t = 1/a$ (Landaw [30]), i.e., at one time constant of the model response, or equivalently where the response has dropped to $e^{-1} = 0.3679$ of its initial value. Hence, minimizing (9) over t gives the individual D-optimal designs as,

$$t_D(\text{fast}) = 1/a_1 = 0.6667 \quad (12)$$

$$t_D(\text{slow}) = 1/a_2 = 4 \quad (13)$$

These two D-optimal designs are plotted with a circle and diamond, respectively, in Figure 1, and indicate the time constants for each response. Clearly, the optimal sample times for the two support points are very different (i.e., 0.6667 versus 4). This exposes a dilemma when using the D-optimal design. What is good for one support point is very different from what is good for the other support point.

ED-Optimal Design—One method to resolve the above dilemma is to use a robust design approach such as ED design. As discussed earlier, ED design makes use of prior information, and is defined by maximizing over the expected value of $|M|$,

$$t_{ED} = \arg \max_t E[|M|] = \arg \max_t .5|M(a_1)| + .5|M(a_2)| \quad (14)$$

Here $|M|$ is given by (9). Optimization of (14) for this example gives the ED design as

$$t_{ED} = 3.9991 \quad (15)$$

The ED cost (14) is a weighted combination of the individual D-optimal costs given earlier in (9), with weights of (0.5, 0.5) corresponding to the prior probabilities (p_1, p_2) . The situation is depicted in Figure 2. The ED-optimal sampling design of $t = 3.9991$, represents a compromise between the Dopt values of $t = .6667$ and $t = 4$. Here, the weighting strongly favors the term $|M(a_2)|$ which is associated with the slow a_2 support point.

Bayes Optimal Design—Computation of the Bayes Optimal Design is usually unwieldy. Fortunately, the current example is simple enough to allow its analytic computation.

Responses for the two model support points (10)(11) are shown in Figure 3, where the response separation is defined by $r(t) = |\mu(t, a_1) - \mu(t, a_2)|$. Since there is additive Gaussian noise associated with each measured response, the Bayes Risk (19) can be represented as the grey area in Figure 4. Here, Figure 4 can be thought of as looking at Figure 3 “sideways”, such that the two Gaussian “bumps” represent the two hypotheses associated with the true patient being Model 1 or Model 2. The Bayes Optimal Design minimizes this Bayes Risk (i.e., grey area), which can be calculated analytically as

$$\text{Bayes Risk} = \text{Prob}(\text{Type I error}) + \text{Prob}(\text{Type II error}) \quad (16)$$

$$= p_1 \int_{y_{\text{cross}}}^{\infty} p(y|a_1) dy + p_2 \int_{-\infty}^{y_{\text{cross}}} p(y|a_2) dy = 1 - \int_{-\infty}^{r(t)/2} N(0, \sigma^2) dy \quad (17)$$

where y_{cross} is where the two Gaussians intersect in Figure 4. Intuitively, this calculation is simply adding the standard Type 1 and Type 2 error probabilities that arise in a classical two-hypothesis test (i.e., defined as if we were treating a_1 as the null hypothesis). The response separation $r(t)$ is an important statistic in the two-support-point example.

Equation (17) indicates that the Bayes Risk decreases monotonically with increasing $r(t)$. This is seen graphically in Figure 4 where the sum of the two grey areas decreases monotonically as the two distributions are “pulled apart”. Hence the Bayes Risk is minimized at the time of maximal response separation $r(t)$ given by,

$$t_B = \arg \max_t r(t) = \arg \max_t |\mu(t, a_1) - \mu(t, a_2)| = 1.4334 \quad (18)$$

MMopt Design—The MMopt design is calculated by minimizing the Bayes Risk overbound (2) as

$$t_{MM} = \arg \min_t p_1^{\frac{1}{2}} p_2^{\frac{1}{2}} e^{-\kappa(1,2)} = \arg \min_t p_1^{\frac{1}{2}} p_2^{\frac{1}{2}} e^{-(\mu(t, a_1) - \mu(t, a_2))^2 / (8\sigma^2)} \quad (19)$$

The minimization is changed to a maximization by taking the negative log of (19) and then is simplified to give,

$$t_{MM} = \arg \max_t |\mu(t, a_1) - \mu(t, a_2)| \quad (20)$$

Since MMopt directly maximizes the response separation $r(t) = |\mu(t, a_1) - \mu(t, a_2)|$, it is identical to the Bayes Optimal Design (18) for this example, and is given by,

$$t_{MM} = 1.4334 \quad (21)$$

Summary of Results—Results are summarized in Table 4.1. The true Bayes Risk for each design is calculated using Monte Carlo simulation with 1.4×10^6 realizations.

MMopt gives the identical sampling time $t = 1.4334$ as the Bayes Optimal design, with the identical Bayes Risk of 0.1659. This is due to the fact that they both maximize response separation, which was seen in (17) to minimize Bayes Risk and be the optimal strategy for discriminating between the two support points. The ED optimal sampling time of 3.9991 is very different from the Bayes optimal, and suffers a Bayes Risk that is 60% larger. (i.e., 0.2712 versus 0.1659). The fact that the ED design discriminates poorly is not surprising since the criteria (14) is a simple average (over the prior) of the Fisher information calculated for each support point separately, which is generally unrelated to response separation. Likewise, the remaining Fisher-Information-based designs (EID, and ELD) do not perform as well as Bopt and MMopt (although ELD is a close second), due to their criteria (11) (12), which ignore response separation.

Here, the MMopt sample times ensure that the MAP Bayesian estimate (i.e., the largest support point of the Bayesian posterior), will correctly determine the true patient 83 percent of the time (i.e., $0.8341 = 1 - 0.1659$). This can be an important advantage in applications to PK control, where the MAP estimate is often used to compute the patient's next drug dose or to decide subsequent treatment.

In summary, it has been shown that the key idea when optimizing experiment design for a nonparametric model (with two support points - the simplest possible case), is to pick a sampling time that maximizes the separation between the two model responses. It is interesting that this notion is completely missing (and should not be expected), from Fisher-Information based designs which are derived and justified based on completely different arguments. This example demonstrates the usefulness of MMopt, and clearly underscores the inherent difference between trying to design an optimal experiment to learn a parametric model versus a nonparametric model.

4.3 Designs for Two-Support-Point Example with Closely-Spaced Parameters

A modification of the two-support-point example is now studied, where the uncertainty in parameter a is chosen very small. In this case, it will be seen that the Bayes Optimal, MMopt, and the ED, EID, and ELD designs all give very similar results.

The two-support-point example (5)(6) is used again as in the previous example. However, in the current example the values for a_1 and a_2 are chosen very close

$$\text{Model 1: } a_1 = 1.1 \text{ (fast support point)} \quad (22)$$

$$\text{Model 2: } a_2 = 1 \text{ (slow support point)} \quad (23)$$

so that there is very little uncertainty in the problem. The two support point responses are plotted in Figure 5. The D-optimal designs for each support point taken separately are calculated as $t = 1/a_1 = 0.9091$ (for fast support point a_1) and $t = 1/a_2 = 1$ (for slow support point a_2). These sample times are relatively close together, in contrast to the previous example. Results are summarized in Table 4.2. As expected by the theory, the MMopt design again agrees exactly with Bayes optimal design. However, somewhat unexpectedly, the remaining designs (ED, EID, ELD, D-Optimal) are also very close to Bayes Optimal design, with essentially the same optimal sampling time at $t = 0.95$ hours.

It is shown in Appendix A that the fact that the various designs give the same optimal sampling time in this example is not an accident, but rather, follows from the asymptotic properties of the robust design cost functions as the parameters a_1 and a_2 become close. Specifically, it is proved in Corollary A.1 (for the Restricted Two-Support-Point Problem which is a slightly more general problem than treated here), that as the parameter uncertainty $a = a_2 - a_1$ becomes vanishingly small, the objectives of maximizing J_{ED} , J_D , J_{ELD} and minimizing J_{EID} asymptotically approach the single common objective of maximizing the response separation $r(t) = |\mu(a_1, t) - \mu(a_2, t)|$ over time t .

In summary, results from this two-support-point example indicate that the ED, EID, ELD, and D-Optimal designs applied to models having closely-spaced parameters (i.e., small model uncertainty), tend to agree with the Bayes optimal and MMopt designs in this special case. The result is not just numerical, but is supported by the theory provided in Appendix A.

5 Four-Support-Point Example

5.1 System Model

In this section the ED, EID, ELD, and D-Optimal designs are compared with MMopt on a more complicated four-support-point example with 2 unknown parameters. A two-parameter exponential measurement model is defined as

$$y_i = \mu(t_i, a, b) + \sigma n_i \quad (24)$$

$$\mu(t, a, b) = be^{-at} \quad (25)$$

$$n_i \sim N(0, \sigma^2) \quad (26)$$

Discrete parameter values are given in Table 5.3, with $\sigma = 0.1$ with prior $p_i = .25$, $i = 1, \dots, 4$. The four support point responses are plotted in Figure 6.

In this example, model parameters are intentionally chosen to help demonstrate the difference between using Fisher information and classification as experiment design objectives. Specifically, Fisher-information-based designs strongly favor putting a sample at the time of the highest output response, which in this example occurs at time $t = 0$. However, from a classification point of view, a sample at time $t = 0$ is partially wasted because it does not help discriminate between all responses. Specifically, 3 of the 4 support points shown in Figure 6 have the same output response at $t = 0$ and are not well separated.

The designs are optimized over a time grid $t \in [0, 4]$, with grid points separated by 0.05 sec apart. The Bayes optimal design is too difficult to compute for this example and is not included. The Bayes cost for each design is evaluated using Monte Carlo analysis with 1.4×10^6 runs per estimate. Results are compared for the two and three sample time cases only, since the model has 2 parameters and the ED, EID and ELD designs do not exist for the one sample design.

5.2 Designs for Four-Support-Point Example

Results for the various metrics are summarized in Table 5.4 for a two sample design, and in Table 5.5 for a three sample design. As expected, the ED, EID, and ELD designs place a sample at $t = 0$ in all cases. The MMopt design improves on these designs by shifting this first sample slightly to the right (to $t = 0.45$ sec), at which time the response curves are better separated, and hence better discriminated. It is seen that MMopt has smallest Bayes Risk for all cases studied.

6 Pharmacokinetic Example

MMopt experiment design is demonstrated on a simulated pharmacokinetic example. The example is based on a one-compartment (first-order) model,

$$\dot{x} = -Kx + d(t) \quad (1)$$

$$\mu_k = \frac{x(t_k)}{V} \quad (2)$$

$$y_k = \mu_k + \sigma_k n_k \quad (3)$$

$$n_k \sim N(0, 1) \quad (4)$$

$$\sigma_k = 0.1 \quad (5)$$

Here, x is the amount of drug in the central compartment, $d(t)$ is the dose input (time-dependent in general), K is the elimination constant, V is the volume of distribution, y_k is the noisy measurement at time t_k , and σ_k is the standard deviation of its noise. The population model is comprised of $m = 10$ randomly chosen support points, whose values of V and K are given in Table 6.1,

The Bayesian prior is defined such that all 10 support points have equal probability, i.e., $p_i = 1/10$, $i = 1, \dots, 10$. The system is driven by a dosing profile $d(t)$ having a value of 300 units that is held constant for 1 hour and then shut off. The ten support point responses to the dose input $d(t)$ are shown in Figure 7. All optimization is performed over a discrete uniform time grid ranging from $t = 0$ to $t = 24$ hours, with time points spaced at $t = 0.25$ hours (i.e., 15 minutes) apart. This time grid is chosen coarser than the grids used previously in order to simplify the optimization process, while at the same time being consistent with what can reasonably be enforced in a practical clinical application. For each case studied, results are computed using a Monte Carlo simulation having 1×10^6 runs. The corresponding 99% confidence bounds are calculated from the Monte Carlo statistics and reported for completeness.

6.1 Design for PK Estimation

MMopt design is applied to the problem of estimating the PK parameters V and K in model (1)–(3) by designing experiments having 1, 2 and 3 optimal sample times. For the 1 sample design, MMopt is compared to Bopt (the optimal Bayesian design), while for the 2 and 3-

sample designs, MMopt is compared to ED. This choice is motivated by the fact that ED is not defined for a 1 sample design, while Bopt becomes computationally unwieldy for the 2 and 3-sample cases.

Performance is summarized in Table 6.2. For the 1-sample design, the true Bayes Risk is computed numerically using 1×10^6 Monte Carlo runs, and is shown in Figure 8. The minimum is attained at $t = 4.25$ hours. The shape of the curve is interesting, having several local minima which would generally complicate its optimization. For the 1-sample design, MMopt also attains an optimal at $T = 4.25$ hours, which is identical to that of Bopt. Hence, MMopt is globally optimal for this example. Of course this result is obtained on a coarse time grid, so they may not agree exactly on a finer grid. For the 2 and 3 sample designs, MMopt has lower Bayes risk than ED design. In both cases, the improvement is statistically significant with p-values smaller than $2.2e-16$ (machine precision). The low p-values are due to the high confidence obtained from using a large number of runs in each Monte Carlo evaluation.

7 Voriconazole Case Study

We wished to use MMopt to develop an optimal first sampling strategy for oral voriconazole in children. Because this drug has a mixture of linear and non-linear elimination, which is related to but not entirely predicted by age, it is a challenge for practitioners to know when steady state is achieved. This in turn makes comparison of a trough concentration to a pre-established target steady-state range of 1 to 5.5 mg/L difficult due to the uncertainty about steady-state conditions. For this optimal sampling project, we used the same dataset that we previously used to build a population model of voriconazole in adults and children (Neely et al. [38]). The data comprised 141 subjects, with 85 children ranging from 2 to 11 years of age, and 56 adults 19 to 55 years of age. The population model had 7 parameters with linear absorption, allometrically scaled Michaelis-Menten clearance and volume of the central compartment, a peripheral tissue compartment, and bioavailability. There were 125 support points in the final nonparametric population model.

We first used the MMopt function that has been incorporated into our Pmetrics nonparametric population modeling and simulation package for R to calculate one and two optimal times to sample after the first maintenance dosage of voriconazole. The MMopt calculations were constrained to 0.5 hour intervals on the range from 0 to 36 hours. We used the population points and assay error polynomial (5) of $c_0 = 0.02$, $c_1 = 0.1$, $c_2 = c_3 = 0$, with loading dosages of 9 mg/kg two times in the first 24 hours, followed by the first maintenance dose of 8 mg/kg at the beginning of day 2, 12 hours after the second loading dose. We combined three sets of 125 time-concentration profiles generated from the support points in the model, with each set corresponding to a child aged 2, 7 or 12 years, and weighing the median for each age, according to CDC growth charts: 12.2, 23.3 and 36.1 kg. The single optimal time point to sample (MMopt1) was 1 hour before the second maintenance dose, i.e. 35 hours into the regimen. The two optimal time points (MMopt2) were a peak and trough, 1 hour after the first maintenance dose, and just before the second maintenance dose, i.e. 25 and 36 hours into the regimen. This is shown in Figure 9.

Next, we used the population model with full covariances to simulate 50 new subjects administered oral voriconazole 9 mg/kg q12h for two doses, followed by a single 8 mg/kg dose. The mean (range) simulated weights and ages were 40.6 (12 to 89) kg and 6.4 (2.2 to 11.1) years. We simulated hourly concentrations until 24 hours after the final dose, and these observations were corrupted with random noise drawn from a Gaussian distribution with mean of 0 and standard deviation based on the assay error coefficients above.

The complete data for each simulated subject was defined as the “full” dataset. We depleted the full data set to include only the MMopt2 observations at times 25 and 36 hours. We also depleted it to include only the MMopt1 observation time at 35 hours. We used the model to calculate the full Bayesian posterior with concentrations every 12 minutes from 0 to 48 hours for each dataset (full, MMopt2, MMopt1). From each posterior, we calculated the area-under-the-curve (AUC) from 24–36 hours by trapezoidal approximation, as well as extracted the concentration at time 25 hours (1 hour post-dose peak) and 36 hours (12 hour post-dose trough).

Using the full dataset to generate a Bayesian posterior, the simulated observations could be predicted as $0.05 + 0.92 \times \text{prediction}$, $R^2 = 0.87$, indicating a good fit. We therefore defined the AUC, peak and trough from the Bayesian posterior based on the full dataset to be the reference, and compared the AUCs, peaks and troughs from the MMopt2 and MMopt1 posteriors using a paired T-test. The results are shown in Table 7.1. Also included in the table is the 95th percentile of the absolute prediction difference as a measure of imprecision.

In conclusion, we could show that using two MMopt samples, one obtained 1 hour after the first maintenance dose and one just prior to the second maintenance dose, permits estimation of voriconazole AUC, peaks and troughs with reasonably low bias. While the peaks are significantly underpredicted, this is not likely to affect clinical outcome since efficacy has been linked to AUC or trough concentrations [1][24][26][34][37], and the absolute underprediction is low. A single sample 1 hour before the second maintenance dose is very similar in accuracy to two samples. Using one or two MMopt samples with this population model results in an error in trough prediction that is 95% likely to be less than 0.38 or 0.93 mg/L, respectively. D-optimal sampling would require seven blood samples, which is not readily feasible even in the inpatient setting. MMopt sampling will permit rapid assessment of the appropriateness of the voriconazole oral dose in children, and subsequent sampling times in individual patients can be guided by further use of MMopt for each patient. Clinicians may now have a tool that overcomes the concerns about steady state, and which uses the patient’s own measured concentrations to control both the dose and the time to best sample.

8 Discussion

Several properties of the various experiment design approaches are summarized in Table 8.2, and are discussed below.

An interesting question arises whether an experiment design is invariant under a change of variables in the model. Let an estimation problem having parameter vector $a \in \mathbb{R}^p$ be reparametrized in terms of the vector $b \in \mathbb{R}^p$ where a and b are related by the mapping,

$$a=f(b) \quad (1)$$

The mapping (1) is assumed to be *regular* in the sense that it is one-to-one and has continuous partial derivatives. In the case of a regular *linear* reparametrization, the mapping (1) takes the form

$$a= Fb \quad (2)$$

where $F \in \mathbb{R}^{p \times p}$ is a square and invertible matrix.

[P1]: *Invariance under regular linear reparametrization.*

[P2]: *Invariance under regular nonlinear reparametrization.*

An obvious property of support point responses is that they are invariant under regular reparametrization. Accordingly, methods such as Bopt and MMopt based purely on support point responses, will also trivially be invariant under a regular reparametrization (of both the linear and nonlinear kind). However the same is not true for the Fisher-based robust designs. An analysis performed in the Supplemental File [58] proves that while all the robust designs are invariant under linear parametrization, only the ELD design (not the ED and EID designs) are invariant under nonlinear parametrization.

This lack of invariance can have significant consequences in practice. For example, a one-compartment PK model is parameterized by volume V and elimination rate K as

$$\dot{x} = -Kx + d \quad (3)$$

$$\mu_i = \frac{x(t_i)}{V} \quad (4)$$

Here, the drug amount in the compartment is x , the dose is d , and μ_i indicates the noiseless part of a drug concentration measurement at time t_i . However, an alternative parametrization is given by using parameters volume V and clearance C (where $C = VK$),

$$\dot{x} = -(C/V)x + d \quad (5)$$

$$\mu_i = \frac{x(t_i)}{V} \quad (6)$$

According to Table 8.2 property (P2), when using the ED or EID methods, a clinician would have to compute separate experiment designs depending on whether he was using the (V, K) or (V, C) parametrization. This behavior is nonintuitive as the responses are identical under either parametrization and should not call for separate designs. In contrast, the ELD, MMopt and Bopt designs are invariant under this change of variables.

[P3]: *Avoids dependence on estimator asymptotic properties.*

[P4]: *Avoids knowing true parameters in nonlinear problems.*

As discussed in Pronzato and Pazman [39], there are two common avenues for circular reasoning to enter into the experimental design process. The first (cf., (P3)), is the dependence on asymptotic properties of the estimator. This represents a contradiction since experiment designs are typically used for placing only a small (non-asymptotic) number of samples, while certain information measures (e.g., Fisher Information) are rigorously only valid asymptotically as the number of samples goes to infinity. The second (cf., (P4)), is the assumption that the true parameters are available beforehand for use in computing the optimal design. This represents a contradiction since it is the whole point of experimental design to find these parameters in the first place.

MMopt and Bopt both satisfy (P3) and (P4) and completely avoid these contradictions. However, the robust designs do not satisfy (P3) because they are defined in terms of Fisher information which involves an asymptotic approximation. Nevertheless, all robust designs avoid the second contradiction (and hence satisfy (P4)), due to the fact that they average over the prior distribution.

[P5]: *Allows taking fewer than p samples.*

Both Bopt and MMopt allow taking fewer than p samples, where p is the number of parameters. The D-optimal robust designs do not enjoy this same property since their cost would be zero for all sub-sampled designs. It is worth noting that certain heuristic modifications of D-optimal designs have been introduced to help overcome this limitation [16].

[P6]: *Can handle heterogeneous model structures.*

The Bopt and MMopt experiment designs are based purely on support point responses and as such, are able (trivially) to handle heterogeneous model structures. In particular, the underlying discrete NP population can be a mix of models having different orders or numbers of compartments, linear or nonlinear dynamics, or any other combination of heterogeneous characteristics. The D-optimal designs do not enjoy this same property. However, several extensions relevant to model discrimination have been developed (cf., Atkinson et al. [2], Vajjah and Duffull [53]).

[P7]: *Gives known optimal solution to 2-support-point example.*

It was shown theoretically and confirmed numerically in the two-support-point example of Section 4.2, that both Bopt and MMopt maximize the response separation and consequently correspond to the Bayes optimal design. Hence both Bopt and MMopt satisfy (P7). Of course the robust designs are not designed to maximize response separation, and as expected, they generally fail to give the Bayes optimal design even in the two-support-point example. However, an interesting property proved theoretically in Appendix A and confirmed by example in Section 4.3 is that the robust designs act to minimize Bayes Risk in the limit as the parameter uncertainty becomes vanishingly small.

[P8]: *Captures main elements of pre-posterior analysis.*

[P9]: *Solves optimal Bayesian classifier problem.*

The idea of pre-posterior analysis is to optimize a “what-if” thought experiment that calculates the Bayesian posterior resulting from “having already implemented” a given candidate experiment design (cf., Raiffa and Schlaifer [41], Lindley [32]). In the current paper, only the Bayesian Optimal Classifier (Bopt) satisfies (P9) since it minimizes Bayes Risk which is a property of the Bayesian posterior. However, MMopt minimizes an overbound on the true Bayes Risk, which acts to indirectly reduce Bayes Risk. Hence MMopt satisfies (P8) and captures important elements of pre-posterior analysis.

It is worth noting that there is a separate literature on Bayesian optimal design (Ryan et al. [42]). Pre-posterior analysis methods from Bayesian optimal design have been applied in the context of PK problems in Merle and Mentre [36]. Interestingly, the second method discussed in [36] can be directly related to MMopt in that it defines informative experiments as those that minimize the covariance of the pre-posterior Bayesian distribution, while MMopt defines informative experiments as those that maximize its mode. Intuitively, both approaches are acting to “hone in” on the parameter values by making the pre-posterior Bayesian distribution “tall” (in the case of MMopt), and “thin” (in the case of [36]).

Generally speaking, Bayesian optimal approaches are promising and may find application to a broad class of PK problems in the future. However for the NP models considered here, such approaches are more complicated than MMopt, and generally require Monte Carlo simulation, multidimensional integration, and/or stochastic analysis which significantly adds to their computation.

9 Conclusions

MMopt is introduced as a novel experiment design method uniquely relevant to nonparametric models. MMopt minimizes an overbound on the Bayes risk, where the overbound is based on a key inequality first proved in [10]. The main idea is to think of individualized therapy with respect to a nonparametric (i.e., discrete) prior, as a classification problem to find the support point that best matches the patient’s behavior. This classification-based approach leads to the important notion of “response separation” as a guiding principle for understanding and optimizing the nonparametric problem. It is interesting that Fisher-based designs completely ignore response separation, being motivated instead by arguments specifically relevant to parametric models. The issue of response separation underscores the essential difference between optimal designs intended for

parametric models versus nonparametric models. Despite this difference, the two approaches were shown to give identical results in the special case where the prior uncertainty becomes vanishingly small. This indicates that the Fisher-based designs may have some value in a restricted local sense when applied to nonparametric models. Bayes Risk minimization has the additional advantage of causing the MAP Bayesian estimate to be more accurate, which is relevant to PK applications where MAP Bayesian control is applied for determining individualized drug doses. Using numerical examples, including a PK voriconazole study, and supporting theoretical analysis in an Appendix and Supplemental File, the usefulness of MMopt is demonstrated and discussed with respect to the ED, EID, and ELD classes of Fisher-Information-based robust designs.

Future work will be directed at comparing MMopt with alternative experimental design approaches, and developing a weighted version of MMopt that incorporates a matrix of misclassification costs (i.e., the cost of misclassifying support point i as j). This latter development will allow the clinician to tailor the experimental design to elicit specific information needed to support a diverse range of clinical goals.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support from NIH grants GM 068968 and HD 070886 is acknowledged.

References

- Andes D, Marchillo K, Stamstad T, Conklin R. In vivo pharmacokinetics and pharmacodynamics of a new triazole, voriconazole, in a murine candidiasis model. *Antimicrobial Agents and Chemotherapy*. 2003; 47(10):3165–3169. [PubMed: 14506026]
- Atkinson, AC., Donev, AN., Tobias, RD. *Optimum Experimental Designs, with SAS*. Oxford University Press; New York: 2007.
- Baek, Y. PhD Thesis. University of Washington; 2006. An Interior Point Approach to Constrained Nonparametric Mixture Models.
- Bayard DS. A Forward Method for Optimal Stochastic Nonlinear and Adaptive Control. *IEEE Transactions on Automatic Control*. Sep; 1991 36(9):1046–1053.
- Bayard DS, Schumitzky A. Implicit Dual Control Based on Particle Filtering and Forward Dynamic Programming. *International Journal of Adaptive Control and Signal Processing*. Mar; 2010 24(3): 155–177. [PubMed: 21132112]
- Bayard, DS., Jelliffe, R., Schumitzky, A., Milman, M., Van Guilder, M. Selected Topics in Mathematical Physics, Prof. R. Vasudevan Memorial Volume. Madras: World Scientific Publishing Co; 1995. Precision drug dosage regimens using multiple model adaptive control: Theory and application to simulated Vancomycin therapy.
- Bayard, DS., Jelliffe, R., Neely, M. Bayes Risk as an Alternative to Fisher Information in Determining Experimental Designs for Nonparametric Models. *PODE 2013 Workshop: Population Optimum Design of Experiments*; Lilly UK. June 15, 2013;
- Bayard, DS., Jelliffe, R., Neely, M. Experiment Design Based on Bayes Risk and Weighted Bayes Risk with Application to Pharmacokinetic Systems. *PODE 2014 Workshop: Population Optimum Design of Experiments*; Roche, Basel. September 11, 2014;
- Bender, CM., Orszag, SA. *Advanced Mathematical Methods for Scientists and Engineers*. New York: McGraw-Hill Book Company; 1978.

10. Blackmore L, Rajamanoharan S, Williams BC. Active estimation for jump Markov linear systems. *IEEE Trans Automatic Control*. Nov; 2008 53(10):2223–2236.
11. Box REP, Lucas HL. Design of experiments in non-linear situations. *Biometrika*. 1959; 46:77–90.
12. Chaloner K, Verdinelli I. Bayesian experimental design: A review. *Statistical Science*. 1995; 10(3): 273–304.
13. D'Argenio DZ. Optimal Sampling Times for Pharmacokinetic Experiments. *J Pharmacokinetics and Biopharmaceutics*. 1981; 9(6):739–756.
14. D'Argenio DZ. Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. *Mathematical Biosciences*. 1990; 99:105–118. [PubMed: 2134510]
15. D'Argenio, DZ., Schumitzky, A., Wang, X. ADAPT 5 User's Guide. Biomedical Simulation Resource, University of Southern California; 2009.
16. Draper NR, Hunter WG. The use of prior distributions in the design of experiments for parameter estimation in non-linear situations. *Biometrika*. 1967; 54(1–2):147–153. [PubMed: 6049531]
17. Duda, RO., Hart, PE., Stork, DG. Pattern Classification. 2. John Wiley & Sons; New York: 2001.
18. Fedorov, VV. Theory of Optimal Experiments. Academic Press; New York: 1972.
19. Fedorov, VV., Leonov, SL. Optimal Design for Nonlinear Response Models. CRC Press; New York: 2013.
20. Gibaldi, M., Perrier, D. Pharmacokinetics. 2. Marcel Dekker, Inc; New York: 1982.
21. Goodwin, GC., Payne, RL. Dynamic System Identification: Experiment Design and Data Analysis. Academic Press; New York: 1977.
22. Hennig S, Nyberg J, Fanta S, Backman JT, Hoppu K, Hooker AC, Karlsson MO. Application of the optimal design approach to improve a pretransplant drug dose finding design for ciclosporin. *J Clin Pharmacol*. 2012 Mar; 52(3):347–60. Epub 2011 May 4. DOI: 10.1177/0091270010397731 [PubMed: 21543664]
23. Hu C, Lovejoy WS, Shafer SL. Comparison of some control strategies for three-compartment PK/PD models. *J Pharmacokinetics and Biopharmaceutics*. 1994; 22(6)
24. Huurneman LJ, Neely M, Veringa A, Prez FD, Ramos-Martin V, Tissing WJ, Alfenaar JWC, Hope W. Pharmacodynamics of voriconazole in children: Further steps along the path to true Individualized therapy. *Antimicrobial Agents and Chemotherapy*. 2016; 60(4):2336–2342. [PubMed: 26833158]
25. 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;
26. Karthaus M, Lehrnbecher T, Lipp HP, Kluge S, Buchheidt D. Therapeutic drug monitoring in the treatment of invasive aspergillosis with voriconazole in cancer patients: an evidence-based approach. *Annals of Hematology*. 2015; 94(4):547–556. [PubMed: 25697592]
27. Katz D, D'Argenio DZ. Discrete approximation of multivariate densities with application to Bayesian estimation. *Computational Statistics & Data analysis*. 1984; 2(1):27–36.
28. Kendall, M., Stuart, A. The Advanced Theory of Statistics. 4. Vol. 2. Charles Griffin & Company; London: 1979.
29. Kistos, CP. Optimal Experimental Design for Non-Linear Models: Theory and Applications. Springer; New York: 2013.
30. Landaw, EM. PhD Dissertation. Los Angeles: Univ. of California; 1980. Optimal Experiment Design for Biologic Compartmental Systems with Applications to Pharmacokinetics.
31. Leary, R., Jelliffe, R., Schumitzky, A., Van Guilder, M. An Adaptive Grid Non-Parametric Approach to Pharmacokinetic and Dynamic (PK/PD) Population Models. Computer-Based Medical Systems, 2001. CBMS 2001. Proceedings. 14th IEEE Symposium on; IEEE; 2001. p. 389–394.
32. Lindley, DV. Bayesian Statistics, A Review. Society for Industrial and Applied Mathematics (SIAM); Philadelphia: 1972.
33. Lindsay B. The Geometry of Mixture Likelihoods: a General Theory. *Ann Statist*. 1983; 11:86–94.

34. Luong ML, Al-Dabbagh M, Groll AH, Racil Z, Nannya Y, Mitsani D, Husain S. Utility of voriconazole therapeutic drug monitoring: a meta-analysis. *Journal of Antimicrobial Chemotherapy*. 2016; 71(7):1786–1799. [PubMed: 27165788]
35. Mallet A. A maximum likelihood estimation method for random coefficient regression models. *Biometrika*. 1986; 73:645–656.
36. Merle Y, Mentre F. Bayesian design criteria: Computation, comparison, and application to a pharmacokinetic and a pharmacodynamic model. *J Pharmacokinetics and Biopharmaceutics*. 1995; 23(1)
37. Neely M, Rushing T, Kovacs A, Jelliffe R, Hoffman J. Voriconazole pharmacokinetics and pharmacodynamics in children. *Clinical Infectious Diseases*. 2010; 50(1):27–36. [PubMed: 19951112]
38. Neely M, Margol A, Fu X, et al. Achieving target voriconazole concentrations more accurately in children and adolescents. *Antimicrobial Agents and Chemotherapy*. 2015; 59:30903097.
39. Pronzato, L., Pazman, A. *Design of Experiments in Nonlinear Models: Asymptotic Normality, Optimality Criteria and Small-Sample Properties*. Springer; New York: 2013.
40. Pronzato L, Walter E. Robust experiment design via stochastic approximation. *Mathematical Biosciences*. 1985; 75:103–120.
41. Raiffa, H., Schlaifer, R. *Applied Statistical Decision Theory*. Wiley Classics Library, John Wiley & Sons (Wiley Classics Library); New York: 2000.
42. Ryan EG, Drovandi CC, McGree JM, Pettitt AN. A review of modern computational algorithms for Bayesian optimal design. *International Statistical Review*. Apr; 2015 84(1):128–154.
43. Rodman, JH., D'Argenio, DZ., Peck, CC. *Analysis of Pharmacokinetic Data for Individualizing Drug Dosage Regimens*. In: Buron, ME, Shaw, LM, Schentag, JJ., Evans, WE., editors. *Applied Pharmacokinetics: Principles of Therapeutic Drug Monitoring*. 4. Spokane, WA: Applied Therapeutics; 2004.
44. Schumitzky, A. Application of stochastic control theory to optimal design of dosage regimens. In: D'Argenio, David Z., editor. *Advanced methods of Pharmacokinetic and Pharmacodynamic Systems Analysis*. Plenum Press; New York: 1991. p. 137–152.
45. Schumitzky. Nonparametric EM Algorithms for estimating prior distributions. *Applied Mathematics and Computation*. Sep; 1991 45(2):143–157.
46. Sheiner LB, Rosenberg B, Melmon KL. Modeling of Individual Pharmacokinetics for Computer-Aided Drug Dosage. *Computers and Biomedical Research*. 1972; 5:441–459.
47. Sheiner LB, Halkin H, Peck C, Rosenberg B, Meimon KL. Improved Computer-Assisted Digoxin Therapy: A Method Using Feedback of Measured Serum Digoxin Concentrations. *Annals of Internal Medicine*. 1975; 82:619–627. [PubMed: 1137256]
48. Sheiner LB, Beal SL. Bayesian Individualization of Pharmacokinetics: Simple Implementation and Comparison with Non-Bayesian Methods. *Journal of Pharmaceutical Sciences*. 1982; 71:1344–1348. [PubMed: 7153881]
49. Silvey, SD. *Optimal Design: An Introduction to the Theory for Parameter Estimation*. Chapman and Hall; London: 1980.
50. Tod M, Rocchisani J-M. Comparison of ED, EID, and API criteria for the robust optimization of sampling times in pharmacokinetics. *J Pharmacokinetics and Biopharmaceutics*. 1997; 25(4)
51. Tod M, Rocchisani J-M. Implementation of OSPOP, an algorithm for the estimation of optimal sampling times in pharmacokinetics by the ED, EID and API criteria. *Computer Methods and Programs in Biomedicine*. Jun; 1996 50(1):13–22. [PubMed: 8835836]
52. Tse E, Bar-Shalom Y, Meier L III. Wide-sense adaptive dual control for nonlinear stochastic systems. *IEEE Transaction on Automatic Control*. 1973; 18(2):98–108.
53. Vajjah P, Duffull SB. A generalisation of T-optimality for discriminating between competing models with an application to pharmacokinetic studies. *Pharmaceutical statistics*. 2012; 11(6):503–510. [PubMed: 23059829]
54. Vozeh S, Berger M, Wenk M, Ritz R, Follath F. Rapid Prediction of Individual Dosage Requirements for Lignocaine. *Clinical Pharmacokinetics*. 1984; 9:353–363.

55. Vozeh S, Steimer J-L. Feedback Control Methods for Drug Dosage Optimisation: Concepts, Classification and Clinical Application. *Clinical Pharmacokinetics*. 1985; 10:457–476. [PubMed: 3905164]
56. Walter E, Pronzato L. Optimal experiment design for nonlinear models subject to large prior uncertainties. *American Journal of Physiology, Regulatory, Integrative and Comparative Physiology*. 1987; 253:R530–R534.
57. BestDose software package. available at <http://www.lapk.org>
58. Supplemental File, 8/13/2016

APPENDIX

A Two-Support-Point Problem with Close Parameters

DEFINITION A.1 (Two-Support-Point Problem)

A general class of two-support-point problems is defined of the form,

$$y(t_k) = \mu(t_k, a) + \sigma_k n_k \quad (1)$$

where $a \in \Omega \triangleq \{a_1, a_2\}$ is a random variable taking on values a_1 and a_2 ; $z(t, a_i)$, $i = 1, 2$ are the support point responses, each assumed to be continuous over the closed interval $t \in \mathcal{T} \triangleq [t_A, t_B]$; $y(t_k)$ is the noisy measurement of $z(t, a)$ taken at discrete time t_k ; $n_k \sim N(0, 1)$ is the measurement noise at time instant t_k ; σ_k scales n_k to a desired level; and the prior support point probabilities are specified as p_1 for a_1 and p_2 for a_2 where $p_1 > 0$, $p_2 > 0$, and $p_1 + p_2 = 1$;

The experiment design problem is to find the set U of n optimal sampling times

$$U = \{t_1, \dots, t_n\} \quad (2)$$

on a specified time interval $t \in \mathcal{T} \triangleq [t_A, t_B]$ that act to best identify the parameter vector $a \in \{a_1, a_2\}$.

DEFINITION A.2 (Restricted Two-Support-Point Problem)

A restricted two-support-point problem is defined by Definition A.1 under the additional conditions that the noise is independent of time, $\sigma_k = \sigma$; there is only one sample to be taken, $n = 1$, $U = \{t_1\}$; and the prior support point probabilities are uniformly distributed, i.e., $p_1 = p_2 = 0.5$. Furthermore, it is assumed that the function $\mu(t, a)$ is sufficiently smooth to admit a first-order Taylor expansion in the vicinity of parameter values a_1 and a_2 for all $t \in \mathcal{T}$.

In what follows, the asymptotic order notation $\mathcal{O}(e^k)$ is used to indicate a term that decreases to zero as e^k . Specifically, $\lim_{e \rightarrow 0} \mathcal{O}(e^k) = 0$ and there exists a constant M such that [9],

$$\lim_{\varepsilon \rightarrow 0} |\mathcal{O}(\varepsilon^k)/\varepsilon^k| = M < \infty \quad (3)$$

Furthermore, the notational dependence of $\mu(t, a)$ on time t will be suppressed and denoted more simply as $\mu(a)$.

THEOREM A.1

The costs associated with the ED,D (evaluated on the mean parameter value $\bar{a} = \frac{1}{2}(a_1 + a_2)$), *EID* and *ELD* optimal designs for the restricted two-support-point problem of Definition A.2 can be written in terms of the response separation $r(t) = |\mu(a_2) - \mu(a_1)|$ for sufficiently small parameter uncertainty $a = a_2 - a_1$ as,

$$J_{ED} = .5|M(a_1)| + .5|M(a_2)| = \frac{r(t)^2}{\sigma^2 \Delta a^2} + \mathcal{O}(\Delta a) \quad (4)$$

$$J_D = |M(\bar{a})| = \frac{r(t)^2}{\sigma^2 \Delta a^2} + \mathcal{O}(\Delta a) \quad (5)$$

$$J_{EID} = .5|M(a_1)|^{-1} + .5|M(a_2)|^{-1} = \frac{\sigma^2 \Delta a^2}{r(t)^2} (1 + \mathcal{O}(\Delta a)) \quad (6)$$

$$J_{ELD} = .5 \ln|M(a_1)| + .5 \ln|M(a_2)| = .5 \ln \left(\frac{r(t)^4}{\sigma^4 \Delta a^4} + \mathcal{O}(\Delta a) \right) \quad (7)$$

Proof of equation (4)—Consider the response $\mu(a)$ as a function of the parameter a . Let $\mu(a_2)$ be expressed in terms of the Taylor expansion about a_1

$$\mu(a_2) = \mu(a_1) + \left. \frac{\partial \mu}{\partial a} \right|_{a_1} (a_2 - a_1) + \mathcal{O}(a_2 - a_1)^2 \quad (8)$$

Letting $a \triangleq a_2 - a_1$ yields upon rearranging,

$$\mu(a_2) - \mu(a_1) = \left. \frac{\partial \mu}{\partial a} \right|_{a_1} \Delta a + \mathcal{O}(\Delta a)^2 \quad (9)$$

Squaring both sides of (9) gives

$$(\mu(a_2) - \mu(a_1))^2 = (\Delta a)^2 \left(\frac{\partial \mu}{\partial a} \Big|_{a_1} \right)^2 + \mathcal{O}(\Delta a)^3 \quad (10)$$

Similarly, let $\mu(a_1)$ be expressed in terms of the Taylor expansion about a_2

$$\mu(a_1) = \mu(a_2) + \frac{\partial \mu}{\partial a} \Big|_{a_2} (a_1 - a_2) + \mathcal{O}(\Delta a)^2 \quad (11)$$

which can be rearranged as

$$\mu(a_2) - \mu(a_1) = \frac{\partial \mu}{\partial a} \Big|_{a_2} \Delta a + \mathcal{O}(\Delta a)^2 \quad (12)$$

Squaring both sides of (12) yields

$$(\mu(a_2) - \mu(a_1))^2 = (\Delta a)^2 \left(\frac{\partial \mu}{\partial a} \Big|_{a_2} \right)^2 + \mathcal{O}(\Delta a)^3 \quad (13)$$

Adding one half of (10) to one half of (13) gives

$$(\mu(a_2) - \mu(a_1))^2 = (\Delta a)^2 \left(\frac{1}{2} \left(\frac{\partial \mu}{\partial a} \Big|_{a_1} \right)^2 + \frac{1}{2} \left(\frac{\partial \mu}{\partial a} \Big|_{a_2} \right)^2 \right) + \mathcal{O}(\Delta a)^3 \quad (14)$$

Dividing both sides by $\sigma^2(\Delta a)^2$ gives,

$$\frac{1}{2\sigma^2} \left(\frac{\partial \mu}{\partial a} \Big|_{a_1} \right)^2 + \frac{1}{2\sigma^2} \left(\frac{\partial \mu}{\partial a} \Big|_{a_2} \right)^2 = \frac{(\mu(a_2) - \mu(a_1))^2}{\sigma^2 \Delta a^2} + \mathcal{O}(\Delta a) \quad (15)$$

Proof of (5)—Combining relations $\bar{a} = \frac{1}{2}(a_1 + a_2)$ and $\Delta a = a_2 - a_1$, gives

$$\bar{a} = a_1 + \frac{1}{2} \Delta a \quad (16)$$

$$\bar{a} = a_2 - \frac{1}{2}\Delta a \quad (17)$$

Let $\mu(a_2)$ be expressed in terms of a Taylor expansion of $\mu(a)$ about

$$\mu(a_2) = \mu(\bar{a}) + \left. \frac{\partial \mu}{\partial a} \right|_{\bar{a}} (a_2 - \bar{a}) + \mathcal{O}(a_2 - \bar{a})^2 \quad (18)$$

Rearranging (17) gives $a_2 - \bar{a} = \frac{1}{2}\Delta a$ which is substituted into (18) to give

$$\mu(a_2) = \mu(\bar{a}) + \left. \frac{\partial \mu}{\partial a} \right|_{\bar{a}} \left(\frac{1}{2}\Delta a \right) + \mathcal{O}(\Delta a)^2 \quad (19)$$

Similarly, let $\mu(a_1)$ be expressed in terms of a Taylor expansion of $\mu(a)$ about

$$\mu(a_1) = \mu(\bar{a}) + \left. \frac{\partial \mu}{\partial a} \right|_{\bar{a}} (a_1 - \bar{a}) + \mathcal{O}(a_1 - \bar{a})^2 \quad (20)$$

Rearranging (16) gives $a_1 - \bar{a} = -\frac{1}{2}\Delta a$ which is substituted into (20) to give

$$\mu(a_1) = \mu(\bar{a}) - \left. \frac{\partial \mu}{\partial a} \right|_{\bar{a}} \left(\frac{1}{2}\Delta a \right) + \mathcal{O}(\Delta a)^2 \quad (21)$$

Subtracting (21) from (19) and rearranging gives

$$\mu(a_2) - \mu(a_1) = \left. \frac{\partial \mu}{\partial a} \right|_{\bar{a}} \Delta a + \mathcal{O}(\Delta a)^2 \quad (22)$$

Squaring both sides of (22) and dividing by $\sigma^2(\Delta a)^2$ gives upon rearranging,

$$\frac{1}{\sigma^2} \left(\left. \frac{\partial \mu}{\partial a} \right|_{\bar{a}} \right)^2 = \frac{(\mu(a_2) - \mu(a_1))^2}{\sigma^2 \Delta a^2} + \mathcal{O}(\Delta a) \quad (23)$$

Proof of (6)—Reciprocating both sides of equation (10), using the relation

$\frac{1}{1+\varepsilon} = 1 - \varepsilon + \mathcal{O}(\varepsilon^2)$ for small ε , and rearranging gives,

$$\frac{\Delta a^2}{(\mu(a_2) - \mu(a_1))^2} = \left(\frac{\partial \mu}{\partial a} \Big|_{a_1} \right)^{-2} (1 - \mathcal{O}(\Delta a)) \quad (24)$$

Likewise, reciprocating both sides of equation (13) and rearranging gives,

$$\frac{\Delta a^2}{(\mu(a_2) - \mu(a_1))^2} = \left(\frac{\partial \mu}{\partial a} \Big|_{a_2} \right)^{-2} (1 - \mathcal{O}(\Delta a)) \quad (25)$$

Adding (24) and (25) and multiplying both sides by $\sigma^2/2$ gives upon rearranging

$$.5 \left(\frac{1}{\sigma^2} \frac{\partial \mu}{\partial a} \Big|_{a_1} \right)^{-2} + .5 \left(\frac{1}{\sigma^2} \frac{\partial \mu}{\partial a} \Big|_{a_2} \right)^{-2} = \frac{\sigma^2 \Delta a^2}{(\mu(a_2) - \mu(a_1))^2} (1 + \mathcal{O}(\Delta a)) \quad (26)$$

Proof of (7)—Multiplying (10) and (13) yields,

$$(\mu(a_2) - \mu(a_1))^4 = \Delta a^4 \left(\frac{\partial \mu}{\partial a} \Big|_{a_1} \right)^2 \left(\frac{\partial \mu}{\partial a} \Big|_{a_2} \right)^2 + \mathcal{O}(\Delta a)^5 \quad (27)$$

Dividing both sides by $\sigma^4 \Delta a^4$ gives upon rearranging

$$\left(\frac{1}{\sigma^2} \frac{\partial \mu}{\partial a} \Big|_{a_1} \right)^2 \left(\frac{1}{\sigma^2} \frac{\partial \mu}{\partial a} \Big|_{a_2} \right)^2 = \frac{(\mu(a_2) - \mu(a_1))^4}{\sigma^4 \Delta a^4} + \mathcal{O}(\Delta a) \quad (28)$$

Taking $\frac{1}{2} \ln(\cdot)$ of both sides gives

$$.5 \ln \left(\frac{1}{\sigma^2} \frac{\partial \mu}{\partial a} \Big|_{a_1} \right)^2 + .5 \ln \left(\frac{1}{\sigma^2} \frac{\partial \mu}{\partial a} \Big|_{a_2} \right)^2 = .5 \ln \left(\frac{(\mu(a_2) - \mu(a_1))^4}{\sigma^4 \Delta a^4} + \mathcal{O}(\Delta a) \right) \quad (29)$$

COROLLARY A.1

As the parameter uncertainty $a = a_2 - a_1$ becomes small in the restricted two-support-point problem of Definition A.2, the objective functions J_{ED} , J_D and J_{ELD} approach a monotonically increasing function of the response separation $r(t)$, while J_{EID} approaches a monotonically decreasing function of the response separation $r(t)$.

Proof—The result follows by the properties of relations (4)(5)(6)(7) as a becomes small, noting that functions $1/x$ and $\ln(x)$ are monotonically decreasing and increasing in x , respectively.

Corollary A.1 indicates that as the parameter uncertainty $a = a_2 - a_1$ becomes small in the restricted two-support-point problem of Definition A.2, the objectives of maximizing J_{ED} , J_D , J_{ELD} and minimizing J_{EID} asymptotically approach the single common objective of maximizing the response separation $r(t) = |\mu(a_2, t) - \mu(a_1, t)|$ over time $t \in \mathcal{T}$. These asymptotic properties are examined numerically in the two-support-point problem of Section 4.3.

Figure 1. Responses for the two-support-point example. The Dopt design is indicated for each support point taken separately.

Figure 2.

The ED measure (solid) is equal to one half the Dopt measure for the slow a_2 support point $|M(a_2)|$ (dot), and one half the Dopt measure for the fast a_1 support point $|M(a_1)|$ (dash). Since the slow support point measure dominates the fast support point, the ED sample at $t = 3.9991$ is nearly identical to that of the slow a_2 model's Dopt sample at $t = 4$, but has been drawn slightly lower by the weighted influence of the fast a_1 model's Dopt sample at $t = 0.6667$.

Figure 3.
Depiction of response separation $r(t)$ between responses $\mu(t, a_1)$ and $\mu(t, a_2)$.

Figure 4.

Bayes Risk is depicted as the grey area between curves, corresponding to the sum of Type I and Type II errors. Bayes Risk decreases as response separation $r(t)$ increases, which “pulls apart” the two Gaussian distributions and shrinks the grey area between them.

Figure 5.
Responses for two-support-point example with closely-spaced parameters.

Figure 6.
Responses for four-support-point example.

Figure 7.
Ten support point responses to dose input $d(t)$.

Figure 8. Bayes Risk computed using 1×10^6 Monte Carlo runs, showing the 1-sample Bayesian design at $B_{opt}=4.25$ hr with an optimal value of $prob=0.5474$.

Figure 9.

Oral dosing responses of time versus concentration, simulated from a voriconazole nonparametric model with 125 support points. Patients are given loading doses at 0 and 12 hours, and a maintenance dose at 24 hours. MMopt samples times are determined on a grid constrained to 0.5 hour intervals. The two optimal sample times (MMopt2) are shown in dashed lines at 25 and 36 hours. The one optimal sample time (MMopt1) is indicated by the dash-dot line at 35 hours.

Table 4.1

Summary of optimal sampling time designs for two-support-point example.

Design Metric	Sample Time (hr)	Bayes Risk (prob)
Bopt	1.4334	0.1659
MMopt	1.4334	0.1659
ED	3.9991	0.2712
EID	0.7498	0.2003
ELD	1.1429	0.1705

Table 4.2

Summary of optimal sampling time designs for two-support-point example with close parameters. All designs give nearly the same optimal sample time at 0.95 hours.

Design Metric	Sample Time (hr)	Bayes Risk (prob)
Bayes Optimal	0.9530	0.4767
MMopt	0.9530	0.4767
ED	0.9570	0.4767
EID	0.9480	0.4767
ELD	0.9520	0.4767
D-Optimal	0.9524	0.4767

Table 5.3Values for model parameters a and b .

Index	a	b
1	2	2.625
2	1	0.6
3	0.7	0.6
4	0.5	0.6

Table 5.4

Summary of optimal 2-sample designs for four-support-point example.

Design Metric	2-Sample Times (hr)		Bayes Risk (prob)	99% Bayes Risk Conf Bound
MMopt	0.45	1.4	0.32839	± 0.00070
ED	0	0.8	0.37028	± 0.00070
EID	0	1	0.36044	± 0.00072
ELD	0	0.95	0.36234	± 0.00072

Table 5.5

Summary of optimal 3-sample designs for four support-point example.

Design Metric	3-Sample Times (hr)			Bayes Risk (prob)	99% Bayes Risk Conf Bound
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
ELD	0	0.85	1.05	0.30990	± 0.00069

Table 6.1

Values of V and K in a pharmacokinetic model with 10 support points.

#	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
Mean	0.0820	102.3852
STD	0.0311	9.8615

Table 6.2

Results for optimal 1, 2 and 3 sample designs applied to PK estimation.

Design Metric	Samples (hr)			Bayes Risk (prob)	99% conf (prob)
	<i>1-Sample Design</i>				
Bopt	4.25			0.5474	±0.0015
MMopt	4.25			0.5474	±0.0015
	<i>2-Sample Design</i>				
MMopt	1	9.5		0.2947	±0.0014
ED	1	24		0.3272	±0.0014
	<i>3-Sample Design</i>				
MMopt	1	1	10.5	0.2325	±0.0013
ED	1	1	24	0.2617	±0.0013

Table 7.1

Bias and Imprecision of predictions from Bayesian posteriors based on two optimal samples (MMopt2) or one (MMopt1) optimal sample compared to posteriors based on hourly samples (Full). The MMopt2 samples are taken 1 hour after and just before the first and second oral maintenance voriconazole doses, respectively. The MMopt1 sample is taken 1 hour before the second oral maintenance voriconazole dose.

	Full	MMopt2	MMopt1
		Bias (MMopt minus Full)	
AUC mg*h/L	18.53	-0.82 (-3.18 to 1.55), P=0.49	-1.86 (-4.13 to 0.40), P=0.11
Peak mg/L	2.58	0.37 (0.16 to 0.58), P< 0.001	0.42 (0.04 to 0.79), P=0.03
Trough mg/L	0.8	0.01 (-0.16 to 0.18), P=0.87	-0.09 (-0.20 to 0.03), P=0.12
		95th Percentile Imprecision (MMopt minus Full)	
AUC mg*h/L		11.5	10.75
Peak mg/L		1.13	2.55
Trough mg/L		0.93	0.38

Table 8.2

Properties of ED, EID, ELD, MMopt, and Bopt designs, where Y stands for “yes,” and N stands for “no”.

#	PROPERTY	ED	EID	ELD	MMopt	Bopt
P1	Invariance under regular linear reparametrization	Y	Y	Y	Y	Y
P2	Invariance under regular nonlinear reparametrization	N	N	Y	Y	Y
P3	Avoids dependence on estimator asymptotic properties	N	N	N	Y	Y
P4	Avoids knowing true parameters in nonlinear problems	Y	Y	Y	Y	Y
P5	Allows taking fewer than p samples	N	N	N	Y	Y
P6	Can handle heterogeneous model structures	N	N	N	Y	Y
P7	Gives known optimal solution to 2-support point example	N	N	N	Y	Y
P8	Captures main elements of pre-posterior analysis	N	N	N	Y	Y
P9	Solves optimal Bayesian classifier problem	N	N	N	N	Y