The Canadian Journal of Statistics
Vol. 37, No. 4, 2009, Pages 495–513
La revue canadienne de statistique

Bayesian optimal design for changepoint problems

Juli ATHERTON[1]*, Benoit CHARBONNEAU[2], David B. WOLFSON[3], Lawrence JOSEPH[4], Xiaojie ZHOU[5] and Alain C. VANDAL[3, 6]

[1]*Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada H3A 1A2*
[2]*Department of Mathematics, Duke University, Durham, NC 27708, USA*
[3]*Department of Mathematics and Statistics, McGill University, Montréal, Québec, Canada H3A 2K6*
[4]*Division of Clinical Epidemiology, McGill University Health Center, Montréal, Québec, Canada H3A 1A1*
[5]*The Proctor & Gamble Company, Mason, OH 45040, USA*
[6]*Center for Clinical Epidemiology and Community Studies, Sir Mortimer B. Davis Jewish General Hospital, Montréal, Québec H3T IE2*

Key words and phrases: Bayesian optimal design; changepoint problems; design measure; estimation; testing.

MSC 2000: Primary 62K05; secondary 62C10.

Abstract: We investigate Bayesian optimal designs for changepoint problems. We find robust optimal designs which allow for arbitrary distributions before and after the change, arbitrary prior densities on the parameters before and after the change, and any log-concave prior density on the changepoint. We define a new design measure for Bayesian optimal design problems as a means of finding the optimal design. Our results apply to any design criterion function concave in the design measure. We illustrate our results by finding the optimal design in a problem motivated by a previous clinical trial. *The Canadian Journal of Statistics* 37: 495–513; 2009 © 2009 Statistical Society of Canada

Résumé: Nous considérons les plans d'expérience bayésiens optimaux pour les problèmes de point de rupture. Nous obtenons un plan d'expérience robuste sous des lois arbitraires avant et aprés le point de rupture ainsi que pour toute loi a priori du point de rupture dont la densité est log-concave. Nous définissons une nouvelle mesure sur l'espace des plans d'expérience bayésiens nous permettant d'y déterminer le plan d'expérience optimal. Nos résultats s'appliquent à toute fonction critère qui soit concave sur le domaine de la mesure des plans d'expérience. Nous illustrons nos résultats en déterminant le plan d'expérience optimal pour un problème suggéré par une étude clinique déjà complétée. *La revue canadienne de statistique* 37: 495–513; 2009 © 2009 Société statistique du Canada

1. INTRODUCTION

Suppose that a sequence of observations (possibly multivariate) is taken on some interval $[0, T]$ of the real axis. Let the distribution of this sequence be parametrically defined and suppose that the distribution changes at some point, called a changepoint, to a second distribution for the remaining observations. If the change occurs at T then by convention no change is said to occur. There is a large literature on inference for data arising in such fixed sample size changepoint

This article is based in part on Juli Atherton's doctoral dissertation, which was selected as a co-winner of the 2007 Pierre Robillard Award of the Statistical Society of Canada. The thesis supervisors were David B. Wolfson and Alain C. Vandal.

* *Author to whom correspondence may be addressed.*
E-mail: juli.atherton@mcgill.ca

settings (see, e.g., Chen & Gupta, 2000). What renders changepoint problems challenging is that the changepoint is unknown.

The literature on changepoint problems appears to be exclusively devoted to data that have already been collected, so that to our knowledge this paper discusses optimal designs for changepoint problems for the first time. We concentrate on solving the so-called single-path design problem, as stage 1 in a two-stage program, where main ideas are developed through a single-path and then extended at stage 2 to a multi-path setting. Multi-path optimal designs may be used in the planning of clinical trials, for example, in which data are to be collected on a group of subjects, each providing their own path with their own changepoint. In the simpler scenario, in which the paths of all subjects have roughly the same changepoint, the optimal designs presented here may be used with no modification.

Most often, fixed sample size changepoint inference has been done retrospectively on data collected at regular spacings throughout an interval (of time or space). Further, an assumption is usually made that the point of change, τ, can only occur at one of the observations where data have been collected. These assumptions deserve some discussion.

In considering optimal designs, where data are still to be collected, in principle there may be no reason to restrict observations to be at regular intervals. Nevertheless, there are two reasons for imposing certain restrictions on the sampling scheme. With no restriction, for the problems that we discuss, the optimal designs will place multiple observations on the same subject at the same instant; such designs are clearly impossible to implement. Indeed practical considerations will almost always demand that observations be taken at least $d > 0$ units apart. Further, if d is sufficiently large, then it may be justifiable to assume that within-path observations are conditionally independent, thus leading to much simpler models. We comment further on this in Section 2.1 and in our concluding remarks. We therefore constrain our sampling points to be least d units apart.

The common changepoint assumption that a change can occur only at one of the observation points, is clearly restrictive and does not make sense from an optimal design point of view since the possible changepoint locations of the model would change as one optimized over the design space. In deriving our optimal designs we allow the change to occur anywhere in the sampling interval. Our purpose in this paper is to present optimal sampling schemes for a single-path of observations that could undergo a change in some interval $[0, T]$. Without loss of generality we may regard $[0, T]$ as representing a time interval.

Formally,

(i) We shall assume that a covariance stationary stochastic process $\{y(t), t \in [0, T]\}$, undergoes a change in its parameter at some unknown time point $\tau \in [0, T]$. This assumption includes the scenario in which there is no change in the process on the interval; by convention this occurs when the changepoint, $\tau = T$.

(ii) We shall assume that for all observation points, t_1 and t_2, conditional on the process parameters, and τ, the observations $y(t_1)$ and $y(t_2)$ are independent.

(iii) We shall assume that for all observation points, $t_i < t_j, t_j - t_i \geq d$, for some pre-specified d.

Bischoff (1989) and Bischoff & Miller (2000), find asymptotic optimal designs in a frequentist regression setting when testing for a change at a known location. The complexity in the biphasic regression problem that results from an *unknown* changepoint is considerable and optimal design in such a setting remains an open problem. In this paper we trade the greater generality of the biphasic regression model with a *known* changepoint for a simpler model with an *unknown* changepoint. We take a Bayesian approach which, conveniently, allows us to introduce a design measure (see Section 2) as both a natural consequence of this paradigm and as a mathematical device.

Often Bayesian design criterion functions are described in terms of their analogous frequentist criterion functions, as reviewed by Chaloner & Verdinelli (1995). DasGupta (1996) and Clyde (2001) also provide comprehensive reviews of Bayesian optimal design.

Our design criterion functions are based on Bayesian decision theory. Specifically, we use a Bayes risk based on a generalized 0–1 loss (see Felsenstein, 1990; Blackmore & Williams, 2005) as well as the Spezzaferri criterion (see Spezzaferri, 1988) to find designs optimal for testing for a change, and a Bayes risk based on squared error loss, to find designs optimal for estimating the parameters before and after τ. As discussed in Clyde & Chaloner (1996), these criterion functions can be combined. Both Chaloner & Verdinelli (1995) and DasGupta (1996) discuss optimal designs that are robust. Here, our designs are robust in the sense that: (i) they apply to any distribution of the data, (ii) they apply to any prior distribution for the parameters before and after the change, and (iii) they apply to all log-concave prior densities for the changepoint.

We introduce a novel design measure, π, constructed by combining a prior density with the design. Our design measure is the lynchpin of our discussion.

In Section 2, we present a motivating example, which will thread its way through the rest of the paper. In Section 3, we present the single-path changepoint problem formally. In Section 4, we define the design space $\mathbb{X}^n$, the design measure π, and the mapping G from the design space to the space of design measures $G(\mathbb{X}^n)$. We discuss the shape of the design measure space $G(\mathbb{X}^n)$ in Section 5.

In Section 6, we find optimal designs for design criterion functions concave in the design measure π. In Section 7, we prove that certain standard design criterion functions for testing and estimation are concave in π, thus enabling us to invoke the results of Section 6 to derive optimal designs. Section 8 contains some concluding remarks. Many of the technical details appear in the Appendix and the listed technical reports Atherton & Wolfson (2009), Atherton, Charbonneau & Wolfson (2009), and Atherton (2009b).

2. BLOOD PRESSURE STUDY

In this section we present an illustrative example which we shall refer to repeatedly throughout the remainder of the paper.

Example 1. Lyle et al. (1987) present the results of a study to investigate the potential of a dietary supplement of calcium to lower blood pressure. Four weekly baseline blood pressure readings were recorded for each of the 75 subjects in the trial. Thirty-seven were then given a diet containing the calcium supplement and 38 a placebo. All were followed for a further 12 weeks. After the 4th week of the trial blood pressure readings were bi-weekly. Each blood pressure observation resulted in four readings: supine systolic, seated systolic, supine diastolic, and seated diastolic.

Joseph et al. (1996) re-analyzed the data from Lyle et al. (1987), using a multi-path changepoint model. The main motivations for doing so were (1) to account for the unknown delay in the dietary effect and (2) this delay would be likely be different for each subject on the diet. In fact, a fraction of people might not respond at all. For example, for supine systolic blood pressure it was estimated that approximately 30% of subjects receiving calcium experienced a change at the 12th week (i.e., 8 weeks after starting supplement) and 40% had no change during the study period.

The above example suggests that sharper inference might have been made had optimal designs been used to collect the data. Specific goals would be, to estimate (i) the changes in blood pressures for each patient as well as the overall change, (ii) the proportion of subjects whose blood pressure would change right after administration of the calcium supplement, and (iii) the proportion of subjects who would not benefit from the calcium supplement.

The optimal designs required to address this multi-path problem are the topic of a later paper. Here we focus on a single subject and consider the equivalent of an "n-of-one" trial where the goals would be to (i) estimate the magnitude of the treatment effect (equivalently, estimate the means before and after the change), (ii) test if a change occurs in the 2 weeks after administration of the drug, and (iii) test if a change occurs at all. It is important to note that the single-path optimal design methods and solutions are critical to solving the multi-path case. They pave the way for the multi-path case, which introduces additional complexities, by providing a methodological framework.

In Lyle et al. (1987), both practical considerations and those relating to the absorbtion of calcium, dictated that the blood pressure observations be taken 1 week apart prior to the treatment and 2 weeks apart after the start of treatment. Under this sampling scheme it was reasonable to assume the readings were, for each subject, conditionally independent given the changepoint and the other model parameters. In constructing our optimal design we make the assumption that observations taken at least $d = 1$ week apart, are conditionally independent. Unlike Joseph et al. (1996), we do not assume that any change, if it occurs, could only occur at one of the design points.

3. THE MODEL

The Bayesian single-path changepoint model consists of a likelihood function for n observations $\mathbf{y} = (y_1, \ldots, y_n)$, design points $x = (x_1, \ldots, x_n)$ at which the observations are taken, prior distributions for the parameters θ_1 and θ_2 before and after the change, and for the changepoint τ. For a changepoint, $(x_k \leq \tau < x_{k+1})$, let the k observations taken before the change be $\mathbf{y_1} = (y_1, \ldots, y_k)$ and the $n - k$ observations taken after the change be $\mathbf{y_2} = (y_{k+1}, \ldots, y_n)$. We further assume that the single-path joint conditional density $f(\mathbf{y}|\theta_1, \theta_2, \tau)$, has marginal densities $f(\mathbf{y_1}|\theta_1)$ and $f(\mathbf{y_2}|\theta_2)$ for observations before and after the change, respectively. We allow the distributions of the data $\mathbf{y}$ and the parameters θ_1 and θ_2 to be either continuous, discrete, or mixed.

Conditional dependence between observations is allowed as long as this dependence is not a function of the distance between observations. One tractable situation, however, where such a dependency arises is in the multi-path problem mentioned in the introduction in which all paths have a common changepoint (see Atherton & Wolfson, 2009).

The likelihood when the observations are conditionally independent given θ_1, θ_2, and τ, has the form

$$f(\mathbf{y}|\theta_1, \theta_2, \tau) = \prod_{x_i \leq \tau} f(y_i|\theta_1) \prod_{x_i > \tau} f(y_i|\theta_2).$$

This is the usual changepoint model except we allow the change to occur anywhere in the observation interval $[0, T]$. To complete the model, we incorporate the prior $f(\theta_1, \theta_2, \tau)$, by assuming that τ is independent of θ_1 and θ_2. Letting $0 \leq p_T < 1$ be the probability of a change at T, we represent the prior density for the changepoint as $(1 - p_T)g(\tau) + p_T I_{(\tau=T)}$, where $g(\tau)$ is a continuous density in $[0, T]$ and I is the indicator function. To simplify our notation we denote $(1 - p_T)g$ by $\bar{g}$. In the case of no mass at T, p_T is zero and the prior density reduces to g. Thus $f(\theta_1, \theta_2, \tau) = f(\theta_1, \theta_2)(\bar{g}(\tau) + p_T I_{(\tau=T)})$.

4. THE DESIGN SPACE $\mathbb{X}^n$, THE DESIGN MEASURE $\boldsymbol{\pi}$, AND THE MAPPING G

The set of allowable designs consists of all the design vectors $x = (x_1, \ldots, x_n)$ such that $x \in [0, T]^n$ with $x_1 \geq 0$, $x_{i-1} + d \leq x_i$ for $2 \leq i \leq n$ and $x_n \leq T$. For each experiment d is selected based on practical considerations and/or so that observations a distance d or more apart can be

assumed to be roughly conditionally independent given θ_1, θ_2, and τ. We denote the set of all allowable designs by $\mathbb{X}^n$, where n is the number of observations that are taken. We refer to the set of all the allowable designs $\mathbb{X}^n$ as the *design space*.

It is easily seen that the design space, $\mathbb{X}^n$ forms a simplex. Let V be the vertex set of $\mathbb{X}^n$. The designs in V are listed as $(0, d, \ldots, (n-1)d)$, $(0, d, \ldots, (n-2)d, T), \ldots, (T-(n-1)d, T-(n-2)d, \ldots, T)$. There are $n+1$ such designs in V and we label them v_0 to v_n, respectively. Thus,

$$V = \{v_0, \ldots, v_n\}. \tag{1}$$

For all experiments we assume that $T > (n-1)d$, so that the n observations fit into the observation interval $[0, T]$ while maintaining at least a distance d apart.

A direct approach to the optimal design problem is to attempt to minimize the design criterion function over the design space itself. However, this approach is difficult since the shape of the design criterion function depends on the prior distribution of the changepoint. Furthermore, unless this prior is uniform with support $[0, T]$ the design criterion function will be multimodal. We overcome this difficulty by introducing a new type of Bayesian design measure. The design criterion function, rewritten in terms of the design measure, is always a concave function of the *design measure* regardless of the shape of the prior distribution for the changepoint.

Formally, let τ_x be a random variable representing the number of observations taken before or at the change. Hence, the event $\{\tau_x = k\}$ is the event that the change occurs at x_k or between the design points x_k and x_{k+1}. If the change occurs at T this is equivalent to no change and $\tau_x = n$. Since there are n design points in the observation interval $[0, T]$ there are $n+1$ intervals in which the changepoint can fall. For notational convenience, ignoring the dummy variable of integration, for prior density $\bar{g} + p_T$ on the changepoint, τ_x has mass function with probabilities

$$\pi_k = P(\tau_x = k) = P(x_k \le \tau < x_{k+1}) = \int_{x_k}^{x_{k+1}} \bar{g},$$

for $(0 \le k \le n-1)$. For $k = n$,

$$\pi_n = P(\tau_x = n) = P(x_n \le \tau) = \int_{x_n}^{T} \bar{g} + p_T.$$

We shall refer to the vector $\pi = (\pi_0, \ldots, \pi_n)$, as the *design measure*. Once the prior distribution for the changepoint has been selected, we consider it to be fixed and the design measure π is then a function of only the design x.

Since the design measure π is discrete with $n+1$ support points, its components are the barycentric coordinates of *any* n-dimensional simplex. We define the specific simplex S^n, to be the convex hull of the standard basis $\{e_1, \ldots, e_n\}$ and the zero vector e_0. We may therefore let the components of π be the barycentric coordinates of S^n so that $\pi_1, \ldots, \pi_n$ are the Cartesian coordinates in n-dimensional space. For $x \in \mathbb{X}^n$, define

$$G(x) = (\pi_1(x), \ldots, \pi_n(x)) = \left(\int_{x_1}^{x_2} \bar{g}, \int_{x_2}^{x_3} \bar{g}, \ldots, \int_{x_{n-1}}^{x_n} \bar{g}, \int_{x_n}^{T} \bar{g} + p_T \right).$$

Then G is a mapping from $\mathbb{X}^n$ to the Cartesian coordinates of $S^n \subset \mathbb{R}^n$. Obviously, if $\bar{g}$ is greater than zero everywhere on $[0, T]$ then the mapping G is one-to-one. In cases where $\bar{g}$ is equal zero

on a sub-interval G is not one-to-one. In Example 1, the changepoint prior with support restricted to the post-treatment sub-interval of $[0, T]$, induces such a case.

4.1. Main Steps in the Sequel

Our strategy is first to establish a general theory of optimal design for changepoint problems. This theory applies to any setting where the design criterion function is a concave function of the design measure.

The main steps are as follows:

(i) The design criterion function is, by definition, a function of the design x. We first re-express it as a function of the design measure π, which leads to a simpler convex optimization problem.
(ii) We then minimize the design measure criterion function over the set of allowable design *measures*, $G(\mathbb{X}^n)$.
(iii) We next return to the problem of original interest, that is, of ascertaining the optimal design itself. This is carried out by associating the optimal design measure with an optimal design, or at worst, in some cases when G is not one-to-one, with a small class of equivalent optimal designs.
(iv) Finally, we focus on the two main classes of changepoint problems: testing for a change and estimation of the parameters before and after the change. We show that the most common design criterion functions (Bayes risk based, on squared error loss for estimation, and Bayes risk based on a generalized 0–1 loss and the Spezzaferri design criterion function, for testing) are convex functions of the design measure. We may therefore apply the general theory of steps (i), (ii), and (iii).

5. THE SHAPE OF $G(\mathbb{X}^n)$

It is important to investigate the shape of $G(\mathbb{X}^n)$, since although $G(\mathbb{X}^n)$ is not a simplex, the hope is that it lies entirely within a simplex, and contains the vertices of the simplex. For such sets the concavity of the design criterion function guarantees that the optimal design measure is at one of these vertices. Figure 1a–c illustrates how G depends on the prior density for the changepoint τ when there are two design points. Only in Figure 1a which corresponds to a truncated normal prior, does $G(\mathbb{X}^2)$ remain inside the simplex. In Figure 1b and c, which correspond to a bimodal mixture of normal densities and a mixture of a uniform and gamma density, respectively, $G(\mathbb{X}^2)$ leaves the simplex.

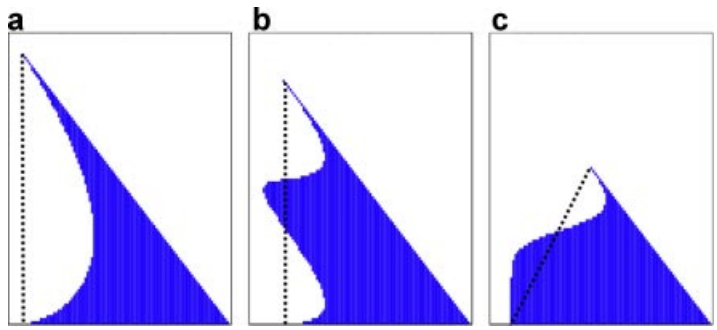

FIGURE 1: (a)–(c) represent $G(\mathbb{X}^2)$ for (a) a truncated normal prior density (b) a bimodal mixture of normal densities, and (c) a mixture of a uniform density and a gamma density. The observation interval is taken to be $T = 10$ and the minimum distance between observations is $d = 2$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

As we shall see in Theorem 1, it is log-concavity that is fundamental, not unimodality; For Figure 1c the prior density is unimodal but not log-concave. Since, the class of log-concave distributions is rich (Bagnoli & Bergström, 1989), the restriction is not great.

We prove Theorem 1 for the case where G is one-to-one. Additions to the proofs for when G is not one-to-one are given in Atherton, Charbonneau & Wolfson (2009).

Theorem 1. *Let $\triangle = \text{Conv}(G(V))$ be the convex hull of $G(V)$. If in the density $\bar{g} + p_T$, the density $\bar{g}$ is log-concave and everywhere greater than zero then $G(\mathbb{X}^n) \subset \triangle$.*

Proof of Theorem 1. The proof of Theorem 1 is given in Appendix A.1 along with a basic lemma.

The converse of Theorem 1 does not hold. It is not true that $G(\mathbb{X}^n) \subset \triangle$ implies that $\bar{g}$ is log-concave, since there are non-log-concave $\bar{g}$, such that $G(\mathbb{X}^n) \subset \triangle$. ■

6. OPTIMAL DESIGNS FOR GENERAL DESIGN CRITERION FUNCTIONS CONCAVE IN π

The discussion in this section applies to minimizing concave design criterion functions. Of course, the same results hold when the goal is to maximize a convex design criterion function. Theorem 1 and its analogues for cases where G is not one-to-one underpin Theorems 2 and 3, which show that the optimal designs are located in the vertex set, V, defined by (1). This means that the optimal design problem is, at worst, reduced to a numerical search over a set of possible designs. Theorems 2 and 3 differ in that the former theorem places no constraint (apart from the d-apart restriction) on the location of the design points while the latter theorem places an additional constraint. As we shall see in Sections 7.2 and 7.3 the designs in Theorem 2 are appropriate for estimation and the designs in Theorem 3 are appropriate for testing.

Theorem 2. *Suppose the density $\bar{g} + p_T$, is constructed from a density g that is log-concave with support $[0, T]$. Suppose the observations may be taken anywhere on $[0, T]$ provided they are at least a distance d apart. Then the optimal design for any changepoint problem with a design criterion function, concave in π, is one of the designs in the set V.*

Proof of Theorem 2. For priors everywhere greater than zero on $[0, T]$, Theorem 1 shows that the image of $\mathbb{X}^n$ under G is a subset of the simplex $\triangle = \text{Conv}(G_f(\mathbb{X}^n))$. Clearly, it is the designs in V that map to the vertices of the simplex $\triangle$. Hence by Jensen's inequality, one of the designs in V is the design that minimizes the concave design criterion function. ■

A similar theorem and proof for cases when G is not one-to-one is given in Atherton, Charbonneau & Wolfson (2009). See Example 1 (Continued—estimation) at the end of Section 7.3 for an illustration of how Theorem 2 is used.

When testing for a change in a sub-interval $[t_1, t_2]$, it is necessary to fix two consecutive design points at t_1 and t_2, respectively. This is to ensure that an event expressed in terms of τ_x will correspond to a statement about the changepoint in $[t_1, t_2]$. In general, by fixing two design points at t_1 and t_2, we are left with the outer intervals $[0, t_1 - d]$ and $[t_2 + d, T]$ in which to place the remaining $n - 2$ design points (Figure 2). The optimal number of points in these two intervals,

FIGURE 2: The outer intervals $[0, t_1 - d]$ and $[t_2 + d, T]$ formed by fixing two design points at t_1 and t_2, respectively. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

respectively, will ultimately be determined by straight enumeration. Given a fixed number of points in each interval, it remains to determine the optimal placement of these points within $[0, t_1 - d]$ and $[t_2 + d, T]$.

Fixing the number of points in one of these two intervals, it is essentially clear from Theorems 1 and 2 that the optimal design for log-concave $\bar{g}$ places the points at either end of the selected interval. It is then easily seen that the global optimal design with two points fixed at t_1 and t_2 places points towards the four ends of the outer intervals.

Formally, fix $x_q = t_1$ and $x_{q+1} = t_2$. The points $x_1, \ldots, x_{q-1}$ fall in the interval $[0, t_1 - d]$ and the points $x_{q+2}, \ldots, x_n$ fall in the interval $[t_2 + d, T]$. We must have q in the set

$$Q = \{q \mid t_1 - d > (q-2)d \quad \text{and} \quad T - (t_2 - d) > (n - q - 2)d\}.$$

For $q \in Q$, let

$$\begin{aligned}
c_0 &= (0, d, \ldots, (q-2)d), & d_0 &= (t_2 + d, \ldots, t_2 + (n-q-1)d),\\
c_1 &= (0, d, \ldots, (q-3)d, t_1 - d), & d_1 &= (t_2 + d, \ldots, t_2 + (n-q-2)d, T),\\
&\vdots & &\vdots\\
c_{q-1} &= (t_1 - (q-1)d, \ldots, t_1 - d), & d_{n-q-1} &= (T - (n-q-2)d, \ldots, T - d, T),
\end{aligned}$$

and set

$$\begin{aligned}
C_q &= \{c_0, c_1, \ldots, c_{q-1}\},\\
D_q &= \{d_0, d_1, \ldots, d_{n-q-1}\}.
\end{aligned}$$

Theorem 3. *Suppose the density $\bar{g} + p_T$ is constructed from a density g that is log-concave with support $[0, T]$. Suppose the design criterion function is concave in π. The optimal design amongst all those designs with two observations fixed at t_1 and t_2, is in the Cartesian products $\{(t_1, t_2)\} \times D_1$, $C_{n-1} \times \{(t_1, t_2)\}$, or $C_q \times \{(t_1, t_2)\} \times D_q$ for some $2 \le q \le n - 2$.*

Proof of Theorem 3. Take any fixed q with $2 \le q \le n$. Noting the analogy of C_q and D_q to V, it is easily seen, by extending Theorem 1, that we must minimize the concave design criterion function over a Cartesian product of two regions, each a subset of a simplex and containing the vertices. It is also easily seen that the designs in C_q map to the vertices of the first simplex and the designs in D_q map to the vertices of the second simplex. The result follows from an extension of Jensen's inequality. For $q = 1$ or $q = n - 1$ the result is obvious. ■

See Example 1 (Continued—testing) at the end of Section 7.2 for an implementation of Theorem 3.

7. CONCAVITY OF SPECIFIC DESIGN CRITERION FUNCTIONS

The optimal designs derived in Section 6 have required the concavity of the design criterion functions as functions of the design measure π. We discuss optimal designs for testing and estimation separately, and show that for the most commonly used loss functions (Bayes risk based on squared error loss for estimation and the Bayes risk based on the generalized 0–1 loss and the Spezzaferri design criterion function for testing), the design criterion functions are, indeed, concave in the design measure.

The key once again is to re-express the model in terms of the design measure π.

7.1. Model and Dependence on π

In general, the model presented in Section 3 can be expressed as

$$\begin{array}{cc} \text{Joint likelihood} & f(\mathbf{y}|\theta_1, \theta_2, \tau_x = k) \\ \text{Marginal likelihoods} & f(\mathbf{y_1}|\theta_1) \quad \text{and} \quad f(\mathbf{y_2}|\theta_2) \end{array}$$

The model also includes specification of the joint prior $f(\theta_1, \theta_2, \tau) = f(\theta_1, \theta_2)(\bar{g} + p_T)$, which as a function of τ_x, becomes $f(\theta_1, \theta_2, \tau_x = k) = f(\theta_1, \theta_2)\pi_k$.

Letting $\eta(\cdot)$ represent an arbitrary probability measure, we have

$$f(\mathbf{y}|\tau_x = k) = \int\int f(\mathbf{y}|\theta_1, \theta_2, \tau_x = k) f(\theta_1, \theta_2)\, \mathrm{d}\eta(\theta_1)\, \mathrm{d}\eta(\theta_2), \tag{2}$$

whence

$$f(\mathbf{y}) = \sum_k f(\mathbf{y}|\tau_x = k)\pi_k \tag{3}$$

and

$$f(\tau_x = k|\mathbf{y}) = \frac{f(\mathbf{y}|\tau_x = k)\pi_k}{\sum_l f(\mathbf{y}|\tau_x = l)\pi_l}. \tag{4}$$

7.2. Design Criterion Functions When Testing for a Change

Referring to Example 1, researchers may want to know if the calcium supplementation takes effect during a sub-interval $[t_1, t_2]$ after administration. To test for a change in $[t_1, t_2]$ it is clearly necessary to guarantee that there are observations at t_1 and t_2. Say we place x_q at t_1 and x_{q+1} at t_2. Then the events $\{\tau \in [t_1, t_2]\}$ and $\{\tau_x = q\}$ are equivalent. For designs having two points fixed at t_1 and t_2, Theorem 3 gives optimal designs for design criterion functions that are concave in π. If either $t_1 = 0$ or $t_2 = T$, that is, we intend to test for a change in $[0, t_2]$ or $[t_1, T]$ then Theorem 3 needs to be modified slightly. For $[0, t_2]$ we need to include the case where the single design point x_1 is fixed at t_2 and similarly for $[t_1, T]$ the case where the single design point x_n is fixed at T. The important case of testing for a change in $[0, T]$ only requires the design point x_n to be fixed at T. It is then easily seen that for a concave design criterion function the optimal design lies in the subset $\{v_1, \ldots, v_n\}$.

Two common design criterion functions used for testing are the Bayes risk based on a generalized 0–1 loss function and the design Spezzaferri criterion function (see Spezzaferri, 1988). The theorems below establish concavity of two functions and, therefore, render our general optimal design results of Section 6 applicable to hypothesis testing.

The generalized 0–1 loss specifies a loss K_0, incurred when the null hypothesis is chosen although the true state is the alternative. Likewise K_1 is the loss incurred when the alternative is chosen but the true state is the null hypothesis. Letting Θ_{H_o} represent the parameter space under the null hypothesis and Θ_{H_1} represent the parameter space under the alternative hypothesis, the generalized 0–1 loss function leads to the following Bayes's rule.

$$\begin{aligned} y \in R_0 \quad \text{if} \quad & K_0 \int\limits_{\theta\in\Theta_{H_1}} f(y|\theta)p(\theta)\, \mathrm{d}\theta < K_1 \int\limits_{\theta\in\Theta_{H_0}} f(y|\theta)p(\theta)\, \mathrm{d}\theta \\ y \in R_1 \quad \text{if} \quad & K_1 \int\limits_{\theta\in\Theta_{H_0}} f(y|\theta)p(\theta)\, \mathrm{d}\theta < K_0 \int\limits_{\theta\in\Theta_{H_1}} f(y|\theta)p(\theta)\, \mathrm{d}\theta. \end{aligned}$$

The 0–1 Bayes risk with the decision rule above has the form

$$K_0 \int_{R_0} \int_{\theta \in \Theta_{H_1}} f(y|\theta) p(\theta)\, \mathrm{d}\theta\, \mathrm{d}y + K_1 \int_{R_1} \int_{\theta \in \Theta_{H_0}} f(y|\theta) p(\theta)\, \mathrm{d}\theta\, \mathrm{d}y. \tag{5}$$

Theorem 4. *Consider $H_0 : \tau \notin [t_1, t_2]$ vs $H_1 : \tau \in [t_1, t_2]$. The Bayes risk based on the generalized 0–1 loss for the changepoint testing problem is concave in π.*

Proof of Theorem 4. After re-expressing the risk (5) in terms of π for $H_0 : \tau \notin [t_1, t_2]$ and $H_1 : \tau \in [t_1, t_2]$, we apply the definition of concavity to show that the Bayes risk based on 0–1 is concave in π. See Atherton (2009b) for a complete proof. ■

The Spezzaferri design criterion function is based on the quadratic scoring rule of Brier (1950) and DeFinetti (1962). This criterion posits that the usefulness of an experiment is measured by the expected relative increase of utility after the experiment is performed (see Spezzaferri, 1988). Atherton (2009a) places the Spezzaferri criterion function in the usual Bayesian theoretic decision setting described as follows: Let the design space be $\mathcal{X}$, the decision space $\mathcal{D}$, the parameter space Θ, and the sample space $\mathcal{Y}$. For a utility, U selected by the experimenter the optimal design is then

$$\arg \left(\max_{x \in \mathcal{X}} \int_{\mathcal{Y}} \max_{d \in \mathcal{D}} \int_{\Theta} U(d, \theta, x, y) p(\theta|y, x) p(y|x)\, \mathrm{d}\theta\, \mathrm{d}y \right);$$

that is, the design $x \in \mathcal{X}$ that maximizes the data-averaged posterior expected utility at the optimal decision, d, is the optimal design (see, e.g., Chaloner & Verdinelli, 1995). Taking the utility U to be the quadratic scoring rule, Atherton (2009a) shows that this posterior expected quadratic scoring rule criterion function is equivalent to the Spezzaferri criterion function. The design criterion function based on the posterior quadratic scoring rule simplifies to the following,

$$1 - 2P(\theta \in \Theta_{H_0}) P(\theta \in \Theta_{H_1}) \int \frac{f(y|\theta \in \Theta_{H_0}) f(y|\theta \in \Theta_{H_1})}{f(y)}\, \mathrm{d}y. \tag{6}$$

Formula (6) has a much simpler form than the original Spezzaferri criterion function, and hence we work in this framework.

Theorem 5. *Consider $H_0 : \tau \notin [t_1, t_2]$ versus $H_1 : \tau \in [t_1, t_2]$. The posterior expected quadratic score criterion function is convex in π.*

Proof of Theorem 5. We prove that the expression

$$\int \frac{f(y|\theta \in \Theta_{H_0}) f(y|\theta \in \Theta_{H_1})}{f(y)}\, \mathrm{d}y \tag{7}$$

is concave. Our approach is to prove the integrand is concave for any value of y using the definition of concavity. It immediately follows that the integral is concave. See Atherton (2009b) for the complete proof. ■

Example 1 (Continued—testing)**.** We derive the optimal design for testing for a change in blood pressure, basing our design choice on the posterior expected score design criterion function in expression (6). Later, we also derive the optimal design for estimating the treatment effect if there is a change. We consider two relevant testing problems—testing for a change in blood pressure

TABLE 1: Values of the posterior expected quadratic score criterion function for the various design measures when testing for a change in $[0, T]$.

v_1	0.6668
v_2	0.6939
v_3	0.7061
v_4	0.7112
v_5	0.7113
v_6	0.7059
v_7	0.6977
v_8	0.6850
v_9	0.6660
v_{10}	0.6152
v_{Actual}	0.6920
v_{Times}	0.6914

The v_i's are defined at the beginning of Section 4.

within the entire study and testing for a change within 2 weeks of the start of the calcium diet. By Theorem 5 the design criterion function is concave in π. The optimal design for testing for a change in $[0, T]$ is one of the designs in $\{v_1, \ldots, v_n\}$. We used the results of Joseph et al. (1996) to suggest prior distributions for our model. We assumed that 10 blood pressure readings would be taken. We assumed the data to be normally distributed about the before-and after-change means with variance 25 mmHg^2. Assuming that the calcium supplementation would start at week 4 we took $(1/2)U(4 \text{ weeks}, 16 \text{ weeks}) + (1/2)I(16 \text{ weeks})$ to be the prior distribution for the changepoint where $U(a, b)$ denotes uniform distribution on (a, b). Normal conjugate priors were used for the before- and after-change means. Based on hyper-parameter values used in Joseph et al. (1996) we took the hyper-parameter means before and after the change each to be 115 mmHg and the before-and after-change hyper-parameter variances to be 100 mmHg^2. We evaluated integral (7) using Monte Carlo integration by taking 100,000 draws from $f(\mathbf{y}|\theta \in \Theta_{H_1})$ and averaging the remaining terms in the integrand. Table 1 lists the values of the posterior expected quadratic design criterion function at the designs $v_1, v_2, \ldots, v_{10}$, the only possible candidates for the optimal design measure. Table 1 also includes the corresponding value of the criterion function at the design v_{Actual} that was used in the original study by Lyle et al. (1987), and the value at the design v_{Times} when observations were taken at equally spaced times. We see that the optimal design is v_5, which places points at weeks (0–4, 12–16). We note that the optimal design gives a small improvement over the other designs: this is in itself informative, since it could be used to justify the simple design protocol used originally by Lyle et al. (1987). Of course, were the times between observations allowed to be smaller, the improvement would be greater.

Naturally, the sooner that a dietary supplement takes effect the better. Hence it might be of interest to test whether this supplement lowers blood pressure within the first 2 weeks. Using Theorem 3 we find that the optimal design places observations at weeks (0–4, 6–9, 16). For simulated criterion function values, see Atherton (2009b).

7.3. Concavity of the Design Criterion Function for Estimating the Parameters Before and After the Change

Suppose that we wish to find an optimal design for estimating the parameters before and after the change. For instance, referring to Example 1, one might wish to estimate the magnitude of

the blood pressure treatment effect, anticipating from past studies, say, that there would be some treatment effect.

Since the goal is to estimate θ_1 and θ_2, the Bayes risk based on squared error loss is,

$$\begin{aligned}
R &= \int\int (\theta_1 - E(\theta_1|\mathbf{y}))^2 f(\mathbf{y})\,\mathrm{d}\eta(\theta_1)\,\mathrm{d}\eta(\mathbf{y}) + \int\int (\theta_2 - E(\theta_2|\mathbf{y}))^2 f(\mathbf{y})\,\mathrm{d}\eta(\theta_2)\,\mathrm{d}\eta(\mathbf{y}) \\
&= \int \mathrm{Var}(\theta_1|\mathbf{y}) f(\mathbf{y})\,\mathrm{d}\eta(\mathbf{y}) + \int \mathrm{Var}(\theta_2|\mathbf{y}) f(\mathbf{y})\,\mathrm{d}\eta(\mathbf{y}) \\
&= \int E_{\tau_x=k|\mathbf{y}}(\mathrm{Var}(\theta_1|\mathbf{y}, \tau_x = k)) f(\mathbf{y})\,\mathrm{d}\eta(\mathbf{y}) + \int \mathrm{Var}_{\tau_x=k|\mathbf{y}}(E(\theta_1|\mathbf{y}, \tau_x = k)) f(\mathbf{y})\,\mathrm{d}\eta(\mathbf{y}) \\
&\quad + \int E_{\tau_x=k|\mathbf{y}}(\mathrm{Var}(\theta_2|\mathbf{y}, \tau_x = k)) f(\mathbf{y})\,\mathrm{d}\eta(\mathbf{y}) + \int \mathrm{Var}_{\tau_x=k|\mathbf{y}}(E(\theta_2|\mathbf{y}, \tau_x = k)) f(\mathbf{y})\,\mathrm{d}\eta(\mathbf{y}).
\end{aligned}$$

Denote the latter four integrals by R_1, R_2, R_3, and R_4, respectively. We first re-express the four integrals on the right hand side as functions of π. Note that R_1 and R_3 describe the within-model variability, while the terms R_2 and R_4 describe the between-model variability. The within-model variability refers to the variability around the parameters θ_1 and θ_2 given a fixed changepoint, and the between-model variability refers to the extra variability induced by the uncertainty of the location of the changepoint.

We show R_1 and R_3 are linear in π and then that R_2 and R_4 are concave in π. Without loss of generality, we establish these properties for R_1 and R_2 only. It then follows that R being a linear combination of concave functions, is concave in π. Once concavity of R is established, Theorem 2 may be used to find the optimal design in this estimation setting. The proof of the concavity of R is preceded by three lemmas. Many of the proofs are provided in the technical report by Atherton (2009b).

Lemma 2. *The terms R_1 and R_3 are linear functions of π.*

Proof of Lemma 2. Without loss of generality, consider only R_1. By Fubini's theorem, R_1 can be re-expressed as

$$R_1 = \sum_k E_{\mathbf{y}|\tau_x=k}(\mathrm{Var}(\theta_1|\mathbf{y}, \tau_x = k))\pi_k.$$

Since the density $f(\theta_1|\mathbf{y}, \tau_x = k)$ is independent of π, the expectation, $E(\theta_1|\mathbf{y}, \tau_x = k)$ and hence the variance $Var(\theta_1|\mathbf{y}, \tau_x = k)$, are both independent of π. Observing that the density, $f(\mathbf{y}|\tau_x = k)$, in Equation (2), does not depend on π, we conclude that $E_{\mathbf{y}|\tau_x=k}(\mathrm{Var}(\theta_1|\mathbf{y}, \tau_x = k))$ does not depend on π.

Next,

$$R_2 = \int \sum_k \left(E(\theta_1|\mathbf{y}, \tau_x = k) - \sum_l E(\theta_1|\mathbf{y}, \tau_x = l) f(\tau_x = l|\mathbf{y}) \right)^2 \cdot f(\tau_x = k|\mathbf{y}) f(\mathbf{y})\,\mathrm{d}\eta(\mathbf{y}).$$

From (3) and (4) we see that $f(\tau_x = k|\mathbf{y})$ and $f(\mathbf{y})$ are functions of π, so that R_2 is also a function of π. It suffices to prove that the integrand defining R_2 is concave in π for any value of $\mathbf{y}$, for then the integral R_2 is also concave in π. ■

Lemma 3. *The integral for R_2 equals*

$$\sum_{k>l} \left(E(\theta_1|\mathbf{y}, \tau_x = k) - E(\theta_1|\mathbf{y}, \tau_x = l)\right)^2 \frac{f(\mathbf{y}|\tau_x = l)\pi_l f(\mathbf{y}|\tau_x = k)\pi_k}{\sum_r f(\mathbf{y}|\tau_x = r)\pi_r}. \tag{8}$$

Proof of Lemma 3. The proof of Lemma 3 is given in Atherton (2009b). ■

Lemma 4. *For $k = 0, \ldots, n$, and for $d_k \in \mathbb{R}$ and $x_k, y_k \in \mathbb{R}^+$, we have*

$$\frac{\sum_{k>l}(d_k - d_l)^2(x_k + y_k)(x_l + y_l)}{\sum_r (x_r + y_r)} \geq \frac{\sum_{k>l}(d_k - d_l)^2 x_k x_l}{\sum_r x_r} + \frac{\sum_{k>l}(d_k - d_l)^2 y_k y_l}{\sum_r y_r}. \tag{9}$$

Proof of Lemma 4. The proof of Lemma 4 is given in Atherton (2009b). ■

Theorem 6. *The terms R_2 and R_4 are concave functions of π.*

Proof of Theorem 6. Consider the integrand of R_2 in the form of (8). Let $\pi_{(a)}$ and $\pi_{(b)}$ be any two values of the design measure and let $\pi_{(a),k}$ and $\pi_{(b),k}$ be the kth components of $\pi_{(a)}$ and $\pi_{(b)}$ respectively. We show that for $\lambda_a, \lambda_b \geq 0$ $(\lambda_a + \lambda_b) = 1$,

$$\begin{aligned}
&\sum_{k>l}(h_k - h_l)^2 \frac{f(\mathbf{y}|\tau_x = l)(\lambda_a\pi_{(a),l} + \lambda_b\pi_{(b),l}) f(\mathbf{y}|\tau_x = k)(\lambda_a\pi_{(a),k} + \lambda_b\pi_{(b),k})}{\sum_r \lambda_a f(\mathbf{y}|\tau_x = r)\pi_{(a),r} + \sum_s \lambda_b f(\mathbf{y}|\tau_x = s)\pi_{(b),s}} \\
&\quad \geq \lambda_a \sum_{k>l}(h_k - h_l)^2 \frac{f(\mathbf{y}|\tau_x = l)\pi_{(a),l} f(\mathbf{y}|\tau_x = k)\pi_{(a),k}}{\sum_r f(\mathbf{y}|\tau_x = r)\pi_{(a),r}} \\
&\qquad + \lambda_b \sum_{k>l}(h_k - h_l)^2 \frac{f(\mathbf{y}|\tau_x = l)\pi_{(b),l} f(\mathbf{y}|\tau_x = k)\pi_{(b),k}}{\sum_s f(\mathbf{y}|\tau_x = s)\pi_{(b),s}}.
\end{aligned}$$

The above inequality is equivalent to

$$\begin{aligned}
&\sum_{k>l}(h_k - h_l)^2 \\
&\quad \times \frac{(\lambda_a f(\mathbf{y}|\tau_x = l)\pi_{(a),l} + \lambda_b f(\mathbf{y}|\tau_x = l)\pi_{(b),l})(\lambda_a f(\mathbf{y}|\tau_x = k)\pi_{(a),k} + \lambda_b f(\mathbf{y}|\tau_x = k)\pi_{(b),k})}{\sum_r \lambda_a f(\mathbf{y}|\tau_x = r)\pi_{(a),r} + \sum_s \lambda_b f(\mathbf{y}|\tau_x = s)\pi_{(b),s}} \\
&\geq \sum_{k>l}(h_k - h_l)^2 \frac{\lambda_a f(\mathbf{y}|\tau_x = l)\pi_{(a),l}\lambda_a f(\mathbf{y}|\tau_x = k)\pi_{(a),k}}{\sum_r \lambda_a f(\mathbf{y}|\tau_x = r)\pi_{(a),r}} \\
&\quad + \sum_{k>l}(h_k - h_l)^2 \frac{\lambda_b f(\mathbf{y}|\tau_x = l)\pi_{(b),l}\lambda_b f(\mathbf{y}|\tau_x = k)\pi_{(b),k}}{\sum_s \lambda_b f(\mathbf{y}|\tau_x = s)\pi_{(b),s}}.
\end{aligned}$$

This inequality is guaranteed by Lemma 4 by setting

$$\begin{aligned}
d_k &= h_k, \\
d_l &= h_l, \\
X_k &= \lambda_a f(\mathbf{y}|\tau_x = k)\pi_{(a),k}, \\
X_l &= \lambda_a f(\mathbf{y}|\tau_x = l)\pi_{(a),l},
\end{aligned}$$

TABLE 2: Values of the Bayes risk based on squared error loss, for the designs in set the V.

v_0	89.7711
v_1	71.2716
v_2	67.8714
v_3	66.6232
v_4	66.1647
v_5	66.2339
v_6	66.4680
v_7	67.2907
v_8	68.8239
v_9	72.1571
v_{10}	96.0371
v_{Actual}	67.7416
v_{Times}	67.7946

$$Y_k = \lambda_b f(\mathbf{y}|\tau_x = k)\pi_{(b),k}, \quad \text{and}$$
$$Y_l = \lambda_b f(\mathbf{y}|\tau_x = l)\pi_{(b),l}.$$

Since the integrand is concave for all $\mathbf{y}$ it follows that the integrals, R_2 and, hence, R_4, are concave. ■

Example 1 (Continued—estimation)**.** With the goal of estimating the calcium supplementation effect on blood pressure, we found the optimal design by invoking Theorem 2, Lemma 2, and Theorem 6. The various designs and their risks (based on a Monte Carlo simulation with 100,000 iterations) are given in Table 2. We see that the optimal design is v_4 with observations at weeks (0, 1, 2, 3, 4, 5, 13, 14, 15, 16). For simulated values of R_1, R_2, R_3, and R_4, see Atherton (2009b).

8. CONCLUDING REMARKS

We stress that the single-path results in Section 6 of this paper apply for all design criterion functions that are concave in the design measure. Our results also apply to data with any distribution and for any prior distributions on the before- and after-change parameters. The only constraint we have imposed is that the prior distribution for the changepoint is log-concave.

Simulations in Zhou (1997) demonstrate the difficulties of finding an optimal design for estimating the changepoint itself since no easily described pattern emerges. Unfortunately, re-expressing the design criterion function in terms of π does not simplify the problem of finding the optimal design.

When the parameters of interest are the means μ_1 and μ_2, researchers sometimes prefer to make inference about $\mu_1 - \mu_2$. In this setting the obvious parameters would be μ and $\mu + \delta$. The inference is then about δ. The Bayes risk based on a squared error loss for δ, remains concave in π, and the optimal design follows from Theorem 2.

When estimating parameters θ_1 and θ_2, the term $R_1 + R_3$ is equivalent to the generalized Läuter's criterion function introduced by Zhou et al. (2003), when a Bayes risk based on squared error loss is used. Suppose that the parameters are the means and their prior distributions are

normal. Suppose further the difference between the hyper-parameter means is large compared to the hyper-parameter- and model-variances. It is a conjecture that the generalized Läuter criterion function gives a good approximation to the Bayes risk based on squared error loss. That is, $R_2 + R_4$ is small, and $R_1 + R_3$ is a good approximation to R.

Although the discussion in this paper has focused on scalar valued parameters, with appropriate modification of the loss functions, these results could be extended to vector valued parameters.

With testing and estimation both as the goals, design criterion functions can be combined. We have suggested three criterion functions (two for testing and one for estimating) which are all concave in π. For example, by taking a weighted convex combination of one of the testing criterion functions and the estimation criterion function, we obtain a criterion function that is again concave in π; the weights would be selected to reflect the perceived relative importance of the two problems. Theorem 3 may then be used to find the optimal design. Of course such an approach would need to be taken with care since the loss functions for estimating and testing are not really comparable.

APPENDIX

A.1. Proof of Theorem 1

The proof is in two parts. First, in Lemma 1, we show that the set $G(V)$ is affinely independent. It follows that the convex hull $\mathrm{Conv}(G(V))$, which we denote as $\triangle$, is a simplex. Second, in the proof of Theorem 1 we consider each $(n-1)$-dimensional boundary of $G(\mathbb{X}^n)$ and show that $G(\mathbb{X}^n) \subset \triangle$ if g has support on $[0, T]$.

Lemma 1. *Let g (piecewise continuous) have support $[0, T]$. Then the set $G(V)$ is affinely independent.*

Proof of Lemma 1. Let

$$\alpha_i = \int_{(i-1)d}^{id} \bar{g}, \qquad \beta_i = \int_{(n-i-1)d}^{T-(i-1)d} \bar{g}, \qquad \gamma_i = \int_{T-id}^{T-(i-1)d} \bar{g},$$

for $i = 1, \ldots, n-1$, and let

$$\bar{\beta} = \int_{(n-1)d}^{T} \bar{g} + p_T.$$

To prove that $G(V)$ is an affinely independent set, we show that the vectors $G(v_0) - G(v_n), G(v_1) - G(v_n), \ldots, G(v_{n-1}) - G(v_n)$ are linearly independent by showing that the determinant

$$\begin{vmatrix} \alpha_1 - \gamma_{n-1} & \alpha_2 - \gamma_{n-2} & \cdots & \alpha_{n-1} - \gamma_1 & \bar{\beta} - p_T \\ \alpha_1 - \gamma_{n-1} & \alpha_2 - \gamma_{n-2} & \cdots & \beta_1 - \gamma_1 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \alpha_1 - \gamma_{n-1} & \beta_{n-2} - \gamma_{n-2} & \cdots & 0 & 0 \\ \beta_{n-1} - \gamma_{n-1} & 0 & \cdots & 0 & 0 \end{vmatrix}$$

is non-zero. It is indeed the case as all the diagonal elements $\beta_j - \gamma_j = \int_{(n-j-1)d}^{T-jd} \bar{g}$ and $\bar{\beta} - p_T = \int_{(n-1)d}^{T} \bar{g}$ are greater than zero.

To prove Theorem 1 the following preliminaries are needed. For $1 \leq i \leq n-1$ define E_i to be the facets $x_{i+1} = x_i + d$ of $\mathbb{X}^n$. Define E_0 to be the facet $x_1 = 0$, and E_n the facet $x_n = T$. Each facet E_i is thus the convex hull of every vertex except the vertex v_{n-i} (which has the largest distance possible between the points x_i and x_{i+1}). The facet E_i is then parametrized by the $(n-1)$-dimensional vector $(x_1, \ldots, x_i, x_{i+2}, \ldots, x_n)$, without x_{i+1}. Hence $G(E_i)$ is an $(n-1)$-dimensional surface lying in $\mathbb{R}^n$.

To simplify notation, let G^i denote the restriction $G|_{E_i}$ of G to the subset E_i of $\mathbb{X}^n$. Then

$$G^i(x_1, \ldots, x_i, x_{i+2}, \ldots, x_n) = G(x_1, \ldots, x_i, x_i + d, x_{i+2}, \ldots, x_n).$$

We denote partial derivatives by subscripts. If $p \in E_i$ is such that $q = G^i(p)$ and $T_q G^i$ is the space tangent to G^i at the point q, then for $j \neq i+1$ the partial derivatives $G^i_j|_p$, lie in the tangent space $T_q G^i$.

It is easy to show that, since the density function g is always greater than zero, the partial derivatives $\{G^i_1, \ldots, G^i_i, G^i_{i+2}, \ldots, G^i_n\}$ are linearly independent in any tangent space $T_q G^i$.

Let $e_1, \ldots, e_n$ denote the usual basis of $\mathbb{R}^n$, using a determinantal formula similar to the one for the cross product (when $n = 3$), we compute the normal vector to the surface G^i

$$\begin{aligned} N^i = {} & (-1)^{n-i} \bar{g}(x_1) \cdots \bar{g}(x_{i-1}) \bar{g}(x_i + d) \bar{g}(x_{i+2}) \cdots \bar{g}(x_n) e_i \\ & + \sum_{j=i+1}^{n} (-1)^{n-i} \bar{g}(x_1) \cdots \bar{g}(x_{i-1}) \left(\bar{g}(x_i + d) - \bar{g}(x_i)\right) \bar{g}(x_{i+2}) \cdots \bar{g}(x_n) e_j. \end{aligned}$$

To ascertain if N^i points inwards or outwards we take the dot product of N^i and a vector known to point inwards. In the facet E_i, the coordinate x_{i+1} is fixed. As we vary it, we penetrate inside $\mathbb{X}^n$ and therefore G_{i+1} points inside $G(\mathbb{X}^n)$. With $\bar{g}(x_{i+1})$ as the ith component and $-\bar{g}(x_{i+1})$ as the $(i+1)$st component we have

$$G_{i+1} = (0, \ldots, 0, \bar{g}(x_{i+1}), -\bar{g}(x_{i+1}), 0, \ldots, 0).$$

Hence,

$$N^i \cdot G_{i+1} = (-1)^{n-i} \bar{g}(x_1) \cdots \bar{g}(x_{i-1}) \bar{g}(x_i) \bar{g}(x_{i+1}) \cdots \bar{g}(x_n).$$

For simplicity, we set

$$\bar{N}^i = (-1)^{n-i} N^i,$$

so that $\bar{N}^i$ always points inwards. Let $\bar{M}^i = \bar{N}^i / \|\bar{N}^i\|$.

We can study the shape of a surface, be it $G(E_i)$ or other surfaces, by studying how the normal vector changes as it moves in various directions. The object that accomplishes this task is the shape operator L. It is a map from the tangent plane at one point to itself. The eigenvalues of this operator are called "principal curvatures."

Details can be found in Chapter 5 of O'Neill (1997), but for our purposes, the shape operator is determined by the relation

$$L(G^i_j) \cdot G^i_k = \bar{M}^i \cdot G^i_{jk}. \tag{10}$$

If the principal curvatures calculated are all non-positive at every point $p \in E_i$, then the shape of $G(E_i)$ is such that $G(E_i)$ is "pulled" inside $\triangle$. ■

For the sake of completeness we restate Theorem 1 below.

Theorem 1. *Let $\triangle = \mathrm{Conv}(G(V))$ be the convex hull of $G(V)$. If in the density $\bar{g} + p_T$, the density g is log-concave and everywhere greater than zero then $G(\mathbb{X}^n) \subset \triangle$.*

Proof of Theorem 1. The set $G(V) \subset G(\mathbb{X}^n)$ lies in $\triangle$. We examine the $n+1$ boundaries of $G(\mathbb{X}^n)$ and show that for any log-concave g, each lies inside $\triangle$. Since g is everywhere greater than zero, G is injective and, thus, maps boundaries of $\mathbb{X}^n$ to boundaries of $G(\mathbb{X}^n)$. We can then pair boundaries of $G(\mathbb{X}^n)$ and $\triangle$, noting that each pair is parametrized by the same facet E_i of $\mathbb{X}^n$.

Since G^0 parametrizes $\pi_0 = 0$ and G^n parametrizes $\pi_n = p_T$, the boundaries parametrized by G^0 and G^n are coincident with their corresponding boundaries in $\triangle$. Consequently, we focus on the boundaries parametrized by G^i for $1 \le i \le n-1$.

Referring to (10) since all mixed partial derivatives are zero, $L(G^i_j) \cdot G^i_k$ is zero for all $j \neq k$. Furthermore, $L(G^i_j) \cdot G^i_j = 0$ for $j \neq i$. We are left with

$$\begin{aligned}\bar{N}^i \cdot G^i_{ii} = (-1)^{n-i} &\bar{g}(x_1) \cdots \bar{g}(x_{i-1}) \bar{g}(x_{i+2}) \cdots \bar{g}(x_n) \\ &\cdot \left(\left(\bar{g}'(x_i + d) - \bar{g}'(x_i) \right) \bar{g}(x_i + d) - \bar{g}'(x_i + d) \left(\bar{g}(x_i + d) - \bar{g}(x_i) \right) \right)\end{aligned}$$

to determine the sign of $L(G^i_i) \cdot G^i_i$. It is obviously determined by the sign of the expression

$$\left(\bar{g}'(x_i + d) - \bar{g}'(x_i) \right) \bar{g}(x_i + d) - \bar{g}'(x_i + d) \left(\bar{g}(x_i + d) - \bar{g}(x_i) \right).$$

To summarize, for some constant A_i, we have

$$L(G^i_j) \cdot G^i_k = \begin{cases} A_i, & \text{when } j = k = i, \\ 0, & \text{otherwise.} \end{cases}$$

Since the basis G^i_j is not necessarily orthonormal, A_i is not necessarily an eigenvalue of L, and hence perhaps not a principal curvature. However, since the shape operator L is self adjoint, there exists an orthonormal basis consisting of eigenvectors $w_1, \ldots, w_{n-1}$ for L with respective eigenvalues $k_1, \ldots, k_{n-1}$. Denoting the Kronecker delta function by δ_{jk}, we have

$$L(w_j) \cdot w_k = \delta_{jk} k_j.$$

Since $\{G^i_i, \ldots, G^i_i, G^i_{i+2}, \ldots, G^i_n\}$ and $\{w_1, \ldots, w_{n-1}\}$ are two bases of the same space, there is a matrix $[h_{lk}]$ such that, $w_j = \sum_k h_{jk} G^i_k$. The principal curvatures, or eigenvalues, of the shape operator L, are

$$\begin{aligned} k_j &= L(w_j) \cdot w_j \\ &= \sum_{k,l} h_{jk} h_{jl} L(G^i_k) \cdot G^i_l \\ &= A_i (h_{ji})^2. \end{aligned} \tag{11}$$

If g is log-concave then $\frac{g'}{g}$ is monotone decreasing and consequently

$$\bar{g}(x_i) \bar{g}'(x_i + d) - \bar{g}'(x_i) \bar{g}(x_i + d) \le 0.$$

Therefore $L(G_i^i) \cdot G_i^i$, and hence the principal curvatures given by Equation (11), are non-positive. The result follows. ■

ACKNOWLEDGEMENTS

This work was supported by the Institut des Sciences Mathématiques in Montréal, le Fonds Québécois de la Recherche sur la Nature et les Technologies, the National Sciences and Engineering Research Council of Canada, and an NIH grant. We thank two reviewers and the associate editor for their helpful comments.

BIBLIOGRAPHY

J. Atherton (2009a). A new simple interpretation of an optimal design criterion. *Statistics & Probability Letters*, 79, 707–710.

J. Atherton (2009b). "*Bayesian optimal design for changepoint problems: Concavity of design criterion functions*," Technical Report, McGill University.

J. Atherton & D. B. Wolfson (2009). "*Bayesian optimal design for changepoint problems: The multi-path normal case*," Technical Report, McGill University.

J. Atherton, B. Charbonneau & D. B. Wolfson (2009). "*Bayesian optimal design for changepoint problems: Further prior distributions*," Technical Report, McGill University.

M. Bagnoli & T. Bergström (1989). "*Log-concave probability and its applications*," Manuscript, University of Michigan.

W. Bischoff (1998). A functional central limit theorem for regression models. *Annals of Statistics*, 26(4), 1398–1410.

W. Bischoff & F. Miller (2000). Asymptotically optimal tests and optimal designs for testing the mean in regression models with applications to change-point problems. *Annals of the Institute of Statistical Mathematics*, 52(4), 658–679.

L. Blackmore & B. Williams (2005). Finite horizon control design for optimal model discrimination. In *Proceedings of the joint 44th IEEE conference on decision and control and European control conference*.

G. Brier (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78, 1–3.

K. Chaloner & I. Verdinelli (1995). Bayesian experimental design: A review. *Statistical Science*, 10, 273–310.

J. Chen & A. K. Gupta (2000). "*Parametric Statistical Changepoint Analysis*," Birkhauser, Boston, MA.

M. Clyde (2001). "*Experimental design: A Bayesian perspective*," Technical Report, Duke University.

M. Clyde & K. Chaloner (1996). The equivalence of constrained and weighted designs in multiple objective design problems. *Journal of the American Statistical Association*, 91(435), 1236–1244.

A. DasGupta (1996). Review of optimal Bayes designs. *Handbook of Statistics*, 13, 1099–1144.

B. DeFinetti (1962). "Does it make sense to speak of good probability appraisers?" in I. Good, A. Mayne & J. Smith, Eds, "*The Scientist Speculates (An Anthology of Partly Baked Ideas)*," Capricon Books, New York, pp. 357–364.

K. Felsenstein (1990). Optimal Bayesian design for discrimination among rival models. *Computational Statistics & Data Analysis*, 14, 427–436.

L. Joseph, D. Wolfson, R. du Berger & R. Lyle (1996). "Change-point analysis of a randomized trial on the effects of calcium supplementation on blood pressure," in D. Berry & D. Stangl, Eds, "*Bayesian Biostatistics*," Marcel Dekker, New York, pp. 617–649.

R. M. Lyle, C. L. Melby, G. C. Hyner, J. W. Edmondson, J. Z. Miller & M. H. Weinberger (1987). Blood pressure and metabolic effects of calcium supplementation in normotensive white and black men. *Journal of American Medical Association*, 257(13), 1773–1777.

B. O'Neill (1997). "*Elementary Differential Geometry*," Academic Press, New York.

F. Spezzaferri (1988). "Nonsequential designs for model discrimination and parameter estimation," in J. Bernardo, M. DeGroot & D. Lindley, Eds, "*Bayesian Statistics 3*," Oxford University Press, pp. 777–783.

X. Zhou (1997). *Optimal design for changepoint problems*. PhD Thesis, McGill University.

X. Zhou, L. Joseph, D. Wolfson & P. Belisle (2003). A Bayesian A-optimal and model robust design criterion. *Biometrics*, 59, 1082–1088.

Received 13 October 2008
Accepted 20 May 2009