A STATISTICAL APPROACH TO ADAPTIVE PARAMETER TUNING IN NATURE-INSPIRED OPTIMIZATION AND OPTIMAL SEQUENTIAL DESIGN OF DOSE-FINDING TRIALS

By

Kwok Pui Choi
Tze Leung Lai
Xin T. Tong
Weng Kee Wong

Technical Report No. 2020-17
December 2020

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065
A STATISTICAL APPROACH TO ADAPTIVE PARAMETER TUNING IN NATURE-INSPIRED OPTIMIZATION AND OPTIMAL SEQUENTIAL DESIGN OF DOSE-FINDING TRIALS

Kwok Pui Choi, Tze Leung Lai, Xin T. Tong and Weng Kee Wong

National University of Singapore, Stanford University and UCLA

Abstract: Nature-inspired metaheuristic algorithms have been steadily rising in popularity in the last couple of decades and now constitute a major toolbox for tackling complex high-dimensional optimization problems. Making use of group sequential experimentation, adaptive design, multi-armed bandits, and bootstrap resampling methods, this paper develops a novel statistical methodology for efficient and systematic group sequential selection of the tuning parameters, which have been widely recognized to be pivotal to the success of metaheuristic optimization algorithms in practice, as new information accumulates during the course of an experiment. The methodology is applied to computation of optimal experimental designs in nonlinear regression models and is illustrated with the solution of long-standing optimal design problems in early-phase dose-finding oncology trials.

Key words and phrases: Adaptive group sequential designs, compound optimality criterion for toxicity and efficacy, locally D-optimal and c-optimal designs.
1. Introduction

Metaheuristic optimization algorithms have become major tools for tackling complex high-dimensional optimization problems of the Information Age in the past two decades since the turn of the millennium. They are essentially free of assumptions, fast, easy to implement and frequently able to find the optimum or a solution close to the optimum after relatively few iterations, but require good tuning parameters. In particular, for the metaheuristic optimization qPSO (quantum Particle Swarm Optimization), Sun, Lai and Wu (2012, Chapter 5) provide convergence analysis and performance comparison for different choices of the tuning parameters to show the fundamental importance of choosing them well. Huang, Li and Yao (2020) have recently emphasized the importance of finding appropriate tuning parameters in metaheuristic optimization algorithms in their survey of the subject. They say that whereas heuristic optimization algorithms are problem-specific and implement some heuristic rules or strategies to solve the specific problems, “metaheuristics are high-level methodologies or general algorithmic templates” and “most of metaheuristics are nature-inspired (inspired from some principles in physics, biology, etc.), contain stochastic components, and often have several free parameters that can be set by users according to the problem(s) at hand” which can have “strong impact
2 ADAPTIVE GROUP SEQUENTIAL SELECTION OF TUNING PARAMETERS IN METAHEURISTIC OPTIMIZATION

on the performance or efficiency of a metaheuristic”. To address the increasing demand for “systematic approaches for metaheuristics’ parameter setting”, Section 2 develops a novel statistical methodology which we apply in Section 3 to a metaheuristic optimization algorithm called particle swarm optimization (PSO). Section 4 shows how this methodology can be applied to solve long-standing optimal sequential design problems in early-phase dose-finding oncology trials, and gives some concluding remarks.

2. Adaptive Group Sequential Selection of Tuning Parameters in Metaheuristic Optimization

The novel approach to adaptive tuning parameter selection for metaheuristic optimization algorithms, presented in Section 2.1, is developed from two statistical ideas, namely, group sequential design of adaptive trials (Section 2.1) and multi-armed bandit schemes for selecting the unknown best population/strategy (Section 2.2). An efficiency theory of this adaptive parameter tuning approach for metaheuristic optimization is given in Section 2.3. As pointed out by Bartroff, Lai and Shih (2013, pp.2-4), the statistical subject of sequential design and analysis “was born in response to the need for more efficient testing of anti-aircraft gunnery during World War II, which led to the development of the sequential probability ratio test”,
and it was recognized a few years after the War that “sequential hypothesis testing might provide a useful tool in clinical trials to test the efficacy of new medical treatments.” However, “in double-blind multi-center clinical trials, it is not possible to arrange for continuous examination of the data as they accumulate.” On the other hand, many trials have Data and Safety Monitoring Committees (DSMC) who conduct periodic reviews of the trial, particularly with respect to the incidence of treatment-related adverse events. One such trial was the Beta-blocker Heart Attack Trial that was terminated early during an interim analysis by its DSMC because of positive results on the efficacy of the treatment. This success story in 1981 paved the way for steadily increasing adoption of group sequential designs and major advances in group sequential methods are summarized in Chapter 4 of [Bartroff, Lai and Shih (2013)]. Since statistical applications of metaheuristic optimization is often related to estimation, or hypothesis testing, of an unknown parameter $\theta$, we use $\lambda$ to denote the vector of tuning parameters in the metaheuristic optimization algorithm. An analog of $\lambda$ is the hyperparameter of the prior distribution in Bayesian estimation.
2.1 Group sequential learning of optimal $\lambda$

Consider the problem of searching for $\mathbf{x}^* \in \mathbb{R}^d$ that minimizes $f(\mathbf{x}) \in \mathbb{R}$ over $\mathbf{x}$ belonging to some bounded region of $\mathbb{R}^d$, where $f$ is a given objective function. Group sequential updating of the optimal tuning parameter vector, which we assume to belong to some bounded region of $\mathbb{R}^m$, for the metaheuristic optimization algorithms used for this search is carried out at user-specified times $t_1 < \cdots < t_J$, with $J = \max\{j \geq 1 : t_j < T\}$, during the running time $T$ of the algorithm, in contrast to offline parameter tuning surveyed by Huang, Li and Yao (2020). Illustrative examples are given in Section 4.2. Moreover, metaheuristic optimization algorithms involve stochastic components, denoted by a multivariate vector $\mathbf{Z}_t$ at time $t < T$. We assume that $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ are independent in the basic algorithm that we describe below.

The metaheuristic optimization algorithm to search for $\mathbf{x}^*$ that minimizes $f$, with tuning parameter $\mathbf{\lambda}$, will be denoted by $A(\mathbf{\lambda})$, and the sample path generated by $A(\mathbf{\lambda})$ will be denoted by $\mathbf{x}_t(\mathbf{\lambda})$. Let $\mathbf{x}_t^*(\mathbf{\lambda}) = \arg\min_{s \leq t} f(\mathbf{x}_s(\mathbf{\lambda}))$. Initialize $\mathbf{\lambda}$ by choosing $\mathbf{\lambda}_0$ at random (or according to some given distribution) from the bounded region $\Lambda \subset \mathbb{R}^m$ to which the tuning parameter belongs. Let $\Lambda_0 = \{\mathbf{\lambda}_0\}$. Run $A(\mathbf{\lambda}_0)$ until $t_1 - 1$, and at time $t_j$ ($j = 1, \ldots, J$), update $\mathbf{\lambda}$ and run $A(\mathbf{\lambda}_j)$ with the updated value $\mathbf{\lambda}_j$.
2.1 Group sequential learning of optimal $\lambda$

to generate $x_t(\lambda_j)$ and $x_t^*(\lambda_j)$ for $t_j \leq t < t_{j+1}$ using the following procedures to handle the stochastic components $Z_t$ in $x_t(\lambda)$ for $\lambda = \lambda_j \in \Lambda_j$ and $t_j \leq t < t_{j+1}$ and to update the choice of $\lambda_j$ and $\Lambda_j$ at time $t_j$.

A. How to deal with the stochastic components $Z_t$ in $x_t(\lambda)$ for given $\lambda$, $t_j \leq t < t_{j+1}$:

A1. Generate $B$ samples of independent $Z_t^{(b)}$, $t_j \leq t < t_{j+1}$ ($b = 1, \ldots, B$).

A2. With the stochastic components $Z_t^{(b)}$ of the metaheuristic optimization algorithm, generate $B$ simulated samples of $(x_{t,b}(\lambda), x_{t,b}^*(\lambda))$, $t_j \leq t < t_{j+1}$ ($b = 1, \ldots, B$).

A3. Taking the average of $f(x_{t,b}^*(\lambda))$ over $b = 1, \ldots, B$ yields an estimate, denoted by $\hat{E}f(x_t^*(\lambda))$, of $Ef(x_t^*(\lambda))$ for $t_j \leq t < t_{j+1}$.

B. How to update $\lambda_j$ and $\Lambda_j$ at $t_j$ ($1 \leq j \leq J$):

B1. Choose $\lambda^*$ from $\Lambda$ according to some given distribution and let $\Lambda^*_j = \Lambda_{j-1} \cup \{\lambda^*\}$.

B2. Letting $\Delta_j(\lambda) = f(x_{t_{j-1}}^*(\lambda)) - f(x_{t_j}^*(\lambda))$, compute $\hat{E}\Delta_j(\lambda)$ for each $\lambda \in \Lambda^*_j$; see Step A3 above.

B3. Let $\lambda^*_j = \arg\max_{\lambda \in \Lambda_j} \hat{E}\Delta_j(\lambda)$. Sample $\lambda_j$ from $\Lambda^*_j$ with probability $1 - \epsilon$ assigned to $\lambda^*_j$ and probability $\epsilon/j$ assigned to each $\lambda \in \Lambda^*_j \setminus \{\lambda^*_j\}$;
2.2 Multi-armed bandits and \( \epsilon \)-greedy randomization

This is the \( \epsilon \)-greedy randomization scheme, with user-specified \( 0 < \epsilon < 1/2 \), that will be explained in Section 2.2. Let \( \Lambda_j = \Lambda_{j-1} \cup \{ \lambda_j \} \).

2.2 Multi-armed bandits and \( \epsilon \)-greedy randomization

The multi-armed bandit problem, introduced by Robbins (1952) as a new direction for sequential design of experiments, subsequently evolved into an important area of reinforcement learning, in which the dilemma between active learning (also called “exploration”) to gather information about unknown system parameters and passive learning (also called “exploitation”) from the outputs that the control system aims at driving toward some prescribed target; see Kaelbling, Littman and Moore (1996). Robbins (1952) considered two populations (arms) to sample sequentially from in order to maximize the expected sum \( E(\sum_{t=1}^{T} Y_t) \) of the observations \( Y_t \) generated when the population means are unknown. If the population with the larger mean \( \mu^* \) were known, then one would sample from it to receive expected reward \( T\mu^* \). By sampling at a sparse set of increasing times \( t_i \) from the population with the smaller total sample size up to time \( t_i \), while sampling from the population with the larger sample mean at other times, Robbins used the law of large numbers to show that \( E(\sum_{t=1}^{T} Y_t) = T(\mu^* + o(1)) \) as \( T \to \infty \), where sparsity means that for \( I = \max\{i : t_i \leq T\} \), \( t_I \to \infty \) but
2.2 Multi-armed bandits and $\epsilon$-greedy randomization

$t_{i}/T \to 0$. Subsequently Lai and Robbins (1985) extended multi-armed bandits from 2 to $K$ arms and provided a definitive theory by introducing the concept of \textit{regret} for adaptive allocation rules, defined by

$$R_T = T\mu^* - E\left(\sum_{t=1}^{T} Y_t\right) = \sum_{k: \mu^k < \mu^*} (\mu^* - \mu^k)ET(k),$$

where $T(k)$ is the total sample size from population $k$ that has mean $\mu^k$; an allocation rule is called “adaptive” if its choice of which population to sample from at time $t$ depends on the observations prior to $t$. They derived an asymptotic lower bound, as $T \to \infty$, for the regret $R_T$ of uniformly good adaptive allocation rules; an adaptive allocation rule is called uniformly good if $R_T = o(T^a)$ for all $a > 0$ and at all values of $(\mu^1, \ldots, \mu^k)$. The asymptotic lower bound is given by

$$R_T \geq (1 + o(1))\sum_{k: \mu^k < \mu^*} \frac{(\mu^* - \mu^k)}{I(\mu^k, \mu^*)} \log T,$$

where $I(\cdot, \cdot)$ denotes the Kullback-Leibler information number, and Lai and Robbins (1985) and Lai (1987) have shown for the exponential family of densities $e^{\theta y - \psi(\theta)}$ (for which $\mu = \psi'(\theta)$) that the asymptotic lower bound can be attained by allocating to the population with the largest upper confidence bound (UCB) at stage $t - 1$. An alternative to the UCB rule for attaining the asymptotic lower bound is the $\epsilon$-greedy randomization algorithm that allocates at stage $t$ to the population with the largest sample
mean at stage $t - 1$ with probability $1 - \epsilon$ and to each remaining population with probability $\epsilon/(K - 1)$, see Auer, Cesa-Bianchi and Fischer (2002).

2.3 Efficiency theory of adaptive hyperparameter tuning

We first make use of multi-armed bandit theory summarized in Section 2.2 to derive optimality of the adaptive selection of the hyperparameter $\hat{\lambda}_j$ at time $t_j$ when the hyperparameter can be updated to run the metaheuristic optimization algorithm for $t_j \leq t < t_{j+1}$, with $t_{j+1} - 1 = T$. Let $\tau_j = t_{j+1} - t_j$. Following Chan and Lai (2006, p.182) who use a function $g$ to incorporate all previous measures of sampling cost in the selection and ranking literature on normal data or more general observations from exponential families, define the total sampling cost

$$C_T(\Lambda_J) = \sum_{j=1}^{J} g(x^* - x^*_j(\lambda_j))\tau_j$$

for a metaheuristic optimization algorithm that updates its tuning hyperparameter at times $t_1, \ldots, t_J$, leading to the set $\Lambda_J = \{\lambda_1, \ldots, \lambda_J\}$ of successive hyperparameter values to run the algorithm as in part A of Section 2.1. The case

(C) $g(x) = 0$ if $\|x\| < \delta$ and $\inf_{\|x\| \geq \epsilon} g(x) > 0$ for $\epsilon > \delta$
2.3 Efficiency theory of adaptive hyperparameter tuning

Corresponds to the “indifference zone” formulation, in which selecting a population (or method) is as good as the best one if its expected outcome is within $\delta$ of the best; see Section 5.2 of Chan and Lai (2006) and in particular its connection to multi-armed bandits.

**Theorem 1.** Assume that $g$ satisfies (C) and that $Z_t$ are independent with density from the exponential family $e^{\theta^\top z - \psi(\theta)}$. The group sequential selection method yielding $\hat{\Lambda}_J$ in part B of Section 2.1 has asymptotically minimal $E_{C_T}(\Lambda_J)$ as $B \to \infty$, among all group sequential procedures that satisfy $E_{C_T}(\Lambda_J) = o(B^r)$ for all $r > 0$ and all $\theta$.

The last paragraph in Section 5.2 of Chan and Lai (2006) mentions the asymptotic lower bound for the regret of uniformly good adaptive allocation rules in the multi-armed bandit problem, which we have reviewed in Section 2.2, and suggests how it can fit into the indifference zone formulation in the selection and ranking literature. Theorem 1 provides concrete details for the problem of selecting the best $\Lambda_J$ within $\delta$ of the best by using the function $g$ that satisfies the “$\delta$-indifference condition” (C) to define the total sampling cost $C_T(\Lambda_J)$. This cost function is easily amenable to the Bayesian treatment of the hyperparameter selection problem. Huang, Li and Yao (2020, p.202) remark that offline parameter tuning “usually requires a large number of runs of the (metaheuristic optimization) algorithm.”
2.3 Efficiency theory of adaptive hyperparameter tuning

to analyze its performance on one instance, or a set of parameter instances with different parameter settings.” To paraphrase their remark in statistical jargon, they essentially consider a family of prior distributions, indexed by a hyperparameter vector $\lambda \in \Lambda$, on the optimal tuning parameter, and compute a dictionary of Bayes procedures $B(\lambda)$, $\lambda \in \Lambda$, whose performance is evaluated on a given problem to find the best one. This Bayesian perspective provides a way to circumvent the “time-consuming” disadvantage of offline parameter tuning; see the Supplementary Material S2 after the proof of Theorem 1.

The $\delta$-indifference zone is commonly used in the probability of correct selection (PCS) constraint on the selected population/method in the selection and ranking literature. A general formulation of PCS for the case of population means is $P(\mu^D > \mu^* - \delta) \geq 1 - \alpha$ for all $\mu^1, \ldots, \mu^K$, where $\mu^D$ denotes the $\mu^j$ selected; see Eq. (1.3) of Chan and Lai (2006, p.181) whose Sections 3 and 4 consider the case of $\inf_{x \leq 0} g(x) > 0$ in the definition of the total sampling cost $C_T$ for the one-parameter exponential family and the asymptotic optimality of $\mu^D$ in this case. Theorem 2 extends Theorem 1 to the PCS formulation; its proof also uses multi-armed bandit theory and is given in supplementary material S1, where the theory underlying $\epsilon$-greedy randomization in Section 2.2 will also be provided. In the current setting
of group sequential hyperparameter tuning, let $\lambda_\theta$ denote the optimal hyperparameter vector when $\theta$ is the parameter of the exponential family of densities for $Z_t$; precise details will be given in the proof of Theorem 2 in Supplementary Material S1.

**Theorem 2.** Among all group sequential procedures that satisfy the PCS constraint $P_\theta(\lambda_\theta \in \Lambda_J) \geq 1 - \alpha$ for all $\theta$, the method yielding $\Lambda_J$ in part B of Section 2.1 has asymptotically minimal $EC_T(\Lambda_J)$ as $\alpha \to 0$; moreover, $EC_T(\Lambda_J) = O(\log \alpha^{-1})$.

In Theorems 1 and 2, $EC_T(\Lambda_J)$ refers to the expectation under the actual probability measure generating $Z_t$. As pointed out in Step A3 of Section 2.1, $\hat{E}$ is an estimate of $E$ based on $B$ simulated samples $Z^{(b)}_t$, $b = 1, \ldots, B$. The sampling variability of $\hat{E}\Delta_j(\lambda)$ in Step B3 of Section 2.1 is why multi-armed bandit theory (implemented via the $\epsilon$-greedy randomization scheme) is needed. The reason why $E\Delta_j(\lambda)$ is not used directly in this step is that the expectation is usually difficult to compute except by Monte Carlo simulations. Moreover, the distribution of $Z_t$ often involves the unknown parameter $\theta$, which has to be estimated sequentially from the observed data up to time $t_j$ ($j = 1, \ldots, J$).
3. PSO and Locally D-Optimal Designs

Given a statistical model, an optimization criterion and the total number \( n \) of observations allowed for the study, consider the problem of finding continuous designs that optimize the criterion. Continuous designs, introduced by Kiefer and Wollowitz \( (1960) \), can be viewed as probability measures defined on the design space; see Atkinson, Donev and Tobias \( (2007) \) and Pukelsheim \( (2006) \). If a continuous design has \( k \) points with weight \( w_i \) at the design point \( x_i, i = 1, \ldots, k \), we implement it by taking \( \lceil nw_i \rceil \) observations at \( x_i, i = 2, \ldots, k \), where \( \lceil nw_i \rceil \) is the nearest rounded integer of \( nw_i \) subject to \( \lceil nw_1 \rceil + \ldots + \lceil nw_k \rceil = n \). Continuous optimal designs are appealing because there is a unified theory for checking whether a continuous design is optimal among all designs; and if not, the theory provides an assessment of its proximity to the optimum without knowing the latter. Although explicit formulas are available for relatively simple models with few regressors, there are no analytical descriptions for the optimal designs for more complex settings and numerical methods must be used. There are algorithms for finding optimal continuous designs; some are ad hoc and some can be shown to converge in theory to the optimum. However, the algorithms can be very slow and get stalled during the search. Others such as the Fedorov-type algorithms require intermittent collapsing of clusters.
of points into a design point. Some also require that the design space be discretized and do not work well for models with many regressors.

For nonlinear regression models, the Fisher information matrix involves the unknown parameter vector $\theta$, hence the D-optimal design that minimizes the logarithm of the determinant of the inverse of the Fisher information matrix requires specification of $\theta$ which the design aims to estimate. To circumvent this circuitous difficulty, Chernoff (1953) introduced the concept of “locally D-optimal design” which replaces $\theta$ by a nominal value arising from the design objective (such as hypothesis testing) or a pilot study; see Atkinson, Donev and Tobias (2007). Federov (1972) introduced an “exchange algorithm” which was recently refined by Huang et al. (2019) into the “point exchange” and “coordinate exchange” algorithms PEA and CEA. In this section we apply the systematic group sequential selection of $\theta$, introduced in Section 2.1 to address the issue of users of metaheuristic optimization algorithms such as PSO being “unlucky with the choice of tuning constants”, mentioned by Huang et al. (2019), to find optimal designs in high-dimensional nonlinear regression models in Section 3.1. For these applications, the distributions of the stochastic components $Z_t$ in Section 2.1 are not completely specified because they depend on the unknown $\theta$, which can be estimated sequentially at times $t_1 < \cdots < t_J$. 
from the observed data in Section 3.2. We estimate the distributions of $Z_t$
directly by applying the bootstrap to these data.

### 3.1 Enhanced PSO with adaptively tuned hyperparameters

Particle swarm optimization (PSO), proposed by [Kennedy and Eberhart (1995)] as a nature-inspired optimization method, was developed from the
model of a swarm of flying birds collaborating to search for food on the
ground. Each bird has its opinion of the food’s position and the birds com-
municate with one another their findings to determine collectively where
the food is. Thus there are two types of positions called respectively the
“personal best” position and the “global best” position found by the flock
to date. Velocities and locations of the birds are updated iteratively. For
the next iteration, each bird flies in a direction that takes into account of
(i) its current direction, (ii) its current known personal best position (cog-
nitive component), and (iii) the flock’s current best known position (social
component). Extending the “birds” to “collaborating particles” to mini-
mize a loss function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, PSO denotes the location (respectively, velocity) of the
$i$th particle at the $t$th iteration by $x_i(t)$ (respectively, $v_i(t)$) and
defines

$$ x_i^*(t) = \arg\min_{0 \leq s \leq t} f(x_i(s)), \quad \mathbf{x}^*(t) = \arg\min_{1 \leq j \leq n, 0 \leq s \leq t} f(x_j(s)), \quad (3.1) $$
which represent the personal best location found by the $i$th particle and the
global best location found by the swarm so far, respectively. We consider
here the following enhancement of the classical PSO algorithm:

$$
\begin{align*}
x_i(t+1) &= \chi_D(x_i(t) + \eta v_i(t + 1)), \\
v_i(t + 1) &= (1 - \eta \omega)v_i(t) + \eta c_1 U_{1,i}(t) \circ (x_i^*(t) - x_i(t)) \\
&\quad + \eta c_2 U_{2,i}(t) \circ (x_i^*(t) - x_i(t)) + \eta Z_i(t + 1),
\end{align*}
$$

where $\chi_D$ is the projection onto a bounded region $D$ that is known to contain
the minimizer $x^*$ of $f$ in its interior, $\circ$ denotes the Hadamard product,
$U_{j,i}(t)$ have i.i.d. components with finite second moments for $j = 1, 2$, and
$Z_i(t + 1)$ have i.i.d. zero-mean random vectors in $\mathbb{R}^d$ with finite second
moments and are independent of $U_{1,i}(t)$ and $U_{2,i}(t)$.

The classical PSO algorithm uses $\eta = 1$, does not include the term
$\eta Z_i(t + 1)$ in (3.3), and assumes the components of $U_{1,i}(t)$ and $U_{2,i}(t)$ to be
i.i.d. Unif(0,1). Choosing $c_1 = c_2 = c$, the tuning parameter vector in this
case is $(\eta, \omega, c)$, in which $\eta$ is the step-size, whereas $1 - \eta \omega$ and $c$ are positive
weights. We can reduce the dimensionality of the search to 2 (instead of 4)
modifying part B in Section 2.1 as follows: Use a fixed value $\eta_+ = 0.95$ of $\eta$
for the initial stages at user-selected times $t_1, \ldots, t_J$. Concerning the choice
of $\eta$ for the later stages $j > J$, the strategy is to choose $(\omega, c)$ first and

\begin{align*}
x_i(t+1) &= \chi_D(x_i(t) + \eta v_i(t + 1)), \\
v_i(t + 1) &= (1 - \eta \omega)v_i(t) + \eta c_1 U_{1,i}(t) \circ (x_i^*(t) - x_i(t)) \\
&\quad + \eta c_2 U_{2,i}(t) \circ (x_i^*(t) - x_i(t)) + \eta Z_i(t + 1),
\end{align*}

where $\chi_D$ is the projection onto a bounded region $D$ that is known to contain
the minimizer $x^*$ of $f$ in its interior, $\circ$ denotes the Hadamard product,
$U_{j,i}(t)$ have i.i.d. components with finite second moments for $j = 1, 2$, and
$Z_i(t + 1)$ have i.i.d. zero-mean random vectors in $\mathbb{R}^d$ with finite second
moments and are independent of $U_{1,i}(t)$ and $U_{2,i}(t)$.

The classical PSO algorithm uses $\eta = 1$, does not include the term
$\eta Z_i(t + 1)$ in (3.3), and assumes the components of $U_{1,i}(t)$ and $U_{2,i}(t)$ to be
i.i.d. Unif(0,1). Choosing $c_1 = c_2 = c$, the tuning parameter vector in this
case is $(\eta, \omega, c)$, in which $\eta$ is the step-size, whereas $1 - \eta \omega$ and $c$ are positive
weights. We can reduce the dimensionality of the search to 2 (instead of 4)
modifying part B in Section 2.1 as follows: Use a fixed value $\eta_+ = 0.95$ of $\eta$
for the initial stages at user-selected times $t_1, \ldots, t_J$. Concerning the choice
of $\eta$ for the later stages $j > J$, the strategy is to choose $(\omega, c)$ first and
3.1 Enhanced PSO with adaptively tuned hyperparameters

then to determine $\eta$. Without introducing additional notation, we simply let $\lambda = (\omega, c)$ and replace the $A(\lambda)$ in Section 2.1 by $A(\lambda, \eta)$, and the $x_t(\lambda_j)$ and $x_t^*(\lambda_j)$ by $x_t(\lambda_j, \eta_j)$ and $x_t^*(\lambda_j, \eta_j)$. In this way, we can replace part B of Section 2.1 by the following steps to update $\lambda_j, \Lambda_j$ and $\eta_j$ at $t_j$ ($1 \leq j \leq J$), after initializing at time $t = 1$ by choosing $\eta = \eta_+,$ $\lambda_0$ from $\Lambda$ according to some given distribution, and $\Lambda_0 = \{\lambda_0\}$:

(i) Choose $\lambda^*$ from $\Lambda$ according to some given distribution and let $\Lambda_j^* = \Lambda_{j-1} \cup \{\lambda^*\}.$

(ii) Letting $\Delta_j(\lambda, \eta) = f(x^*_{t_{j-1}}(\lambda, \eta)) - f(x^*_t(\lambda, \eta))$, compute $\hat{E}\Delta_j(\lambda, \eta_+)$ for $\lambda \in \Lambda_j$.

(iii) Let $\lambda^*_j = \arg\max_{\lambda \in \Lambda_j^*} \hat{E}\Delta_j(\lambda, \eta_+)$ and carry out the $\epsilon$-greedy randomization scheme to define $\lambda_j$ and $\Lambda_j$.

(iv) Switch to a smaller, adaptively chosen step-size at stage $I$ defined by

$$I = \inf\{2 \leq j \leq K : \frac{\hat{E}\Delta_j(\lambda_j, \eta_+)}{\hat{E}[f(x^*_{t_{j-1}}(\lambda_j, \eta_+)) - f(x^*_t(\lambda_j, \eta_+))]} \leq \delta\},$$

(3.4)

with $\inf \emptyset = K$. Here $\delta$ is user-selected and $K + 1$ is a prespecified upper bound on the size of $\Lambda_j$. The basic idea underlying the initial stages (up to stage $I$) is to use a larger step-size $\eta_+$ to attain fast descent of the expected loss from $\hat{E}f(x^*_t(\cdot))$ prior to time $t_1$. In
(3.4), $\delta$ is the threshold that signals relatively small incremental improvement at stage $j$, hence we switch to a smaller $\eta$ suggested by the theory in Section 2.3.

(v) For updates at times $t_j$ with $j > I$, carry out Step (i) to generate $\lambda^*$ and define $\Lambda_j^* = \Lambda_{j-1} \cup \{\lambda^*\}$. Let $\lambda_j^* = \arg\max_{\lambda \in \Lambda_j^*} \hat{E}\Delta_j(\lambda, \eta_{j-1})$. Then use the $\epsilon$-greedy randomization scheme to define $\lambda_j$ and $\Lambda_j$. Let $\eta_j = \arg\max_{\eta \leq 0.95} \hat{E}\Delta_j(\lambda_j, \eta)$.

3.2 Locally optimal designs in continuation-ratio model

Fan and Chaloner (2004) consider locally D-optimal designs for trinomial response in the regression model defined by

$$\log \left( \frac{\pi_3}{1 - \pi_3} \right) = \theta_1 + \theta_2 x, \quad \log \left( \frac{\pi_2}{\pi_1} \right) = \theta_1 + a + bx, \quad (3.5)$$

where $a \geq 0, \theta_2 > 0, b > 0$. This is the continuation-ratio model, in contrast to the “proportional odds model” that replaces $\log(\pi_2/\pi_1)$ by $\log ((\pi_2 + \pi_3)/\pi_1)$ and $b$ by $\theta_2$ in the second equation of (3.5). In particular, locally D-optimal designs $\xi_D$ maximize $\log \det(I_\xi)$ of the Fisher information matrix $I_\xi$ over design measures $\xi$, and can be “found numerically” to have a fixed number (which increases with $a$) of design points. Locally c-optimal designs $\xi_\psi$ minimize the asymptotic variance of the MLE.
of a real-valued function $\psi$ of the parameters and can be found similarly by numerical methods that need to address additional issues such as singularity of the asymptotic covariance matrix, as explained in Fan and Chaloner (2004, pp.352-354) who also specify the choice of the “function of interest” for the c-optimality criterion. They consider the maximum tolerated dose (MTD), which is the highest dose $x_T$ that produces a user-specified proportion $p$ of subjects with dose-limiting toxicity (DLT) in dose-response studies, i.e., $\pi_3(x_T; \theta_1, \theta_2) = p$ in the continuation-ratio model (3.5). Solving this equation yields $x_T = (\logit p - \theta_1)/\theta_2$, hence $\nabla x_T(\theta_1, \theta_2) = (-1/\theta_2, \theta_2^{-2}(\theta_1 - \logit p))^\top$, and the locally c-optimal design minimizes the asymptotic variance of the MLE of $\psi(\theta_1, \theta_2) = \log(\nabla x_T(\theta_1, \theta_2) I^-_\xi(\theta_1, \theta_2) \nabla x_T(\theta_1, \theta_2))$ for dose-response studies of the MTD, where $I^-_\xi$ is the generalized inverse of the Fisher information matrix $I_\xi$ which may be singular. Denote this locally c-optimal design by $\xi_T$.

The proportional odds model mentioned at the beginning of the preceding paragraph was introduced by Thall and Russell (1998) for dose-finding “based on efficacy and adverse outcomes” in Phase I/II oncology trials. They propose a trinomial outcome with the outcome 0 representing “no toxicity and no efficacy”, outcome 1 representing moderate toxicity and outcome 2 representing severe toxicity in the bone marrow transplant
3.2 Locally optimal designs in continuation-ratio model

treatment of lymphoma. Instead of using this proportional odds model which Fan and Chaloner (2004, p.349 and Fig.1) find unlikely to be valid and the continuation-ratio model to fit actual data “much better”, we use the continuation-ratio model to define not only MTD but also MED (most efficacious dose), which is defined as the dose with the highest probability of efficacy without severe toxicity, as illustrated by Thall and Russell (1998, Fig.1) for the outcome of moderate GVHD (graft-versus-host disease) and no severe toxicity. In the continuation-ratio model for toxicity and efficacy, MED is defined as the maximizer \( x_E \) of

\[
\pi_3 := \frac{1}{1 + e^{\theta_3 + \theta_4 x}},
\]

where \( \pi_3 \) is given by the first equation of (3.5) which is equivalent to

\[1 - \pi_3 = \left(1 + e^{\theta_1 + \theta_2 x}\right)^{-1}, \quad \theta = (\theta_1, \theta_2, \theta_3, \theta_4), \]

and \( \theta_3 + \theta_4 x \) is the analog of \( \theta_1 + \theta_2 x \) for the logarithm of the odds ratio for the probability of efficacy, with \( \theta_4 < 0 \) (similar to \( \theta_2 > 0 \)) because a dose with severe toxicity is no longer efficacious. Solving \( (d/dx) \log p(x; \theta) = 0 \) yields \( x_E(\theta) \) from which we can (a) use the delta method to derive the asymptotic variance of the MLE of \( x_E(\theta) \), (b) apply the implicit function theorem to evaluate \( \nabla x_E(\theta) \), and (c) derive the locally c-optimal design \( \xi_E \) that minimizes \( \psi(\theta) := \log \left( \nabla^T x_E(\theta) I_{\xi}(\theta) \nabla x_E(\theta) \right) \).

Clyde and Chaloner (1996) considered a compound optimality criterion for designs to estimate both MTD and MED, extending earlier work of
Cook and Wong (1994) in this direction. We consider here a more general compound optimality criterion

\[
\Psi(\xi; \lambda) = \lambda_1 \Psi_T(\xi) + \lambda_2 \Psi_E(\xi) + \lambda_3 \Psi_D(\xi),
\]

with nonnegative \( \lambda_i \) that sum to 1. Section 4 will regard \( \lambda = (\lambda_1, \lambda_2, \lambda_3) \) as a tuning parameter such that \( \lambda_3 = 0 \) and \( \lambda_1 + \lambda_2 = 1 \), and apply the ideas in Section 2.1 to the multi-objective optimization problem that is implicit in the compound optimization criterion (3.6) as an objective function, which we explain below. As pointed out by Chen, Heyse and Lai (2018, Section 3.3 and pp.90-91), multi-objective optimization typically involves conflicting objectives, such as benefit versus risk (or efficacy versus toxicity) of a treatment. Let \( f : S \to \mathbb{R}^m \) be a vector-valued function such that \( f_i(\mathbf{x}) \) represents the \( i \)th objective function in a multi-objective optimization problem, where \( S \subset \mathbb{R}^d \). We say that \( \mathbf{x}' \) dominates \( \mathbf{x} \) if \( f_i(\mathbf{x}') \geq f_i(\mathbf{x}) \) for every \( i = 1, \ldots, m \), with strict inequality for some \( i \), and that \( \mathbf{x} \) is “Pareto optimal” if there does not exist \( \mathbf{x}' \in S \) that dominates it. If \( \mathbf{x} \) is a random variable, then the \( f_i \) are expected functionals of \( \mathbf{x} \). The set of Pareto optimal elements of \( S \) is called the “Pareto boundary”, which is the solution of the multi-objective optimization problem.

**Example 1.** As an illustration, Figure 1 in the Supplementary Material S3 plots the Pareto surface of the multi-objective optimization problem of
minimizing the compound optimality criterion (3.6), for which

\[
\begin{align*}
    f_1(\xi; \lambda) &= \lambda_1 \left\{ \log \left( \nabla^T x_T I_{\xi T} \nabla x_T \right) - \log \left( \nabla^T x_T I_{\xi} \nabla x_T \right) \right\}, \\
    f_2(\xi; \lambda) &= \lambda_2 \left\{ \log \left( \nabla^T x_E I_{\xi E} \nabla x_E \right) - \log \left( \nabla^T x_E I_{\xi} \nabla x_E \right) \right\}, \\
    f_3(\xi; \lambda) &= (1 - \lambda_1 - \lambda_2) \left( \log \det(I_{\xi D}) - \log \det(I_{\xi}) \right).
\end{align*}
\]

Here we consider the locally D-optimal design \(\xi_D\) that maximizes \(\log \det(I_{\xi})\) at \((\theta_1, \theta_2, \theta_3, \theta_4) = (-3.3, 0.5, -3.4, -1)\) in (3.5) and the locally c-optimal designs \(\xi_T\) (for MTD, with \(p = 0.3\)) and \(\xi_E\) (for MED). Including these designs for the individual criteria in defining \((f_1, f_2, f_3)^T\) amounts to subtracting the constant \(\lambda_1 \Psi_T(\xi_T) + \lambda_2 \Psi_E(\xi_E) + \lambda_3 \Psi_D(\xi_D)\) from (3.6), which does not change the Pareto optimal boundary. Note that minimizing \(f_i\) corresponds to maximizing \(-f_i\), hence \(f_3\) considers \(-(1-\lambda_1-\lambda_2)p^{-1}(\log \det(I_{\xi}) - \log \det(I_{\xi_D}))\). For given \(\lambda_1, \lambda_2 \in [0, 1]\) with \(0 < \lambda_1 + \lambda_2 < 1\), we can use the equivalence theorem to show that the optimal compound design is a weighted sum of the optimal designs \(\xi_T, \xi_E\) and \(\xi_D\) under the individual criteria, thereby reducing the problem to a single-objective design after determining the weights.

**Refinement 1.** Consider the compound optimality criterion (3.6) with \(\lambda_3 = 0\) and \(\lambda_2 = 1 - \lambda_1\), reducing the problem to only finding \(\xi_T\) and \(\xi_E\).

**Example 2.** For the parameter configuration \((\theta_1, \theta_2, \theta_3, \theta_4)\) in Example
3.2 Locally optimal designs in continuation-ratio model

1, consider the compound criterion (3.6) with $\lambda_3 = 0$ and $\lambda_2 = 1 - \lambda_1$. Figure 2 in the Supplementary Material S3 plots the Pareto curve of the two-objective compound criterion in this case.

In applications such as those in Section 4, the complication is that $\lambda_1$ is unspecified and so is $\mathbf{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$, which is needed to define the locally c-optimal designs $\xi_T$ and $\xi_E$. We update the values of $\lambda$ and $\mathbf{\theta}$ at times $t_1 < \cdots < t_J$, with $J = \max\{j \geq 1 : t_j < T\}$ and $T$ being the time that Phase II of the trial begins, applying Section 2.1 to the metaheuristic optimization algorithm PSO. The PSO recursions (3.1)–(3.3) can be used to compute efficiently, as explained below, an optimal discrete set $D_j$ of doses at time $t_j$ when the information set $\mathcal{F}_{t_j}$ generated by the doses and toxicity-efficacy outcomes up to that time is used to update $\mathbf{\theta}, \lambda$ and $D_j$. First we explain how, for given $\mathbf{\theta}$, PSO can be used to compute the Pareto boundary for optimizing the compound criterion (3.6) with $\lambda_3 = 0$ and $\lambda_1 = \lambda$. The directional derivative of the design $\xi_{\lambda}$ that minimizes $\lambda \Psi_T(\xi) + (1 - \lambda) \Psi_E(\xi)$, in the direction off the design measure $\delta_x$ degenerate at $x$ is given by

$$
\lambda \text{tr} \left\{ I_{\delta_x}(\mathbf{\theta}) I_{\xi_{\lambda}}(\mathbf{\theta}) (\nabla x_T(\mathbf{\theta}) \nabla x_T(\mathbf{\theta})^\top) I_{\xi_{\lambda}}^{-1}(\mathbf{\theta}) \right\} / \text{tr} \left\{ (\nabla x_T(\mathbf{\theta}) \nabla x_T(\mathbf{\theta})^\top) I_{\xi_{\lambda}}^{-1}(\mathbf{\theta}) \right\}
+ (1 - \lambda) \text{tr} \left\{ I_{\delta_x}(\mathbf{\theta}) I_{\xi_{\lambda}}(\mathbf{\theta}) (\nabla x_E(\mathbf{\theta}) \nabla x_E(\mathbf{\theta})^\top) I_{\xi_{\lambda}}^{-1}(\mathbf{\theta}) \right\} / \text{tr} \left\{ (\nabla x_E(\mathbf{\theta}) \nabla x_E(\mathbf{\theta})^\top) I_{\xi_{\lambda}}^{-1}(\mathbf{\theta}) \right\}. 
$$

(3.7)

The equivalence theorem for optimal designs (Pukelsheim, 2006) yields that $\xi_{\lambda}$ is optimal for (3.6) if and only if (3.7) is bounded above by 1 for all $x \in [x_{\min}, x_{\max}]$, with equality at the support points of $\xi_{\lambda}$. For each value $\lambda$, carry out PSO to minimize $\lambda \Psi_T(\xi) + (1 - \lambda) \Psi_E(\xi)$, yielding the mini-
mizer $\xi_\lambda$; the Pareto boundary is \{\$\xi_\lambda : 0 < \lambda < 1\$. Note that $m = 2$ and $S \subset [x_{\text{min}}, x_{\text{max}}]$ in this case. Actually we only need to approximate the Pareto boundary over a discrete grid $D_J$ for the doses in Phase II, with $\lambda$ belonging to a discrete subset $\Lambda_J$ of $(0, 1)$, rather than the entire Pareto boundary \{\$\xi_\lambda : 0 < \lambda < 1\$. Proceeding as in Section 2.1 to generate for the PSO algorithm a discrete set $\Lambda_j$ of hyperparameter values $\lambda$ and the corresponding $\xi_\lambda, I_{\xi_\lambda}(\theta_j)$ and the tuning parameters $\omega, c$ and $\eta_j$ of PSO described in the second paragraph of Section 3.1, we choose \(Z_{1}(s), \ldots, Z_{n}(s)\)^T in (3.3) for $t_j < s \leq t_{j+1}$ to be i.i.d. replicates of the random vector in (3.7) with $\theta = \theta_j$ and $\lambda \in \Lambda_j$; see parts A and B of the basic algorithm in the second paragraph of Section 2.1.

Concerning the choice of $\theta_j$, one can use the `glm` function in CRAN to fit generalized linear models by maximum likelihood and then substitutes $\theta_j$ in $Z(s) := (Z_{1}(s), \ldots, Z_{n}(s))^T$ for $t_j < s \leq t_{j+1}$ by $\hat{\theta}_j$. This has the drawback of neglecting the sampling validity of $\hat{\theta}_j$ in carrying out the multi-armed bandit scheme (part B of Section 2.1) to update the set $\Lambda_j$ of selected hyperparameters. We use the following better approach.

**Refinement 2 (bootstrap).** [Liu 1988] introduced bootstrap resampling, which “is known to be a good general procedure for estimating a sampling
4 OPIMAL SEQUENTIAL DESIGN OF DOSE-FINDING TRIALS

distribution under i.i.d. models”, for independent but non-identically distributed data, focusing in particular on the setting of $Z(1), \ldots, Z(T)$ drawn from distributions $G_1, \ldots, G_T$ and on the regression case that $Z(t)$ depends on covariates. For i.i.d. data, the population distribution $G$ can be approximated by empirical distribution $\hat{G}$ that puts weight $1/T$ at each of the observed $Z(t)$, and the bootstrap resamples $Z^*(t)$ from $\hat{G}$. For the independent $Z(s)$ that are identically distributed if $s \in \{t_j + 1, \ldots, t_{j+1}\}$, the bootstrap procedure resamples $Z^*(s)$ from the empirical distribution that puts weight $(t_{j+1} - t_j)^{-1}$ at each item of the observed sample $\{Z(t) : t_j < t \leq t_{j+1}\}$, which is basically the regression case discussed in Section 4B of [Liu (1988)].

4. Optimal Sequential Design of Dose-Finding Trials

The locally D-optimal and c-optimal designs for the continuation-ratio or proportional odds model in Section 3.2 are related to design of dose-finding studies of treatments that have dose-limiting toxicities, e.g., cytotoxic chemotherapies for cancer patients, for which risk-benefit modeling and analysis play an important role as discussed in the last paragraph of the preceding section. In practice, these dose-finding trials are early-phase open-label clinical trials and there are ethical, informed consent, and sam-
ple size constraints in giving experimental doses to patients accrued to a trial. In a special issue on this topic in *Statistical Science* 25(2) in 2010, overviews of the progress and emerging trends in the design of Phase I (or Phase I/II) trials to date were presented. In Section 4.1, we (a) summarize these methods and subsequent developments in the past decade, (b) integrate them with locally c-optimal designs for MTD and MED estimation and weighting them in the compound optimality criterion (3.6) with $\lambda_3 = 0$ so that $\lambda_2 = 1 - \lambda_1$ as in Example 2, and (c) formulate an optimal adaptive choice of the weight $\lambda_1$ (as a tuning parameter). In Section 4.2, we use the methods in Sections 2.1, 2.2, 3.2 and 4.1 to develop an optimal sequential design of early-phase trials to determine the dose for a late-phase confirmatory trial of the treatment. We also give some concluding remarks.

### 4.1 Model-based sequential designs of dose-finding trials

Cheung (2010) contrasts model-based designs which “make dose decisions based on the explicit use of dose-toxicity models” with commonly used “algorithm-based designs whereby a set of dose-escalation rules are pre-specified for any given dose”. He also highlights the difference between dose-finding studies in bioassays of laboratory animals and those in Phase I clinical trials involving human subjects, for whom he lists the ethical con-
4.1 Model-based sequential designs of dose-finding trials

A group sequential design $\xi$ with group size $m$ and dose $x_i$ is called coherent if there is a threshold $p$ such that

$$P_\xi(x_i - x_{i-1} > 0|\tilde{Y}_{i-1} \geq p) = 0 \quad \text{and} \quad P_\xi(x_i - x_{i-1} < 0|\tilde{Y}_{i-1} \leq p) = 0 \quad (4.8)$$

for all $i$, where $\tilde{Y}_{i-1}$ is the observed proportion of toxicities in group $i-1$ and $P_\xi$ denotes the probability measure induced by the design $\xi$, for which (4.8) is tantamount to dose de-escalation if $\tilde{Y}_{i-1} \geq p$, and dose escalation if $\tilde{Y}_{i-1} \leq p$. Cheung says that “an algorithm-based design can explicitly incorporate dose decision rules that respect coherence”, such as the traditional $3 + 3$ design for which $p = 0.33$. He also mentions the desirable properties of consistency and unbiasedness (i.e., $P_\xi(x_T = d_k)$ is nonincreasing in $\pi(d_i)$ for $i \leq k$, and nondecreasing in $\pi(d_j)$ for $j > k$, where $\pi(d_i)$ is the true toxicity probability at dose $d_i$, $\pi(d_k) = p$, and $x_T$ is the selected dose when the trial terminates after $T$ dose-finding moves).

O’Quigley and Conaway (2010) and Thall (2010) consider model-based sequential designs for Phase I or Phase I/II trials. Thall considers estimation of toxicity and efficacy responses for trinomial outcomes from the proportional odds model of Thall and Russell (1998) in a Phase I/II trial. O’Quigley and Conaway consider the continual reassessment model CRM (O’Quigley et al., 1990; Shen and O’Quigley, 1996) in Phase I trials to es-
timate MTD, or the more general MSD (most successful dose) presented
in their Section 7 and illustrated with HIV treatment for which “failure is
either a toxicity (that causes) inability to maintain treatment, or an unac-
ceptably low therapeutic response” so that the probability \( P(d_i) \) of success
is \( Q(d_i)(1 - R(d_i)) \), where \( R(d_i) \) is the probability of toxicity at dose \( d_i \)
and \( Q(d_i) \) is the probability of therapeutic response given no toxicity at
that dose. “A successful trial would identify the dose level \( \ell \) such that
\( P(d_\ell) > P(d_i) \) for \( i \neq \ell \)”, and “CRM can be readily adapted to address
these kinds of questions.”

[Tighiouart and Rogatko (2010) and Bartroff and Lai (2010)] consider
model-based sequential designs of Phase I trials involving the Bayesian
model EWOC (Escalation With Overdose Control) introduced by [Babb,
Rogatko and Zacks (1998)]. Both papers consider MTD estimation and
choice of prior distribution. Whereas Theorem 2.2 of [Tighiouart and Ro-
gatko (2010)] gives conditions on a reparameterized prior distribution to
ensure coherence of EWOC, [Bartroff and Lai (2010)] incorporate the risk
to patients in the trial with a “global risk” \( E[\sum_{k=1}^{T} h(x_k, \eta) + g(\hat{\eta}_T, \eta)] \),
where \( \eta \) denotes the MTD and \( \hat{\eta}_T \) is the MTD estimate at the termi-
nation of the trial, \( g(\hat{\eta}_T, \eta) \) measures the effect of \( \hat{\eta}_T \) on future patients,
4.1 Model-based sequential designs of dose-finding trials

\[ h(x, \eta) = \omega(\eta - x)_+ + (1 - \omega)(x - \eta)_+ \]
as the EWOC doses are the \( \omega \)th quantiles of the posterior distributions of \( \eta \). They use the rollout algorithm in approximate dynamic programming to minimize the global risk; the idea of rollout is “to approximate the optimal policy \( x_k^* \) by the minimizer (over \( x \)) of \( h_{k-1}(x) + E[\sum_{i=k+1}^{T} h_i(\hat{x}_i)|\mathcal{F}_{k-1}, \hat{x}_k = x] \) (that replaces) \( x_{k+1}^*, \ldots, x_T^* \) by some base policy \( \hat{x}_{k+1}, \ldots, \hat{x}_T \), which ideally is some easily computed policy that is not far from the optimum”, where \( \mathcal{F}_{k-1} \) denotes the “information set generated by \((x_1, y_1), \ldots, (x_{k-1}, y_{k-1})\)“.

Bartroff, Lai and Narasimhan (2014) propose a novel group sequential design of early-phase clinical trials for cytotoxic chemotherapies. They note that “much (recent) work on early phase cancer trials incorporate both toxicity and efficacy data” but “do not explicitly address the Phase II hypothesis test of \( H_0 : p \leq p_0 \), where \( p \) is the probability of efficacy at the estimated MTD dose \( \hat{\eta} \) and \( p_0 \) is the baseline efficacy rate.” Their new design “addresses the uncertainty in the estimate \( p = p(\hat{\eta}) \) in \( H_0 \) by using sequential generalized likelihood ratio theory” and “can be used all the way from the first dose of Phase I through the final accept/reject (go/no go) decision about \( H_0 \) at the end of Phase II, utilizing both toxicity and efficacy data throughout” and allowing for “early stopping to show treatment
effect or futility” in Phase II hypothesis testing. In the next subsection, we integrate this idea with those in the preceding paragraphs of this subsection and in the last paragraph of Section 3.3 to formulate a new optimal group sequential model-based design of early-phase clinical trials of a cytotoxic chemotherapy to decide if the treatment should proceed to Phase III for confirmatory testing and the dose to be used if the decision is positive. As in [Bartroff and Lai (2010)], we use the EWOC scheme for Phase I to determine the MTD and its coherence property established by Tighiouart and Rogatko (2010), rather than incorporate the risk to patients in the trial into the global risk of the preceding paragraph to circumvent that difficult dynamic programming problem. After Phase I, we use the compound optimality criterion (3.6) with $\lambda_3 = 0$ and treat $\lambda_1 = 1 - \lambda_2$ as a tuning parameter for which we can use the ideas of Section 2.1 to find the optimal one, where Bayesian c-optimal designs are used to define optimality. Since the solution to the compound optimality criterion is a Pareto boundary, this entails loop optimization for which PSO in Section 3.2, with adaptively chosen tuning parameters, is particularly effective, as will be shown in the next subsection.
4.2 Efficient group sequential design of early-phase trials

We decompose the group sequential design into three stages, labeled as Phase I (for MTD estimation), Phase I/II (for finding a discrete set $D$ of doses, belonging to the Pareto boundary, that are c-optimal for estimating MED subject to probability constraints on DLT), and Phase II (for adaptively choosing the dose from $D$ and testing efficacy at the dose). Implementation details and the underlying rationale for each stage are given separately under Phase I, Phase I/II and Phase II.

**Phase I.** The groups refer to successive cohorts, each of size $m$, and the model-based design is EWOC which [Bartroff and Lai (2010)](Bartroff and Lai (2010)) reparameterize as follows and choose the dose for each cohort in the interval $[x_{\text{min}}, x_{\text{max}}]$ (without discretization) to circumvent rigidity. Assuming an upper bound $q > 0$ on the probability $\rho$ of toxicity at $x_{\text{min}}$, uniform distributions over $[x_{\text{min}}, x_{\text{max}}]$ and $[0, q]$ are taken as the prior distributions for $\eta$ (the MTD) and $\rho$, respectively. EWOC assumes a logistic regression model $1/(1 + e^{-(\alpha + \beta x)})$ for the probability of DLT at dose level $x$. Since the logistic regression parameters can be expressed in terms of $\eta$ and $\rho$ via

$$\alpha = (x_{\text{min}} \logit p - \eta \logit \rho)/(x_{\text{min}} - \eta), \quad \beta = (\logit \rho - \logit p)/(x_{\text{min}} - \eta),$$
4.2 Efficient group sequential design of early-phase trials

the $\mathcal{F}_t$-posterior density of $(\rho, \eta)$ is equal to

\begin{equation}
C^{-1} \prod_{i=1}^{t} \left[ e^{y_i \psi(x_i; \rho, \eta)} / (1 + e^{y_i \psi(x_i; \rho, \eta)}) \right] \text{ on } [x_{\min}, x_{\max}] \times [0, q],
\end{equation}

where $y_i$ is the binary indicator of DLT for subject $i$, $\psi(x; \rho, \eta) = \alpha + \beta x$

and

\begin{equation}
C = \int_{x_{\min}}^{x_{\max}} \int_{\eta_{min}}^{\eta_{max}} \prod_{i=1}^{t} \left[ e^{y_i \psi(x_i; \rho, \eta)} / (1 + e^{y_i \psi(x_i; \rho, \eta)}) \right] d\rho d\eta
\end{equation}

is the normalizing constant that can be determined by numerical evaluation

of a double integral using MATLAB or other software packages. Letting

$F(x, \eta)$ denote the probability of DLT at dose level $x$ when the MTD is

$\eta$ (at which the probability of DLT is $p$) in the logistic regression model,

Tighiouart and Rogatko (2010) Theorem 2.2 show that EWOC is coherent

if $F(x, \eta)$ in non-increasing in $\eta$ for fixed $x$, which is implicitly assumed by

Bartroff and Lai (2010) in their rollout algorithm to minimize the global

risk mentioned in the penultimate paragraph of Section 4.1 that explains

the role of the $\mathcal{F}_t$-posterior density (4.9) in the rollout algorithm. The

Phase I design here simply uses the coherence of EWOC and applies (4.9)

to the current cohort of $m$ subjects for determining the dose of the next

cohort as the $p$th quantile of the posterior distribution of $\eta$ given the binary
4.2 Efficient group sequential design of early-phase trials

indicators \( y_i \) of DLT in the current cohort; the \( p \)th quantile \( \eta_i \) is defined by

\[
\int_0^q \prod_{i=1}^t \left[ \frac{e^{y_i \psi(x_i; \rho, \eta_i)}}{1 + e^{y_i \psi(x_i; \rho, \eta_i)}} \right] \frac{d\rho}{C} = p,
\]

where \( \{(x_i, y_i) : 1 \leq i \leq m\} \) are the observations of the cohort and \( C \) is given by (4.10).

_Phase I/II._ This stage uses the toxicity and efficacy outcomes of the subjects treated at doses that are adaptively chosen by particle swarm optimization (PSO in Section 3.2) to minimize the compound optimality criterion (3.6) with \( \lambda_3 = 0 \) and adaptively chosen \( \lambda_1 = 1 - \lambda_2 \), assuming the continuation-ratio model for toxicity and efficacy outcomes as in Refinements 1 and 2 in Section 3.3.

_Phase II._ With the discrete subset \( D_j \) of doses determined in Phase I/II, we can proceed as in Section 3.2 of Bartroff, Lai and Narasimhan (2014), abbreviated by BLN for Bartroff, Lai, Narasimhan. For \( t_j > T \), we perform order-restricted GLR (generalized likelihood ratio) tests at the \( j \)th interim analysis. Let \( \pi(x) \) and \( p(x) \) denote the probability of DLT and that of efficacious response and no DLT, respectively, as in the first and second paragraphs of Section 3.3 in which the dependence of these probabilities on \( \theta = (\theta_1, \theta_2, \theta_3, \theta_4) \) is also highlighted. Relabeling the doses in the discrete set \( D_j \) with cardinality \( \nu \) as \( d_1 < d_2 < \cdots < d_\nu \), assume the order restrictions
4.2 Efficient group sequential design of early-phase trials

\[ \pi(d_1) \leq \cdots \leq \pi(d_\nu) \text{ and } p(d_1) \leq \cdots \leq p(d_\nu), \]
which yield order-restricted MLE of \( \theta \) (Section 2.2 of BLN). Taking advantage of the discrete dose set \( D_J \), Section 3.2 of BLN first introduces \( \pi_i(y, z) = P(y_t = y, z_t = z \mid x_t = d_i) \) to relate the binary indicator \( z_t \) of efficacious response without DLT for \( t_j < t \leq t_{j+1} \) when the dose \( x_t \) (determined at time \( t_j \) of the \( j \)th interim analysis) is \( d_i \). It then expresses the log-likelihood ratio at the \( j \)th interim analysis as

\[ \ell_j(\theta) = \sum_{t=1}^{t_j} \log \pi_{i(x_t)}(y_t, z_t), \]
where \( i(x_t) = d_i \) if \( x_t \) is assigned dose \( d_i \), and introduces \( \nu \) parameters \( \rho_i = \pi_{i(x_t)}(0,0)\pi_{i(x_t)}(1,1)/\{\pi_{i(x_t)}(1,0)\pi_{i(x_t)}(0,1)\} \), \( i = 1, \ldots, \nu \). This yields the log GLR statistic at the \( j \)th interim analysis, for testing efficacy of the selected doses up to that time, in terms of MLEs of \( \pi(d_1), \ldots, \pi(d_\nu), p(d_1), \ldots, p(d_\nu), \rho_1, \ldots, \rho_\nu \) subject to these order restrictions, and stopping rules in Section 3.1 of BLN, whose Section 4 (particularly Section 4.2) has provided extensive simulation studies showing the advantages of this approach. The software to implement Phase II, together with examples and data, are available at https://med.stanford.edu/cisd/research/software.html

In conclusion, nature-inspired metaheuristic optimization algorithms are important tools in machine intelligence (commonly called AI), which has increasingly permeated lives in modern society. Yet despite the recent
advances and promises of these algorithms, a long-standing open problem is how to tune them effectively to achieve maximal performance for particular systems and problems. We have made use of recent advances in Statistics to address this open problem in AI, and have demonstrated in Section 4 the usefulness of our solution in tackling challenging optimal design problems in early-phase oncology trials.

Acknowledgement

Choi’s research is supported by the Singapore MOE Academic Research Funds R-155-000-222-114. Lai’s research is supported by the National Science Foundation under DMS-1811818. Tong’s research is supported by MOE Academic Research Funds R-146-000-292-114. Wong’s research is supported by the National Institutes of Health under R01GM107639.

References


REFERENCES


Department of Statistics and Applied Probability, National University of Singapore

E-mail: choikp@nus.edu.sg
Department of Statistics, Stanford University

E-mail: lait@stanford.edu

Department of Mathematics, National University of Singapore

E-mail: mattxin@nus.edu.sg

Biostatistics Department, University of California, Los Angeles

E-mail: wkwong@ucla.edu
A STATISTICAL APPROACH TO ADAPTIVE PARAMETER TUNING IN NATURE-INSPIRED OPTIMIZATION AND OPTIMAL SEQUENTIAL DESIGN OF DOSE-FINDING TRIALS

Kwok Pui Choi, Tze Leung Lai, Xin T. Tong and Weng Kee Wong

National University of Singapore, Stanford University and UCLA

Supplementary Material

This is the supplementary material for the article “A statistical approach to adaptive parameter tuning in nature-inspired optimization and optimal sequential design of dose-finding trials”. All sections and equations refer to the main text.

S1 Proof of Theorem 2.

Section 1 of Chan and Lai (2006) reviews the literature on selection and ranking, and in particular the δ-difference zone approach [Bechhofer, Kiefer and Sobel [1968]], which we now apply to multi-armed bandits. For the one-parameter exponential family, UCB of an individual arm based on observations up to time $t$ from that arm is derived by inverting a GLR test; see Lai (1987, Section 2), where $\log \alpha^{-1} \sim \log B$ for the confidence level $1 - \alpha$ under δ-indifference and $B$ is the sample size in Theorem 1. The asymptotic lower bound for the regret that we have reviewed in Section 2.3 can be translated in terms of the total sampling cost $C_T$ since $\inf_x g(x) > 0$ is assumed in Theorem 2. Extension of this argument to the multiparameter exponential
family, which is the setting of Theorems 1 and 2, is straightforward, particularly since we use $\epsilon$-greedy sampling instead of upper confidence bounds (UCB) that Lai (1987) has defined only in the one-parameter case. Auer, Cesa-Bianchi and Fischer (2002, Theorem 3) have shown that the $\epsilon$-greedy sampling method introduced by Sutton and Barto (1998) can also attain the asymptotic lower bound for the regret.$^1$

S2 Proof of Theorem 1.

Section 5.2 of Chan and Lai (2006) consider the one-dimensional case of condition (C) but does not provide details$^2$ on how the procedure (which we have just described in the case $\inf_x g(x) > 0$) still yields under (C) “an asymptotically optimal selection procedure with expected total sampling cost of the order of $|\log \alpha|$”. Since condition (C) is the counterpart of the assumption $\inf_x g(x) > 0$ in Theorem 2, the preceding proof of Theorem 2 has already provided these details even for the multiparameter exponential family, thereby proving Theorem 1.

We now explain the “empirical Bayes hyperparameter tuning” refor-

---

$^1$The result stated in that paper is for time-varying $\epsilon_t$. It is also applicable to time-invariant $\epsilon$ when $t$ does not exceed a finite upper bound $T$, which is assumed in Theorems 1 and 2.

$^2$Chan and Lai (2006) only mention the asymptotic lower bound for the regret in multi-armed bandit problem or exponential families and refer to Lai (1987) for the UCB rule that attains the bound.
S2. PROOF OF THEOREM 1.3

Assume that $g$ satisfies condition (C) and that $Z_t$ are independent with density from a multiparameter exponential family. Let $\lambda_{opt}$ denote the optimal hyperparameter of a metaheuristic optimization algorithm to minimize $C_T(\lambda) = \sum_{j=1}^{J} g \left( x^* - x_t_j(\lambda) \right) \tau_j$ for a particular problem or system, and let $\Pi_\lambda$ be a prior distribution on $\lambda_{opt}$ so that $B(\lambda)$ is the Bayes rule that minimizes the Bayes risk $\int E C_T(\lambda) \, d\Pi_\lambda$. The Empirical Bayes (EB) approach to hyperparameter tuning uses empirical performance to choose $\lambda \in \Lambda$ for the Bayes rule $B(\lambda)$. In the group sequential setting of Section 2.1, an efficient group sequential EB hyperparameter tuning procedure is given by sequence $\{ \hat{\lambda}_1, \ldots, \hat{\lambda}_J \}$ of Theorem 1, which shows that the sequence has asymptotically minimal Bayes risk, of order $O(\log B)$ as $B \to \infty$, among all group sequential hyperparameter tuning procedures with Bayes risks of order $o(B^r)$ for any $r > 0$. 

mulation of Theorem 1 mentioned in the paragraph following the theorem.
S3 Figures for illustrations

Figure 1: Pareto surface for compound optimality criterion (3.6), with colormap (using Matplotlib in MATLAB) explaining the colors of the surface.

Figure 2: Pareto curve for (3.6) with $\lambda_3 = 0$ and known $\theta$.

Additional References
