Statistica Sinica **31** (2021), 2381-2401
doi:https://doi.org/10.5705/ss.202020.0412

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

Kwok Pui Choi[1], Tze Leung Lai[2], Xin T. Tong[1] and Weng Kee Wong[3]

[1]*National University of Singapore,* [2]*Stanford University and* [3]*University of California, Los Angeles*

Abstract: Nature-inspired metaheuristic algorithms have become increasingly popular in the last couple of decades, and now constitute a major toolbox for tackling complex high-dimensional optimization problems. Using group sequential experimentation, adaptive design, multi-armed bandits, and bootstrap resampling methods, this study develops a novel statistical methodology for efficient and systematic group sequential selection of the tuning parameters, which are widely recognized as 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 compute optimal experimental designs in nonlinear regression models, and is illustrated with solutions 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. 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 quantum particle swarm optimization (qPSO), Sun, Lai and Wu (2012, Chap. 5) provide a 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) recently emphasized the importance of finding appropriate tuning parameters in metaheuristic

Corresponding author: Kwok Pui Choi, Department of Statistics and Data Science, National University of Singapore, Singapore 117546. E-mail: choikp@nus.edu.sg.

optimization algorithms in their survey on the subject. They say that whereas heuristic optimization algorithms are problem-specific and implement 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 a "strong impact on the performance or efficiency of a metaheuristic." To address the increasing demand for systematic approaches for the metaheuristics' parameter setting, Section 2 develops a novel statistical methodology. In Section 3, we apply the methodology 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. Section 5 concludes the paper with further discussion about recent advances, some of which are related to adaptive platform trials for COVID-19 vaccine development, and provides additional references to the background literature.

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

Our novel approach to adaptive tuning parameter selection for metaheuristic optimization algorithms is presented in Section 2.1. The approach is developed from two statistical ideas, namely, the 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 for this adaptive parameter tuning approach for metaheuristic optimization is given in Section 2.3. As pointed out by Bartroff, Lai and Shih (2013, p.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." Then, a few years after the War, it was recognized 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 trials, particularly with respect to the incidence of treatment-related adverse events. One such trial was the Beta-blocker Heart Attack Trial, which

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 the adoption of group sequential designs. The major advances in group sequential methods are summarized in Chapter 4 of Bartroff, Lai and Shih (2013). Because statistical applications of metaheuristic optimization are often related to estimation, or hypothesis testing, of an unknown parameter $\boldsymbol{\theta}$, we use $\boldsymbol{\lambda}$ to denote the vector of tuning parameters in the metaheuristic optimization algorithm. An analog of $\boldsymbol{\lambda}$ is the hyperparameter of the prior distribution in Bayesian inference.

2.1. Group sequential learning of optimal $\boldsymbol{\lambda}$

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. In contrast to offline parameter tuning, surveyed by Huang, Li and Yao (2020), group sequential updating of the optimal tuning parameter vector is carried out during the running time T of the algorithm. Specifically, we assume that the optimal tuning parameter vector belongs to some bounded region of $\mathbb{R}^m$, and search for it at user-specified $t_1 < \cdots < t_J$, with $J = \max\{j \geq 1 : t_j < T\}$. 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 described below.

The metaheuristic optimization algorithm to search for $\mathbf{x}^*$ that minimizes f, with tuning parameter $\boldsymbol{\lambda}$, is denoted by $A(\boldsymbol{\lambda})$, and the sample path generated by $A(\boldsymbol{\lambda})$ is denoted by $\mathbf{x}_t(\boldsymbol{\lambda})$. Let $\mathbf{x}_t^*(\boldsymbol{\lambda}) = \text{argmin}_{s \leq t} f(\mathbf{x}_s(\boldsymbol{\lambda}))$. Initialize $\boldsymbol{\lambda}$ by choosing $\boldsymbol{\lambda}_0$ at random (or according to some given distribution) from the bounded region $\Lambda \subset \mathbb{R}^m$ to which the tuning parameter belongs. Let $\Lambda_0 = \{\boldsymbol{\lambda}_0\}$. Run $A(\boldsymbol{\lambda}_0)$ until $t_1 - 1$, and at time t_j $(j = 1, \ldots, J)$, update $\boldsymbol{\lambda}$ and run $A(\boldsymbol{\lambda}_j)$ with the updated value $\boldsymbol{\lambda}_j$ to generate $\mathbf{x}_t(\boldsymbol{\lambda}_j)$ and $\mathbf{x}_t^*(\boldsymbol{\lambda}_j)$, for $t_j \leq t < t_{j+1}$. Use the following procedures to handle the stochastic components $\mathbf{Z}_t$ in $\mathbf{x}_t(\boldsymbol{\lambda})$, for $\boldsymbol{\lambda} = \boldsymbol{\lambda}_j \in \Lambda_j$ and $t_j \leq t < t_{j+1}$, and to update the choice of $\boldsymbol{\lambda}_j$ and Λ_j at time t_j.

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

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

   A2. With the stochastic components $\mathbf{Z}_t^{(b)}$ of the metaheuristic optimization algorithm, generate B simulated samples of $(\mathbf{x}_{t,b}(\boldsymbol{\lambda}), \mathbf{x}_{t,b}^*(\boldsymbol{\lambda}))$, for $t_j \leq$

$t < t_{j+1}$ $(b = 1, \ldots, B)$.

A3. Taking the average of $f(\mathbf{x}^*_{t,b}(\boldsymbol{\lambda}))$ over $b = 1, \ldots, B$ yields an estimate, denoted by $\widehat{E}f(\mathbf{x}^*_t(\boldsymbol{\lambda}))$, of $Ef(\mathbf{x}^*_t(\boldsymbol{\lambda}))$, for $t_j \le t < t_{j+1}$.

B. How to update $\boldsymbol{\lambda}_j$ and Λ_j at t_j $(1 \le j \le J)$:

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

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

B3. Let $\boldsymbol{\lambda}^*_j = \operatorname{argmax}_{\boldsymbol{\lambda} \in \Lambda^*_j} \widehat{E}\Delta_j(\boldsymbol{\lambda})$. Sample $\boldsymbol{\lambda}_j$ from Λ^*_j with probability $1-\epsilon$ assigned to $\boldsymbol{\lambda}^*_j$ and probability ϵ/j assigned to each $\boldsymbol{\lambda} \in \Lambda^*_j \setminus \{\boldsymbol{\lambda}^*_j\}$; this is the ϵ-greedy randomization scheme, with user-specified $0 < \epsilon < 1/2$, explained in Section 2.2. Let $\Lambda_j = \Lambda_{j-1} \cup \{\boldsymbol{\lambda}_j\}$.

2.2. Multi-armed bandits and ϵ-greedy randomization

The multi-armed bandit problem, introduced by Robbins (1952) as a new direction for the nascent field of sequential design of experiments, subsequently evolved into an important area of reinforcement learning. Such learning combines active learning (also called "exploration"), to gather information about unknown system parameters, with passive learning (also called "exploitation") from the outputs, which the control system aims at driving toward some prescribed target; see Kaelbling, Littman and Moore (1996). Robbins (1952) considered two populations (arms) from which to sample sequentially 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 μ^* were known, then one would sample from it to receive the 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 $I = \max\{i : t_i \le T\}$, $t_I \to \infty$, but $t_I/T \to 0$. Subsequently Lai and Robbins (1985) extended multi-armed bandits from 2 to K arms, and provided a definitive theory by introducing the concept of *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 μ^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. In this case, 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

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

where $I(\cdot, \cdot)$ denotes the Kullback–Leibler information number. 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 ϵ-greedy randomization algorithm, which 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). Further discussion, including recent developments and additional references are given in the Supplementary Material S4.

2.3. Efficiency theory of adaptive hyperparameter tuning

We first use of the multi-armed bandit theory summarized in Section 2.2 to derive the optimality of the adaptive selection of the hyperparameter $\widehat{\boldsymbol{\lambda}}_j$ 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 the sampling cost in the selection, and the ranking literature on normal data or more general observations from exponential families, define the total sampling cost

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

for the metaheuristic optimization algorithm in Section 2.1. The group sequential updating of the tuning hyperparameter at times $t_1, \ldots, t_J$ leads to the set $\Lambda_J = \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_J\}$ of successive hyperparameter values to run the algorithm, as in part A of Section 2.1. The case

$$\text{(C)} \qquad g(\mathbf{x}) = 0 \ \text{ if } \ \|\mathbf{x}\| < \delta \quad \text{and} \quad \inf_{\|\mathbf{x}\| \geq \epsilon} g(\mathbf{x}) > 0 \ \text{ for } \ \epsilon > \delta$$

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 δ of the best.

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

Chan and Lai (2006) mention in their last paragraph of Section 5.2 the asymptotic lower bound for the regret of uniformly good adaptive allocation rules in the multi-armed bandit problem, which we reviewed in Section 2.2, and suggest 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 Λ_J within δ of the best by using the function g that satisfies the "δ-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 to analyze its performance on one instance, or a set of parameter instances with different parameter settings." In other words, they essentially consider a family of prior distributions, indexed by a hyperparameter vector $\boldsymbol{\lambda} \in \Lambda$, on the optimal tuning parameter, and compute a dictionary of Bayes procedures $B(\boldsymbol{\lambda})$, for $\boldsymbol{\lambda} \in \Lambda$, the performance of which 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 δ-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 the PCS for the case of population means is $P(\mu^D > \mu^* - \delta) \geq 1 - \alpha$, for all $\mu^1, \ldots, \mu^K$, where μ^D denotes the μ^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 μ^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 the Supplementary Material S1, where the theory underlying ϵ-greedy randomization in Section 2.2 is also provided. In the current setting of group sequential hyperparameter tuning, let

$\boldsymbol{\lambda}_{\boldsymbol{\theta}}$ denote the optimal hyperparameter vector when $\boldsymbol{\theta}$ is the parameter of the exponential family of densities for $\mathbf{Z}_t$. Precise details are given in the proof of Theorem 2 in the Supplementary Material S1, and extensions beyond the exponential family are provided in S4.

Theorem 2. *Among all group sequential procedures that satisfy the PCS constraint* $P_{\boldsymbol{\theta}}(\boldsymbol{\lambda}_{\boldsymbol{\theta}} \in \Lambda_J) \geq 1 - \alpha$ *for all* $\boldsymbol{\theta}$*, the method yielding* Λ_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 $\mathbf{Z}_t$. As pointed out in Step A3 of Section 2.1, $\widehat{E}$ is an estimate of E based on B simulated samples $\mathbf{Z}_t^{(b)}$, for $b = 1, \ldots, B$. The sampling variability of $\widehat{E}\Delta_j(\boldsymbol{\lambda})$ in Step B3 of Section 2.1 is why multi-armed bandit theory (implemented using the ϵ-greedy randomization scheme) is needed. The reason why $E\Delta_j(\boldsymbol{\lambda})$ 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 $\mathbf{Z}_t$ often involves the unknown parameter $\boldsymbol{\theta}$, 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 Wolfowitz (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, for $i = 1, \ldots, k$, we implement it by taking $[nw_i]$ observations at x_i, for $i = 2, \ldots, k$, where $[nw_i]$ is the nearest rounded integer of nw_i subject to $[nw_1]+\cdots+[nw_k] = n$. 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 formulae are available for relatively simple models with few regressors, there are no analytical descriptions for optimal designs for more complex settings, and thus 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 may stall 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 $\boldsymbol{\theta}$. Hence the D-optimal design that minimizes the logarithm of the determinant of the inverse of the Fisher information matrix requires a specification of $\boldsymbol{\theta}$ that the design aims to estimate. To circumvent this circuitous difficulty, Chernoff (1953) introduced the concept of "locally D-optimal design," which replaces $\boldsymbol{\theta}$ with 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, respectively. In this section we apply the systematic group sequential selection of $\boldsymbol{\theta}$, introduced in Section 2.1, to address the issue of users of metaheuristic optimization algorithms 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 $\mathbf{Z}_t$ in Section 2.1 are not completely specified, because they depend on the unknown $\boldsymbol{\theta}$, 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 $\mathbf{Z}_t$ directly by applying the bootstrap method to these data.

3.1. Enhanced PSO with adaptively tuned hyperparameters

PSO, proposed by Kennedy and Eberhart (1995) as a nature-inspired optimization method, was developed from a 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 communicate their findings with one another to determine collectively where the food is. Thus, there are two types of positions, called the "personal best" position and the "global best" position, respectively, found by the flock to date. The velocities and locations of the birds are updated iteratively. For the next iteration, each bird flies in a direction that takes into account (i) its current direction, (ii) its current known personal best position (cognitive component), and (iii) the flock's current best known position (social component). Extending the "birds" to "collaborating particles" to minimize a loss function $f : \mathbb{R}^d \to \mathbb{R}$, PSO denotes the location (respectively, velocity) of the ith particle at the tth iteration by $\mathbf{x}_i(t)$ (respectively, $\mathbf{v}_i(t)$), and defines

$$\mathbf{x}_i^*(t) = \underset{0\le s\le t}{\operatorname{argmin}} f(\mathbf{x}_i(s)), \quad \mathbf{x}^*(t) = \underset{1\le j\le n, 0\le s\le t}{\operatorname{argmin}} f(\mathbf{x}_j(s)), \tag{3.1}$$

representing the personal best location found by the ith particle and the global best location found by the swarm so far, respectively. We consider here the following enhancement of the classical PSO algorithm:

$$\mathbf{x}_i(t+1) = \chi_{\mathcal{D}}\left(\mathbf{x}_i(t) + \eta \mathbf{v}_i(t+1)\right), \tag{3.2}$$

$$\begin{aligned}\mathbf{v}_i(t+1) = {} & (1-\eta\omega)\mathbf{v}_i(t) + \eta c_1 \mathbf{U}_{1,i}(t) \circ (\mathbf{x}_i^*(t) - \mathbf{x}_i(t)) \\ & + \eta c_2 \mathbf{U}_{2,i}(t) \circ (\mathbf{x}^*(t) - \mathbf{x}_i(t)) + \eta \mathbf{Z}_i(t+1),\end{aligned} \tag{3.3}$$

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

The classical PSO algorithm uses $\eta = 1$, does not include the term $\eta\mathbf{Z}_i(t+1)$ in (3.3), and assumes the components of $\mathbf{U}_{1,i}(t)$ and $\mathbf{U}_{2,i}(t)$ to be i.i.d. Unif$(0,1)$. Choosing $c_1 = c_2 = c$, the tuning parameter vector in this case is (η, ω, c), in which η is the step size, whereas $1 - \eta\omega$ and c are positive weights. We can reduce the dimensionality of the search to two (instead of four) by modifying part B in Section 2.1 as follows. Use a fixed value $\eta_+ = 0.95$ of η for the initial stages at user-selected times $t_1, \ldots, t_J$. With regard to the choice of η for the later stages $j > J$, the strategy is to choose (ω, c) first, and then to determine η. Without introducing additional notation, we simply let $\boldsymbol{\lambda} = (\omega, c)$, and replace $A(\boldsymbol{\lambda})$ in Section 2.1 with $A(\boldsymbol{\lambda}, \eta)$, and replace $\mathbf{x}_t(\boldsymbol{\lambda}_j)$ and $\mathbf{x}_t^*(\boldsymbol{\lambda}_j)$ with $\mathbf{x}_t(\boldsymbol{\lambda}_j, \eta_j)$ and $\mathbf{x}_t^*(\boldsymbol{\lambda}_j, \eta_j)$, respectively. In this way, we replace part B of Section 2.1 with the following steps to update $\boldsymbol{\lambda}_j, \Lambda_j$, and η_j at t_j $(1 \le j \le J)$, after initializing at time $t = 1$ by choosing $\eta = \eta_+$, $\boldsymbol{\lambda}_0$ from Λ according to some given distribution, and $\Lambda_0 = \{\boldsymbol{\lambda}_0\}$:

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

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

(iii) Let $\boldsymbol{\lambda}_j^* = \operatorname{argmax}_{\boldsymbol{\lambda}\in\Lambda_j^*} \widehat{E}\Delta_j(\boldsymbol{\lambda}, \eta_+)$ and carry out the ϵ-greedy randomization scheme to define $\boldsymbol{\lambda}_j$ and Λ_j.

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

$$I = \inf\left\{2 \le j \le K : \frac{\widehat{E}\Delta_j(\boldsymbol{\lambda}_j, \eta_+)}{\widehat{E}[f(\mathbf{x}^*_{t_{j-1}}(\boldsymbol{\lambda}_j, \eta_+)) - f(\mathbf{x}^*_{t_1-1}(\boldsymbol{\lambda}_j, \eta_+))]} \le \delta\right\}, \quad (3.4)$$

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

(v) For updates at times t_j, with $j > I$, carry out Step (i) to generate $\boldsymbol{\lambda}^*$, and define $\Lambda^*_j = \Lambda_{j-1} \cup \{\boldsymbol{\lambda}^*\}$. Let $\boldsymbol{\lambda}^*_j = \text{argmax}_{\boldsymbol{\lambda}\in\Lambda^*_j} \widehat{E}\Delta_j(\boldsymbol{\lambda}, \eta_{j-1})$. Then, use the ϵ-greedy randomization scheme to define $\boldsymbol{\lambda}_j$ and Λ_j. Let $\eta_j = \text{argmax}_{\eta\le 0.95}\widehat{E}\Delta_j(\boldsymbol{\lambda}_j, \eta)$.

3.2. Locally optimal designs in continuation-ratio model

Fan and Chaloner (2004) consider locally D-optimal designs for trinomial responses 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 $a \ge 0, \theta_2 > 0$, and $b > 0$. This is the *continuation-ratio model*. In contrast, the "proportional odds model", replaces $\log(\pi_2/\pi_1)$ with $\log((\pi_2+\pi_3)/\pi_1)$ and b with θ_2 in the second equation of (3.5). In particular, locally D-optimal designs ξ_D maximize $\log\det(\mathbf{I}_\xi)$ of the Fisher information matrix $\mathbf{I}_\xi$ over the design measures ξ, and can be found numerically to have a fixed number (that increases with a) of design points. Locally c-optimal designs ξ_ψ minimize the asymptotic variance of the MLE of a real-valued function ψ of the parameters, and can be found similarly using numerical methods that need to address additional issues, such as the singularity of the asymptotic covariance matrix, as explained in Fan and Chaloner (2004, p.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; that is, $\pi_3(x_T; \theta_1, \theta_2) = p$ in the continuation-ratio model (3.5). Solving this equation yields $x_T = (\text{logit}\, p - \theta_1)/\theta_2$. Hence, $\boldsymbol{\nabla} x_T(\theta_1, \theta_2) = \left(-1/\theta_2, \theta_2^{-2}(\theta_1 - \text{logit}\, p)\right)^\top$, and the locally c-optimal design minimizes the asymptotic variance of the MLE of $\psi(\theta_1, \theta_2) = \log(\boldsymbol{\nabla}^\top x_T(\theta_1, \theta_2)\mathbf{I}^-_\xi(\theta_1, \theta_2)\boldsymbol{\nabla} x_T(\theta_1, \theta_2))$ for dose-response studies of the MTD, where $\mathbf{I}^-_\xi$ is the generalized inverse of the Fisher information matrix $\mathbf{I}_\xi$, which

may be singular. Denote this locally c-optimal design by ξ_T.

The proportional odds model 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 zero representing "no toxicity and no efficacy," outcome one representing moderate toxicity, and outcome two representing severe toxicity in the bone marrow transplant 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 the actual data, we use the continuation-ratio model to define the MTD and the most efficacious dose (MED). The MED 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 graft-versus-host disease and no severe toxicity. In the continuation-ratio model for toxicity and efficacy, the MED is defined as the maximizer x_E of $p(x;\boldsymbol{\theta}) := (1-\pi_3)\left(1+e^{\theta_3+\theta_4 x}\right)^{-1}$, where π_3 is given by the first equation of (3.5), which is equivalent to $1-\pi_3 = \left(1+e^{\theta_1+\theta_2 x}\right)^{-1}$, $\boldsymbol{\theta} = (\theta_1,\theta_2,\theta_3,\theta_4)$, 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 $(d/dx)\log p(x;\boldsymbol{\theta}) = 0$ yields $x_E(\boldsymbol{\theta})$, from which we can (a) use the delta method to derive the asymptotic variance of the MLE of $x_E(\boldsymbol{\theta})$, (b) apply the implicit function theorem to evaluate $\boldsymbol{\nabla} x_E(\boldsymbol{\theta})$, and (c) derive the locally c-optimal design ξ_E that minimizes $\psi(\boldsymbol{\theta}) := \log(\boldsymbol{\nabla}^\top x_E(\boldsymbol{\theta})\mathbf{I}_\xi^-(\boldsymbol{\theta})\boldsymbol{\nabla} x_E(\boldsymbol{\theta}))$.

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

$$\Psi(\xi;\boldsymbol{\lambda}) = \lambda_1\Psi_T(\xi) + \lambda_2\Psi_E(\xi) + \lambda_3\Psi_D(\xi), \tag{3.6}$$

with nonnegative λ_i that sum to one. Section 4 views $\boldsymbol{\lambda} = (\lambda_1,\lambda_2,\lambda_3)$ as a tuning parameter such that $\lambda_3 = 0$ and $\lambda_1+\lambda_2 = 1$, and applies 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, Sec. 3.3 and p.90–91), multi-objective optimization typically involves conflicting objectives, such as the benefit versus the risk (or the efficacy versus the toxicity) of a treatment. Let $f : S \to \mathbb{R}^m$ be a vector-valued function such that $f_i(\mathbf{x})$ represents the ith 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

$$\begin{aligned} f_1(\xi; \boldsymbol{\lambda}) &= \lambda_1 \left\{ \log \left(\boldsymbol{\nabla}^\top x_T \mathbf{I}_{\xi_T}^- \boldsymbol{\nabla} x_T \right) - \log \left(\boldsymbol{\nabla}^\top x_T \mathbf{I}_{\xi}^- \boldsymbol{\nabla} x_T \right) \right\}, \\ f_2(\xi; \boldsymbol{\lambda}) &= \lambda_2 \left\{ \log \left(\boldsymbol{\nabla}^\top x_E \mathbf{I}_{\xi_E}^- \boldsymbol{\nabla} x_E \right) - \log \left(\boldsymbol{\nabla}^\top x_E \mathbf{I}_{\xi}^- \boldsymbol{\nabla} x_E \right) \right\}, \\ f_3(\xi; \boldsymbol{\lambda}) &= (1 - \lambda_1 - \lambda_2) \left(\log \det(\mathbf{I}_{\xi_D}) - \log \det(\mathbf{I}_{\xi}) \right). \end{aligned}$$

Here, we consider the locally D-optimal design ξ_D that maximizes $\log \det(\mathbf{I}_\xi)$ 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 ξ_T (for MTD, with $p = 0.3$) and ξ_E (for MED). Including these designs for the individual criteria in defining $(f_1, f_2, f_3)^\top$ 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 $-(1-\lambda_1-\lambda_2)p^{-1}(\log \det(\mathbf{I}_\xi) - \log \det(\mathbf{I}_{\xi_D}))$. 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 ξ_T, ξ_E, and ξ_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 one of finding only ξ_T and ξ_E.

Example 2. For the parameter configuration $(\theta_1, \theta_2, \theta_3, \theta_4)$ in Example 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 λ_1 and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$ are unspecified. The latter is needed to define the locally c-optimal designs ξ_T and ξ_E. We update the values of λ and $\boldsymbol{\theta}$ 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 efficiently compute an optimal

discrete set $\mathcal{D}_j$ of doses at time t_j when the information set $\mathcal{F}_{t_j}$ generated by the doses and the toxicity-efficacy outcomes up to that time is used to update $\boldsymbol{\theta}, \boldsymbol{\lambda}$, and $\mathcal{D}_j$ (see below). First, we explain how, for a given $\boldsymbol{\theta}$, the 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 ξ_λ that minimizes $\lambda\Psi_T(\xi) + (1-\lambda)\Psi_E(\xi)$, in the direction off the design measure δ_x degenerate at x, is given by

$$\lambda \frac{\operatorname{tr}\left\{\mathbf{I}_{\delta_x}(\boldsymbol{\theta})\mathbf{I}^-_{\xi_\lambda}(\boldsymbol{\theta})\left(\boldsymbol{\nabla} x_T(\boldsymbol{\theta})\boldsymbol{\nabla} x_T(\boldsymbol{\theta})^\top\right)\mathbf{I}^-_{\xi_\lambda}(\boldsymbol{\theta})\right\}}{\operatorname{tr}\left\{\left(\boldsymbol{\nabla} x_T(\boldsymbol{\theta})\boldsymbol{\nabla} x_T(\boldsymbol{\theta})^\top\right)\mathbf{I}^-_{\xi_\lambda}(\boldsymbol{\theta})\right\}} \tag{3.7}$$
$$+(1-\lambda)\frac{\operatorname{tr}\left\{\mathbf{I}_{\delta_x}(\boldsymbol{\theta})\mathbf{I}^-_{\xi_\lambda}(\boldsymbol{\theta})\left(\boldsymbol{\nabla} x_E(\boldsymbol{\theta})\boldsymbol{\nabla} x_E(\boldsymbol{\theta})^\top\right)\mathbf{I}^-_{\xi_\lambda}(\boldsymbol{\theta})\right\}}{\operatorname{tr}\left\{\left(\boldsymbol{\nabla} x_E(\boldsymbol{\theta})\boldsymbol{\nabla} x_E(\boldsymbol{\theta})^\top\right)\mathbf{I}^-_{\xi_\lambda}(\boldsymbol{\theta})\right\}}.$$

The equivalence theorem for optimal designs (Pukelsheim (2006)) yields that ξ_λ is optimal for (3.6) if and only if (3.7) is bounded above by one for all $x \in [x_{\min}, x_{\max}]$, with equality at the support points of ξ_λ. For each value λ, carry out the PSO to minimize $\lambda\Psi_T(\xi)+(1-\lambda)\Psi_E(\xi)$, yielding the minimizer ξ_λ; the Pareto boundary is $\{\xi_\lambda : 0 < \lambda < 1\}$. Note that $m = 2$ and $S \subset [x_{\min}, x_{\max}]$ in this case. Actually, we need only approximate the Pareto boundary over a discrete grid $\mathcal{D}_J$ for the doses in Phase II, with λ belonging to a discrete subset Λ_J of $(0, 1)$, rather than the entire Pareto boundary $\{\xi_\lambda : 0 < \lambda < 1\}$. For the PSO algorithm, we generate a discrete set Λ_j of hyperparameter values λ and the corresponding $\xi_\lambda, \mathbf{I}_{\xi_\lambda}(\boldsymbol{\theta}_j)$ and the tuning parameters ω, c, and η_j of the PSO described in the second paragraph of Section 3.1. Then, we choose $(Z_1(s), \ldots, Z_n(s))^\top$ in (3.3), for $t_j < s \le t_{j+1}$, to be i.i.d. replicates of the random vector in (3.7), with $\boldsymbol{\theta} = \boldsymbol{\theta}_j$ and $\lambda \in \Lambda_j$; see parts A and B of the basic algorithm in the second paragraph of Section 2.1.

With regard to the choice of $\boldsymbol{\theta}_j$, one can use the `glm` function in CRAN to fit generalized linear models using the maximum likelihood, and then substitute $\boldsymbol{\theta}_j$ in $\mathbf{Z}(s) := (Z_1(s), \ldots, Z_n(s))^\top$, for $t_j < s \le t_{j+1}$, with $\hat{\boldsymbol{\theta}}_j$. This has the drawback of neglecting the sampling validity of $\hat{\boldsymbol{\theta}}_j$ in carrying out the multi-armed bandit scheme (part B of Section 2.1) to update the set Λ_j of the 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 distribution under i.i.d. models," for independent but non-identically distributed data, fo-

cusing in particular on the setting of $\mathbf{Z}(1), \ldots, \mathbf{Z}(T)$ drawn from distributions $G_1, \ldots, G_T$, and on the regression case that $\mathbf{Z}(t)$ depends on covariates. For i.i.d. data, the population distribution G can be approximated by an empirical distribution $\widehat{G}$ that places a weight $1/T$ on each of the observed $\mathbf{Z}(t)$, and by the bootstrap resamples $\mathbf{Z}^*(t)$ from $\widehat{G}$. For the independent $\mathbf{Z}(s)$ that are identically distributed if $s \in \{t_j + 1, \ldots, t_{j+1}\}$, the bootstrap procedure resamples $\mathbf{Z}^*(s)$ from the empirical distribution that places the weight $(t_{j+1} - t_j)^{-1}$ on each item of the observed sample $\{\mathbf{Z}(t) : t_j < t \le t_{j+1}\}$, 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 designs of dose-finding studies of treatments that have dose-limiting toxicities. These include 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 sample 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 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 the MTD and MED estimations, and weight 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 λ_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 that "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 prespecified 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 constraint of *coherence*. A group sequential

design ξ with group size m and dose x_i is called coherent if there is a threshold p such that

$$P_\xi(x_i - x_{i-1} > 0 | \widetilde{Y}_{i-1} \geq p) = 0 \quad \text{and} \quad P_\xi(x_i - x_{i-1} < 0 | \widetilde{Y}_{i-1} \leq p) = 0 \quad (4.1)$$

for all i. Here, $\widetilde{Y}_{i-1}$ is the observed proportion of toxicities in group $i-1$ and P_ξ denotes the probability measure induced by the design ξ, for which (4.1) is tantamount to a dose de-escalation if $\widetilde{Y}_{i-1} \geq p$, and to a dose escalation if $\widetilde{Y}_{i-1} \leq p$. 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 estimates the 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, Pepe and Fisher (1990); Shen and O'Quigley (1996)) in Phase I trials to estimate the MTD, or the more general most successful dose presented in their Section 7 and illustrated using HIV treatment, for which "failure is either a toxicity (that causes) inability to maintain treatment, or an unacceptably low therapeutic response." As such, 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 a therapeutic response given no toxicity at that dose. "A successful trial would identify the dose level ℓ 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 escalation with overdose control (EWOC) model, introduced by Babb, Rogatko and Zacks (1998). Both papers consider MTD estimation and the choice of prior distribution. Theorem 2.2 of Tighiouart and Rogatko (2010) gives conditions on a reparameterized prior distribution to ensure coherence of the EWOC. In contrast, Bartroff and Lai (2010) incorporate the risk to patients in the trial using a "global risk" $E[\sum_{k=1}^{T} h(x_k, \eta) + g(\hat{\eta}_T, \eta)]$, where η denotes the MTD, $\hat{\eta}_T$ is the MTD estimate at the termination of the trial, $g(\hat{\eta}_T, \eta)$ measures the effect of $\hat{\eta}_T$ on future

patients, and $h(x,\eta) = \omega(\eta - x)_+ + (1-\omega)(x-\eta)_+$, because the EWOC doses are the ωth quantiles of the posterior distributions of η. They use the rollout algorithm in approximate dynamic programming to minimize the global risk, the idea of the rollout is "to approximate the optimal policy x_k^* by the minimizer (over x) of $h_{k-1}(x) + E[\sum_{i=k+1}^{T} h_{i-1}(\hat{x}_i)|\mathcal{F}_{k-1},\ \hat{x}_k = x]$ (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 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 whether 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 to circumvent the 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. Section 2.1 describes how t select the optimal parameter value, where Bayesian c-optimal designs are used to define optimality. Because the solution to the compound optimality criterion is a Pareto boundary, this entails loop optimization. Here, the PSO in Section 3.2, with adaptively chosen tuning parameters, is particularly effective, as shown in the next subsection.

4.2. Efficient group sequential design of early-phase trials

We decompose the group sequential design into three stages: Phase I (for MTD estimation), Phase I/II (for finding a discrete set $\mathcal{D}$ of doses, belonging

to the Pareto boundary, that are c-optimal for estimating the MED, subject to probability constraints on the DLT), and Phase II (for adaptively choosing the dose from $\mathcal{D}$ and testing the 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 the EWOC, which Bartroff and Lai (2010) reparameterize as follows, and choose the dose for each cohort in the interval $[x_{\min}, x_{\max}]$ (without discretization) to circumvent rigidity. Assuming an upper bound $q > 0$ on the probability ρ of toxicity at $x_{\min}$, uniform distributions over $[x_{\min}, x_{\max}]$ and $[0, q]$ are taken as the prior distributions for η (the MTD) and ρ, respectively. The EWOC assumes a logistic regression model $1/(1+e^{-(\alpha+\beta x)})$ for the probability of DLT at dose level x. Because the logistic regression parameters can be expressed in terms of η and ρ as

$$\alpha = \frac{x_{\min}\operatorname{logit} p - \eta \operatorname{logit} \rho}{x_{\min} - \eta}, \quad \beta = \frac{\operatorname{logit} \rho - \operatorname{logit} p}{x_{\min} - \eta},$$

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

$$C^{-1} \prod_{i=1}^{t} \left[\frac{e^{y_i \psi(x_i;\rho,\eta)}}{1 + e^{y_i \psi(x_i;\rho,\eta)}} \right] \quad \text{on } [x_{\min}, x_{\max}] \times [0, q], \tag{4.2}$$

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

$$C = \int_{x_{\min}}^{x_{\max}} \int_0^q \prod_{i=1}^{t} \left[\frac{e^{y_i \psi(x_i;\rho,\eta)}}{1 + e^{y_i \psi(x_i;\rho,\eta)}} \right] d\rho \, d\eta \tag{4.3}$$

is the normalizing constant, which 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 η (at which the probability of DLT is p) in the logistic regression model, Tighiouart and Rogatko (2010, Thm. 2.2) show that the EWOC is coherent if $F(x, \eta)$ is non-increasing in η for fixed x. This is implicitly assumed by Bartroff and Lai (2010) in their rollout algorithm to minimize the global risk, which explains the role of the $\mathcal{F}_t$-posterior density (4.2) in the rollout algorithm. The Phase I design here simply uses the coherence of the EWOC, and applies (4.2) to the current cohort of m subjects to determine the dose of the next cohort as the pth quantile of the posterior distribution of η, given the binary indicators y_i of DLT in the current cohort.

The pth quantile η_* is defined by

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

where $\{(x_i, y_i) : 1 \le i \le m\}$ are the observations of the cohort, and C is given by (4.3).

Phase I/II. This stage uses the toxicity and efficacy outcomes of the subjects treated at doses that are chosen adaptively using PSO 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 $\mathcal{D}_j$ of doses determined in Phase I/II, we can proceed as in Section 3.2 of Bartroff, Lai and Narasimhan (2014, BLN). For $t_j > T$, we perform order-restricted generalized likelihood ratio (GLR) tests at the jth 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 $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$ is also highlighted. Relabeling the doses in the discrete set $\mathcal{D}_J$ with cardinality ν as $d_1 < d_2 < \cdots < d_\nu$, assume the order restrictions $\pi(d_1) \le \cdots \le \pi(d_\nu)$ and $p(d_1) \le \cdots \le p(d_\nu)$, which yield an order-restricted MLE of $\boldsymbol{\theta}$ (Section 2.2 of BLN). Taking advantage of the discrete dose set $\mathcal{D}_J$, Section 3.2 of BLN first introduces $\pi_i(y, z) = P(y_t = y, z_t = z | x_t = d_i)$ to relate the binary indicator z_t of an efficacious response without DLT, for $t_j < t \le t_{j+1}$, when the dose x_t (determined at time t_j of the jth interim analysis) is d_i. It then expresses the log-likelihood ratio at the jth interim analysis as $\ell_j(\boldsymbol{\theta}) = \sum_{t=1}^{t_j} \log \pi_{i(x_t)}(y_t, z_t)$, where $i(x_t) = d_i$ if x_t is assigned dose d_i, and introduces ν parameters $\rho_i = \pi_{i(x_t)}(0,0)\pi_{i(x_t)}(1,1)/\{\pi_{i(x_t)}(1,0)\pi_{i(x_t)}(0,1)\}$, for $i = 1, \ldots, \nu$. This yields the log GLR statistic at the jth interim analysis for testing the efficacy of the selected doses up to that time, in terms of the 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 the stopping rules in Section 3.1 of BLN, whose Section 4 (particularly Section 4.2) provides extensive simulation studies showing the advantages of this approach. The software to implement Phase II, together with examples, data, simulation studies, and real-world applications, is available at `https://med.stanford.edu/cisd/research/software.html`.

5. 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 use recent advances in statistics to address this open problem in AI, showing the usefulness of our solution in tackling challenging optimal design problems in early-phase oncology trials. The Supplementary Material S4 describes further advances that are still ongoing, including (a) precision-guided development of personalized therapies and master protocols of confirmatory clinical trials to test them for regulatory approval, (b) valid statistical inference from the data in these adaptive designs of group sequential trials, and further discussion of Phase II and its extension to Phase II/III, (c) metaheuristic optimization algorithms other than PSO that Section 3 focuses on, and the convergence theory of these adaptively tuned algorithms, and (d) extensions of Theorems 1 and 2 beyond the exponential family, and an extension of the multi-armed bandit theory in Section 2.2 to nonparametric multi-armed bandits with covariates (which are important for personalized medicine and recommendation systems).

Acknowledgments

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

References

Atkinson, A., Donev, A. and Tobias, R. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford.

Auer, P., Cesa-Bianchi, N. and Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning* **47**, 235–256.

Babb, J., Rogatko, A. and Zacks, S. (1998). Cancer phase I clinical trials: Efficient dose escalation with overdose control. *Statistics in Medicine* **17**, 1103–1120.

Bartroff, J. and Lai, T. L. (2010). Approximate dynamic programming and its applications to the design of Phase I cancer trials. *Statistical Science* **25**, 245–257.

Bartroff, J., Lai, T. L. and Narasimhan, B. (2014). A new approach to designing Phase I-II cancer trials for cytotoxic chemotherapies. *Statistics in Medicine* **33**, 2718–2735.

Bartroff, J., Lai, T. L. and Shih, M.-C. (2013). *Sequential Experimentation in Clinical Trials: Design and Analysis*. Springer, New York.

Chan, H. P. and Lai, T. L. (2006). Sequential generalized likelihood ratios and adaptive treatment allocation for optimal sequential selection. *Sequential Analysis* **25**, 179–201.

Chen, J., Heyse, J. and Lai, T. L. (2018). *Medical Product Safety Evaluation: Biological Models and Statistical Methods*. Chapman & Hall/CRC, Boca Raton, FL.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics* **24**, 586–602.

Cheung, Y. K. (2010). Stochastic approximation and modern model-based designs for dose-finding clinical trials. *Statistical Science* **25**, 191–201.

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

Cook, R. D. and Wong, W. K. (1994). On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association* **89**, 687–692.

Fan, S. and Chaloner, K. (2004). Optimal designs and limiting optimal designs for a trinomial response. *Journal of Statistical Planning and Inference* **126**, 347–360.

Federov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York. Translated from Russian and edited by W. J. Studden and E. M. Kilimmo.

Huang, C., Li, Y. and Yao, X. (2020). A survey of automatic parameter tuning methods for metaheuristics. *IEEE Transactions on Evolutionary Computation* **24**, 201–216.

Huang, Y., Gilmour, S. G., Mylona, K. and Goos, P. (2019). Optimal design of experiments for non-linear response surface models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **68**, 623–640.

Kaelbling, L. P., Littman, M. L. and Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research* **4**, 237–285.

Kennedy, J. and Eberhart, R. C. (1995). Particle swarm optimization. In *Proceedings of the IEEE International Conference on Neural Networks IV*, 1942–1948. IEEE.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics* **12**, 363–366.

Lai, T. L. (1987). Adaptive treatment allocation and the multi-armed bandit problem. *The Annals of Statistics* **15**, 1091–1114.

Lai, T. L. and Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics* **6**, 4–22.

Liu, R. Y. (1988). Bootstrap procedures under some non-iid models. *The Annals of Statistics* **16**, 1696–1708.

O'Quigley, J. and Conaway, M. (2010). Continual reassessment and related dose-finding designs. *Statistical Science* **25**, 202–216.

O'Quigley, J., Pepe, M. and Fisher, L. (1990). Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* **46**, 33–48.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society* **58**, 527–535.

Shen, L. Z. and O'Quigley, J. (1996). Consistency of continual reassessment method under model misspecification. *Biometrika* **83**, 395–405.

Sun, J., Lai, C.-H. and Wu, X.-J. (2012). *Particle Swarm Optimisation: Classical and Quantum Perspectives*. Chapman & Hall/CRC, Boca Raton, FL.

Thall, P. F. (2010). Bayesian models and decision algorithms for complex early phase clinical trials. *Statistical Science* **25**, 227–244.

Thall, P. F. and Russell, K. E. (1998). A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in Phase I/II clinical trials. *Biometrics* **54**, 251–264.

Tighiouart, M. and Rogatko, A. (2010). Dose finding with escalation with overdose control (EWOC) in cancer clinical trials. *Statistical Science* **25**, 217–226.

Kwok Pui Choi

Department of Statistics and Data Science, National University of Singapore, Singapore 117546.

E-mail: choikp@nus.edu.sg

Tze Leung Lai

Department of Statistics, Stanford University, Stanford, CA 94305, USA.

E-mail: lait@stanford.edu

Xin T. Tong

Department of Mathematics, National University of Singapore, Singapore 119077.

E-mail: mattxin@nus.edu.sg

Weng Kee Wong

Biostatistics Department, University of California, Los Angeles, LA 90095, USA.

E-mail: wkwong@ucla.edu

(Received October 2020; accepted December 2020)