LABORATOIRE

INFORMATIQUE, SIGNAUX ET SYSTÈMES
DE SOPHIA ANTIPOLIS
UMR 6070

ASYMPTOTIC PROPERTIES OF ADAPTIVE PENALIZED OPTIMAL DESIGN WITH APPLICATION TO DOSE-FINDING

Luc Pronzato

Equipe SYSTEMES

Rapport de recherche
ISRN I3S/RR–2008-19–FR

Octobre 2008

Laboratoire d'Informatique de Signaux et Systèmes de Sophia Antipolis - UNSA-CNRS
2000, rte.des Lucioles – Les Algorithmes – Bât Euclide B – B.P. 121 – 06903 Sophia-Antipolis Cedex – France
Tél.: 33 (0)4 92 94 27 01 – Fax: 33 (0)4 92 94 28 98 – www.i3s.unice.fr
UMR6070

RÉSUMÉ :

MOTS CLÉS :

ABSTRACT:

Adaptive optimal design with a cost constraint is considered, both for nonlinear regression and Bernoulli type experiments, with application in clinical trials. The strong consistency and asymptotic normality of the estimators is proved for designs over a finite space, both when the cost level is fixed, and the adaptive design converges to an optimum constrained design, and when the objective is to minimize the cost.

KEY WORDS :

Adaptive design; Sequential design; Optimal experimental design; Penalized experimental design; Constrained experimental design; Dose finding; Consistency; Asymptotic normality

Asymptotic properties of adaptive penalized optimal design with application to dose-finding*

Luc Pronzato

Laboratoire I3S, CNRS/Université de Nice-Sophia Antipolis

Bât Euclide, Les Algorithmes

2000 route des lucioles, BP 121

06903 Sophia Antipolis cedex, France

pronzato@i3s.unice.fr

October 21, 2008

Abstract Adaptive optimal design with a cost constraint is considered, both for nonlinear regression and Bernoulli type experiments, with application in clinical trials. The strong consistency and asymptotic normality of the estimators is proved for designs over a finite space, both when the cost level is fixed, and the adaptive design converges to an optimum constrained design, and when the objective is to minimize the cost.

Key words: Adaptive design; Sequential design; Optimal experimental design; Penalized experimental design; Constrained experimental design; Dose finding; Consistency; Asymptotic normality

*This work was partly accomplished while the author was invited at the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK. The support of the Newton Institute and of CNRS are gratefully acknowledged.

1 Introduction and motivation

This work has been motivated by recent papers [5, 6] where the authors use constrained optimal design to make a compromise between individual and collective ethics in dose-finding studies. Their idea is to use a cost function that accounts for poor efficacy and for toxicity, and to maximize information-per-cost-unit, which can be put in the form of a standard (unconstrained) optimal design problem. Using a parametric model for the dose/efficacy-toxicity responses (Gumbel or Cox model as in [5] or a bivariate probit model as in [6]), the Fisher information matrix can be calculated and optimal designs can be constructed.

In a companion paper to the present one, we introduce some flexibility in setting the balance between the information gained (in terms of precision of parameter estimation) and the cost of the experiment (in terms of poor success for the patients enroled in the trial). This is obtained by maximizing information-per-observation under a constraint on the cost or, equivalently, by optimizing a penalized design criterion where the penalty is related to the cost of the experiment.

The aim of the present paper is to investigate the asymptotic properties of such penalized optimal designs when they are constructed sequentially, the choice of each new design point being based on the current estimated value of the model parameters. Although [5, 6] advocate the use of adaptive experimental design, the convergence of the procedure (strong consistency of the estimator of the model parameters and convergence of the design to the optimal non-sequential design for the true value of the model parameters) is left as an open issue. This difficulty is usually overcome by considering an initial experiment (non adaptive) that grows in size when the total number of observations increases. Although this number is often severely limited in practise, especially for clinical trials, we think that it is reassuring to know that, *for a given initial experiment*, adaptive design guarantees suitable asymptotic properties under reasonable conditions. Using simple arguments, we show that this is indeed the case when *the design space is finite*, which forms a rather natural assumption in the context of clinical trials. Our results concern penalized D-optimal design but also cover the case, more classical, of fully adaptive D-optimal design. Therefore, they also apply for the method used in [5, 6]. Additionally to strong consistency, we also show that the asymptotic distribution of the parameter estimates is normal, with variance-covariance matrix given by inverse of the usual information matrix, similarly to the non-adaptive case. Moreover, we show that, for suitable penalty functions, when the weight given to the cost for bad treatment increases with the number of patients enroled, the doses allocated converge to the optimal one while the parameters are still estimated consistently.

The results presented are of rather general applicability and, although this work has been motivated by considerations in the context of clinical trials, we show that they also cover the case of least-squares estimation in nonlinear regression models, for which there exist even less consistency results than for maximum likelihood estimation when the design is adaptive. In particular, the results of Sect. 3 and 4 form a major improvement over those in [24] where only linear regression models were considered.

Adaptive penalized D-optimal design is introduced in Sect. 2. Sect. 3 concerns the case where the penalty coefficient is bounded. We show that when both the design and the penalty coefficient are adapted to the current estimated value of the model parameters, one can obtain strong consistency and asymptotic normality of the estimator and strong convergence of the design to the optimum. In Sect. 4, the penalty coefficient is allowed to grow to infinity. We show that when the increase is not too fast, the strong consistency and asymptotic normality of the parameter estimator can be preserved, while at the same time the design asymptotically concentrates around points of minimum cost. Sect. 5 concludes and suggests several directions for further developments. The proofs of lemmas and theorems are collected in an Appendix.

2 An adaptive penalized D-optimal design

Let $\mathscr{X}$, a compact subset of $\mathbb{R}^d$, denote the admissible domain for the experimental variables x (design points) and $\theta \in \mathbb{R}^p$ denote the p-dimensional vector of parameters of interest in a parametric model. The information matrix for parameters θ and design measure ξ (a probability measure on $\mathscr{X}$) is denoted $\mathbf{M}(\xi,\theta) = \int_{\mathscr{X}} \mu(x,\theta)\, \xi(dx)$, with $\mu(x,\theta)$ the contribution of the design point x, see Sect. 3.1.

In a nonlinear situation, $\mathbf{M}(\xi,\theta)$ depends on θ and local optimal design maximizes a concave function $\Psi(\cdot)$ of $\mathbf{M}(\xi,\theta)$. Here we shall only consider D-optimal design, for which $\Psi(\mathbf{M}) = \log\det(\mathbf{M})$. The extension to other global optimality criteria, such as $[\text{trace}(\mathbf{M}^{-1})]^{-1}$ for instance, can be obtained by following a similar route. A rather common approach to overcome the difficulty caused by the dependence of a local optimal design in the unknown value of the model parameters is to design the experiment sequentially.

In fully-adaptive D-optimal design, next design point after N observations is taken as

$$x_{N+1} = \arg\max_{x\in\mathcal{X}} \text{trace}[\mu(x,\hat{\theta}^N)\mathbf{M}^{-1}(\xi_N,\hat{\theta}^N)]\,, \tag{1}$$

where $\hat{\theta}^N \in \Theta \subset \mathbb{R}^p$ is the current estimated value for θ, based on $x_1,\ldots,x_N$ and the associated observations $Y_1,\ldots,Y_N$, and $\xi_N = (1/N)\sum_{i=1}^N \delta_{x_i}$ is the current empirical design measure, with δ_z the delta measure that puts mass 1 at z. We leave aside initialisation issues and simply assume that $x_1,\ldots,x_p$ are such that $\mathbf{M}(\xi_p,\theta)$ is nonsingular for any $\theta\in\Theta$. Note that (1) can only be considered as an algorithm for choosing design points, in the sense that $\mathbf{M}(\xi_N,\theta)$ is not the information matrix for parameters θ due to the sequential construction of the design (we shall see, however, in Sect. 3 that when $\mathcal{X}$ is finite, from the same repeated sampling principle as in [32], one can still use $\mathbf{M}(\xi_N,\hat{\theta}^N)$ to characterize the precision of the estimation of θ as $N\to\infty$).

Constrained local D-optimal design maximizes $\log\det[\mathbf{M}(\xi,\theta)]$ under a cost constraint $\Phi(\xi,\theta)\le C$. We suppose that the cost function $\Phi(\xi,\theta)$ is linear in ξ, that is

$$\Phi(\xi,\theta) = \int_{\mathcal{X}} \phi(x,\theta)\,\xi(dx)\,,$$

and that $\phi(x,\theta)$ is bounded on $\mathcal{X}$. The extension to nonlinear constraints is considered, e.g., in [3] and [9, Chap. 4]. We shall restrict our attention to the case where a single — scalar — constraint is present, some of the issues caused by the presence of several constraints are addressed in the same references; see also Sect. 5. A necessary and sufficient condition for the optimality of ξ^* satisfying $\Phi(\xi^*,\theta)\le C$ is the existence of a Lagrange coefficient $\lambda^* = \lambda^*(\theta)\ge 0$ such that

$$\lambda^*[C-\Phi(\xi^*,\theta)] = 0 \;\text{ and }\; \forall x\in\mathcal{X}\,,\; \text{trace}[\mu(x,\theta)\mathbf{M}^{-1}(\xi^*,\theta)] \le p + \lambda^*[\phi(x,\theta)-\Phi(\xi^*,\theta)]\,.$$

In practice, ξ^* can be determined by maximizing

$$H_\theta(\xi,\lambda) = \Psi[\mathbf{M}(\xi,\theta)] - \lambda\,\Phi(\xi,\theta) \tag{2}$$

for an increasing sequence (λ_i) of Lagrange coefficients λ, starting at $\lambda_0 = 0$ and stopping at the first λ_i such that the associated optimal design ξ^* satisfies $\Phi(\xi^*, \theta) \leq C$, see, e.g., [22] (for C large enough, the unconstrained optimal design for $\Psi(\cdot)$ is optimal for the constrained problem). The penalty coefficient λ can thus be used to set the tradeoff between the maximization of $\Psi[\mathbf{M}(\xi, \theta)]$ (gaining information) and minimization of $\Phi(\xi, \theta)$ (reducing cost). One may refer to [4] for the equivalence between constrained and compound optimal designs.

For constrained D-optimal design, we take x_{N+1} that gives the steepest ascent direction for $H_{\hat{\theta}^N}(\xi_N, \lambda_N)$,

$$x_{N+1} = \arg\max_{x \in \mathcal{X}} \left\{ \mathrm{trace}[\mu(x, \hat{\theta}^N)\mathbf{M}^{-1}(\xi_N, \hat{\theta}^N)] - \lambda_N \, \phi(x, \hat{\theta}^N) \right\} , \tag{3}$$

where the choice of λ_N is discussed below. Since (1) can be considered as a special case of (3), the results in Sect. 3 also cover the case of classical (unconstrained) sequential D-optimal design (1) for which $\lambda_N = 0$ for all N (they therefore also cover the situation considered in [5, 6], which can be formulated as a standard D-optimal design problem). One can notice the similarity between (3) and the construction used in [24] for optimizing a parametric function, the parameters of which being estimated by least-squares in a linear regression model.

When $\hat{\theta}^N$ is frozen to a fixed value θ and λ_N is constant, the iterations (1) and (3) correspond to one step of a steepest-ascent vertex-direction algorithm with step-length $1/N$ at step N. Convergence to an optimal design measure is proved in [34] for iterations given by (1) and in [24] for (3) (using a general argument developed in [33]).

The fact that $\hat{\theta}^N$ is estimated in adaptive design creates dependency among observations and makes the investigation of the asymptotic behavior of the design and estimator a much more complicated issue for which few results are available: [10, 32, 23] concern a particular example with least-squares estimation; [16] is specific of Bayesian estimation by posterior mean and does not use a fully sequential design of the form (1); [20] and [2] require the introduction of a subsequence of non-adaptive design points to ensure consistency of the estimator and [1] requires that the size of the initial experiment (non-adaptive) grows with the increase in size of the total experiment. Intuitively, the almost sure convergence of $\hat{\theta}^N$ to some $\hat{\theta}^\infty$ would be enough to imply the convergence of ξ_N to an optimal design measure for $\hat{\theta}^\infty$ (this will be shown in Theorem 3) and, conversely, convergence of ξ_N to a design ξ_∞ such that $\mathbf{M}(\xi_\infty, \theta)$ is nonsingular for any θ would be enough in general to make an estimator consistent. It is thus the interplay between estimation and design iterations (which implies that each design point depends on previous observations) that creates difficulties. As shown in the next sections, those difficulties about consistency disappear when $\mathscr{X}$ is a finite set (notice that the assumption that $\mathscr{X}$ is finite is seldom limitative since practical considerations often impose such a restriction on possible choices for the design points; this can be contrasted with the much less natural assumption that would consist in considering the feasible parameter set as finite).

Two situations will be considered concerning the choice of the sequence (λ_N) in (3), respectively in Sect. 3 and 4. In the first one, the objective is to obtain an optimal design with a specified cost: we adapt λ_N to $\hat{\theta}^N$ and take $\lambda_N = \lambda_N^* = \lambda^*(\hat{\theta}^N)$, the optimal Lagrange coefficient for the constrained D-optimal design problem with parameters $\hat{\theta}^N$. The second situation corresponds to the case where (λ_N) forms an increasing sequence, which gives more and more importance to the constraint in the construction of the design. When $\phi(x, \theta)$ has a single minimum, by letting the Lagrange coefficient λ_N increase with N one may hope to be able to force the design to concentrate at the minimizer of ϕ associated with the true value of θ (that is, for clinical trials, to focuss more and more on individual ethics by allocating treatments with increasing efficiency).

The developments presented in the next sections rely on simple arguments based on three ideas. First, the sequence $(\hat{\theta}^N)$ in (3) is taken as *any sequence* in Θ. The asymptotic design properties obtained within this framework will thus also apply when $\hat{\theta}^N$ corresponds to some estimator of θ in Θ. Second, when $\mathscr{X}$ is finite we obtain a lower bound on the sampling rate of a subset of points of $\mathscr{X}$ associated with a nonsingular information matrix. Third, we show that this bound guarantees the strong consistency of the estimator of θ, both for least-squares estimation in nonlinear regression and maximum-likelihood estimation for Bernoulli trials. With a few additional technicalities, this yields almost sure convergence results for the adaptive designs constructed via (3). These results apply to a wide range of situations and we try to keep the presentation general enough. However, to avoid unnecessary complications we only treat the univariate case where $\mu(x,\theta)$ has rank one, and write $\mu(x,\theta) = \mathbf{f}_\theta(x)\mathbf{f}_\theta^\top(x)$, with $\mathbf{f}_\theta(x)$ a p-dimensional vector. The extension to the multivariate case does not raise particular difficulties. Matrices are denoted by bold capital letters and we denote by $\|\mathbf{A}\|$ the usual norm of $\mathbf{A}$, $\|\mathbf{A}\| = [\mathrm{trace}(\mathbf{A}^\top\mathbf{A})]^{1/2} = (\sum_{i,j}\{\mathbf{A}\}_{i,j}^2)^{1/2}$, and by $\Lambda_{\min}(\mathbf{A})$ the minimum eigenvalue of $\mathbf{A}$.

3 Asymptotic properties of adaptive design with bounded penalty

We shall use the following assumptions on the design space $\mathcal{X}$, vector $\mathbf{f}(x,\theta)$, constraint function $\phi(x,\theta)$ and Lagrange coefficients λ_N.

$\mathbf{H}_{\mathcal{X}}$-(i): The design space $\mathcal{X}$ is finite, $\mathcal{X} = \{x^{(1)}, x^{(2)}, \ldots, x^{(K)}\}$.

$\mathbf{H}_{\mathcal{X}}$-(ii): $\inf_{\theta\in\Theta} \Lambda_{\min}\left[\sum_{i=1}^K \mathbf{f}_\theta(x^{(i)})\mathbf{f}_\theta^\top(x^{(i)})\right] > \gamma > 0$.

$\mathbf{H}_{\phi}$-(i): $0 \leq \phi(x,\theta) < \bar{\phi}\,,\ \ \forall x \in \mathcal{X}$ and $\theta \in \Theta$.

$\mathbf{H}_{\lambda}$-(i): $0 \leq \lambda_N < \bar{\lambda} < \infty\,,\ \ \forall N$.

When $\lambda_N = \lambda_N^* = \lambda^*(\hat{\theta}^N)$, the optimal Lagrange coefficient for the constrained D-optimal design problem with parameters $\hat{\theta}^N \in \Theta$, the following condition guarantees that H_λ-(i) is satisfied.

$\mathbf{H}_{\lambda}$-(i'): There exists $C' < C$ such that $\forall\theta \in \Theta$, $\exists\hat{\xi}(\theta) \in \Xi$ with $\Phi[\hat{\xi}(\theta),\theta] \leq C'$ and $\mathbf{M}[\hat{\xi}(\theta),\theta]$ has full rank.

We first obtain a lower bound on the sampling rate of nonsingular designs, which will be the cornerstone for proving the consistency and asymptotic normality of estimators. The proof is given in Appendix.

Lemma 1 *Let $(\hat{\theta}^N)$ be an arbitrary sequence in Θ used to generate design points according to (3) in a design space satisfying $H_{\mathcal{X}}$-(i), $H_{\mathcal{X}}$-(ii), with an initialisation such that $\mathbf{M}(\xi_N,\theta)$ is non-singular for all θ in Θ and all $N \geq p$. Let $r_{N,i} = r_N(x^{(i)})$ denote the number of times $x^{(i)}$ appears in the sequence $x_1,\ldots,x_N$, $i = 1,\ldots,K$, and consider the associated order statistics $r_{N,1:K} \geq r_{N,2:K} \geq \cdots \geq r_{N,K:K}$.*

Define

$$q^* = \max\{j : \textit{ there exists } \alpha > 0 \textit{ such that } \liminf_{N\to\infty} r_{N,j:K}/N > \alpha\}\,.$$

Then, H_ϕ-(i) and H_λ-(i) imply $q^* \geq p$. When the sequence $(\hat{\theta}^N)$ is random, the statement holds with probability one.

3.1 Consistency of estimators

Least-squares estimation in nonlinear regression. Consider a regression model with observations

$$Y_i = Y(x_i) = \eta(x_i, \bar{\theta}) + \varepsilon_i\,, \tag{4}$$

with $\bar{\theta}$ in the interior of Θ, a compact subset of $\mathbb{R}^p$, $x_i \in \mathcal{X} \subset \mathbb{R}^d$, and $\{\varepsilon_i\}$ a sequence of independently and identically distributed random variables with $\mathbb{E}\{\varepsilon_1\} = 0$ and $\mathbb{E}\{\varepsilon_1^2\} = \sigma^2 < \infty$ (with $\sigma = 1$ without loss of generality). We assume that the model response $\eta(x, \theta)$ is differentiable with respect to $\theta \in \mathrm{int}(\Theta)$ for any $x \in \mathcal{X}$.

Denote

$$S_N(\theta) = \sum_{k=1}^{N} [Y(x_k) - \eta(x_k, \theta)]^2 \tag{5}$$

and let $\hat{\theta}^N_{LS} = \arg\min_{\theta\in\Theta} S_N(\theta)$ be the least-squares (LS) estimator of θ. The contribution of the design point x to the information matrix is then $\mu(x, \theta) = \mathbf{f}_\theta(x)\,\mathbf{f}_\theta^\top(x)$ with $\mathbf{f}_\theta(x) = \partial\eta(x, \theta)/\partial\theta$. The results can easily be extended to non stationary errors and weighted least-squares. In the case of maximum-likelihood estimation, the contribution of x to the Fisher information matrix only differs by a multiplicative constant.

Define

$$D_N(\theta, \bar{\theta}) = \sum_{k=1}^{N} [\eta(x_k, \theta) - \eta(x_k, \bar{\theta})]^2\,. \tag{6}$$

Next theorem shows that the consistency of the LS estimator is a consequence of $D_N(\theta, \bar{\theta})$ tending to infinity fast enough for $\|\theta - \bar{\theta}\| \geq \delta > 0$. The fact that the design space $\mathcal{X}$ is finite makes the required rate of increase for $D_N(\theta, \bar{\theta})$ quite slow. The result is valid whether the x_k's are non-random constants or are generated via a sequential design algorithm such as (3).

Theorem 1 *Let (x_i) be a non-random design sequence* on a finite set $\mathcal{X}$. *If $D_N(\theta,\bar{\theta})$ given by (6) satisfies*

$$\text{for all} \quad \delta > 0\,, \quad \left[\inf_{\|\theta-\bar{\theta}\|\geq\delta} D_N(\theta,\bar{\theta})\right]/(\log\log N) \to \infty\,, \;\; N \to \infty\,, \tag{7}$$

then $\hat{\theta}^N_{LS} \stackrel{\text{a.s.}}{\to} \bar{\theta}$ as $N \to \infty$ (almost sure convergence). The result remains valid for (x_i) a random sequence on $\mathcal{X}$ finite when (7) holds almost surely.

The proof is given in [25] and is based on Lemma 1 in [31]. Note that the condition (7) is much less restrictive than the classical one for strong consistency of LS estimation in nonlinear regression ($D_N(\theta,\bar{\theta}) = \mathcal{O}(N)$ for $\theta \neq \bar{\theta}$), see [18]. It is also less restrictive than the condition obtained in [20] for sequential design.

Consider the following identifiability assumption on the regression model (4).

$\mathbf{H}_{\mathcal{X}}$-(iii): For all $\delta > 0$ there exists $\epsilon(\delta) > 0$ such that for any subset $\{i_1,\ldots,i_p\}$ of distinct elements of $\{1,\ldots,K\}$, $\inf_{\|\theta-\bar{\theta}\|\geq\delta} \sum_{j=1}^p [\eta(x^{(i_j)},\theta) - \eta(x^{(i_j)},\bar{\theta})]^2 > \epsilon(\delta)$.

For any sequence $(\hat{\theta}^N)$ used in (3), the conditions of Lemma 1 ensure the existence of N_1 and $\alpha > 0$ such that $r_{N,j:K} > \alpha N$ for all $N > N_1$ and all $j = 1,\ldots,p$. Under the additional assumption $\mathrm{H}_{\mathcal{X}}$-(iii) we thus obtain that $D_N(\theta,\bar{\theta})$ given by (6) satisfies $\inf_{\|\theta-\bar{\theta}\|\geq\delta} D_N(\theta,\bar{\theta}) > \alpha N\epsilon(\delta)$, $N > N_1$. Therefore, $\hat{\theta}^N_{LS} \stackrel{\text{a.s.}}{\to} \bar{\theta}$ ($N \to \infty$) from Theorem 1. Since this holds for any sequence $(\hat{\theta}^N)$ in Θ, it is true in particular when $\hat{\theta}^N_{LS}$ is substituted for $\hat{\theta}^N$ in (3).

Maximum-likelihood estimation in Bernoulli trials. Consider now the case of dose-response experiments with

$$Y \in \{0,1\}\,, \quad \text{with} \;\; \mathrm{Prob}\{Y = 1|x_i,\theta\} = \pi(x_i,\theta)\,. \tag{8}$$

We suppose that Θ is a compact subset of $\mathbb{R}^p$, that $\bar{\theta}$, the 'true' value of θ that generates the observations, lies in the interior of Θ, and that $\pi(x,\theta) \in (0,1)$ for any $\theta \in \Theta$ and $x \in \mathcal{X}$. We also assume that $\pi(x,\theta)$ is differentiable with respect to $\theta \in \mathrm{int}(\Theta)$ for any $x \in \mathcal{X}$.

The log-likelihood for the observation Y at the design point x is given by $l(Y,x;\theta) = Y\log[\pi(x,\theta)] + (1-Y)\log[1-\pi(x,\theta)]$ and the contribution of the point x to the Fisher information matrix can be written as $\mu(x,\theta) = \mathbf{f}_\theta(x)\mathbf{f}_\theta^\top(x)$ with

$$\mathbf{f}_\theta(x) = \frac{1}{\sqrt{\pi(x,\theta)[1-\pi(x,\theta)]}} \, \frac{\partial \pi(x,\theta)}{\partial \theta} \,. \tag{9}$$

Suppose that N observations $Y_1,\ldots,Y_N$ are performed at the design points $X_1,\ldots,X_N$, with the Y_i's independent conditionally on the X_i's, so that the conditional likelihoods satisfy $\mathcal{L}(Y_i|X_i, Y_{j\neq i}, X_{j\neq i}, \theta) = \mathcal{L}(Y_i|X_i,\theta)$ for all i. Suppose that X_k is a non-random function of $Y_1,\ldots,Y_{k-1}$, $X_1,\ldots,X_{k-1}$ for all k. The log-likelihood for these observations is then $L_N(\theta) = \sum_{i=1}^N l(Y_i, x_i; \theta)$. We shall denote $\hat\theta_{ML}^N$ the Maximum-Likelihood (ML) estimator of θ, given by $\hat\theta_{ML}^N = \arg\max_{\theta\in\Theta} L_N(\theta)$.

Lemma 2 *If for any $\delta > 0$*

$$\liminf_{N\to\infty} \inf_{\|\theta-\bar\theta\|\geq\delta} [L_N(\bar\theta) - L_N(\theta)] > 0 \quad \textit{almost surely},$$

then $\hat\theta_{ML}^N \overset{\text{a.s.}}{\to} \bar\theta$ as $N \to \infty$.

The proof is identical to that of Lemma 1 in [31]. We then obtain a property similar to Theorem 1, see [25].

Theorem 2 *Let (x_i) be a non-random design sequence* on a finite set $\mathcal{X}$. *Assume that*

$$D_N(\theta,\bar\theta) = \sum_{i=1}^N \pi(x_i,\bar\theta)\, \log\left[\frac{\pi(x_i,\bar\theta)}{\pi(x_i,\theta)}\right] + [1-\pi(x_i,\bar\theta)]\, \log\left[\frac{1-\pi(x_i,\bar\theta)}{1-\pi(x_i,\theta)}\right] \tag{10}$$

satisfies (7). Then, $\hat\theta_{ML}^N \overset{\text{a.s.}}{\to} \bar\theta$ as $N\to\infty$ in the model (8). The same is true for (x_i) a random sequence such that (7) holds almost surely.

Consider the following identifiability assumption for the Bernoulli model.

$\mathbf{H}_{\mathcal{X}}$-(iii'): For all $\delta > 0$ there exists $\epsilon(\delta) > 0$ such that for any subset $\{i_1,\ldots,i_p\}$ of distinct elements of $\{1,\ldots,K\}$, $\inf_{\|\theta-\bar\theta\|\geq\delta} \sum_{j=1}^p \left[\pi(x^{(i_j)},\bar\theta) - \pi(x^{(i_j)},\theta)\right]^2 > \epsilon(\delta)$.

Defining $g(a,b) = a\log(a/b) + (1-a)\log[(1-a)/(1-b)]$, $a,b \in (0,1)$, we can easily check that, for any fixed $a \in (0,1)$, $g(a,b) > 2(a-b)^2$ with $g(a,a) = 0$, so that each term of the sum (10) is positive. Also, $\mathrm{H}_{\mathcal{X}}$-(iii') implies that

$$\inf_{\|\theta-\bar\theta\|\geq\delta} \sum_{j=1}^p \pi(x^{(i_j)},\bar\theta)\, \log\left[\frac{\pi(x^{(i_j)},\bar\theta)}{\pi(x^{(i_j)},\theta)}\right] + [1-\pi(x^{(i_j)},\bar\theta)]\, \log\left[\frac{1-\pi(x^{(i_j)},\bar\theta)}{1-\pi(x^{(i_j)},\theta)}\right] > \epsilon(\delta) > 0$$

for any $\delta > 0$ and any subset $\{i_1, \ldots, i_p\}$ of distinct elements of $\{1, \ldots, K\}$. Similarly to the case of LS estimation in nonlinear regression, but using now Theorem 2 instead of Theorem 1, we thus obtain that under the conditions of Lemma 1 and with the additional assumption H$_{\mathcal{X}}$-(iii') the ML estimator satisfies $\hat{\theta}^N_{ML} \stackrel{\text{a.s.}}{\to} \bar{\theta}$, $N \to \infty$, when $\hat{\theta}^N_{ML}$ is substituted for $\hat{\theta}^N$ in (3).

3.2 Asymptotic optimality of adaptive penalized D-optimal design

We consider the adaptive design algorithm (3) with $\lambda_N = \lambda^*_N = \lambda^*(\hat{\theta}^N)$, the optimal Lagrange coefficient for the constrained D-optimal design problem with parameters $\hat{\theta}^N$. Define

$$H^*_{\bar{\theta}} = \max_{\xi \in \Xi} \left\{ \log \det \mathbf{M}(\xi, \bar{\theta}) + \lambda^*(\bar{\theta}) \left[C - \Phi(\xi, \bar{\theta})\right] \right\} . \tag{11}$$

By asymptotic optimality we mean that the empirical design measure ξ_N is such that $\hat{\theta}^N \stackrel{\text{a.s.}}{\to} \bar{\theta}$ and

$$H_{\bar{\theta}}(\xi_N) = \log \det \mathbf{M}(\xi_N, \bar{\theta}) + \lambda^*(\bar{\theta}) \left[C - \Phi(\xi_N, \bar{\theta})\right] \stackrel{\text{a.s.}}{\to} H^*_{\bar{\theta}} \, , \; N \to \infty \, . \tag{12}$$

(We thus do not consider the true optimal design for sequential dependent observations, which is extremely difficult to construct, see, e.g., [11, 12] for suboptimal attempts.) One can easily check that the optimal design matrix $\mathbf{M}(\xi^*, \bar{\theta})$ is unique, and (12) is thus equivalent to $\mathbf{M}(\xi_N, \bar{\theta}) \stackrel{\text{a.s.}}{\to} \mathbf{M}[\xi^*(\bar{\theta}), \bar{\theta}]$, $N \to \infty$, that is, ξ_N tends to be an optimal constrained design for $\bar{\theta}$. We state this property below as a theorem (the proof is given in Appendix). We use the following additional assumptions.

H$_{\mathcal{X}}$-(iv): For any subset $\{i_1, \ldots, i_p\}$ of distinct elements of $\{1, \ldots, K\}$,

$$\Lambda_{\min} \left[\sum_{j=1}^{p} \mathbf{f}_{\bar{\theta}}(x^{(i_j)}) \mathbf{f}^\top_{\bar{\theta}}(x^{(i_j)}) \right] \geq \bar{\gamma} > 0 \, .$$

H$_f$-(i): For all x in $\mathcal{X}$, $\mathbf{f}_\theta(x)$ is a continuous function of θ in the interior of Θ.

H$_\phi$-(ii): For all x in $\mathcal{X}$, $\phi(x, \theta)$ is a continuous function of θ in the interior of Θ.

Theorem 3 *Suppose that in the regression model (4) (respectively, in the Bernoulli model (8)) the design points for $N > p$ are generated sequentially according to (3), where $\lambda_N = \lambda^*(\hat{\theta}^N)$, the optimal Lagrange coefficient for the constrained D-optimal design problem with parameters $\hat{\theta}^N$, with $\hat{\theta}^N = \hat{\theta}^N_{LS}$ (respectively,*

$\hat{\theta}^N = \hat{\theta}^N_{ML}$). Suppose, moreover, that the first p design points are such that the information matrix is nonsingular for any $\theta \in \Theta$. Then, under $H_{\mathcal{X}}$-(i), $H_{\mathcal{X}}$-(ii), $H_{\mathcal{X}}$-(iii) (respectively, $H_{\mathcal{X}}$-(iii')), $H_{\mathcal{X}}$-(iv), H_λ-(i'), H_ϕ-(i), H_ϕ-(ii) and H_f-(i) we have $\hat{\theta}^N_{LS} \stackrel{\text{a.s.}}{\to} \bar{\theta}$ (respectively, $\hat{\theta}^N_{ML} \stackrel{\text{a.s.}}{\to} \bar{\theta}$) and $\mathbf{M}(\xi_N, \bar{\theta}) \stackrel{\text{a.s.}}{\to} \mathbf{M}[\xi^*(\bar{\theta}), \bar{\theta}]$, $N \to \infty$, with $\xi^*(\bar{\theta})$ a constrained D-optimal design for $\bar{\theta}$.

Notice that Theorem 3 also applies in the case where the sequence (λ_N) is not adapted to $\hat{\theta}^N$ but is simply controlled so as to satisfy a suitable compromise between designing for precise estimation of θ and cost-minimization, and satisfies $\lambda_N \stackrel{\text{a.s.}}{\to} \lambda$, $N \to \infty$, for some λ.

Under a fixed design (penalized D-optimal for instance) the information matrix can be considered as a large sample approximation for the variance-covariance matrix of the estimator, thus allowing straightforward statistical inference from the trial. The situation is more complicated for adaptive designs and has been intensively discussed in the literature. Intuitively, the usual asymptotic normality of $\hat{\theta}^N$ ($N \to \infty$) should hold when the sequence (x_N) is such that $\hat{\theta}^N$ is strongly consistent and $\mathbf{M}(\xi_N, \bar{\theta})$ converges to a nonsingular matrix. The corollary below shows that this is indeed the case in the present situation. One may refer e.g. to [8, 28, 14, 13, 35] for statistical inference in dose-finding problems when using up-and-down [19] or bandit methods [15], randomized Pólya-urn [7] etc. In contrast with those approaches, a guaranteed level of precision for the estimation of the model parameters can easily be imposed by choosing the targeted cost C or the value of λ in (2), see [26, Proposition 1]. Our result is a corollary of the lemma below (its proof is given in Appendix), which uses the following regularity assumption for the model.

$\mathbf{H}_f$-(ii): For all x in $\mathcal{X}$, the components of $\mathbf{f}_\theta(x)$ are continuously differentiable with respect to θ in some open neighborhood of $\bar{\theta}$.

Lemma 3 *Assume that H_f-(ii) is satisfied, that the design points belong to a finite set, see $H_{\mathcal{X}}$-(i), and are such that*

$$\liminf_{N\to\infty} \tau_N \Lambda_{\min}[\mathbf{M}(\xi_N, \bar{\theta})] > \Lambda > 0 \ a.s.$$

for some sequence (τ_N) satisfying $\lim_{N\to\infty} \tau_N / N^{1/4} = 0$. Then, if $\hat{\theta}^N_{ML} \stackrel{\text{a.s.}}{\to} \bar{\theta}$ ($N \to \infty$) in the Bernoulli *model (8), we also have*

$$\sqrt{N}\,\mathbf{M}^{1/2}(\xi_N,\hat{\theta}^N_{ML})(\hat{\theta}^N_{ML}-\bar{\theta}) \stackrel{\rm d}{\rightarrow} \omega \sim \mathscr{N}(\mathbf{0},\mathbf{I})\,,\ N\rightarrow\infty\,. \tag{13}$$

The same is true in the regression model (4): when $\hat{\theta}^N_{LS} \stackrel{\rm a.s.}{\rightarrow} \bar{\theta}$ *we also have*

$$\sqrt{N}\,\mathbf{M}^{1/2}(\xi_N,\hat{\theta}^N_{LS})(\hat{\theta}^N_{LS}-\bar{\theta}) \stackrel{\rm d}{\rightarrow} \omega \sim \mathscr{N}(\mathbf{0},\mathbf{I})\,,\ N\rightarrow\infty\,, \tag{14}$$

under the additional condition $\lim_{N\rightarrow\infty}\tau_N\,N^{-\delta/(2+\delta)}=0$ *for some* δ *such that* $\mathbb{E}\{|\varepsilon_1|^{2+\delta}\}<\infty$*.*

Note that when $\lim_{N\rightarrow\infty}\tau_N/N^{1/4}=0$, the condition $\lim_{N\rightarrow\infty}\tau_N\,N^{-\delta/(2+\delta)}=0$ for regression models is not restrictive when moments $\mathbb{E}\{|\varepsilon_1|^\alpha\}$ exist for $\alpha>8/3$. One may also notice that, compared to [28] we do not require that $\mathbf{M}(\xi_n,\bar{\theta})$ tends to a constant matrix, compared to [31] we do not require that $\tau_N\mathbf{M}(\xi_N,\bar{\theta})$ tends to some positive definite matrix, and compared to [21, 20] we do not require the existence of non-random matrices $\mathbf{C}_N$ such that $\mathbf{C}_N\mathbf{M}^{1/2}(\xi_N,\bar{\theta}) \stackrel{\rm p}{\rightarrow} \mathbf{I}$, $N\rightarrow\infty$. On the other hand, we need that $\Lambda_{\min}[\mathbf{M}(\xi_N,\bar{\theta})]$ decreases more slowly than $N^{-1/4}$. The lemma applies to more general designs than (1, 3). In particular, adaptive rules on a finite design space that have a non degenerate limiting distribution ξ_∞ (such that $\det\mathbf{M}(\xi_\infty,\theta)\neq 0$) satisfy the conditions of the lemma with $\tau_N\equiv 1$. This is the case in particular for up-and-down methods in clinical trials, see, e.g., [8, 14, 17].

Under the conditions of Theorem 3, there exist N_0 and $\alpha>0$ such that, for all $N>N_0$ we have $\Lambda_{\min}[\mathbf{M}(\xi_N,\bar{\theta})]>\alpha\bar{\gamma}$, with $\bar{\gamma}$ as in $\mathrm{H}_{\mathscr{X}}$-(iv). We can thus take $\tau_N\equiv 1$ in Lemma 3 and obtain the following property, which indicates that it is legitimate (asymptotically) to characterize the precision of the estimation by the inverse information matrix $\mathbf{M}^{-1}(\xi_N,\hat{\theta}^N)$ when using the adaptive design scheme (3) on a finite design set $\mathscr{X}$.

Corollary 1 *Under the conditions of Theorem 3, and assuming that, moreover,* H_f*-(ii) is satisfied, the ML estimator in the model (8) satisfies (13) and the LS estimator in the model (4) satisfies (14).*

4 Asymptotic properties of adaptive design with increasing penalty

We consider now the case where the sequence (λ_N) of Lagrange coefficients in (3) is unbounded and satisfies

H$_\lambda$-(ii): (λ_N) is a non-decreasing positive sequence and $\lim_{N\to\infty}\lambda_N=\infty$.

Replacing H$_\lambda$-(i) by H$_\lambda$-(ii) in the assumptions of Sect. 3, we obtain the following lower bound on the sampling rate of nonsingular designs. The proof is given in Appendix.

Lemma 4 *Let $(\hat{\theta}^N)$ be an arbitrary sequence in Θ used to generate design points according to (3) in a design space satisfying $H_\mathscr{X}$-(i), $H_\mathscr{X}$-(ii), with an initialisation such that $\mathbf{M}(\xi_N,\theta)$ is non-singular for all θ in Θ and all $N\geq p$. Let $r_{N,j:K}$ be defined as in Lemma 1, $j=1,\ldots,K$, and define*

$$q^*=\max\{j:\ \textit{there exists } \alpha>0 \textit{ such that } \liminf_{N\to\infty}\lambda_N\, r_{N,j:K}/N>\alpha\}\,.$$

Then, H_ϕ-(i) and H_λ-(ii) imply $q^*\geq p$. When the sequence $(\hat{\theta}^N)$ is random, the statement holds with probability one.

4.1 Consistency of estimators

As in Sect. 3.1, we consider the consistency of the LS and ML estimators in regression and dose-response experiments respectively, but now in the case where (λ_N) is an unbounded increasing sequence of penalty coefficients. We show that strong consistency is ensured provided that this sequence tends to infinity slowly enough.

Least-squares estimation in nonlinear regression. For any sequence $(\hat{\theta}^N)$ used in (3), the conditions of Lemma 4 ensure the existence of N_1 and $\alpha>0$ such that $r_{N,j:K}>\alpha N/\lambda_N$ for all $N>N_1$ and all $j=1,\ldots,p$. Under the additional assumption H$_\mathscr{X}$-(iii) we thus obtain that $D_N(\theta,\bar{\theta})$ given by (6) satisfies

$$\frac{1}{\log\log N}\inf_{\|\theta-\bar{\theta}\|\geq\delta}D_N(\theta,\bar{\theta})>\frac{\alpha N\epsilon(\delta)}{\lambda_N\,\log\log N}\,,\quad N>N_1\,.$$

Therefore, if $(\lambda_N \log\log N)/N \to 0$ when $N \to \infty$, $\hat{\theta}^N_{LS} \stackrel{\text{a.s.}}{\to} \bar{\theta}$ from Theorem 1. Since this holds for any sequence $(\hat{\theta}^N)$ in Θ, it is true in particular when $\hat{\theta}^N_{LS}$ is substituted for $\hat{\theta}^N$ in (3).

Maximum-likelihood estimation in Bernoulli trials. The situation is similar to previous one. Using now Theorem 2 instead of Theorem 1, we obtain that under the conditions of Lemma 4 and with the additional assumption $\mathrm{H}_{\mathcal{X}}$-(iii') the ML estimator satisfies $\hat{\theta}^N_{ML} \stackrel{\text{a.s.}}{\to} \bar{\theta}$, $N \to \infty$, when $\hat{\theta}^N_{ML}$ is substituted for $\hat{\theta}^N$ in (3) with $(\lambda_N \log\log N)/N \to 0$ when $N \to \infty$.

4.2 Convergence to minimum-cost design and asymptotic normality

The following theorem shows that using the following assumptions

$\mathbf{H}_\lambda$-(iii): the sequence (λ_N) is such that λ_N/N is non-increasing with $(\lambda_N \log\log N)/N \to 0$, $N \to \infty$;

$\mathbf{H}_\phi$-(iii): $\phi(x,\bar{\theta})$ has a unique global minimizer x^*: $\forall \beta > 0$, $\exists \epsilon > 0$ such that $\phi(x,\bar{\theta}) < \phi(x^*,\bar{\theta}) + \epsilon$ implies $\|x - x^*\| < \beta$;

in complement of H_λ-(ii), the adaptive design algorithm (3) is such that (x_N) tends to accumulate at the point of minimum cost for $\bar{\theta}$. The proof is given in Appendix.

Theorem 4 *Suppose that in the regression model (4) (respectively, in the Bernoulli model (8)) the design points for $N > p$ are generated sequentially according to (3), where λ_N satisfies H_λ-(ii) and H_λ-(iii). Suppose, moreover, that the first p design points are such that the information matrix is nonsingular for any $\theta \in \Theta$. Then, under $H_{\mathcal{X}}$-(i), $H_{\mathcal{X}}$-(ii), $H_{\mathcal{X}}$-(iii) (respectively, $H_{\mathcal{X}}$-(iii')), $H_{\mathcal{X}}$-(iv), H_ϕ-(i), H_ϕ-(ii), and H_f-(i) we have $\hat{\theta}^N_{LS} \stackrel{\text{a.s.}}{\to} \bar{\theta}$ (respectively, $\hat{\theta}^N_{ML} \stackrel{\text{a.s.}}{\to} \bar{\theta}$) and*

$$\Phi(\xi_N,\bar{\theta}) \stackrel{\text{a.s.}}{\to} \phi^*_{\bar{\theta}} = \min_{x\in\mathcal{X}} \phi(x,\bar{\theta})\,,\ N \to \infty\,. \tag{15}$$

If, moreover, H_ϕ-(iii) is satisfied, then

$$\xi_N \stackrel{\text{w}}{\to} \delta_{x^*} \text{ almost surely}\,,\ N \to \infty\,, \tag{16}$$

with $\stackrel{\text{w}}{\to}$ denoting the weak convergence of probability measures and δ_{x^*} the delta measure at $x^* = \arg\min_{x\in\mathcal{X}} \phi(x,\bar{\theta})$.

The property (16) does not imply that the x_k's generated by (3) converge to x^*. However, Proposition 2 of [26] states a condition that guarantees that the support points of an optimal design measure converge to x^* as $\lambda \to \infty$. When $\mathscr{X}$ is obtained by the discretization of a compact set $\mathscr{X}'$ and this condition is satisfied at $\theta = \bar{\theta}$ for design measures on $\mathscr{X}'$, then the points x_N will gather around x^* as $N \to \infty$, see the example in Section 4 of [26].

Convergence results similar to those in Theorem 4 are obtained in [24] for LS estimation in a *linear* regression model, without the assumption that $\mathscr{X}$ is finite, but under more restrictive conditions than H_λ-(ii), H_λ-(iii) on the growth rate of the sequence (λ_N).

Under the conditions of Theorem 4, there exist N_0 and $\alpha > 0$ such that, for all $N > N_0$, $\Lambda_{\min}[\mathbf{M}(\xi_N, \bar{\theta})] > \alpha\bar{\gamma}/\lambda_N$, with $\bar{\gamma}$ as in $\text{H}_{\mathscr{X}}$-(iv). We use again Lemma 3, with now $\tau_N = \lambda_N$ and obtain the following.

Corollary 2 *Under the conditions of Theorem 4 and assuming that, moreover, H_f-(ii) is satisfied and* $\lim_{N\to\infty} \lambda_N/N^{1/4} = 0$*, the ML estimator in the Bernoulli model (8) satisfies (13). Also, the LS estimator in the regression model (4) satisfies (14) under the additional assumption* $\lim_{N\to\infty} \lambda_N\, N^{-\delta/(2+\delta)} = 0$ *for some δ such that* $\mathbb{E}\{|\varepsilon_1|^{2+\delta}\} < \infty$.

5 Conclusion and further developments

We have shown that for finite design spaces, adaptive penalized D-optimal design possess the asymptotic optimality properties one may expect: strong consistency and asymptotic normality of the estimators, convergence to an optimal design for the true value of the model parameters. We mention below some extensions of this work, some straightforward, others more challenging, and indicate a motivating objective concerning the design of non-stationary dose-finding experiments preserving individual ethics.

Bayesian estimation The extension of the results presented to Bayesian estimators should not raise particular difficulties, especially since consistency is usually easier to obtained than for LS or ML estimation using martingales properties, see, e.g., [16]. A straightforward modification of (3) is to replace $\mathbf{M}(\xi_N, \hat{\theta}^N)$ by $[\mathbf{M}^{-1}(\xi_N, \hat{\theta}^N) + \Omega/N]^{-1}$, with Ω the prior covariance matrix for the model parameters.

Multiple constraints The results obtained in Sect. 3 and 4 easily generalize to the case when several constraints are present and one maximizes $\log\det \mathbf{M}(\xi,\theta)$ with respect to $\xi \in \Xi$ subject to $\Phi(\xi,\theta) \leq C_i$, $i = 1,\ldots,m$. A Lagrange coefficient is then associated with each constraint and the design algorithm (3) becomes

$$x_{N+1} = \arg\max_{x\in\mathcal{X}} \left\{ \mathrm{trace}[\mu(x,\hat{\theta}^N)\mathbf{M}^{-1}(\xi_N,\hat{\theta}^N)] - \sum_{i=1}^{m} \lambda_N^{(i)}\, \phi_i(x,\hat{\theta}^N) \right\} .$$

When the $\lambda_N^{(i)}$'s are controlled to increase to infinity, define $\rho_N^{(j)} = \lambda_N^{(j)} / \sum_i \lambda_N^{(i)}$ and suppose that a limit $\bar{\rho}_j$ exists for each $\rho_N^{(j)}$, $j = 1,\ldots,m$. Then, if all cost functions $\phi_i(\cdot,\theta)$ are bounded on $\mathcal{X}$, the asymptotic behaviors of the design and estimators are the same as in Sect. 4 for $\lambda_N = \sum_i \lambda_N^{(i)}$ and $\phi(x,\theta) = \sum_j \bar{\rho}_j \phi_j(x,\theta)$. Also, the developments of Sect. 3 remain valid when the $\lambda_N^{(i)}$'s are kept constant, or when they are adapted to $\hat{\theta}^N$, that is, when they correspond to the optimal coefficients for the constrained problem with parameters $\hat{\theta}^N$. Note, however, that the presence of several constraints makes this optimal solution more difficult to determine since it can no longer be obtained by solving a series of unconstrained problems with an increasing sequence of coefficients, contrary to the single constraint case. The Lagrangian approach proposed in Sect. 2.3 of [3] can then provide a solution at reasonable cost.

Finite horizon The results of Sect. 4 indicate that, when λ_N increases to infinity at suitable speed, the design converges to the delta measure located at the optimum. In clinical trials, this means that more and more patients receive doses close to the optimal one (and none receives extreme doses if the penalty function satisfies the condition given in [26]). This is an asymptotic result, however, and approaching the optimal solution over a finite horizon is a challenging task. In the case of LS estimation in linear regression, design strategies are suggested in [27] that are shown to be close to the optimum control (stochastic dynamic programming) solution. Is is then tempting to replace the Bernoulli model by a regression type model (observation Y_k at design point x_k has mean value $\pi(x_k,\theta)$ and variance $\pi(x_k,\theta)[1-\pi(x_k,\theta)]$), as suggested in [30], with a linear parametrization, and then try to apply the finite-horizon results of [27].

Non stationary clinical trials Strong consistency of the estimator is obtained in Sect. 4 when the penalty coefficient λ_N in (3) tends to infinity. Moreover, when the growth of λ_N is not too fast, the estimator is asymptotically normal. This means that, although the design is non-stationary in the sense that patients enrolled in the trial receive better and better treatments, the information collected at the end of the experiment can be used to set future treatments. In particular, the minimum effective and maximum tolerated doses can be estimated and confidence intervals can be given. At the same time, the fact that patients receive unequal treatments in the trial raises ethical issues: there is no randomization, a new patient tends to receive a better treatment than patients previously treated (since when λ_N increases, the design points tend to get closer to the optimal dose). This emphasizes the importance of constructing a fair rule for choosing the increasing sequence (λ_N). Trying to give equal probabilities of success at patients P_N and P_{N+1} when patient P_N is treated first seems to be an honest ambition, and the increase of λ_N could then be used to compensate for the late treatment of patient P_{N+1}. This requires a model for the evolution of the probability of success as a function of the delay in treatment, to be combined with a suitable characterization of the improvement of treatment that can be expected when increasing λ_N.

Appendix

Proof of Lemma 1. First note that $q^* \geq 1$ since $\mathcal{X}$ is finite. Suppose that $p \geq 2$ and $q^* < p$. We show that this leads to a contradiction.

For any N we can write

$$\begin{aligned}\mathbf{M}(\xi_N, \theta) &= \frac{1}{N} \sum_{k=1}^{N} \mathbf{f}_\theta(x_k) \mathbf{f}_\theta^\top(x_k) \\ &= \frac{1}{N} \sum_{i=1}^{q^*} r_{N,i:K}\, \mathbf{f}_\theta(x^{(i_N)}) \mathbf{f}_\theta^\top(x^{(i_N)}) + \frac{1}{N} \sum_{x_k \notin \mathcal{X}_N(q^*)} \mathbf{f}_\theta(x_k) \mathbf{f}_\theta^\top(x_k)\,, \end{aligned} \quad (17)$$

where i_N is the index (depending on N) of a design point appearing $r_{N,i:K}$ times in $x_1, \ldots, x_N$ and $\mathcal{X}_N(q^*) = \{x^{(1_N)}, \ldots, x^{(q_N^*)}\}$ is the set of such points for $i \leq q^*$. Let $\mathbf{M}_N(\theta)$ denote the first matrix on the right-hand side of (17). For any $x^{(i_N)} \in \mathcal{X}_N(q^*)$ we have

$$\mathbf{f}_\theta^\top(x^{(i_N)})\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{f}_\theta(x^{(i_N)}) \leq \mathbf{f}_\theta^\top(x^{(i_N)})\mathbf{M}_N^-(\theta)\mathbf{f}_\theta(x^{(i_N)}) = \frac{N}{r_{N,i:K}}\,,$$

with $\mathbf{M}_N^-(\theta)$ any g-inverse of $\mathbf{M}_N(\theta)$. Therefore, from the definition of q^*, there exists N_1 such that

$$\text{for all} \ \ i \leq q^*\,,\ N > N_1 \text{ and } \theta \in \Theta\,,\ \mathbf{f}_\theta^\top(x^{(i_N)})\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{f}_\theta(x^{(i_N)}) - \lambda_N\,\phi(x^{(i_N)},\theta) \leq \frac{1}{\alpha}\,. \tag{18}$$

Let $\beta_N = r_{N,(q^*+1):K}/N$. Showing that $\liminf_{N\to\infty} \beta_N \geq \underline{\beta}$ for some $\underline{\beta} > 0$ will contradict the definition of q^*.

Define

$$\mathbf{M}_N^{(1)}(\theta) = \sum_{i=1}^{q^*} \mathbf{f}_\theta(x^{(i_N)})\mathbf{f}_\theta^\top(x^{(i_N)})\,,\ \ \mathbf{M}_N^{(2)}(\theta) = \sum_{i=1}^{K} \mathbf{f}_\theta(x^{(i_N)})\mathbf{f}_\theta^\top(x^{(i_N)})\,. \tag{19}$$

We have $(1-\beta_N)\,\mathbf{M}_N^{(1)}(\theta)+\beta_N\,\mathbf{M}_N^{(2)}(\theta)-\mathbf{M}(\xi_N,\theta) \in \mathbb{M}^{\geq}$, where $\mathbb{M}^{\geq}$ is the set of symmetric nonnegative definite $p\times p$ matrices. For any $\mathbf{u}\in\mathbb{R}^p$,

$$\begin{aligned}
\mathbf{u}^\top\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{u} \ \ &\geq \ \ \mathbf{u}^\top[(1-\beta_N)\,\mathbf{M}_N^{(1)}(\theta)+\beta_N\,\mathbf{M}_N^{(2)}(\theta)]^{-1}\mathbf{u} \\
&= \max_{\mathbf{z}\in\mathbb{R}^p} 2\mathbf{z}^\top\mathbf{u} - \mathbf{z}^\top[(1-\beta_N)\,\mathbf{M}_N^{(1)}(\theta)+\beta_N\,\mathbf{M}_N^{(2)}(\theta)]\mathbf{z} \\
&\geq \max_{\mathbf{z}\in\mathcal{N}[\mathbf{M}_N^{(1)}(\theta)]} 2\mathbf{z}^\top\mathbf{u} - \mathbf{z}^\top[(1-\beta_N)\,\mathbf{M}_N^{(1)}(\theta)+\beta_N\,\mathbf{M}_N^{(2)}(\theta)]\mathbf{z}
\end{aligned}$$

with $\mathcal{N}(\mathbf{M}) = \{\mathbf{v} : \mathbf{M}\mathbf{v} = \mathbf{0}\}$ the null-space of the matrix $\mathbf{M}$. Direct calculations then give

$$\mathbf{u}^\top\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{u} \geq \frac{1}{\beta_N}\,\mathbf{u}^\top[\mathbf{M}_N^{(2)}(\theta)]^{-1}\,[\mathbf{I}-\mathbf{P}_N(\theta)]\,\mathbf{u} \tag{20}$$

with $\mathbf{I}$ the p-dimensional identity matrix and $\mathbf{P}_N(\theta)$ the projector

$$\mathbf{P}_N(\theta) = \mathbf{M}_N^{(1)}(\theta)\left[\mathbf{M}_N^{(1)}(\theta)[\mathbf{M}_N^{(2)}(\theta)]^{-1}\mathbf{M}_N^{(1)}(\theta)\right]^-\mathbf{M}_N^{(1)}(\theta)[\mathbf{M}_N^{(2)}(\theta)]^{-1}\,.$$

Note that the right-hand side of (20) is zero when $\mathbf{u}\in\mathcal{M}[\mathbf{M}_N^{(1)}(\theta)]$ (i.e., when $\mathbf{u} = \mathbf{M}_N^{(1)}(\theta)\mathbf{v}$ for some $\mathbf{v}\in\mathbb{R}^p$). When $\mathbf{u} = \mathbf{f}_\theta(x^{(i_N)})$ for some $i\in\{q^*+1,\ldots,K\}$ we can construct a lower bound for this term, of the form A/β_N with A constant. Indeed, from (19) and $\mathrm{H}_{\mathcal{X}}$-(ii),

$$\text{for all} \ \ \theta\in\Theta \text{ and } \mathbf{v}\in\mathbb{R}^p\,,\ \mathbf{v}^\top\left[\mathbf{M}_N^{(1)}(\theta) + \sum_{i=q^*+1}^{K} \mathbf{f}_\theta(x^{(i_N)})\mathbf{f}_\theta^\top(x^{(i_N)})\right]\mathbf{v} > \gamma\,\|\mathbf{v}\|^2$$

so that for all $\theta \in \Theta$ and $\mathbf{z} \in \mathcal{N}[\mathbf{M}_N^{(1)}(\theta)]$,

$$\max_{i=q^*+1,\ldots,K} [\mathbf{z}^\top \mathbf{f}_\theta(x^{(i_N)})]^2 > \frac{\gamma}{K-q^*} \, \|\mathbf{z}\|^2 \, . \tag{21}$$

Take $\mathbf{z} = \mathbf{z}_{\theta,i_N} = [\mathbf{M}_N^{(2)}(\theta)]^{-1}[\mathbf{I} - \mathbf{P}_N(\theta)]\mathbf{f}_\theta(x^{(i_N)})$ for some $i \in \{q^*+1,\ldots,K\}$, so that $\mathbf{z}_{\theta,i_N} \in \mathcal{N}[\mathbf{M}_N^{(1)}(\theta)]$ and $\mathbf{f}_\theta^\top(x^{(i_N)})[\mathbf{M}_N^{(2)}(\theta)]^{-1}\,[\mathbf{I} - \mathbf{P}_N(\theta)]\,\mathbf{f}_\theta(x^{(i_N)}) = \mathbf{z}_{\theta,i_N}^\top \mathbf{f}_\theta(x^{(i_N)}) = \mathbf{z}_{\theta,i_N}^\top \mathbf{M}_N^{(2)}(\theta)\mathbf{z}_{\theta,i_N}$. We obtain

$$\begin{aligned}
\max_{i=q^*+1,\ldots,K} \mathbf{f}_\theta^\top(x^{(i_N)})[\mathbf{M}_N^{(2)}(\theta)]^{-1}\,[\mathbf{I} - \mathbf{P}_N(\theta)]\,\mathbf{f}_\theta(x^{(i_N)}) &= \max_{i,j=q^*+1,\ldots,K} \mathbf{z}_{\theta,i_N}^\top \mathbf{M}_N^{(2)}(\theta)\mathbf{z}_{\theta,j_N} \\
&= \max_{i,j=q^*+1,\ldots,K} \mathbf{z}_{\theta,i_N}^\top \mathbf{f}_\theta(x^{(j_N)})\,,
\end{aligned}$$

and thus from (21),

$$\text{for all } \ \theta \in \Theta\,, \ \max_{i=q^*+1,\ldots,K} \mathbf{f}_\theta^\top(x^{(i_N)})[\mathbf{M}_N^{(2)}(\theta)]^{-1}\,[\mathbf{I} - \mathbf{P}_N(\theta)]\,\mathbf{f}_\theta(x^{(i_N)}) > \left(\frac{\gamma}{K-q^*}\right)^{1/2} \max_{i=q^*+1,\ldots,K} \|\mathbf{z}_{\theta,i_N^*}\|\,.$$

Let i_N^* denote the argument of the maximum on the left-hand side, for which we have, $\mathbf{z}_{\theta,i_N^*}^\top \mathbf{f}_\theta(x^{(i_N^*)}) = \mathbf{z}_{\theta,i_N^*}^\top \mathbf{M}_N^{(2)}(\theta)\mathbf{z}_{\theta,i_N^*} \ \leq \ K\,L\,\|\mathbf{z}_{\theta,i_N^*}\|^2$ with $L = \max_{x\in\mathcal{X},\,\theta\in\Theta} \|\mathbf{f}_\theta(x)\|^2$. This finally gives: for all $\theta \in \Theta$, $\max_{i=q^*+1,\ldots,K} \mathbf{f}_\theta^\top(x^{(i_N)})[\mathbf{M}_N^{(2)}(\theta)]^{-1}\,[\mathbf{I} - \mathbf{P}_N(\theta)]\,\mathbf{f}_\theta(x^{(i_N)}) > \gamma/[LK(K-q^*)]$, and thus, from (20) and H_ϕ-(i),

$$\text{for all } \ \theta \in \Theta\,, \ \max_{i=q^*+1,\ldots,K} \left\{\mathbf{f}_\theta^\top(x^{(i_N)})\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{f}_\theta(x^{(i_N)}) - \lambda_N\,\phi(x^{(i_N)},\theta)\right\} > \frac{1}{\beta_N}\,\frac{\gamma}{LK(K-q^*)} - \bar{\lambda}\bar{\phi}\,. \tag{22}$$

Together with (18), it gives: $N > N_1$ and $\beta_N < \beta^* = \gamma\alpha/[LK(K-q^*)(1+\alpha\bar{\lambda}\bar{\phi})] \Rightarrow x_{N+1} \notin \mathcal{X}_N(q^*)$ in the sequence (3). Define

$$\beta_N^* = \frac{\sum_{i=q^*+1}^K r_{N,i:K}}{(K-q^*)N}\,,$$

so that $\beta_N \geq \beta_N^* \geq \beta_N/(K-q^*)$. Also, when $N > N_1$, $(\sum_{i=1}^{q^*} r_{N,i:K})/N > q^*\alpha$, and therefore $\beta_N^* < (1-q^*\alpha)/(K-q^*)$. By construction, $\beta_N < \beta^*$ and $N > N_1$ imply

$$\begin{aligned}
\beta_{N+1} \geq \beta_{N+1}^* = \frac{N\beta_N^*(K-q^*)+1}{(K-q^*)(N+1)} &= \beta_N^* + \frac{1}{N+1}\left(\frac{1}{K-q^*} - \beta_N^*\right) \\
&> \beta_N^* + \frac{1}{N+1}\,\frac{q^*\alpha}{K-q^*} \geq \frac{\beta_N}{K-q^*} + \frac{1}{N+1}\,\frac{q^*\alpha}{K-q^*}\,.
\end{aligned}$$

By induction, this lower bound on β_{N+k} increases with k,

$$\beta_{N+k} > \frac{\beta_N}{K-q^*} + \frac{q^*\alpha}{K-q^*}\,\sum_{i=1}^{k}\frac{1}{N+i}\,,$$

until β_{N+k} becomes larger that β^*. Suppose that the threshold β^* is crossed downwards at $N_2 > N_1$, i.e., $\beta_{N_2-1} \geq \beta^*$ and $\beta_{N_2} < \beta^*$. This implies $\beta_{N_2} = \beta_{N_2-1}(N_2-1)/N_2$ and thus $\beta^*(N_2-1)/N_2 \leq \beta_{N_2} < \beta^*$, so that β_{N_2} tends to β^* when $N_2 \to \infty$. We thus obtain $\liminf_{N\to\infty} \beta_N \geq \underline{\beta} = \beta^*/(K-q^*)$, showing that $q^* \geq p$, which concludes the proof. ■

Proof of Theorem 3. We have already seen that the conditions of Lemma 1 with $\mathrm{H}_{\mathcal{X}}$-(iii) (respectively, $\mathrm{H}_{\mathcal{X}}$-(iii')) imply $\hat{\theta}^N_{LS} \overset{\text{a.s.}}{\to} \bar{\theta}$ (respectively, $\hat{\theta}^N_{ML} \overset{\text{a.s.}}{\to} \bar{\theta}$) as $N \to \infty$. All what we need to obtain the asymptotic optimality of ξ_N is thus a continuity property of the form: for all $\epsilon > 0$, there exists $\beta > 0$ such that $[\|\hat{\theta}^N - \bar{\theta}\| < \beta$ for all N larger than some $N_0]$ implies $\liminf_{N\to\infty} H_{\bar{\theta}}(\xi_N) > H^*_{\bar{\theta}} - \epsilon$. We show that this is indeed true under the additional assumptions $\mathrm{H}_{\mathcal{X}}$-(iv), H_{ϕ}-(ii) and H_{f}-(i).

First note that $\lambda^*(\theta) = \arg\min_{\lambda\geq 0} \max_{\xi\in\Xi} \{\log\det \mathbf{M}(\xi,\theta) + \lambda\,[C - \Phi(\xi,\theta)]\}$ is continuous in θ as the argument of the minimum of a convex function (given by the maximum of convex functions) that is continuous in θ. Therefore, $\forall\epsilon_0$, $\exists\beta_1$ such that $\|\theta - \bar{\theta}\| < \beta_1 \Rightarrow |\lambda^*(\theta) - \lambda^*(\bar{\theta})| < \epsilon_0$. Also, H_{f}-(i), H_{ϕ}-(ii) and $\mathrm{H}_{\mathcal{X}}$-(i) imply: $\forall\epsilon_0$, $\exists\beta_2$ such that $\|\theta - \bar{\theta}\| < \beta_2 \Rightarrow \max_{x\in\mathcal{X}} \|\mathbf{f}_\theta(x) - \mathbf{f}_{\bar{\theta}}(x)\| < \epsilon_0$ and $\exists\beta_3$ such that $\|\theta - \bar{\theta}\| < \beta_3 \Rightarrow \max_{x\in\mathcal{X}} |\phi(x,\theta) - \phi(x,\bar{\theta})| < \epsilon_0$.

From Lemma 1, there exits N_1 and $\alpha > 0$ such that for all $N > N_1$, $r_{N,j:K} > \alpha N$, $j = 1,\ldots,q^*$, with $q^* > p$, and thus from $\mathrm{H}_{\mathcal{X}}$-(iv), $\Lambda_{\min}[\mathbf{M}(\xi_N,\bar{\theta})] > \alpha\bar{\gamma}$. Direct calculations then give

$$\max_{x\in\mathcal{X}} \max_{\|\theta-\bar{\theta}\|<\beta_2} |\mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{f}_\theta(x) - \mathbf{f}_{\bar{\theta}}^\top(x)\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{f}_{\bar{\theta}}(x)| < B\epsilon_0$$

for ϵ_0 small enough and $N > N_1$, with B depending on $\alpha, \bar{\gamma}$ and $\bar{f} = \max_{x\in\mathcal{X}} \|\mathbf{f}_{\bar{\theta}}(x)\|$.

Therefore, for all ϵ_0 small enough, we can take $\beta = \min\{\beta_1,\beta_2,\beta_3\}$ and obtain, for all $N > N_1$ and $\|\hat{\theta}^N - \bar{\theta}\| < \beta$ in (3),

$$\begin{aligned}
&\mathbf{f}_{\bar{\theta}}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{f}_{\bar{\theta}}(x_{N+1}) - \lambda^*(\bar{\theta})\phi(x_{N+1},\bar{\theta}) \\
&\qquad > \mathbf{f}_{\hat{\theta}^N}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\hat{\theta}^N)\mathbf{f}_{\hat{\theta}^N}(x_{N+1}) - \lambda^*(\hat{\theta}^N)\phi(x_{N+1},\hat{\theta}^N) - \epsilon_0(B + \bar{\lambda} + \bar{\phi}) - \epsilon_0^2 \\
&\qquad = \max_{x\in\mathcal{X}} \left[\mathbf{f}_{\hat{\theta}^N}^\top(x)\mathbf{M}^{-1}(\xi_N,\hat{\theta}^N)\mathbf{f}_{\hat{\theta}^N}(x) - \lambda^*(\hat{\theta}^N)\phi(x,\hat{\theta}^N)\right] - \epsilon_0(B + \bar{\lambda} + \bar{\phi}) - \epsilon_0^2 \\
&\qquad > \max_{x\in\mathcal{X}} \left[\mathbf{f}_{\bar{\theta}}^\top(x)\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{f}_{\bar{\theta}}(x) - \lambda^*(\bar{\theta})\phi(x,\bar{\theta})\right] - 2\epsilon_0(B + \bar{\lambda} + \bar{\phi})\,.
\end{aligned}$$

For a given ϵ, take $\epsilon_0 = \epsilon/(B + \bar{\lambda} + \bar{\phi})$ and β as above, and suppose that there exists $\delta > 0$ such that

$$H_{\bar{\theta}}(\xi_N) < H_{\bar{\theta}}^* - \delta - \epsilon \tag{23}$$

for all N larger than some N_2, with $H_{\bar{\theta}}(\xi_N)$ and $H_{\bar{\theta}}^*$ respectively given by (12) and (11). From the concavity of the design criterion, this implies

$$\max_{x\in\mathcal{X}} \left[\mathbf{f}_{\bar{\theta}}^\top(x)\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{f}_{\bar{\theta}}(x) - \lambda^*(\bar{\theta})\phi(x,\bar{\theta})\right] > p - \lambda^*(\bar{\theta})\Phi(\xi_N,\bar{\theta}) + \epsilon + \delta$$

and thus

$$\mathbf{f}_{\bar{\theta}}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{f}_{\bar{\theta}}(x_{N+1}) - \lambda^*(\bar{\theta})\phi(x_{N+1},\bar{\theta}) > p - \lambda^*(\bar{\theta})\Phi(\xi_N,\bar{\theta}) + \delta \tag{24}$$

for all $N > \max\{N_1, N_2, N_0\}$ when $\|\hat{\theta}^N - \bar{\theta}\| < \beta$ for all $N > N_0$. Direct calculations give

$$\begin{aligned} H_{\bar{\theta}}(\xi_{N+1}) - H_{\bar{\theta}}(\xi_N) \quad &= \quad \log\left[1 + \frac{\mathbf{f}_{\bar{\theta}}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{f}_{\bar{\theta}}(x_{N+1})}{N}\right] - p\log\left(1+\frac{1}{N}\right) \\ &\quad + \frac{\lambda^*(\bar{\theta})}{N+1}\left[\Phi(\xi_N,\bar{\theta}) - \phi(x_{N+1},\bar{\theta})\right] \end{aligned} \tag{25}$$

and thus from (24),

$$\begin{aligned} H_{\bar{\theta}}(\xi_{N+1}) - H_{\bar{\theta}}(\xi_N) \quad &> \quad \log\left[1 + \frac{\mathbf{f}_{\bar{\theta}}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{f}_{\bar{\theta}}(x_{N+1})}{N}\right] - p\log\left(1+\frac{1}{N}\right) \\ &\quad + \frac{p + \delta - \mathbf{f}_{\bar{\theta}}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{f}_{\bar{\theta}}(x_{N+1})}{N+1}\,. \end{aligned}$$

Since $\mathbf{f}_{\bar{\theta}}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{f}_{\bar{\theta}}(x_{N+1})$ is bounded by $\bar{f}^2/(\alpha\bar{\gamma})$, we obtain that $H_{\bar{\theta}}(\xi_{N+1}) - H_{\bar{\theta}}(\xi_N) > \delta/(2N)$ for N large enough, which implies that $H_{\bar{\theta}}(\xi_N) \to \infty$ as $N \to \infty$, in contradiction with H$_\phi$-(i), H$_\lambda$-(i') and H$_f$-(i), see the definition of $H_{\bar{\theta}}(\xi)$ in (12).

Therefore, for any $\epsilon > 0$, $\|\hat{\theta}^N - \bar{\theta}\| < \beta$ for all $N > N_0$ implies the existence of a subsequence ξ_{N_t} such that $\limsup_{t\to\infty} H_{\bar{\theta}}(\xi_{N_t}) \geq H_{\bar{\theta}}^* - \epsilon$. From (25), for all $\delta > 0$ there exists N_3 such that for all $N > N_3$, $H_{\bar{\theta}}(\xi_{N+1}) > H_{\bar{\theta}}(\xi_N) - \delta$. Also, from the developments just above, there exists N_4 such that for all $N > N_4$, (23) implies $H_{\bar{\theta}}(\xi_{N+1}) > H_{\bar{\theta}}(\xi_N)$. Take any $N_t > \max(N_0, N_3, N_4)$ satisfying $H_{\bar{\theta}}(\xi_{N_t}) > H_{\bar{\theta}}^* - \epsilon - \delta$, we obtain $H_{\bar{\theta}}(\xi_N) > H_{\bar{\theta}}^* - \epsilon - 2\delta$, for all $N > N_t$. Since δ is arbitrary, $\liminf_{N\to\infty} H_{\bar{\theta}}(\xi_N) \geq H_{\bar{\theta}}^* - \epsilon$, which concludes the proof. ■

Proof of Lemma 3. We consider the case of LS estimation in the regression model (4). The proof is similar (but simpler) for ML estimation in the model (8) and we shall only indicate the adaptations that are required.

A first-order series development of the gradient of the LS criterion (5) around $\bar{\theta}$ gives

$$\nabla_\theta S_N(\hat{\theta}^N_{LS}) = \mathbf{0} = \nabla_\theta S_N(\bar{\theta}) + \nabla^2_\theta S_N(\tilde{\theta}^N)(\hat{\theta}^N_{LS} - \bar{\theta}) \,, \tag{26}$$

where $\nabla_\theta S_N(\bar{\theta}) = -2\sum_{k=1}^N \varepsilon_k \mathbf{f}_{\bar{\theta}}(x_k)$, $\nabla^2_\theta S_N(\theta)$ is the (Hessian) matrix of second-order derivatives of $S_N(\theta)$, given by

$$\nabla^2_\theta S_N(\theta) = 2N\,\mathbf{M}(\xi_N,\theta) - 2\sum_{k=1}^N [Y_k - \eta(x_k,\theta)]\,\frac{\partial \mathbf{f}_\theta(x_k)}{\partial \theta^\top}\,,$$

and $\tilde{\theta}^N = (1-\gamma_N)\bar{\theta} + \gamma_N \hat{\theta}^N_{LS}$, $\gamma_N \in (0,1)$, with $\tilde{\theta}^N$ measurable, see [18].

We first try to bound $\|\hat{\theta}^N_{LS} - \bar{\theta}\|$. We have $(1/N)\nabla^2_\theta S_N(\theta) = 2\mathbf{M}(\xi_N,\theta) + 2\mathbf{A}(\xi_N,\theta,\bar{\theta}) - 2\mathbf{B}_N(\xi_N,\theta)$, with

$$\mathbf{A}(\xi_N,\theta,\bar{\theta}) = \frac{1}{N}\sum_{k=1}^N [\eta(x_k,\theta) - \eta(x_k,\bar{\theta})]\,\frac{\partial \mathbf{f}_\theta(x_k)}{\partial \theta^\top} \quad \text{and} \quad \mathbf{B}_N(\xi_N,\theta) = \frac{1}{N}\sum_{k=1}^N \varepsilon_k\,\frac{\partial \mathbf{f}_\theta(x_k)}{\partial \theta^\top}\,.$$

From H_f-(ii), there exist $A_1 > 0$ and $A_2 > 0$ such that $\sup_{x\in\mathcal{X}} \sup_{\|\theta-\bar{\theta}\|\leq\delta} |\eta(x,\theta) - \eta(x,\bar{\theta})| < A_1\delta$ and $\sup_{x\in\mathcal{X}} \sup_{\|\theta-\bar{\theta}\|\leq\delta} \|\partial \mathbf{f}_\theta(x)/\partial\theta^\top\| < A_2$. This implies $\lim_{\delta\to 0} \sup_{\|\theta-\bar{\theta}\|\leq\delta} \|\mathbf{A}(\xi_N,\theta,\bar{\theta})\| = 0$. We also have

$$\sup_{\|\theta-\bar{\theta}\|\leq\delta} \|\mathbf{B}(\xi_N,\theta)\| < \sum_{x\in\mathcal{X}} \frac{\left|\sum_{k=1,\,x_k=x}^N \varepsilon_k\right|}{r_N(x)}\,A_2\,\frac{r_N(x)}{N} \tag{27}$$

where $r_N(x)$ denotes the number of times x appears in the sequence $x_1,\ldots,x_N$. Either $r_N(x)$ is bounded or $r_N(x)$ tends to infinity (but remains smaller than N), in any case $\lim_{N\to\infty} \sup_{\|\theta-\bar{\theta}\|\leq\delta} \left\|\mathbf{B}(\xi_N,\theta,\bar{\theta})\right\| = 0$ a.s. We have similarly $\mathbf{M}(\xi_N,\theta) - \mathbf{M}(\xi_N,\bar{\theta}) = \sum_{x\in\mathcal{X}} [r_N(x)/N]\,[\mathbf{f}_\theta(x)\mathbf{f}_\theta^\top(x) - \mathbf{f}_{\bar{\theta}}(x)\mathbf{f}_{\bar{\theta}}^\top(x)]$, and therefore $\lim_{\delta\to 0}\sup_{\|\theta-\bar{\theta}\|\leq\delta} \|\mathbf{M}(\xi_N,\theta) - \mathbf{M}(\xi_N,\bar{\theta})\| = 0$. Since $\hat{\theta}^N_{LS} \overset{\text{a.s.}}{\to} \bar{\theta}$, $N\to\infty$, we obtain that there exists a.s. N_0 such that for all $N > N_0$, $\Lambda_{\min}[\nabla^2_\theta S_N(\tilde{\theta}^N)/N] > \Lambda/\tau_N$. From the series development (26), it implies $\|\hat{\theta}^N_{LS} - \bar{\theta}\| < [\tau_N/(\Lambda\sqrt{N})]\,\|\nabla_\theta S_N(\bar{\theta})\|/\sqrt{N}$. We have $\nabla_\theta S_N(\bar{\theta})/\sqrt{N} = -2\sum_{x\in\mathcal{X}} \mathbf{v}_N(x)$, where $\mathbf{v}_N(x) = \zeta_N(x)\alpha_N(x)\mathbf{f}_{\bar{\theta}}(x)$ with $\zeta_N(x) = (\sum_{k=1,\,x_k=x}^N \varepsilon_k)/\sqrt{r_N(x)}$ and $\alpha_N(x) = \sqrt{r_N(x)/N} < 1$, therefore $\|\nabla_\theta S_N(\bar{\theta})\|/\sqrt{N} < 2\bar{f}\sum_{x\in\mathcal{X}} |\zeta_N(x)|$ with $\bar{f} = \max_{x\in\mathcal{X}} \|\mathbf{f}_{\bar{\theta}}(x)\|$ and $\zeta_N(x) = \mathcal{O}_p(1)$ (that is, $\zeta_N(x)$ is bounded in probability) for all x. We thus obtain $\|\hat{\theta}^N_{LS} - \bar{\theta}\| < (2\bar{f}\tau_N\omega_N)/(\Lambda\sqrt{N})$ for $N > N_0$ with $\omega_N = \sum_{x\in\mathcal{X}} |\zeta_N(x)| = \mathcal{O}_p(1)$.

Now, for N large enough we have $\|\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{A}(\xi_N,\theta,\bar{\theta})\| < (2\tau_N/\Lambda)\|\mathbf{A}(\xi_N,\theta,\bar{\theta})\|$ and therefore $\|\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{A}(\xi_N,\tilde{\theta}^N,\bar{\theta})\| < (4A_1A_2\bar{f}/\Lambda^2)\,(\tau_N^2\omega_N/\sqrt{N}) \xrightarrow{\text{p}} 0$, $N\to\infty$ (since $\tau_N^2/\sqrt{N}\to 0$ and $\omega_N = \mathcal{O}_p(1)$). Also, $\|\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{B}(\xi_N,\theta)\| < (2\tau_N/\Lambda)\|\mathbf{B}(\xi_N,\theta)\|$ for N large enough, which implies that $\sup_{\|\theta-\bar{\theta}\|\leq\delta}\left\|\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{B}(\xi_N,\theta)\right\| < (2A_2/\Lambda)\,(\tau_N\omega_N/\sqrt{N})$, see (27), and $\|\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{B}(\xi_N,\tilde{\theta}^N)\| \xrightarrow{\text{p}} 0$, $N\to\infty$. Moreover, $\|\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{M}(\xi_N,\theta)-\mathbf{I}\| = \|\mathbf{M}^{-1}(\xi_N,\bar{\theta})[\mathbf{M}(\xi_N,\theta)-\mathbf{M}(\xi_N,\bar{\theta})]\|$ and, for N large enough, $\|\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{M}(\xi_N,\theta)-\mathbf{I}\| < (2\tau_N/\Lambda)\,\sup_{x\in\mathcal{X}}\|\mathbf{f}_\theta(x)\mathbf{f}_\theta^\top(x)-\mathbf{f}_{\bar{\theta}}(x)\mathbf{f}_{\bar{\theta}}^\top(x)\|$. Since $\|\mathbf{f}_\theta(x)\mathbf{f}_\theta^\top(x)-\mathbf{f}_{\bar{\theta}}(x)\mathbf{f}_{\bar{\theta}}^\top(x)\|^2 = \|\mathbf{f}_\theta(x)-\mathbf{f}_{\bar{\theta}}(x)\|^4 + 2\{[\mathbf{f}_\theta(x)-\mathbf{f}_{\bar{\theta}}(x)]^\top\mathbf{f}_{\bar{\theta}}(x)\}^2 + 4\|\mathbf{f}_\theta(x)-\mathbf{f}_{\bar{\theta}}(x)\|^2[\mathbf{f}_\theta(x)-\mathbf{f}_{\bar{\theta}}(x)]^\top\mathbf{f}_{\bar{\theta}}(x)$ and $\sup_{x\in\mathcal{X}}\sup_{\|\theta-\bar{\theta}\|\leq\delta}\|\mathbf{f}_\theta(x)-\mathbf{f}_{\bar{\theta}}(x)\| < A_2\delta$, there exists $A_3>0$ such that for δ small enough we have $\sup_{x\in\mathcal{X}}\sup_{\|\theta-\bar{\theta}\|\leq\delta}\|\mathbf{f}_\theta(x)\mathbf{f}_\theta^\top(x)-\mathbf{f}_{\bar{\theta}}(x)\mathbf{f}_{\bar{\theta}}^\top(x)\| < A_3\delta$. We thus obtain $\|\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{M}(\xi_N,\tilde{\theta}^N)-\mathbf{I}\| < (4A_3\bar{f}/\Lambda^2)(\tau_N^2\omega_N/\sqrt{N}) \xrightarrow{\text{p}} 0$, $N\to\infty$ and finally $\|\mathbf{M}^{-1/2}(\xi_N,\bar{\theta})[\nabla_\theta^2 S_N(\tilde{\theta}^N)/N]\mathbf{M}^{-1/2}(\xi_N,\bar{\theta})-2\mathbf{I}\| \xrightarrow{\text{p}} 0$, so that (26) gives

$$[2+o_p(1)]\sqrt{N}\mathbf{M}^{1/2}(\xi_N,\bar{\theta})(\hat{\theta}_{LS}^N-\bar{\theta}) = -\mathbf{M}^{-1/2}(\xi_N,\bar{\theta})\nabla_\theta S_N(\bar{\theta})/\sqrt{N}\,,\ N\to\infty\,. \tag{28}$$

We show now that $\mathbf{M}^{-1/2}(\xi_N,\bar{\theta})\nabla_\theta S_N(\bar{\theta})/\sqrt{N}$ is asymptotically normal.

Denote by $\mathcal{F}_k$ the σ-field generated by $\{Y_1,\ldots,Y_k\}$. Notice that from the adaptive construction of the design, x_k is $\mathcal{F}_{k-1}$-measurable. Take any $\mathbf{u}\in\mathbb{R}^p$ with $\|\mathbf{u}\|=1$ and consider

$$R_N = -\frac{1}{2\sqrt{N}}\,\mathbf{u}^\top\mathbf{M}^{-1/2}(\xi_N,\bar{\theta})\nabla_\theta S_N(\bar{\theta}) = \sum_{k=1}^N \zeta_{Nk}$$

where $\zeta_{Nk} = \varepsilon_k z_{Nk}/\sqrt{N}$ with $z_{Nk} = \mathbf{u}^\top\mathbf{M}^{-1/2}(\xi_N,\bar{\theta})\mathbf{f}_{\bar{\theta}}(x_k)$. Using [29, Th. 1, p. 541], to prove that $R_N \xrightarrow{\text{d}} r \sim \mathcal{N}(0,s^2)$ it is enough to show that, for any $\gamma\in(0,1)$,

$$\sum_{k=1}^N \text{Prob}\{|\zeta_{Nk}|>\gamma|\mathcal{F}_{k-1}\} \xrightarrow{\text{p}} 0 \tag{29}$$

$$\sum_{k=1}^N \mathbb{E}\{\zeta_{Nk}\mathbb{I}(|\zeta_{Nk}|\leq 1)|\mathcal{F}_{k-1}\} \xrightarrow{\text{p}} 0 \tag{30}$$

$$\sum_{k=1}^N \text{Var}\{\zeta_{Nk}\mathbb{I}(|\zeta_{Nk}|\leq\gamma)|\mathcal{F}_{k-1}\} \xrightarrow{\text{p}} s^2 \tag{31}$$

as $N\to\infty$, with $\mathbb{I}(\cdot)$ the indicator function. First consider (29). Define $t_{Nk} = \text{Prob}\{|\zeta_{Nk}|>\gamma|\mathcal{F}_{k-1}\} = \int_{|\varepsilon|>\gamma\sqrt{N}/|z_{Nk}|} dF(\varepsilon)$ with $F(\cdot)$ the probability distribution of the errors in the model (4). Since $|z_{Nk}| \leq$

$\bar{f}\Lambda_{\min}^{-1/2}\mathbf{M}(\xi_N,\bar{\theta}) < \bar{f}\sqrt{2\tau_N}/\sqrt{\Lambda}$ for N large enough (a.s.), with $\bar{f} = \max_{x\in\mathcal{X}}\|\mathbf{f}_{\bar{\theta}}(x)\|$, we get $t_{Nk} < \int_{|\varepsilon|>\gamma\rho\sqrt{N/\tau_N}} dF(\varepsilon) < (\gamma\rho)^{-(2+\delta)}(\tau_N/N)^{1+\delta/2}\int_{|\varepsilon|>\gamma\rho\sqrt{N/\tau_N}}|\varepsilon|^{2+\delta}\,dF(\varepsilon)$, where $\rho = \sqrt{\Lambda}/(\bar{f}\sqrt{2})$. Since $\lim_{N\to\infty}\tau_N\,N^{-\delta/(2+\delta)} = 0$ and $\mathbb{E}\{|\varepsilon_1|^{2+\delta}\} < \infty$, $Nt_{Nk} \to 0$ as $N \to \infty$ and (29) is satisfied. Consider (30) and define $t'_{Nk} = \mathbb{E}\{\zeta_{Nk}\mathbb{I}(|\zeta_{Nk}| \leq 1)|\mathcal{F}_{k-1}\} = (z_{Nk}/\sqrt{N})\int_{|\varepsilon|\leq\sqrt{N}/|z_{Nk}|}\varepsilon\,dF(\varepsilon)$, so that $|t'_{Nk}| = (|z_{Nk}|/\sqrt{N})\left|\int_{|\varepsilon|>\sqrt{N}/|z_{Nk}|}\varepsilon\,dF(\varepsilon)\right|$ (since $\mathbb{E}\{\varepsilon_1\} = 0$). Therefore, for N large enough, $|t'_{Nk}| \leq (1/\rho)\sqrt{\tau_N/N}\int_{|\varepsilon|>\rho\sqrt{N/\tau_N}}|\varepsilon|\,dF(\varepsilon) < \rho^{-(2+\delta)}(\tau_N/N)^{1+\delta/2}\int_{|\varepsilon|>\rho\sqrt{N/\tau_N}}|\varepsilon|^{2+\delta}\,dF(\varepsilon)$, which implies that $N|t'_{Nk}| \to 0$ as $N \to \infty$ and (30) is satisfied. We proceed in a similar way for (31) and define $t''_{Nk} = \mathrm{Var}\{\zeta_{Nk}\mathbb{I}(|\zeta_{Nk}| \leq \gamma)|\mathcal{F}_{k-1}\} = (z_{Nk}^2/N)\left[\int_{|\varepsilon|\leq\gamma\sqrt{N}/|z_{Nk}|}\varepsilon^2\,dF(\varepsilon) - \left(\int_{|\varepsilon|\leq\gamma\sqrt{N}/|z_{Nk}|}\varepsilon\,dF(\varepsilon)\right)^2\right]$. Since $\mathbb{E}\{\varepsilon_1^2\} = 1$, we obtain $\sum_k t''_{Nk} = (1/N)\sum_k z_{Nk}^2 + Q_N$ where, for N large enough, Q_N satisfies $|Q_N| < \tau_N/(N\rho^2)\sum_k\left[\int_{|\varepsilon|>\gamma\sqrt{N}/|z_{Nk}|}\varepsilon^2\,dF(\varepsilon) + \left(\int_{|\varepsilon|>\gamma\sqrt{N}/|z_{Nk}|}|\varepsilon|\,dF(\varepsilon)\right)^2\right]$ (using $\mathbb{E}\{\varepsilon_1\} = 0$). Therefore, for N large enough, $|Q_N| < (\tau_N/\rho^2)\left[\int_{|\varepsilon|>\gamma\rho\sqrt{N/\tau_N}}\varepsilon^2\,dF(\varepsilon) + \left(\int_{|\varepsilon|>\gamma\rho\sqrt{N/\tau_N}}|\varepsilon|\,dF(\varepsilon)\right)^2\right] < \gamma^{-\delta}\rho^{-(2+\delta)})(\tau_N^{1+\delta/2}/N^{\delta/2})\left[\int_{|\varepsilon|>\gamma\rho\sqrt{N/\tau_N}}|\varepsilon|^{2+\delta}\,dF(\varepsilon) + \left(\int_{|\varepsilon|>\gamma\rho\sqrt{N/\tau_N}}|\varepsilon|^{1+\delta/2}\,dF(\varepsilon)\right)^2\right] \to 0$ as $N \to \infty$. Moreover, $(1/N)\sum_k z_{Nk}^2 = \mathbf{u}^\top\mathbf{u} = 1$ and we have thus proved that $R_N \xrightarrow{\mathrm{d}} r \sim \mathcal{N}(0,1)$ as $N \to \infty$. Since this is true for any $\mathbf{u}$, we have $\mathbf{M}^{-1/2}(\xi_N,\bar{\theta})\nabla_\theta S_N(\bar{\theta})/\sqrt{N} \xrightarrow{\mathrm{d}} \mathbf{v} \sim \mathcal{N}(\mathbf{0},4\mathbf{I})$, and therefore from (28), $\sqrt{N}\mathbf{M}^{1/2}(\xi_N,\bar{\theta})(\hat{\theta}^N_{LS} - \bar{\theta}) \xrightarrow{\mathrm{d}} \mathbf{w} \sim \mathcal{N}(\mathbf{0},\mathbf{I})\,,\ N \to \infty$. Since $\|\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{M}(\xi_N,\hat{\theta}^N_{LS}) - \mathbf{I}\| \xrightarrow{\mathrm{p}} 0$ as $N \to \infty$, we obtain the announced result (14).

Consider now ML estimation in the model (8). We substitute $L_N(\theta)$ for $S_N(\theta)$ is the series development (26) (with the difference that $L_N(\theta)$ is now *maximum* at $\hat{\theta}^N_{ML}$). Denoting $\pi_k = \pi(x_k,\theta)$ and $\bar{\pi}_k = \pi(x_k,\bar{\theta})$, we have $L_N(\theta) = \sum_k Y_k\log\pi_k + (1-Y_k)\log(1-\pi_k)$, $\nabla_\theta L_N(\theta) = \sum_k(Y_k-\pi_k)/\sqrt{\pi_k(1-\pi_k)}\mathbf{f}_\theta(x_k)$ and $\nabla^2_\theta L_N(\theta)/N = -\mathbf{M}(\xi_N,\theta) + \mathbf{A}_N(\xi_N,\theta,\bar{\theta}) + \mathbf{B}_N(\xi_N,\theta,\bar{\theta})$, where $\mathbf{f}_\theta(x)$ is given by (9),

$$\mathbf{A}_N(\xi_N,\theta,\bar{\theta}) = \frac{1}{N}\sum_{k=1}^N \frac{\bar{\pi}_k - \pi_k}{\sqrt{\pi_k(1-\pi_k)}}\,\mathbf{Q}_k(x_k,\theta) \ \text{ and } \mathbf{B}_N(\xi_N,\theta,\bar{\theta}) = \frac{1}{N}\sum_{k=1}^N Z_k\,\frac{\sqrt{\bar{\pi}_k(1-\bar{\pi}_k)}}{\sqrt{\pi_k(1-\pi_k)}}\,\mathbf{Q}_k(x_k,\theta)\,,$$

with

$$\mathbf{Q}_k(x_k,\theta) = \frac{1}{\sqrt{\pi_k(1-\pi_k)}}\left[\frac{\partial^2\pi_k}{\partial\theta\partial\theta^\top} + (2\pi_k-1)\mathbf{f}_\theta(x_k)\mathbf{f}_\theta^\top(x_k)\right] \ \text{ and } \ Z_k = \frac{Y_k - \bar{\pi}_k}{\sqrt{\bar{\pi}_k(1-\bar{\pi}_k)}}\,.$$

Notice that $\mathbb{E}\{Z_k\} = 0$ and $\mathbb{E}\{Z_k^2\} = 1$. Similarly to the case of LS estimation, we obtain that exists a.s. N_0 such that, for all $N > N_0$, $\Lambda_{\min}[-\nabla^2_\theta L_N(\tilde{\theta}^N)/N] > 1/(2\tau_N)$, where $\tilde{\theta}^N$ is now a point between $\bar{\theta}$ and $\hat{\theta}^N_{ML}$. This implies that $\|\hat{\theta}^N_{ML} - \bar{\theta}\| < c\tau_N\omega_N/\sqrt{N}$ for some constant c and $\omega_N = \mathcal{O}_p(1)$. We consider then $\|\mathbf{M}^{-1}(\xi_N,\bar{\theta})\nabla^2_\theta L_N(\theta)/N + \mathbf{I}\|$ and obtain, since $\lim_{N\to\infty}\tau_N/N^{1/4} = 0$, $\|\mathbf{M}^{-1/2}(\xi_N,\bar{\theta})[\nabla^2_\theta L_N(\tilde{\theta}^N)/N]\mathbf{M}^{-1/2}(\xi_N,\bar{\theta})+\mathbf{I}\| \stackrel{\text{p}}{\to} 0$, $N\to\infty$. This gives $[1+o_p(1)]\sqrt{N}\mathbf{M}^{1/2}(\xi_N,\bar{\theta})(\hat{\theta}^N_{ML}-\bar{\theta}) = \mathbf{M}^{-1/2}(\xi_N,\bar{\theta})\nabla_\theta L_N(\bar{\theta})/\sqrt{N}$, $N\to\infty$. We consider finally $R_N = (1/\sqrt{N})\mathbf{u}^\top\mathbf{M}^{-1/2}(\xi_N,\bar{\theta})\nabla_\theta L_N(\bar{\theta})$ for some vector $\mathbf{u}$ with norm 1, and write $R_N = \sum_k \zeta_{N,k}$ with now $\zeta_{N,k} = Z_k\mathbf{u}^\top\mathbf{M}^{-1/2}(\xi_N,\bar{\theta})\mathbf{f}_{\bar{\theta}}(x_k)/\sqrt{N}$. The properties (29, 30, 31) directly follow from the fact that Z_k is bounded, $|Z_k| < \max_{x\in\mathcal{X}}\max\{([1-\pi(x,\bar{\theta})]/\pi(x,\bar{\theta}))^{1/2},(\pi(x,\bar{\theta})/[1-\pi(x,\bar{\theta})])^{1/2}\}$ for all k. The rest of the proof is similar to the case of LS estimation. ■

Proof of Lemma 4. First note that $\liminf_{N\to\infty} r_{N,1:K}/N > 1/K$ since $\mathcal{X}$ is finite, so that $q^* \geq 1$. Suppose that $p\geq 2$ and $q^* < p$, we show that we arrive at a contradiction. The proof follows the same lines as for Lemma 1.

The property (18) is replaced by

$$\text{for all } \; i\leq q^*\,,\; N>N_1 \text{ and } \theta\in\Theta\,,\; \mathbf{f}_\theta^\top(x^{(i_N)})\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{f}_\theta(x^{(i_N)}) - \lambda_N\,\phi(x^{(i_N)},\theta) \leq \frac{\lambda_N}{\alpha}\,. \tag{32}$$

Define ρ_N as $\rho_N = \lambda_N\beta_N$, with $\beta_N = r_{N,(q^*+1):K}/N$ as in Lemma 1. We show that $\liminf_{N\to\infty}\rho_N \geq \underline{\rho}$ for some $\underline{\rho} > 0$, which contradicts the definition of q^*.

The property (22) becomes

$$\text{for all } \; \theta\in\Theta\,,\; \max_{i=q^*+1,\ldots,K}\left\{\mathbf{f}_\theta^\top(x^{(i_N)})\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{f}_\theta(x^{(i_N)}) - \lambda_N\,\phi(x^{(i_N)},\theta)\right\} > \frac{1}{\beta_N}\,\frac{\gamma}{LK(K-q^*)} - \lambda_N\bar{\phi}\,.$$

Together with (32), it gives $N > N_1$ and $\rho_N < \rho^* = \gamma\alpha/[LK(K-q^*)(1+\alpha\bar{\phi})] \Rightarrow x_{N+1}\notin\mathcal{X}_N(q^*)$. Define

$$\rho^*_N = \lambda_N\,\frac{\sum_{i=q^*+1}^K r_{N,i:K}}{(K-q^*)N}\,,$$

so that $\rho_N \geq \rho^*_N \geq \rho_N/(K-q^*)$. Also, when $N>N_1$, $\sum_{i=1}^{q^*} r_{N,i:K} > q^*\alpha N/\lambda_N$, so that $\sum_{i=q^*+1}^K r_{N,i:K} < N(1-q^*\alpha/\lambda_N)$ and $\rho^*_N < \lambda_N(1-q^*\alpha/\lambda_N)/(K-q^*)$. By construction, $\rho_N<\rho^*$ and $N>N_1$ imply

$$\begin{aligned}\rho_{N+1} \;\;\geq\;\; & \rho^*_{N+1} = \lambda_{N+1}\,\frac{N\rho^*_N(K-q^*)/\lambda_N + 1}{(K-q^*)(N+1)} = \frac{\lambda_{N+1}}{\lambda_N}\,\rho^*_N + \frac{\lambda_{N+1}}{N+1}\left(\frac{1}{K-q^*} - \frac{\rho^*_N}{\lambda_N}\right) \\ & > \frac{\lambda_{N+1}}{\lambda_N}\,\left[\rho^*_N + \frac{1}{N+1}\,\frac{q^*\alpha}{K-q^*}\right] \geq \rho^*_N + \frac{1}{N+1}\,\frac{q^*\alpha}{K-q^*} \geq \frac{\rho_N}{K-q^*} + \frac{1}{N+1}\,\frac{q^*\alpha}{K-q^*}\,.\end{aligned}$$

The end of the proof is strictly identical to that of Lemma 1. By induction, the lower bound on ρ_{N+k} increases with k,

$$\rho_{N+k} > \frac{\rho_N}{K-q^*} + \frac{q^*\alpha}{K-q^*} \sum_{i=1}^{k} \frac{1}{N+i}\,,$$

until ρ_{N+k} becomes larger that ρ^* and $\liminf_{N\to\infty} \rho_N > \underline{\rho} = \rho^*/(K-q^*)$, which shows that $q^* \geq p$ and concludes the proof. ■

Proof of Theorem 4. We have already seen that the conditions of Lemma 4 with H$_{\mathcal{X}}$-(iii) (respectively, H$_{\mathcal{X}}$-(iii')) imply $\hat{\theta}^N_{LS} \overset{\text{a.s.}}{\to} \bar{\theta}$ (respectively, $\hat{\theta}^N_{ML} \overset{\text{a.s.}}{\to} \bar{\theta}$) as $N \to \infty$. Therefore, what we first need to show is that for all $\delta > 0$, there exist some $N_0 > 0$ and $\beta > 0$ such that $\|\hat{\theta}^N - \bar{\theta}\| < \beta$ for all $N > N_0$ implies $\limsup_{N\to\infty} \Phi(\xi_N, \bar{\theta}) < \phi^*_{\bar{\theta}} + \delta$. This will prove (15).

As in the proof of Theorem 3, H$_f$-(i), H$_\phi$-(ii) and H$_{\mathcal{X}}$-(i) imply: $\forall \epsilon_0$, $\exists \beta_1, \beta_2$ such that $\|\theta - \bar{\theta}\| < \beta_1 \Rightarrow \max_{x\in\mathcal{X}} \|\mathbf{f}_\theta(x) - \mathbf{f}_{\bar{\theta}}(x)\| < \epsilon_0$ and $\|\theta - \bar{\theta}\| < \beta_2 \Rightarrow \max_{x\in\mathcal{X}} |\phi(x,\theta) - \phi(x,\bar{\theta})| < \epsilon_0$. Also, from Lemma 4, there exits N_1 and $\alpha > 0$ such that for all $N > N_1$, $r_{N,j:K} > \alpha N/\lambda_N$, $j = 1, \ldots, q^*$, with $q^* > p$, and thus from H$_{\mathcal{X}}$-(iv), $\Lambda_{\min}[\mathbf{M}(\xi_N, \bar{\theta})] > \alpha\bar{\gamma}/\lambda_N$. Direct calculations give

$$\max_{x\in\mathcal{X}} \max_{\|\theta-\bar{\theta}\|<\beta_1} |\mathbf{f}_\theta^\top(x)\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{f}_\theta(x) - \mathbf{f}_{\bar{\theta}}^\top(x)\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{f}_{\bar{\theta}}(x)| < B\lambda_N\epsilon_0$$

for ϵ_0 small enough and $N > N_1$, with B depending on $\alpha, \bar{\gamma}$ and $\bar{f} = \max_{x\in\mathcal{X}} \|\mathbf{f}_{\bar{\theta}}(x)\|$.

Therefore, for all ϵ_0 small enough, we can take $\beta = \min\{\beta_1, \beta_2\}$ and obtain, for all $N > N_1$ and $\|\hat{\theta}^N - \bar{\theta}\| < \beta$ in (3),

$$\begin{aligned}
&\mathbf{f}_{\bar{\theta}}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\theta)\mathbf{f}_{\bar{\theta}}(x_{N+1}) - \lambda_N\phi(x_{N+1},\bar{\theta}) \\
&\qquad > \mathbf{f}_{\hat{\theta}^N}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\hat{\theta}^N)\mathbf{f}_{\hat{\theta}^N}(x_{N+1}) - \lambda_N\phi(x_{N+1},\hat{\theta}^N) - \epsilon_0(B+1)\lambda_N \\
&\qquad = \max_{x\in\mathcal{X}} \left[\mathbf{f}_{\hat{\theta}^N}^\top(x)\mathbf{M}^{-1}(\xi_N,\hat{\theta}^N)\mathbf{f}_{\hat{\theta}^N}(x) - \lambda_N\phi(x,\hat{\theta}^N)\right] - \epsilon_0(B+1)\lambda_N \\
&\qquad > \max_{x\in\mathcal{X}} \left[\mathbf{f}_{\bar{\theta}}^\top(x)\mathbf{M}^{-1}(\xi_N,\bar{\theta})\mathbf{f}_{\bar{\theta}}(x) - \lambda_N\phi(x,\bar{\theta})\right] - 2\epsilon_0(B+1)\lambda_N\,. \qquad (33)
\end{aligned}$$

Suppose now that exists $\delta > 0$ such that $\liminf_{N\to\infty} \Phi(\xi_N, \bar{\theta}) > \phi^*_{\bar{\theta}} + \delta$, that is, there exists N_2 such that

$$\Phi(\xi_N, \bar{\theta}) > \phi^*_{\bar{\theta}} + \delta \qquad (34)$$

for all $N > N_2$. This implies

$$\mathbf{f}_{\bar\theta}^\top(x^*)\mathbf{M}^{-1}(\xi_N,\bar\theta)\mathbf{f}_{\bar\theta}(x^*) - \lambda_N\phi(x^*,\bar\theta) > -\lambda_N\phi(x^*,\bar\theta) > \lambda_N\left[\delta - \Phi(\xi_N,\bar\theta)\right] \,,\ N > N_2\,,$$

with x^* such that $\phi(x^*,\bar\theta) = \min_{x\in\mathcal{X}}\phi(x,\bar\theta)$, and therefore,

$$\max_{x\in\mathcal{X}}\left[\mathbf{f}_{\bar\theta}^\top(x)\mathbf{M}^{-1}(\xi_N,\bar\theta)\mathbf{f}_{\bar\theta}(x) - \lambda_N\phi(x,\bar\theta)\right] > \lambda_N\left[\delta - \Phi(\xi_N,\bar\theta)\right] \,,\ N > N_2\,.$$

For $\epsilon = \delta/2$, take $\epsilon_0 = \epsilon/[2(B+1)]$ and β as above and define

$$G_{\bar\theta}(\xi_N,\lambda_N) = \frac{1}{\lambda_N}\,\log\det\mathbf{M}(\xi_N,\bar\theta) + [C - \Phi(\xi_N,\bar\theta)]\,. \tag{35}$$

From (25, 33), we obtain

$$\begin{aligned} G_{\bar\theta}(\xi_{N+1},\lambda_N) - G_{\bar\theta}(\xi_N,\lambda_N) \quad > \quad & \frac{1}{\lambda_N}\log\left[1 + \frac{\mathbf{f}_{\bar\theta}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\bar\theta)\mathbf{f}_{\bar\theta}(x_{N+1})}{N}\right] - \frac{p}{\lambda_N}\log\left(1+\frac{1}{N}\right) \\ & + \frac{\delta}{2(N+1)} - \frac{1}{N+1}\,\frac{\mathbf{f}_{\bar\theta}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\bar\theta)\mathbf{f}_{\bar\theta}(x_{N+1})}{\lambda_N}\,, \end{aligned}$$

for $\|\hat\theta^N - \bar\theta\| < \beta$ and $N > \max\{N_1,N_2\}$. Since $\Lambda_{\min}[\mathbf{M}(\xi_N,\bar\theta)] > \alpha\bar\gamma/\lambda_N$ for $N > N_1$, we also have $\mathbf{f}_{\bar\theta}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\bar\theta)\mathbf{f}_{\bar\theta}(x_{N+1}) < \bar f^2\lambda_N/(\lambda\bar\gamma)$, with $\bar f = \max_{x\in\mathcal{X}}\|\mathbf{f}_{\bar\theta}(x)\|$. Since $\lambda_N \to \infty$ from H$_\lambda$-(ii) and $\lambda_N/N \to 0$ from H$_\lambda$-(iii) when $N\to\infty$, for any constant $D < 1$ there exists N_3 such that for all $N > N_3$,

$$\log\left[1 + \frac{\mathbf{f}_{\bar\theta}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\bar\theta)\mathbf{f}_{\bar\theta}(x_{N+1})}{N}\right] > D\,\frac{\mathbf{f}_{\bar\theta}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\bar\theta)\mathbf{f}_{\bar\theta}(x_{N+1})}{N+1}$$

and

$$\frac{p}{\lambda_N}\log\left(1+\frac{1}{N}\right) < \frac{\delta}{4(N+1)}\,.$$

This gives

$$\begin{aligned} G_{\bar\theta}(\xi_{N+1},\lambda_N) - G_{\bar\theta}(\xi_N,\lambda_N) \quad > \quad & \frac{\delta}{4(N+1)} + (D-1)\,\frac{\mathbf{f}_{\bar\theta}^\top(x_{N+1})\mathbf{M}^{-1}(\xi_N,\bar\theta)\mathbf{f}_{\bar\theta}(x_{N+1})}{(N+1)\lambda_N} \\ & > \frac{1}{N+1}\left[\frac{\delta}{4} + (D-1)\,\frac{\bar f^2}{\alpha\bar\gamma}\right] > \frac{\delta}{8(N+1)} \end{aligned}$$

for $N > \max\{N_1,N_2,N_3\}$ and $\|\hat\theta^N - \bar\theta\| < \beta$ when choosing $D > 1 - \delta\alpha\bar\gamma/(8\bar f^2)$. Consider now

$$G_{\bar\theta}(\xi_{N+1},\lambda_{N+1}) - G_{\bar\theta}(\xi_N,\lambda_N) = \left[\frac{1}{\lambda_{N+1}} - \frac{1}{\lambda_N}\right]\log\det\mathbf{M}(\xi_{N+1},\bar\theta) + G_{\bar\theta}(\xi_N,\lambda_{N+1}) - G_{\bar\theta}(\xi_N,\lambda_N)\,.$$

For N large enough and $\|\hat{\theta}^N - \bar{\theta}\| < \beta$, it satisfies

$$G_{\bar{\theta}}(\xi_{N+1}, \lambda_{N+1}) - G_{\bar{\theta}}(\xi_N, \lambda_N) > \left[\frac{1}{\lambda_{N+1}} - \frac{1}{\lambda_N}\right] \log\det \mathbf{M}(\xi_{N+1}, \bar{\theta}) + \frac{\delta}{8(N+1)}\,.$$

Denote $L^* = \max_{\xi\in\Xi} \log\det \mathbf{M}(\xi, \bar{\theta})$. Since λ_N is non-decreasing from H$_\lambda$-(ii) and λ_N/N is non-increasing from H$_\lambda$-(iii), we get

$$G_{\bar{\theta}}(\xi_{N+1}, \lambda_{N+1}) - G_{\bar{\theta}}(\xi_N, \lambda_N) > \frac{\delta}{8(N+1)} - \left[\frac{1}{\lambda_N} - \frac{1}{\lambda_{N+1}}\right] L^* \geq \frac{\delta}{8(N+1)} - \frac{L^*}{\lambda_N(N+1)} > \frac{\delta}{16(N+1)}$$

for N large enough, in contradiction with the fact that $G_{\bar{\theta}}(\xi, \lambda)$ is bounded for $\lambda > 1$, see (35).

Therefore, for all $\delta > 0$ we can find a β such that $\|\hat{\theta}^N - \bar{\theta}\| < \beta$ for all $N > N_0$ implies that there exists a subsequence ξ_{N_t} such that $\liminf_{t\to\infty} \Phi(\xi_{N_t}, \bar{\theta}) < \phi^*_{\bar{\theta}} + \delta$. First note that H$_\phi$-(i) implies $\Phi(\xi_{N+1}, \bar{\theta}) - \Phi(\xi_N, \bar{\theta}) = 1/(N+1)\,[\phi(x_{N+1}, \bar{\theta}) - \Phi(\xi_N, \bar{\theta})] < \bar{\phi}/(N+1)$ so that there exists N_4 such that $\Phi(\xi_{N+1}, \bar{\theta}) < \Phi(\xi_N, \bar{\theta}) + \delta/2$ for all $N > N_4$. Also, from the developments above, there exists N_5 such that for $N > N_5$, $\|\hat{\theta}^N - \bar{\theta}\| < \beta$ and $\Phi(\xi_N, \bar{\theta}) > \phi^*_{\bar{\theta}} + \delta$ imply $G_{\bar{\theta}}(\xi_{N+1}, \lambda_{N+1}) > G_{\bar{\theta}}(\xi_N, \lambda_N)$. Moreover, since $\Lambda_{\min}[\mathbf{M}(\xi_N, \bar{\theta})] > \alpha\bar{\gamma}/\lambda_N$ for all $N > N_1$, we have $p\,\log(\alpha\bar{\gamma}) - p\,\log\lambda_N < \log\det \mathbf{M}(\xi_N, \bar{\theta}) < L^*$ and thus

$$\frac{|\log\det \mathbf{M}(\xi_N, \bar{\theta})|}{\lambda_N} < \frac{\delta}{4} \tag{36}$$

for N larger than some N_6. Take any $N_t > \max(N_0, N_4, N_5, N_6)$ and such that $\Phi(\xi_{N_t}, \bar{\theta}) \leq \phi^*_{\bar{\theta}} + \delta$. We show that $\Phi(\xi_N, \bar{\theta}) < \phi^*_{\bar{\theta}} + 2\delta$ for all $N > N_t$ until the next N' such that $\Phi(\xi_{N'}, \bar{\theta}) \leq \phi^*_{\bar{\theta}} + \delta$. We have $\Phi(\xi_{N_t+1}, \bar{\theta}) < \phi^*_{\bar{\theta}} + (3/2)\delta$. If $\Phi(\xi_{N_t+1}, \bar{\theta}) \leq \phi^*_{\bar{\theta}} + \delta$ we take $N' = N_t + 1$. Suppose that $\Phi(\xi_{N_t+1}, \bar{\theta}) > \phi^*_{\bar{\theta}} + \delta$. Then, $G_{\bar{\theta}}(\xi_{N_t+2}, \lambda_{N_t+2}) > G_{\bar{\theta}}(\xi_{N_t+1}, \lambda_{N_t+1})$ and, from the definition of $G_{\bar{\theta}}(\xi, \lambda)$ and (36), $\Phi(\xi_{N_t+2}, \bar{\theta}) < \Phi(\xi_{N_t+1}, \bar{\theta}) + \delta/2 < \phi^*_{\bar{\theta}} + 2\delta$. If $\Phi(\xi_{N_t+2}, \bar{\theta}) \leq \phi^*_{\bar{\theta}} + \delta$ we take $N' = N_t + 2$. Otherwise, we have $G_{\bar{\theta}}(\xi_{N_t+3}, \lambda_{N_t+3}) > G_{\bar{\theta}}(\xi_{N_t+2}, \lambda_{N_t+2}) > G_{\bar{\theta}}(\xi_{N_t+1}, \lambda_{N_t+1})$ and thus $\Phi(\xi_{N_t+3}, \bar{\theta}) < \Phi(\xi_{N_t+1}, \bar{\theta}) + \delta/2 < \phi^*_{\bar{\theta}} + 2\delta$. By induction, $\Phi(\xi_N, \bar{\theta}) < \phi^*_{\bar{\theta}} + 2\delta$ for all $N > N_t$. This concludes the proof of (15).

Define $B_N(\beta) = \{x_i : \|x_i - x^*\| > \beta\,,\ i = 1, \ldots, N\}$ and denote by $b_N(\beta)$ the number of elements of $B_N(\beta)$. Assume that there exists $\beta > 0$ such that $\limsup_{N\to\infty} b_N(\beta)/N > \gamma > 0$. From H$_\phi$-(iii), this implies that there exists $\epsilon > 0$ such that $\limsup_{N\to\infty} \Phi(\xi_N, \bar{\theta}) - \phi^*_{\bar{\theta}} > \gamma\epsilon$, which contradicts (15). Therefore, for all $\beta > 0$, $\lim_{N\to\infty} b_N(\beta)/N = 0$ which gives (16). ■

Acknowledgements The author wishes to thank V. Fedorov for his comments and N. Flournoy for her careful reading of the paper, her detailed comments and advice.

References

[1] P. Chaudhuri and P.A. Mykland. Nonlinear experiments: optimal design and inference based likelihood. *Journal of the American Statistical Association*, 88(422):538–546, 1993.

[2] P. Chaudhuri and P.A. Mykland. On efficiently designing of nonlinear experiments. *Statistica Sinica*, 5:421–440, 1995.

[3] D. Cook and V.V. Fedorov. Constrained optimization of experimental design (invited discussion paper). *Statistics*, 26:129–178, 1995.

[4] D. Cook and W.K. Wong. On the equivalence between constrained and compound optimal designs. *Journal of the American Statistical Association*, 89(426):687–692, 1994.

[5] V. Dragalin and V. Fedorov. Adaptive designs for dose-finding based on efficacy-toxicity response. *Journal of Statistical Planning and Inference*, 136:1800–1823, 2006.

[6] V. Dragalin, V. Fedorov, and Y. Wu. Adaptive designs for selecting drug combinations based on efficacy-toxicity response. *Journal of Statistical Planning and Inference*, 138:352–373, 2008.

[7] S. Durham, N. Flournoy, and W. Li. A sequential design for maximizing the probability of a favorable response. *Can. Journal Statist.*, 26:479–495, 1998.

[8] S.D. Durham and N. Flournoy. Random walks for quantile estimation. In J.O. Berger and S.S. Gupta, editors, *Statistical Decision Theory and Related Topics V*, pages 467–476. Springer, New York, 1994.

[9] V.V. Fedorov and P. Hackl. *Model-Oriented Design of Experiments.* Springer, Berlin, 1997.

[10] I. Ford and S.D. Silvey. A sequentially constructed design for estimating a nonlinear parametric function. *Biometrika*, 67(2):381–388, 1980.

[11] R. Gautier and L. Pronzato. Sequential design and active control. In N. Flournoy, W.F. Rosenberger, and W.K. Wong, editors, *New Developments and Applications in Experimental Design, Lecture Notes — Monograph Series, vol. 34*, pages 138–151. IMS, Hayward, 1998.

[12] R. Gautier and L. Pronzato. Adaptive control for sequential design. *Discussiones Mathematicae, Probability & Statistics*, 20(1):97–114, 2000.

[13] A. Giovagnoli. Asymptotic properties of biased coin designs for treatment allocation. In A. Di Bucchianico, H. Läuter, and H.P. Wynn, editors, *mODa'7 – Advances in Model–Oriented Design and Analysis, Proceedings of the 7th Int. Workshop, Heeze (Netherlands)*, pages 81–88, Heidelberg, June 2004. Physica Verlag.

[14] A. Giovagnoli and N. Pintacuda. Properties of frequency distributions induced by general 'up-and-down' methods for estimating quantiles. *Journal of Statistical Planning and Inference*, 74:51–63, 1998.

[15] J. Hardwick and Q. Stout. Optimizing a unimodal response function for binary variables. In A. Atkinson, B. Bogacka, and A. Zhigljavsky, editors, *Optimum Design 2000*, chapter 18, pages 195–210. Kluwer, Dordrecht, 2001.

[16] I. Hu. On sequential designs in nonlinear problems. *Biometrika*, 85(2):496–503, 1998.

[17] A. Ivanova. A new dose-finding design for bivariate outcomes. *Biometrics*, 59:1001–1007, 2003.

[18] R.I. Jennrich. Asymptotic properties of nonlinear least squares estimation. *Annals of Math. Stat.*, 40:633–643, 1969.

[19] E.E. Kpamegan and N. Flournoy. An optimizing up-and-down design. In A. Atkinson, B. Bogacka, and A. Zhigljavsky, editors, *Optimum Design 2000*, chapter 19, pages 211–224. Kluwer, Dordrecht, 2001.

[20] T.L. Lai. Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *Annals of Statistics*, 22(4):1917–1930, 1994.

[21] T.L. Lai and C.Z. Wei. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics*, 10(1):154–166, 1982.

[22] J. Mikulecká. On a hybrid experimental design. *Kybernetika*, 19(1):1–14, 1983.

[23] W.G. Müller and B.M. Pötscher. Batch sequential design for a nonlinear estimation problem. In V.V. Fedorov, W.G. Müller, and I.N. Vuchkov, editors, *Model-Oriented Data Analysis II, Proceedings 2nd IIASA Workshop, St Kyrik (Bulgaria), May 1990*, pages 77–87. Physica Verlag, Heidelberg, 1992.

[24] L. Pronzato. Adaptive optimisation and D-optimum experimental design. *Annals of Statistics*, 28(6):1743–1761, 2000.

[25] L. Pronzato. One-step ahead adaptive D-optimal design on a finite design space is asymptotically optimal. Technical Report I3S/RR-2008-14-FR, 22 pages, Laboratoire I3S, CNRS–Université de Nice-Sophia Antipolis, 06903 Sophia Antipolis, France, 2008. `http://www.i3s.unice.fr/~mh/RR/rapports.html`.

[26] L. Pronzato. Penalized optimal designs for dose-finding. Technical Report I3S/RR-2008-18-FR, Laboratoire I3S, CNRS–Université de Nice-Sophia Antipolis, 06903 Sophia Antipolis, France, 2008. `http://www.i3s.unice.fr/~mh/RR/rapports.html`.

[27] L. Pronzato and E. Thierry. Sequential experimental design and response optimisation. *Statistical Methods and Applications*, 11(3):277–292, 2003.

[28] W.F. Rosenberger, N. Flournoy, and S.D. Durham. Asymptotic normality of maximum likelihood

estimators for multiparameter response-driven designs. *Journal of Statistical Planning and Inference*, 60:69–76, 1997.

[29] A.N. Shiryaev. *Probability.* Springer, Berlin, 1996.

[30] J. Volaufová, S.M. Redmann, Jr., and L. LaMotte. Is logistic regression the best for binomial response? In *Tatra Mountains Mathematical Publications*, volume 26, pages 237–246, 2003.

[31] C.F.J. Wu. Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, 9(3):501–513, 1981.

[32] C.F.J. Wu. Asymptotic inference from sequential design in a nonlinear situation. *Biometrika*, 72(3):553–558, 1985.

[33] C.F.J. Wu and H.P. Wynn. The convergence of general step–length algorithms for regular optimum design criteria. *Annals of Statistics*, 6(6):1273–1285, 1978.

[34] H.P. Wynn. The sequential generation of D-optimum experimental designs. *Annals of Math. Stat.*, 41:1655–1664, 1970.

[35] Y. Yi and X. Wang. Asymptotically efficient estimation in response adaptive trials. *Journal of Statistical Planning and Inference*, 138:2899–2905, 2008.