The Annals of Statistics
1991, Vol. 19, No. 4, 1978–2002

ONE-ARMED BANDIT PROBLEMS WITH COVARIATES[1]

BY JYOTIRMOY SARKAR

University of Michigan

As does Woodroofe, we consider a Bayesian sequential allocation between two treatments that incorporates a covariate. The goal is to maximize the total discounted expected reward from an infinite population of patients. Although our model is more general than Woodroofe's, we are able to duplicate his main result: The myopic rule is asymptotically optimal.

1. Introduction.

1.1. *Statement of the problem.* Consider a population of patients who arrive sequentially for treatment of a disease. Suppose each patient may be treated with either a standard treatment whose statistical characteristics are known, or a new treatment whose characteristics are unknown. Also suppose that before deciding to assign a given patient to a treatment, we observe a covariate X, such as age, severity of disease or general physical status, which is specific to the patient.

Let Y^0 and Y^1 denote the potential rewards from the standard and the new treatment, respectively. Let $\delta = 0$ and $\delta = 1$ denote the choice of standard or new treatment, respectively. We would like to assign patients to treatments in such a manner that the total discounted expected reward over the whole population of patients is maximized. The discount sequence is geometric with discount factor $\alpha \in (0, 1)$.

1.2. *Rationale for covariate model.* In clinical trials, the goals an experimenter would like to attain are diverse and often conflicting. Ethical considerations are prominent in all experimentation involving human subjects. Conflicts are invariably generated by the obligation of a researcher to balance the well-being of the current patient (individualistic view) with that of the future patients (utilitarian view) who stand to benefit from new advances in medical treatment. This long-standing dilemma has received considerable attention in both statistical and medical literature such as Anscombe (1963), Weinstein (1974), Byar, Simon, Friedewald, DeMets, Ellenberg, Gail and Ware (1976), Bartlett, Roloff, Cornell, Andrews, Dillon and Zwischenberger (1985) and Woodroofe and Hardwick (1990).

Received November 1989; revised January 1991.

[1] Research supported by U.S. Army Research Office under DAAL 03-88-0122 and NSF under DMS 89-02188.

AMS 1980 *subject classification*. 62L10.

Key words and phrases. Sequential allocation, one-armed bandit problem, Bayesian analysis, regret, myopic policy, asymptotically optimal.

Classical randomized clinical trials, which allocate approximately equal numbers of patients to different treatments, exemplify an extreme utilitarian goal. The individualistic goal, on the other hand, is exemplified by the "myopic" allocation rule in which patients are assigned to the treatment that has the highest current expected reward.

The main result of this article is that in the presence of a suitable covariate, the myopic procedure attains the utilitarian objective in addition to the individualistic one. For the use of covariate models, however, one must ensure the availability of concomitant information. In many medical trials, such information is likely to be present at little or no extra cost.

1.3. *Summary of results.* We formulate the maximization of the total discounted expected reward problem as a variation of the classical one-armed bandit problem, in which the decision to choose arm 1 or arm 0 depends on the present covariate as well as the previous covariates, allocations and responses that have been observed up to the present time. Let Y be the difference between the reward from the new treatment and the expected reward from the standard treatment. We have obtained a reasonably explicit asymptotic solution (as $\alpha \to 1$) to the maximization problem, in the case of a *one-parameter exponential family*, for which the conditional distribution of Y given X and θ may be described by

$$g(y|x,\theta) = \exp\{\theta t(x,y) - \psi(x,\theta)\}. \tag{1}$$

For this family of distributions, we describe the structure of the optimal Bayesian policy $\delta = (\delta_1, \delta_2, \dots)$ for a given prior distribution π_0 over Θ, using the dynamic programming equation. The description is nonconstructive but leads to easily verified conditions that prescribe whether arm 0 or arm 1 should be chosen.

Under some regularity conditions we show that the myopic rule is asymptotically (as $\alpha \to 1$) optimal, in the sense described in Section 2. The proofs of these results require detailed analysis of the sequence of likelihood functions and the sequence of posterior distributions. The style is similar in spirit to that of Woodroofe (1979) but the formulation is more general and there are technical differences in the proofs.

In Section 2, we give the mathematical formulation of the problem. Section 3 contains the main results, Theorems 1 and 2, together with an example. Some preliminary lemmas and propositions, including the strong consistency of MLEs (Proposition 3) and the asymptotic normality of the sequence of posterior distributions (Proposition 4), are presented in Section 4. Finally, in Section 5, we prove Theorems 1 and 2.

1.4. *Summary of references.* Bandit problems have been studied by various authors. A thorough discussion of bandit models appears in Berry and Fristedt (1985). The pioneering work in the realm of covariate models for a one-armed bandit problem was done by Woodroofe (1979) who studied an extremely simple model for the geometrically discounted responses from an infinite population. He established the asymptotic optimality of the myopic rule. Woodroofe (1982) studied the optimal allocation policy in the case of a covariate model for uniformly discounted responses from a finite population and investigated the asymptotic properties in the case of a large population. In this article we extend Woodroofe's (1979) model to a more general and hopefully more realistic model, preserving most of its good features.

Clayton (1989) investigated a discrete time, finite horizon uniformly discounted Bernoulli bandit with covariate. In his article the probability of success depended on the covariate through a "link" function such as logit or log-linear. He also developed a notion of Gitten's index for the covariate model.

2. Mathematical formulation.

2.1. *The model.* Let X, Y^0, Y^1 denote the covariate, the potential responses from arm 0 (standard treatment) and arm 1 (new treatment), respectively. Suppose that:

1. X has a known distribution F;
2. the conditional distribution $G^0(\cdot\,|x)$ of Y^0 given $X = x$ is known;
3. the conditional distribution $G^1(\cdot\,|x, \theta)$ of Y^1 given $X = x$ is specified by an unknown parameter θ, $\theta \in \Theta$;
4. Θ has a known prior distribution π.

Also suppose that Y^0, Y^1 are real valued; initially X, Θ may be quite general taking values in Polish spaces $\mathscr{X}$ and Θ, respectively. Also assume that F and π yield finite expectations for Y^0 and Y^1. Let (X_k, Y_k^0, Y_k^1), $k \geq 1$, be conditionally independent and identically distributed (iid) as (X, Y^0, Y^1) given $\Theta = \theta$.

2.2. *Policies and their worth.* Suppose further that $X_1, X_2, \ldots$ are observed sequentially and that for each k, we may observe either Y_k^0 or Y_k^1, but not both. By a sequential allocation, we shall mean a sequence $\delta = (\delta_1, \delta_2, \ldots)$ in which each δ_k takes the value 0 or 1 according as we observe Y_k^0 or Y_k^1 as a function of X_k and $\mathscr{F}_{k-1}^\delta$, where $\mathscr{F}_{k-1}^\delta$ denotes the σ-field generated by the relevant data available at time k, that is,

$$\mathscr{F}_k^\delta := \sigma\big(X_1, \ldots, X_k, \delta_1, \ldots, \delta_k, \delta_1 Y_1^1 + (1 - \delta_1) Y_1^0, \ldots, \delta_k Y_k^1 + (1 - \delta_k) Y_k^0\big).$$

$\mathscr{F}_k^\delta$ may be denoted by $\mathscr{F}_k$ if the dependence on δ is clear from the context. The expected α-worth of a policy δ when the prior is π is defined to be

$$V_\alpha(\delta, \pi) = E^\pi\left[\sum_{k=1}^{\infty} \alpha^{k-1}\{\delta_k Y_k^1 + (1 - \delta_k) Y_k^0\}\right], \tag{2}$$

where E^π denotes the expectation with respect to the joint distribution of Θ and (X_k, Y_k^0, Y_k^1), $k \geq 1$. The series in (2) converges almost everywhere and may be integrated termwise in view of the assumption that F, G^0, G^1 and π have finite expectations.

Given α and π, we seek to maximize (2) by choosing δ.

As in Woodroofe (1979), we may reduce the problem to the case $Y^0 \equiv 0$, by letting $Y_k = Y_k^1 - E\{Y_k^0|X_k\}$ for $k = 1, 2, \dots$. Then

$$V_\alpha(\delta, \pi) = \frac{1}{1-\alpha} E[Y^0] + U_\alpha(\delta, \pi),$$

where

$$U_\alpha(\delta, \pi) = E^\pi\left[\sum_{k=1}^{\infty} \alpha^{k-1}\delta_k Y_k\right] \tag{3}$$

for all δ and π. Thus it is sufficient to maximize $U_\alpha(\delta, \pi)$ with respect to δ; and $U_\alpha(\delta, \pi)$ is of the same form as $V_\alpha(\delta, \pi)$ with Y^1 replaced by Y and Y^0 replaced by 0. Moreover, for any $k \geq 1$, the posterior distribution of Θ given $\mathscr{F}_k^\delta$ is the same as the posterior distribution of Θ given $X_1, \dots, X_k$; $\delta_1, \dots, \delta_k$; $\delta_1 Y_1, \dots, \delta_k Y_k$.

From now on, we shall suppose that $Y^0 = 0$ and write Y for Y^1. Let

$$U_\alpha(\pi) = \sup_{\delta'} U_\alpha(\delta', \pi), \tag{4}$$

where the supremum extends over all δ'. Then U_α is a convex function of π because $U_\alpha(\delta, \pi)$ is linear in π for each δ [cf. DeGroot (1970), pages 125–128]. We call a policy δ^α optimal if and only if δ^α attains the supremum in (4), and we call the supremum itself the value of the bandit problem.

2.3. *Notation and assumptions.* Let $G(\cdot|x, \theta)$ denote the conditional distribution of Y given $X = x$, $\Theta = \theta$. Assume for each $x \in \mathscr{X}$ the family of distributions $\{G(\cdot|x, \theta)\colon \theta \in \Theta\}$ is dominated by a σ-finite measure Λ_x, with versions of conditional densities $\{g(\cdot|x, \theta)\colon \theta \in \Theta\}$ such that $G(dy|x, \theta) = g(y|x, \theta)\, d\Lambda_x(y)$. We suppose that g may be chosen measurably with respect to x, y and θ. Let $\pi^* = \pi^*(\cdot|x, y)$ denote the posterior distribution of Θ given $X = x$ and $Y = y$. Then

$$\pi^*(d\theta|x, y) \propto g(y|x, \theta)\, d\pi(\theta) \quad \text{a.e. } y(\Lambda_x), \text{ a.e. } x(F).$$

The conditional expectation of Y given $X = x$, $\Theta = \theta$ and that given $X = x$ are

$$\mu(x, \theta) = \int_{\mathbb{R}} y G(dy|x, \theta)$$

and

$$\bar{\mu}(x, \pi) = \int_{\Theta} \mu(x, \theta)\, d\pi(\theta)$$

for all x, θ and π. Let $y^+ = \max(0, y)$ for $-\infty < y < \infty$ and define

$$\nu(\theta) = \int_{\mathscr{X}} \mu(x, \theta)^+\, dF(x)$$

and

$$\bar{\nu}(\pi) = \int_{\Theta} \nu(\theta)\, d\pi(\theta).$$

If θ were revealed at each allocation time, one would get a reward of $\bar{\nu}(\pi)$ per patient on the average. Therefore, the maximum total discounted reward would be $\bar{\nu}(\pi)/(1-\alpha)$. Clearly then for each δ,

$$U_\alpha(\delta, \pi) \le U_\alpha(\pi) \le \frac{1}{1-\alpha}\bar{\nu}(\pi). \tag{5}$$

2.4. *Construction of optimal policies.* Suppose we have observed $X_1 = x$. If we observe Y_1 and then proceed optimally, our expected conditional (given $X_1 = x$) gain is $\bar{\mu}(x, \pi) + \alpha E^{\pi}\{U_\alpha(\pi^*(\cdot|x, Y_1))|X_1 = x\}$. On the other hand, if we do not observe Y_1 and then continue optimally, our expected gain is $\alpha U_\alpha(\pi)$. Clearly, it is optimal to observe Y_1 if and only if the former is larger than the latter. So the optimal policy stipulates

$$\delta_1^\alpha(x;\pi) = 1 \quad \text{if and only if} \quad \bar{\mu}(x,\pi) + \alpha\rho_\alpha(x,\pi) > 0, \tag{6}$$

where

$$\rho_\alpha(x,\pi) = E^{\pi}\{U_\alpha(\pi^*(\cdot|x, Y_1)) - U_\alpha(\pi)|X_1 = x\}$$

is a nonnegative functional, by the convexity of U_α [cf. Woodroofe (1982), Appendix].

The description of optimal policy δ^α may be completed by replacing X_1 and π by X_k and the posterior given $\mathscr{F}_{k-1}^{\delta^\alpha}$ at later times $k = 2, 3, \dots$.

Remark 2.1. Observe that $\rho_\alpha(x, \pi)$ is a version of the conditional expectation of $U_\alpha(\pi^*(\cdot|x, Y_1)) - U_\alpha(\pi)$ given X_1, evaluated at $X_1 = x$. Thus it seems reasonable to regard $\alpha\rho_\alpha(x,\pi)$ as the expected gain in relevant information that would result from observing Y_1. It is hard to compute $\rho_\alpha(x,\pi)$ without imposing further restrictions such as strong ancillarity [cf. Woodroofe (1982)]. We shall not address the issue any further in this article.

2.5. *A class of strategies.* Let

$$\mathscr{C} = \{\delta = (\delta_1, \delta_2, \dots): \delta_k = 1 \text{ if } \bar{\mu}(X_k, \pi_{k-1}) > 0\}.$$

Notice that the myopic rule δ^0 given by

$$\delta_k^0 = 1 \quad \text{if and only if} \quad \bar{\mu}\left(X_k, \pi_{k-1}^{\delta^0}\right) > 0 \tag{7}$$

and the optimal rule δ^α given in (6) both belong to $\mathscr{C}$.

3. One-parameter exponential family model.

3.1. *Assumptions.* Let Π_0 be the class of all prior distributions on Θ and endow Π_0 with the topology of weak convergence. Let (X_k, Y_k), $k \ge 1$, be a conditionally iid sequence of random variables, given $\theta \in \Theta$. The following conditions are needed.

CONDITION 1. Θ is a compact interval of $\mathbb{R}$.

CONDITION 2. The cumulative distribution function for Y_k given x and θ is of the form

$$G(dy|x,\theta) = \exp\{\theta t(x,y) - \psi(x,\theta)\}\Lambda_x(dy), \tag{8}$$

where t is a measurable function on $\mathscr{X} \times \mathbb{R}$ and Λ_x is a nondegenerate sigma-finite measure on the Borel sets of $\mathbb{R}$ for each x. Moreover, there is an open set Θ_1 containing Θ such that

$$\exp\{\psi(x,\theta)\} = \int \exp\{\theta t(x,y)\}\Lambda_x(dy) < \infty$$

for all $\theta \in \Theta_1$ and all $x \in \mathscr{X}$.

CONDITION 3. The initial prior distribution π_0 has a positive continuous density ξ_0 on Θ^0, the interior of Θ.

CONDITION 4. For each $\theta \in \Theta_1$, we have the following two conditions.

CONDITION 4a. $\mu(x,\theta) := E[Y|x,\theta] = \int_{\mathbb{R}} yG(dy|x,\theta)$ is finite $\forall\ x \in \mathbb{R}$.

CONDITION 4b. $\mu(x,\theta)$ is differentiable with respect to x on an open set containing $\{x\colon \mu(x,\theta) = 0\}$ and $\{x\colon \mu(x,\theta) = 0\} \subseteq \{x\colon \mu_{10}(x,\theta) > 0\}$, where $\mu_{ij}(x,\theta) = \partial^{i+j}\mu(x,\theta)/\partial x^i \partial\theta^j$.

CONDITION 5. $\mathscr{X} \subseteq \mathbb{R}$ and F has a bounded continuous density f with respect to Lebesgue measure, and satisfies the following conditions.

CONDITION 5a. $F\{x\colon \mu(x,\theta) = 0\} = 0,\ \forall\ \theta \in \Theta$.

CONDITION 5b. $F\{x\colon \bar{\mu}(x,\pi) > 0\} > 0,\ \forall\ \pi \in \Pi_0$.

CONDITION 5c. $\int_{\mathbb{R}} \sup_\theta|\mu_{01}(x,\theta)|\, dF(x) < \infty$.

CONDITION 5d. $\int_{\mathbb{R}} \sup_\theta|\mu_{02}(x,\theta)|\, dF(x) < \infty$.

CONDITION 5e. $\int_{\mathbb{R}} \sup_\theta|\psi_{02}(x,\theta)^2|\, dF(x) < \infty$.

REMARK 3.1. By Condition 1, Π_0 is compact [cf. Billingsley (1968), Theorem 6.1].

REMARK 3.2. By Corollary 2.3 of Brown (1986), $\psi_{01}(x,\theta) = E[t(x,Y)|x,\theta]$ and $\psi_{02}(x,\theta) = \operatorname{Var}[t(x,Y)|x,\theta] > 0$ for all x and θ. Therefore, the log-likelihood function of θ given $X_1 = x$ and $Y_1 = y$, namely, $\log(g(y|x,\theta)) = \theta t(x,y) - \psi(x,\theta)$ is strictly concave in θ.

REMARK 3.3. Note the relationship between $\psi_{01}(x,\theta)$ and $\mu(x,\theta)$. By Lemma 2.8 of Brown (1986), $\psi_{01}(x,\theta)$ is analytic in θ for each x. So is $\mu(x,\theta)$.

REMARK 3.4. Condition 5a holds when μ is strictly increasing in x for each θ and Condition 5b holds when μ increases without bound as $x \to \infty$, for each θ. Condition 5b means that no matter what the prior distribution is, for some nontrivial covariate values, the new treatment is preferable to the standard treatment. Furthermore, if $\mu(x,\theta) > 0$ for all x, then the allocation problem is trivial: Always use the new treatment. Hence of interest is the case when $\mu(x,\theta)$ equals 0 for some x. By Condition 4b, there can be at most one such x.

REMARK 3.5. Conditions 4b and 5 are not satisfied by the model without covariate. Conditions 5a and 5b account for the simplicity of the solution to the allocation problem in the covariate model.

REMARK 3.6. Conditions 1 and 3 are used in proving Proposition 4. Condition 4b is essential to invoke the implicit function theorem in the proof of Lemma T1.1. Condition 5a is used in Lemma T1.6; Condition 5b in Proposition 1; Condition 5c in Lemma T1.1; Condition 5d in Lemmas T1.4, T1.5, T1.6 and T1.7 to apply the dominated convergence theorem; and Condition 5e in Propositions 1(ii) and 6.

3.2. *Example* 1. Suppose the reward for the standard treatment, given covariate x, is described by $Y^0 = 1 + \sqrt{x}Z$, $Z \sim N(0,1)$; and the reward for the new treatment, given covariate x and parameter θ, is described by $Y^1 = xV$, $V \sim N(\theta, 1)$, where $\theta \in \Theta$, a compact subset of $(0,\infty)$. Let the initial prior distribution on Θ be π_0 having a continuous density ξ_0.

In particular, for small values of x $(x < \theta^{-1})$, the standard treatment has a higher expected reward, whereas for large values of x $(x > \theta^{-1})$, the new treatment has not only higher expected reward but also large standard deviation. This may be an appropriate assumption if very little is known about the side effects of the new treatment. Therefore,

$$Y = Y^1 - E\left[Y^0 | X = x\right] = xV - 1 \sim N(x\theta - 1, x^2).$$

Hence

$$g(y; x, \theta) \propto \exp\left\{\theta \frac{y+1}{x} - \frac{\theta^2}{2}\right\}.$$

So that $\psi(x,\theta) = \theta^2/2$, $\mu(x,\theta) = x\theta - 1$ and $\sigma^{-2}(\theta) = 1 - F(\theta^{-1})$, where σ^2 is defined by (10) below. Note that, Conditions 1–4 hold. The myopic policy simplifies to $\delta_k^0 = 1$ if and only if $x > 1/\zeta_{k-1}$, where $\zeta_k = \zeta(\pi_k) = \int \theta \, d\pi_k(\theta)$.

Furthermore, $\psi_{02}(x,\theta) = 1$. So, with $K_n = \Sigma_{k=1}^n \delta_k$, the likelihood function of θ given $\mathscr{F}_n$ is

$$L_n(\delta, \theta) \propto \exp\left\{\frac{K_n}{2}\left(\theta - \frac{1}{K_n}\sum_{k=1}^n \delta_k \frac{Y_k + 1}{X_k}\right)^2\right\}.$$

Hence the MLE of θ given $\mathscr{F}_n$ is

$$\hat{\theta}_n = \frac{1}{K_n} \sum_{k=1}^{n} \delta_k \frac{Y_k + 1}{X_k}.$$

3.3. *Results.* In view of (5), we define the regret for a strategy δ as

$$R_\alpha(\pi, \delta) := \frac{1}{1-\alpha} \bar{\nu}(\pi) - U_\alpha(\pi, \delta). \tag{9}$$

That ν is twice continuously differentiable is shown in Lemma T1.1. Define

$$\sigma^{-2}(\theta) := \int_{\bar{\mu}(x,\theta)>0} \psi_{02}(x, \theta)\, dF(x), \tag{10}$$

$$c(\theta) := \tfrac{1}{2}\sigma^2(\theta) \left\{ \nu''(\theta) - \int_{\mu(x,\theta)>0} \mu_{02}(x, \theta)\, dF(x) \right\} \tag{11}$$

and

$$\bar{c}(\pi) = \int_\Theta c(\theta)\, d\pi(\theta). \tag{12}$$

For the *one-parameter exponential family* model, as described in (8), our two major results are Theorems 1 and 2.

THEOREM 1. *Under Conditions* 1–5e

$$R_\alpha(\delta^0, \pi) \sim \ln\left(\frac{1}{1-\alpha}\right) \bar{c}(\pi) \quad \textit{as } \alpha \to 1.$$

THEOREM 2. *Under Conditions* 1–5e

$$\inf_{\delta \in \mathscr{C}} R_\alpha(\delta, \pi) = R_\alpha(\delta^\alpha, \pi) \sim \ln\left(\frac{1}{1-\alpha}\right) \bar{c}(\pi) \quad \textit{as } \alpha \to 1.$$

3.4. *Comments.* These two theorems assert that the myopic policy δ^0 is asymptotically optimal, provided the regularity conditions 1–5e hold. An interesting comparison of this approximate solution to the allocation problem with the approximate solution to the stopping problem that arises when F is degenerate at 0 is given in Woodroofe (1979). For the purpose of this article, we wish to highlight the following two features:

1. The myopic and the optimal procedures are in very close agreement with each other in the covariate problem. Thus the myopic procedure does, in fact, fulfill the utilitarian goal in addition to the individualistic one.
2. While we have assumed F to be known, the myopic rule which is asymptotically optimal does not require the knowledge of F for its implementation.

In contrast, when F is degenerate the myopic and the optimal procedures have regrets of different orders of magnitude. This was established in

TABLE 1
Expected number of successes: π = *Beta*(1, 1)

	Uniform covariate		Noncovariate	
N	Optimal	Myopic	Optimal	Myopic
3	1.89	1.89	1.67	1.68
5	3.18	3.18	2.85	2.85
10	6.43	6.43	5.80	5.82
20	13.00	12.99	11.91	11.78
25	16.30	16.29	14.89	14.77
40	26.22	26.20	24.17	23.71
50	32.84	32.83	30.32	29.68
60	39.47	39.46	36.49	35.64
80	52.75	52.73	48.87	47.57
100	66.04	66.02	61.33	59.51

Woodroofe (1976) for the Normal model and extended to the general exponential family models in Lai (1987).

3.5. *Numerical illustration*. Consider, as in Woodroofe (1982), a Bernoulli response problem where the probability of success with the standard treatment defines the covariate X and that with the new treatment is a parameter θ. In other words, for the standard treatment, $P(Y^0 = 1|x, \theta) = x = 1 - P(Y^0 = 0|x, \theta)$, and for the new treatment, $P(Y^1 = 1|x, \theta) = \theta = 1 - P(Y^1 = 0|x, \theta)$. Let the initial prior distribution of θ be $\pi = \text{Beta}(a, b)$. The goal is to maximize the total number of successes among a population of N patients.

Notice that whereas the rest of the article is concerned with geometric discounting with rate α, this particular example deals with uniform discounting of successes from a finite number N of patients. The asymptotic worth of a policy in the geometric discounting problem as $\alpha \to 1$ can be approximated fairly well by the worth of a policy in the uniform discounting problem with a large horizon N.

For this latter problem, it is possible to compute the optimal worth numerically by dynamic programming. We compare a uniformly distributed covariate model [X is distributed as Uniform(0, 1)] with a noncovariate model (X is degenerate at 1/2). Table 1 gives the expected number of successes for the special case $\pi = \text{Beta}(1, 1)$. We see that the myopic procedure is nearly optimal in the uniform covariate model. For details see Sarkar (1990), Chapter 2.

4. Preliminary lemmas and propositions.

4.1. *Preliminaries*. First we need a definition.

DEFINITION 1 (Uniform convergence in probability). A sequence of measurable functions $\{U_n(\delta): n \geq 1\}$ is said to converge to 0 in P^π-probability, uniformly with respect to $\delta \in \mathscr{C}$ if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup_{\delta\in\mathscr{C}} P^{\pi}\big[|U_n(\delta)| > \varepsilon\big] = 0.$$

If π is degenerate at θ_0, we shall write P_{θ_0} instead of P^{π} in the above. For $\delta \in \mathscr{C}$, define

$$\begin{aligned} C_{k-1}(\delta) &= \{x\colon \delta_k = 1\},\\ K_n(\delta) &= \sum_{k=1}^{n} \delta_k, \qquad (13)\\ \Psi_n(\delta,\theta) &= \sum_{k=1}^{n} \delta_k \psi(X_k,\theta). \end{aligned}$$

Among other things we shall show that the covariate bandit problem is not a stopping problem. That is, for any prior distribution π and any rule $\delta \in \mathscr{C}$, the long-range proportion of patients assigned to the new treatment is bounded away from 0. This is proved in Proposition 1(i). To justify the data-dependent transformation of the parameter θ in (16), we need to prove Proposition 1(ii), equicontinuity of $\Psi_n''(\delta,\theta)/n$ and the strong consistency of the sequence of MLEs.

PROPOSITION 1.

$$\begin{aligned} \text{(i)}\qquad \liminf_{n\to\infty}\,\inf_{\delta\in\mathscr{C}} \frac{1}{n}K_n(\delta) &= \liminf_{n\to\infty}\,\inf_{\delta\in\mathscr{C}} \frac{1}{n}\sum_{k=1}^{n} F\{C_{k-1}(\delta)\}\\ (14)\qquad &\geq \inf_{\pi\in\Pi_0} F\{x\colon \bar{\mu}(x,\pi) > 0\} := q, \quad \textit{say}, \end{aligned}$$

and

$$\begin{aligned} \text{(ii)}\quad \liminf_{n\to\infty}\,\inf_{\delta\in\mathscr{C}} \frac{1}{n}\Psi_n''(\delta,\theta_0) &= \liminf_{n\to\infty}\,\inf_{\delta\in\mathscr{C}} \frac{1}{n}\sum_{k=1}^{n}\int_{C_{k-1}(\delta)} \psi_{02}(x,\theta_0)\,dF(x)\\ (15)\qquad &\geq \inf_{\pi\in\Pi_0}\int_{\{x\colon \bar{\mu}(x,\pi)>0\}}\left[\int_{\Theta}\psi_{02}(x,\theta)\,d\pi(\theta)\right]dF(x)\\ &:= r, \quad \textit{say}, \end{aligned}$$

in P_{θ_0}-probability, for all $\theta_0 \in \Theta$. Furthermore, $q > 0$ and $r > 0$.

PROOF. The equalities in the above are obtained by specializing Lemma P1.1 below by choosing $w(x) \equiv 1$ for (i) and $w(x) = \psi_{02}(x,\theta_0)$ for (ii) and using Condition 5e. Since the indicator function of an open set is lower semicontinuous, an adaptation of Example 17 of Pollard (1984) proves that $q > 0$ and $r > 0$. □

Lemma P1.1. *Let w be a measurable function for which $\int w^2(x)\,dF(x) < \infty$. Then*

$$\frac{1}{n}\sum_{k=1}^{n}\left[\delta_k w(X_k) - \int_{C_{k-1}(\delta)} w(x)\,dF(x)\right] \to 0$$

in P^{π}-probability, uniformly with respect to $\delta \in \mathscr{E}$.

Proof. For any $\delta \in \mathscr{E}$,

$$E^{\pi}\left[\left\{\frac{1}{n}\sum_{k=1}^{n}\left[\delta_k w(X_k) - \int_{C_{k-1}(\delta)} w(x)\,dF(x)\right]\right\}^2\right]$$

$$= \frac{1}{n^2}\sum_{k=1}^{n} E^{\pi}\left[\left\{\delta_k w(X_k) - \int_{C_{k-1}(\delta)} w(x)\,dF(x)\right\}^2\right] \le \frac{1}{n}\int w^2(x)\,dF(x),$$

which is independent of δ. So by Chebychev's inequality we get the result. □

4.2. *Equicontinuity of $\Psi_n''(\delta,\theta)/n$.*

Proposition 2. *For all $\theta_0 \in \Theta$,*

$$\lim_{\eta\to 0}\ \sup_{|\theta-\theta_0|\le\eta}\ \sup_{n\ge 1}\ \sup_{\delta\in\mathscr{E}} \frac{1}{n}\left|\Psi_n''(\delta,\theta) - \Psi_n''(\delta,\theta_0)\right| = 0,$$

a.s. (P_{θ_0}).

Proof. For each fixed n,

$$\sup_{|\theta-\theta_0|\le\eta} \frac{1}{n}\left|\Psi_n''(\delta,\theta) - \Psi_n''(\delta,\theta_0)\right| \le \frac{1}{n}\sum_{k=1}^{n}\ \sup_{|\theta-\theta_0|\le\eta}\left|\psi_{02}(X_k,\theta) - \psi_{02}(X_k,\theta_0)\right|,$$

which is independent of δ and converges to

$$\int_{\mathscr{X}}\ \sup_{|\theta-\theta_0|\le\eta}\left|\psi_{02}(x,\theta) - \psi_{02}(x,\theta_0)\right| dF(x)$$

a.s. (P_{θ_0}) as $n \to \infty$, by the strong law of large numbers. Now by Remark 3.3, $\psi_{02}(x,\theta)$ is continuous at θ_0 for all x. Therefore, the last integral converges to 0 as $\eta \to 0$. The result follows. □

4.3. *Strong consistency of the sequence of MLEs.* In Remark 3.2, we noted that for any δ, the log-likelihood function is strictly concave. Hence it attains its maximum at a unique point $\hat{\theta}_n$, the MLE of θ given $\mathscr{F}_n$.

Proposition 3. *There exist $\eta, \varepsilon_0 > 0$ such that*

$$P_\theta\left[\left|\hat{\theta}_n - \theta\right| > \varepsilon\right] \le 2\exp\{-\eta\varepsilon^2 n\},$$

$\forall\ n \ge 1, \forall\ 0 < \varepsilon \le \varepsilon_0, \forall\ \theta \in \Theta$ and $\forall\ \delta \in \mathscr{E}$.

Corollary P3.1. *For every $\varepsilon > 0$,*

$$\lim_{m\to\infty} \sup_{\delta\in\mathscr{C}} \sup_{\theta\in\Theta} P_\theta\left\{ \sup_{n\geq m} \left|\hat{\theta}_n - \theta\right| > \varepsilon \right\} = 0.$$

The proofs of Proposition 3 and Corollary P3.1 are given in the Appendix.

Remark 4.1. Propositions 1(ii), 2 and 3 imply that

$$\lim_{n\to\infty} \inf_{\delta\in\mathscr{C}} \frac{1}{n} \Psi_n''\left(\delta, \hat{\theta}_n\right) \geq r,$$

in P_{θ_0}-probability for all $\theta_0 \in \Theta$. So the following transformation makes sense:

$$Z_n := \sqrt{\Psi_n''\left(\delta, \hat{\theta}_n\right)}\left(\theta - \hat{\theta}_n\right). \tag{16}$$

4.4. *Asymptotic normality of posterior distributions of Z_n.* From Condition 2, the log-likelihood function of θ given $\mathscr{F}_n$ is

$$l_n(\delta, \theta) = \theta \sum_{k=1}^{n} \delta_k t(X_k, Y_k) - \Psi_n(\delta, \theta). \tag{17}$$

For a given policy δ, let $\pi_n = \pi_n^\delta$ be the posterior distribution of θ given $\mathscr{F}_n$. Let

$$\zeta_n = \zeta_n(\delta) = \int_\Theta \theta \pi_n(d\theta)$$

and

$$\sigma_n^2 = \sigma_n^2(\delta) = \int_\Theta (\theta - \zeta_n)^2 \pi_n(d\theta)$$

be, respectively, the posterior mean and variance of θ given $\mathscr{F}_n$. Let $\phi_n(\delta,z)$ denote the posterior density of Z_n given $\mathscr{F}_n$. Then

$$\phi_n(\delta, z) = \frac{1}{c_n} \exp\left\{l_n(\delta, \theta) - l_n\left(\delta, \hat{\theta}_n\right)\right\}\xi_0(\theta) \tag{18}$$

with

$$c_n = \int_\Theta \exp\left\{l_n(\delta, \theta) - l_n\left(\delta, \hat{\theta}_n\right)\right\}\xi_0(\theta)\, dz, \tag{19}$$

where θ and z are related by $z = \sqrt{\Psi_n''\left(\delta, \hat{\theta}_n\right)}(\theta - \hat{\theta}_n)$. Let $\phi(z)$ denote the density of a standard normal distribution.

Proposition 4. *For all $\theta_0 \in \Theta$,*

$$\int (1 + z^4)\left|\phi_n(\delta, z) - \phi(z)\right| dz \to 0$$

as $n \to \infty$, in P_{θ_0}-probability, uniformly with respect to $\delta \in \mathscr{C}$.

The proof of Proposition 4 is presented in the Appendix. The following corollaries are immediate.

COROLLARY P4.1. *For all $\theta_0 \in \Theta$, the posterior distribution π_n of Θ converges to the degenerate distribution at θ_0 in P_{θ_0}-probability, uniformly with respect to $\delta \in \mathscr{E}$.*

COROLLARY P4.2. *For all $\theta_0 \in \Theta$, $\sqrt{n}(\hat{\theta}_n - \zeta_n)$ converges to 0 in P_{θ_0}-probability, uniformly with respect to $\delta \in \mathscr{E}$.*

4.5. *Evaluation of $K_n(\delta)/n$ and $\Psi_n''(\delta, \hat{\theta}_n)/n$ for the myopic policy δ^0.* By virtue of Corollary P4.1, specializing Lemma P1.1 for δ^0, we obtain, for all $\theta \in \Theta$,

$$\begin{aligned}\frac{K_n(\delta^0)}{n} &= \frac{1}{n}\sum_{k=1}^{n} F\left\{x : \bar{\mu}\left(x, \pi_{k-1}^{\delta^0}\right) > 0\right\} + o_P(1) \\ &\to F\{x : \mu(x, \theta) > 0\}\end{aligned} \tag{20}$$

and

$$\begin{aligned}\frac{\Psi_n''\left(\delta^0, \hat{\theta}_n\right)}{n} &= \frac{1}{n}\sum_{k=1}^{n} \int_{\bar{\mu}(x, \pi_{k-1}^{\delta^0}) > 0} \psi_{02}\left(x, \hat{\theta}_n\right) dF(x) + o_P(1) \\ &\to \int_{\mu(x, \theta) > 0} \psi_{02}(x, \theta)\, dF(x) := \sigma^{-2}(\theta)\end{aligned} \tag{21}$$

in P_θ-probability.

PROPOSITION 5. *For all $\theta \in \Theta$, $\lim_{k \to \infty} k\sigma_k^2(\delta^0) = \sigma^2(\theta)$ in P_θ-probability.*

PROOF. Notice that, for any $\delta \in \mathscr{E}$,

$$\begin{aligned}k\sigma_k^2(\delta) &= k\int (\theta - \zeta_k)^2\, d\pi_k^\delta(\theta) \\ &= k\left\{\int \left(\theta - \hat{\theta}_k\right)^2 d\pi_k^\delta(\theta) - \left[\int \left(\theta - \hat{\theta}_k\right) d\pi_k^\delta(\theta)\right]^2\right\} \\ &= \frac{k}{\Psi_k''\left(\delta, \hat{\theta}_k\right)}\left\{\int z^2 \phi_k(\delta, z)\, dz - \left[\int z \phi_k(\delta, z)\, dz\right]^2\right\}.\end{aligned}$$

The proof is completed by (21) and Lemma A.1 (with $l = 1, 2$) of the Appendix. □

5. Proof of the theorems.

5.1. *Evaluation of the regret for the myopic policy.* For the myopic policy δ^0, let

$$B_k^0 := \left\{x \colon \bar{\mu}\left(x, \pi_k^{\delta^0}\right) > 0\right\}.$$

Therefore,

$$U_\alpha(\delta^0, \pi) = \sum_{k=1}^{\infty} \alpha^{k-1} E^{\pi}\left[\int_{B_{k-1}^0} \bar{\mu}\left(x, \pi_{k-1}^{\delta^0}\right) dF(x)\right]$$

and

$$R_\alpha(\delta^0, \pi) = \sum_{k=1}^{\infty} \alpha^{k-1} E^{\pi}\left[\bar{\nu}\left(\pi_{k-1}^{\delta^0}\right) - \int_{B_{k-1}^0} \bar{\mu}\left(x, \pi_{k-1}^{\delta^0}\right) dF(x)\right]. \tag{22}$$

5.2. *Proof of Theorem* 1. Since $\sum_{k=1}^{\infty} \alpha^k / k = -\ln(1-\alpha)$ for $\alpha \in (0,1)$, it is enough to show that

$$E^{\pi}\left[\left| k\left\{\bar{\nu}\left(\pi_k^{\delta^0}\right) - \int_{B_k^0} \bar{\mu}\left(x, \pi_k^{\delta^0}\right) dF(x)\right\} - c(\Theta)\right|\right] \to 0.$$

By Lemma T1.1 below, ν is twice continuously differentiable. So, for any $\delta \in \mathscr{C}$, we may write

$$\left[k\left\{\bar{\nu}\left(\pi_k^{\delta}\right) - \int_{B_k(\delta)} \bar{\mu}\left(x, \pi_k^{\delta}\right) dF(x)\right\}\right] = \mathrm{I} - \mathrm{II} - \mathrm{III} + \mathrm{IV},$$

where (dropping "δ" for notational simplicity)

$$\mathrm{I} := k\left[\bar{\nu}(\pi_k) - \left\{\nu(\zeta_k) + \tfrac{1}{2}\sigma_k^2 \nu''(\zeta_k)\right\}\right],$$

$$\mathrm{II} := k\left[\int_{A_k} \left\{\bar{\mu}(x, \pi_k) - \mu(x, \zeta_k) - \tfrac{1}{2}\mu_{02}(x, \zeta_k)\sigma_k^2\right\} dF(x)\right],$$

$$\mathrm{III} := k\left[\int_{B_k} \bar{\mu}(x, \pi_k)\, dF(x) - \int_{A_k} \bar{\mu}(x, \pi_k)\, dF(x)\right],$$

$$\mathrm{IV} := \tfrac{1}{2} k \sigma_k^2 \left[\nu''(\zeta_k) - \int_{A_k} \mu_{02}(x, \zeta_k)\, dF(x)\right]$$

with

$$A_k = \{x \colon \mu(x, \zeta_k) > 0\}$$

and

$$B_k = \{x \colon \bar{\mu}(x, \pi_k) > 0\}.$$

By Lemmas T1.4, T1.5 and T1.6 below, $E^{\pi}[|\mathrm{I}| + |\mathrm{II}| + |\mathrm{III}|] \to 0$ uniformly in $\delta \in \mathscr{E}$. Also by Lemma T1.7, $E^{\pi}[|\mathrm{IV}(\delta^0) - c(\Theta)|] \to 0$. This completes the proof of Theorem 1. □

The proofs of the lemmas cited in the above discussion are given below.

LEMMA T1.1. *ν is twice continuously differentiable.*

PROOF. We shall explicitly evaluate $\nu'(\theta)$ and $\nu''(\theta)$. In view of Condition 4b and Remark 3.4, using the theorem of implicit functions [cf. Courant (1937), page 114], for each $\theta \in \Theta$, there exists a unique $x(\theta)$ such that

$$\mu(x(\theta), \theta) = 0$$

and the function $x(\theta)$ is differentiable with derivative

$$x'(\theta) = -\frac{\mu_{01}(x(\theta), \theta)}{\mu_{10}(x(\theta), \theta)}.$$

So that

$$\{\mu(x, \theta) < 0 < \mu(x, \theta')\} = \{x(\theta') < x < x(\theta)\}.$$

Next, we apply Theorem 16.8 of Billingsley (1979) successively twice to get

$$\nu'(\theta) = \int_{\mu(x, \theta) > 0} \mu_{01}(x, \theta)\, dF(x) \tag{23}$$

and

$$\nu''(\theta) = \int_{\mu(x, \theta) > 0} \mu_{02}(x, \theta)\, dF(x) + f(x(\theta)) \frac{\mu_{01}^2(x(\theta), \theta)}{\mu_{10}(x(\theta), \theta)}. \tag{24}$$

The analyticity of $\mu(x, \theta)$ implies the continuity of $\nu''(\theta)$. □

REMARK. The explicit formula (24) for $\nu''(\theta)$ simplifies (11) to give

$$c(\theta) = \frac{1}{2}\sigma^2(\theta) f(x(\theta)) \frac{\mu_{01}^2(x(\theta), \theta)}{\mu_{10}(x(\theta), \theta)}. \tag{25}$$

EXAMPLE 1 (Continued). Suppose further that $F(x) = 1 - e^{-x}$, $x \in (0, \infty)$. Then Conditions 5–5e hold and $\nu(\theta) = \theta \exp\{-\theta^{-1}\}$, $\nu'(\theta) = (1 + \theta^{-1})\exp\{-\theta^{-1}\}$ and $\nu''(\theta) = \theta^{-3}\exp\{-\theta^{-1}\}$. One may easily check that the formulas (23) and (24) give the same results.

LEMMA T1.2. *Let $\delta^k = (\delta_1^k, \delta_2^k, \dots) \in \mathscr{E}$, $\forall\, k \geq 1$ and write $\sigma_k^2 = \sigma_k^2(\delta^k)$. Then $\{k\sigma_k^2;\ k \geq 1\}$ are uniformly integrable.*

PROOF. Note that $k\sigma_k^2 \leq E^{\pi}[k(\hat{\theta}_k - \theta)^2 | \mathscr{F}_{k-1}]$. So it is enough to show that $\{k(\hat{\theta}_k - \theta)^2;\ k \geq 1\}$ are uniformly integrable. This follows easily from

Proposition 3 which implies that

$$E_\theta\left[k^2\left(\hat{\theta}_k - \theta\right)^4\right] \le Ck^2P_\theta\left[\left|\hat{\theta}_k - \theta\right| \ge \varepsilon_0\right] + 4\int_0^{\varepsilon_0\sqrt{k}} s^3P_\theta\left[\left|\hat{\theta}_k - \theta\right| \ge \frac{s}{\sqrt{k}}\right] ds$$

$$\le 2Ck^2\exp\left(-\eta\varepsilon_0^2k\right) + 8\int_0^\infty s^3\exp\left(-\eta s^2\right) ds$$

for all $k \ge 3$, where η and ε_0 are as in Proposition 3 and $C = (\sup\theta - \inf\theta)^4 < \infty$ by Condition 1. Since these remain bounded as $k \to \infty$ uniformly in θ, $E^\pi[k(\hat{\theta}_k - \theta)^2]$ remains bounded as $k \to \infty$, so that $\{k(\hat{\theta}_k - \theta)^2; k \ge 1\}$ are uniformly integrable. □

LEMMA T1.3. *For $\eta > 0$, let*

$$Q\left(\pi_k^\delta, \eta\right) := \int_{|\theta - \zeta_k| > \eta} (\theta - \zeta_k)^2 d\pi_k^\delta(\theta). \tag{26}$$

Then

$$\lim_{k\to\infty} \sup_{\delta\in\mathscr{E}} kE^\pi\left[Q\left(\pi_k^\delta, \eta\right)\right] = 0$$

for all $\eta > 0$.

PROOF. It is enough to show the result for arbitrary $\delta^k \in \mathscr{E}$ since we can choose δ^k to be near where the supremum is attained. Then letting $\delta = \delta^k$, it suffices to show that $kQ(\pi_k^\delta, \eta) \to 0$ in P_π-probability [since $Q(\pi_k^\delta, \eta) \le \sigma_k^2$, for all $k \ge 1$; and $\{k\sigma_k^2; k \ge 1\}$, are uniformly integrable]. Notice that

$$kQ\left(\pi_k^\delta, \eta\right) \le 4k\int_{|\theta - \hat{\theta}_k| > \eta/2} \left(\theta - \hat{\theta}_k\right)^2 d\pi_k^\delta(\theta) + \mathrm{I}\left(\left|\hat{\theta}_k - \zeta_k\right| > \frac{\eta}{2}\right)k\sigma_k^2(\delta)$$

$$\le \frac{4k}{\Psi_k''\left(\delta, \hat{\theta}_k\right)}\int_{|\theta - \hat{\theta}_k| > \eta/2} z^2\phi_k(\delta, z)\, dz + \mathrm{I}\left(\left|\hat{\theta}_k - \zeta_k\right| > \frac{\eta}{2}\right)k\sigma_k^2(\delta),$$

which approaches 0 in P_{θ_0}-probability as $k \to \infty$, by Proposition 4, Corollary P4.2 and the remark in Section 4.3. □

LEMMA T1.4.

$$\lim_{k\to\infty} \sup_{\delta\in\mathscr{E}} E^\pi[|\mathrm{II}|] = 0.$$

PROOF. Let $\varepsilon > 0$ be given. By Lemma T1.1, there exists an $\eta_0 > 0$ such that

$$\sup_{|\theta_1 - \theta_2| < \eta_0} \left|\nu''(\theta_1) - \nu''(\theta_2)\right| < \varepsilon.$$

Then

$$|\mathrm{I}| = k\left|\int_{\Theta}\left\{\nu(\theta) - \nu(\zeta_k) - \tfrac{1}{2}\nu''(\zeta_k)(\theta - \zeta_k)^2\right\} d\pi_k^{\delta}(\theta)\right|$$
$$\leq \tfrac{1}{2}\varepsilon k \sigma_k^2 + kQ(\pi_k, \eta_0)\sup_{\theta}|\nu''(\theta)|.$$

Therefore, Lemma T1.4 follows from Condition 1 and Lemmas T1.2 and T1.3. □

LEMMA T1.5.

$$\lim_{k\to\infty}\sup_{\delta\in\mathscr{E}} E^{\pi}[|\mathrm{II}|] = 0.$$

PROOF. For $\eta > 0$, let

$$\varepsilon_{\eta}(x) = \sup_{|\theta' - \theta|\leq\eta}|\mu_{02}(x,\theta') - \mu_{02}(x,\theta)|$$

for $x \in \mathbb{R}$. Then

$$E^{\pi}[|\mathrm{II}|] \leq \tfrac{1}{2}\int\varepsilon_{\eta}(x)\, dF(x) \times E^{\pi}\left[k\sigma_k^2(\delta)\right]$$
$$+ \int\sup_{\theta}|\mu_{02}(x,\theta)|\, dF(x) \times E^{\pi}\left[kQ\left(\pi_k^{\delta}, \eta\right)\right],$$

which approaches 0 uniformly in $\delta \in \mathscr{E}$, as $k \to \infty$ and then $\eta \to 0$; by Lemmas T1.2 and T1.3, Condition 5d and the analyticity of μ. □

LEMMA T1.6.

$$\lim_{k\to\infty}\sup_{\delta\in\mathscr{E}} E^{\pi}[|\mathrm{III}|] = 0.$$

PROOF.

$$\mathrm{III} = k\left[\int_{\mu(x,\zeta_k^{\delta})<0<\bar{\mu}(x,\pi_k^{\delta})}\bar{\mu}\left(x,\pi_k^{\delta}\right) dF(x) - \int_{\bar{\mu}(x,\pi_k^{\delta})<0<\mu(x,\zeta_k^{\delta})}\bar{\mu}\left(x,\pi_k^{\delta}\right) dF(x)\right]$$
$$= \mathrm{III}^1 - \mathrm{III}^2, \quad \text{say}.$$

Now observe that

$$|\mathrm{III}^1| \leq k\int_{\mu(x,\zeta_k^{\delta})<0<\bar{\mu}(x,\pi_k^{\delta})}\left[\bar{\mu}\left(x,\pi_k^{\delta}\right) - \mu\left(x,\zeta_k^{\delta}\right)\right] dF(x)$$
$$\leq k\sigma_k^2(\delta)\int_{\mu(x,\zeta_k^{\delta})<0<\bar{\mu}(x,\pi_k^{\delta})}\sup_{\theta\in\Theta}\left|\tfrac{1}{2}\mu_{02}(x,\theta)\right| dF(x).$$

By Corollary P4.1 and Condition 5a

$$F\left\{x\colon \mu\left(x,\zeta_k^{\delta}\right) < 0 < \bar{\mu}\left(x,\pi_k^{\delta}\right)\right\} \to 0$$

uniformly in $\delta \in \mathscr{C}$. Therefore, by Condition 5d and Lemma T1.2, the dominated convergence theorem gives

$$\lim_{k\to\infty} \sup_{\delta\in\mathscr{C}} E^{\pi}[|\mathrm{III}^1|] = 0.$$

Similarly,

$$\lim_{k\to\infty} \sup_{\delta\in\mathscr{C}} E^{\pi}[|\mathrm{III}^2|] = 0. \qquad \square$$

Lemma T1.7.

$$\lim_{k\to\infty} E^{\pi}\big[\,|\mathrm{IV}(\delta^0) - c(\Theta)|\,\big] = 0.$$

Proof. Lemma T1.1 and Corollary P4.1 together imply that $\nu''(\zeta_k^{\delta^0}) \to \nu''(\theta)$. By Proposition 5, $k\sigma_k^2(\delta^0) \to \sigma^2(\theta)$. Finally, by Corollary P4.1 and the dominated convergence theorem,

$$\int_{\mu(x,\zeta_k^\delta)>0} \mu_{02}(x,\zeta_k^\delta)\,dF(x) \to \int_{\mu(x,\theta)>0} \mu_{02}(x,\theta)\,dF(x).$$

So, $\mathrm{IV}(\delta^0) \to c(\theta)$ a.s. (P_θ). Next, see that for all $\delta \in \mathscr{C}$,

$$|\mathrm{IV}(\delta)| \le \tfrac{1}{2}\sup_{\Theta}\left[\nu''(\theta) - \int_{\mu(x,\theta)>0} \mu_{02}(x,\theta)\,dF(x)\right] k\sigma_k^2.$$

Therefore, by Conditions 1 and 5d, Lemmas T1.1 and T1.2 and the dominated convergence theorem, we complete the proof. □

5.3. *Lower bound on the regret for the optimal policy δ^α.* For the optimal procedure δ^α let $B_k = \{x\colon \bar{\mu}(x,\pi_k^{\delta^\alpha}) > 0\}$ and recall from (13) that $C_k = \{x\colon \bar{\mu}(x,\pi_k^{\delta^\alpha}) + \alpha\rho_\alpha(x,\pi_k^{\delta^\alpha}) > 0\}$. Then

$$\begin{aligned} R_\alpha(\delta^\alpha,\pi) &= \sum_{k=1}^{\infty} \alpha^{k-1}E^{\pi}\left[\bar{\nu}(\pi_{k-1}^{\delta^\alpha}) - \int_{B_{k-1}} \bar{\mu}(x,\pi_{k-1}^{\delta^\alpha})\,dF(x)\right] \\ &\quad + \sum_{k=1}^{\infty} \alpha^{k-1}E^{\pi}\left[\int_{C_{k-1}-B_{k-1}} |\bar{\mu}(x,\pi_{k-1}^{\delta^\alpha})|\,dF(x)\right]. \end{aligned} \tag{27}$$

Notice that all the terms on the right-hand side of (27) are nonnegative. Therefore,

$$R_\alpha(\delta^\alpha,\pi) \ge E^{\pi}\left[\sum_{k\in J_\alpha} \frac{\alpha^{k-1}}{k}(\mathrm{I} - \mathrm{II} - \mathrm{III} + \mathrm{IV})\right],$$

where I, II, III and IV are as in the proof of Theorem 1 with $\delta = \delta^\alpha$ and

$$J_\alpha := \left\{k\colon \ln^2\!\left(\frac{1}{1-\alpha}\right) \le k \le \frac{1}{1-\alpha}\right\}.$$

5.4. *Proof of Theorem* 2. By Lemmas T1.4, T1.5 and T1.6, which hold uniformly with respect to $\delta \in \mathscr{E}$,

$$\limsup_{\alpha \to 1} \max_{k \in J_\alpha} E^\pi[|\mathrm{I}| + |\mathrm{II}| + |\mathrm{III}|] = 0.$$

Therefore, the main interest centers on IV. First note that, as $\alpha \to 1$,

$$\sum_{k > 1/(1-\alpha)} \frac{\alpha^{k-1}}{k} \le 1$$

and

$$\sum_{k < \ln^2[1/(1-\alpha)]} \frac{\alpha^{k-1}}{k} \sim \ln\left[\ln^2\left(\frac{1}{1-\alpha}\right)\right] = o\left(\ln \frac{1}{1-\alpha}\right).$$

Therefore, as $\alpha \to 1$,

$$\sum_{k \in J_\alpha} \frac{\alpha^{k-1}}{k} \sim \ln\left(\frac{1}{1-\alpha}\right).$$

By Proposition 6 below and arguments similar to Lemma T1.7,

$$\limsup_{\alpha \to 1} \max_{k \in J_\alpha} E^\pi\big[|\mathrm{IV}(\delta^\alpha) - c(\Theta)|\big] = 0.$$

So that

$$\liminf_{\alpha \to 1} \frac{R_\alpha(\delta^\alpha, \pi)}{\ln(1/(1-\alpha))} \ge \bar{c}(\pi).$$

This completes the proof of Theorem 2. □

Proposition 6.

$$\limsup_{\alpha \to 1} \max_{k \in J_\alpha} E^\pi\big[|k\sigma_k^2(\delta^\alpha) - \sigma^2(\theta)|\big] = 0.$$

Proof. For $n \in J_\alpha$,

$$\begin{aligned} &E^\pi\left[\frac{1}{n}\left\{\Psi_n''(\delta^\alpha, \hat{\theta}_n) - \Psi_n''(\delta^0, \hat{\theta}_n)\right\}\right] \\ &\quad \le \frac{1}{n} \sum_{k=1}^{n} \int_{\bar{\mu}(x,\pi_k) < 0 < \bar{\mu}(x,\pi_k) + \alpha\rho_\alpha(x,\pi_k)} \left|\psi_{02}(x, \hat{\theta}_n)\right| dF(x) \\ &\quad \le \left[\int \sup_\theta \psi_{02}(x,\theta)^2\, dF(x)\right]^{1/2} \\ &\qquad \times \left[\frac{1}{n} \sum_{k=1}^{n} F\{\bar{\mu}(x,\pi_k) < 0 < \bar{\mu}(x,\pi_k) + \alpha\rho_\alpha(x,\pi_k)\}\right]^{1/2} \end{aligned} \tag{28}$$

by the Cauchy–Schwarz inequality. It is shown below that the last sum approaches 0 uniformly in $n \in J_\alpha$ as $\alpha \to 1$. Therefore, by Condition 5e, we have

$$\limsup_{\alpha \to 1} \max_{n \in J_\alpha} E^\pi\left[\frac{1}{n}\left\{\Psi_n''\left(\delta^\alpha, \hat{\theta}_n\right) - \Psi_n''\left(\delta^0, \hat{\theta}_n\right)\right\}\right] = 0.$$

Hence, by Corollary P3.2 and Proposition 5,

$$\limsup_{\alpha \to 1} \max_{k \in J_\alpha} E^\pi\left[\left|\frac{k}{\Psi_k''\left(\delta^\alpha, \hat{\theta}_k\right)} - \sigma^2(\theta)\right|\right] = 0.$$

An argument similar to that in the proof of Proposition 5 now completes the proof of Proposition 6. It remains to show that the last sum in (28) approaches 0. Given $\varepsilon > 0$, choose $\eta > 0$ such that

$$\limsup_{\alpha \to 1} \max_{n \in J_\alpha} E^\pi\left[\frac{1}{n} \sum_{k=1}^{n} F\{-\eta < \bar{\mu}(x, \pi_k) < 0\}\right] < \varepsilon.$$

This is possible since there exists $\eta > 0$ such that $F\{-\eta/2 < \mu(x,\theta) < 0\} < \varepsilon$ for all $\theta \in \Theta$ and by Corollary P4.1, $|\bar{\mu}(x, \pi_k) - \mu(x,\theta)| < \eta/2$ for all sufficiently large k. By (27), for α close to 1 and $n \in J_\alpha$,

$$\begin{aligned} E^\pi\left[\frac{1}{n}\sum_{k=1}^{n} F\{C_k - B_k\}\right] &\le E^\pi\left[\frac{1}{n}\sum_{k=1}^{n} F\{-\eta < \bar{\mu}(x,\pi_k) < 0\}\right. \\ &\qquad \left. + \frac{1}{\eta n \alpha^n}\sum_{k=1}^{n} \alpha^{k-1}\int_{C_k - B_k} |\bar{\mu}(x,\pi_k)|\, dF(x)\right] \\ &\le 2\varepsilon + \frac{1}{\eta n \alpha^n} R_\alpha(\delta^\alpha, \pi) \le 2\varepsilon + \frac{4e\bar{c}(\pi)}{\eta \ln(1/(1-\alpha))} \end{aligned}$$

since for α close to 1 and $n \in J_\alpha$,

$$n\alpha^n \ge \frac{1}{2e}\ln^2\left(\frac{1}{1-\alpha}\right)$$

and

$$R_\alpha(\delta^\alpha, \pi) \le R_\alpha(\delta^0, \pi) \le 2\bar{c}(\pi)\ln\left(\frac{1}{1-\alpha}\right).$$

Since ε is arbitrary, the proof is complete. □

APPENDIX

Strong consistency of the MLEs.

PROOF OF PROPOSITION 3. Recall the definitions of q and r from (14) and (15). Also recall from Remark 3.2 that the log-likelihood function is strictly concave. Therefore,

$$P_\theta\left[\hat{\theta}_n > \theta + \varepsilon\right] = P_\theta\left[l_n'(\delta, \theta + \varepsilon) > 0\right]$$

for all $\theta \in \Theta$, $\varepsilon > 0$ for which $\theta + \varepsilon \in \Theta$ and all $n \geq 1$. From (17),

$$S_n := l'_n(\delta, \theta + \varepsilon) = \sum_{k=1}^{n} \delta_k [t(X_k, Y_k) - \psi_{01}(X_k, \theta + \varepsilon)]$$

and

$$w_+(x,y) = t(x,y) - \psi_{01}(x, \theta + \varepsilon).$$

Then

$$\begin{aligned}
\Delta_+(\varepsilon, \theta) &:= \Delta\left(w_+, \frac{\varepsilon}{2}, \theta\right) \\
&= \int_{\mathscr{X}} \int_{\mathscr{Y}} \exp\left\{\frac{\varepsilon}{2}[t(x,y) - \psi_{01}(x, \theta + \varepsilon)]\right\} g(y; x, \theta)\, d\Lambda_x(Y)\, dF(x) \\
&= \int_{\mathscr{X}} \left[\int_{\mathscr{Y}} \exp\left\{\left(\theta + \frac{\varepsilon}{2}\right) t(x,y)\right\} d\Lambda_x(Y)\right] \\
&\qquad \times \exp\left\{-\frac{\varepsilon}{2}\psi_{01}(x, \theta + \varepsilon) - \psi(x, \theta)\right\} dF(x) \\
&= \int_{\mathscr{X}} \exp\left\{\psi\left(x, \theta + \frac{\varepsilon}{2}\right) - \psi(x, \theta) - \frac{\varepsilon}{2}\psi_{01}(x, \theta + \varepsilon)\right\} dF(x) \\
&\leq \int_{\mathscr{X}} \exp\left\{-\frac{\varepsilon}{2}\left[\psi_{01}(x, \theta + \varepsilon) - \psi_{01}\left(x, \theta + \frac{\varepsilon}{2}\right)\right]\right\} dF(x).
\end{aligned}$$

The exponent in the last integrand is negative since $\psi_{01}(x, \theta)$ is monotonically increasing in θ (assuming $\theta + \varepsilon \in \Theta_1$). Hence $0 < \Delta_+(\varepsilon, \theta) < 1$, $\forall\, \varepsilon > 0$. Also

$$\begin{aligned}
&\frac{1 - \Delta_+(\varepsilon, \theta)}{\varepsilon^2} \\
&\qquad \geq \int_{\mathscr{X}} \frac{1 - \exp\{-(\varepsilon/2)[\psi_{01}(x, \theta + \varepsilon) - \psi_{01}(x, \theta + \varepsilon/2)]\}}{\varepsilon^2}\, dF(x).
\end{aligned}$$

Using L'Hospital's rule the limit of the integrand on the right-hand side as $\varepsilon \to 0$ can be shown to be $\psi_{02}(x, \theta)/4$. Hence

$$\liminf_{\varepsilon \to 0} \frac{1 - \Delta_+(\varepsilon, \theta)}{\varepsilon^2} \geq \frac{1}{4}\int_{\mathscr{X}} \psi_{02}(x, \theta)\, dF(x) \geq \frac{r}{4},$$

which does not depend on θ and is positive. In fact, by the compactness of Θ, the last relation holds uniformly in θ. So there exists $\varepsilon_+ > 0$ such that for all $0 < \varepsilon < \varepsilon_+$ and all $\theta \in \Theta$,

$$1 - \Delta_+(\varepsilon, \theta) \geq \frac{r}{8}\varepsilon^2.$$

Therefore, for all $0 < \varepsilon \le \varepsilon_+$ and all $\theta \in \Theta$,

$$\begin{aligned} P_\theta\big[\hat{\theta}_n > \theta + \varepsilon\big] &= P_\theta[S_n > 0] \le E_\theta\big[\exp\{\tfrac{1}{2}\beta S_n\}\big] \\ &\le \sqrt{E_\theta\big[\Delta_+(\varepsilon,\theta)^{K_n}\big]} \le \exp\{-\tfrac{1}{2}q[1-\Delta_+(\varepsilon,\theta)]n\} \\ &\le \exp\left\{-\frac{qr}{16}\varepsilon^2 n\right\}. \end{aligned}$$

The first inequality above is the Bernstein inequality, the second and the third inequalities hold by Lemmas A2 and A1, respectively, of Woodroofe (1979).

A similar bound for $P_\theta[\hat{\theta}_n < \theta - \varepsilon]$ can be obtained for $0 < \varepsilon < \varepsilon_-$ by replacing w_+ by w_- and taking $\Delta_-(\varepsilon,\theta) = \Delta(w_-, -\varepsilon/2, \theta)$, where $w_-(x,y) = t(x,y) - \psi_{01}(x, \theta - \varepsilon)$.

Proposition 3 follows with $\varepsilon_0 = \min(\varepsilon_+, \varepsilon_-)$ and $\eta = qr/16$. □

Proof of Corollary P3.1. The proof follows easily by summing.

$$\begin{aligned} P_\theta\left[\sup_{n\ge m}\big|\hat{\theta}_n - \theta\big| > \varepsilon\right] &\le \sum_{n=m}^{\infty} P_\theta\big[\big|\hat{\theta}_n - \theta\big| > \varepsilon\big] \\ &\le 2\sum_{n=m}^{\infty}\exp\{-\eta n\varepsilon^2\} = \frac{2\exp\{-\eta m\varepsilon^2\}}{1-\exp\{-\eta\varepsilon^2\}}, \end{aligned}$$

which converges to 0, as $m \to \infty$. □

Asymptotic normality of the posteriors. The proof of Proposition 4 is presented here. The following lemma is needed. For notational simplicity "δ" will be dropped in the sequel.

Lemma A.1. *For each $l \ge 0$,*

$$\int_{-\infty}^{\infty}|z|^l\left|\frac{\xi(\theta)}{\xi(\theta_0)}\exp\big\{l_n(\theta)-l_n(\hat{\theta}_n)\big\} - \exp\left\{-\frac{1}{2}z^2\right\}\right|dz \to 0$$

in P_{θ_0}-probability, uniformly in $\delta \in \mathscr{E}$ for all $\theta_0 \in \Theta^0$.

Proof. Given θ_0 and $\varepsilon > 0$, let η be so small that $[\theta_0 - 2\eta, \theta_0 + 2\eta] \in \Theta^0$ and

$$\sup_{|\theta_1-\theta_2|\le 2\eta}|\xi(\theta_1) - \xi(\theta_2)| \le \varepsilon\xi(\theta_0)$$

and let $\gamma > 0$ be so small that

$$\int_{-\infty}^{\infty}|z|^l\big[\exp\{-\tfrac{1}{2}(1-\gamma)z^2\} - \exp\{-\tfrac{1}{2}(1+\gamma)z^2\}\big]\,dz \le \varepsilon.$$

Let

$$A_n = \left\{\left|\frac{l_n''(\theta)}{l_n''(\theta_0)} - 1\right| \le \gamma, \forall\, |\theta-\theta_0| \le 2\eta\right\} \cap \big\{\big|\hat{\theta}_n - \theta_0\big| \le \eta\big\}.$$

The integral in the statement of Lemma A.1 is the sum of three terms I_n, II_n

and III_n with

$$\mathrm{I}_n = \int_{|z| \le \eta\sqrt{\Psi_n''(\hat{\theta}_n)}}, \qquad \mathrm{II}_n = \int_{z > \eta\sqrt{\Psi_n''(\hat{\theta}_n)}}, \qquad \mathrm{III}_n = \int_{z < -\eta\sqrt{\Psi_n''(\hat{\theta}_n)}},$$

the integrand being the same as in the statement of Lemma A.1. We shall show that each of these tends to 0 as $n \to \infty$.

If A_n occurs and $|\hat{\theta}_n - \theta_0| \le \eta$, then

$$\begin{aligned}
\mathrm{I}_n \le & \int_{|\theta - \hat{\theta}_n| \le \eta} |z|^l \left| \frac{\xi(\theta)}{\xi(\theta_0)} - 1 \right| \exp\left\{ l_n(\theta) - l_n(\hat{\theta}_n) \right\} dz \\
& + \int_{|\theta - \hat{\theta}_n| \le \eta} |z|^l \left| \exp\left\{ l_n(\theta) - l_n(\hat{\theta}_n) \right\} - \exp\left\{ -\tfrac{1}{2}(1 + \gamma) z^2 \right\} \right| dz \\
& + \int_{|\theta - \hat{\theta}_n| \le \eta} |z|^l \left| \exp\left\{ -\tfrac{1}{2}(1 + \gamma) z^2 \right\} - \exp\left\{ -\tfrac{1}{2} z^2 \right\} \right| dz \\
\le & \int_{|\theta - \hat{\theta}_n| \le \eta} \varepsilon |z|^l \exp\left\{ -\tfrac{1}{2}(1 - \gamma) z^2 \right\} dz \\
& + 2 \int_{|\theta - \hat{\theta}_n| \le \eta} |z|^l \left| \exp\left\{ -\tfrac{1}{2}(1 - \gamma) z^2 \right\} - \exp\left\{ -\tfrac{1}{2}(1 + \gamma) z^2 \right\} \right| dz \\
\le & \ \varepsilon \int |z|^l \exp\left\{ -\tfrac{1}{2}(1 - \gamma) z^2 \right\} dz + 2\varepsilon.
\end{aligned}$$

To show II_n is arbitrarily small we argue as follows. If A_n occurs and $z > \eta\sqrt{\Psi_n''(\hat{\theta}_n)}$, then $\theta > \hat{\theta}_n + \eta$ and

$$\begin{aligned}
l_n(\theta) - l_n(\hat{\theta}_n) & \le \frac{l_n(\hat{\theta}_n + \eta) - l_n(\hat{\theta}_n)}{\eta} (\theta - \hat{\theta}_n) \\
& \le \tfrac{1}{2}\eta l_n''(\theta_n^*)(\theta - \hat{\theta}_n) \le b_n z,
\end{aligned}$$

where $\hat{\theta}_n < \hat{\theta}_n^* < \hat{\theta}_n + \eta$ and

$$b_n = \tfrac{1}{2}\eta \left[\sup_{\hat{\theta}_n \le t \le \hat{\theta}_n + \eta} l''(t) \right] \Big/ \sqrt{\Psi_n''(\hat{\theta}_n)} .$$

Notice that on A_n, $b_n < -\eta(1 - \gamma)/4$. Therefore,

$$\mathrm{II}_n \le \frac{M}{\xi(\theta_0)} \int_{z > n\sqrt{\Psi_n''(\hat{\theta}_n)}} z^l \exp\left\{ -\frac{1}{4}\eta(1 - \gamma) z \right\} dz,$$

where $M = \sup_{\theta \in \Theta} \xi(\theta) < \infty$, by Conditions 1 and 3. Therefore, by Corollary P2.2, II_n is arbitrarily small on A_n, for all large n.

Likewise, we can show that III_n is arbitrarily small on A_n for all large n. Now, by Lemma 2, Proposition 3 and the compactness of Θ,

$$\lim_{n\to\infty} P_{\theta_0}[A_n] = 1$$

uniformly in $\delta \in \mathscr{C}$. This completes the proof of Lemma A.1. □

Proof of Proposition 4. From (18) and (19), we get the following inequalities:

$$|\phi_n(z) - \phi(z)| \le \left|\frac{\xi(\theta_0)}{c_n} - \frac{1}{\sqrt{2\pi}}\right| \frac{\xi(\theta)}{\xi(\theta_0)} \exp\{l_n(\theta) - l_n(\hat\theta_n)\}$$
$$+ \frac{1}{\sqrt{2\pi}}\left|\frac{\xi(\theta)}{\xi(\theta_0)} \exp\{l_n(\theta) - l_n(\hat\theta_n)\} - \exp\left\{-\frac{1}{2}z^2\right\}\right|$$

and

$$\left|\frac{c_n}{\xi(\theta_0)} - \sqrt{2\pi}\right| \le \int_{-\infty}^{\infty}\left|\frac{\xi(\theta)}{\xi(\theta_0)} \exp\{l_n(\theta) - l_n(\hat\theta_n)\} - \exp\left\{-\frac{1}{2}z^2\right\}\right| dz.$$

Proposition 4 follows easily. □

Acknowledgments. The author is indebted to Prof. Woodroofe for his constant help, guidance, encouragement and care. Special thanks are due to Prof. Bélisle, Jeffrey Eisele, Mauro Gasparini, Prof. Jeganathan, Prof. Rauch and Prof. Thelen. Thanks also to two unknown reviewers whose comments and suggestions resulted in substantial modification in the presentation of this article. The referee is to be credited for the concise summary.

REFERENCES

Anscombe, F. (1963). Sequential medical trials, *J. Amer. Statist. Assoc.* **58** 365–383.

Bartlett, R. H., Roloff, D. W., Cornell, R. G., Andrews, A. F., Dillon, P. W. and Zwischenberger, J. B. (1985). Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study. *Pediatrics* **76** 479–487.

Berry, D. A. and Fristedt, B. (1985). *Bandit Problems: Sequential Allocation of Experiments*. Chapman and Hall, London.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families*. IMS, Hayward, Calif.

Byar, D. P., Simon, R. M., Friedewald, W. T., DeMets, D. L., Ellenberg, J. J., Gail, M. H. and Ware, J. H. (1976). Randomized clinical trials: Perspectives on some recent ideas. *New England Journal of Medicine* **295** 74–80.

Clayton. M. K. (1989). Covariate models for Bernoulli bandits. *Sequential Anal.* **8** 405–426.

Courant, R. (1937). *Differential and Integral Calculus* **2**. Interscience, New York.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

Lai, T. L. (1987). Adaptive treatment allocation and the multi-armed bandit problem. *Ann. Statist.* **15** 1091–1114.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

Sarkar, J. (1990). Bandit problems with covariates: Sequential allocation of experiments. Ph.D. dissertation, Univ. Michigan.

WEINSTEIN, M. (1974). Allocation of subjects in medical experiments. *New England Journal of Medicine* **291** 1278–1285.

WOODROOFE, M. B. (1976). On the one-armed bandit problem. *Sankhyā Ser. A* **38** 79–91.

WOODROOFE, M. B. (1979). A one-armed bandit problem with a concomitant variable. *J. Amer. Statist. Assoc.* **74** 799–806.

WOODROOFE, M. B. (1982). Sequential allocation with covariates. *Sankhyā Ser. A* **44** 403–414.

WOODROOFE, M. B. and HARDWICK, J. (1990). Sequential allocation for an estimation problem with ethical costs. *Ann. Statist.* **18** 1358–1377.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
1444 MASON HALL
ANN ARBOR, MICHIGAN 48109