The Annals of Statistics
1990, Vol. 18, No. 3, 1358–1377

SEQUENTIAL ALLOCATION FOR AN ESTIMATION PROBLEM WITH ETHICAL COSTS[1]

BY MICHAEL WOODROOFE AND JANIS HARDWICK

The University of Michigan

The problem of designing an experiment to estimate the difference between the means of two normal populations with unit variances is considered, when the cost of drawing a sample from either population may depend on unknown parameters. A quasi-Bayesian approach is adopted in which the mean difference is estimated by its maximum likelihood estimator, but the design (allocation rule) is evaluated in Bayesian, decision-theoretic terms. A three-stage procedure is proposed and its risk evaluated, up to terms which are small compared to the cost of a single observation. This procedure is shown to be optimal to second order for squared error loss.

1. Introduction.

Consider the problem of designing an experiment to estimate the difference between the means of two normal populations, called the treatment and control groups below, when the cost of drawing a sample from either group may depend on unknown parameters. For example, in a clinical study to estimate the effect of a new treatment, when compared to a control, there are ethical costs inherent in assigning a patient to the control if the early results indicate the treatment is superior, and conversely.

To formalize the problem, let $X_1, X_2 \ldots$ and $Y_1, Y_2 \ldots$ denote independent normally distributed random variables for which

$$X_1, X_2, \ldots \sim N(\mu, 1) \quad \text{and} \quad Y_1, Y_2 \ldots \sim N(\nu, 1), \tag{1}$$

where $\mu, \nu \in \mathbb{R}$ are unknown parameters. Here $X_1, X_2 \ldots$ denote potential responses to the control, and $Y_1, Y_2 \ldots$ to the treatment. It is convenient to let

$$\theta = \nu - \mu.$$

Suppose that a total of t observations are to be taken from the two populations and that the objective of the study is to estimate θ with a loss function of the form

$$L_t(\theta, \hat{\theta}) = tK\left[\sqrt{t}\,(\theta - \hat{\theta})\right], \qquad \hat{\theta}, \theta \in \mathbb{R},$$

where K is a nonnegative, nonconstant, symmetric function of polynomial growth for which $K(0) = 0$, $K(z)$ is nondecreasing and right continuous in $z \geq 0$. Suppose further that the cost of drawing a sample from the control, or treatment, group is $a(\theta)$, or $b(\theta)$, where a and b are positive, locally integrable functions on $\mathbb{R}$. Here the loss has been so normalized that the expected loss

Received August 1988; revised August 1989.
[1]Research supported by the NSF, under DMS 84-13452.
AMS 1980 *subject classification*. 62L12.
Key words and phrases. Loss function, sampling costs, integrated risk, invariance, sequential designs, posterior distributions, asymptotic normality.

due to estimation error and the cost of sampling are both of order t. It is convenient to let

$$c(\theta) = b(\theta) - a(\theta), \qquad \theta \in \mathbb{R}, \tag{2}$$

and assume throughout that c is continuously differentiable on $\mathbb{R}$. The loss structure is sufficiently general to include both point and interval estimation. In particular, if $K(z) = 0$ or 1 for $|z| < c$ or $|z| \geq c$, where $c > 0$, then the loss is t times the indicator of the event $\theta \notin (\hat{\theta} - c/\sqrt{t}, \hat{\theta} + c/\sqrt{t})$. If c is any continuously differentiable function, then the conditions imposed on the cost structure are satisfied by $a(\theta) = 1 + c(\theta)^-$ and $b(\theta) = 1 + c(\theta)^+$, where $+$ and $-$ denote positive and negative parts (and "1" represents the cost of performing one replication of the experiment).

The model, the cost structure and the loss structure are all invariant with respect to translations of $X_1, X_2 \dots$ and $Y_1, Y_2 \dots$ by a common constant. The invariance of the loss has been criticized, since it appears to disallow having higher losses for $\theta > 0$ (or large $|\theta|$) than for $\theta < 0$ (or small $|\theta|$). To some extent this criticism may be ameliorated by reparameterizing the problem. See Remark 2 below for the details. The invariance is implicitly used in the construction of the estimator and explicitly used in the consideration of designs, below. Convenient choices of the maximal invariants are $X_i - X_1$, $i = 2, 3 \dots$, and $Y_j - X_1$, $j = 1, 2 \dots$. Of course, their distributions depend only on θ. Let

$$\mathscr{D}_{m,n} = \sigma\{X_2 - X_1, \dots, X_m - X_1, Y_1 - X_1, \dots, Y_n - X_1\},$$

for $m, n = 1, 2 \dots$.

For fixed $t \geq 3$, a (translation-invariant) sequential design δ is (defined to be) a sequence $\delta_1, \dots, \delta_t$ of indicator variables, taking the values of 0 and 1 (for control and treatment), for which $\delta_1 = 0$, $\delta_2 = 1$ and δ_{k+1} is $\mathscr{D}_{m_k, n_k}$-measurable for all $k = 2, \dots, t-1$, where $n_k = \delta_1 + \cdots + \delta_k$, and $m_k = k - n_k$ for $k = 1, \dots, t$. To avoid second-order subscripts, m_t and n_t are denoted by M and N below, and the dependence of M and N on δ is suppressed in the notation.

If $m, n \geq 1$, then the likelihood function, given $\mathscr{D}_{m,n}$, is

$$\mathscr{L}_{mn}(\theta) = \exp\left\{-\frac{1}{2}\left(\frac{mn}{m+n}\right)(\theta - \hat{\theta}_{mn})^2\right\}, \qquad \theta \in \mathbb{R},$$

where

$$\hat{\theta}_{mn} = \overline{Y}_n - \overline{X}_m$$

and $\overline{X}_m$ and $\overline{Y}_n$ denote the sample means of $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$. Then $\hat{\theta}_{mn}$ is the maximum likelihood estimator. Moreover, if δ is any sequential design, then m_k and n_k may be substituted for m and n, since the likelihood function is unaffected by the use of a sequential design.

A quasi-Bayesian, decision-theoretic approach to the problem is adopted in which θ is estimated by its maximum likelihood estimator, but the design is evaluated in Bayesian, decision-theoretic terms. Thus, the risk function of a sequential design δ is defined to be

$$r_t(\delta;\theta) = E_\theta\{tK[\sqrt{t}(\hat{\theta}_{MN} - \theta)] + b(\theta)N + a(\theta)M\}, \tag{3}$$

where E_θ denotes expectation in the model (1) with $\nu - \mu = \theta$, say $\mu = 0$ and $\nu = \theta$; and the problem is to find a design for which the risk is small.

To anticipate the nature of the solution, consider the fixed sample size design δ^f in which M and N are constants. Then $\hat{\theta}_{MN} - \theta$ has a normal distribution with mean 0 and variance $1/M + 1/N$, so that

$$E_\theta\{K[\sqrt{t}(\hat{\theta}_{MN} - \theta)]\} = \kappa(\sigma_t),$$

where

$$\sigma_t^2 = \frac{t}{M} + \frac{t}{N} = \frac{t^2}{MN}, \tag{4}$$

$$\kappa(\tau) = \int_{\mathbb{R}} K(\tau z)\Phi(dz), \qquad \tau > 0, \tag{5}$$

and Φ denotes the standard normal distribution. Since $Ma(\theta) + Nb(\theta) = Nc(\theta) + ta(\theta)$, it follows easily that

$$r_t(\delta^f;\theta) = t\psi[q_t, c(\theta)] + ta(\theta),$$

where

$$q_t = \frac{N}{t} \quad \text{and} \quad \psi(p,c) = \kappa\left[\sqrt{\left(\frac{1}{p} + \frac{1}{1-p}\right)}\right] + cp, \tag{6}$$

for $0 < p < 1$ and $c \in \mathbb{R}$. By Proposition 2 below, $\psi(p,c)$ attains its minimum at a unique point $p = q(c)$ for each $c \in \mathbb{R}$. Let

$$\varphi(c) = \psi[q(c), c] = \inf_p \psi(p,c), \qquad c \in \mathbb{R}. \tag{7}$$

Then, for all θ, $r_t(\delta^f;\theta) \geq t\varphi \circ c(\theta) + ta(\theta)$ with equality iff $N = tq \circ c(\theta)$.

The above derivation suggests an adaptive procedure in which $q \circ c(\theta)$ is estimated sequentially as data accumulate. A three-stage version of this procedure is studied in Section 4.

If ξ is a (prior) density on $\mathbb{R}$ for which $\int_{\mathbb{R}} (a+b)\xi\, d\theta < \infty$, then the integrated risk of a design δ with respect to ξ is (defined to be)

$$\bar{r}_t(\delta;\xi) = \int_{\mathbb{R}} r_t(\delta;\theta)\xi(\theta)\, d\theta.$$

Then

$$\bar{r}_t(\delta;\xi) = E^\xi\{tK[\sqrt{t}(\Theta - \hat{\theta}_{MN})] + b(\Theta)N + a(\Theta)M\},$$

where E^ξ denotes expectation in the Bayesian model in which: There is a random variable Θ with density ξ; and (1) holds conditionally with $\mu = 0$ and $\nu = \theta$, given $\Theta = \theta$, for all $\theta \in \mathbb{R}$. The main results show that the regret, $\bar{r}_t(\delta^t;\xi) - t\int_{\mathbb{R}}[\varphi \circ c(\theta) + a(\theta)]\xi(\theta)\, d\theta$, of the three-stage procedures δ^t, $t \geq 3$, remains bounded as $t \to \infty$ for a large class of ξ. Moreover, it is shown that these designs are asymptotically second-order efficient for squared error loss.

This paper is similar in spirit to Robbins and Siegmund (1974) and Louis (1975) in that optimal allocation rules are sought, although the focus here is on estimation instead of testing. For recent work on allocation and estimation, see Shapiro (1983). At a technical level, some ideas are adapted from recent work on sequential estimation—notably, Hall (1981) and Woodroofe (1985).

2. An expression for the posterior risk. Let $\mathscr{H}$ be the class of all Borel measurable functions h of polynomial growth; and let

$$\Phi h = \int h \, d\Phi$$

and

$$Uh(z) = e^{z^2/2} \int_z^\infty [h(y) - \Phi h] e^{-y^2/2} \, dy,$$

for all $z \in \mathbb{R}$ and $h \in \mathscr{H}$. Then U is easily seen to be a linear operator from $\mathscr{H}$ back into $\mathscr{H}$. See Stein (1986), Chapter 2. For example, if $h_i(z) = z^i$ for $z \in \mathbb{R}$ and $i = 1, 2$, then $Uh_1(z) = 1$ and $Uh_2(z) = z$ for all z. The composition of U with itself is denoted by $U^2 = U \circ U$.

For $p = 0, 1, 2, \ldots$, let $\mathscr{H}_p$ denote the collection of all $h \in \mathscr{H}$ for which $|h(z)| \leq 1 + |z|^p$ for all $z \in \mathbb{R}$.

LEMMA 1. *There are finite positive constants* $J_0, J_1 \ldots$ *for which* $U(\mathscr{H}_p) \subseteq J_p \mathscr{H}_{p'}$, *where* $p' = \max\{0, p - 1\}$, *for all* $p = 0, 1, 2 \ldots$.

PROOF. For $p = 0$, the assertion is proved by Stein (1986), pages 27 and 28; and a similar proof works for $p \geq 1$. $\square$

If ξ is any prior density and δ is any sequential design, then the posterior density, given $\mathscr{D}_{M,N}$, of

$$Z_t = \sqrt{\left(\frac{MN}{t}\right)} (\Theta - \hat{\theta}_{MN})$$

is

$$\zeta_t(z) = \exp\{-\tfrac{1}{2}z^2\} f_t(z),$$

where

$$f_t(z) = \frac{1}{\chi_t} \xi\left(\hat{\theta}_{MN} + \frac{\sigma_t z}{\sqrt{t}}\right), \qquad z \in \mathbb{R},$$

χ_t is a normalizing constant, and σ_t is as in (4).

Below, Ξ_1 denotes the class of all absolutely continuous ξ for which $\int_{\mathbb{R}}(a + b)\xi \, d\theta < \infty$ and $\int_{\mathbb{R}} |\xi'| \, d\theta < \infty$, where $'$ denotes derivative; and Ξ_2 denotes the class of all $\xi \in \Xi_1$ for which ξ' is absolutely continuous and $\int_{\mathbb{R}} |\xi''| \, d\theta < \infty$.

PROPOSITION 1. *For any design, any $\xi \in \Xi_1$ and any $h \in \mathscr{H}$,*

$$E^{\xi}\{h(Z_t)|\mathscr{D}_{M,N}\} = \Phi h + \frac{1}{\sqrt{t}}\sigma_t E^{\xi}\left\{Uh(Z_t)\left(\frac{\xi'}{\xi}\right)(\Theta)\middle|\mathscr{D}_{M,N}\right\};$$

moreover,

$$E^{\xi}\{h(Z_t)|\mathscr{D}_{M,N}\} = \Phi h + \frac{1}{t}\sigma_t^2 E^{\xi}\left\{U^2 h(Z_t)\left(\frac{\xi''}{\xi}\right)(\Theta)\middle|\mathscr{D}_{M,N}\right\},$$

if $\xi \in \Xi_2$ and h is a symmetric function.

PROOF. The proof is similar to the proof of Proposition 1 in Woodroofe (1989). It is included for completeness.

There is no loss of generality in supposing that $\Phi h = 0$. Then, since $f_t'(z) = \sigma_t \xi'(\theta)/\chi_t\sqrt{t}$, and θ and z are related by $\theta = \hat{\theta}_{MN} + \sigma_t z/\sqrt{t}$,

$$\begin{aligned} E^{\xi}\{h(Z_t)|\mathscr{D}_{M,N}\} &= \int_{\mathbb{R}} h(z)\, f_t(z)\exp\left(-\frac{1}{2}z^2\right) dz \\ &= \int_{\mathbb{R}} Uh(z)\, f_t'(z)\exp\left(-\frac{1}{2}z^2\right) dz \\ &= \frac{1}{\chi_t\sqrt{t}}\sigma_t \int_{\mathbb{R}} Uh(z)\xi'(\theta)\exp\left(-\frac{1}{2}z^2\right) dz \\ &= \frac{1}{\sqrt{t}}\sigma_t E^{\xi}\left\{Uh(Z_t)\left(\frac{\xi'}{\xi}\right)(\Theta)\middle|\mathscr{D}_{M,N}\right\}, \end{aligned}$$

by a simple integration by parts. If $\xi \in \Xi_2$ and h is symmetric, then f_t is twice differentiable and $\Phi(Uh) = 0$ by symmetry. So the argument may be repeated, with f_t replaced by f_t', and

$$\begin{aligned} E^{\xi}\{h(Z_t)|\mathscr{D}_{M,N}\} &= \int_{\mathbb{R}} U^2 h(z)\, f_t''(z)\exp\left(-\frac{1}{2}z^2\right) dz \\ &= \frac{1}{t}\sigma_t^2 E^{\xi}\left\{U^2 h(Z_t)\left(\frac{\xi''}{\xi}\right)(\Theta)\middle|\mathscr{D}_t\right\}. \end{aligned} \qquad \square$$

For $\tau > 0$, let $K_\tau(z) = K(\tau z)$ for $z \in \mathbb{R}$, so that $\kappa(\tau) = \Phi K_\tau$ for $\tau > 0$ in (5).

COROLLARY 1. *For any design δ and any $\xi \in \Xi_2$, the posterior risk is*

$$\begin{aligned} R_t(\xi;\delta) &:= E^{\xi}\{tK[\sqrt{t}(\Theta - \hat{\theta}_{MN})] + Nb(\Theta) + Ma(\Theta)|\mathscr{D}_{M,N}\} \\ &= t\psi(q_t, C_t) + \Gamma_t + tE^{\xi}\{a(\Theta)|\mathscr{D}_{M,N}\}, \end{aligned} \tag{8}$$

where

$$\Gamma_t = \sigma_t^2 E^{\xi}\left\{U^2 K_{\sigma_t}(Z_t)\frac{\xi''}{\xi}(\Theta)\middle|\mathscr{D}_{M,N}\right\},$$

$$C_t = E^{\xi}\{c(\Theta)|\mathscr{D}_{M,N}\},$$

q_t and ψ are as in (6), *and* $:=$ *indicates a definition.*

Proof. Since $Nb(\Theta) + Ma(\Theta) = Nc(\Theta) + ta(\Theta)$ and $\sqrt{t}(\Theta - \hat{\theta}_{MN}) = \sigma_t Z_t$,

$$\begin{aligned} E^{\xi}&\{tK\left[\sqrt{t}(\Theta - \hat{\theta}_{MN})\right] + Nb(\Theta) + Ma(\Theta)|\mathscr{D}_{M,N}\} \\ &= E^{\xi}\{tK_{\sigma_t}(Z_t) + Nc(\Theta) + ta(\Theta)|\mathscr{D}_{M,N}\} \\ &= t\kappa(\sigma_t) + \Gamma_t + NC_t + tE^{\xi}\{a(\Theta)|\mathscr{D}_{M,N}\}, \end{aligned}$$

by Proposition 1. □

Of course, the integrated risk may be obtained by integrating R_t with respect to P^{ξ}.

Some properties of the function ψ, defined in (6), are needed below. These are stated here and proved in Section 7.

Proposition 2. *For every $c \in \mathbb{R}$, $\psi(p, c)$ is twice continuously differentiable and strictly convex in $0 < p < 1$ and $\psi(p, c)$ attains its minimum at a unique point $p = q(c)$. Moreover, q defines a twice continuously differentiable function from $\mathbb{R}$ into* $(0, 1)$.

Corollary 2. *The function φ of* (7) *is twice continuously differentiable and concave on $\mathbb{R}$.*

Proof. The differentiability is clear from Proposition 2. The concavity follows from (7), since $\psi(p, c)$ is linear in c for each $0 < p < 1$. □

3. An asymptotic expression for the integrated risk. Asymptotic normality of Z_t is needed in the analysis of the integrated risk.

Proposition 3. *Let $\xi \in \Xi_1$ and let $\delta = \delta^t$, $t \geq 3$, be any sequence of sequential designs for which* $\min(M, N) \to \infty$ *in P^{ξ}-probability as $t \to \infty$. Then $\hat{\theta}_{MN} \to \Theta$ in P^{ξ}-probability and the P^{ξ}-distribution of $(\hat{\theta}_{MN}, Z_t)$ converges weakly to that of (Θ, Z), where Θ has density ξ and Z denotes a standard normal random variable which is independent of Θ. Moreover, all positive powers of $|Z_t|$ are uniformly integrable on* $\min(M, N) \geq \eta t$ *for any $\eta > 0$.*

PROOF. To begin, let g be a continuous function and h be a measurable function with $|h| \le 1$ and $|g| \le 1$. Then, by Proposition 1 and Lemma 1,

$$
\begin{aligned}
&\left|E^{\xi}\{g(\hat{\theta}_{MN})h(Z_t)\} - E^{\xi}\{g(\hat{\theta}_{MN})\}\Phi h\right| \\
&\quad = \left|E^{\xi}\{g(\hat{\theta}_{MN})\{E^{\xi}[h(Z_t)|\mathscr{D}_{M,N}] - \Phi h\}\}\right| \\
&\quad = \left|E^{\xi}\left\{g(\hat{\theta}_{MN})\frac{\sigma_t}{\sqrt{t}}E^{\xi}\left[Uh(Z_t)\frac{\xi'}{\xi}(\Theta)\middle|\mathscr{D}_{M,N}\right]\right\}\right| \\
&\quad \le \int \frac{2J_0}{\min(\sqrt{N},\sqrt{M})}\left|\frac{\xi'}{\xi}(\Theta)\right| dP^{\xi} \to 0,
\end{aligned}
$$

as $t \to \infty$, where J_0 is as in Lemma 1. Letting $g = 1$, shows that Z_t is asymptotically normal. That $\hat{\theta}_{MN} \to \Theta$ in P^{ξ}-probability is an easy consequence of the asymptotic normality; and the limiting joint distribution of $(\hat{\theta}_{MN}, Z_t)$ then follows easily since $E^{\xi}\{g(\hat{\theta}_{MN})\} \to E^{\xi}\{g(\Theta)\}$.

For the uniform integrability, first we observe that

$$
\begin{aligned}
&P_{\theta}\{|\hat{\theta}_{jk} - \theta| > \varepsilon, \exists\, j \ge r \text{ or } k \ge r\} \\
&\qquad \le P_{\theta}\{|\bar{X}_j| > \tfrac{1}{2}\varepsilon, \exists\, j \ge r\} + P_{\theta}\{|\bar{Y}_k - \theta| > \tfrac{1}{2}\varepsilon, \exists\, k \ge r\} \\
&\qquad \le 4\exp\left(-\tfrac{1}{8}r\varepsilon^2\right),
\end{aligned} \tag{9}
$$

for all $\varepsilon > 0$, $r = 1, 2, \dots,$ and $\theta \in \mathbb{R}$, by Bernstein's inequality for martingales, applied to the reverse martingales $\bar{X}_j$, $j \ge 1$, and $\bar{Y}_k - \theta$, $k \ge 1$, for fixed θ. So, for any design δ,

$$
\begin{aligned}
&P_{\theta}\{|Z_t| > z,\, M \ge \eta t \text{ and } N \ge \eta t\} \\
&\qquad \le P_{\theta}\left\{|\hat{\theta}_{jk} - \theta| > \frac{2z}{\sqrt{t}}\ \exists\, j \ge \eta t \text{ or } k \ge \eta t\right\} \\
&\qquad \le 4\exp\left(-\tfrac{1}{2}\eta z^2\right),
\end{aligned} \tag{10}
$$

for all $z > 0$, $\eta > 0$ and $\theta \in \mathbb{R}$, since $\sigma_t \ge 2$ for all t. Of course, P_{θ} may be replaced by P^{ξ} in these inequalities for any density ξ, by simply integrating over θ; and the uniform integratability is an easy consequence of this. □

Let Ξ_0 denote the class of all twice continuously differentiable densities ξ with compact support in $\mathbb{R}$ for which

$$
\int_{\mathbb{R}} \frac{\xi'^2}{\xi}\,d\theta < \infty \quad \text{and} \quad \int_{\mathbb{R}} \left|\frac{\xi''}{\xi}\right|^{\alpha} \xi\,d\theta < \infty, \tag{11}
$$

for some $\alpha > 1$. Further, recall the definition of φ from (7) and define a function $\sigma\colon \mathbb{R} \to (0, \infty)$ by

$$
\sigma^2(\theta) = \frac{1}{q \circ c(\theta)} + \frac{1}{1 - q \circ c(\theta)} = \frac{1}{q \circ c(\theta)[1 - q \circ c(\theta)]},
$$

for $\theta \in \mathbb{R}$, where q and c are as in Proposition 2 and (2).

THEOREM 1. *Let $\xi \in \Xi_0$; and let $\delta = \delta^t$, $t \geq 3$, denote any sequence of sequential designs for which*

$$\sigma_t^2 \to \sigma^2(\Theta) \quad \text{in } P^\xi\text{-probability as } t \to \infty \tag{12}$$

and

$$\lim_{t\to\infty} t^\alpha P^\xi\{\min(M,N) < \eta t\} = 0, \qquad \forall\, \alpha > 0, \exists\, \eta \in (0,1). \tag{13}$$

Then

$$\begin{aligned} &\liminf_{t\to\infty}\left\{\bar{r}_t(\delta;\xi) - t\int_{\mathbb{R}}[\varphi\circ c(\theta) + a(\theta)]\xi(\theta)\,d\theta\right\} \\ &\qquad \geq \tfrac{1}{2}\int_{\mathbb{R}}\{H''(\theta) - \varphi''\circ c(\theta)c'(\theta)^2\sigma^2(\theta)\}\xi(\theta)\,d\theta, \end{aligned} \tag{14}$$

where

$$H(\theta) = \sigma^3(\theta)\kappa'\circ\sigma(\theta), \qquad \theta\in\mathbb{R}.$$

Moreover, the limit exists and there is equality in (14) *if*

$$\lim_t t\int_B\{\psi[q_t, C_t] - \varphi(C_t)\}\,dP^\xi = 0, \tag{15}$$

where ψ is as in (6), *φ is as in* (7) *and B is defined in* (16) *below.*

PROOF. Let $\delta = \delta^t$, $t \geq 3$, denote any sequence of designs; fix a $\xi \in \Xi_0$; let $\underline{\theta}$ and $\bar{\theta}$ denote the minimum and maximum of its support; and let $B = B_t(\delta;\xi)$ be the event

$$B = \{\hat{\theta}_{MN} \in [\underline{\theta} - 1, \bar{\theta} + 1]\} \cap \{\min(M,N) \geq \eta t\}, \tag{16}$$

where η is as in the statement of the theorem. Then B is $\mathscr{D}_{M,N}$-measurable; and

$$\begin{aligned} P^\xi(B') &\leq P\{\min(M,N) < \eta t\} \\ &\qquad + P^\xi\left\{\left|\hat{\theta}_{jk} - \Theta\right| > 1, \exists\, j \geq \eta t \text{ or } k \geq \eta t\right\} = o(t^{-\alpha}), \end{aligned}$$

as $t \to \infty$ for all $\alpha > 0$, by (9) and (13).

Recall the expression for the posterior risk from Corollary 1, and write

$$\begin{aligned} &\bar{r}_t(\delta;\xi) - t\int_{\mathbb{R}}[\varphi\circ c(\theta) + a(\theta)]\xi(\theta)\,d\theta \\ &\quad = t\int_B\{\psi[q_t, C_t] - \varphi(C_t)\}\,dP^\xi \\ &\qquad + t\int_B[\varphi(C_t) - \varphi\circ c(\Theta)]\,dP^\xi + \int_B \Gamma_t\,dP^\xi \\ &\qquad + \int_{B'}\left\{tK\left[\sqrt{t}\left(\Theta - \hat{\theta}_{MN}\right)\right] + c(\Theta)N - t\varphi\circ c(\Theta)\right\}dP^\xi. \end{aligned} \tag{17}$$

The first term on the right side of (17) is nonnegative and approaches 0, by assumption, if (15) holds. The limits of the other terms are computed in Lemmas 3, 4 and 5 below. In these, $\delta = \delta^t$, $t \geq 3$, denotes any sequence of designs for which (12) and (13) hold; ξ denotes a fixed member of Ξ_0; and all limits are taken as $t \to \infty$. The theorem is a direct consequence of these three lemmas. □

LEMMA 2.

$$\lim_t \int_B \left| \sqrt{t}\left[C_t - c(\hat{\theta}_{MN})\right]\right|^p dP^\xi = 0, \qquad \forall p > 0.$$

PROOF. It suffices to prove the lemma for $p \geq 2$, by Holder's inequality. Let $h_t(z) = \sqrt{t}[c(\hat{\theta}_{MN} + \sigma_t z/\sqrt{t}) - c(\hat{\theta}_{MN}) - c'(\hat{\theta}_{MN})\sigma_t z/\sqrt{t}]$ for $z \in \mathbb{R}$, so that

$$\sqrt{t}\left[C_t - c(\hat{\theta}_{MN})\right] = c'(\hat{\theta}_{MN})\sigma_t E^\xi\{Z_t | \mathscr{D}_{M,N}\} + E^\xi\{h_t(Z_t) | \mathscr{D}_{M,N}\},$$

for all $t \geq 1$. Next let J denote an upper bound for $|c'|$ on $[\underline{\theta} - 1, \bar{\theta} + 1]$. Then $h_t(Z_t)$ converges to 0 in probability, by Proposition 3 and the assumed differentiability of c; and $|h_t(Z_t)| \leq J|Z_t|$ on B. So

$$\lim_t \int_B |h_t(Z_t)|^p dP^\xi = 0,$$

by Proposition 3 and the form of B. Next, by Propositions 1 and 3 and the form of B, there is a constant J' for which

$$\int_B \left|\sigma_t c'(\hat{\theta}_{MN}) E^\xi\{Z_t | \mathscr{D}_{M,N}\}\right|^p dP^\xi \leq \int_B J' \left| E^\xi\{Z_t | \mathscr{D}_t\}\right|^p dP^\xi \to 0,$$

since the integrand approaches 0 in P^ξ-probability, by Proposition 1, and is bounded by $J'E^\xi[|Z_t|^p | \mathscr{D}_{M,N}]$, which is uniformly integrable, by Proposition 3. □

LEMMA 3.

$$\lim_t t\int_B [\varphi(C_t) - \varphi \circ c(\Theta)]\, dP^\xi = -\tfrac{1}{2}\int_{\mathbb{R}} \varphi'' \circ c(\theta) c'(\theta)^2 \sigma^2(\theta) \xi(\theta)\, d\theta.$$

PROOF. By a simple Taylor series expansion,

$$\varphi[c(\Theta)] - \varphi(C_t) = \varphi'(C_t)[c(\Theta) - C_t] + \tfrac{1}{2}\varphi''(C_t^*)[c(\Theta) - C_t]^2,$$

where C_t^* denotes an intermediate point. So, since $B \in \mathscr{D}_{M,N}$ and $E^\xi\{c(\Theta) - C_t | \mathscr{D}_{M,N}\} = 0$,

$$t\int_B [\varphi \circ c(\Theta) - \varphi(C_t)]\, dP^\xi = \int_B \tfrac{1}{2} t \varphi''(C_t^*)[c(\Theta) - C_t]^2\, dP^\xi.$$

Now, since $\Theta - \hat{\theta}_{MN} = \sigma_t Z_t/\sqrt{t}$ and $\sigma_t \to \sigma(\Theta)$ in P^ξ-probability, the integrand on the right converges in distribution to $\frac{1}{2}\varphi'' \circ c(\Theta)c'(\Theta)^2\sigma^2(\Theta)Z^2$, where Z denotes a standard normal random variable which is independent of Θ, by Lemma 2, Proposition 3 and another simple Taylor series expansion. Moreover, it is dominated on B by a constant multiple of $t|\hat{\theta}_{MN} - \Theta|^2 + t|C_t - c(\hat{\theta}_{MN})|^2$, which is uniformly integrable on B, by Proposition 3 and Lemma 2. So

$$\lim_t \int_B \tfrac{1}{2} t\varphi''(C_t^*)\left[c(\Theta) - C_t\right]^2 dP^\xi = \tfrac{1}{2}\int_{\mathbb{R}} \varphi'' \circ c(\theta) c'(\theta)^2 \sigma^2(\theta)\xi(\theta)\, d\theta. \quad \square$$

LEMMA 4.

$$\lim_t \int_B \Gamma_t \, dP^\xi = \tfrac{1}{2}\int_{\mathbb{R}} H''(\theta)\xi(\theta)\, d\theta.$$

PROOF. Let $g_t = U \cdot_{\sigma_t}$. Then, since B is $\mathscr{D}_{M,N}$-measurable,

$$\int_B \Gamma_t \, dP^\xi = \int_B g_t(Z_t)\sigma_t^2 \frac{\xi''}{\xi}(\Theta)\, dP^\xi = \gamma_{1,t} + \gamma_{2,t},$$

where

$$\gamma_{1,t} = \int_B E^\xi\{g_t(Z_t)|\mathscr{D}_{M,N}\}\sigma_t^2 E^\xi\left\{\frac{\xi''}{\xi}(\Theta)\middle|\mathscr{D}_{M,N}\right\} dP^\xi$$

and

$$\gamma_{2,t} = \int_B g_t(Z_t)\sigma_t^2\left\{\frac{\xi''}{\xi}(\Theta) - E^\xi\left[\frac{\xi''}{\xi}(\Theta)\middle|\mathscr{D}_{M,N}\right]\right\} dP^\xi.$$

Let α be as in (11) and β be the conjugate value $(1/\alpha + 1/\beta = 1)$. Then

$$|\gamma_{2,t}| \le \left\{\int_B \sigma_t^{2\alpha}\left|\frac{\xi''}{\xi}(\Theta) - E^\xi\left(\frac{\xi''}{\xi}(\Theta)\middle|\mathscr{D}_{M,N}\right)\right|^\alpha dP^\xi\right\}^{1/\alpha}\left\{\int_B |g_t(Z_t)|^\beta \, dP^\xi\right\}^{1/\beta}.$$

By Lemma 1 and the conditions imposed on K, there are positive constants J and p for which $|g_t(z)| \le J(1 + |z|^p)$ for all $z \in \mathbb{R}$ on B for all sufficiently large t. So, the second factor is bounded, by Proposition 3. Moreover, since σ_t^2 is bounded on B, the first factor approaches 0 as $t \to \infty$, by the martingale convergence theorem. So $\gamma_{2,t} \to 0$.

For the analysis of $\gamma_{1,t}$, it follows from Lemma 1, Proposition 1 and the consistency of σ_t, that w.p.1 (P^ξ),

$$\lim_t E^\xi\{g_t(Z_t)|\mathscr{D}_{M,N}\} = \Phi\left[U^2 K_{\sigma(\Theta)}\right];$$

and

$$\Phi\left[U^2 K_{\sigma(\theta)}\right] = \tfrac{1}{2}\int_{\mathbb{R}} (z^2 - 1)K[\sigma(\theta)z]\varphi(z)\, dz = \tfrac{1}{2}\sigma(\theta)\kappa'[\sigma(\theta)],$$

for all θ by (fairly) routine calculations. It is clear from the analysis of $\gamma_{2,t}$

that $E^{\xi}\{g_t(Z_t)|\mathscr{D}_{M,N}\}\sigma_t^2 E^{\xi}\{(\xi''/\xi)(\Theta)|\mathscr{D}_{M,N}\}$, $t \geq 3$, are uniformly integrable. So

$$\begin{aligned}\lim_t \gamma_{1,t} &= \int_{\mathbb{R}} \tfrac{1}{2}\sigma(\theta)\kappa'[\sigma(\theta)]\sigma^2(\theta)\frac{\xi''}{\xi}(\theta)\xi(\theta)\,d\theta \\ &= \tfrac{1}{2}\int_{\mathbb{R}} H(\theta)\xi''(\theta)\,d\theta \\ &= \tfrac{1}{2}\int_{\mathbb{R}} H''(\theta)\xi(\theta)\,d\theta. \qquad \square\end{aligned}$$

Lemma 5.

$$\lim_t \int_{B'} \left\{tK\left[\sqrt{t}\left(\Theta - \hat{\theta}_{MN}\right)\right] + c(\Theta)N - t\varphi\circ c(\Theta)\right\} dP^{\xi} = 0.$$

Proof. Since ξ has compact support, there is a constant J for which $|c(\theta)| + |\varphi \circ c(\theta)| \leq J$ for all θ in the support of ξ. Moreover, $K(z)^2 \leq J[1 + |z|^{2p}]$ for all z for some $p \geq 1$ and $J \geq 0$; and $E^{\xi}\{|\Theta - \hat{\theta}_{MN}|^{2p}\}$ is bounded, by a simple application of (9). So there is a J'' for which

$$\begin{aligned}&\int_{B'} \left|tK\left[\sqrt{t}\left(\Theta - \hat{\theta}_{MN}\right)\right] + Nc(\Theta) - t\varphi(\Theta)\right| dP^{\xi} \\ &\quad\leq \sqrt{E^{\xi}\left\{\left|tK\left[\sqrt{t}\left(\Theta - \hat{\theta}_{MN}\right)\right] + Nc(\Theta) - t\varphi(\Theta)\right|^2\right\}}\sqrt{P^{\xi}(B')} \\ &\quad\leq J''t^{1+p}\sqrt{P^{\xi}(B')} \to 0. \qquad \square\end{aligned}$$

4. A three-stage procedure. An ad hoc procedure which takes observations in three batches is investigated next. Such procedures were introduced by Hall (1981) for sequential estimation and have been studied by Woodroofe (1988) and Meslem (1987) in related contexts.

Let $r = r_t$, $t \geq 5$, and $s = s_t$, $t \geq 5$, be any two sequences of positive integers for which $2r + 2s < t$ for all $t \geq 5$,

$$\lim_{t\to\infty} \frac{r+s}{t} = 0 \quad \text{and} \quad \lim_{t\to\infty} \frac{t\log t}{s\sqrt{r}} = 0. \tag{18}$$

In the first batch, r observations are taken from each population. Then $\hat{\theta}_{rr}$ may be computed. Let

$$n = \min\left\{t - r - 2s, \max\left\{r, \left[q\circ c\left(\hat{\theta}_{rr}\right)(t - 2s)\right]\right\}\right\} \quad \text{and} \quad m = (t - 2s) - n,$$

where $[\cdot]$ denotes the greatest integer function. In the second batch, $t - 2(r + s)$ additional observations are taken of which $n - r$ are from the treatment group and $m - r$ are from the control group. The $\hat{\theta}_{mn}$ may be computed. Let

$$N = \min\left\{t - m, \max\left\{n, \left[q\circ c\left(\hat{\theta}_{mn}\right)t\right]\right\}\right\} \quad \text{and} \quad M = t - N.$$

The remaining $2s$ observations are taken in the final batch. Of these $N - n$ are from the treatment group, and $M - m$ from the control.

THEOREM 2. *If $\delta = \delta^t$, $t \geq 5$, is the three-stage procedure, then the limit exists and there is equality in* (14) *for any $\xi \in \Xi_0$.*

PROOF. Fix a $\xi \in \Xi_0$. Then it suffices to show that (12), (13) and (15) are satisfied by δ.

Since $r \to \infty$, $\hat{\theta}_{mn} \to \Theta$ in P^{ξ}-probability, by Proposition 3; and, therefore, $\sigma_t^2 \to \sigma^2(\Theta)$ in P^{ξ}-probability. Moreover, there is an η for which $\eta < q \circ c(\theta) < 1 - \eta$ and $|(q \circ c)'(\theta)| \leq 1/\eta$ for all $\theta \in [\underline{\theta} - 1, \bar{\theta} + 1]$, where $\underline{\theta}$ and $\bar{\theta}$ denote the minimum and maximum of support(ξ). With this choice of η and large t,

$$P^{\xi}\{\min(M, N) < \eta t\} \leq P^{\xi}\{|\hat{\theta}_{rr} - \Theta| \geq 1\} \leq 4\exp\left(-\tfrac{1}{8}r\right) = o(t^{-\alpha}),$$

as $t \to \infty$ for all $\alpha > 0$, by (9) and (18). So, (12) and (13) are satisfied.

That (15) is satisfied is the content of Lemma 8 below. In Lemmas 6, 7 and 8, δ denotes the three-stage procedure, ξ denotes an arbitrary element of Ξ_0, η is as in the previous paragraph and all limits are taken as $t \to \infty$. □

LEMMA 6. *For $t \geq 5$, let $A = A_t$ denote the event*

$$A = \left\{\hat{\theta}_{rr} \in \left[\underline{\theta} - 1, \bar{\theta} + 1\right]\right\} \cap \left\{\hat{\theta}_{mn} \in \left[\underline{\theta} - 1, \bar{\theta} + 1\right]\right\} \cap \left\{\left|\hat{\theta}_{mn} - \hat{\theta}_{rr}\right| \leq \frac{\eta^2 s}{t}\right\}.$$

Then

$$P^{\xi}(A') = o(t^{-\alpha}), \qquad \forall\, \alpha > 0,$$

and

$$n = \left[(t - 2s)q \circ c(\hat{\theta}_{rr})\right] \quad \textit{and} \quad N = \left[tq \circ c(\hat{\theta}_{mn})\right], \quad \textit{on } A,$$

for all sufficiently large t.

PROOF. The first assertion follows directly from (9) and (18); and the second is clear, since $\eta < q \circ c(\hat{\theta}_{rr}) < 1 - \eta$ on A and $(r + s)/t \to 0$. For the third assertion, it suffices to show that $[tq \circ c(\hat{\theta}_{mn})] > n$ and $[tq \circ c(\hat{\theta}_{mn})] < t - m$ on A for all large t. Since $|(q \circ c)'(\theta)| \leq 1/\eta$ for all $\underline{\theta} - 1 \leq \theta \leq \bar{\theta} + 1$,

$$\begin{aligned} tq \circ c(\hat{\theta}_{mn}) - n &\geq 2sq \circ c(\hat{\theta}_{rr}) + t\{q \circ c(\hat{\theta}_{mn}) - q \circ c(\hat{\theta}_{rr})\} \\ &\geq 2s\eta - \frac{t}{\eta}\left|\hat{\theta}_{mn} - \hat{\theta}_{rr}\right| \geq \eta s, \end{aligned} \tag{19}$$

on A for all large t. The first relation follows easily; and the second may be established similarly. □

LEMMA 7.

$$\lim_t t\int_A \left(\hat{\theta}_{MN} - \hat{\theta}_{mn}\right)^2 dP^{\xi} = 0.$$

PROOF. Let $\bar{Y}_{n,N}$ denote the average of Y_i for $i = n+1, \ldots, N$. Then

$$\bar{Y}_N - \bar{Y}_n = \left(\frac{N-n}{N}\right)\{(\bar{Y}_{n,N} - \theta) - (\bar{Y}_n - \theta)\}.$$

So, since N is $\mathscr{D}_{m,n}$-measurable and $N \geq n \geq \eta t$ on A for all sufficiently large t, there is a constant J for which

$$\begin{aligned} E_\theta\left\{(\bar{Y}_N - \bar{Y}_n)^2 \middle| X_1, \mathscr{D}_{m,n}\right\} &= \left(\frac{N-n}{N^2}\right) + \left(\frac{N-n}{N}\right)^2 (\bar{Y}_n - \theta)^2 \\ &\leq J\left\{\frac{s}{t^2} + \frac{s^2}{t^4} n^2 (\bar{Y}_n - \theta)^2\right\} \end{aligned}$$

w.p.1 on A for all sufficiently large t. So, by the $\mathscr{D}_{m,n}$-measurability of A, Wald's lemma for second moments and (18),

$$t\int_A (\bar{Y}_N - \bar{Y}_n)^2 \, dP^\xi \leq Jt\int_A \left\{\frac{s}{t^2} + \frac{s^2}{t^4} n^2 (\bar{Y}_n - \theta)^2\right\} dP^\xi \leq Jt\left\{\frac{s}{t^2} + \frac{s^2}{t^3}\right\} \to 0,$$

as $t \to \infty$. A dual argument shows that $\int_A (\bar{X}_M - \bar{X}_m)^2 \, dP^\xi \to 0$ to complete the proof of the lemma. □

LEMMA 8.

$$\lim_t t\int_B \{\psi[q_t, C_t] - \varphi(C_t)\} \, dP^\xi = 0.$$

PROOF. It suffices to prove the lemma with B replaced by $A \cap B$, since $\psi(q_t, C_t)$ and $\varphi(C_t)$, $t \geq 5$, are uniformly bounded on B and, therefore,

$$t\int_{B-A} [\psi(q_t, C_t) - \varphi(C_t)] \, dP^\xi \leq JtP^\xi(A') \to 0,$$

as $t \to \infty$ for some constant J.

Since $\partial\psi(p,c)/\partial p$ vanishes at $p = q(c)$ and $|\partial^2\psi(p,c)/\partial p^2|$ is bounded on compact subsets of $(0,1) \times \mathbb{R}$, there is a constant J for which

$$\psi[q_t, C_t] - \varphi(C_t) = \psi[q_t, C_t] - \psi[q(C_t), C_t] \leq J[q_t - q(C_t)]^2$$

w.p.1 on $A \cap B$ for all large t. Next, since $|N - tq \circ c(\hat{\theta}_{mn})| \leq 1$, on A for large t, by Lemma 6, there is a constant J' for which

$$\begin{aligned} |q_t - q(C_t)| &\leq \frac{1}{t} + \left|q \circ c(\hat{\theta}_{mn}) - q \circ c(\hat{\theta}_{MN})\right| + \left|q \circ c(\hat{\theta}_{MN}) - q(C_t)\right| \\ &\leq J'\left\{\frac{1}{t} + \left|\hat{\theta}_{mn} - \hat{\theta}_{MN}\right| + \left|c(\hat{\theta}_{MN}) - C_t\right|\right\} \end{aligned}$$

w.p.1 on A for all sufficiently large t. So, by Lemmas 2 and 7,

$$\lim_t t\int_{A\cap B} |q_t - q(C_t)|^2 \, dP^\xi = 0. \qquad \square$$

5. Global optimality. Theorems 1 and 2 assert that the three-stage procedure is asymptotically optimal (to second order) in the class of all procedures which satisfy conditions (12) and (13), but leave open the possibility that it is asymptotically suboptimal in the class of all procedures. In this section, the three-stage procedure is shown to be asymptotically optimal in the class of all procedures for the special case of squared error loss, $K(z) = z^2$ for $z \in \mathbb{R}$, in which case $U^2K(z) = 1$ for $z \in \mathbb{R}$.

A preliminary result is needed.

PROPOSITION 4. *Let ξ be a continuous density which has compact support, is positive on the interior of its support and is monotone near the endpoints of its support. Then there is an $\varepsilon = \varepsilon(\xi) > 0$ for which*

$$P^{\xi}\{|Z_t| \le \varepsilon | \mathscr{D}_{M,N}\} \le \tfrac{1}{2} \tag{20}$$

*w.p.*1 (P^{ξ}) *for all designs δ and all $t = 3, 4 \dots$.*

PROOF. Let ξ be as in the statement of the proposition; and let δ be any design. Further, let $i^2 = MN/t$, the information in the sample; observe that $i^2 \ge 1/2$; and write $\hat{\theta}$ for $\hat{\theta}_{MN}$. Then

$$P^{\xi}\{|Z_t| \le \varepsilon | \mathscr{D}_{M,N}\} = \frac{\int_{-\varepsilon}^{\varepsilon} \xi\left(\hat{\theta} + \dfrac{z}{i}\right) e^{-z^2/2}\, dz}{\int_{\mathbb{R}} \xi\left(\hat{\theta} + \dfrac{w}{i}\right) e^{-w^2/2}\, dw}, \tag{21}$$

for all $\varepsilon > 0$.

To identify ε, let $[\underline{\theta}, \bar{\theta}]$ denote the support of ξ. Then ξ must be nondecreasing near $\underline{\theta}$ and nonincreasing near $\bar{\theta}$. So there are $\underline{\theta} < \theta' < \theta'' < \bar{\theta}$ for which ξ is nondecreasing on $[\underline{\theta}, \theta']$ and nonincreasing on $[\theta'', \bar{\theta}]$. Let $\Delta = \min\{(\theta'' - \theta')/4, 1\}$; let J be so large that $1/J \le \xi(\theta) \le J$ for $(\underline{\theta} + \theta')/2 \le \theta \le (\theta'' + \bar{\theta})/2$; and let $\varepsilon > 0$ such that $36\varepsilon < \min\{(\theta' - \underline{\theta}), (\bar{\theta} - \theta''), \Delta/J^2\}$.

If $\hat{\theta} \le \underline{\theta} - \varepsilon/i$, then the right side of (21) is 0. If $\underline{\theta} - \varepsilon/i < \hat{\theta} < \theta' - 18\varepsilon$, then the numerator is at most $2\varepsilon\xi(\hat{\theta} + \varepsilon/i)$, and the denominator is at least $(8\varepsilon/\sqrt{e})\xi(\hat{\theta} + \varepsilon/i)$ (since $\int_{\mathbb{R}} \ge \int_{\varepsilon}^{9\varepsilon}$), so that the ratio is at most $\sqrt{e}/4 \le 1/2$. If $\theta' - 18\varepsilon \le \hat{\theta} \le (\theta' + \theta'')/2$, then the numerator is at most $2J\varepsilon$, and the denominator at least $\Delta/J\sqrt{e}$ (since $\int_{\mathbb{R}} \ge \int_0^{\Delta}$), so that the ratio is at most $2\varepsilon J^2\sqrt{e}/\Delta \le 1/2$. The remaining three cases may be handled similarly to complete the proof. □

In the final theorem, Ξ_{00} denotes the class of $\xi \in \Xi_0$ for which ξ is positive on the interior of its support, monotone near the endpoints of its support and

$$\inf_{\theta} \frac{\xi''}{\xi}(\theta) > -\infty, \tag{22}$$

where $0/0$ is to be interpreted as 0. Then Ξ_{00} contains all densities of the form

$\xi(\theta) = \{(\theta - \underline{\theta})^+(\bar{\theta} - \theta)^+\}^\alpha \zeta(\theta)$, where $-\infty < \underline{\theta} < \bar{\theta} < \infty$, $\alpha > 2$ and ζ is positive and twice continuously differentiable on $\mathbb{R}$.

THEOREM 3. *Suppose that $K(z) = z^2$ for all $z \in \mathbb{R}$; and let δ^0 denote the three-stage procedure. If $\xi \in \Xi_{00}$, then*

$$\lim_{t\to\infty} \lim_{\gamma} \{\bar{r}_t(\gamma; \xi) - \bar{r}_t(\delta^0; \xi)\} = 0,$$

where the infimum extends over all sequential designs γ.

PROOF: Since the infimum is nonpositive for all t, it suffices to show that its limit inferior is nonnegative as $t \to \infty$. For a given $\xi \in \Xi_{00}$, it is easily seen that there is an optimal procedure which may, in principle, be computed from backward induction. See, for example, Whittle (1982), Chapter 11 and/or Haggstrom (1966), Theorem 4.1. So, it suffices to show that conditions (12) and (13) are satisfied by the optimal procedure. Condition (13) is established first.

Let δ denote the optimal procedure and let $\mathbb{R}_t$ denote the posterior risk, using the optimal procedure. Then the principle of optimality (dynamic programming) requires that for all $m \geq 1$ and $n \geq 1$,

$$E^\xi(R_t|\mathscr{D}_{m,n}) = \operatorname*{ess\,inf}_{\delta'} E^\xi\left(R_t'|\mathscr{D}_{m,n}\right) \tag{23}$$

a.e. on $\{m_k = m,\ n_k = n\}$ for all $k = 3, \dots, t-1$, where $n_k = \delta_1 + \cdots + \delta_k$, $m_k = k - n_k$, the essential infimum extends over all designs δ' for which $\delta_j' = \delta_j$ for $j = 1, \dots, k$, and R_t' denotes the posterior risk of the design δ'.

By Proposition 4, there is an $\varepsilon > 0$ for which $E^\xi(Z_t^2|\mathscr{D}_{M,N}) \geq 12\varepsilon$ w.p.1 for all t, using the optimal procedure. Let J_0 be an upper bound for $|c(\theta)|$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$, the support of ξ; and let $0 < \eta \leq \min\{\varepsilon/2J_0, 1/8\}$. It is shown below that $P^\xi\{M < \eta t\} = 0$ for all sufficiently large t. Condition (13) follows directly from this result and its dual (in which M is replaced by N). Let

$$\tau = \inf\{k \geq 1 \colon m_k \leq \eta t - (t - k - 1) \text{ and } \delta_{k+1} = 1 \text{ or } k = t\}.$$

Then τ is a stopping time with respect to $\mathscr{D}_{m_k, n_k}$, $k = 1, \dots, t$; and $\{M < \eta t\} \subseteq \{\tau < t\}$ for all $t > 1/\eta$. In fact, if k is the largest j for which $\delta_j = 1$, then $\tau \leq k - 1$ on $\{M > \eta t\}$ for all $t > 1/\eta$. Moreover, it follows from the minimality of τ that $m_\tau \geq \eta t - (t - \tau)$ for all $t > 1/\eta$.

Next let δ' be the design for which $\delta_k' = \delta_k$ for all $k \leq \tau$ and $\delta_k' = 0$ for all $k = \tau + 1, \dots, t$. Then $M' \geq M + 1$ and $\eta t \leq M' \leq \eta t + 1$ on $\{\tau < t\}$ for all $t > 1/\eta$. Now

$$R_t - R_t' = \mathrm{I} + \mathrm{II} + \mathrm{III},$$

where

$$\mathrm{I} = t\left(\sigma_t^2 - \sigma_t'^2\right) E^\xi\left[Z_t^2|\mathscr{D}_{M,N}\right],$$

$$\mathrm{II} = t\sigma_t'^2\left\{E^\xi\left(Z_t^2|\mathscr{D}_{M,N}\right) - E^\xi\left(Z_t'^2|\mathscr{D}_{M',N'}\right)\right\},$$

and

$$\mathrm{III} = NC_t - N'C_t'.$$

Let $t \geq 4/\eta$; and let m and n be positive integers for which $m + n < t$. Then, a.e. on $\{m_\tau = m\} \cap \{n_\tau = n\}$,

$$\sigma_t^2 - \sigma_t'^2 = \frac{t^2}{MNM'N'}(N' - M)(M' - M)$$

and

$$\mathrm{I} \geq 12\varepsilon \frac{t^3}{MNM'N'}(N' - M)(M' - M) \geq \frac{3\varepsilon t}{\eta}\left(\frac{M' - M}{M}\right).$$

Next

$$\mathrm{II} = \mathrm{II}_1 + \mathrm{II}_2,$$

where

$$\mathrm{II}_1 = \sigma_t'^2\left\{\left(\sigma_t^2 - \sigma_t'^2\right)E^{\xi}\left[\frac{\xi''}{\xi}(\Theta)\middle|\mathscr{D}_{M,N}\right]\right\}$$

and

$$\mathrm{II}_2 = \sigma_t'^4\left\{E^{\xi}\left[\frac{\xi''}{\xi}(\Theta)\middle|\mathscr{D}_{M,N}\right] - E^{\xi}\left[\frac{\xi''}{\xi}(\Theta)\middle|D_{M',N'}\right]\right\}.$$

Here

$$E^{\xi}\left(\mathrm{II}_2|\mathscr{D}_{m,n}\right) = 0,$$

a.e. on $\{m_\tau = m\} \cap \{n_\tau = n\}$, since σ_t' is measurable with respect to $\mathscr{D}_{m_\tau,n_\tau}$ and $\mathscr{D}_{m_\tau,n_\tau} \subseteq \mathscr{D}_{M,N} \cap \mathscr{D}_{M',N'}$. Moreover, there is a constant J for which $\sigma_t'^2 \leq J$ w.p.1 and $\xi''(\theta) \geq -J\xi(\theta)$ for all θ [see (22)]. So

$$\mathrm{II}_1 \geq -J^2\frac{t^2}{NM'N'}(N' - M)\left(\frac{M' - M}{M}\right) \geq -\frac{2J^2}{\eta}\left(\frac{M' - M}{M}\right).$$

Similarly,

$$\mathrm{III} = \mathrm{III}_1 + \mathrm{III}_2,$$

where

$$\mathrm{III}_1 = C_t(N - N') \geq -J_0\eta t\left(\frac{M' - M}{M}\right)$$

and

$$E^{\xi}\left(\mathrm{III}_2|\mathscr{D}_{m,n}\right) = N'\left\{E^{\xi}\left(C_t|\mathscr{D}_{m,n}\right) - E^{\xi}\left(C_t'|\mathscr{D}_{m,n}\right)\right\} = 0$$

a.e. on $\{m_\tau = m\} \cap \{n_\tau = n\}$. So, for all $t \ge \min\{4/\eta, 2J^2/\varepsilon\}$,

$$\begin{aligned} E^\xi[R_t|\mathscr{D}_{m,n}] - E^\xi\big[R'_t|\mathscr{D}_{m,n}\big] &= E^\xi(\mathrm{I} + \mathrm{II}_1 + \mathrm{III}_1|\mathscr{D}_{m,n}) \\ &\ge \left\{\frac{3\varepsilon t}{\eta} - \frac{2J^2}{\eta} - J_0\eta t\right\} E^\xi\left[\left(\frac{M'-M}{M}\right)\middle|\mathscr{D}_{m,n}\right] \\ &\ge \frac{\varepsilon t}{\eta} E^\xi\left[\left(\frac{M'-M}{M}\right)\middle|D_{m,n}\right] > 0 \end{aligned}$$

a.e. on $\{m_\tau = m\} \cap \{n_\tau = n\}$ for all m and n for which $m + n < t$. Since m and n were arbitrary integers for which $m + n < t$, it follows easily from (23) that $P^\xi\{\tau < t\} = 0$ for all sufficiently large t. Condition (13) follows.

Condition (12) is now easily established. To see how, write

$$\frac{1}{t}\bar{r}_t(\delta;\xi) - \int_{\mathbb{R}} [\varphi \circ c(\theta) + a(\theta)]\xi(\theta)\, d\theta$$

$$= \int [\psi(q_t, C_t) - \phi(C_t)]\, dP^\xi + \int [\varphi(C_t) - \varphi \circ c(\Theta)]\, dP^\xi + \frac{1}{t}\int \Gamma_t\, dP^\xi.$$

Now $E^\xi(|\Gamma_t|^\alpha)$ remains bounded as $t \to \infty$ for some $\alpha > 1$, by Proposition 1, (11) and (13). Moreover, $C_t \to c(\Theta)$ in P^ξ-probability and φ is bounded on the (compact) image $\mathscr{C} = c([\underline{\theta}, \bar{\theta}])$ of $[\underline{\theta}, \bar{\theta}]$ under c. So the last two terms approach 0 as $t \to \infty$. Next the limit superior of the left side is at most 0, by Theorem 2, since the risk of the optimal procedure is at most the risk of the three-stage procedure, So $\psi(q_t, C_t) - \varphi(C_t)$ approaches 0 in probability; and this requires $q_t \to q \circ c(\Theta)$ in P^ξ-probability, since $\inf\{|\psi(p,b) - \varphi(b)|\colon |p - q(b)| \ge \delta, b \in \mathscr{C}\} > 0$ for every $\delta > 0$. Relation (12) follows immediately. □

6. Remarks.

1. Together, Theorems 2 and 3 assert that one design, the three-stage procedure, is asymptotically second-order optimal with respect to a large class of prior distributions for squared error loss. This may be contrasted with the estimation problem, where second-order efficiency requires knowledge of the prior. A similar phenomenon was observed by Woodroofe (1985) for one-sample stopping problems with squared error loss.

2. If the loss structure is changed to $L_t(\theta, \hat{\theta}) = tw(\theta)K[\sqrt{t}(\theta - \hat{\theta})]$, where K is as above and w is a smooth positive weight function, then Theorems 1, 2 and 3 may be applied after reparameterization. To see this, let $a^*(\theta) = a(\theta)/w(\theta)$ and $b^*(\theta) = b(\theta)/w(\theta)$ for $\theta \in \mathbb{R}$. Then for any density $\xi \in \Xi_1$,

$$\int_{\mathbb{R}} E_\theta\big\{tw(\theta)K\big[\sqrt{t}(\theta - \hat{\theta}_{MN})\big] + Ma(\theta) + Nb(\theta)\big\}\xi(\theta)\, d\theta$$

$$= \kappa_\xi \int_{\mathbb{R}} E_\theta\big\{tK\big[\sqrt{t}(\theta - \hat{\theta}_{MN})\big] + Ma^*(\theta) + Nb^*(\theta)\big\}\xi^*(\theta)\, d\theta,$$

where $\xi^*(\theta) \propto w(\theta)\xi(\theta)$ for $\theta \in \mathbb{R}$ and $\kappa_\xi = \int w\xi\, d\theta$. The latter integral is the integrated risk for the problem in which a, b and ξ have been replaced by a^*, b^* and ξ^*. So Theorems 1 and 2 are applicable to it; and Theorem 3 is applicable if K is the square function.

3. We conjecture that Theorem 3 (and hence Remark 1) are valid for a much larger class of loss functions.

4. We also conjecture that analogues of Theorems 1, 2 and 3 exist for the fully Bayesian formulation of the problem in which θ is estimated by a Bayes estimator, although some additional terms may appear in the expansions.

5. The three-stage procedure does not specify how subjects are to be allocated within batches. This could be done by drawing balls from an urn, in order to preserve some blindness. It would be interesting to know what kinds of biased coin designs could be used within batches without affecting the expansion for the risk.

6. There is also a fully sequential procedure which may be described as follows: First take $r \geq 1$ observations for each population; after m observations have been taken from the control and n from the treatment, take the next observation on the treatment iff $n \leq (m + n)q \circ c(\hat{\theta}_{mn})$. It is conjectured that this procedure has the same risk as the three-stage procedure, asymptotically to second order. It would be interesting to know whether the fully sequential procedure is better than the three-stage procedure is any important sense.

7. There is also a fully sequential formulation of the problem, in which a stopping time is sought, along with the allocation rule. For squared error loss, this degenerates into two one-sample sequential problems; but for other loss functions, it does not.

8. In the proof of Theorem 1, it is not necessary that (13) hold for all $\alpha > 0$. If $K(z) \leq J(1 + |z|^p)$ for all z for some $J \geq 0$ and even integer p, then it is sufficient for (13) to hold for some $\alpha > 2 + 2p$.

9. There is a corresponding first-order theory [obtainable by dividing (14) by t] for which a two-stage procedure is asymptotically (first-order) optimal. This theory is not pursued here, because it is nearly trivial in the present context. It could be developed in a much more general context, however.

7. The proof of Proposition 2.

The notation of Sections 1 and 2 is used without comment. Two lemmas are needed.

LEMMA 9. *κ is strictly increasing and infinitely differentiable on* $(0, \infty)$. *Moreover,*

$$\sigma^2\kappa'(\sigma) = 2\int_0^{\infty} z\varphi\left(\frac{z}{\sigma}\right)\, dK(z) > 0$$

and

$$\kappa''(\sigma) \geq -\frac{2}{\sigma}\kappa'(\sigma), \qquad \forall\, \sigma > 0.$$

PROOF. The differentiability follows easily from the relation

$$\kappa(\sigma) = 2\int_0^\infty K(z)\frac{1}{\sigma}\varphi\left(\frac{z}{\sigma}\right)dz, \qquad \sigma > 0.$$

In fact, differentiation under the integral sign is permitted. See, for example, Brown (1987), pages 34 and 35. Now $\partial\{\sigma^{-1}\varphi(\sigma^{-1}z)\}/\partial\sigma = \partial^2\varphi(\sigma^{-1}z)/\partial^2 z$ for all σ and z. So

$$\begin{aligned}\kappa'(\sigma) &= 2\int_0^\infty K(z)\frac{\partial^2}{\partial z^2}\varphi\left(\frac{z}{\sigma}\right)dz\\ &= -2\int_0^\infty \frac{\partial}{\partial z}\varphi\left(\frac{z}{\sigma}\right)dK(z)\\ &= 2\sigma^{-2}\int_0^\infty z\varphi\left(\frac{z}{\sigma}\right)dK(z),\end{aligned}$$

for all $\sigma > 0$. This establishes the third assertion of the lemma. That $\kappa(\sigma)$ and $\sigma^2\kappa'(\sigma)$ are strictly increasing follows immediately; and the final assertion then follows from

$$0 \le \frac{d}{d\sigma}\{\sigma^2\kappa'(\sigma)\} = \sigma^2\kappa''(\sigma) + 2\sigma\kappa'(\sigma), \qquad \sigma > 0. \qquad \square$$

LEMMA 10. *$\kappa\{\sqrt{[1/p + 1/(1-p)]}\}$ is strictly convex in $0 < p < 1$.*

PROOF. Let

$$\tau = \tau(p) = \sqrt{\left[\frac{1}{p} + \frac{1}{1-p}\right]}.$$

Then

$$\tau'(p) = (p - \tfrac{1}{2})\tau^3(p), \qquad \tau''(p) = \tau^3(p) + 3(p - \tfrac{1}{2})^2\tau^5(p)$$

and

$$\begin{aligned}\frac{d^2}{dp^2}\kappa\left\{\sqrt{\left[\frac{1}{p} + \frac{1}{1-p}\right]}\right\} &= \kappa''(\tau)\tau'^2 + \kappa'(\tau)\tau''\\ &\ge \kappa'(\tau)\left\{\tau'' - 2\frac{\tau'^2}{\tau}\right\}\\ &= \kappa'(\tau)\left\{\tau^3(p) + (p - \tfrac{1}{2})^2\tau^5(p)\right\},\end{aligned}$$

where the inequality follows from Lemma 9. Since the last line is positive, the lemma follows. $\square$

PROOF OF PROPOSITION 2. That $\psi(p, c) = \kappa(\tau) + cp$ is strictly convex and twice continuously differentiable in p for each fixed $c \in \mathbb{R}$ follows directly from Lemma 10. Clearly,

$$\frac{\partial}{\partial p}\psi(p,c) = (p - \tfrac{1}{2})\tau^3\kappa'(\tau) + c,$$

for $0 < p < 1$. This is increasing in p by Lemma 10. Since $\tau \to \infty$ as $p \to 0$ or 1, $\partial\psi(p,c)/\partial p$ approaches $-\infty$ or ∞ as p approaches 0 or 1, by (the third assertion in) Lemma 9. Thus, the equation $\partial\psi(p,c)/\partial p = 0$ has a unique solution $p = q(c)$ for all $c \in \mathbb{R}$. Here $q(c)$ minimizes $\psi(p,c)$ by the convexity; and the differentiability of q is easily established. □

REFERENCES

BROWN, L. (1989). *Fundamentals of Statistical Exponential Families*. IMS, Hayward, Calif.

HAGGSTROM, G. (1966). Optimal stopping and experimental design. *Ann. Math. Statist.* **37** 7–29.

HALL, P. (1981). Asymptotic theory of triple sampling for sequential estimation of the mean. *Ann. Statist.* **9** 1229–1238.

LOUIS, T. (1975). Optimal allocation in sequential tests comparing the means of two Gaussian populations. *Biometrika* **62** 359–370.

MESLEM, A. (1987). Asymptotic expansions for confidence intervals with fixed proportional accuracy. Ph.D. thesis, Univ. Michigan.

ROBBINS, H. and SIEGMUND, D. (1974). Sequential test involving two populations. *J. Amer. Statist. Assoc.* **69** 132–139.

SHAPIRO, C. (1983). Sequential allocation and optimal stopping in Bayesian simultaneous estimation. *J. Amer. Statist. Assoc.* **78** 396–402.

STEIN, C. (1986). *The Approximate Computation of Expectations*. IMS, Hayward, Calif.

WHITTLE, P. (1982). *Optimization over Time*. Wiley, New York.

WOODROOFE, M. (1985). Asymptotic local minimaxity in sequential estimation. *Ann. Statist.* **13** 676–688.

WOODROOFE, M. (1988). Fixed proportional accuracy in three stages. *Statistical Decision Theory and Related Topics IV* (S.S. Gupta and J.O. Berger, eds.) **2** 209–221.

WOODROOFE, M. (1989). Very weak expansions for sequentially designed experiments: Linear models. *Ann. Statist.* **17** 1087–1102.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109