Conditional Inference in Two-Stage Adaptive Experiments via the Bootstrap

Adam Lane, HaiYing Wang, and Nancy Flournoy

Abstract We study two-stage adaptive designs in which data accumulated in the first stage are used to select the design for a second stage. Inference following such experiments is often conducted by ignoring the adaptive procedure and treating the design as fixed. Alternative inference procedures approximate the variance of the parameter estimates by approximating the inverse of expected Fisher information. Both of these inferential methods often rely on normal distribution assumptions in order to create confidence intervals. In an effort to improve inference, we develop bootstrap methods that condition on a non-ancillary statistic that defines the second stage design.

1 Introduction

In many experiments, the evaluation of designs with respect to a specific objective requires knowledge of the true model parameters. Examples include optimal designs for the estimation of nonlinear functions of the parameters in linear models; optimal designs for the estimation of linear functions of parameters in nonlinear models; and dose-finding designs where it is desired to treat patients at a pre-specified quantile of the response function. In the absence of perfect knowledge of the model parameters, it is appealing to update initial parameter estimates using data accumulated from all previous stages and to allocate observations in the current stage by an assessment of designs evaluated at these estimates. Such procedures result in designs that are functions of random variables whose distributions depend on the model parameters, i.e., the designs are not ancillary statistics.

A. Lane (✉)
Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229, USA
e-mail: Adam.Lane@cchmc.org

H. Wang
University of New Hampshire, 33 Academic Way, Durham, NH 03824, USA
e-mail: HaiYing.Wang@unh.edu

N. Flournoy
University of Missouri, 146 Middlebush Hall, Columbia, MO 65211, USA
e-mail: flournoyn@missouri.edu

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_20

In practice it is common to conduct inference ignoring the adaptive nature of the experiment and treating the design as if it were an ancillary statistic. The potential issues resulting from such procedures are well known; see [1]. Briefly, when a design is ancillary, it is non-informative with respect to model parameters. Analysing adaptive experiments with non-ancillary designs as if the design were ancillary does not account for all the information contained in the design. An alternative often suggested is to approximate the variance of the maximum likelihood estimate (MLE) with an approximation of the inverse of the expected Fisher information (defined below). Regardless of the method, the distribution of the MLE resulting from adaptive experiments are often assumed to follow a normal distribution when used to create confidence intervals.

We develop three bootstrap procedures that approximate the distribution of the MLE conditional on a non-ancillary statistic that defines the second stage design. The bootstrap reduces the reliance on the assumption of normality. A bootstrap procedure in the context of adaptive experiments with binary outcomes, primarily in the context of urn models, has been developed previously [7].

2 The Model and an Illustration of the Problem

Consider an experiment conducted to measure a constant θ. Suppose it is possible to obtain unbiased observations y of θ from two sources; each source $k = r, s$ produces errors from a $\mathcal{N}(0, \sigma_k^2)$, where $\sigma_r^2 \neq \sigma_s^2$ are known constants. Throughout this paper Y and y denote the random variable and its observed realization, respectively. A function $\psi(\theta)$ is defined to determine which of the two sources should be used based on maximizing efficiency, ethics, cost or other considerations. For illustration, we use $\psi(\theta) = r$ if $\theta < c$ and s otherwise, given some constant c.

The adaptive procedure is as follows: obtain an initial cohort of n_1 independent observations, $y_1 = (y_{11}, \ldots, y_{1n_1})$, from source r. Evaluate ψ at the MLE of θ based on first stage observations; $\hat{\theta}_1 = \bar{y}_1 = \sum_{j=1}^{n_1} y_{1j}/n_1$. The function $\psi(\bar{y}_1)$ determines the source from which a second cohort of n_2 independent observations, $y_2 = (y_{21}, \ldots, y_{2n_2})$, will be obtained and induces a correlation between Y_1 and Y_2. This model has been used to illustrate conditional inference in experiments with ancillary designs; see [2] and [4]. Here we use the model to illustrate conditional inference with non-ancillary designs in adaptive experiments.

Let $N = n_1 + n_2$ represent the total sample size; and let $S_r = (-\infty, c)$ and $S_s = (c, \infty)$ be the regions of the parameter space that selects the second stage source. The second stage design is completely determined by the non-ancillary variable $I(\overline{Y}_1 \in S_k)$, $k = r, s$, where $I(\cdot)$ is the indicator function. Therefore, $Y_2|\overline{Y}_1 \in S_k$ is independent of $\overline{Y}_1$, $k = r, s$. The log likelihood is $l_\theta = -n_1 (\bar{y}_1 - \theta)^2 / 2\sigma_r^2 - n_2 (\bar{y}_2 - \theta)^2 / 2\sigma_{\psi(\bar{y}_1)}^2 +$ constant. Defining $w(\bar{y}_1) = (n_1/\sigma_r^2 + n_2/\sigma_{\psi(\bar{y}_1)}^2)^{-1}$, the MLE based on both stages is $\hat{\theta} = w(\bar{y}_1)(n_1\bar{y}_1/\sigma_r^2 + n_2\bar{y}_2/\sigma_{\psi(\bar{y}_1)}^2)$.

To approximate the distribution of $\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_k$, $k = r, s$, three bootstrap methods are developed. Resulting parameter estimates and confidence intervals are compared to analyses where the variance of the MLE is approximated using the inverse of the observed and expected Fisher information: $\mathscr{I}(\hat{\theta}) = -[\partial^2 l_\theta/\partial\theta^2]_{\theta=\hat{\theta}} = w^{-1}(\bar{y}_1)$ and $\mathscr{F}_\theta = \mathrm{E}[\mathscr{I}(\hat{\theta})]$, respectively. Confidence intervals for comparative methods are then constructed using normal quantiles. If the design were ancillary, then the variance of the MLE would be $w(\bar{y}_1)$. Using $\mathscr{F}_\theta^{-1}$ to approximate the variance, does not treat the design as ancillary. But it is a function of θ and must be estimated; $\mathscr{F}_{\hat{\theta}}^{-1}$ is used here. In contrast, since $Y_2|\overline{Y}_1 \in S_k$ is independent of $\overline{Y}_1$, $\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_k$ has variance

$$\mathrm{Var}\left[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_k\right] = w_k^2\left(\frac{n_1^2}{\sigma_r^4}\mathrm{Var}\left[\overline{Y}_1|\overline{Y}_1 \in S_k\right] + \frac{n_2}{\sigma_k^2}\right),$$

where $w_k = (n_1/\sigma_r^2 + n_2/\sigma_k^2)^{-1}$. The design of the second stage is completely determined by $I(\overline{Y}_1 \in S_k)$. Therefore $\{\psi(\overline{Y}_1)|\overline{Y}_1 \in S_k\} = k$. The function $w(\overline{Y}_1)$ depends on $\overline{Y}_1$ only through $\psi(\overline{Y}_1)$. Hence $\{w(\overline{Y}_1)|\overline{Y}_1 \in S_k\} = w_k$, for $k = r, s$.

Let $p_k = P(\overline{Y}_1 \in S_k)$, $E_k = \mathrm{E}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_k]$, $V_k = \mathrm{Var}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_k]$ and $\mathscr{I}_k = \mathscr{I}(\hat{\theta})$, $k = r, s$ for short.

From a series of 10,000 simulations, Table 1 presents, for $k = r, s$, p_k, E_k, $N\widehat{se}^2$ and tail probabilities, $\mathrm{P}[\theta < C_l]$ and $\mathrm{P}[\theta > C_u]$, where $\widehat{se}^2$ is either $\mathrm{Var}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)]$, V_k, $\mathscr{I}_k^{-1}$ or $\mathscr{F}_\theta^{-1}$; $C_l = \hat{\theta} - Z_{1-\alpha/2}\widehat{se}$, $C_u = \hat{\theta} + Z_{1-\alpha/2}\widehat{se}$ and Z_α is the α quantile of the standard normal distribution. Values of $\theta = 1$, $\sigma_r = 1$, $\sigma_s = 3$, $n_1 = 50$, $n_2 = 150$, $c = \theta + 1/\sqrt{n_1}$ and $\alpha = 0.05$ are used throughout. Both $\mathscr{I}_k^{-1}$ and $\mathscr{F}_\theta^{-1}$ are significantly greater than V_k, $k = r, s$. Both V_k, $k = r, s$, are significantly less than $\mathrm{Var}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)]$. If $\overline{Y}_1 \in S_r$, then E_r is nearly unbiased, and despite slightly unbalanced tail probabilities, overall coverage would be adequate if V_r were known and could be used. When $\overline{Y}_1 \in S_s$, the bias is considerable and the coverage is unacceptable. We focus throughout on improvements when $\overline{Y}_1 \in S_s$.

Table 1 Conditional probability, conditional expectation, variance approximation method along with its estimate and tail probabilities by the source of the second stage observations

Source	p_k	E_k	Variance approximation	$N\widehat{se}^2$	$\mathrm{P}[\theta < C_l]$	$\mathrm{P}[\theta > C_u]$
r	0.77	0.99	V_r	0.88	0.02	0.04
			$\mathrm{Var}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)]$	1.77	0.00	0.01
			$\mathscr{I}_r^{-1}$	1.00	0.01	0.03
			$\mathscr{F}_\theta^{-1}$	1.47	0.00	0.01
s	0.23	1.14	V_s	1.24	0.39	0.00
			$\mathrm{Var}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)]$	1.77	0.26	0.00
			$\mathscr{I}_s^{-1}$	3.00	0.10	0.00
			$\mathscr{F}_\theta^{-1}$	1.47	0.33	0.00

3 Conditional Bootstrap Inference

In an effort to improve inference following adaptive experiments, three bootstrap methods are developed. We begin with a straightforward and intuitive conditional bootstrap procedure. Unfortunately, as will be discussed, the conditions required for this procedure to give accurate inference are extremely restrictive.

Conditional Bootstrap Method 1 (BM1)

1. Construct a probability distribution, $\hat{F}_i$, by putting mass $1/n_i$ at each point $y_{i1}, \ldots, y_{in_i}$, $i = 1, 2$. With fixed $\hat{F}_1$ and $\hat{F}_2$, draw random samples of size n_1 and n_2 from $\hat{F}_1$ and $\hat{F}_2$, respectively. Denote the vector of bootstrap samples as $y_i^*, i = 1, 2$.
2. Repeat step 1 B times. For the bootstrap samples satisfying $\bar{y}_1^* \in S_{\psi(\bar{y}_1)}$, use $\bar{y}_1^*$ and $\bar{y}_2^*$ to find $\hat{\theta}^*$. Use $(C_l, C_u) = (Q^*_{\alpha/2}, Q^*_{1-\alpha/2})$ as the $(1-\alpha)$ confidence interval, where Q^*_α is the α quantile of the bootstrapped sample distribution $\hat{\theta}^*(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in S_{\psi(\bar{y}_1)}$.

Let P^* be the empirical probability measure given Y_1 and let $[\cdot]_b$ represent the bth sampled bootstrap. Then $P^*(\overline{Y}_1^* \in S_{\psi(\bar{y}_1)}) \approx \sum_{b=1}^{B} I([\bar{y}_1^*]_b \in S_{\psi(\bar{y}_1)})/B$ is the probability the mean of a bootstrap sample is in $S_{\psi(\bar{y}_1)}$.

Table 2 presents the simulation results p_s, E_s, V_s, C_l, C_u, $\mathrm{P}[\theta < C_l]$ and $\mathrm{P}[\theta > C_u]$. The first row results are for the distribution of $\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s$. The second and third row results use $\mathscr{I}_s^{-1}$ and $\mathscr{F}_{\hat{\theta}}^{-1}$, and tail probabilities found using

Table 2 Results are conditional on the region $\overline{Y}_1 \in S_s$. Method describes the procedure by which the approximate distribution was obtained; p_s is approximated by $\sum_{b=1}^{B} I([\bar{y}_1^*]_b \in S_s)/B$ for the bootstrap and by $P(\overline{Y}_1 \in S_s|\theta = \hat{\theta})$ for the expected information. E_s, V_s, the confidence limits and tail probabilities are also provided. All values are averages from 10,000 simulations except the column p_s which uses the median. "True" refers to either the empirical distribution of $\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s$ or $\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s$ obtained from simulation

Estimate	Method	p_s	E_s	N*V$_s$	C_l	C_u	P[$\theta < C_l$]	P[$\theta > C_u$]
$\hat{\theta}$	True	0.23	1.14	1.24	0.99	1.30	–	–
$\hat{\theta}$	$\mathscr{F}_{\hat{\theta}}^{-1}$	0.59	1.14	2.17	0.93	1.34	0.14	0.00
$\hat{\theta}$	$\mathscr{I}_s^{-1}$	–	1.14	3.00	0.90	1.38	0.10	0.00
$\hat{\theta}^*$	BM1	0.67	1.19	1.87	1.02	1.39	0.63	0.00
$\hat{\theta}^+$	BM2	0.25	1.13	1.36	0.99	1.30	0.41	0.00
$\tilde{\theta}$	True	0.23	1.00	5.25	0.66	1.28	–	–
$\tilde{\theta}$	$\Delta_{\tilde{\theta}}$	0.74	1.00	6.47	0.65	1.35	0.02	0.01
$\tilde{\theta}^+$	BM3	0.25	0.98	5.60	0.65	1.29	0.02	0.02

quantiles of the standard normal distribution. The fourth row results are from BM1 using $B = 1000$ bootstrap samples. All values are averages except p_s whose values are skewed and for which the median provides a better summary of the data. Both $P^*(\overline{Y}_1^* \in S_s)$ and $P(\overline{Y}_1 \in S_s|\theta = \hat{\theta})$ are significantly greater than p_s. The methods BM1 and $\mathscr{F}_{\hat{\theta}}^{-1}$ produce similar variances, both of which are less than $\mathscr{I}_s^{-1}$. All three approximations are significantly greater than V_s. However, the mean of the bootstrap distribution has greater bias than $\hat{\theta}$. Thus BM1 does not improve inference in comparison to using $\mathscr{F}_{\hat{\theta}}^{-1}$.

3.1 Technical Details for the Conditional Bootstrap

In this section we consider the MLE conditional on an indicator function of the first stage sample mean. This illustrates the implications of conditioning in two-stage adaptive experiments with non-ancillary designs and why the inference produced by BM1 was poor. Let $T_k = (a_k, d_k)$, where $a_k = \theta + a'_k/\sqrt{n_1}$, $d_k = \theta + d'_k/\sqrt{n_1}$, a'_k and d'_k are constants. This is a slightly more general division of the parameter space than S_k which has only one of a'_k and d'_k finite. Note, we consider the parameter space defined in a local neighborhood of θ. This is done so that $P(\overline{Y}_1 \in S_k)$, $k = r, s$ has positive probability not equal to 1 for large n_1; otherwise the design would be deterministic for large first stage sample sizes and not adaptive. Consider

$$\begin{aligned}
&P^*\left[\sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) < x | \sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\right] \\
&\quad = \frac{P^*\left[\{\sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) < x\} \cap \{\sqrt{n_1}(\overline{Y}_1^* - \overline{Y}_1) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\}\right]}{P^*\left[\sqrt{n_1}(\overline{Y}_1^* - \overline{Y}_1) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\right]} \\
&\quad = \frac{P\left[\{\sqrt{n_1}\left(\overline{Y}_1 - \theta\right) < x\} \cap \{\sqrt{n_1}(\overline{Y}_1 - \theta) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\}\right] + o(1)}{P\left[\sqrt{n_1}(\overline{Y}_1 - \theta) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\right] + o(1)} \\
&\quad = P\left[\sqrt{n_1}\left(\overline{Y}_1 - \theta\right) < x | \sqrt{n_1}\left(\overline{Y}_1 - \theta\right) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\right] + o(1),
\end{aligned} \tag{1}$$

where $x \in \mathbb{R}$, and the equalities hold almost surely under P. Note, from the left hand side of equation (1), that a conditional bootstrap should include bootstrap samples satisfying $\sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)$, i.e., $\overline{Y}_1^* \in (a_k - \overline{\varepsilon}_1, d_k - \overline{\varepsilon}_1)$, where $\overline{\varepsilon}_1 = \overline{Y}_1 - \theta$. BM1 considers bootstrap samples satisfying $\overline{Y}_1^* \in (a_k, d_k)$ and hence its poor performance. Unfortunately, θ is unknown and must be estimated. In (1), if one can replace θ with an estimator (say $\tilde{\theta}$) that converges to θ at a rate faster than $\sqrt{n_1}$, then

$$P^*\left[\sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) < x | \sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) \in \sqrt{n_1}(a_k - \tilde{\theta}, d_k - \tilde{\theta})\right]$$
$$= P\left[\sqrt{n_1}\left(\overline{Y}_1 - \theta\right) < x | \sqrt{n_1}\left(\overline{Y}_1 - \theta\right) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\right] + o_P(1). \tag{2}$$

In a single stage experiment this would not be possible. However, in a two-stage adaptive experiment in which $n_1 = o(n_2)$, such an estimator may exist. Theoretically this is somewhat unsatisfactory, but practically this is interesting especially since it has been shown that two-stage experiments are optimized when $n_1 = O(\sqrt{n_2})$; see [5] and [6]. It is also common, for practical or logistical reasons, that a small pilot study precedes a much larger follow-up.

Note $\overline{Y}_1 | \overline{Y}_1 \in T_k \sim \mathcal{TN}(\theta, \sigma_k^2/n_1; T_k)$, where $\mathcal{TN}$ denotes a truncated normal distribution. Suppose an estimator, $\tilde{\theta}$, exists as described and let $\tilde{T}_k = (a_k - \tilde{\varepsilon}_1, d_k - \tilde{\varepsilon}_1)$, where $\tilde{\varepsilon}_1 = \overline{Y}_1 - \tilde{\theta}$. Then (2) implies $\{\overline{Y}_1^* | \overline{Y}_1^* \in \tilde{T}_k\}$ can be approximated by a $\mathcal{TN}(\overline{y}_1, \sigma_k^2/n_1; \tilde{T}_k)$ and therefore

$$\mathrm{E}^*_{\overline{Y}_1}\left[\overline{Y}_1^* | \overline{Y}_1^* \in \tilde{T}_k\right] \approx \overline{Y}_1 + b_{1k}(\tilde{\theta}); \tag{3}$$

$$\mathrm{Var}_{\overline{Y}_1}\left[\overline{Y}_1^* | \overline{Y}_1^* \in \tilde{T}_k\right] \approx \frac{\sigma_k^2}{n_1}\left[1 + \frac{\gamma_{\tilde{\theta}}(a_k)\phi\{\gamma_{\tilde{\theta}}(a_k)\} - \gamma_{\tilde{\theta}}(d_k)\phi\{\gamma_{\tilde{\theta}}(d_k)\}}{U}\right] - \left\{b_{1k}(\tilde{\theta})\right\}^2.$$

where $b_{1k}(\theta) = \sigma_k\left[\phi\{\gamma_\theta(a_k)\} - \phi\{\gamma_\theta(d_k)\}\right]/(\sqrt{n_1}U)$, $\gamma_\theta(a_k) = \sqrt{n_1}(a_k - \theta)/\sigma_k$, $U = \Phi\{\gamma_\theta(b)\} - \Phi\{\gamma_\theta(a)\}$; $\phi(\cdot)$ and $\Phi(\cdot)$ represent the probability density function (PDF) and cumulative distribution function (CDF) of the standard normal distribution, respectively. Note a conditional bootstrap leads to an additional bias term in the expectation in equation (3). This must be accounted for in any conditional bootstrap procedure.

4 Adjusted Conditional Bootstrap Methods

The bootstrap methods developed in this section are predicated on the existence of an estimator that converges to θ at a rate faster than $\sqrt{n_1}$. Provided standard regularity conditions hold and $n_1 = o(n_2)$, such an estimator will always exists in the form of the second stage MLE conditional on the second stage design. However, in cases where the bias has an explicit form, it is possible to obtain $\tilde{\theta}$ by adapting the bias reduction method suggested in [8]. Note

$$\begin{aligned} \mathrm{E}\left[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_k\right] &= w_k \mathrm{E}\left[n_1\overline{Y}_1/\sigma_k^2 + n_2\overline{Y}_2/\sigma_k^2|\overline{Y}_1 \in S_k\right] \\ &= w_k\left(n_1/\sigma_r^2 + n_2/\sigma_k^2\right)\theta + w_k n_1 b_{1k}(\theta)/\sigma_r^2 \\ &= \theta + w_k n_1 b_{1k}(\theta)/\sigma_r^2. \end{aligned}$$

Let $b_k(\theta) = w_k n_1 b_{1k}(\theta)/\sigma_r^2$. Then a bias corrected estimate of θ is $\tilde{\theta} = \hat{\theta} - b_k(\tilde{\theta})$. The Newton-Raphson method was used to solve this equation. One iteration was sufficient, that is, $\tilde{\theta} = \hat{\theta} - \lambda_k(\hat{\theta})$, where $\lambda_k(\hat{\theta}) = b_k(\hat{\theta})/[1 + (\partial b_k(\theta)/\partial\theta)_{\theta=\hat{\theta}}]$ was used for analytic expressions and numeric calculations.

Now we develop a bootstrap method that adjusts the conditioning region per the discussion in Sect. 3.1. Let $\tilde{S}_r = \{-\infty, c - \tilde{\varepsilon}_1\}$ and $\tilde{S}_s = \{c - \tilde{\varepsilon}_1, \infty\}$, where $\tilde{\varepsilon}_1 = \overline{Y}_1 - \tilde{\theta}$.

Adjusted Conditional Bootstrap Method (BM2): Repeat BM1 keeping only bootstrap samples satisfying $\bar{y}_1^* \in \tilde{S}_{\psi(\bar{y}_1)}$. Let $\hat{\theta}^+ = \hat{\theta}^* - b_{\psi(\bar{y}_1)}(\tilde{\theta})$ and $(C_l^+, C_u^+) = (Q_{\alpha/2}^+, Q_{1-\alpha/2}^+)$ as the $(1-\alpha)$ confidence interval, where Q_α^+ is the α quantile of the bootstrap sample distribution of $\hat{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_{\psi(\bar{y}_1)}$.

Note $\hat{\theta}^+$ is used in place of $\hat{\theta}^*$ to correct for the additional bias term previously discussed.

A concern in inference when the variance of the MLE depends on the parameter estimates is how sensitive the approximate distribution is to this estimate. Figure 1 (top left) plots the histogram of the simulated distribution of $\{\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s\} - E_s$ (solid line). In the same figure, for simulations that correspond to the 0.025, 0.50 and 0.975 quantiles of $\hat{\theta}$, histograms of the bootstrap sample distributions of $\{\hat{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_s\} - \hat{\theta}$ (dotted, dashed and dot-dashed line) are plotted. This figure illustrates how well BM2 works across the domain of θ at approximating the distribution $\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s$. Compare this to Fig. 1 (top right) which plots the PDF of $\mathscr{N}(0, \mathscr{F}_{\hat{\theta}}^{-1})$ for $\mathscr{F}_{\hat{\theta}}^{-1}$ evaluated at the same quantiles of $\hat{\theta}$. The bootstrap distribution is less sensitive to the value of $\hat{\theta}$ than $N(0, \mathscr{F}_{\hat{\theta}}^{-1})$ and it provides a better approximation to the shape of the target distribution.

The fifth row of Table 2 shows that with BM2 $P^*(\overline{Y}_1^* \in \tilde{S}_s)$ is nearly equal to p_s; $\mathrm{E}^*[\hat{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_s]$ is approximately equal to E_s; $\mathrm{Var}^*[\hat{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_s]$ is only slightly greater than V_s; and the confidence interval endpoints correspond to the quantiles of $\hat{\theta}$. It is clear that BM2 provides a more accurate representation of the conditional distribution of $\hat{\theta}$ than the alternatives. However, BM2 still fails to provide adequate coverage. This is not a problem with the bootstrap but rather a problem with the distribution of $\hat{\theta}$ being tightly centered around the wrong value.

Considering the poor coverage of BM2, we develop a bootstrap procedure for the bias corrected estimator $\tilde{\theta}$. Recall $\tilde{\theta} = \hat{\theta} - \lambda_k(\hat{\theta})$, $k = r, s$, is simply a function of $\hat{\theta}$. Thus BM2 can be adapted to estimate its distribution. Because the distribution of the bias is skewed, bias corrected bootstrap confidence intervals were used; see [3].

Bias Adjusted Conditional Bootstrap Method (BM3): Replace $\hat{\theta}^+$ in BM2 with $\tilde{\theta}^+ = \hat{\theta}^+ - \lambda_{\psi(\bar{y}_1)}(\hat{\theta}^+)$. Use $(C_l, C_u) = [\widehat{CDF}^{-1}\{\Phi(2v_0 + Z_{\alpha/2})\}, \widehat{CDF}^{-1}\{\Phi(2v_0 - Z_{\alpha/2})\}]$, where $\widehat{CDF}(t) = P^*[\tilde{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_{\psi(\bar{y}_1)} < t]$ and $v_0 = \Phi^{-1}\widehat{CDF}(\tilde{\theta})$.

Since $\tilde{\theta}$ is a function of $\hat{\theta}$, for comparison we approximate the variance of $\tilde{\theta}$ with $\Delta_{\tilde{\theta}} = [\{\partial(\theta - \lambda_{\psi(\bar{y}_1)}(\theta))/\partial\theta\}^2 \mathscr{F}_{\theta}^{-1}]_{\theta=\hat{\theta}}$. Figure 1 (bottom left) plots the histogram of the simulated distribution of $\{\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s\} - \mathrm{E}[\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s]$ (solid line). In the same figure, for simulations that correspond to the 0.025, 0.50 and 0.975 quantiles of $\tilde{\theta}$, histograms of the bootstrap sample distributions of $\{\tilde{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_s\} - \tilde{\theta}$ (dotted, dashed and dot-dashed line) are plotted. For comparison, Fig. 1 (bottom right) plots the probability density functions of a $\mathscr{N}(0, \Delta_{\tilde{\theta}})$ with $\Delta_{\tilde{\theta}}$ evaluated at the same quantiles of $\tilde{\theta}$. Once again we see that

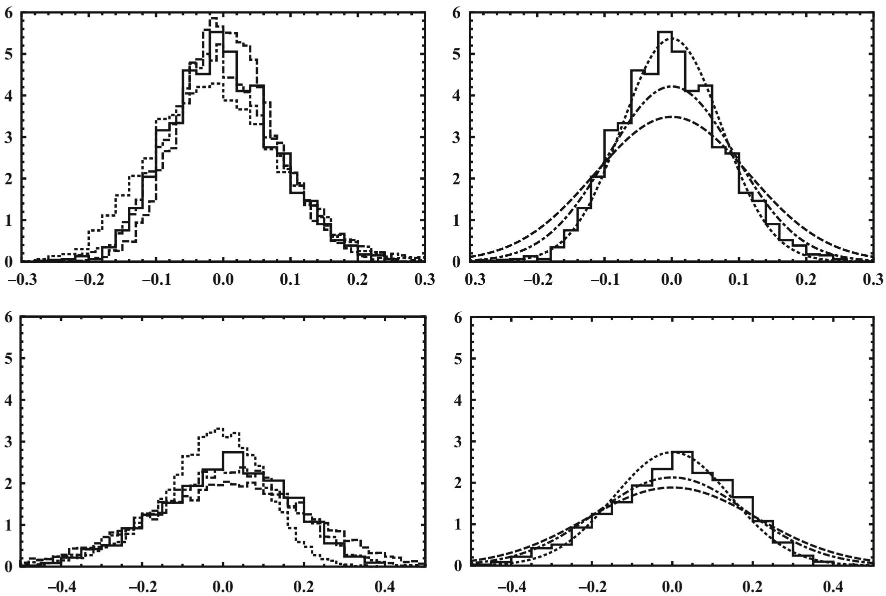

Fig. 1 Histograms of bootstrap distributions $\{\hat{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_s\} - \hat{\theta}$ (*top left*) and $\{\tilde{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_s\} - \tilde{\theta}$ (*bottom left*). Probability density functions of $N(0, \mathscr{F}_{\hat{\theta}}^{-1})$ (*top right*) and $N(0, \Delta_{\tilde{\theta}})$ (*bottom right*). The *dotted*, *dot dashed* and *dashed lines* correspond to the bootstrap distribution or expected information for the 0.025, 0.50, and 0.975 of quantiles of $\hat{\theta}$ or $\tilde{\theta}$. In each case the *solid line* is the histogram of the distribution of $\{\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s\} - E_s$ or $\{\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s\} - \mathrm{E}[\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s]$

Table 3 Mean square error (MSE) for the true distribution of $\hat{\theta}$ and $\tilde{\theta}$ along with the bootstrap MSE of $\hat{\theta}^+$ or $\tilde{\theta}^+$ for the case when $\overline{Y}_1 \in s$

Estimate	Method	MSE
$\hat{\theta}$	True	4.92
$\hat{\theta}^+$	BM2	4.98
$\tilde{\theta}$	True	5.25
$\tilde{\theta}^+$	BM3	5.71

the BM3 well approximates the distribution $\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s$ across most of the domain of $\tilde{\theta}$, perhaps with the exception of the 0.025 quantile.

The sixth row of Table 2 shows the simulation results the distribution of $\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s$. The seventh and eighth rows are results for $\Delta_{\tilde{\theta}}$ and BM3, respectively. BM3 provides an unbiased estimate of the mean; slightly overestimates the $\text{Var}[\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s]$; and provides confidence limits which coincide with the correct quantiles of $\tilde{\theta}$. Coverage is slightly less than the nominal level, but is a significant improvement over inference methods for $\hat{\theta}$. Note using $\Delta_{\tilde{\theta}}$ significantly overestimates $\text{Var}[\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s]$ and slightly skews the confidence limits.

The performance of $\hat{\theta}^+$ and $\tilde{\theta}^+$ reflects the bias versus variance tradeoff. Table 3 compares the mean square error (MSE) for the simulated distribution of $\hat{\theta}$ and $\tilde{\theta}$ along with the bootstrap MSE of $\hat{\theta}^+$ and $\tilde{\theta}^+$ for the case when $\psi(\bar{y}_1) = s$. Despite the shortcomings of the procedure for $\hat{\theta}^+$, it still provides a lower MSE than $\tilde{\theta}^+$ in this example.

References

1. Baldi Antognini, A., Giovagnoli, A.: On the asympotic inference for response-adaptive experiments. METRON – Int. J. Stat. **49**, 29–46 (2006)
2. Dette, H., Bornkamp, B., Bretz, F.: On the efficiency of two-stage response-adaptive designs. Stat. Med. **32**, 1646–1660 (2013)
3. Efron, B.: Nonparametric standard errors and confidence intervals. Can. J. Stat. **9**, 139–158 (1981)
4. Efron, B., Hinkley, D.V.: Assessing the accuracy of the maximum likelihood estimate: observed versus expected Fisher information. Biometrika **65**, 457–483 (1978)
5. Lane, A., Yao, P., Flournoy, N.: Information in a two-stage adaptive optimal design. J. Stat. Plan. Inference **144**, 173–187 (2014)
6. Pronzato, L., Pazman, A.: Design of Experiments in Nonlinear Models. Springer, New York (2013)
7. Rosenberger, W.F., Hu, F.: Bootstrap methods for adaptive designs. Stat. Med. **18**, 1757–1767 (1999)
8. Whitehead, J.: On the bias of maximum likelihood estimation following a sequential test. Biometrika **73**, 573–581 (1986)