



Published in final edited form as:

J Stat Plan Inference. 2013 March ; 143(3): 593–602. doi:10.1016/j.jspi.2012.08.014.

On the efficiency of nonparametric variance estimation in sequential dose-finding

Chih-Chi Hu and Ying Kuen Cheung

Department of Biostatistics, Columbia University, 722 West 168th Street, New York, New York 10032, U.S.A.

Abstract

Dose-finding in clinical studies is typically formulated as a quantile estimation problem, for which a correct specification of the variance function of the outcomes is important. This is especially true for sequential study where the variance assumption directly involves in the generation of the design points and hence sensitivity analysis may not be performed after the data are collected. In this light, there is a strong reason for avoiding parametric assumptions on the variance function, although this may incur efficiency loss. In this article, we investigate how much information one may retrieve by making additional parametric assumptions on the variance in the context of a sequential least squares recursion. By asymptotic comparison, we demonstrate that assuming homoscedasticity achieves only a modest efficiency gain when compared to nonparametric variance estimation: when homoscedasticity in truth holds, the latter is at worst 88% as efficient as the former in the limiting case, and often achieves well over 90% efficiency for most practical situations. Extensive simulation studies concur with this observation under a wide range of scenarios.

Keywords

Homoscedasticity; Least squares estimate; Phase I trials; Quantile estimation; Stochastic approximation

1. Introduction

We consider quantile estimation in the context of dose-finding study where patients are tested in successive groups of size m . Precisely, let X_i denote the dose given to the patients in the i th group, and Y_{ij} denote a continuous biomarker from the j th patient in the group. A response is said to occur if the outcome Y_{ij} exceeds a threshold t_0 . The objective is to estimate the dose θ such that $\pi(\theta) = p$ for some pre-specified p , where $\pi(x) := \text{pr}(Y_{ij} > t_0 \mid X_i = x)$. This clinical setting is not uncommon, and there is also a wide range of applications in other areas such as reliability testing and bioassay. However, quantile estimation based on continuous data has received relatively little attention in the literature. In practice, this problem is often dealt with by using sequential methods based on the dichotomised data $V_{ij} := \mathcal{I}(Y_{ij} > t_0)$, where $\mathcal{I}(A)$ is indicator of the event A , such as the logit-MLE (Wu, 1985) or the continual reassessment method (O'Quigley et al., 1990). These methods, using the

© 2012 Elsevier B.V. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

binary data to estimate θ , provide general solutions without imposing strong assumptions on the characteristics of Y_{ij} . On the other hand, this approach can result in substantial information loss due to dichotomisation. Cheung (2010) demonstrates that, with group size $m = 3$ and normal data, the asymptotic efficiency of an optimal logit-MLE using the dichotomised data V_{ij} is at most 80% of a corresponding Robbins-Monro (1951) procedure using the continuous data Y_{ij} ; and the efficiency loss becomes more substantial with a larger m or a more extreme target p . Having said this, we acknowledge that there is an ongoing need for designs and models for clinical situations with truly binary outcomes that are not results of dichotomisation of continuous outcomes. This paper, however, focuses on the relative efficiency of a least squares recursion using the continuous data under various assumptions on Y_{ij} . Generally, we consider the regression model

$$Y_{ij} = M(X_i) + \sigma(X_i)Z_{ij}, \quad (1)$$

where the noise Z_{ij} is standard normal. Among the earliest proposals to address this problem, Eichhorn and Zacks (1973) study sequential search procedures for θ under the assumptions that the mean function $M(x)$ is linear in x and the standard deviation is known and is constant, i.e., $\sigma(x) = \sigma$. Recently, Cheung and Elkind (2010) describe a novel application of the stochastic approximation method that leaves both $M(x)$ and $\sigma(x)$ unspecified subject to the constraint that θ is uniquely defined, and propose to estimate $\sigma(x)$ nonparametrically. These two sets of assumptions represent two extreme approaches, and raise the question whether there is a reasonable middle ground. Specifically, this article focuses on the estimation of the standard deviation function, and investigates how much efficiency may be retrieved by imposing stronger assumptions on $\sigma(x)$ than that in Cheung and Elkind (2010) while keeping the mean $M(x)$ unspecified. Our investigation will be conducted in the context of a sequential least squares recursion described in Section 2. Section 3 derives the asymptotic distribution of an proposed estimator for θ . Section 4 reviews Wu's (1985) logit-MLE as a comparison method of the least squares recursion. Efficiency comparison is given in Section 5, and concluding remarks in Section 6. Technical details are put in the Appendix.

2. Least squares recursion

Under model (1), Cheung and Elkind (2010) show that solving $\pi(\theta) = p$ is equivalent to solving $f(\theta) = t_0$, where $f(x) = M(x) + z_p\sigma(x)$ and z_p is the upper p th percentile of standard normal. For brevity in discussion, we may assume here that the objective function f is continuous and strictly increasing so that the solution θ exists uniquely. An important class of models that satisfies this assumption is models with increasing mean $M(x)$ and constant coefficient of variation across doses. Conditions 1–3 below make precise statements of the assumptions that are much less restrictive.

Now, pretend that $f(x) = t_0 + b(x - \theta)$ for some $b > 0$, and suppose also that we can observe an asymptotically unbiased variable $U_{i,n}$ of $f(X_i)$ for group i . A least squares estimate $\hat{\theta}_n$ of θ based on the first n groups of observations can be obtained by solving

$$\frac{1}{n} \sum_{i=1}^n [U_{i,n} - \{t_0 + b(X_i - \hat{\theta}_n)\}] = 0. \quad (2)$$

Then we may set the next dose

$$X_{n+1} = \hat{\theta}_n. \quad (3)$$

The least squares recursion formed by (2) and (3) in essence is identical to the adaptive design proposed by Lai and Robbins (1979). A subtle difference is that the unbiased variable $U_{i,n}$ is chosen based on the assumption about the variance function $\sigma(x)$.

Case 1 (known variance): When $\sigma(x)$ is completely known, a natural choice is to define $U_{i,n} = \bar{Y}_i + z_p \sigma(X_j)$, where $\bar{Y}_i = m^{-1} \sum_{j=1}^m Y_{ij}$ is the average of the measurements in group i .

Case 2 (heteroscedasticity): When $\sigma(x)$ is unknown and unspecified, we may define $U_{i,n} = \bar{Y}_i + z_p \lambda_m^{1/2} s_i$, where s_i^2 is the sample variance of the measurements in group i ,

$$\lambda_m = \frac{(m-1)\Gamma^2\{(m-1)/2\}}{2\Gamma^2(m/2)} \quad (4)$$

and $\Gamma(\cdot)$ is the gamma function. Note that the form of λ_m in (4) ensures $E(\lambda_m^{1/2} s_i) = \sigma(X_i)$ so that $U_{i,n}$ is unbiased for $f(X_j)$.

Under both Cases 1 and 2, the observed variable $U_{i,n}$ is unbiased for $f(X_j)$, and $U_{i,n}$ and $U_{j,n}$ are mutually independent for $i \neq j$. Therefore, using the same techniques as in Lai and Robbins (1979), we can then verify that the least squares recursion formed by (2) and (3) is identical to the nonparametric Robbins-Monro procedure under these two cases: $X_{n+1} = X_n - (nb)^{-1}(U_{n,n} - t_0)$, where $b > 0$ is the same as the assumed slope used in the least squares estimation (2). Hence, the standard convergence results of stochastic approximation apply so that $X_n \rightarrow \theta$ with probability one; for example, see Sacks (1958). In addition, if $b < 2f'(\theta)$, the distribution of $\sqrt{n}(X_n - \theta)$ will converge weakly to a mean zero normal with variance equal to $\alpha_1 \sigma^2(\theta)$ under Case 1 and $\alpha_1 \alpha_2 \sigma^2(\theta)$ under Case 2, where $\alpha_1 = [mb\{2f'(\theta) - b\}]^{-1}$ and $\alpha_2 = 1 + mz_p^2(\lambda_m - 1)$. In other words, the asymptotic relative efficiency due to the knowledge of $\sigma(x)$ is equal to α_2 . To illustrate the magnitude, the efficiency $\alpha_2 = 2.87, 2.35, 2.17$ for $m = 2, 3, 4$ and $p = 0.10$. The efficiency gain is quite substantial, and is not surprising because Cases 1 and 2 in a sense represent two extremities of assumptions.

Case 3 (homoscedasticity): When $\sigma(x)$ is identical to an unknown constant σ for all x , we may choose $U_{i,n} = \bar{Y}_i + z_p \hat{\sigma}_n$ where $\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n s_i^2$.

Under Case 3, we can rewrite the least squares recursion as follows:

$$X_{n+1} = \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{b} \left(\frac{1}{n} \sum_{i=1}^n \bar{Y}_i - \hat{\mu}_n \right), \quad (5)$$

where $\hat{\mu}_n = t_0 - z_p \hat{\sigma}_n$.

We note that homoscedasticity may not be a viable assumption in many practical situations, and it is arguably the strongest parametric assumption one can impose on $\sigma(x)$ besides complete knowledge assumed under Case 1. The consideration of Case 3 is intended to serve as a reference for Case 2, so as to shed light on how much efficiency one may lose due to nonparametric estimation of $\sigma(x)$.

3. Asymptotic normality under homoscedasticity

Convergence for the recursion formed by (2) and (3) does not trivially follow the standard results of stochastic approximation under Case 3, because the summands are correlated in a complex way via $U_{i,n}$. The following lemma is a key result that transforms the least squares

recursion (5) into a Robbins-Monro-type recursion with a target $\mu := t_0 - z_p\sigma$. Note that the estimand θ also solves $M(\theta) = \mu$ under homoscedasticity.

Lemma 1. *The design sequence $\{X_n\}$ generated by (5) under homoscedasticity can be represented as*

$$X_{n+1} = X_n - \frac{1}{nb} \left\{ \bar{Y}_n + \frac{z_p}{2\sigma} (s_n^2 - \sigma^2) - \mu \right\} + \xi_n \quad (6)$$

where $\sum_{n=1}^{\infty} E(|\xi_n| | \mathcal{F}_{n-1}) < \infty$ a.s. and \mathcal{F}_{n-1} denotes the σ -field generated by $(X_i, Z_{i1}, Z_{i2}, \dots, Z_{im})$ for $i = 1, \dots, n-1$.

In words, the recursion (6) is generated by the mean function $M(x)$ and independent errors

ξ_n with bias $\bar{\xi}_n = \sigma \left\{ \bar{Z}_n + 0.5z_p(s_n^2/\sigma^2 - 1) \right\}$ and $\bar{Z}_n = m^{-1} \sum_{j=1}^m Z_{nj}$. It is also easy to verify that $E(\xi_n) = 0$ and $\text{var}(\xi_n) = \alpha_3 \sigma^2 / m$, where $\alpha_3 = 1 + mz_p^2 \{2(m-1)\}^{-1}$. Hence, if the bias $\bar{\xi}_n$ is adequately small, we expect the convergence properties of (6) will be similar to that of the Robbins-Monro procedure without the bias term.

Condition 1. The mean function $M(x)$ is weakly increasing in that $(x - \theta)\{M(x) - \mu\} > 0$ for all $x \neq \theta$.

Condition 2. There exists a constant $C_1 > 0$ such that $|M(x) - \mu| \leq C_1|x - \theta|$ for all x .

Theorem 1. *Suppose Conditions 1 and 2 hold and $\sigma(x) \equiv \sigma$. The sequence $\{X_n\}$ generated by the least squares recursion (5) converges to θ with probability one.*

Condition 1 is weaker than requiring an increasing mean $M(x)$, and is often reasonable in dose-finding study. Condition 2 puts a bound on the tails of $M(x)$ and requires it to be flat on the tails. Particularly, while lab measurements of a bioassay in theory take on values from the real line (after taking log), they are typically confined to a finite range in practice. This implies that the mean function is bounded, which in turn satisfies Condition 2. Thus, the conditions for the consistency of the least squares recursion are quite mild and can often be verified from the clinicians. To obtain asymptotic normality of X_m , we also need:

Condition 3. The mean function can be expressed as $M(x) = \mu + \beta(x - \theta) + \tau(x, \theta)$ for all x such that $\beta > 0$ and $\tau(x, \theta) = o(|x - \theta|)$ as $x \rightarrow \theta$.

Condition 3 ensures that the local slope of $M(x)$ around θ is equal to β , while allows very flexible form of functions via $\tau(x, \theta)$. Note that under homoscedasticity, $\beta = M'(\theta) = f'(\theta)$ because $\sigma'(x) \equiv 0$.

Theorem 2. *Suppose Conditions 1–3 hold and $\sigma(\theta) \equiv \sigma$. If $b < 2\beta$, the distribution of $\sqrt{n}(X_n - \theta)$ converges weakly to a mean zero normal with variance $\alpha_1 \alpha_3 \sigma^2$.*

4. Dosing finding with dichotomised data

Instead of using the continuous data Y_{ij} , a convenient alternative using the dichotomised data V_{ij} by the logit-MLE recursion that solves

$$\sum_{i=1}^n \left[\sum_{j=1}^m V_{ij} - \frac{mp \exp\{\tilde{b}(\tilde{X}_i - \tilde{\theta}_n)\}}{1 - p + p \exp\{\tilde{b}(\tilde{X}_i - \tilde{\theta}_n)\}} \right] = 0 \quad (7)$$

and sets $\tilde{X}_{n+1} = \tilde{\theta}_n$. In practice, we need to consider a two-stage approach (Cheung, 2005) that assigns doses initially via the stochastic approximation based on the dichotomised data: $\tilde{X}_{n+1} = \tilde{X}_n - (n\tilde{b})^{-1} (m^{-1} \sum_{j=1}^m V_{ij} - p)$ and switches to logit-MLE when a unique solution to (7) exists, i.e., when $V_{ij} > V_{i'j'}$ for some $i > i'$ or $j > j'$.

Using the results in Ying and Wu (1997), we can show that $\tilde{X}_n \rightarrow \theta$ with probability 1, and that if $\tilde{b} < 2\beta\phi(z_p)\{\sigma p(1-p)\}^{-1}$, $\sqrt{n}(\tilde{X}_n - \theta)$ converges weakly to a mean zero normal distribution with variance

$$\frac{\sigma}{m\tilde{b}\{2\beta\phi(z_p) - \sigma\tilde{b}p(1-p)\}} \quad (8)$$

where $\phi(z)$ is the standard normal pdf. The asymptotic variance (8) achieves its minimum when $\tilde{b} = \tilde{\beta} := \beta\phi(z_p)\{\sigma p(1-p)\}^{-1}$.

5. Efficiency comparisons

As a consequence of Theorem 2, the asymptotic efficiency of the least squares recursion assuming heteroscedasticity (Case 2) relative to that assuming homoscedasticity (Case 3) equals α_3/α_2 , when homoscedasticity in truth holds. As shown in the left panel of Fig. 1, the ratio α_3/α_2 is uniformly less than 1. Such efficiency loss is not surprising because no parametric assumption on $\sigma(x)$ is made under Case 2, whereas homoscedasticity amounts to a single-parameter model. Generally, the efficiency worsens as p becomes extreme, and converges to $\{2(m-1)(\lambda_m-1)\}^{-1}$ in the limiting case $p \rightarrow 0$ or 1 where the ratio reaches a minimum of 0.88 when $m = 2$. The efficiency improves as the group size m increases, and always stays above 0.90 with $m \geq 3$.

In contrast, the efficiency of the least squares recursion assuming homoscedasticity (Case 3) against that assuming a known σ (Case 1) is plotted in the right panel of Fig. 1, which shows a great efficiency loss. The efficiency is about 0.40 for $p = 0.10$ and becomes arbitrarily close to 0 as $p \rightarrow 0$ or 1. These comparisons demonstrate that efficiency loss due to incomplete knowledge about σ is far more substantial than that due to relaxing the parametric assumptions on the variance function.

We conducted a series of simulation studies to compare efficiency in finite-sample settings. The outcomes are generated with mean

$$M(x) = 2\{1 + \exp(\theta - x)\}^{-1} \{-c_p + 2 \log(x - \theta + 1) + 1.5(x - \theta)^3\} \quad (9)$$

where c_p is the upper p th percentile of the cdf of Z_{ij} . We note that all methods in the simulation make the working assumption that Z_{ij} arises from a standard normal, even though we may generate noise from other distributions (see details below). This allows us to evaluate the impact of violation of the normality assumption. In the simulation, we set the variance $\sigma^2(x) \equiv 1$ for $x \in [0, 1]$ and $t_0 = 0$ so that θ as specified in (9) is the target p th percentile that we want to estimate under model (1). We consider $p = 0.1, 0.2$ and $\theta = 0.25, 0.50, 0.75$.

The simulation include the least squares recursion procedures described in Section 2 and the logit-MLE in Section 4. For the least squares recursion, we consider $b = \beta$, which corresponds to the optimal choice in terms of asymptotic variance, and $b = \beta/2$ in order to investigate the relative performance of the methods when we fail to choose a good b . For the logit-MLE, we set $\tilde{b} = \tilde{\beta}$ and $\tilde{\beta}/2$ respectively.

In the first set of simulations, we ran the four procedures with $m = 3$ and $n = 15$, and also considered the fully sequential version of the logit-MLE, i.e., $m = 1$ and $n = 45$. Each simulated trial will have a starting dose $X_1 = 0.25$ or 0.50 . We apply truncation to the subsequent doses and set the next dose at $X_{n+1} = \max\{\min(\hat{\theta}_n, 1), 0\}$ instead of (3) for the least squares recursion. Likewise, for the logit-MLE, we set $\tilde{X}_1 = 0.25$ or 0.50 , and set $\tilde{X}_{n+1} = \max\{\min(\tilde{\theta}_n, 1), 0\}$. Such truncation does not affect the asymptotic property of the recursion (see appendix), and is often done in practice.

Table 1 summarizes the results of the first simulation study. Overall, the biases are small when compared to the variances for all methods. In line with the asymptotic comparison in Fig. 1, the efficiency against assuming known σ is quite low for the other procedures, especially when the target percentile is extreme, i.e., $p = 0.1$. Also as expected, assuming heteroscedasticity instead of homoscedasticity yields further drop in efficiency—but the drop is slight. In contrast, the logit-MLE shows a marked efficiency loss when compared to the least squares recursion procedures that use the continuous data. The fully sequential logit-MLE retrieves some information loss from the small-group logit-MLE, but the gain in efficiency does not completely recover the loss due to the use of dichotomised data.

Also in line with the asymptotic theory, setting $b = \beta$ and $\tilde{b} = \tilde{\beta}$ respectively for the least squares recursion and the logit-MLE generally yields better results than $b = \beta/2$ and $\tilde{b} = \tilde{\beta}/2$. The only exception is when $p = 0.1$ and $\theta = 0.75$, the logit-MLE with a low starting dose ($X_1 = 0.25$) has worse mean squared error when $\tilde{b} = \tilde{\beta}$ than when $\tilde{b} = \tilde{\beta}/2$. It is known that logit-MLE with a large \tilde{b} corresponds to small changes in subsequent doses; therefore, with a finite sample size, it will have difficulty climbing to a high θ if the starting dose is low, and can be improved with the use of a smaller \tilde{b} .

The impact of the starting dose X_1 on the operating characteristics is comparatively nuanced, although the logit-MLE tends to have smaller variance when the starting dose X_1 is closer to the target dose θ .

The second simulation study further studies the effects of group sizes. Specifically, we consider designs with a bigger group size, namely, $m = 5$ and $n = 9$, so that the total sample size remains 45. Using bigger group sizes can be appealing in practice because it reduces the study duration and administrative burdens. We also consider random group sizes generated by permuting $\{2, 2, 2, 3, 3, 3, 4, 4, 5, 5, 6, 6\}$ so that there are $n = 12$ groups and a total of 45 subjects in each simulated trial.

A bigger group size seems to have a slightly negative effect on the logit-MLE when comparing the results in Table 1 ($m = 1, 3$) and Table 2 ($m = 5$). Specifically, the fully sequential logit-MLE seems to outperform the small-group logit-MLE in finite sample size. Note that the asymptotic variance of the logit-MLE does not depend on group as long as the total sample size nm is the same. In contrast, the impact of group size on least squares recursion is relatively small. There is in fact slight improvement in relative efficiency of Case 2 and Case 3 against Case 1 when $m = 5$: this is in line with the fact that bigger group size improves asymptotic efficiency of the least squares recursion with unknown variance; cf. Fig. 1. The relative performance of the four procedures follows the same pattern under varying group sizes.

The third simulation study aims to examine the robustness of the least squares recursion when Z_{ij} is non-normal. While the methods use normality as the working assumption, we generated noises from other distributions with mean 0 and unit variance. Table 3 summarizes the results under the logistic distribution with mean 0 and scale 0.55, and the t -distribution with 6 degrees of freedom (scaled to have unit variance).

Overall, the least squares recursion procedures induce larger biases under misspecified distribution (cf. Table 1). However, the biases are generally small when compared to the variances. Interestingly, assuming heteroscedasticity (Case 2) seems to mitigate the increase in bias due to misspecification, and as a result, leads to smaller mean squared error than the procedure assuming homoscedasticity (Case 3). It is also important to note that the least squares recursion procedures are generally superior to the logit-MLE in terms of mean squared error, even though the latter did not require normality to be valid. This suggests that variability, rather than bias, is the limiting factor of performance when sample size ranges from small to moderate. In other words, the information retrieved via the use of continuous data outweighs the potential bias induced by misspecification. Having said this, we recommend using pilot data to assess the noise distribution in the planning stage; see Cheung and Elkind (2010) for example.

6. Concluding remarks

The contribution of this paper is two-fold. First, it provides a unified least squares recursion approach (2) for sequential quantile estimation using continuous data. Second, and importantly, in the context of this least squares recursion (2), we investigate the issue of variance modeling in the context of an important biomedical application in dose finding. By asymptotic comparison and simulation studies, we show that the efficiency loss due to nonparametric variance estimation is small when compared to parametric estimation under the correct model (i.e. variance is an identity function of dose). Furthermore, the simulation study suggests that nonparametric variance estimation leads to improved robustness when the normality assumption is violated.

For the non-sequential settings, Fedorov and Leonov (2004) give a detailed and insightful discussion on parameter estimation for normal data with unknown variance, and study the behaviors of an iterated estimator under a parametric model. They show that the iterated least squares estimators may not be efficient without adjustment; this may bear implications on the use of nonparametric variance estimation for which no adjustment is needed. Having said this, the focus of this paper differ from that of Fedorov and Leonov (2004) in two ways. First, we avoid parametric assumptions on the mean $M(x)$. Second, we focus on situations with sequential accrual of the data. The sequential nature of our problem renders the correctness of the parametric assumptions all the more crucial for the validity of statistical inference: in reality where modeling the variance function is difficult, the working assumption on $\sigma(x)$ has a direct impact on the design $\{X_j\}$ so that it is not possible to perform sensitivity analysis after the data are collected. As such, a misspecified $\sigma(x)$ will affect the final estimate of θ in an irreconcilable way, and thus parametric structure on the variance function should be avoided unless there are compelling reasons—and, as we show in this paper, the advantage of parametric estimation is very modest even when the assumption is correct.

Acknowledgments

This work was supported by NIH/NINDS grants R01 NS055809.

Appendix A. Proofs

This section provides the proofs of Lemma 1, Theorems 1 and 2.

Proof of Lemma 1. Applying Taylor's expansion in $\hat{\sigma}_n$ about σ gives

$$\hat{\sigma}_n - \sigma = \frac{1}{2\sigma}(\hat{\sigma}_n^2 - \sigma^2) - \frac{1}{8\sigma^3}(\hat{\sigma}_n^2 - \sigma^2)^2 + \frac{1}{16\sigma_n^{*5}}(\hat{\sigma}_n^2 - \sigma^2)^3$$

where σ_n^* is between σ and $\hat{\sigma}_n$. Therefore, we can rewrite (5) as

$$X_{n+1} = \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{b} \left\{ \frac{1}{n} \sum_{i=1}^n \bar{Y}_i + \frac{z_p}{2\sigma}(\hat{\sigma}_n^2 - \sigma^2) - \mu \right\} + \eta_n \quad (A1)$$

where

$$\eta_n = \frac{z_p}{8\sigma^3 b} \left\{ 1 - \frac{\sigma^3}{2\sigma_n^{*5}}(\hat{\sigma}_n^2 - \sigma^2) \right\} (\hat{\sigma}_n^2 - \sigma^2)^2.$$

Next, consider the design $\{X'_i\}$ generated with $X'_1 = X_1$ and

$$X'_{i+1} = X'_i - \frac{1}{ib} \left[\bar{Y}'_i + \frac{z_p}{2\sigma}(s_i^2 - \sigma^2) - \mu \right] + \xi_i$$

where $\bar{Y}'_i = M(X'_i) + \sigma \bar{Z}_i$. Multiplying i on both sides then gives

$$iX'_{i+1} - (i-1)X'_i = X'_i - \frac{1}{b} \left\{ \bar{Y}'_i + \frac{z_p}{2\sigma}(s_i^2 - \sigma^2) - \mu \right\} + i\xi_i.$$

Iterating the above equation, we get

$$X'_{n+1} = \frac{1}{n} \sum_{i=1}^n X'_i - \frac{1}{b} \left\{ \frac{1}{n} \sum_{i=1}^n \bar{Y}'_i + \frac{z_p}{2\sigma}(\hat{\sigma}_n^2 - \sigma^2) - \mu \right\} + \frac{1}{n} \sum_{i=1}^n i\xi_i. \quad (A2)$$

Matching the last terms in (A1) and (A2) gives

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 1 & 2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 2 & \cdots & n-1 & n \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix} = \begin{pmatrix} \eta_1 \\ 2\eta_2 \\ \vdots \\ n\eta_n \end{pmatrix}, \quad (A3)$$

and inverting (A3), we have $X_n \equiv X'_n$ and $\bar{Y}_n \equiv \bar{Y}'_n$ with $n\xi_n = m\eta_n - (n-1)\eta_{n-1}$. To complete the proof and show $\sum_{n=1}^{\infty} E(|\xi_n| | \mathcal{F}_{n-1}) < \infty$ a.s., it suffices to show that

$$\sum_{n=1}^{\infty} E(|\xi_n|) = E \left\{ \sum_{n=1}^{\infty} E(|\xi_n| | \mathcal{F}_{n-1}) \right\} < \infty.$$

Let $D_n = \hat{\sigma}_n^2 - \sigma^2$. Thus, $nD_n = (n-1)D_{n-1} + (s_n^2 - \sigma^2)$ and

$$\begin{aligned} & m\eta_n - (n-1)\eta_{n-1} \\ &= \frac{(n-1)z_p}{8n\sigma^3 b} \left[-D_{n-1}^2 + 2(s_n^2 - \sigma^2)D_{n-1} + \frac{(s_n^2 - \sigma^2)^2}{n-1} - \left\{ \frac{\sigma^3}{2\sigma_n^{*5}} \left(\frac{n-1}{n} \right)^2 - \frac{\sigma^3}{2\sigma_{n-1}^{*5}} \right\} nD_{n-1}^3 - \frac{3\sigma^3}{2n\sigma_n^{*5}} \left\{ (n-1)D_{n-1}^2 (s_n^2 - \sigma^2) + D_n \right. \right. \end{aligned}$$

Thus, $E(|m\eta_n - (n-1)\eta_{n-1}| | \mathcal{F}_{n-1}) \leq K_0|D_{n-1}| + K_1D_{n-1}^2 + K_2n|D_{n-1}|^3 + O(n^{-1})$ for some $K_0, K_1, K_2 > 0$. Since

$$E(D_n^2) = O(n^{-1}), E(|D_n|) \leq [E(D_n^2)]^{1/2} = O(n^{-1/2}), \text{ and } E(|D_n|^3) \leq [E(D_n^4)]^{3/4} = O(n^{-3/2}),$$

$$\sum_{n=1}^{\infty} E(|\xi_n|) = \sum_{n=1}^{\infty} O(n^{-3/2}) < \infty.$$

Proof of Theorem 1. Following from Condition 2 and recursion (6) in Lemma 1, there exist $K_3, K_4 > 0$ such that $E\{(X_{n+1} - \theta)^2 | \mathcal{F}_{n-1}\}$

$$(X_n - \theta)^2 - \frac{2}{nb}(X_n - \theta)(M(X_n) - \mu) + K_3E(|\xi_n| | \mathcal{F}_{n-1}) + K_4E(\xi_n^2 | \mathcal{F}_{n-1}) + O(n^{-2}).$$

It is easy to show that $\sum_{n=1}^{\infty} E(\xi_n^2 | \mathcal{F}_{n-1}) < \infty$ a.s. by verifying $\sum_n E(\xi_n^2) < \infty$. Therefore, Theorem 1 of Robbins and Siegmund (1971) implies that $\lim_{n \rightarrow \infty} (X_{n+1} - \theta)^2$ exists and $\sum_{n=1}^{\infty} n^{-1}(X_n - \theta)(M(X_n) - \mu) < \infty$ a.s. Since $(X_n - \theta)(M(X_n) - \mu) > 0$ under Condition 1, we conclude that $X_n \rightarrow \theta$ a.s.

Proof of Theorem 2. We will follow the approach of Sacks (1958). First, define

$$\gamma_{in} = \prod_{j=i+1}^n (1 - \rho^{j-1}) \text{ and } h_n = \left(\sum_{i=1}^n \rho^{2i-2} \gamma_{in}^2 \right)^{-1/2}, \quad (A4)$$

where $\rho := \beta/b$, with the following properties:

Property 1. $(1 + \varepsilon_i) i^\rho n^{-\rho} \leq \gamma_{in} \leq (1 + \varepsilon'_i) i^\rho n^{-\rho}$ where $\varepsilon_i, \varepsilon'_i \rightarrow 0$ as $i \rightarrow \infty$.

Property 2. $\lim_{n \rightarrow \infty} h_n \gamma_{in} = 0$ for fixed i and $\rho > 1/2$.

Property 3. $h_n \sim \{b(2\beta - b)\}^{1/2} \beta^{-1} n^{1/2}$.

Properties 1–3 are respectively Equation (2.3), Lemma 2, and Lemma 5 of Sacks (1958). Now, under Condition 3, we can rewrite (6) as

$$\begin{aligned} X_{n+1} - \theta &= (X_n - \theta) - \frac{1}{nb} \{ \beta(X_n - \theta) + \tau(X_n, \theta) + \bar{e}_n \} + \xi_n \\ \iff X_{n+1} - \theta &= \left(1 - \frac{\beta}{nb} \right) (X_n - \theta) - \frac{\tau(X_n, \theta)}{nb} - \frac{\bar{e}_n}{nb} + \xi_n \quad (A5) \end{aligned}$$

$$\iff X_{n+1} - \theta = \gamma_{0n}(X_1 - \theta) - \frac{1}{b} \sum_{i=1}^n i^{-1} \gamma_{in} \tau(X_i, \theta) - \frac{1}{b} \sum_{i=1}^n i^{-1} \gamma_{in} \bar{e}_i + \sum_{i=1}^n \gamma_{in} \xi_i. \quad (A6)$$

Equation (A6) is obtained by iteration of (A5). If we show that $h_n(X_{n+1} - \theta)$ is asymptotically normal with mean 0 and variance $\alpha_3 \sigma^2 (m\beta^2)^{-1}$, then the desired result will follow from Property 3. Corresponding to the terms in (A6), the limiting results

- i. $h_n \gamma_{0n}(X_1 - \theta) \rightarrow 0$ a.s.
- ii. $h_n b^{-1} \sum_i i^{-1} \gamma_{in} \tau(X_i, \theta) \rightarrow 0$ in probability
- iii. $h_n b^{-1} \sum_i i^{-1} \gamma_{in} \bar{e}_i \xrightarrow{\mathcal{L}} N(0, \alpha_3 \sigma^2 m^{-1} \beta^{-2})$

can be derived by mimicking the proof of Theorem 1 in Sacks (1958) under an additional assumption:

Condition 2b. There exists $0 < C_0 < C_1$ such that $C_0|x - \theta| \leq |M(x) - \mu| \leq C_1|x - \theta|$. Finally, following from Properties 1–3,

$$h_n \sum_{i=1}^n \gamma_{in} \xi_i \sim \{b(2\beta - b)\}^{1/2} \beta^{-1} n^{-1/2} n^{1-\rho} \sum_{i=1}^n i^{\rho-1} (i\xi_i) \rightarrow 0$$

when $\rho > 1$ by Kronecker Lemma; and,

$$h_n \sum_{i=1}^n \gamma_{in} \xi_i \sim \{b(2\beta - b)\}^{1/2} \beta^{-1} n^{-1/2-\rho} n \eta_n \sum_{i=1}^n \left\{ \frac{i\eta_i - (i-1)\eta_{i-1}}{n\eta_n} \right\} i^{\rho-1} \rightarrow 0$$

if $1/2 < \rho < 1$, because $\sum_{i=1}^n i\eta_i - (i-1)\eta_{i-1} = n\eta_n$. The desired result is thus obtained under Conditions 1–3 and 2b.

Suppose now that $M(x)$ satisfies Condition 1–3 but not 2b. Since $b < 2\beta$, there exists $t > 0$ such that $b < 2(\beta - t)$. Let $C_0 = \beta - t < C_1$. Then under Condition 3, we can find $\delta > 0$ such that $C_0|x - \theta| \leq |M(x) - \mu| \leq C_1|x - \theta|$ for $|x - \theta| \leq \delta$.

Next, define $M_\delta(x) = M(x)$ if $|x - \theta| \leq \delta$ and $M_\delta(x) = \mu + C_0(x - \theta)$ if $|x - \theta| > \delta$, and let $X_1^{(\delta)} = X_{N_\delta+1}$ and

$$X_{n+1}^{(\delta)} - \theta = (X_n^{(\delta)} - \theta) - \frac{1}{(n+N_\delta)b} \{M_\delta(X_n^{(\delta)}) - \mu\} - \frac{\bar{e}_n + N_\delta}{(n+N_\delta)b} + \xi_{n+N_\delta}$$

where $N_\delta > 0$ is determined such that, for a given u , $\text{pr}(|X_n - \theta| \leq \delta \text{ for all } n \leq N_\delta) > 1 - u$. We can find such an N_δ because $X_n \rightarrow \theta$ a.s. under Conditions 1 and 2. Observing that $M_\delta(x)$ satisfies Condition 2b, we can verify $n^{1/2}(X_n^{(\delta)} - \theta) \xrightarrow{\mathcal{L}} N(0, \alpha_1 \alpha_3 \sigma^2)$. Furthermore, $X_{n+N_\delta}^{(\delta)} \equiv X_n^{(\delta)}$ a.s. on the event $\{|X_{n+N_\delta} - \theta| \leq \delta \text{ for all } n \leq N_\delta\}$. Thus,

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \Pr\{n^{1/2}(X_n - \theta) < w\} &= \limsup_{n \rightarrow \infty} \Pr\{(n+N_\delta)^{1/2}(X_{n+N_\delta} - \theta) < w\} \\
&\leq \limsup_{n \rightarrow \infty} \Pr\{(n+N_\delta)^{1/2}(X_n^{(\delta)} - \theta) < w, |X_{n+N_\delta} - \theta| \leq \delta, n \geq 1\} + u \\
&\leq \lim_{n \rightarrow \infty} \Pr\{(n+N_\delta)^{1/2}(X_n^{(\delta)} - \theta) < w\} + u.
\end{aligned}$$

Similarly, we obtain

$$\liminf_{n \rightarrow \infty} \Pr\{n^{1/2}(X_n - \theta) < w\} \geq \lim_{n \rightarrow \infty} \Pr\{(n+N_\delta)^{1/2}(X_n^{(\delta)} - \theta) < w\} - u.$$

Since u is arbitrary, we conclude that $n^{1/2}(X_n - \theta)$ and $n^{1/2}(X_n^{(\delta)} - \theta)$ have the same limiting distribution. This completes the proof of Theorem 2.

Appendix B. Asymptotic irrelevance of truncation

This section discusses the asymptotic equivalence of the truncated least squares recursion and its non-truncated counterpart.

Under Cases 1 and 2, the least squares recursion is identical to the Robbins-Monro procedure, whose truncated version is known to be asymptotically equivalent to the non-truncated design; see Lai and Robbins (1981) for example.

Under Case 3, the true standard deviation σ is consistently estimated by $\hat{\sigma}_n$ and we can rewrite (2) as

$$\frac{1}{n} \sum_{i=1}^n \tilde{Y}_i + z_p \sigma - \{t_0 + b(X_i - \hat{\theta}_n)\} = -z_p(\hat{\sigma}_n - \sigma).$$

Since $\hat{\sigma}_n - \sigma \rightarrow 0$ with probability one, by Martingale convergence theorem, we have

$$\frac{1}{n} \sum_{i=1}^n f(X_i) - \{t_0 + b(X_i - \hat{\theta}_n)\} \rightarrow 0$$

with probability one. Therefore, we can use the arguments in the proof of Theorem 2 in Ying and Wu (1997) to show that $X_{n+1} = \hat{\theta}_n$ eventually with probability 1. Thus, in view of asymptotic efficiency, we can focus on the non-truncated design formed by (2) and (3) recursively.

References

Cheung YK. Coherence principles in dose-finding studies. *Biometrika*. 2005; 92:863–873.

- Cheung YK. Stochastic approximation and modern model-based designs for dose-finding clinical trials. *Statistical Science*. 2010; 25:191–201. [PubMed: 21197369]
- Cheung YK, Elkind MSV. Stochastic approximation with virtual observations for dose-finding on discrete levels. *Biometrika*. 2010; 97:109–121. [PubMed: 23049118]
- Eichhorn BH, Zacks S. Sequential search of an optimal dosage, I. *J. Am. Statist. Assoc.* 1973; 68:594–598.
- Fedorov VV, Leonov S. Parameter estimation for models with unknown parameters in variance. *Comm. Statist.* 2004; 33:2627–2657.
- Lai TL, Robbins H. Adaptive design and stochastic approximation. *Ann. Statist.* 1979; 7:1196–1221.
- Lai TL, Robbins H. Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*. 1981; 56:329–360.
- O’Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics*. 1990; 46:33–48. [PubMed: 2350571]
- Robbins H, Monro S. A stochastic approximation method. *Ann. Math. Statist.* 1951; 22:400–407.
- Robbins, H.; Siegmund, D. A convergence theorem for non-negative almost supermartingales and some applications. In: Rustagi, J., editor. *Optimizing Methods in Statistics*. New York: Academic Press; 1971. p. 237-257.
- Sack J. Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.* 1958; 29:373–405.
- Wu CFJ. Efficient sequential designs with binary data. *J. Am. Statist. Assoc.* 1985; 80:974–984.
- Ying Z, Wu CFJ. An asymptotic theory of sequential designs based on maximum likelihood recursion. *Statistica Sinica*. 1997; 7:75–91.

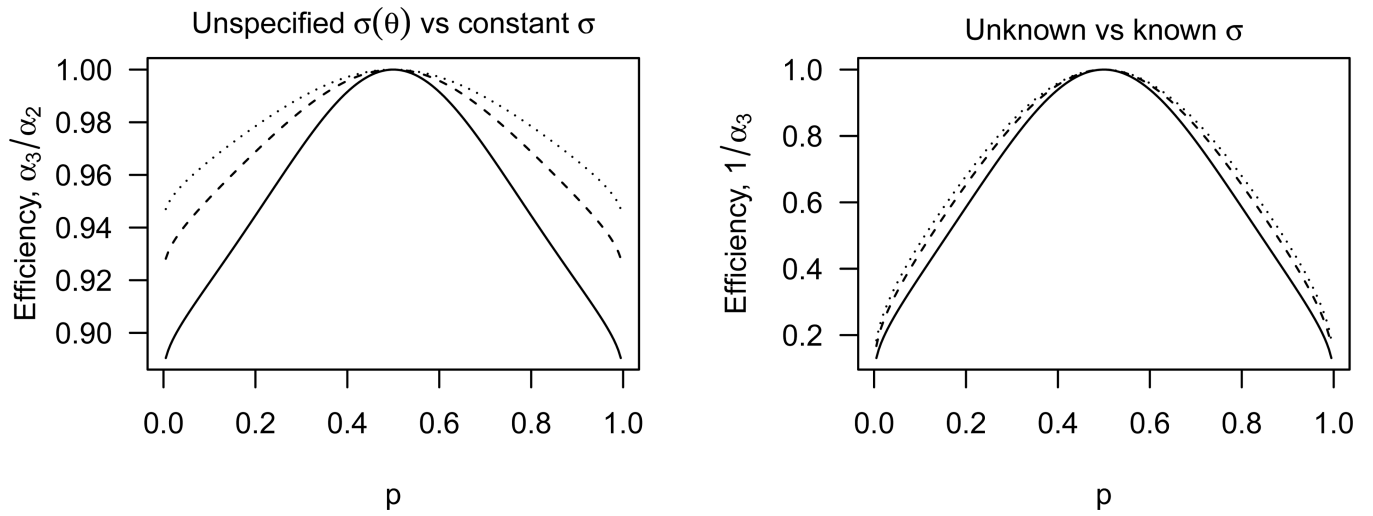


Fig. 1. Asymptotic relative efficiencies under homoscedasticity for $m = 2$ (solid), 3 (dashed), and 4 (dotted).

Bias($\times 10^2$) and variance($\times 10^2$) of the least squares recursion and the logit-MLE with $m = 3$, $n = 15$, and the fully sequentially logit-MLE(f), i.e., $m = 1$, $n = 45$. The mean squared error ratio (rmse) is calculated relative to the method assuming known σ .

Table 1

p	θ	Method assumes	$X_1 = 0.25, b = \beta^T$			$X_1 = 0.50, b = \beta^T$			$X_1 = 0.50, b = 0.5\beta^T$			$X_1 = 0.50, b = 0.5\beta^T$		
			bias	var	rmse	bias	var	rmse	bias	var	rmse	bias	var	rmse
0.1	0.25	Known σ	-0.05	1.05	—	-0.10	1.51	—	-0.06	1.05	—	-0.10	1.52	—
		Unspecified σ	-0.09	2.11	0.50	-0.20	2.99	0.50	-0.09	2.11	0.50	-0.21	2.99	0.50
		Constant σ	-0.04	2.03	0.52	-0.18	2.88	0.52	-0.05	2.03	0.52	-0.19	2.88	0.52
		logit-MLE	-0.04	2.71	0.39	-0.44	3.09	0.46	-0.12	2.38	0.44	-0.54	3.01	0.46
		logit-MLE(f)	-0.02	2.59	0.41	-0.35	3.19	0.46	-0.11	2.41	0.44	-0.43	3.11	0.46
0.50	0.25	Known σ	0.00	1.03	—	0.00	1.48	—	-0.01	1.03	—	0.00	1.48	—
		Unspecified σ	-0.01	2.25	0.46	-0.01	3.33	0.45	-0.01	2.25	0.46	-0.01	3.33	0.45
		Constant σ	0.04	2.14	0.48	0.00	3.15	0.47	0.03	2.14	0.48	-0.01	3.15	0.47
		logit-MLE	-0.11	3.60	0.29	-0.20	4.88	0.30	-0.09	3.13	0.33	-0.43	4.27	0.33
		logit-MLE(f)	-0.09	3.39	0.30	-0.21	4.58	0.32	-0.09	3.09	0.33	-0.40	4.17	0.34
0.75	0.25	Known σ	0.10	1.04	—	0.12	1.49	—	0.04	1.03	—	0.09	1.48	—
		Unspecified σ	0.13	2.12	0.49	0.23	3.01	0.49	0.09	2.12	0.49	0.19	3.01	0.49
		Constant σ	0.18	2.02	0.51	0.24	2.88	0.51	0.13	2.02	0.51	0.20	2.88	0.51
		logit-MLE	-1.35	4.96	0.15	-0.29	4.74	0.31	-0.19	3.11	0.33	-0.04	4.11	0.36
		logit-MLE(f)	-0.98	4.32	0.20	-0.09	4.24	0.35	-0.15	2.96	0.35	-0.11	4.04	0.37
0.2	0.25	Known σ	-0.03	0.81	—	-0.06	1.16	—	-0.04	0.81	—	-0.07	1.16	—
		Unspecified σ	-0.05	1.23	0.66	-0.11	1.77	0.65	-0.06	1.23	0.66	-0.12	1.78	0.65
		Constant σ	-0.02	1.19	0.68	-0.10	1.72	0.67	-0.02	1.19	0.68	-0.11	1.73	0.67
		logit-MLE	-0.04	1.59	0.51	-0.24	2.13	0.53	-0.06	1.53	0.53	-0.27	2.17	0.52
		logit-MLE(f)	-0.04	1.57	0.51	-0.21	2.13	0.54	-0.07	1.58	0.51	-0.24	2.14	0.53
0.50	0.25	Known σ	0.00	0.79	—	0.01	1.13	—	0.00	0.79	—	0.00	1.13	—
		Unspecified σ	0.00	1.22	0.65	0.00	1.77	0.64	0.00	1.22	0.65	0.00	1.77	0.64
		Constant σ	0.03	1.18	0.67	0.01	1.70	0.66	0.02	1.18	0.67	0.00	1.70	0.66
logit-MLE	0.00	1.84	0.43	-0.15	2.37	0.47	-0.05	1.65	0.48	-0.19	2.32	0.48		

p	θ	Method assumes	$X_1 = 0.25, b = \beta^\dagger$		$X_1 = 0.25, b = 0.5\beta^\ddagger$		$X_1 = 0.50, b = \beta^\dagger$		$X_1 = 0.50, b = 0.5\beta^\ddagger$					
			bias	var	bias	var	bias	var	bias	var	bias	var	rmse	rmse
		logit-MLE(f)	0.00	1.81	0.44	-0.16	2.32	0.48	-0.05	1.68	0.47	-0.17	2.33	0.48
0.75	Known σ		0.08	0.80	—	0.09	1.15	—	0.03	0.80	—	0.07	1.14	—
	Unspecified σ		0.10	1.23	0.66	0.14	1.77	0.65	0.05	1.22	0.65	0.11	1.76	0.65
	Constant σ		0.13	1.19	0.67	0.15	1.71	0.67	0.08	1.18	0.67	0.11	1.70	0.67
	logit-MLE		0.11	2.09	0.39	0.19	2.52	0.45	0.00	1.77	0.45	-0.01	2.34	0.49
	logit-MLE(f)		0.09	1.90	0.42	0.06	2.44	0.47	0.00	1.75	0.46	-0.06	2.33	0.49

$^\dagger b \sim \beta$ for logit-MLE;

$^\ddagger b \sim 0.5 \beta$.

Bias($\times 10$) and variance($\times 10^2$) of the least squares recursion and the logit-MLE with starting dose $X_1 = 0.25$. The mean squared error ratio (rmse) is calculated relative to the method assuming known σ .

Table 2

p	θ	Method assumes	$m = 5, n = 9$ $b = \beta^*$				$m = 5, n = 9$ $b = 0.5\beta^*$				varying group sizes $b = 0.5\beta^*$			
			bias	var	rmse		bias	var	rmse		bias	var	rmse	
0.1	0.25	Known σ	-0.04	1.04	—	-0.04	1.17	—	-0.15	1.55	—	-0.11	1.73	—
		Unspecified σ	-0.06	1.94	0.53	-0.05	2.42	0.48	-0.22	2.80	0.55	-0.20	3.40	0.51
		Constant σ	-0.02	1.90	0.55	0.00	2.30	0.51	-0.21	2.74	0.57	-0.17	3.24	0.54
		logit-MLE	-0.02	2.65	0.39	-0.03	2.71	0.43	-0.56	3.12	0.46	-0.48	3.32	0.49
0.50	0.25	Known σ	0.01	1.03	—	0.02	1.16	—	-0.01	1.53	—	0.01	1.72	—
		Unspecified σ	0.02	2.03	0.50	0.02	2.68	0.43	0.00	3.03	0.50	0.00	4.04	0.43
		Constant σ	0.06	1.98	0.52	0.08	2.52	0.46	0.01	2.96	0.52	0.01	3.81	0.45
		logit-MLE	-0.09	3.71	0.28	-0.07	3.73	0.31	-0.22	5.17	0.29	-0.22	5.26	0.32
0.75	0.25	Known σ	0.15	1.04	—	0.14	1.18	—	0.18	1.57	—	0.17	1.75	—
		Unspecified σ	0.19	1.97	0.53	0.18	2.45	0.48	0.30	2.80	0.55	0.28	3.48	0.50
		Constant σ	0.23	1.92	0.54	0.23	2.32	0.51	0.31	2.74	0.56	0.30	3.33	0.52
		logit-MLE	-1.57	5.26	0.14	-1.49	5.16	0.16	-0.48	5.25	0.29	-0.34	5.05	0.34
0.2	0.25	Known σ	-0.02	0.80	—	-0.02	0.91	—	-0.10	1.19	—	-0.08	1.35	—
		Unspecified σ	-0.03	1.15	0.70	-0.03	1.42	0.64	-0.13	1.70	0.70	-0.12	2.07	0.65
		Constant σ	0.00	1.13	0.71	0.01	1.36	0.67	-0.13	1.68	0.71	-0.11	1.99	0.67
		logit-MLE	-0.04	1.60	0.50	-0.02	1.57	0.58	-0.34	2.16	0.53	-0.29	2.18	0.60
0.50	0.25	Known σ	0.01	0.79	—	0.02	0.90	—	-0.01	1.16	—	0.01	1.31	—
		Unspecified σ	0.02	1.15	0.69	0.03	1.43	0.63	0.01	1.67	0.69	0.01	2.13	0.62
		Constant σ	0.04	1.13	0.70	0.06	1.37	0.65	0.01	1.65	0.71	0.01	2.04	0.64
		logit-MLE	0.01	1.85	0.43	0.03	1.84	0.49	-0.19	2.51	0.46	-0.18	2.54	0.51
0.75	0.25	Known σ	0.13	0.81	—	0.12	0.92	—	0.14	1.21	—	0.14	1.36	—
		Unspecified σ	0.15	1.16	0.69	0.14	1.43	0.64	0.20	1.71	0.70	0.20	2.11	0.64
		Constant σ	0.17	1.14	0.70	0.18	1.37	0.66	0.21	1.69	0.71	0.20	2.04	0.66

Watermark-text

Watermark-text

Watermark-text

p	θ	Method assumes	$m = 5, n = 9$ $b = \beta^{\dagger}$		varying group sizes $b = \beta^{\dagger}$		$m = 5, n = 9$ $b = 0.5\beta^{\ddagger}$		varying group sizes $b = 0.5\beta^{\ddagger}$					
			bias	var	rmse	bias	var	rmse	bias	var	rmse			
		logit-MLE	0.09	2.37	0.35	0.09	2.23	0.42	0.28	2.66	0.45	0.22	2.67	0.51

$\dagger b_{\sim} = \beta_{\sim}$ for logit-MLE;

$\ddagger b_{\sim} = 0.5 \beta_{\sim}$.

Bias($\times 10^2$) and variance($\times 10^2$) of the least squares recursion and the logit-MLE with $m = 3$, $n = 15$, and the fully sequentially logit-MLE(f), i.e., $m = 1$, $n = 45$, with starting dose $X_1 = 0.25$. The mean squared error ratio (rmse) is calculated relative to the method assuming known σ .

Table 3

p	θ	Method	Logistic, $b = \beta^T$			$t_6, b = \beta^T$			Logistic, $b = 0.5\beta^T$			$t_6, b = 0.5\beta^T$		
			bias	var	rmse	bias	var	rmse	bias	var	rmse	bias	var	rmse
0.1	0.25	Known σ	-0.50	1.00	—	-0.74	0.90	—	-0.56	1.43	—	-0.81	1.29	—
		Unspecified σ	-0.30	2.11	0.57	-0.40	2.02	0.67	-0.40	2.87	0.58	-0.50	2.78	0.64
		Constant σ	-0.40	2.10	0.55	-0.54	2.02	0.62	-0.53	2.79	0.57	-0.68	2.70	0.62
		logit-MLE	-0.04	2.74	0.46	-0.08	2.57	0.56	-0.49	3.16	0.51	-0.52	3.12	0.58
		logit-MLE(f)	-0.06	2.64	0.47	-0.11	2.56	0.56	-0.43	3.11	0.53	-0.46	3.04	0.60
		Known σ	-0.45	1.01	—	-0.69	0.93	—	-0.46	1.45	—	-0.70	1.40	—
0.50	0.25	Unspecified σ	-0.25	2.40	0.49	-0.36	2.36	0.57	-0.26	3.51	0.46	-0.38	3.53	0.51
		Constant σ	-0.37	2.49	0.46	-0.55	2.58	0.49	-0.43	3.63	0.44	-0.62	3.75	0.46
		logit-MLE	-0.18	3.71	0.32	-0.19	3.68	0.38	-0.31	5.06	0.32	-0.36	4.91	0.38
		logit-MLE(f)	-0.18	3.62	0.33	-0.16	3.47	0.40	-0.37	4.76	0.34	-0.35	4.61	0.40
		Known σ	-0.36	1.03	—	-0.60	0.95	—	-0.37	1.50	—	-0.63	1.43	—
		Unspecified σ	-0.09	2.30	0.50	-0.20	2.31	0.56	-0.02	3.29	0.50	-0.15	3.36	0.54
0.75	0.25	Constant σ	-0.21	2.40	0.47	-0.39	2.57	0.48	-0.18	3.42	0.47	-0.39	3.71	0.47
		logit-MLE	-1.19	4.89	0.18	-1.22	4.89	0.21	-0.27	4.92	0.33	-0.35	4.90	0.36
		logit-MLE(f)	-0.90	4.38	0.22	-0.95	4.36	0.25	-0.12	4.50	0.36	-0.15	4.46	0.41
		Known σ	-0.47	0.78	—	-0.62	0.72	—	-0.52	1.12	—	-0.68	1.05	—
		Unspecified σ	-0.35	1.23	0.74	-0.43	1.17	0.82	-0.43	1.73	0.72	-0.51	1.66	0.78
		Constant σ	-0.42	1.24	0.71	-0.53	1.20	0.75	-0.52	1.72	0.70	-0.64	1.67	0.72
0.2	0.25	logit-MLE	-0.08	1.47	0.68	-0.11	1.35	0.82	-0.30	1.98	0.67	-0.32	1.92	0.74
		logit-MLE(f)	-0.09	1.46	0.68	-0.11	1.34	0.82	-0.28	1.94	0.69	-0.28	1.90	0.76
		Known σ	-0.44	0.78	—	-0.58	0.72	—	-0.44	1.11	—	-0.60	1.07	—
		Unspecified σ	-0.31	1.27	0.71	-0.39	1.22	0.77	-0.32	1.83	0.68	-0.40	1.81	0.72
		Constant σ	-0.38	1.30	0.67	-0.50	1.33	0.67	-0.42	1.87	0.64	-0.56	1.95	0.63
		logit-MLE	-0.05	1.72	0.56	-0.05	1.58	0.67	-0.23	2.26	0.56	-0.25	2.18	0.64

p	θ	Method	Logistic, $b = \beta^\dagger$			$t_6, b = \beta^\dagger$			Logistic, $b = 0.5\beta^\ddagger$			$t_6, b = 0.5\beta^\ddagger$		
			bias	var	rmse	bias	var	rmse	bias	var	rmse	bias	var	rmse
		logit-MLE(f)	-0.08	1.67	0.58	-0.10	1.52	0.69	-0.23	2.15	0.59	-0.25	2.04	0.68
0.75	Known σ		-0.36	0.79	—	-0.51	0.74	—	-0.38	1.14	—	-0.54	1.10	—
	Unspecified σ		-0.22	1.28	0.69	-0.29	1.23	0.76	-0.20	1.85	0.68	-0.30	1.84	0.72
	Constant σ		-0.28	1.31	0.66	-0.40	1.33	0.67	-0.30	1.89	0.65	-0.45	1.96	0.64
	logit-MLE		0.04	2.06	0.45	0.04	1.98	0.50	0.10	2.47	0.52	0.07	2.39	0.58
	logit-MLE(f)		0.05	1.85	0.50	0.04	1.76	0.57	-0.04	2.29	0.56	-0.06	2.20	0.63

$^\dagger b_{\sim} = \beta_{\sim}$ for logit-MLE;

$^\ddagger b_{\sim} = 0.5\beta_{\sim}$.