AN ADAPTIVE GROUP-TESTING PROCEDURE FOR ESTIMATING PROPORTIONS

by

Jacqueline M. Hughes-Oliver and William H. Swallow

Institute of Statistics Mimeo Series Number 2244

February, 1993

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

An Adaptive Group-Testing Procedure for Estimating Proportions

Jacqueline M. HUGHES-OLIVER and William H. SWALLOW

Department of Statistics, North Carolina State University,
Raleigh, North Carolina 27695–8203, U.S.A.

February 1993

Abstract

Group-testing procedures begin by testing units in groups or batches rather than one at a time. This scheme benefits from the possibility that a single measurement/response may simultaneously indicate the satisfactory state of several individual units. Indeed, the savings afforded by group-testing procedures can be substantial, provided a good choice of the group size is made. Unfortunately, careful choice of the group size for estimating the proportion of defectives/infected in a population requires some *a priori* knowledge of that proportion, which is not always available.

This paper describes an adaptive group-testing estimation scheme that is not as sensitive to having good *a priori* information about the proportion (p) to be estimated as is the currently-used group-testing estimation scheme. The comparison between the two procedures is based on asymptotic and small-sample relative efficiency. In large samples ($\geq$ 100), the adaptive estimator is at least comparable to the nonadaptive estimator, and can be much more efficient than the nonadaptive estimator. In small samples, the adaptive estimator is the more efficient when the *a priori* value for p is less than the true value (possibly thousands of times as efficient as the nonadaptive estimator), with the cost being a slight loss in efficiency when the *priori* value is equal to p (and hence no adaptation is needed) or larger. Thus, the adaptive estimator is recommended with the restriction that adaptations (adjustments in group size as data are collected) are based on at least ten measurements/responses to provide adequate information

on which to base a change.

Furthermore, Monte Carlo simulations show that the adaptive estimator is sufficiently well-behaved that a normal approximation is adequate for performing hypothesis tests, as well as generating confidence intervals with confidence coefficients as high as 0.95.

KEY WORDS: mean squared error; asymptotic variance; multistage procedure; maximum likelihood estimator; relative efficiency.

1 Introduction

Group testing has long been recognized as a sampling scheme that can provide substantial benefits. Under this sampling scheme, rather than obtaining measurements on individuals, simultaneous measurement is obtained for a group of, say, five individuals. The measurement is usually taken to be dichotomous with a positive test indicating the presence of one or more positive individuals in the group; some scheme for retesting individuals in positive groups will be required if the positive individuals are to be identified. This type of sampling has been shown to reduce the cost of classifying all members of a population according to whether they possess a certain trait or not (see, for example, Dorfman, 1943). It has also been used to reduce the mean squared error of estimates of proportions in cases where the true proportion does not exceed $\frac{2}{3}$ (see, for example, Thompson (1962), Sobel and Elashoff (1975), Swallow (1985, 1987)).

Although group testing has been shown to have the potential to reduce "costs" (monetary, mean squared error, or some other measure), perhaps considerably, realizing those potential benefits depends on using an appropriate group size. (The properties of group-testing procedures are very sensitive to the choice of group size, and an unfortunate choice can be disastrous; see Thompson (1962), Swallow (1985), Hughes-Oliver and Swallow (1992)). In using group testing to estimate a proportion p, several recommendations have been made that attempt to overcome this deficiency. These include:

1. use several group sizes and combine the data to produce a single estimator;

2. employ double sampling, using information from the first stage in choosing the group size to be used throughout the second, and using only the second stage in determining the final estimate of p;

3. perform Bayesian analysis;

4. use an *a priori* upper bound for p in selecting the group size, i.e., use the group size that would be optimal if p equaled the *a priori* upper bound;

5. adapt/adjust the group size from time to time throughout the testing phase, using all accumulated data, and obtaining a final estimate of p that is based on all the data collected.

The first two recommendations offer protection against a very poor estimate, but cannot be optimal since they require a division and use of the allotted number of tests in ways that cannot maximally benefit the final estimate. The third recommendation will only be advantageous if enough information is available to determine an adequate prior distribution for p (Griffiths, 1972). The recommendation to use an *a priori* upper bound on p seems, at this point, to be dominant (Thompson, 1962; Swallow, 1985). Indeed, provided the *a priori* upper bound is not less than the actual value of p, the worst case is that individual-testing (a group size of one) would be used, and the best case is that the optimal group size (the group size that minimizes the mean squared error, for a given sample size and the true p) would be selected. This recommendation follows the age-old statistical "rule" that it is better to be conservative. There are, however, two problems with this recommendation. The first is that this conservative approach can seriously reduce the realized benefits of group-testing. The second, and more serious, occurs when this *a priori* upper bound isn't an upper bound at all, but an underestimate; in this case too large a group size will be used, and the mean squared error of the estimator will be (perhaps grossly) inflated. In light of the deficiencies of the first four recommendations, the fifth—adaptation—seems worth exploring.

In this paper we describe an adaptive scheme and compare it to the usual nonadaptive scheme.

Section 2 contains descriptions of the usual nonadaptive and the (new) adaptive estimation schemes, and their asymptotic properties. Section 3 provides an illustration of how the procedures would be used to estimate a proportion. Section 4 compares the two estimators presented in Section 2; this comparison is based on both large-sample (asymptotic) and small-sample results. Section 5 investigates the tail behavior of the standardized adaptive estimator and its effects on the coverage probability of large-sample, normal-theory confidence intervals and hypothesis tests for the proportion p.

2 The group-testing estimators and their asymptotic properties

The proportion p, $0 < p < 1$, of individuals in an infinite population that possess a certain trait is to be estimated under the assumptions that: 1) the probability of having this trait is the same for all members of the population; and 2) the test used to reveal this trait is 100% accurate. The first assumption rules out situations where, e.g., clustering of the trait is possible, as in the case of an infectious disease. The second assumption ensures that individuals (or groups) will not be incorrectly labeled as having the trait, and individuals (groups) with the trait will be identified as such; that is, there are no false negatives and no false positives.

2.1 The usual nonadaptive estimator

The nonadaptive group-testing estimation scheme tests N groups, each of size k, then labels each of these groups according to whether one or more individuals in the group possesses the trait. (Actually, Walter, Hildreth and Beaty (1980) and Le (1981) also consider the case of testing N groups with, possibly, different group sizes. However, their group sizes are not determined sequentially; rather, they are chosen beforehand, or determined by uncontrolled circumstances.) If a group is labeled as possessing the trait, the interpretation is that at least one member of the group has the trait. Under the assumptions stated above, the number of groups, X, possessing the trait has a binomial distribution with parameters

N and $1-(1-p)^k$. Using maximum likelihood estimation, the nonadaptive estimator of p is

$$\widehat{p} = 1 - \left(1 - \frac{X}{N}\right)^{1/k}, \tag{1}$$

which is positively biased when $k > 1$ and has mean squared error

$$MSE(\widehat{p}; k, p, N) = (1-p)^2 + \sum_{i=0}^{N} \left(1 - \frac{i}{N}\right)^{\frac{1}{k}} \left[\left(1 - \frac{i}{N}\right)^{\frac{1}{k}} - 2(1-p)\right] \binom{N}{i} \delta_k^i (1-\delta_k)^{N-i}, \tag{2}$$

where $\delta_k = 1-(1-p)^k$. Asymptotically, $\widehat{p}$ is unbiased, normally distributed and efficient. That is, for fixed k and $N \to \infty$,

$$\sqrt{N}(\widehat{p} - p) \xrightarrow{d} Normal\left(0, \frac{1-(1-p)^k}{k^2(1-p)^{k-2}}\right) \tag{3}$$

(see Thompson, 1962). Notice that if the group size is one, that is, $k = 1$, individual units are being tested, and (1) is the usual binomial estimator $\widehat{p} = X/N$, which is unbiased with variance $p(1-p)/N$.

In practice, some prior value for the true p, say p_0, can be used to determine the group size, taking k to be the group size that minimizes the mean squared error of $\widehat{p}$, evaluated at $p = p_0$. That is, choose

$$\begin{aligned} k = k(N) &= \arg\min_l MSE(\widehat{p}; l, p_0, N) \\ &= \arg\min_l Q_N(l; p_0), \end{aligned}$$

where $Q_N(l; p_0) = N \times MSE(\widehat{p}; l, p_0, N)$. Clearly, choosing k in this manner will alter the distributional result in (3), since k is now a function of N. Hence, it is of interest to determine the limiting value of the k chosen in this manner, which is given by the following result:

Result 1 *Define*

$$\overline{k} = \arg\min_l \left\{\frac{1-(1-p_0)^l}{l^2(1-p_0)^{l-2}}\right\}. \tag{4}$$

If there exists N_{p_0} such that $k(N) \geq 1$ for all $N > N_{p_0}$, and the first derivative of $Q_N(l; p_0)$ with respect to l vanishes at most once, then $lim_{N\to\infty} k(N) = \overline{k}$.

[Appendix A contains the proof for the general multistage procedure. Result 1 is obtained as a special case by letting $i = 0$, $N_1 = N$ and $p = p_0$.] This result has been used in the literature as true, but (to our knowledge) never proved. It says that, under mild conditions, as N tends to infinity, the group sizes obtained from minimizing the mean squared error converge to the group size obtained from minimizing the asymptotic variance. Indeed, this is an expected, but not immediate, result that has the following desirable implication. Although for each N, $k(N)$ has the potential for being different, the asymptotic distribution of the estimator $\widehat{p}$ has the same form as in (3), with $\overline{k}$ in place of k. That is,

$$\sqrt{N}(\widehat{p} - p) \xrightarrow{d} Normal\left(0, \frac{1-(1-p)^{\overline{k}}}{\overline{k}^2(1-p)^{\overline{k}-2}}\right).$$

(Note that $\overline{k}$ is the asymptotically optimal group size only if $p_0 = p$.)

In order to illustrate the mildness of the last assumption in Result 1, Figure 1 shows the typical shape of $Q_N(l; p_0)$ as a function of l, for $1 \leq l \leq 500$, $p_0 = 0.01$ and $N = 20$. [Actually, the graph shows $\log[Q_N(l; p_0)]$, not $Q_N(l; p_0)$. In this way the shape of the curve, and not its particular values, is emphasized.] The function is not convex, but it is convex in a region containing the minimum. Outside of this region the function is strictly increasing. Thus, it is quite reasonable to assume that the function has a derivative that vanishes at most once.

2.2 The adaptive estimator

The adaptive estimator is obtained by testing groups in stages, and updating the group size from one stage to the next. Hence, one might have N_1 groups, each of size k_1, tested in the first stage; N_2 groups, each of size k_2 (k_2 depending on the results of the first stage), tested in the second stage; N_3 groups, each of size k_3 (k_3 depending on the results of the first two stages), tested in the third stage, and so on (Griffiths, 1972; Sobel and Elashoff, 1975). Indeed, this is a very general scheme that allows the number N_i of groups for each stage to be arbitrary, but known before the experiment begins, while the group sizes k_i are determined sequentially during experimentation. The challenge comes in deciding how one should update the group size in a manner that will lead to using a group size as close to the optimal group size as possible. Should one determine the next group size using only the information from the most recent stage, or the information from all previous stages? Should the updating be based on a maximum likelihood estimate of the proportion, or should it be based on a more robust estimate? Should the updating be based on the number N_i of tests to be performed in the next stage, or should it use the total number N of tests over all stages?

The adaptive scheme presented here updates the group size between stages based on (i) the maximum likelihood estimate of p obtained using the data from all previous stages and (ii) the number of tests to be performed in the next stage. The group size selected is the one that minimizes the mean squared error of the estimate of p that would be obtained if only the data from the next stage were used, with p replaced in the mean squared error formula by the most recent maximum likelihood estimate of p. The group size for the first stage is still based on an *a priori* value p_0. An exact description of a two-stage adaptive estimator follows. The description and consequent discussion is limited to two stages partly for simplicity and partly because asymptotic results suggest that two stages yield optimal results.

To allow for direct comparison with the nonadaptive estimator, a total of N tests will be performed in two stages, such that $N_1 = \lambda N$ tests are performed in the first stage and $N_2 = (1-\lambda)N$ tests in the second stage, where λ is assumed to be known before the experiment begins. As with the nonadaptive scheme, the group size k_1 for the first stage is based on some prior value p_0 for the true p, and the number of tests λN to be performed in the first stage, and is calculated as

$$k_1 = \arg\min_l Q_{\lambda N}(l; p_0),$$

where the function Q is defined in Section 2.1. Stage 1 then proceeds by testing λN groups, each of size k_1, for the trait of interest. The number X_1 of the λN groups with the trait has a binomial distribution with parameters λN and $1-(1-p)^{k_1}$. At the end of the first stage, an intermediate maximum likelihood estimate of p is determined as

$$\widehat{p}_1 = \widehat{p}_1(x_1) = \widehat{p}_1(X_1 = x_1) = 1 - \left[1 - \frac{x_1}{\lambda N}\right]^{1/k_1}.$$

The group size for the second stage is then obtained as

$$k_2 = k_2(x_1) = k_2(N, \lambda, X_1 = x_1) = \arg\min_l Q_{(1-\lambda)N}(l; \widehat{p}_1(x_1)).$$

That is, for each realization of $X_1 = x_1$, k_2 minimizes the mean squared error of an estimate of p, where the true p in the mean squared error formula is replaced by $\hat{p}_1(x_1)$. Hence, $k_2(X_1)$ is a random variable that derives its randomness from X_1.

Stage 2 proceeds by testing $(1-\lambda)N$ groups, each of size $k_2(x_1)$, and X_2 is the number of these groups showing the trait. Hence, conditioned on X_1, X_2 has a binomial distribution. Specifically, $X_2|X_1 = x_1$ has a binomial distribution with parameters $(1-\lambda)N$ and $1-(1-p)^{k_2(x_1)}$.

The (final) two-stage adaptive estimator of p, $\hat{p}_A$, is the maximum likelihood estimator based on the joint distribution of X_1 and X_2, and is the solution to

$$\frac{k_1 x_1}{1-(1-p)^{k_1}} + \frac{k_2 x_2}{1-(1-p)^{k_2}} = N\left[k_1\lambda + k_2(1-\lambda)\right]. \tag{5}$$

The mean squared error can be calculated as

$$MSE(\hat{p}_A) = E_{X_1}\left\{E_{X_2}\left[(\hat{p}_A - p)^2 |X_1\right]\right\}.$$

In order to assess the asymptotic behavior of $\hat{p}_A$, one must first assess the asymptotic behavior of the nonrandom $k_1(N)$ and the random $k_2(N, X_1)$. Result 1, with λN in place of N, yields $lim_{N\to\infty} k_1(N) = \bar{k}$. The asymptotic behavior of $k_2(N, X_1)$ is addressed by the following result:

Result 2 *Define*

$$\tilde{k} = \arg\min_l \left\{\frac{1-(1-p)^l}{l^2(1-p)^{l-2}}\right\}. \tag{6}$$

If there exists N_p such that with probability one $k_2(N, X_1) \geq 1$ for all $N > N_p$, and with probability one the first derivative of $Q_{(1-\lambda)N}(l;\hat{p}_1)$ with respect to l vanishes at most once, then $k_2(N, X_1) \overset{wp1}{\to} \tilde{k}$ as $N \to \infty$.

[Appendix A contains the proof for the general multistage procedure. Result 2 is obtained as a special case by letting $i = 1$, $N_1 = \lambda N$, $N_2 = (1-\lambda)N$ and by noting that $\hat{p}_1$ is strongly consistent for p.] The condition on $Q_{(1-\lambda)N}(l;\hat{p}_1)$ is similar to the condition found in Result 1, and is, indeed, a fairly mild condition. The implication of this theorem is that the second stage group size approaches the asymptotically optimal group size, regardless of the value of the initial p_0. Use of this fact, and Martingale convergence theorems, provides the following result, whose proof may be found in Appendix B:

Result 3 *Let λ, N, and p_0 be given. If the group sizes k_1 and k_2 are chosen as described above, and $\widehat{p}_A$ is the solution to (5), then*

(i) $\widehat{p}_A$ is strongly consistent for p, and

(ii) $\sqrt{N}(\widehat{p}_A - p) \xrightarrow{d} Normal\left(0, \left\{\lambda \frac{(\overline{k})^2(1-p)^{\overline{k}-2}}{1-(1-p)^{\overline{k}}} + (1-\lambda)\frac{(\widetilde{k})^2(1-p)^{\widetilde{k}-2}}{1-(1-p)^{\widetilde{k}}}\right\}^{-1}\right)$,

where $\overline{k}$ and $\widetilde{k}$ are given by (4) and (6) above.

3 Illustrating the group-testing estimation procedures

To demonstrate the use of the nonadaptive and adaptive group-testing estimation procedures, we consider the case with $\lambda = 0.5$, $N = 30$, $p = 0.05$ (unknown to the experimenter), and $p_0 = 0.025$. That is, we shall perform 30 tests (use 30 groups), and the initial value p_0 is exceeded by the true p. For the adaptive procedure, we will perform 15 of the $N = 30$ tests in Stage 1 and the remaining 15 in Stage 2.

Let us first consider the usual nonadaptive procedure. The group size, k, is chosen to be the integer part of $\arg\min_l Q_{N=30}(l; p_0 = 0.025)$, which is $k = 38$. Then, if X counts the number of the 30 groups that test positive, X is distributed as a binomial random variable with parameters 30 and $1-(1-p)^k = 0.8576$. Suppose, for example, we observe $x = 25$. The nonadaptive estimate, as per (1), is $\widehat{p} = 0.046057$, with the true mean squared error 0.009156 (equation (2)), and estimated mean squared error 0.003973 (equation (2) with p replaced by $\widehat{p}$ everywhere).

For the adaptive procedure, we must first determine k_1, the size of the groups for the 15 tests in Stage 1. This is determined as the integer part of $\arg\min_l Q_{\lambda N=15}(l; p_0 = 0.025)$, i.e., $k_1 = 23$. If X_1 counts the number of the 15 groups that test positive in Stage 1, X_1 is distributed as a binomial random variable with parameters 15 and $1-(1-p)^{k_1} = 0.6926$. Suppose, for example, we observe $x_1 = 10$. The intermediate estimate, $\widehat{p}_1$, required for determining a new group size, is calculated as $\widehat{p}_1 = 0.046643$.

The second stage group size k_2 is the integer part of $\arg\min_l Q_{(1-\lambda)N=15}(l;\widehat{p}_1(x_1)=0.046643)$, and is found to be 13. If X_2 counts the number of the 15 groups that test positive in Stage 2, X_2 is distributed as a binomial random variable with parameters 15 and $1-(1-p)^{k_2}=0.4867$. Now suppose we observe $x_2=7$. The adaptive estimate, $\widehat{p}_A$, is then the solution to (5), and is found to be 0.046882, with the true mean squared error 0.000215, and the estimated mean squared error 0.000182.

4 Asymptotic and small-sample comparison of the estimators

The comparison of the nonadaptive and adaptive estimators is based on the total number N of tests to be performed, not on the number of individuals to be used. In fact, for a fixed total number N of tests, the nonadaptive estimator uses Nk individuals, whereas the adaptive estimator uses $\lambda N k_1+(1-\lambda)Nk_2$ individuals, which could be larger or smaller than Nk. The idea is that the number of individuals is not of overwhelming concern. Rather, the cost of testing is the main constraint on the budget. This is exactly the situation in which group testing offers the most advantage (see Thompson, 1962; Sobel and Elashoff, 1975).

4.1 Asymptotic comparison

Asymptotic comparison of the group-testing estimators will be done through the analysis of the asymptotic efficiency of $\widehat{p}_A$ relative to $\widehat{p}$, defined as the asymptotic variance of $\widehat{p}$ divided by the asymptotic variance of $\widehat{p}_A$:

$$ARE(\widehat{p}_A \text{ to } \widehat{p}) = \lambda + (1-\lambda)\left\{ \frac{(\tilde{k})^2(1-p)^{\tilde{k}-2}}{1-(1-p)^{\tilde{k}}} \Big/ \frac{(\bar{k})^2(1-p)^{\bar{k}-2}}{1-(1-p)^{\bar{k}}} \right\},$$

where $\bar{k}$ and $\tilde{k}$ are defined in (4) and (6), respectively.

Result 4 *(i) When the initial value $p_0=p$, $ARE(\widehat{p}_A$ to $\widehat{p})=1$, otherwise $ARE(\widehat{p}_A$ to $\widehat{p})\geq 1$.*

(ii) Let the initial values $p_{0-}=p-\delta$ and $p_{0+}=p+\delta$, for $\delta>0$, correspond to ARE_- and ARE_+, respectively. Then $ARE_- \geq ARE_+$.

(iii) $ARE(\widehat{p}_A$ to $\widehat{p})$ increases with decreasing λ.

Proof: (i) Clearly, if $p_0 = p$, then $\overline{k} = \widetilde{k}$, so that $ARE(\widehat{p}_A$ to $\widehat{p}) = 1$. By definition of $\widetilde{k}$,

$$\frac{1-(1-p)^k}{k^2(1-p)^{k-2}} \Big/ \frac{1-(1-p)^{\widetilde{k}}}{(\widetilde{k})^2(1-p)^{\widetilde{k}-2}} \geq 1$$

for all values of k. Hence, $ARE(\widehat{p}_A$ to $\widehat{p}) \geq 1$ always.

(ii) Define $f_p(l) = \frac{1-(1-p)^l}{l^2(1-p)^{l-2}}$. Using a second order Taylor's approximation, $f_p(l) \doteq f_p(\widetilde{k}) + \frac{1}{2}f_p''(\widetilde{k})(l-\widetilde{k})^2$, since $f_p'(\widetilde{k}) = 0$. So we may write the asymptotic relative efficiency as

$$ARE \doteq \lambda + (1-\lambda)\left[1 + \frac{1}{2}\frac{f_p''(\widetilde{k})}{f_p(\widetilde{k})}(\overline{k}-\widetilde{k})^2\right].$$

Let $\overline{k}_-$ and $\overline{k}_+$ be the group sizes associated with p_{0-} and p_{0+}, respectively. Then

$$ARE_- - ARE_+ \doteq \frac{1}{2}(1-\lambda)\frac{f_p''(\widetilde{k})}{f_p(\widetilde{k})}[(\overline{k}_- - \widetilde{k})^2 - (\overline{k}_+ - \widetilde{k})^2].$$

Since $|\overline{k}_- - \widetilde{k}| > |\overline{k}_+ - \widetilde{k}|$ (see Thompson (1962), or use the approximate closed-form expressions $\overline{k}_- \doteq \frac{1.5936}{p_{0-}} - 1$ and $\overline{k}_+ \doteq \frac{1.5936}{p_{0+}} - 1$), and $f_p''(\widetilde{k}) \geq 0$, $f_p(\widetilde{k}) > 0$ (see Hughes-Oliver, 1991), $ARE_- - ARE_+ \geq 0$.

(iii) Given the comments in (i), this is obvious. □

This result assures us that as N gets very large, the adaptive estimator will perform at least as well as the nonadaptive estimator. Furthermore, the adaptive estimator has greatest ARE benefits when p_0 is exceeded by the true p; as mentioned earlier, this is precisely the situation wherein the nonadaptive estimator does poorly. So, for large N, not only will we have better performance of the adaptive estimator, but this improvement is greatest for cases wherein the nonadaptive estimator is weakest.

The degree of improvement of the nonadaptive estimator depends on λ as well as p_0. For large λ, that is, more data in the first stage, the adaptive estimator does not have as much opportunity to overcome any bias introduced in the first stage (where we began with a possibly poor p_0). On the other hand, for small λ, that is, more data in the second stage, there is more opportunity to benefit from the adjustment after the first stage. Although ARE increases as λ decreases, in applications λN cannot be so small that the first stage provides insufficient information for a reasonable adjustment of the group size; more will be said about this in the next section.

Figure 2 graphically shows the implications of Result 4. Asymptotic relative efficiency is plotted as a function of p ($< \frac{1}{2}$), for $p_0 = p, p \pm 0.25p, p \pm 0.50p$ and $\lambda = 0.25, 0.50, 0.75$.

4.2 Small sample comparison

The small-sample comparison of the group-testing estimators is done through their relative efficiency, analogous to the comparison based on asymptotic relative efficiency in Section 4.1. The (small-sample) relative efficiency of $\widehat{p}_A$ to $\widehat{p}$ is defined as

$$RE(\widehat{p}_A \text{ to } \widehat{p}) = \frac{MSE(\widehat{p})}{MSE(\widehat{p}_A)}.$$

Table 1 contains mean squared errors of the nonadaptive estimator, $MSE(\widehat{p})$, for $p = 0.01$, 0.025, 0.05, 0.10, 0.20, $p_0 = p, p \pm 0.25p, p \pm 0.50p, p \pm 0.75p$ and $N = 5$, 10, 20, 30, 40, 50, 100. For example, when $p = 0.05$, $p_0 = p - 0.75p = 0.0125$ and $N = 10$, then $MSE(\widehat{p}) = 0.0600$. Only integer values were used for the group sizes, since this necessarily would be the case in practice; non-integer group sizes would increase the mean squared error in some cases, and decrease it in others. A ceiling on the allowable group size was set at 255; this is much higher than would likely be considered in practice because of concerns about dilution effects (resulting in false negatives) and other violations of the binomial model.

Figures 3–9 plot the relative efficiency, $RE(\widehat{p}_A \text{ to } \widehat{p})$, versus N, for all the cases considered in Table 1, and for $\lambda = 0.25, 0.50, 0.75$. The effect of limiting the maximum allowable group size is clearly seen in Figure 3 ($p_0 = 0.25p$), and, to some extent, also in Figure 4 ($p_0 = 0.5p$). By restricting the group size to a maximum of 255, we affect the performance of the estimators (for small p particularly), because the group size that would minimize the mean squared error expressions can be much larger than 255. For example, when $p = 0.0025$ and $N = 50$, the group size that minimizes (2) is 441. This restriction on group size results in similarity of the behavior of the nonadaptive and adaptive estimators sooner (for smaller N) than would be expected; it is seen in Figures 3 and 4 as the curves for increasing values of p being out of order.

These figures also indicate the necessity of careful choice of the values of λ and N. The choice of λ and N must be such that the information collected in Stage 1 is reliable, that is, sufficient for a meaningful adjustment of the group size. Figures 5–9 show the dip in relative efficiency below 1.0 for small N, especially when λ is small. In using the adaptive estimator, we recommend that $\lambda N \geq 10$. If the *a priori* p_0 is very small, a larger value of λN would be desirable; if p_0 is large, λN can be smaller.

When N is large (≥ 100), provided p_0 is not much smaller than p (Figures 5–9), the estimators differ little. In contrast, when p_0 is $0.25p$ (Figure 3) or $0.5p$ (Figure 4) the adaptive estimator is far better unless N is huge.

When N is small (< 100) and λN is chosen as discussed above, the relative performance of the two estimators again depend on whether p_0 is larger than, or smaller than, p. When $p_0 < p$, the adaptive estimator is clearly superior. When $p_0 \geq p$, the nonadaptive estimator is superior, with its best performance providing a mean squared error that is 0.75 times the mean squared error of the adaptive estimator; the differences are much less than suggested by Figures 5–9, which did not require $\lambda N \geq 10$. So, the adaptive estimator yields much better performance when $p_0 < p$, but at the cost of slight loss in performance when $p_0 \geq p$ and N is small.

5 Tail probabilities of the standardized adaptive estimator

Monte Carlo simulations were performed to assess the tail behavior of

$$\tilde{p} = \frac{\hat{p}_A - p}{\sqrt{\widehat{MSE}(\hat{p}_A)}},$$

for the cases $\lambda = 0.25, 0.50, 0.75$, $p = 0.01, 0.025, 0.050, 0.10, 0.20$, $p_0 = p \pm 0\%, 25\%, 50\%, 75\%$ of p, $N = 5, 10, 20, 30, 40, 50$ and $\lambda N \geq 10$ (resulting in a total of 350 cases). Based on Monte Carlo simulations of 500 replications each, we estimated $Pr(|\tilde{p}| > 1.645)$, $Pr(|\tilde{p}| > 1.960)$, and $Pr(|\tilde{p}| > 2.575)$. If $\tilde{p}$ has approximately a standard normal distribution (which it does, in the limit as N approaches infinity), then these probabilities should be close to 0.10, 0.05, 0.01, respectively. Binomial random variates were generated by KBIN (Ahrens and Dieter, 1974) and the random number generator was RAN (Schrage, 1979).

Histograms of the estimates of these tail probabilities are presented in Figure 10. The estimates of $Pr(|\tilde{p}| > 1.645)$ are generally smaller than the standard normal value of 0.10, with about 13 percent of the estimates exceeding 0.10 (2.3 percent exceed at a significance level of 0.05). The estimates of $Pr(|\tilde{p}| > 1.960)$ are also generally smaller than 0.05, with about 20 percent of the estimates exceeding 0.05 (3.4 percent exceed at a significance level of 0.05). On the other hand, about 47 percent of the estimates of $Pr(|\tilde{p}| > 2.575)$ exceed 0.01 (18 percent exceed at a significance level of 0.05). So, in general, the normal approximation may be used to produce confidence intervals with confidence coefficients no less than 0.90, or no less than 0.95 (corresponding to 90% and 95% confidence intervals, respectively). However, the normal approximation is not recommended for obtaining 99% confidence intervals, since the true confidence coefficient is likely to be less than 0.99. Note, however, that 95% of the estimates of $Pr(|\tilde{p}| > 2.575)$ are less than or equal to 0.034.

Smaller simulation studies (whose results are not presented here) indicate that the density of $\tilde{p}$ is truncated on the right hand side, and elongated on the left hand side. This is due to the fact that: i) a large $\hat{p}_A$ results in a large $\widehat{MSE}(\hat{p}_A)$, so that $\tilde{p}$ "shrinks" large positive differences; and ii) a small $\hat{p}_A$ results in a small $\widehat{MSE}(\hat{p}_A)$, so that $\tilde{p}$ "explodes" large negative differences. This kind of behavior causes slow convergence of $\tilde{p}$ to a standard normal. However, $p^* = \frac{\hat{p}_A - p}{\sqrt{MSE(\hat{p}_A)}}$ [using the true $MSE(\hat{p}_A)$ instead of the estimate $\widehat{MSE}(\hat{p}_A)$] does not experience these problems, thus allowing it fairly fast convergence to a standard normal. For this reason it is recommended that p^* be used for approximate hypothesis tests and $\tilde{p}$ be used for approximate confidence intervals.

6 Discussion and conclusions

An adaptive group-testing estimation scheme for proportions has been presented and its properties and advantages discussed. This scheme allows one update/adjustment of the group size, to be based on all currently available information. An estimate of variability has been presented, as well as approximate methods for confidence interval generation and hypothesis testing. These approximations are based on the asymptotic normal distribution of the adaptive estimator.

Asymptotic and small-sample comparisons were made between the usual nonadaptive and adaptive group-testing estimators. The adaptive estimator is always at least as efficient, in an asymptotic sense, as the nonadaptive estimator, with increased efficiency when the initial p_0 is a poor choice. Small-sample comparison showed that, given enough information to provide a reasonable basis for adjusting the group size (10 or more observations in the first stage), the adaptive estimator still performs much better than the nonadaptive estimator in cases where too small a p_0 is used; when $p_0 \geq p$ the adaptive estimator is slightly less efficient than the nonadaptive estimator when N is small.

In estimating a proportion p when one is unsure of the quality of the initial p_0, it is advisable to use adaptive rather than nonadaptive group testing. If p_0 is a good choice, there is little loss of information. On the other hand, if p_0 is a poor choice, the adaptive estimator can be far better.

References

[1] Ahrens, J. H. and Dieter, U. (1974). Computer methods for sampling from gamma, beta, poisson and binomial distributions. *Computing* **12**, 223–246.

[2] Billingsley, P. (1961). *Statistical Inference for Markov Processes*, The University of Chicago Press, Chicago and London.

[3] Dorfman, R. (1943). The detection of defective members of large populations. *Annals of Mathematical Statistics* **14**, 436–440.

[4] Griffiths, D. A. (1972). A further note on the probability of disease transmission. *Biometrics* **28**, 1133–1139.

[5] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*, Academic Press.

[6] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58**, 13–30.

[7] Hughes-Oliver, J. M. (1991). Estimation using group-testing procedures: Adaptive iteration. Ph.D. dissertation, North Carolina State University, Raleigh, N.C.

[8] Hughes-Oliver, J. M. and Swallow, W. H. (1992). Choosing the group size for group testing to estimate a proportion. Institute of Statistics Mimeograph Series No. 2209.

[9] Le, C. T. (1981). A new estimator for infection rates using pools of variable size. *American Journal of Epidemiology* **14**, 132–136.

[10] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, second edition, John Wiley and Sons.

[11] Schrage, L. (1979). A more portable fortran random number generator. *Association for Computing Machinery: Transactions on Mathematical Software* **5**, 132–138.

[12] Sobel, M. and Elashoff, R. M. (1975). Group testing with a new goal, estimation. *Biometrika* **62**, 181–193.

[13] Swallow, W. H. (1985). Group testing for estimating infection rates and probability of disease transmission. *Phytopathology* **75**, 882-889.

[14] Swallow, W. H. (1987). Relative mean squared error and cost considerations in choosing group size for group testing to estimate infection rates and probabilities of disease transmission. *Phytopathology* **77**, 1376–1381.

[15] Thompson, K. H. (1962). Estimation of the proportion of vectors in a natural population of insects. *Biometrics* **18**, 568–578.

[16] Walter, S. D., Hildreth, S. W. and Beaty, B. J. (1980). Estimation of infection rates in populations of organisms using pools of variable size. *American Journal of Epidemiology* **112**, 124–128.

APPENDIX A: ASYMPTOTIC BEHAVIOR OF MULTISTAGE GROUP SIZES

Let N represent the total number of tests to be performed over all stages (known *a priori*); N_i represent the number of tests performed in the i^{th} stage; $\widehat{p}_i$ is the estimate of the proportion obtained at the end of the i^{th} stage; and k_{i+1} is the group size for stage $i+1$, obtained using the estimate $\widehat{p}_i$, for $i \geq 0$ and $\widehat{p}_0 = p_0$. That is,

$$k_{i+1} = k_{i+1}(N, \widehat{p}_i) = \arg\min_l \left\{ Q_{N_{i+1}}(l; \widehat{p}_i) \right\},$$

where

$$Q_n(l;p) = n\left\{(1-p)^2 + \sum_{i=0}^{n}\left(1-\frac{i}{n}\right)^{\frac{1}{l}}\left[\left(1-\frac{i}{n}\right)^{\frac{1}{l}} - 2(1-p)\right]\binom{n}{i}\delta_l^i(1-\delta_l)^{n-i}\right\},$$

and $\delta_l = 1-(1-p)^l$. Result 5 describes the limiting behavior of these group sizes, when the intermediate estimates $\widehat{p}_i$, $i \geq 0$, are strongly consistent for p.

Result 5 *Suppose the estimator $\widehat{p}_i$ is such that, with probability one, $0 < \widehat{p}_i < 1$ for all $N > n_i$ for some n_i, and $\widehat{p}_i \overset{wp1}{\rightarrow} p$ as $N \rightarrow \infty$. Then, provided that with probability one the first derivative of $Q_{N_{i+1}}(l; \widehat{p}_i)$ with respect to l equals zero at most once, and $k_{i+1}(N, \widehat{p}_i) \geq 1$ for all $i \geq 0$ and $N > n_i$,*

$$k_{i+1}(N, \widehat{p}_i) \overset{wp1}{\rightarrow} \widetilde{k} \quad as \quad N \rightarrow \infty,$$

where $i \geq 0$ and $\widetilde{k}$ is as defined in (6).

Result 5 says that, even if the proportion used to obtain the group size varies with N in a random fashion (rather than the deterministic p_0), we still have convergence of the group sizes as N tends to infinity, provided the proportions used to obtain the group sizes converge with probability one to the true proportion as N tends to infinity. Moreover, the limit is independent of the initial value p_0. In other words, if the intermediate estimates are strongly consistent, then, as the total number of tests increases, the group size for the next stage approaches the optimal group size with probability one, whatever the value of p_0.

Proof: Define the function $Q(l;p) = \frac{1-(1-p)^l}{l^2(1-p)^{l-2}}$. The proof consists of two main parts. First, it will be shown that

$$Q_{N_{i+1}}(l;\widehat{p}_i) \overset{wp1}{\rightarrow} Q(l;p) \quad \text{as} \quad N \rightarrow \infty.$$

Then it will be shown that

$$\arg\min_l \{Q_{N_{i+1}}(l;\widehat{p}_i)\} \overset{wp1}{\rightarrow} \arg\min_l \{Q(l;p)\} \quad \text{as} \quad N \rightarrow \infty.$$

Part 1: Show that for all $i \geq 0$

$$Q_{N_{i+1}}(l;\widehat{p}_i) \overset{wp1}{\rightarrow} Q(l;p) \quad \text{as} \quad N \rightarrow \infty.$$

Proof: It is convenient to write the function $Q_{N_{i+1}}(l;\widehat{p}_i)$ as a conditional expectation:

$$Q_{N_{i+1}}(l;\widehat{p}_i) = (1-\widehat{p}_i)^2 E_{Y_{N_{i+1}}|\widehat{p}_i}\left[f\left(Y_{N_{i+1}}\right)\right],$$

where

$$Y_{N_{i+1}}|\widehat{p}_i \sim Binomial(N_{i+1},\widehat{\delta}_i), \quad \widehat{\delta}_i = 1-(1-\widehat{p}_i)^l$$

and

$$f\left(Y_{N_{i+1}}\right) = N_{i+1}\left[1-\left(\frac{1-Y_{N_{i+1}}/N_{i+1}}{1-\widehat{\delta}_i}\right)^{1/l}\right]^2.$$

Let $\widehat{\gamma}_i = \frac{1}{2}(1-\widehat{\delta}_i)$ for $N \geq 1$. Then $0 < \widehat{\gamma}_i < 1$, $\widehat{\delta}_i + \widehat{\gamma}_i < 1$ for all $N \geq 1$, and $\widehat{\gamma}_i \overset{wp1}{\rightarrow} \frac{1}{2}(1-\delta)$ as $N \rightarrow \infty$, where $\delta = 1-(1-p)^l$. Then

$$E_{Y_{N_{i+1}}|\widehat{p}_i}\left[f\left(Y_{N_{i+1}}\right)\right] = E_{Y_{N_{i+1}}|\widehat{p}_i}\left[f\left(Y_{N_{i+1}}\right) I\left(Y_{N_{i+1}} \geq (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1}\right)\right]$$

$$+E_{Y_{N_{i+1}}|\widehat{p}_i}\left[f\left(Y_{N_{i+1}}\right)I\left(Y_{N_{i+1}} < (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1}\right)\right]$$

<u>Case 1</u>: $Y_{N_{i+1}} \geq (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1}$

Then

$$0 < f\left(Y_{N_{i+1}}\right) \leq N_{i+1}$$

which implies that

$$0 \leq E_{Y_{N_{i+1}}|\widehat{p}_i}\left[f\left(Y_{N_{i+1}}\right)I\left(Y_{N_{i+1}} \geq (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1}\right)\right]$$
$$\leq N_{i+1}P\left(Y_{N_{i+1}} \geq (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1} \quad | \quad \widehat{p}_i\right) \overset{wp1}{\to} 0 \text{ as } N \to \infty,$$

by Hoeffding (1963).

<u>Case 2</u>: $Y_{N_{i+1}} < (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1}$

By Taylor's expansion of $\left(\frac{1-x}{1-\widehat{\delta}_i}\right)^{1/l}$ around $x = \widehat{\delta}_i$,

$$0 \leq \left[1-\left(\frac{1-Y_{N_{i+1}}/N_{i+1}}{1-\widehat{\delta}_i}\right)^{1/l}\right]^2 \leq \frac{(1-\widehat{\delta}_i-\widehat{\gamma}_i)^{2/l-2}}{l^2(1-\widehat{\delta}_i)^{2/l}}\left(\frac{Y_{N_{i+1}}}{N_{i+1}}-\widehat{\delta}_i\right)^2$$
$$< \frac{1}{l^2(1-\widehat{\delta}_i-\widehat{\gamma}_i)^2}\left(\frac{Y_{N_{i+1}}}{N_{i+1}}-\widehat{\delta}_i\right)^2.$$

This implies that

$$0 \leq (1-\widehat{\delta}_i-\widehat{\gamma}_i)^2 f\left(Y_{N_{i+1}}\right)I\left(Y_{N_{i+1}} < (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1}\right) < \frac{N}{l^2}\left(\frac{Y_{N_{i+1}}}{N_{i+1}}-\widehat{\delta}_i\right)^2.$$

Moreover,

$$E\left\{\left|(1-\widehat{\delta}_i-\widehat{\gamma}_i)^2 f\left(Y_{N_{i+1}}\right)I\left(Y_{N_{i+1}} < (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1}\right)\right|^2\right\}$$
$$< \frac{N^2}{l^4}E\left\{\left(\frac{Y_{N_{i+1}}}{N_{i+1}}-\widehat{\delta}_i\right)^4\right\} \leq M < \infty.$$

So

$$(1-\widehat{\delta}_i-\widehat{\gamma}_i)^2 f\left(Y_{N_{i+1}}\right)I\left(Y_{N_{i+1}} < (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1}\right)$$

is uniformly integrable. But $(1-\widehat{\delta}_i-\widehat{\gamma}_i)^2 \xrightarrow{wp1} \frac{1}{4}(1-\delta)^2$, $I\left(Y_{N_{i+1}} < (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1}\right) \xrightarrow{p} 1$ and $f(Y_{N_{i+1}}) \xrightarrow{d} \left[\frac{\delta}{l^2(1-\delta)}\right]\chi_1^2$. Hence

$$(1-\widehat{\delta}_i-\widehat{\gamma}_i)^2 f\left(Y_{N_{i+1}}\right) I\left(Y_{N_{i+1}} < (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1}\right) \xrightarrow{d} \frac{\delta(1-\delta)}{4l^2}\chi_1^2,$$

which implies

$$E_{Y_{N_{i+1}}|\widehat{p}_i}\left[f\left(Y_{N_{i+1}}\right) I\left(Y_{N_{i+1}} < (\widehat{\delta}_i+\widehat{\gamma}_i)N_{i+1}\right)\right] \xrightarrow{wp1} \frac{\delta}{l^2(1-\delta)}.$$

Therefore,

$$Q_{N_{i+1}}(l;\widehat{p}_i) = (1-\widehat{p}_i)^2 E_{Y_{N_{i+1}}|\widehat{p}_i}\left[f\left(Y_{N_{i+1}}\right)\right] \xrightarrow{wp1} \frac{(1-p)^2\delta}{l^2(1-\delta)} = Q(l;p).$$

Part 2: Show that

$$k_{i+1} = \arg\min_l \{Q_{N_{i+1}}(l;\widehat{p}_i)\} \xrightarrow{wp1} \arg\min_l \{Q(l;p)\} = \widetilde{k} \quad \text{as} \quad N \to \infty.$$

Proof: Since $Q(\widetilde{k};p) < Q(l;p)$ if $l \neq \widetilde{k}$, then for all $\delta > 0$, there exists $\epsilon > 0$ (ϵ depends on δ) such that

$$Q(\widetilde{k}-\delta;p) - Q(\widetilde{k};p) > \epsilon$$

and

$$Q(\widetilde{k}+\delta;p) - Q(\widetilde{k};p) > \epsilon.$$

Furthermore, by Part 1 above, with probability one there exists N_ϵ such that for all $N > N_\epsilon$,

$$\left|Q_{N_{i+1}}(\widetilde{k};\widehat{p}_i) - Q(\widetilde{k};p)\right| < \frac{\epsilon}{2},$$

$$\left|Q_{N_{i+1}}(\widetilde{k}-\delta;\widehat{p}_i) - Q(\widetilde{k}-\delta;p)\right| < \frac{\epsilon}{2},$$

$$\left|Q_{N_{i+1}}(\widetilde{k}+\delta;\widehat{p}_i) - Q(\widetilde{k}+\delta;p)\right| < \frac{\epsilon}{2}.$$

Hence, with probability one, for all $N > N_\epsilon$,

$$-\epsilon < Q_{N_{i+1}}(\widetilde{k};\widehat{p}_i) - Q_{N_{i+1}}(\widetilde{k}-\delta;\widehat{p}_i) + Q(\widetilde{k}-\delta;p) - Q(\widetilde{k};p) < \epsilon$$

which implies

$$Q_{N_{i+1}}(\tilde{k}-\delta;\hat{p}_i) - Q_{N_{i+1}}(\tilde{k};\hat{p}_i) > Q(\tilde{k}-\delta;p) - Q(\tilde{k};p) - \epsilon > 0.$$

Similarly,

$$Q_{N_{i+1}}(\tilde{k}+\delta;\hat{p}_i) - Q_{N_{i+1}}(\tilde{k};\hat{p}_i) > 0.$$

Since $Q_{N_{i+1}}(l;\hat{p}_i)$ is continuously differentiable in l, then its derivative with respect to l equals zero for some $l \in (\tilde{k}-\delta, \tilde{k}+\delta)$, for all $N > N_\epsilon$, [which implies $k_{i+1}(N,\hat{p}_i) \in (\tilde{k}-\delta, \tilde{k}+\delta)$]. That is, $\lim_{N\to\infty} k_{i+1}(N,\hat{p}_i) \in (\tilde{k}-\delta, \tilde{k}+\delta)$ with probability one. But δ is arbitrary, so

$$k_{i+1}(N,\hat{p}_i) \overset{wp1}{\to} \tilde{k} \quad \text{as} \quad N \to \infty.$$

APPENDIX B: PROOF OF RESULT 3

Notation:

Let

$$X_i = \begin{cases} 1 & i^{th} \text{ group shows the trait} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, N$. Also, let k_1 be the group size for the first λN groups and $k_2 = k_2(X_1, X_2, \ldots, X_{\lambda N})$ be the group size for the last $(1-\lambda)N$ groups.

The likelihood is $\mathcal{L}(p) = \prod_{i=1}^{N} f(x_i|p, x_1, x_2, \ldots, x_{i-1})$, where $f(x_i|p, x_1, x_2, \ldots, x_{i-1}) = [1-(1-p)^{k_j}]^{x_i}[(1-p)^{k_j}]^{1-x_i}$ for $x_i = 0, 1$, $j = 1$ for $i = 1, 2, \ldots, \lambda N$ and $j = 2$ for $i = \lambda N + 1, \lambda N + 2, \ldots, N$. Denote the average log-likelihood as $l(p) = \frac{1}{N}\ln \mathcal{L}(p) = \frac{1}{N}\sum_{i=1}^{N} Y_i$, where $Y_i = \ln f(X_i|p, X_1, \ldots, X_{i-1})$. Then $\hat{p}_A$ is such that $\frac{\partial}{\partial p} l(p)|_{p=\hat{p}_A} = 0$.

Also, let p_t denote the true (unknown) value of the proportion (this distinction was not made in the paper), and $\mathcal{F}_i = \sigma(X_1, X_2, \ldots, X_i)$. $E_{p_t}(\cdot)$ means expectation with respect to the true p_t, and define

$$I(p) = \lambda \frac{(\bar{k})^2(1-p)^{\bar{k}-2}}{1-(1-p)^{\bar{k}}} + (1-\lambda)\frac{(\tilde{k})^2(1-p)^{\tilde{k}-2}}{1-(1-p)^{\tilde{k}}}.$$

Lemmas

Lemma 1: $\frac{1}{N}\sum_{i=1}^{N}\{Y_i - E_{p_t}[Y_i|\mathcal{F}_{i-1}]\} \overset{wp1}{\to} 0$ for all $p \in (0,1)$.

Proof: $S_N = \sum_{i=1}^{N}\{Y_i - E_{p_t}[Y_i|\mathcal{F}_{i-1}]\}$ is a zero-mean martingale. Moreover, since $0 < p < 1$, $f(X_i|p, X_1, \ldots, X_{i-1})$ is bounded away from zero and infinity with probability one, then Y_i is bounded with probability one. Therefore, the conditional variance of Y_i is bounded with probability one for all i. That is, $Var[Y_i|\mathcal{F}_{i-1}] \leq c$ with probability one for all i. Hence, with probability one,

$$\sum_{i=1}^{\infty}\frac{1}{i^2}E\left\{(Y_i - E_{p_t}[Y_i|\mathcal{F}_{i-1}])^2|\mathcal{F}_{i-1}\right\} \leq \sum_{i=1}^{\infty}\frac{c}{i^2} < \infty.$$

Thus, by Theorem 2.18 of Hall and Heyde (1980), $\frac{1}{N}S_N \overset{wp1}{\to} 0$. Since this result only required $0 < p < 1$, it is true for all such p, and we have the desired result.

Lemma 2: $l(p)$ is strictly concave with probability one.

Proof:

$$\frac{\partial^2}{\partial p^2}l(p) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2}{\partial p^2}Y_i = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial p}\left[\frac{\partial}{\partial p}Y_i\right].$$

But

$$\frac{\partial}{\partial p}Y_i = \frac{\frac{\partial}{\partial p}f(X_i|p, X_1, \ldots, X_{i-1})}{f(X_i|p, X_1, \ldots, X_{i-1})},$$

and

$$\begin{aligned}\frac{\partial}{\partial p}f(X_i|p, X_1, \ldots, X_{i-1}) &= X_i[1-(1-p)^{k_j}]^{X_i-1}[k_j(1-p)^{k_j-1}][(1-p)^{k_j}]^{1-X_i} \\ &\quad +[1-(1-p)^{k_j}]^{X_i}(1-X_i)[(1-p)^{k_j}]^{-X_i}[-k_j(1-p)^{k_j-1}] \\ &= f(X_i|p, X_1, \ldots, X_{i-1})g(X_i|p, X_1, \ldots, X_{i-1}),\end{aligned}$$

where

$$g(X_i|p, X_1, \ldots, X_{i-1}) = \frac{k_j[X_i - [1-(1-p)^{k_j}]]}{(1-p)[1-(1-p)^{k_j}]}.$$

Hence

$$\frac{\partial^2}{\partial p^2}Y_i = \frac{\partial}{\partial p}g(X_i|p, X_1, \ldots, X_{i-1})$$

$$= \frac{k_j}{(1-p)^2[1-(1-p)^{k_j}]^2}\left\{X_i[1-(k_j+1)(1-p)^{k_j}]-[1-(1-p)^{k_j}]^2\right\}$$
$$< 0 \text{ for } X_i = 0,1, \text{ since } 1 \le k_j < K < \infty \text{ with probability one and } 0 < p < 1.$$

Therefore, $\frac{\partial^2}{\partial p^2} l(p) < 0$ and $l(p)$ is strictly concave with probability one.

Lemma 3: $\frac{\partial^3}{\partial p^3} l(p)$ is bounded in probability for all $p \in (0,1)$.

Proof:

$$\frac{\partial^3}{\partial p^3} l(p) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial^3}{\partial p^3} Y_i$$

and

$$\begin{aligned}\frac{\partial^3}{\partial p^3} Y_i &= \frac{\partial^2}{\partial p^2} g(X_i|p, X_1, \ldots, X_{i-1}) \text{ (see Lemma 2)}\\ &= \frac{k_j}{(1-p)^3[1-(1-p)^{k_j}]^3}\left\{X_i(k_j+1)k_j(1-p)^{k_j}[1-(1-p)^{k_j}] - 2[1-(1-p)^{k_j}]^3\right.\\ &\qquad \left. +2X_i[1-(k_j+1)(1-p)^{k_j}]^2\right\}.\end{aligned}$$

Since X_i is bounded with probability one, $1 \le k_j < K < \infty$ with probability one, and $0 < p < 1$, then $\frac{\partial^3}{\partial p^3} Y_i$ is bounded in probability for all $p \in (0,1)$. Moreover, the same bound applies to all $\frac{\partial^3}{\partial p^3} Y_i, i = 1, 2, \ldots, N$. Hence, $\frac{\partial^3}{\partial p^3} l(p)$ is bounded in probability for all $p \in (0,1)$.

Lemma 4: $\frac{\partial^2}{\partial p^2} l(p) \overset{wp1}{\to} -I(p)$ for all $p \in (0,1)$.

Proof: Let $W_i = \frac{\partial^2}{\partial p^2} Y_i$ for $i = 1, 2, \ldots, N$. Then $\frac{\partial^2}{\partial p^2} l(p) = \frac{1}{N}\sum_{i=1}^{N} W_i$,

$$\begin{aligned}E[W_i|\mathcal{F}_{i-1}] &= E\left[\frac{k_j}{(1-p)^2[1-(1-p)^{k_j}]^2}\left\{X_i[1-(k_j+1)(1-p)^{k_j}]-[1-(1-p)^{k_j}]^2\right\}|\mathcal{F}_{i-1}\right] \text{ (Lemma 2)}\\ &= -\frac{k_j^2(1-p)^{k_j-2}}{1-(1-p)^{k_j}},\end{aligned}$$

and

$$\begin{aligned}E\{(W_i - E[W_i|\mathcal{F}_{i-1}])^2|\mathcal{F}_{i-1}\} &= \frac{k_j^2}{[1-(1-p)^{k_j}]^3}(1-p)^{k_j-4}[1-(k_j+1)(1-p)^{k_j}]^2\\ &\le c \text{ for all } i \text{ with probability one.}\end{aligned}$$

Moreover, $S_N = \sum_{i=1}^N \{W_i - E[W_i|\mathcal{F}_{i-1}]\}$ is a zero-mean martingale with

$$\sum_{i=1}^{\infty} \frac{1}{i^2} E\left\{(W_i - E[W_i|\mathcal{F}_{i-1}])^2|\mathcal{F}_{i-1}\right\} \leq \sum_{i=1}^{\infty} \frac{c}{i^2} < \infty \text{ with probability one.}$$

Thus, by Theorem 2.18 of Hall and Heyde (1980), $\frac{1}{N}S_N \stackrel{wp1}{\rightarrow} 0$ for all $p \in (0,1)$. That is,

$$\frac{1}{N}\sum_{i=1}^{N} W_i - \frac{1}{N}\sum_{i=1}^{N} E[W_i|\mathcal{F}_{i-1}] \stackrel{wp1}{\rightarrow} 0 \text{ for all } p \in (0,1).$$

But

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} E[W_i|\mathcal{F}_{i-1}] &= \frac{1}{N}\left\{\lambda N \cdot \frac{-k_1^2(1-p)^{k_1-2}}{1-(1-p)^{k_1}} + (1-\lambda)N \cdot \frac{-k_2^2(1-p)^{k_2-2}}{1-(1-p)^{k_2}}\right\} \\ &= -\left\{\lambda \cdot \frac{k_1^2(1-p)^{k_1-2}}{1-(1-p)^{k_1}} + (1-\lambda)\cdot \frac{k_2^2(1-p)^{k_2-2}}{1-(1-p)^{k_2}}\right\},\end{aligned}$$

and $k_1 \rightarrow \overline{k}$ and $k_2 \stackrel{wp1}{\rightarrow} \widetilde{k}$, implies that $\frac{1}{N}\sum_{i=1}^N E[W_i|\mathcal{F}_{i-1}] \stackrel{wp1}{\rightarrow} -I(p)$. Since

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} W_i &= \left[\frac{1}{N}\sum_{i=1}^{N} W_i - \frac{1}{N}\sum_{i=1}^{N} E[W_i|\mathcal{F}_{i-1}]\right] + \left[\frac{1}{N}\sum_{i=1}^{N} E[W_i|\mathcal{F}_{i-1}]\right] \\ &= A_N + B_N,\end{aligned}$$

and $A_N \stackrel{wp1}{\rightarrow} 0$ and $B_N \stackrel{wp1}{\rightarrow} -I(p)$, then $\frac{1}{N}\sum_{i=1}^N W_i \stackrel{wp1}{\rightarrow} -I(p)$.

Lemma 5: $\sqrt{N}\frac{\partial}{\partial p}l(p) \stackrel{d}{\rightarrow} Normal(0, I(p))$ for all $p \in (0,1)$.

Proof: Let $W_i = \frac{\partial}{\partial p}Y_i$ for $i = 1, 2, \ldots, N$. Then $\frac{\partial}{\partial p}l(p) = \frac{1}{N}\sum_{i=1}^N W_i$,

$$E[W_i|\mathcal{F}_{i-1}] = E\left\{\frac{k_j[X_i - [1-(1-p)^{k_j}]]}{(1-p)[1-(1-p)^{k_j}]}|\mathcal{F}_{i-1}\right\} = 0,$$

$$E[W_i^2|\mathcal{F}_{i-1}] = \frac{k_j^2[1-(1-p)^{k_j}](1-p)^{k_j}}{(1-p)^2[1-(1-p)^{k_j}]^2} = \frac{k_j^2(1-p)^{k_j-2}}{1-(1-p)^{k_j}},$$

and

$$E[W_i^4|\mathcal{F}_{i-1}] = \frac{k_j^4}{(1-p)^4[1-(1-p)^{k_j}]^4}c_1,$$

where c_1 is the fourth central moment of a binomial random variable. Hence $E[W_i^4|\mathcal{F}_{i-1}] \leq c$ for all i with probability one.

So $S_N = \sum_{i=1}^N W_i$ is a zero-mean martingale with

$$\frac{1}{N}\sum_{i=1}^N E[W_i^2|\mathcal{F}_{i-1}] = \lambda\frac{k_1^2(1-p)^{k_1-2}}{1-(1-p)^{k_1}} + (1-\lambda)\frac{k_2^2(1-p)^{k_2-2}}{1-(1-p)^{k_2}} \stackrel{wp1}{\rightarrow} I(p)$$

(see Lemma 4), and

$$0 \leq \frac{1}{N^2}\sum_{i=1}^N E[W_i^4|\mathcal{F}_{i-1}] \leq \frac{1}{N^2}\sum_{i=1}^N c = \frac{c}{N} \stackrel{wp1}{\rightarrow} 0.$$

Therefore, $\frac{1}{N^2}\sum_{i=1}^N E[W_i^4|\mathcal{F}_{i-1}] \stackrel{wp1}{\rightarrow} 0$. Hence, by Billingsley's central limit theorem for martingales (1961, p. 52),

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N W_i = \sqrt{N}\frac{\partial}{\partial p}l(p) \stackrel{d}{\rightarrow} Normal(0, I(p)).$$

Strong Consistency:

Define $h(p) = \frac{1}{N}\sum_{i=1}^N E_{p_t}[Y_i|\mathcal{F}_{i-1}]$. By Lemma 1,

$$l(p) - h(p) \stackrel{wp1}{\rightarrow} 0 \quad \text{for all } p \in (0,1). \tag{7}$$

Moreover, Rao's inequality (1973, p.59) yields

$$E_{p_t}[\ln f(X_i|p_t, X_1, \ldots, X_{i-1}) \mid \mathcal{F}_{i-1}] > E_{p_t}[\ln f(X_i|p, X_1, \ldots, X_{i-1}) \mid \mathcal{F}_{i-1}] \text{ for all } p \neq p_t.$$

Hence

$$h(p_t) - h(p) > 0 \quad \text{for all } p \neq p_t. \tag{8}$$

Using (7) and (8), for all $\delta > 0$, there exists $\epsilon = \epsilon(\delta) > 0$ and N_ϵ such that with probability one,

$$\begin{aligned} h(p_t) - h(p_t - \delta) &> \epsilon \\ h(p_t) - h(p_t + \delta) &> \epsilon \\ |l(p_t) - h(p_t)| &< \frac{\epsilon}{2} \quad \text{for all } N > N_\epsilon \\ |l(p_t - \delta) - h(p_t - \delta)| &< \frac{\epsilon}{2} \quad \text{for all } N > N_\epsilon \\ |l(p_t + \delta) - h(p_t + \delta)| &< \frac{\epsilon}{2} \quad \text{for all } N > N_\epsilon \end{aligned}$$

Hence, with probability one,

$$l(p_t) - l(p_t - \delta) > 0$$

$$l(p_t) - l(p_t + \delta) > 0$$

for all $N > N_\epsilon$. That is, with probability one, $\frac{\partial}{\partial p} l(p)$ changes sign in the interval $(p_t - \delta, p_t + \delta)$. Since with probability one, $\frac{\partial}{\partial p} l(p)$ changes sign at most once [$l(p)$ is strictly concave, with probability one, Lemma 2], then $\widehat{p}_A \in (p_t - \delta, p_t + \delta)$ with probability one. But δ is arbitrary, so $\widehat{p}_A \stackrel{wp1}{\rightarrow} p_t$.

Asymptotic Normality:

By Taylor's Theorem,

$$0 = l'(\widehat{p}_A) = l'(p_t) + l''(p_t)(\widehat{p}_A - p_t) + \frac{l'''(p^*)}{2}(\widehat{p}_A - p_t)^2,$$

where p^* is between p_t and $\widehat{p}_A$. Hence,

$$\sqrt{N}(\widehat{p}_A - p_t) = \frac{-\sqrt{N} l'(p_t)}{l''(p_t) + \frac{l'''(p^*)}{2}(\widehat{p}_A - p_t)}.$$

Since $\frac{l'''(p^*)}{2}$ is bounded in probability (Lemma 3) and $\widehat{p}_A - p_t \stackrel{wp1}{\rightarrow} 0$ (strong consistency), then $\frac{l'''(p^*)}{2}(\widehat{p}_A - p_t) \stackrel{p}{\rightarrow} 0$. Moreover, $l''(p_t) \stackrel{wp1}{\rightarrow} -I(p_t)$ (Lemma 4) implies $l''(p_t) + \frac{l'''(p^*)}{2}(\widehat{p}_A - p_t) \stackrel{p}{\rightarrow} -I(p_t)$. Finally, $\sqrt{N} l'(p_t) \stackrel{d}{\rightarrow} Normal(0, I(p_t))$ (Lemma 5) implies $\sqrt{N}(\widehat{p}_A - p_t) \stackrel{d}{\rightarrow} Normal\left(0, \frac{1}{I(p_t)}\right)$.

Table 1. Mean Squared Error of the Nonadaptive Group-Testing Estimator

		N							
p	% away[1]	5	10	20	30	40	50	100	500
0.01	−75	2.09(−3)[2]	1.03(−2)	9.38(−2)	8.83(−2)	3.96(−2)	1.78(−2)	3.24(−4)	3.78(−7)
	−50	3.87(−4)	2.75(−4)	7.41(−4)	1.74(−3)	3.31(−3)	5.03(−3)	3.24(−4)	3.78(−7)
	−25	2.49(−4)	5.93(−5)	2.23(−5)	1.66(−5)	1.46(−5)	1.19(−5)	1.87(−6)	3.29(−7)
	0	2.38(−4)	4.58(−5)	1.28(−5)	6.92(−6)	4.71(−6)	3.58(−6)	1.65(−6)	3.10(−7)
	25	2.56(−4)	4.89(−5)	1.38(−5)	7.41(−6)	5.00(−6)	3.75(−6)	1.70(−6)	3.20(−7)
	50	2.72(−4)	5.45(−5)	1.54(−5)	8.10(−6)	5.41(−6)	4.01(−6)	1.81(−6)	3.38(−7)
	75	2.95(−4)	6.05(−5)	1.70(−5)	8.83(−6)	5.88(−6)	4.36(−6)	1.94(−6)	3.62(−7)
0.025	−75	6.69(−3)	2.83(−2)	1.66(−1)	3.50(−1)	4.96(−1)	6.05(−1)	6.96(−1)	3.72(−1)
	−50	1.49(−3)	1.25(−3)	2.29(−3)	4.85(−3)	6.81(−3)	9.53(−3)	2.64(−3)	2.93(−6)
	−25	1.13(−3)	2.92(−4)	1.12(−4)	7.36(−5)	5.57(−5)	4.39(−5)	1.09(−5)	2.02(−6)
	0	1.08(−3)	2.43(−4)	7.37(−5)	4.12(−5)	2.85(−5)	2.18(−5)	1.01(−5)	1.91(−6)
	25	1.16(−3)	2.63(−4)	8.04(−5)	4.37(−5)	2.99(−5)	2.28(−5)	1.05(−5)	1.97(−6)
	50	1.30(−3)	2.95(−4)	8.74(−5)	4.77(−5)	3.24(−5)	2.44(−5)	1.12(−5)	2.10(−6)
	75	1.30(−3)	3.17(−4)	9.74(−5)	5.27(−5)	3.49(−5)	2.64(−5)	1.20(−5)	2.26(−6)
0.05	−75	1.48(−2)	6.00(−2)	2.31(−1)	4.24(−1)	5.40(−1)	6.27(−1)	6.66(−1)	3.63(−1)
	−50	4.33(−3)	3.32(−3)	5.62(−3)	9.16(−3)	1.10(−2)	1.33(−2)	2.74(−3)	1.15(−5)
	−25	3.24(−3)	9.17(−4)	3.79(−4)	2.48(−4)	1.70(−4)	1.24(−4)	4.20(−5)	7.84(−6)
	0	3.41(−3)	8.42(−4)	2.75(−4)	1.57(−4)	1.10(−4)	8.44(−5)	3.94(−5)	7.44(−6)
	25	3.41(−3)	9.41(−4)	3.01(−4)	1.66(−4)	1.16(−4)	8.76(−5)	4.08(−5)	7.67(−6)
	50	4.05(−3)	1.04(−3)	3.35(−4)	1.84(−4)	1.25(−4)	9.53(−5)	4.37(−5)	8.14(−6)
	75	4.05(−3)	1.20(−3)	3.58(−4)	2.02(−4)	1.35(−4)	1.02(−4)	4.78(−5)	8.75(−6)

Table 1 (continued). Mean Squared Error of the Nonadaptive Group-Testing Estimator

p	% away	N							
		5	10	20	30	40	50	100	500
0.10	−75	3.46(−2)	1.05(−1)	3.09(−1)	4.66(−1)	5.49(−1)	6.08(−1)	6.16(−1)	3.61(−1)
	−50	1.03(−2)	7.98(−3)	8.92(−3)	1.71(−2)	2.02(−2)	1.99(−2)	4.00(−3)	4.41(−5)
	−25	9.23(−3)	2.88(−3)	1.09(−3)	7.01(−4)	5.06(−4)	3.63(−4)	1.58(−4)	3.00(−5)
	0	9.23(−3)	2.81(−3)	9.87(−4)	5.79(−4)	4.10(−4)	3.17(−4)	1.49(−4)	2.82(−5)
	25	1.10(−2)	3.09(−3)	1.10(−3)	6.15(−4)	4.30(−4)	3.38(−4)	1.57(−4)	2.95(−5)
	50	1.10(−2)	3.72(−3)	1.22(−3)	7.08(−4)	4.79(−4)	3.54(−4)	1.72(−4)	3.18(−5)
	75	1.10(−2)	3.72(−3)	1.42(−3)	7.92(−4)	5.21(−4)	4.13(−4)	1.85(−4)	3.37(−5)
0.20	−75	5.40(−2)	1.53(−1)	3.13(−1)	4.53(−1)	5.06(−1)	5.30(−1)	5.28(−1)	3.45(−1)
	−50	3.04(−2)	1.77(−2)	1.89(−2)	2.30(−2)	1.91(−2)	1.95(−2)	3.08(−3)	1.53(−4)
	−25	2.31(−2)	8.36(−3)	3.37(−3)	2.04(−3)	1.51(−3)	1.25(−3)	5.38(−4)	1.04(−4)
	0	2.31(−2)	8.36(−3)	3.28(−3)	1.97(−3)	1.43(−3)	1.12(−3)	5.23(−4)	9.96(−5)
	25	3.20(−2)	9.80(−3)	3.66(−3)	2.08(−3)	1.53(−3)	1.21(−3)	5.42(−4)	1.06(−4)
	50	3.20(−2)	9.80(−3)	4.68(−3)	2.37(−3)	1.76(−3)	1.40(−3)	5.90(−4)	1.16(−4)
	75	3.20(−2)	1.60(−2)	4.68(−3)	3.08(−3)	1.76(−3)	1.40(−3)	6.88(−4)	1.36(−4)

[1] $p_0 = (1 + \%\ \text{away}/100)p$

[2] Mean squared errors are given in exponential form, with the power in parentheses: "2.09(−3)" means 2.09×10^{-3}.

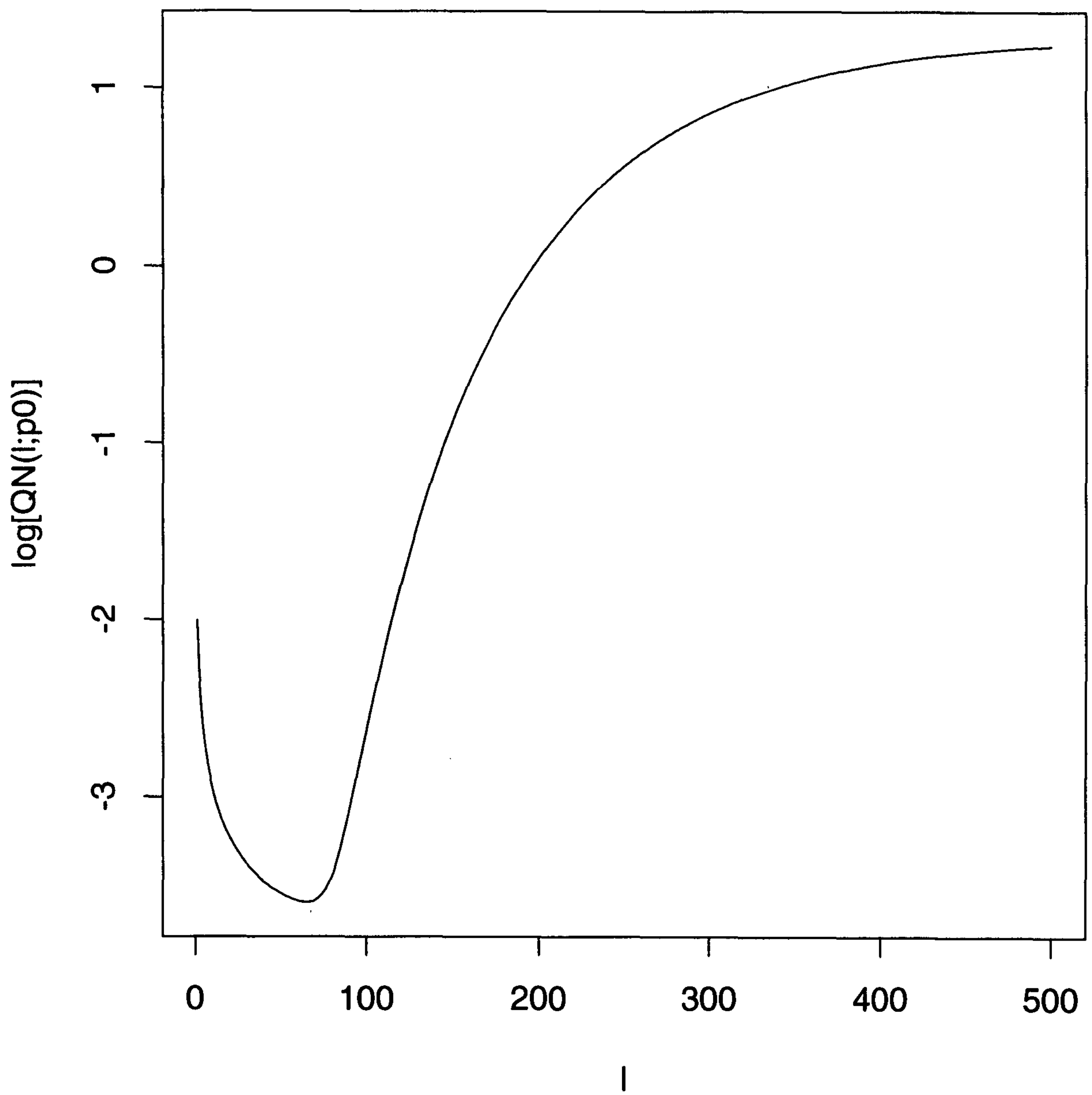

Figure 1. The log of the normalized mean squared error function versus group size, for N=20 and p0=0.01.

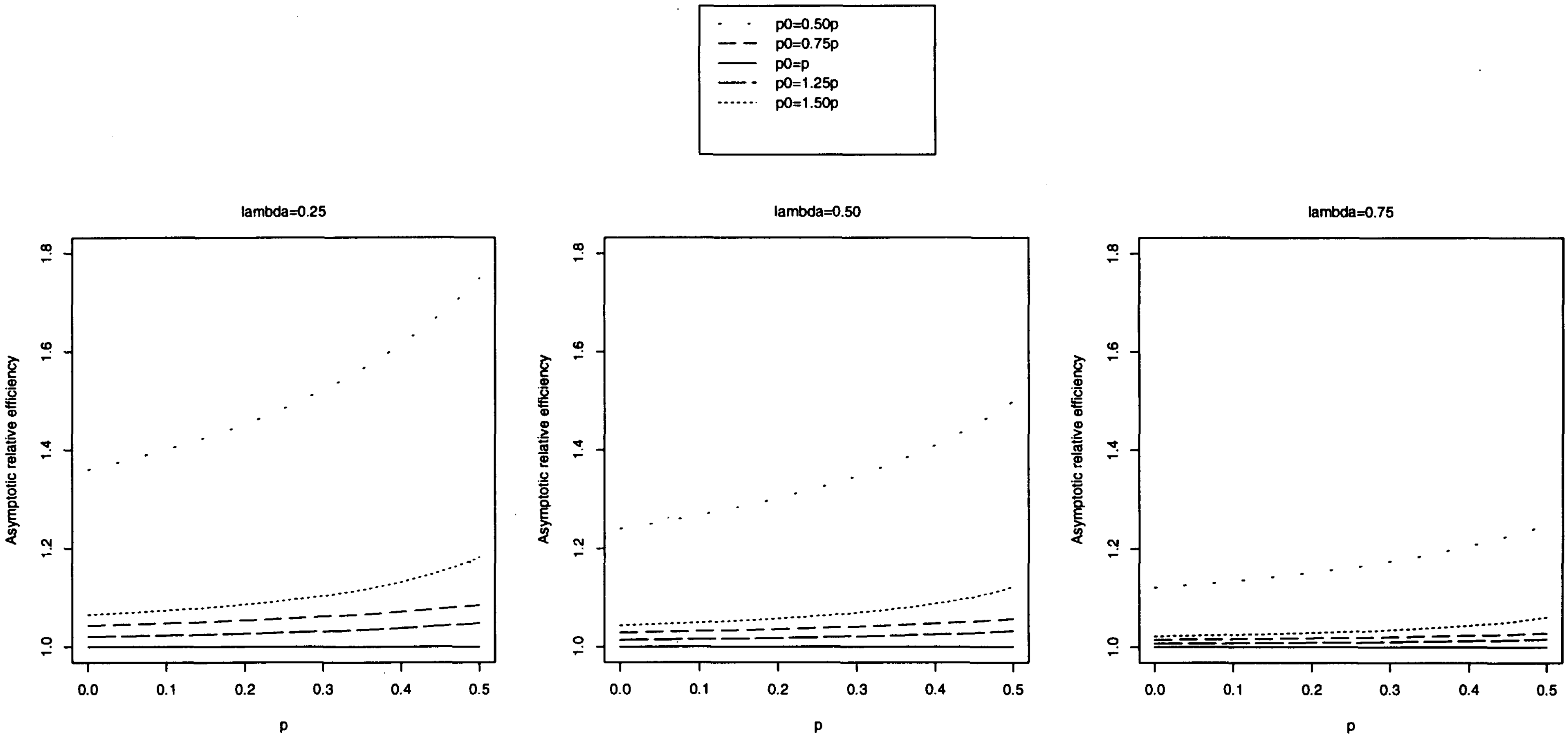

Figure 2. Asymptotic efficiency of the adaptive estimator relative to the nonadaptive estimator

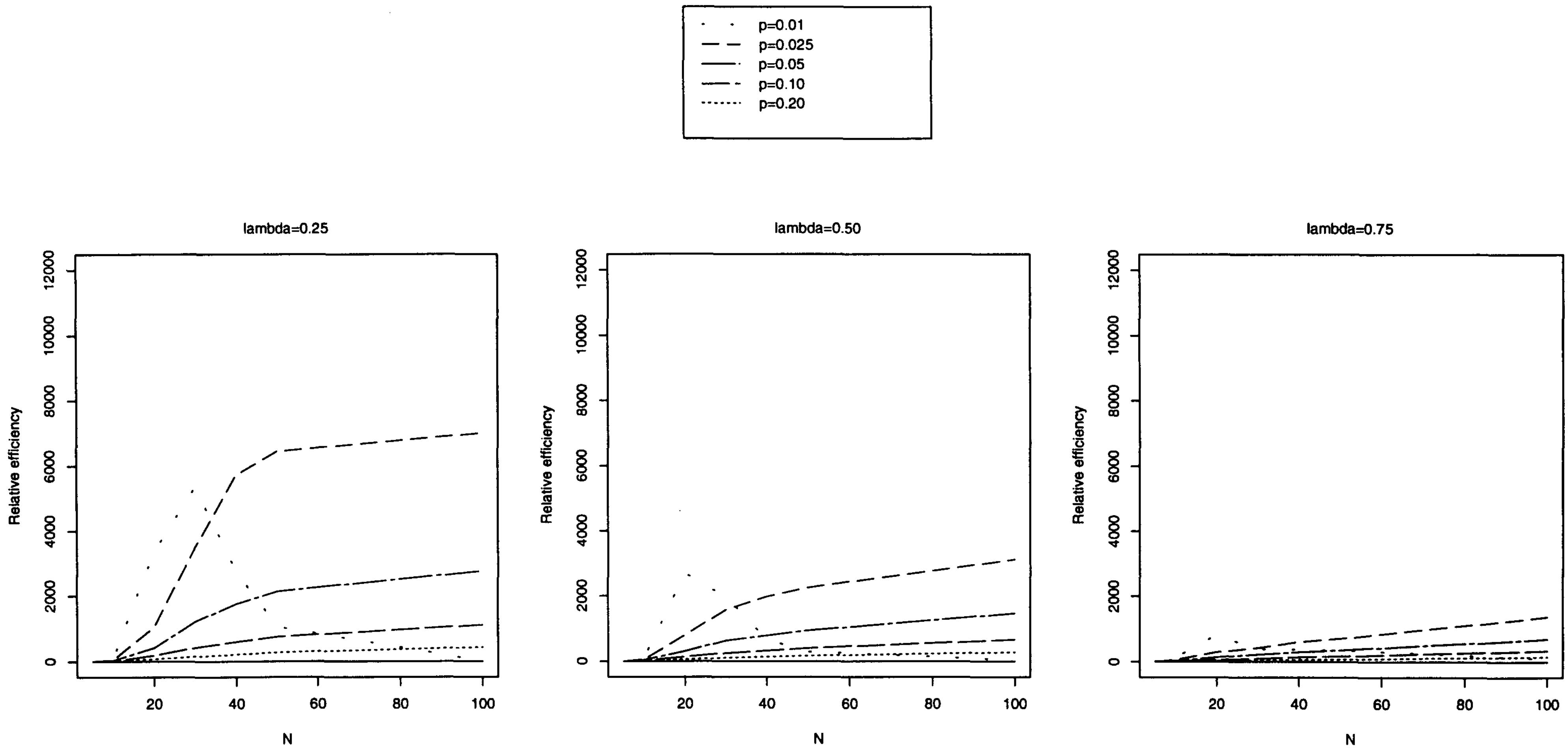

Figure 3. Small-sample efficiency of the adaptive estimator relative to the nonadaptive estimator, p0=0.25p.

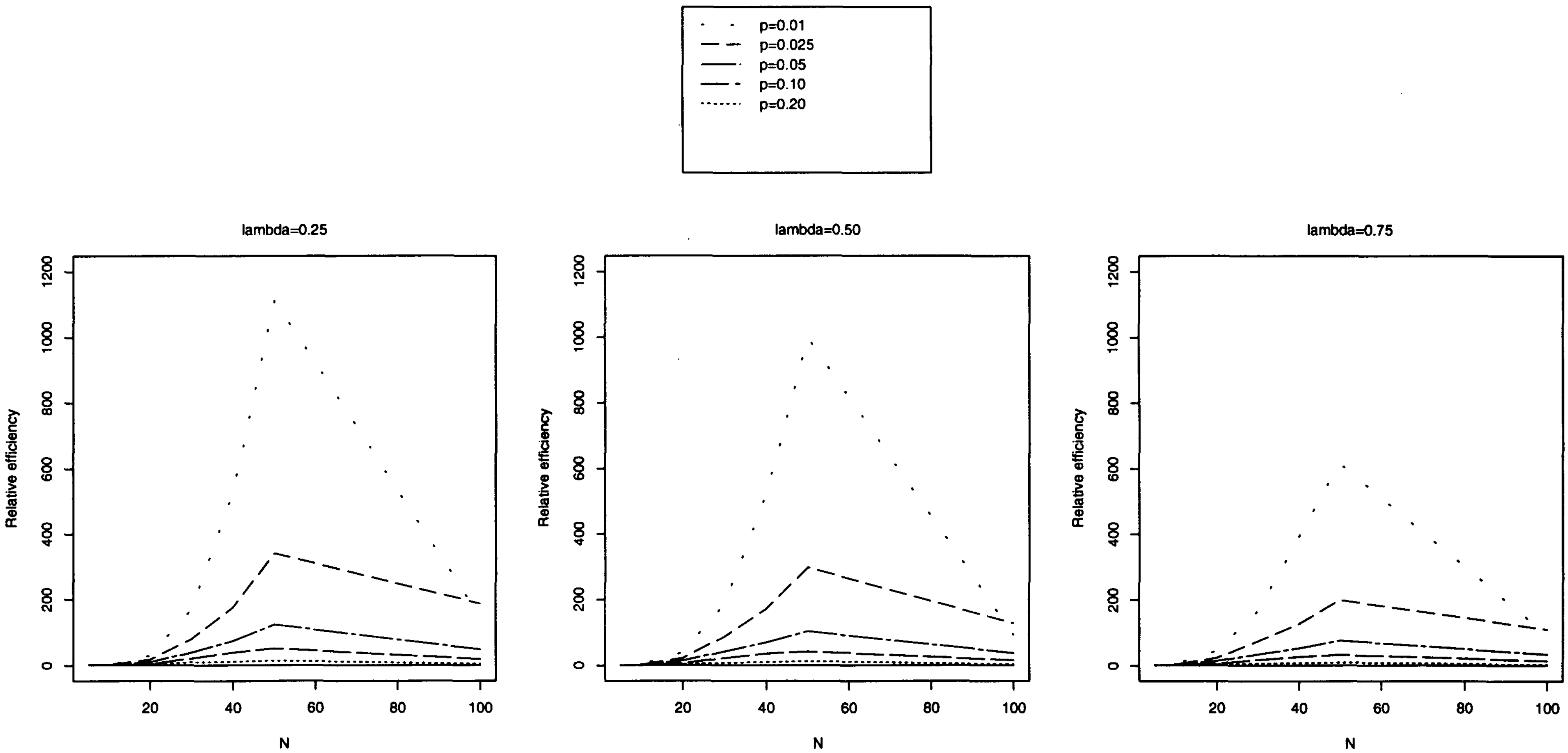

Figure 4. Small-sample efficiency of the adaptive estimator relative to the nonadaptive estimator, p0=0.5p.

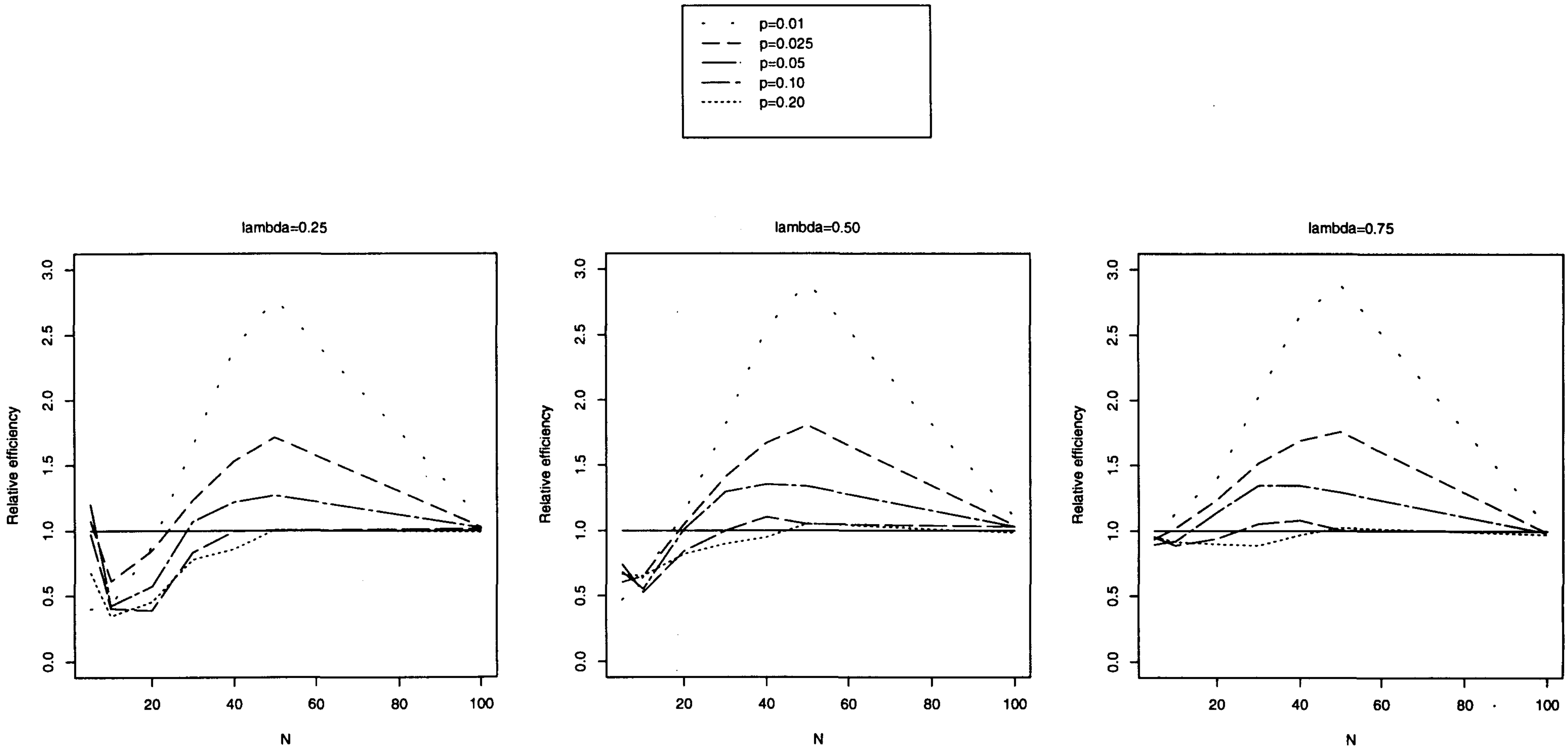

Figure 5. Small-sample efficiency of the adaptive estimator relative to the nonadaptive estimator, p0=0.75p.

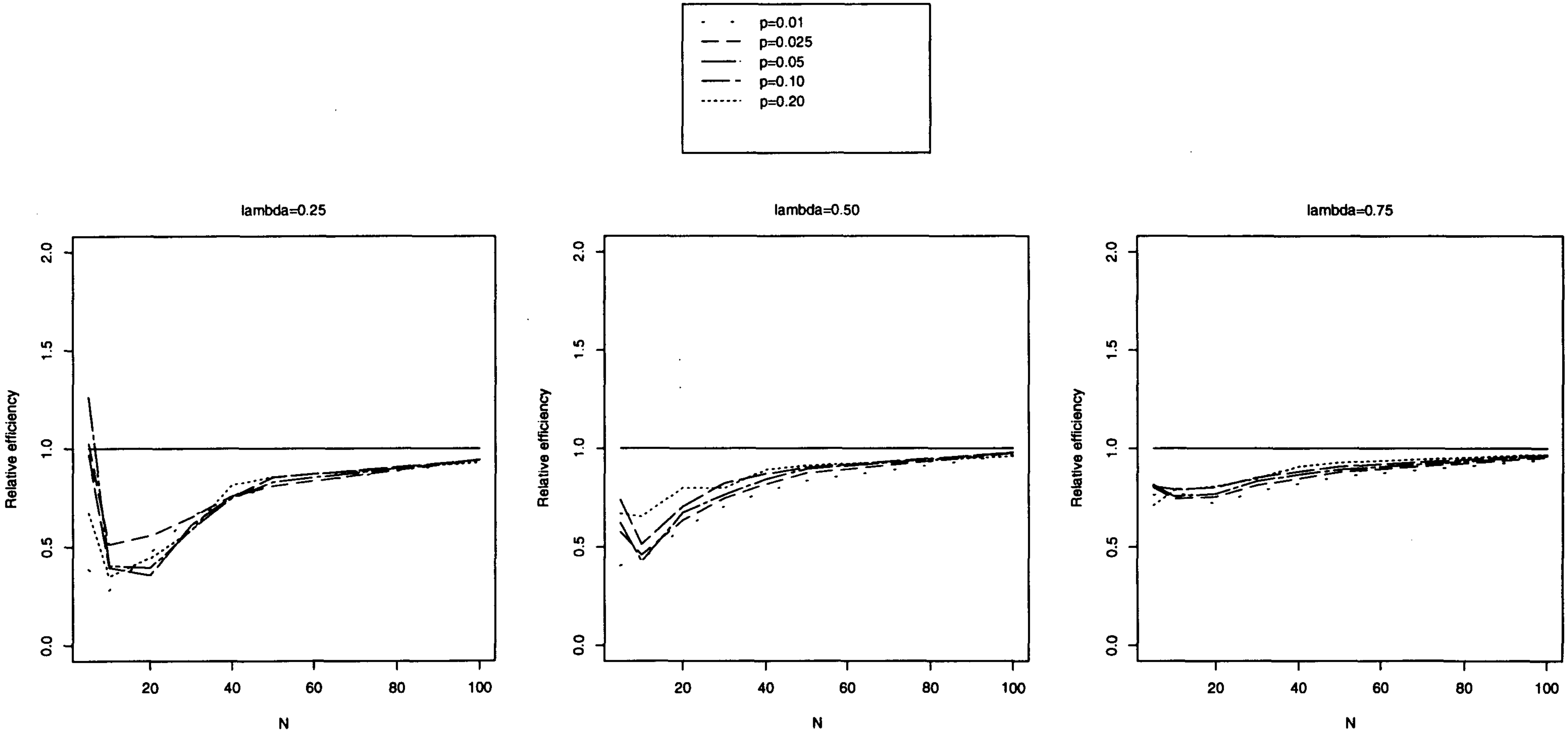

Figure 6. Small-sample efficiency of the adaptive estimator relative to the nonadaptive estimator, p0=p.

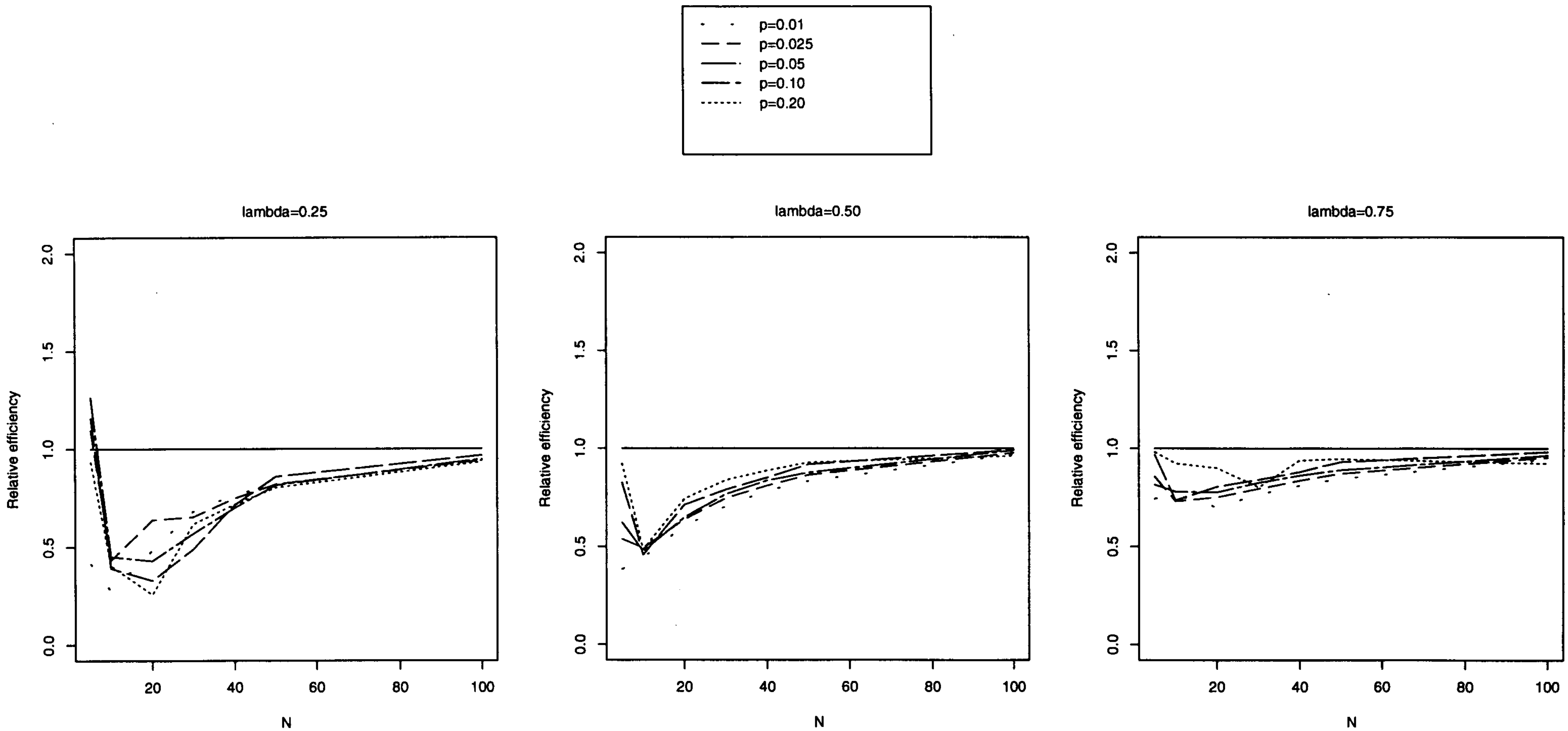

Figure 7. Small-sample efficiency of the adaptive estimator relative to the nonadaptive estimator, p0=1.25p.

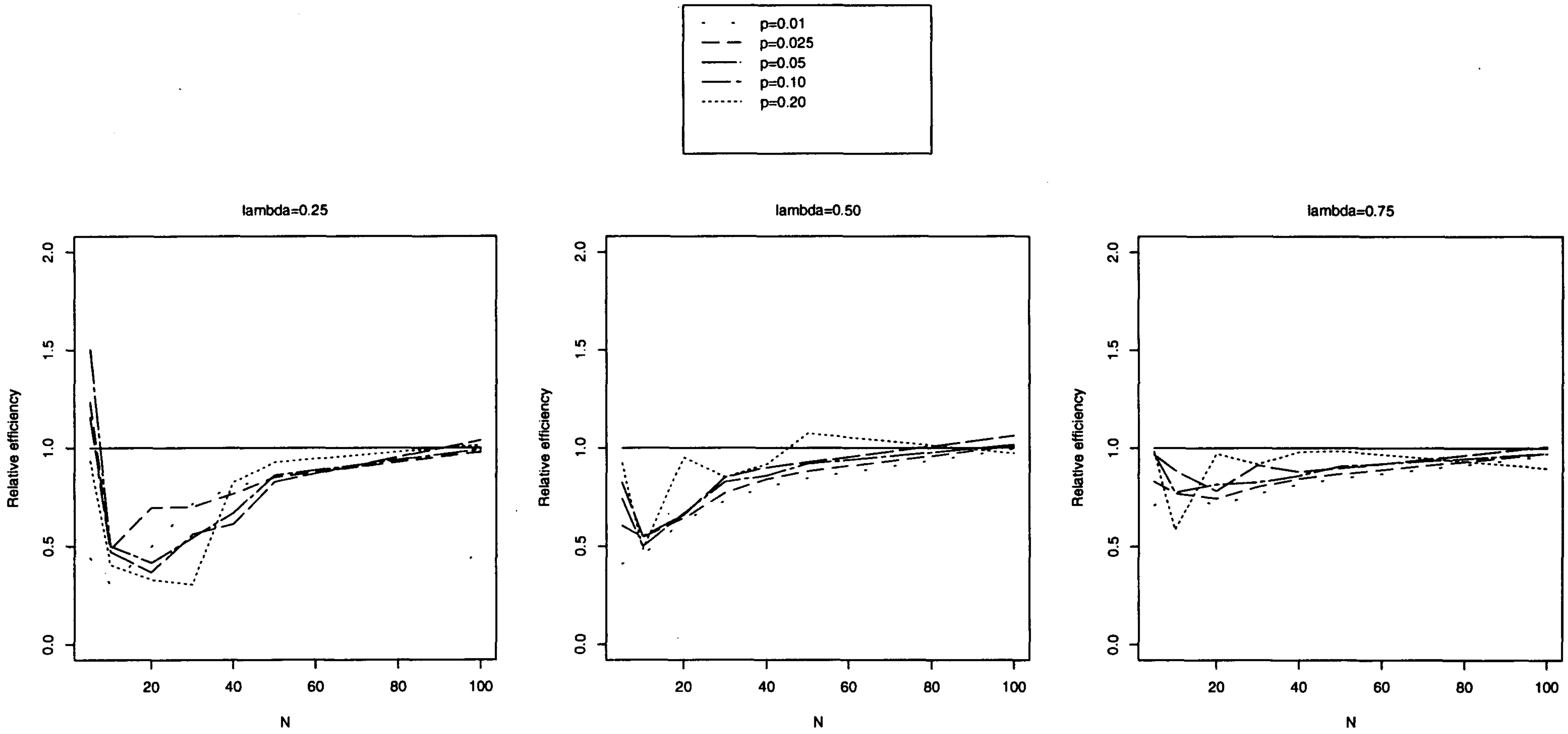

Figure 8. Small-sample efficiency of the adaptive estimator relative to the nonadaptive estimator, p0=1.5p.

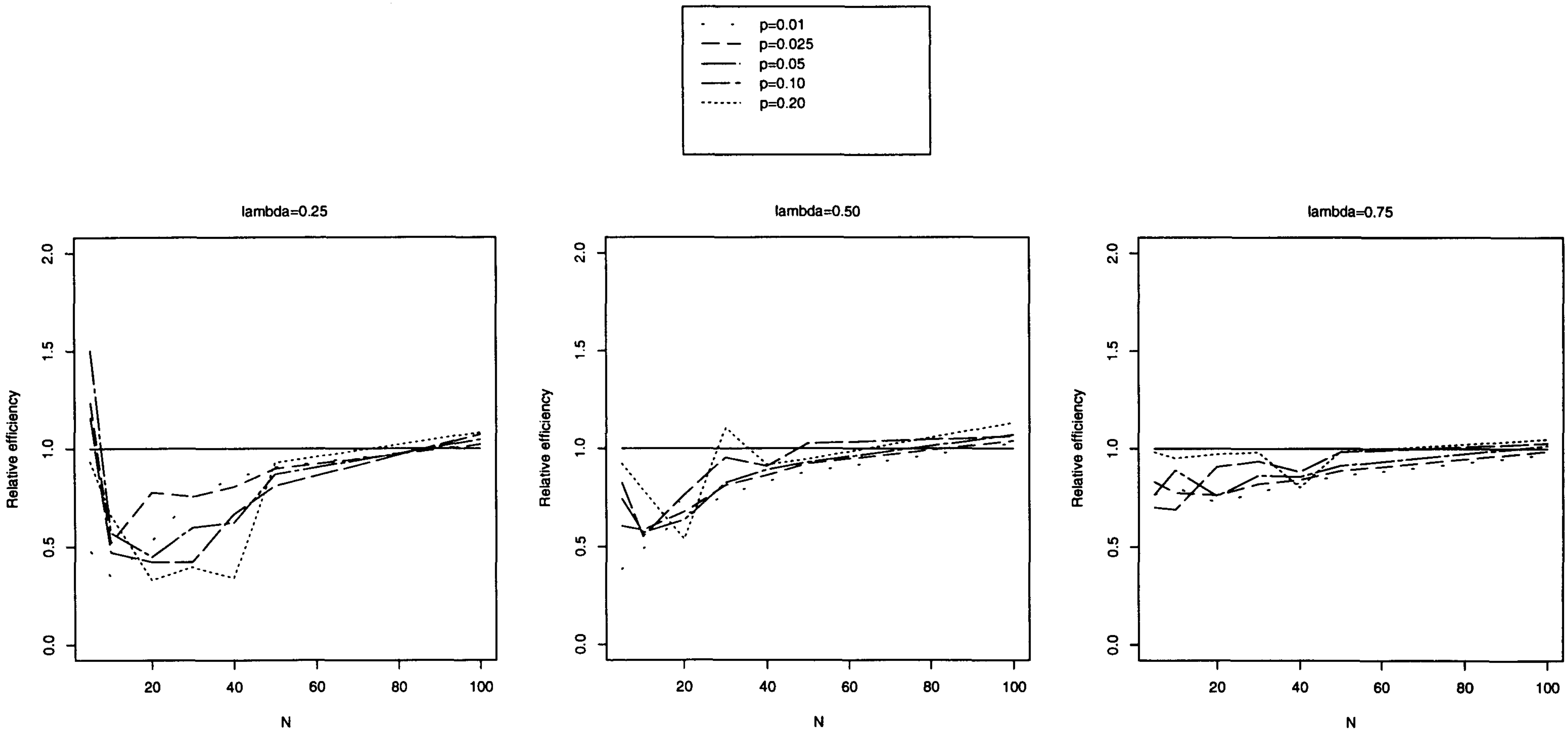

Figure 9. Small-sample efficiency of the adaptive estimator relative to the nonadaptive estimator, p0=1.75p.

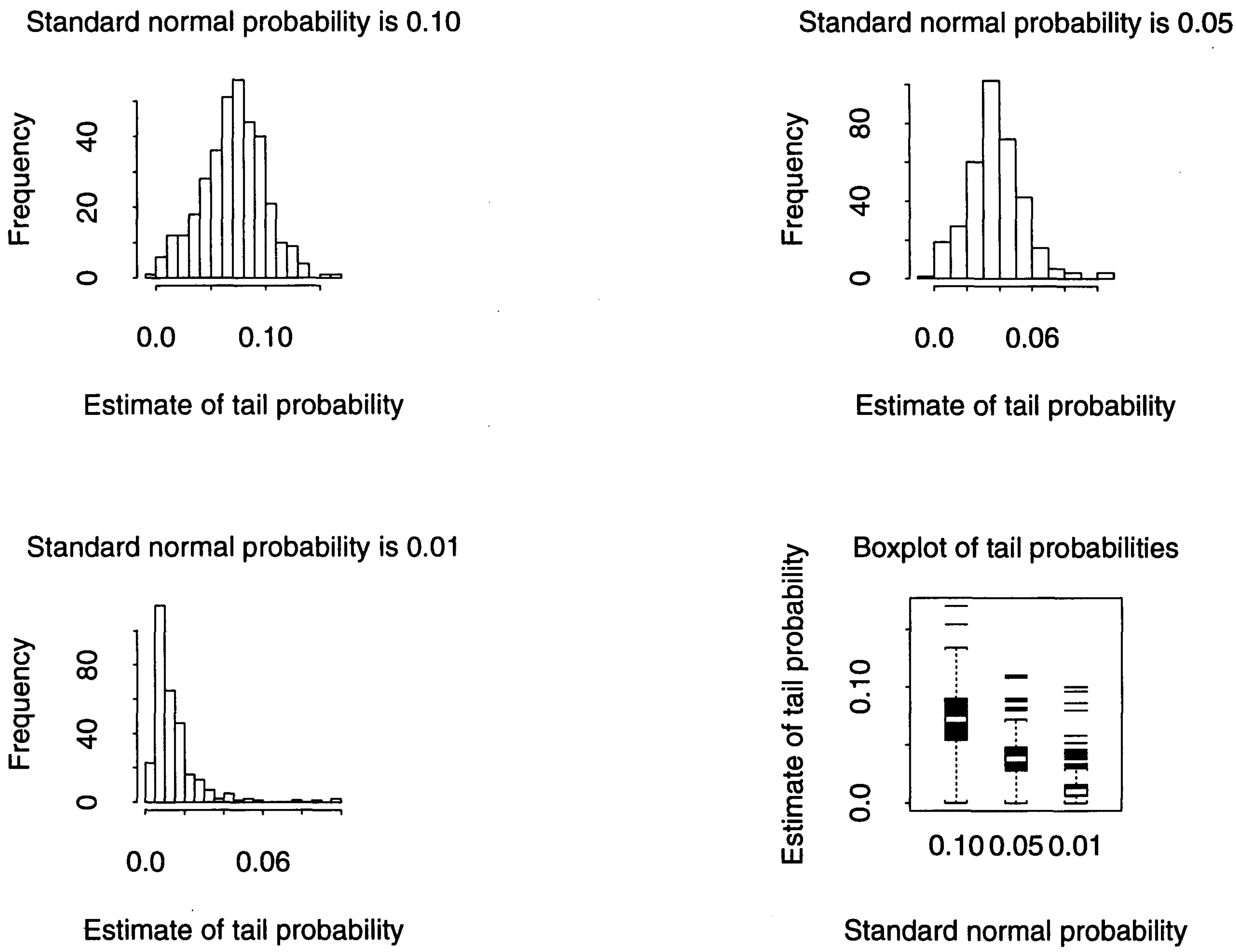

Figure 10. Estimates of the tail probability of the 'normalized' adaptive estimator.