In *Journal American Statistical Association* **93** (1998), pp. 1502–1511.

Sequentially Deciding Between Two Experiments for Estimating a Common Success Probability

Janis Hardwick[1]
University of Michigan

Connie Page
Michigan State University

Quentin F. Stout[1]
University of Michigan

Abstract: To estimate a success probability p, two experiments are available: individual Bernoulli(p) trials or the product of r individual Bernoulli(p) trials. This problem has its roots in reliability where either single components can be tested or a system of r identical components can be tested. A total of N experiments can be performed, and the problem is to sequentially select some combination (allocation) of these two experiments, along with an estimator of p, to achieve low mean square error of the final estimate. This scenario is similar to that of the better-known group testing problem, but here the goal is to estimate failure rates rather than to identify defective units. The problem also arises in epidemiological applications such as estimating disease prevalence.

Information maximization considerations, and analysis of the asymptotic mean square error of several estimators, lead to the following adaptive procedure: use the maximum likelihood estimator to estimate p, and if this estimator is below (above) the cut-point a_r, then observe an individual (product) trial at the next stage. In a Bayesian setting with squared error estimation loss and suitable regularity conditions on the prior distribution, this adaptive procedure, replacing the maximum likelihood estimator with the Bayes estimator, will be asymptotically Bayes.

Exact computational evaluations of the adaptive procedure for fixed sample sizes show that it behaves roughly as the asymptotics predict. The exact analyses also show parameter regions for which the adaptive procedure achieves negative regret, as well as regions for which it achieves normalized mean squared error superior to that asymptotically possible.

An example and a discussion of extensions conclude the work.

Keywords: group testing, adaptive design, composite sampling, pooled testing, omniscient allocation, prevalence estimation, reliability

 Last modified: 23 April 1998.

[1]Research supported in part by National Science Foundation under grants DMS-9157715 and DMS-9504980.

1 Introduction

Consider a sequence of Bernoulli trials with success probability p. To estimate p, two experiments can be performed: either an individual trial outcome can be observed (the *p-experiment*) or the product of r individual trial outcomes can be observed (the *p^r-experiment*), where r is an integer ≥ 2. A total of N experiments (tests) can be performed, and the problem is to select some combination (allocation) of these two experiments along with an estimator of p to achieve low mean square error of the terminal estimator. The experiments can be selected sequentially, so that at each stage, information available at that stage can be used to determine which experiment to carry out at the next stage.

Before continuing, some comments on the origin and application of this problem are in order. The p^r-experiment is a slightly disguised version of the well-studied grouped data experiment with groups of size r. In the grouped data setting, the goal is to estimate the failure probability, $q = 1 - p$, and using groups of sizes other than 1 can reduce the cost of testing (Sobel and Elashoff, 1975) and can lower the variance of the resulting estimator (Chen and Swallow, 1990). Reliability settings, in which components can be tested either individually or as a system of r identical components in series, are prime examples of situations in which group testing can be useful (Easterling and Prairie, 1971). Other group testing scenarios arise in environmental monitoring where sample units of soil or plant matter are combined and tested for toxins. In these settings, the term "group testing" is often replaced by "composite sampling". See Lancaster and Keller-McNulty (1996) for a review of composite sampling methods.

A further application, examined in some detail in Section 9, is that of estimating prevalence. Gastwirth and Hammick (1989), for example, apply group testing methods to estimate the prevalence of HIV antibodies among subpopulations. In screening scenarios of this sort, group testing is particularly desirable because it provides for donor privacy, an issue of serious concern among individuals at risk for HIV. The "pooled testing" of Tu, Litvak and Pagano (1995) is another example in which group testing is used to estimate HIV prevalence.

In the problem considered here, it is assumed that a "natural" r exists for the grouping of items. In this case, the sampling options are restricted to only the two experiments: the p-experiment and the p^r-experiment. The reliability setting, in which a system requires that r units in series be tested, represents a scenario in which such an assumption is clearly viable. Note, however, that it is sometimes useful to seek the optimal r for a specific application. This point is discussed briefly in Section 9 and addressed more thoroughly via the two-stage sampling procedures of Hughes-Oliver and Swallow (1994).

In Section 2, notation is given and the problem is precisely defined. In Section 3, allocation that maximizes information is derived, and in Section 4, pooling data across experiments is discussed and several estimators are analyzed. In Section 5, allocation that minimizes asymptotic mean square error is derived for each of the estimators. In Section 6, an ad-hoc adaptive allocation procedure is proposed, and it is shown that not all of the estimators are consistent when combined with an arbitrary allocation procedure. In Section 7, the problem is placed in a Bayesian framework with squared error estimation loss. Then, under regularity conditions, the adaptive allocation of Section 6, replacing the maximum likelihood estimator with the Bayes estimator, is shown to be asymptotically Bayes. In Section 8, an exact approach is taken to evaluating and optimizing allocation policies for fixed values of N. Here it is shown that the ad-hoc adaptive rule of Section 6 has negative regret with respect to the optimal best fixed sample size rule that can be generated when the parameter is known. It is also shown that, for some values of the parameter, the ad-hoc adaptive rule achieves a normalized mean squared error that is smaller than the asymptotic limit. For comparison purposes, the optimal adaptive allocation rule is also computed, assuming that the parameter is known. In Section 9, an application of the methods in Section 6 is discussed. There, the group testing approach of Gastwirth and Hammick (1989) is compared with the individual testing approach of Nusbacher et al. (1986) and with the methods proposed in the present article. Finally, in Section 10 the extension from 2 experiments to arbitrarily many is briefly examined.

2 Notation and Problem Statement

The problem is set up in its fully sequential form although much of the development in the next sections will not use all of this notation. Let $X_{11}, X_{12}, \ldots$ be a sequence of independent and identically distributed Bernoulli(p) random variables that are independent of $X_{r1}, X_{r2}, \ldots$, independent and identically distributed Bernoulli(p^r) random variables. The success probability p is restricted to (0,1) throughout the paper.

A total of N tests (experiments) will be done, where at each stage the decision to observe an X_1 or an X_r can be made based on past observations. More precisely, an allocation rule is a sequence $\mathbf{d} = (d_1, \ldots, d_N)$ such that for $k = 1, \ldots, N$, d_k takes values 0 or 1, and is measurable $\{Z_1, \ldots, Z_{k-1}\}$, where $Z_i = d_i X_{1i} + (1 - d_i) X_{ri}$. Thus, d_i indicates the population from which the i^{th} observation or test is sampled, with "1" indicating an X_1-observation or the p-experiment, and "0" indicating a X_r-observation or the p^r-experiment. A *terminal estimator of* p must be measurable $\{Z_1, \ldots, Z_N\}$. Finally, let $m_k = \sum_{i=1}^{k} d_i$ be the total number of observations taken from the p-experiments at stage k, and let $n_k = k - m_k$ be the total number of observations taken from the p^r-experiment at stage k, where $k = 1, 2, \ldots, N$. Sometimes the k subscript will be dropped.

Since the final goal is estimation of p, an allocation scheme and estimator will be evaluated as a pair by the mean square error of the terminal estimator. That is, the mean square error of using allocation $\mathbf{d}$ and estimator $\hat{p}$ is given by $\text{MSE}_p(\mathbf{d}, \hat{p}) = E_p(\hat{p} - p)^2$. In the Bayesian framework, the problem of selecting both the allocation rule and the estimator reduces to selecting only the allocation rule and using the Bayes estimator. However, in the non-Bayesian framework, the choice of estimator is not so obvious, and as will be noted in Section 6, the allocation rule and the estimator can interact.

3 Maximum Information Allocation

In this section, the problem of selecting an estimator of p is ignored, and only allocation is considered. The criterion used for allocation will be maximizing the Fisher information, and the best nonrandom allocation will be found. As is typical with such optimal nonrandom allocations, the rule will depend on the unknown parameter p, but will suggest the form of an adaptive rule, and the relationship between the MSE's of estimators and Fisher information will make the adaptive rule efficient.

In typical sequential allocation problems, the different experiments give information about different parameters. However, in this problem both experiments give information about the same parameter,

r	2	5	10	20	50	100
a_r	0.333	0.536	0.679	0.792	0.892	0.937

Table 1: Cut-point Values

although one experiment will give more information depending on the actual value of the parameter. In particular, the Fisher information about p contained in a single observation of the p-experiment is

$$\mathcal{I}_{X_1}(p) = \frac{1}{pq}, \qquad \text{where} \quad q = 1 - p,$$

and the Fisher information about p contained in a single observation of the p^r-experiment is

$$\mathcal{I}_{X_r}(p) = \frac{r^2 p^{r-2}}{1 - p^r}.$$

It is easy to show that $\mathcal{I}_{X_1}(p) > \mathcal{I}_{X_r}(p)$ if and only if $p < a_r$, where a_r is the unique root in (0,1) of the equation in p,

$$p^r(1 - r^2) + r^2 p^{r-1} - 1 = 0.$$

Note that a_r is a function only of r, which is known, and hence a_r can be determined explicitly.

Proposition 3.1 *If N tests are available, then the allocation that maximizes the information about p is*

$$m_N = \begin{cases} N \text{ (observe all } X_1\text{'s)} & \text{if } p < a_r \\ 0 \text{ (observe all } X_r\text{'s)} & \text{if } p > a_r \\ \text{arbitrary} & \text{if } p = a_r \end{cases}$$

(The informations are equal at $p = a_r$). □

Thus, the maximum information allocation is to observe only X_1's (the p-experiment) if $p < a_r$, or only X_r's (the p^r-experiment) if $p > a_r$. This will be denoted as *a_r-cut allocation*, and recall that a_r depends only on r. However, the region where one experiment is better than the other depends on the unknown parameter p. Thus, the obvious adaptive rule is suggested where p is estimated at each stage, and the next observation is allocated depending the relationship between the estimated p and the cut a_r.

Some of the values of a_r (first reported in Loyer, 1983) are noted in Table 1. As r increases to infinity, the cut-point a_r tends to 1, and the region over which the p^r-experiment is better shrinks.

This section is concluded by evaluating the need for sequential allocation. This is done by comparing the information in each experiment. Consider the ratio

$$\frac{\min\{\mathcal{I}_{X_1}(p), \mathcal{I}_{X_r}(p)\}}{\max\{\mathcal{I}_{X_1}(p), \mathcal{I}_{X_r}(p)\}}$$

over the range of p. Were this ratio bounded below by, say, .98, then using the nonoptimal experiment would never result in more than a 2% loss of information, greatly limiting the worth of adaptive allocation. Since adaptive allocation is somewhat more complicated than fixed allocation, there needs to be sufficient benefit to justify its utilization.

Proposition 3.2

$$\frac{\mathcal{I}_{X_1}(p)}{\max\{\mathcal{I}_{X_1}(p), \mathcal{I}_{X_r}(p)\}} = \begin{cases} 1 & \text{for } p \le a_r \\ (1-p^r)/r^2qp^{r-1} \quad \textit{(which is} \ge 1/r\textit{)} & \text{for } p > a_r \end{cases}$$

$$\frac{\mathcal{I}_{X_r}(p)}{\max\{\mathcal{I}_{X_1}(p), \mathcal{I}_{X_r}(p)\}} = \begin{cases} r^2qp^{r-1}/(1-p^r) \quad \textit{(which tends to 0 as p tends to 0)} & \text{for } p \le a_r \\ 1 & \text{for } p > a_r \end{cases}$$

Proof. This is simply algebra. □

Proposition 3.2 indicates that if the p^r-experiment is used, then for p sufficiently small, the information obtained can be arbitrarily close to 0% of that possible when the p-experiment is used. On the other hand, if the p-experiment is used, then the information obtained never falls below (100/r)% of the maximal information obtainable, approaching this bound when p tends to 1. Thus, adaptive allocation can be worthwhile for increasing information. Also, using the p-experiment would be the more conservative approach since one never loses more than (100/r)% of the maximal information obtainable. Note that the bound decreases as r increases, so that at $r = 2$, no more than 50% of the optimal can be lost, but at $r = 10$, one might get only 10% of the optimal.

4 Estimators of p

Since both experiments give information about p, and an adaptive allocation procedure would typically allocate to both experiments, there is the question of how to combine or pool data across experiments. Several estimators are presented here and their properties are derived.

Throughout this section, m is the number of observations from the p-experiment, n is the number of observations from the p^r-experiment, and $m + n = N$ is the fixed total number of experiments.

If m observations from only the p-experiment are used to estimate p, then the best estimator (uniform minimum variance unbiased and the maximum likelihood) is the sample mean $\bar{X}_1 = \frac{1}{m}\sum_{i=1}^{m} X_{1i}$, and its mean square error is equal to its variance, pq/m. Let $\hat{p}_{x_1}$ denote this estimator.

If n observations from only the p^r-experiment are used to estimate p, then there is no unbiased estimator. However, the maximum likelihood estimator is $\bar{X}_r^{1/r}$, the rth root of the sample mean, $\bar{X}_r = \frac{1}{n}\sum_{j=1}^{n} X_{rj}$, which is equivalent to the usual estimator of p in grouped data experiments using groups of size r. Let $\hat{p}_{x_r}$ denote this estimator. The MSE of $\hat{p}_{x_r}$ can be computed for different values of p and n using binomial probabilities, and as n tends to infinity, $n\mathrm{MSE}_p(\hat{p}_{x_r})$ tends to $1/\mathcal{I}_{X_r}(p) = (1-p^r)/r^2p^{r-2}$. This has been noted and studied by Sobel and Elashoff (1975). Some comparisons of the exact mean square and this asymptotic form are made in Loyer (1983).

Next, the maximum likelihood estimator is derived for the general situation with observations of both experiments.

Theorem 4.1 *Given m observations of the p-experiment and n observations of the p^r-experiment, the maximum likelihood estimator of p, denoted $\hat{p}_{ml}$, is the unique root in $[0,1]$ of the equation*

$$p^r(rn + m) + (m - x_1)(p^{r-1} + \ldots + p) - (x_1 + ry) = 0,$$

where $x_1 = \sum_{i=1}^{m} X_{1i}$ and $x_r = \sum_{j=1}^{n} X_{rj}$.

Proof. Differentiating the logarithm of the joint likelihood function with respect to p gives the equation above. It is then straightforward to show the existence and uniqueness of a root in [0,1]. $\square$

For the case of $r = 2$, the maximum likelihood estimator can be given in closed form:

$$\hat{p}_{ml} = \frac{\sqrt{a^2(1-\hat{p}_{x_1})^2 + 4a\hat{p}_{x_1} + 4(1-a)\hat{p}_{x_r}{}^2} - a(1-\hat{p}_{x_1})}{2},$$

where $a = m/(2n+m)$.

Other natural estimators of p are weighted averages of $\hat{p}_{x_1}$ and $\hat{p}_{x_r}$, where the weights could depend on m, n, and N. Two particular weight choices are considered below:

- The *constant weights estimator:* $\hat{p}_\alpha = (\alpha)\hat{p}_{x_1} + (1-\alpha)\hat{p}_{x_r}$.
- The *weight proportional to sample size estimator:* $\hat{p}_s = (m/N)\hat{p}_{x_1} + (n/N)\hat{p}_{x_r}$.

Finally, we mention briefly that ratio estimators such as $\hat{p}_R = (\bar{X}_r/\bar{X}_1)^{\frac{1}{r-1}}$ have been considered for estimating p, in the context of model validation (Chen and Swallow, 1990). However, these estimators are neither efficient nor consistent (in the allocation sense), and thus, are not included in this work. Details on the properties of such ratio estimators in this setting can be found in Hardwick, Page, and Stout (1996).

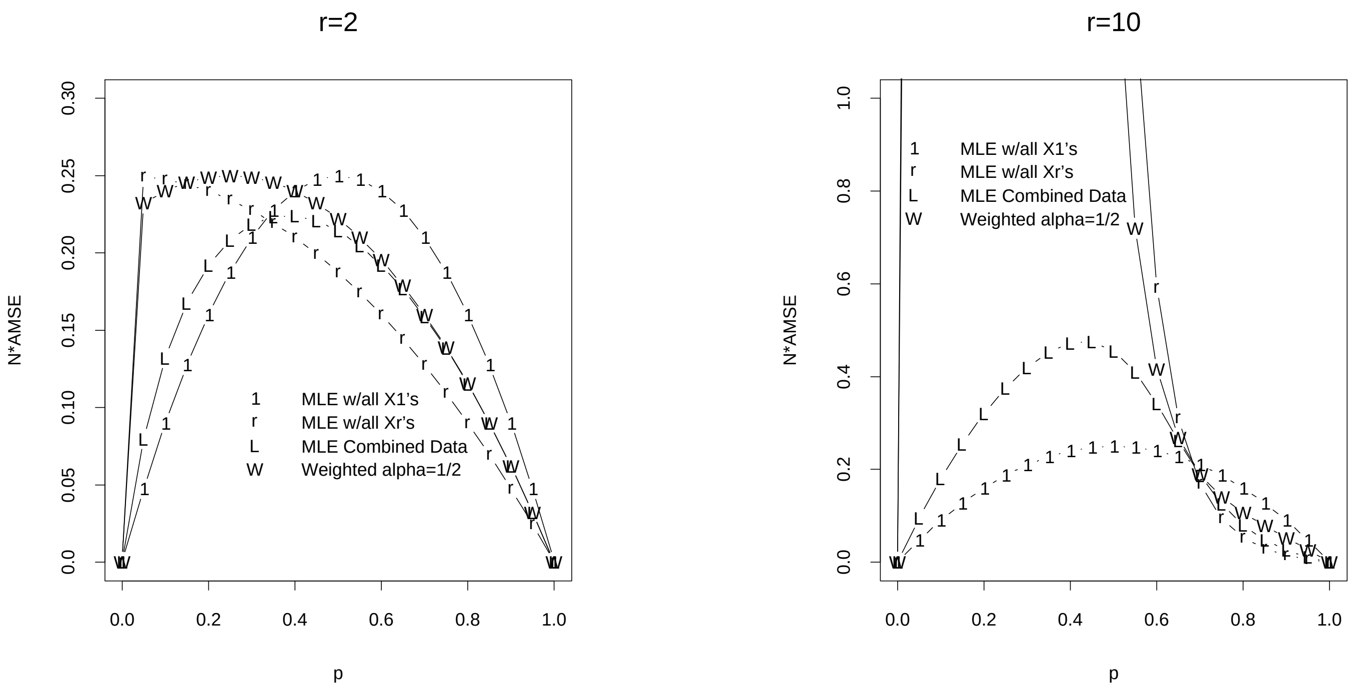
To determine the MSE's of the estimators, note that for the weighted average estimators, since $\hat{p}_{x_1}$ is unbiased for p, $\mathrm{MSE}_p(\hat{p}_\alpha) = \alpha^2 pq/m + (1-\alpha)^2 \mathrm{MSE}_p(\hat{p}_{x_r})$. For the other estimators, their MSE's can be computed exactly for any values of m, n, and p, but there is no apparent closed form for them. However, their asymptotic expressions are very tractable. Define the *asymptotic mean square error of estimator* $\hat{p}$ as $\mathrm{AMSE}_p(\hat{p}) = \mathrm{MSE}_p(\hat{p}) + o(1/n) + o(1/m)$ for p in (0,1).

Theorem 4.2 *The asymptotic MSE's of the estimators are as follows:*

$$\begin{aligned}
\mathrm{AMSE}_p(\hat{p}_{x_1}) &= \frac{pq}{m} = \mathrm{MSE}_p(\hat{p}_{x_1}) \\
\mathrm{AMSE}_p(\hat{p}_{x_r}) &= \frac{1-p^r}{nr^2p^{r-2}} \\
\mathrm{AMSE}_p(\hat{p}_{ml}) &= \frac{pq(1-p^r)}{m(1-p^r) + nr^2p^{r-1}q} \\
\mathrm{AMSE}_p(\hat{p}_\alpha) &= \frac{\alpha^2 pq}{m} + \frac{(1-\alpha)^2(1-p^r)}{nr^2p^{r-2}} \\
\mathrm{AMSE}_p(\hat{p}_s) &= \frac{mpq + n(1-p^r)/r^2p^{r-2}}{N^2}
\end{aligned}$$

Proof: This is straightforward. $\square$

To illustrate the relative performance of the estimators when either degenerate or balanced allocations are used, the values of the limit of $N*$AMSE as N tends to infinity are plotted in Figure 1 for $r = 2$ and 10. Note that $m = N$ allocation is used for $\hat{p}_{x_1}$, $n = N$ is used for $\hat{p}_{x_r}$, and $m = n = 0.5N$ is used for $\hat{p}_{ml}$ and $\hat{p}_\alpha$.



Allocation for L and W is 50/50 on p and p^r.

Figure 1: Normalized $N{\cdot}$AMSE's of Estimators, $r = 2, 10$. Note change in vertical scales.

5 Allocation Minimizing Asymptotic Mean Square Error

While fixed allocations performance of different estimators may be important in some instances, the problem of interest here is the pairing of an estimator and an allocation rule to lower the MSE of the "terminal" estimator of p. In this section, the estimators proposed in Section 4 are considered. Since their exact MSE's are analytically intractable while their first order asymptotic forms are simple, allocation rules that minimize the asymptotic MSE among nonrandom allocations are considered for each estimator. As is typical with optimal nonrandom rules, the allocation will depend on the unknown p, but will suggest an adaptive rule.

For each estimator in Section 4, the following theorem gives a nonrandom allocation, as a function of p, that minimizes the AMSE.

Theorem 5.1 *The AMSE's of estimators are minimized as follows:*

1. $\mathrm{AMSE}_p(\hat{p}_{ml})$ *and* $\mathrm{AMSE}_p(\hat{p}_s)$ *are minimized by the* a_r*-cut allocation described in Section 3.*
2. $\mathrm{AMSE}_p(\hat{p}_\alpha)$ *is minimized by allocating* m/n *in proportion* $\{r\alpha/(1-\alpha)\}\sqrt{p^{r-1}q/(1-p^r)}$.

Proof: This is straightforward algebra, from Theorem 4.2. □

For each estimator considered, the allocations reported in Theorem 5.1 will give the lowest AMSE for that estimator. These are referred to as *asymptotic omniscient fixed allocations* corresponding to the estimators. To compare the estimators based on the omniscient fixed allocations, the minimum AMSE's are needed. Let '*' on the AMSE denote the minimum asymptotic mean square error when the allocations of Theorem 5.1 are used.

Corollary 5.2 *The normalized* AMSE_p^* *values are as follows:*

$$N \cdot \mathrm{AMSE}_p^*(\hat{p}_{ml}) = N \cdot \mathrm{AMSE}_p^*(\hat{p}_s) = \min\{pq, (1-p^r)/r^2p^{r-2}\}$$
$$N \cdot \mathrm{AMSE}_p^*(\hat{p}_\alpha) = \left[\alpha\sqrt{pq} + (1-\alpha)\sqrt{(1-p^r)/(r^2p^{r-2})}\right]^2$$

Hence, for p *in* $(0,1)$*, for any weight* α *in* $(0,1)$*,*

$$\mathrm{AMSE}_p^*(\hat{p}_{ml}) = \mathrm{AMSE}_p^*(\hat{p}_s) \leq \mathrm{AMSE}_p^*(\hat{p}_\alpha),$$

with equality occurring only at $p = a_r$*.*

Proof. This follows easily from Theorem 5.1. □

Corollary 5.2 shows that the AMSE* for the maximum likelihood and weighted average estimators is the lowest among those considered. Thus, according to the asymptotic MSE criterion, the a_r-cut allocation, along with either the maximum likelihood estimator or the sample size weighted average estimator, should be used.

Note that caution should be exercised when using the AMSE formulas. These are first order approximations, ignoring terms $o(1/m) + o(1/n)$, and thus are valid when both m and n tend to infinity. However, the point of allocation here is to eventually end up on the better experiment. As will be seen in Section 6, the AMSE for the maximum likelihood estimator and the sample size weighted average estimator remains valid if one sample size does not tend to infinity. However, the constant weighted average is not even consistent if one of the sample sizes does not tend to infinity.

6 Adaptive Allocation

As mentioned, the allocation rules considered thus far are motivated by the nonrandom "optimal" allocations of Sections 3 and 5. In Section 3, the maximum information rule was the a_r-cut rule. In Section 5, this same allocation rule came out of minimizing the asymptotic MSE. Thus, the a_r-cut allocation rule is a natural choice for an adaptive rule, where an estimator of p is inserted in the a_r-cut form, and that estimator is updated at each stage for sequential allocation. Other results in Section 5 indicated that the estimator of p and the allocation should be compatible in the sense that the a_r-cut allocation should minimize the asymptotic MSE of the estimator in use.

To define an adaptive a_r-cut allocation, consider an estimator of p. Here, the term *estimator* denotes a sequence of estimators $\{\hat{p}_k\}_{k=1}^{\infty}$, where $\hat{p}_k$ is measurable $\{Z_1, \ldots, Z_k\}$. For $k = 1, \ldots, N$, $m_k + n_k = k$, where, at stage k, m_k and n_k are the number of observations on the p-experiment and on the p^r-experiment, respectively. The estimators suggested in Section 4 fill this requirement since they are defined for each pair (m, n). Once an estimator is selected, the ad-hoc a_r-cut allocation is defined in the obvious way:

The a_r-cut allocation with estimated p: At stage 1, take an observation from the p-experiment. At stage k, $1 < k \leq N$, observe from the p-experiment if and only if $\hat{p}_{k-1} \leq a_r$.

The tie at a_r has been decided in favor of the p-experiment because the worst of the consequences of a wrong decision is less (see Proposition 3.2).

The a_r-cut allocation aims eventually to allocate to the better experiment depending on the value of p. This requires that the estimator of p used with the allocation be consistent even if the number of observations on one of the experiments does not tend to infinity. This consistency is called "allocation" consistency:

Estimator $\{\hat{p}_k\}$ is (strongly) allocation consistent if $\hat{p}_k$ tends to p a.s. as k tends to infinity, for all p in $(0,1)$.

Theorem 6.1 *Estimators $\hat{p}_{ml}$ and $\hat{p}_s$ are allocation consistent, while all $\hat{p}_\alpha$ estimators are not.*

Proof. See the Appendix. □

The estimator used with the a_r-cut should be allocation consistent. However, while use of a consistent estimator is prudent, it does not guarantee low MSE of the terminal estimator. That requires efficiency of an estimator, as described below.

Ideally, an adaptive allocation should select the better experiment quickly (high precision of the estimator), and once the better experiment is being used, the estimator should approximate the individual maximum likelihood estimator for that experiment. That is, the goal is to use an estimator $\hat{p}$ with the property that

$$\mathrm{MSE}_p(\hat{p}) \approx \min\{\mathrm{MSE}_p(\hat{p}_{x_1}), \mathrm{MSE}_p(\hat{p}_{x_r})\}.$$

Thus, $\hat{p}$ should approximate $\hat{p}_{x_1}$ over the range where the p-experiment is better, and should approximate $\hat{p}_{x_r}$ over the range where the p_r-experiment is better. Of course, adaptive allocation requires that some of the N observations be used to identify the better experiment. Thus, the desired MSE would not be attained. However, the larger the N, the smaller the proportion expected on the inferior experiment, and the MSE would tend to be nearer to the desired MSE.

Asymptotically, (as N tends to infinity), $\mathrm{MSE}_p(\hat{p}_{x_r})$ can be replaced by $(1-p^r)/Nr^2p^{r-2}$, and the desired limiting MSE can be given as

$$\mathrm{MSE}_p(\hat{p}) \approx \frac{\min\{pq, (1-p^r)/r^2p^{r-2}\}}{N}. \tag{1}$$

Define $H_r(p) = \min\{pq, (1-p^r)/r^2p^{r-2}\}$, and note that the MSE should approximate $H_r(p)/N$ for N large. In Section 8 the normalized MSE will be compared to $H_r(p)$. Before this comparison is made, $H_r(p)$ is shown in Section 7 to arise as the limit of normalized posterior expected loss in a Bayesian setting.

7 Asymptotic Bayes Properties

The focus of this work is frequentist, and both estimators and allocations are evaluated by MSE. However, the major roles of the Fisher information and the maximum likelihood estimator lead to asymptotic Bayes properties of the a_r cut rule (when used with the Bayes estimator), and to a limiting normalized Bayes risk equal to $H_r(p)$.

We set up the present estimation problem in a Bayesian framework by assuming a prior distribution $f(p)$ on p, squared error estimation loss $L(p,\hat{p}) = (p-\hat{p})^2$, and likelihood function to match the previous work, $p^{m\bar{X}_1}(1-p)^{m-m\bar{X}_1}p^{rn\bar{X}_r}(1-p^r)^{n-n\bar{X}_r}$.

Let $\mathbf{E}_N L$ denote the posterior expected loss given $Z_1,\ldots Z_N$ where $Z_i = d_i X_{1i} + (1-d_i)X_{ri}$ as defined in Section 2. Under sufficient regularity conditions on the prior (Kass, Tierney, and Kadane, 1990), the posterior expected loss is approximated by the reciprocal of the Fisher information evaluated at the maximum likelihood estimator:

$$\mathbf{E}_N L = \mathcal{I}_N(\hat{p}_{ml})^{-1}\{1 + O(1/N)\},$$

where $\mathcal{I}_N(p)$ denotes the Fisher information after N observations and $\hat{p}_{ml}$ is the maximum likelihood estimator defined in Section 4. Note that when m X_1's and n X_r's are observed, then $\mathcal{I}_N(p) = m\mathcal{I}_{X_1}(p) + n\mathcal{I}_{X_r}(p)$, where $\mathcal{I}_{X_1}$ and $\mathcal{I}_{X_r}$ are given in Section 3.

Recall from Section 6 that $H_r(p) = \min\{pq, (1-p^r)/r^2p^{r-2}\}$, and note that

$$\begin{aligned} H_r(p) &= p(1-p)I_{\{p\le a_r\}} + (1-p^r)/r^2p^{r-2}I_{\{p>a_r\}} \\ &= \mathcal{I}_{X_1}(p)^{-1}I_{\{p\le a_r\}} + \mathcal{I}_{X_r}(p)^{-1}I_{\{p>a_r\}}, \end{aligned}$$

where $I_{\{.\}}$ is the set indicator function. $H_r(p)$ is a continuous bounded function of p on [0,1], and thus is uniformly continuous. Also, for any p,

$$\begin{aligned} \mathcal{I}_N(p) &= m\mathcal{I}_{X_1}(p) + n\mathcal{I}_{X_r}(p) \\ &\le N[\mathcal{I}_{X_1}(p)I_{\{p\le a_r\}} + \mathcal{I}_{X_r}(p)I_{\{p>a_r\}}]. \end{aligned}$$

This implies that

$$N\mathcal{I}_N(p)^{-1} \ge [\mathcal{I}_{X_1}(p)I_{\{p\le a_r\}} + \mathcal{I}_{X_r}(p)I_{\{p>a_r\}}]^{-1}.$$

But this last term is equal to $\mathcal{I}_{X_1}(p)^{-1}I_{\{p\le a_r\}} + \mathcal{I}_{X_r}(p)^{-1}I_{\{p>a_r\}} = H_r(p)$ since set indicators are used. This proves the following lemma.

Lemma 7.1 $N\mathcal{I}_N(\hat{p}_{ml})^{-1} \ge H_r(\hat{p}_{ml})$. □

Theorem 7.2 *Under sufficient regularity conditions giving the approximation of the posterior variance in terms of Fisher information,*

i. $\liminf_{N\to\infty} N\mathbf{E}_N L \ge H_r(p)$ *a.s.*

ii. $\liminf_{N\to\infty} N\mathbf{E}L \ge \mathbf{E}H_r(p)$.

Proof. For (i), note that by the approximation and by Lemma 7.1,

$$N\mathbf{E}_N L = N\mathcal{I}_N(\hat{p}_{ml})^{-1}\{1+O(1/N)\} \ge H_r(\hat{p}_{ml})\{1+O(1/N)\}\ .$$

But $\hat{p}_{ml}$ converges to p a.s. and H_r is a uniformly continuous function, so thus the lower bound tends a.s. to $H_r(p)$. For (ii), use the inequality above to get $\liminf N\mathbf{E}L \geq \liminf \mathbf{E}H_r(\hat{p}_{ml})$. Then note that H_r is bounded and apply the bounded convergence theorem to deduce the limit to be $\mathbf{E}H_r(p)$. □

This theorem along with sufficient conditions to insure uniform integrability imply that the a_r-cut rule used with the Bayes estimator will have limiting Bayes risk, $\mathbf{E}H_r(p)$, and thus, be asymptotically Bayes.

8 Fixed Sample Size Behavior

Up to this point, the development and evaluation of allocation rules and estimators has been based solely on asymptotic arguments. In this section, the estimators and allocation rules are examined for their behavior based on fixed sample sizes. For moderate sample sizes, some of the estimators do not behave as expected, and there are several adjustments that need to be made when implementing the estimators and allocation rules. Despite these problems, it will be shown that the a_r-cut allocation rule using the MLE estimator does very well.

8.1 Adjusting Cut-points

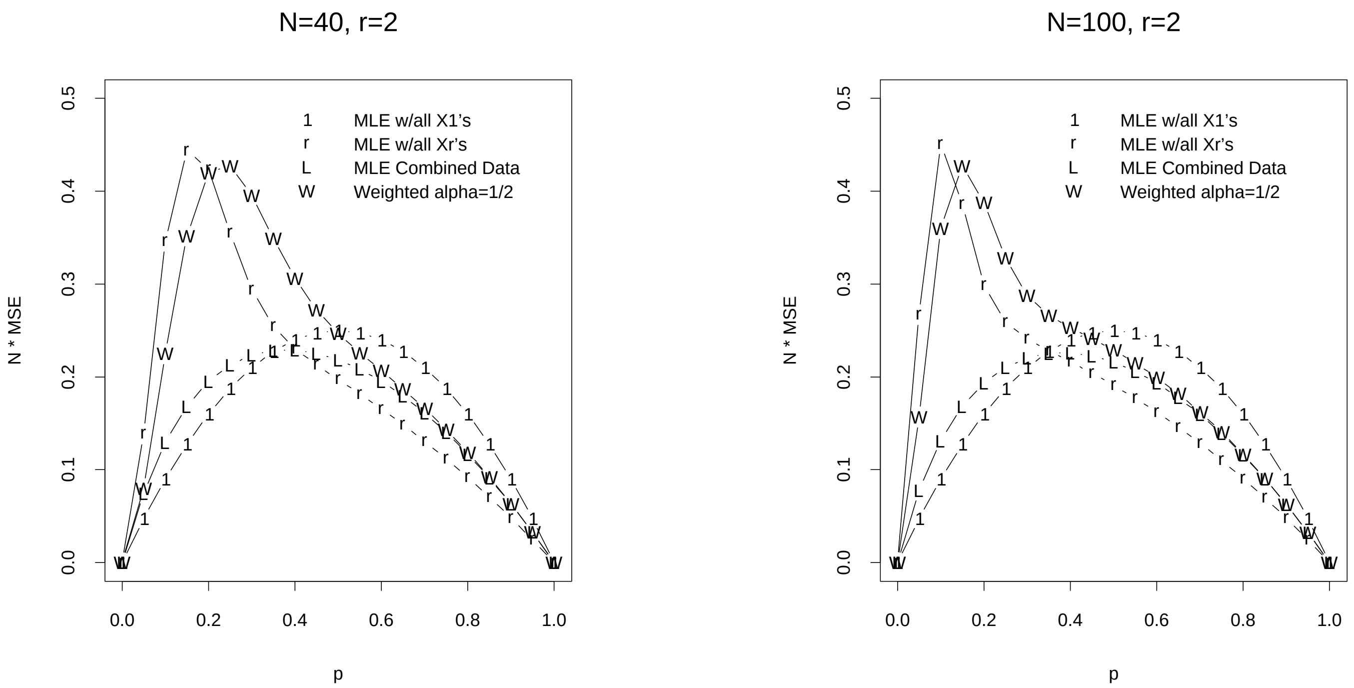
The a_r value was derived from asymptotic considerations. If the exact MSE for the estimator $\hat{p}_{x_r}$ is compared with the exact MSE for $\hat{p}_{x_1}$ using the same number of tests, the cut-point below which the p-experiment does better depends on N as well as r. Determining the cut-point is a straightforward computation, but it does not have a simple closed form. This cut-point differs from a_r, and N may need to be quite large (say ≥ 100) before it is approximately equal to its asymptotic value of a_r. For example, Loyer (1983) showed that for $r = 2$ the value of a_2 is 1/3, but for $N = 20$, the p-experiment is better when $p < 0.445$. For $r = 5$, the value of a_5 is 0.536, but for $N = 20$, the p-experiment is better when $p < 0.729$. Thus, if the total sample size is moderate, then the a_r-cut should be modified to account for the lack of asymptotic fit. This adjustment is more pronounced the larger r is.

8.2 Exact Mean Squared Errors of Estimators Using Fixed Allocations

As noted earlier, the MSE's of most of the estimators under consideration don't have convenient analytic forms. However, given p and N, the MSE's can be calculated for each estimator. In this section, performance of the estimators based on fixed allocations are reviewed. Representative behavior is illustrated in Figure 2 for $N = 40$ and $N = 100$, for $r = 2$.

The labels and relative sample sizes used for the different estimators are the same as in Figure 1. The values in Figure 2 have been scaled by a factor of N, so they can be compared across sample sizes as well as between estimators. It seems clear from both figures that the best estimators for small and large values of p, respectively, are $\hat{p}_{x_1}$ and $\hat{p}_{x_r}$. As noted, as the total sample size increases, the point at which $\hat{p}_{x_r}$ begins to improve on $\hat{p}_{x_1}$ moves towards the value $a_2 = 1/3$.

Note that, even with predetermined allocations of $N/2$ observations from each experiment, the maximum likelihood estimator, $\hat{p}_{ml}$, does very well across the entire range of parameter values regardless of the total sample size. Also, while it may seem odd that the MSE of $\hat{p}_\alpha$ or $\hat{p}_s$ could be larger than the MSE's of



Allocations for L and W are 50/50 on p and p^r.

Figure 2: Exact N∗MSE Values Using Fixed Allocations.

---

either of the two components that comprise the estimators, the plots indicate that this is true. However, there is no contradiction here because the MSE's for $\hat{p}_{x_1}$ and $\hat{p}_{x_r}$ are each based on N observations while the $\hat{p}_{x_1}$ and $\hat{p}_{x_r}$ components of the averaged estimators are each based on fractions of N, and thus are not averages of the X_1 and X_r values shown in the plots.

Comparing Figure 2 to Figure 1, one sees a decrease in the MSE of $\hat{p}_{x_r}$ and the estimators that depend on it as N goes to infinity. Unfortunately, the MSE's of these estimators don't converge very rapidly to the AMSE's. We examined MSE's for $N = 200$ and they were not much closer to the AMSE's than are the MSE's for $N = 100$. On the positive side, the MSE of the maximum likelihood estimator rapidly converges to its AMSE. The MSE's and AMSE's for $\hat{p}_{x_1}$ are, of course, the same.

8.3 MSE's of Adaptive Procedures

The performance of the sequential a_r-cut allocation rule is illustrated in Figure 3 for $N = 100$, $r = 2$, and $a_2 = 1/3$. In this figure, C denotes the exact MSE of the adaptive a_r-cut rule and E represents the normalized lower envelope from Equation (1) of Section 6, namely $H_r(p)$.

Note that the cut-point rule has essentially the same MSE as the lower envelope of the AMSE's, except for the region of approximately $0.3 < p < 0.4$, as shown in the enlargement on the right. In this interval, the MSE of $\hat{p}_{ml}$ actually improves upon the best values gotten using the AMSE's. This behavior leads to questions about the comparison of the exact MSE of the cut-point adaptive rule with other lower envelopes derived for fixed N. For fixed N, the $\text{MSE}_p(\hat{p}_{ml})$ can be computed as a function of m and p, then minimized

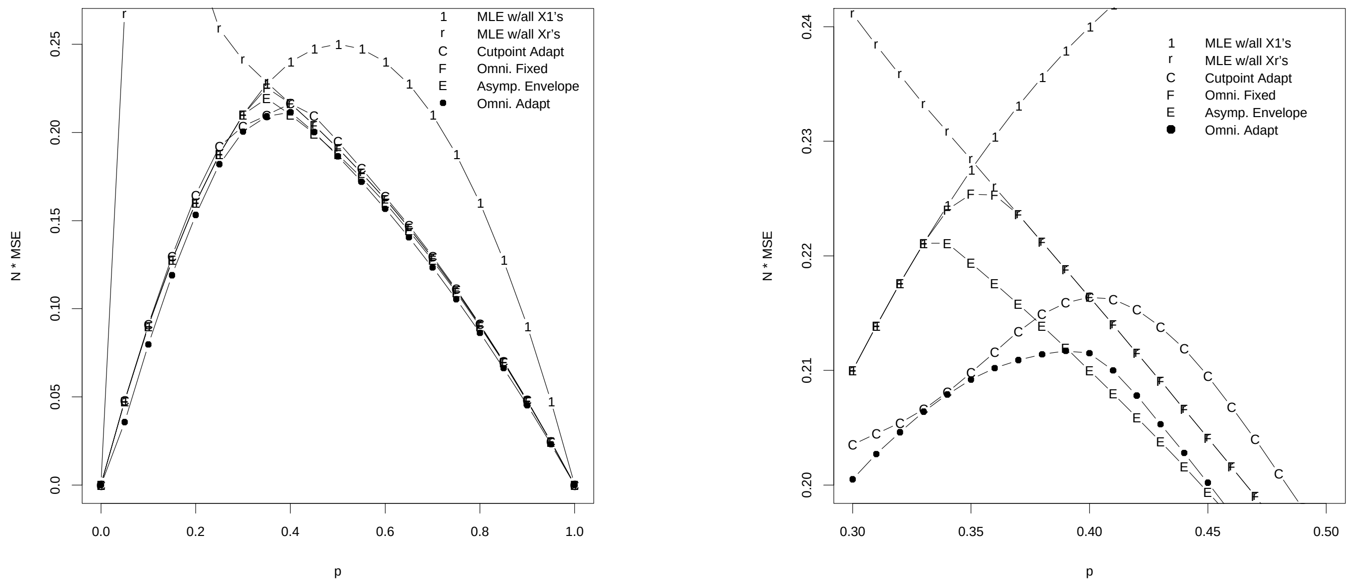


Figure 3: Adaptive Allocation using the MLE, $N = 100$, $r = 2$, cut-point = 1/3.

---

as a function of m. Call the allocation thus obtained "omniscient fixed". This minimizer is dependent on the unknown p, but gives a lower envelope for this MSE. It also is interesting to note that the minimizing m is not necessarily degenerate, i.e., equal to N or 0. For example, at $N = 100$, $r = 2$, and $p = 0.35$, the omniscient fixed allocation to the p-experiment is $m = 57$.

In Figure 3, for $N = 100$ and $r = 2$, the values of normalized MSE for the omniscient fixed allocation are denoted by F. The plot on the left provides normalized MSE's over the entire range of p and the plot on the right is a blow-up of the parameter region in which the MSE of the omniscient fixed rule is better than the asymptotic lower envelope. The region in which the omniscient fixed rule improves on the asymptotic lower bound (E) is quite small and shrinks to zero width as $N \to \infty$. Also note that the cut-point adaptive rule (C) performs better than the omniscient fixed rule in the area of the cut-point. This is a region of negative regret, as discussed for example in Woodroofe (1977) and Martinsek (1983), where the ability to adapt is so beneficial that it overtakes allocation that has advanced knowledge of the parameter but which must be fixed in advance of any experiments.

For fixed N, the true lower bound for the MSE of all rules using the maximum likelihood estimator as terminal estimator can be computed as a function of p and N. This fully sequential rule, which we call "omniscient adaptive", assumes knowledge of the parameter p. Its normalized MSE, denoted by $\bullet$, is shown in Figure 3. Apparently the only way to obtain the omniscient adaptive rule is through dynamic programming calculations.

Finally, computations similar to Figure 3 are shown in Figure 4 for $r = 10$ and $N = 100$. The cut-point used was 0.76, rather than $a_{10} = 0.679$, because this is the point at which the p^{10}-experiment is superior to the p-experiment when $N = 100$. The basic behavior is similar to that seen for $r = 2$, but many of

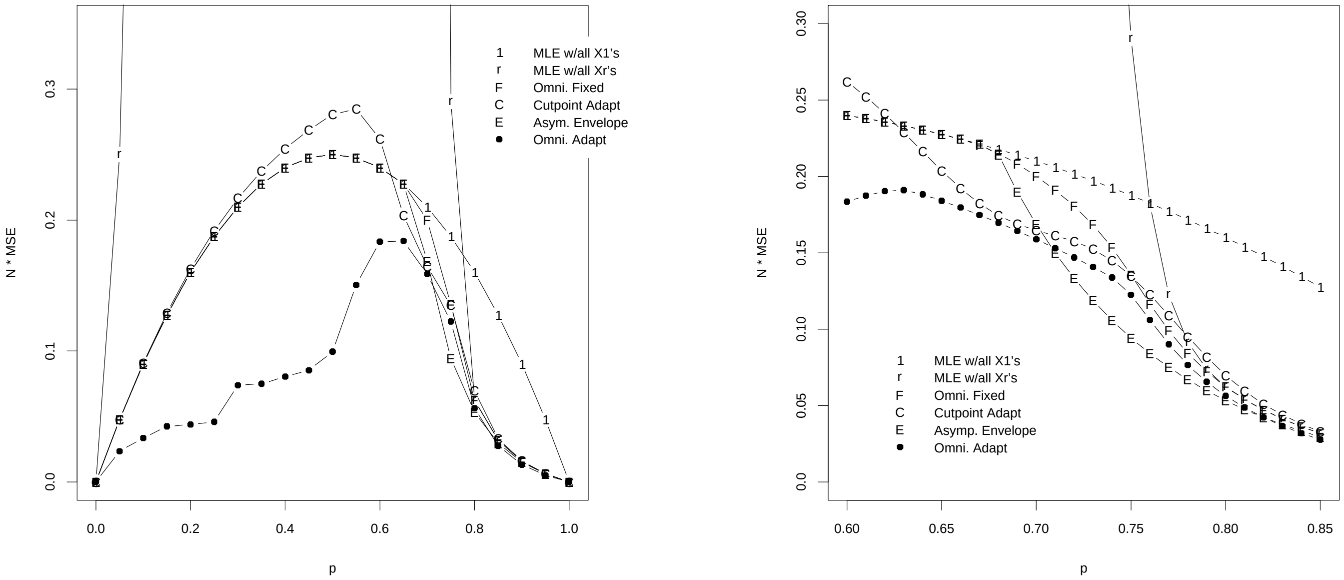


Figure 4: Adaptive Allocation using the MLE, $N = 100$, $r = 10$, cut-point $= 0.76$.

the differences between asymptotic behavior and that for moderate N are more pronounced. For example, for small p, the extent to which the p^r-experiment is worse than the p-experiment is far more extreme. For the same value of N, the region in which the omniscient fixed allocation rule is a mixture of p- and p^r-experiments is far larger than it was for $r = 2$, and the ability of the omniscient adaptive rule to improve upon the omniscient fixed rule is greatly enhanced. The MSE of the omniscient adaptive rule is a more irregular function of p, and, though not evident in Figure 4, it is no longer unimodal.

9 Example: Estimating Prevalence

To provide some insight as to how the adaptive group testing methods presented in Section 6 may be used, we apply them to the Gastwirth and Hammick (1989) reanalysis of a blood screening study of Nusbacher et al. (1986). The authors of the latter work examined whether one could effectively inhibit HIV carriers from donating to a transfusion blood pool. In their study, blood donors who participated in high risk activities were asked to designate their donation to a "research" blood pool rather than to the usual transfusion blood pool. Of the 627 donations to the research blood pool, 11 were found to carry HIV antibodies.

The problem of estimating the prevalence of HIV antibodies motivated Gastwirth and Hammick (1989) to utilize group testing methods on the research blood pool data. However, since screening tests tend to cost considerably less than confirmatory tests, these authors incorporated the sensitivity and specificity of the screening test into their estimators. In this way, they were able to do realistic cost analyses of the group testing approach and the individual testing method. Further, while one of their aims was to provide an accurate prevalence estimator, they also sought a testing method that would preserve the anonymity of the donors. Note that this latter goal is in opposition to the one that motivated Dorfman (1943) to propose group testing methods in the first place. Dorfman's objective was to reduce the cost of detecting

Batch Sizes	No. Samples	E(No. Individuals)	Cost Advantage
1	1020	1020	$11.4\,C_s \leq C_d$
10	110	1100	$0.3\,C_s \leq C_d \leq 11.4\,C_s$
Adapt. 10, 20	63	1250	$C_d \leq 0.3\,C_s$

Figure 5: 3 Methods for Achieving MSE= $1.46 \cdot 10^{-6}$

all positive cases (see also Hwang (1972)).

Our goal here is to compare the accuracy of the adaptive cut-point estimator for prevalence with those obtained from individual testing and fixed group size testing. To simplify comparisons, the strong assumption that the sensitivity and specificity of the screening test are one is made, although adaptive cut-point methods can be optimized for more general settings. It is also assumed that, for each method examined, the observations are sampled from a large population in which the underlying prevalence rate is $11/627 = 0.0175$, the rate observed by Nusbacher et al. (1986).

Gastwirth and Hammick (1989) used batches of size 10, but there seems to be no particular reason to believe that 10 is better than some other group size. We consider what would happen if the adaptive cut-point method is applied for group sizes of $r = 10$ and $s = 20$, assuming, as did Gastwirth and Hammick (1989), that there is no dilution effect. Note that this is equivalent to the $r = 1$ versus $s = 2$ problem considered earlier, in the sense that the $r = 1$ observations are sampled from a Bernoulli population with success rate q^{10}. Taking a sample size of $N = 63$, as in Gastwirth and Hammick (1989), we obtain a MSE of $1.46 \cdot 10^{-6}$. To achieve the same MSE using only batches of size 10, one would need a sample of size $N = 110$; and, using batches of size 1 would require a sample size of $N = 1020$. These results are summarized in Figure 5.

As noted in Gastwirth and Hammick (1989), there are significant advantages to using groups larger than 1, and an adaptive grouped allocation provides yet further advantages. However, there is the consideration that the number of individual samples required increases slightly. Gastwirth and Hammick (1989) addressed this concern via cost analyses, and here cost analyses are carried out using a model consisting of two components. Let C_s be the cost of one screening test and let C_d be the cost of obtaining a single blood donation. Then the cost of achieving of an MSE of $1.46 \cdot 10^{-6}$ using only batches of size 1 is $1020(C_s + C_d)$; the cost using only batches of size 10 is $110\,C_s + 1100\,C_d$; and the expected cost using the adaptive method is $63\,C_s + 1250\,C_d$. The final column of Figure 5 shows the ranges of relative C_s and C_d values for which each method is the most cost-effective.

Typically it happens that $C_d \ll C_s$, and in these situations, adaptive group testing appears to be significantly superior. If total cost is the appropriate consideration, it can be directly incorporated into the adaptive cut-point method. For example, one could compare the asymptotic cost per unit of information using the p-experiment, versus the cost using the p^r-experiment, to decide which to perform. That is, one would determine the cut-point by solving

$$\frac{\mathcal{I}_{X_1}(p)}{C_s + C_d} = \frac{\mathcal{I}_{X_r}(p)}{C_s + rC_d}$$

Finally, one may wonder why $r = 10$ and $s = 20$ were selected for the adaptive version of this example. The only reason for this is that it corresponds well to the main case, p versus p^2, studied in this paper. The present problem was also solved when r and s were taken to be 10 and 100 respectively (which corresponds to the p versus p^{10} case). In this latter case, 63 samples result in a significantly smaller MSE of $3.2 \cdot 10^{-7}$ for $\hat{p}$. The fact that the p^{10} versus p^{100} experiment provides an even greater reduction in MSE for this problem leads one to wonder what the optimal group sizes are for specific problems. As mentioned earlier, Hughs-Oliver and Swallow (1994) consider this question using a two-stage approach. One can extend the present work to include the fully sequential case in which one seeks to estimate not only p, but also the value of r that will optimize a group testing scenario.

10 Extensions

The problem considered in the previous sections was to choose from 2 experiments, the p-experiment and the p^r-experiment, and was motivated by reliability applications. The results can be extended to J experiments and need not necessarily include the p-experiment. Let the available experiments be defined by integers $r(1), r(2), \ldots, r(J)$ such that $1 \leq r(1) < r(2) \ldots < r(J)$, where the i^{th} experiment is a $p^{r(i)}$-experiment. As before, the Fisher information about p contained in a p^r-experiment is

$$\mathcal{I}_{X_r} = r^2 p^{r-2}/(1-p^r) \; .$$

Now, define $G_{r,s}(p) = \mathcal{I}_{X_r}(p) - \mathcal{I}_{X_s}(p)$, and note that $G_{r,s}(p) > 0$ if and only if $\mathcal{I}_{X_r}(p) > \mathcal{I}_{X_s}(p)$, i.e., if and only if the p^r-experiment has more information about p than does the p^s-experiment.

$$G_{r,s}(p) = [(s^2 - r^2)p^s - s^2 p^{s-r} + r^2] \frac{p^{2-r}}{(1-p^r)(1-p^s)} \; ,$$

so for p in (0,1), the sign of $G_{r,s}(p)$ is determined by the first factor. Using derivatives, one can show that $G_{r,s}(p) = 0$ has a unique root, $a_{r,s}$, in (0,1), and, for $r < s$, $\mathcal{I}_{X_r}(p) > \mathcal{I}_{X_s}(p)$ if and only $p < a_{r,s}$.

Cut the unit interval into J parts using cuts

$$0 = a_{r(0),r(1)} < a_{r(1),r(2)} < \ldots < a_{r(J-1),r(J)} < a_{r(J),r(J+1)} = 1 \; ,$$

where for notational convenience we introduce $r(0) = 0$ and $r(J+1) = \infty$. Notice that these cuts are defined using the $r(i)$ in increasing order. Then it can be shown that the $p^{r(i)}$-experiment has maximum information when p is in the i^{th} interval $(a_{r(i-1),r(i)}, a_{r(i),r(i+1)})$, for $i = 1, 2, \ldots, J$. This motivates the very simple adaptive rule which allocates to the $p^{r(i)}$-experiment at stage $k+1$ if the estimator of p based on the data up to and including stage k is in the i^{th} interval. As before, this can also be modified by noting that the $a_{r,s}$ values are based on an asymptotic analysis and can be adjusted for given sample sizes.

There are other useful extensions of the problem examined in this paper. One is to allow the sample size to be a random variable which depends on some stopping criterion. Another is to incorporate unequal costs when sampling from the different experiments. Yet another is to take the Bayesian perspective when sample sizes are fixed. This latter problem requires a significantly different approach than the one taken here.

Appendix: Proof of Theorem 6.1

The lack of consistency of $\hat{p}_\alpha$ occurs when one of the sample sizes does not tend to infinity. That is, if, say, $1 \le n < n_o < \infty$, then $\bar{X}_r{}^{1/r}$ does not converge in probability or a.s., and the weight $1-\alpha$ does not tend to zero. On the other hand, the $\hat{p}_s$ estimator has weights m/N and n/N and if either sample size is bounded as N tends to infinity, the weight tends to zero, implying consistency.

The consistency of $\hat{p}_{ml}$ requires more work. Let $x_1 = \sum_{i=1}^m X_{1_i}$ and $x_r = \sum_{j=1}^n X_{r_j}$. Then the logarithm of the joint likelihood function is equal to

$$h_{m,n}(p) = f_m(p) + g_n(p), \text{ where}$$

$$f_m(p) = x_1 \log(p) + (m - x_1)\log(1-p) \quad \text{and} \quad g_n(p) = x_r \log(p^r) + (n - x_r)\log(1-p^r).$$

The derivative with respect to p is

$$h'_{m,n}(p) = f'_m(p) + g'_n(p), \text{ where}$$

$$f'_m(p) = \tfrac{x_1}{p} - \tfrac{m-x_1}{1-p} \quad \text{and} \quad g'_n(p) = r\tfrac{x_r}{p} - \tfrac{(n-x_r)p^{r-1}}{1-p^r}.$$

The maximum likelihood estimator is the unique root in [0,1] of $h'_{m,n}(p) = 0$, and because f'_m and g'_n are both decreasing in p, it follows that the root of $h'_{m,n}$ is between the roots of f'_m and g'_n. Thus, $\min\{\hat{p}_{x_1}, \hat{p}_{x_r}\} \le \hat{p}_{ml} \le \max\{\hat{p}_{x_1}, \hat{p}_{x_r}\}$. Therefore, if both m and n tend to infinity, then both $\hat{p}_{x_1}$ and $\hat{p}_{x_r}$ tend to p a.s., and thus $\hat{p}_{ml}$ tends to p a.s.

Now, suppose that $n \le n_o < \infty$. Then $\frac{-rn_o}{1-p} \le g'n(p) \le \frac{rn_o}{1-p}$, which in turn bounds $h'_{m,n}$ by

$$A_-(p) = \frac{x_1}{p} - \frac{m - x_1 + rn_o}{q} \ \le\ h'_{m,n}(p) \ \le\ \frac{x_1 + rn_o}{p} - \frac{m - x_1}{q} = A_+(p).$$

But $A_-(p) = 0$ at $p_- = \frac{x_1}{m+rn_o}$ and $A_+(p) = 0$ at $p_+ = \frac{x_1+rn_o}{m+rn_o}$. Then, $p_- \le \hat{p}_{ml} \le p_+$. But p_- and p_+ both tend to p a.s. as m tends to infinity, and thus $\hat{p}_{ml}$ tends to p a.s. as m tends to infinity even though $n \le n_o$. A similar argument holds for $m \le m_o < \infty$. $\square$

References

[1] Chen, C. L. and Swallow, W. H. (1990), "Using Group Testing to Estimate a Proportion, and to Test the Binomial Model," *Biometrics,* **46**, 1035–1046.

[2] Dorfman, R. (1943) "The Detection of Defective Members of Large Populations," *Annals of Mathematical Statistics,***14:** 436–440.

[3] Easterling, R. G. and Prairie, R. R. (1971), "Combining Component and System Information," *Technometrics,* **13**, 271–280.

[4] Gastwirth, J. and Hammick, P. (1989), "Estimation of the Prevalence of a Rare Disease, Preserving the Anonymity of the Subjects by Group Testing: Application to Estimating the Prevalence of AIDS Antibodies in Blood Donors," *Journal of Statistical Planning and Inference,* **22**, 15–27.

[5] Hardwick, J., Page, C. and Stout, Q. F. (1996), "Sequentially Deciding Between Two Experiments for Estimating a Common Success Probability," Technical report, Michigan State University, Department of Statistics and Probability.

[6] Hughs-Oliver, J. and Swallow, W. H. (1994), "A Two-Stage Adaptive Group-Testing Procedure for Estimating Small Proportions," *Journal of the American Statistical Association,* **89**, 982–993.

[7] Hwang, F. K. (1972), "A Method for Detecting all Defective Members in a Population by Group Testing", *Journal of the American Statistical Association,* **67**, 605–608.

[8] Kass, R., Tierney, L., and Kadane, J. (1990), "The Validity of Posterior Expansions Based on Laplace's Method," *Bayesian and Likelihood Methods in Statistics and Econometrics*, S. Geisser, J. S. Hodges, S. J. Press and A. Zellner (eds.), 473–478. Amsterdam: North-Holland.

[9] Lancaster, V. A. and Keller-McNulty, S. (1996), "A Review of Composite Sampling Methods", Technical Report 96-01, Kansas State University, Dept. of Statistics.

[10] Loyer, M. W. (1983), "Bad Probability, Good Statistics, and Group Testing for Binomial Estimation," *The American Statistician,* **37**, 57–59.

[11] Martinsek, A. (1983), "Second Order Approximation to the Risk of a Sequential Procedure," *Annals of Statistics,* **11**, 827–836.

[12] Noble, W. (1990), *First Order Allocation*, Ph. D. Thesis, Michigan State University.

[13] Nusbacher, J., Chiavetta, J. et al. (1986) "Evaluation of a Confidential Method of Excluding Blood Donors Exposed to Human Immunodeficiency Virus", *Transfusion,* **27**, 539–541.

[14] Sobel, M. and Elashoff, R. M. (1975), "Group Testing with a New Goal, Estimation," *Biometrika,* **62**, 181–193.

[15] Tu, X. M., Litvak, E. and Pagano, M. (1995), "On the Informativeness and Accuracy of Pooled Testing in Estimating Prevalence of a Rare Disease: Application to HIV Screening", *Biometrika,* **82**, 287–297.

[16] Woodroofe, M. (1977), "Second Order Approximations for Sequential Point and Interval Estimation," *Annals of Statistics,* **5**, 984–995.