BAYESIAN LOOK-AHEAD SAMPLING ALLOCATIONS FOR SELECTING THE BEST BERNOULLI POPULATION

by

S. S. Gupta and K. J. Miescke
Purdue University University of Illinois at Chicago

Technical Report #95-45C

Department of Statistics
Purdue University

December 1995
Revised February 1996

◇

BAYESIAN LOOK-AHEAD SAMPLING ALLOCATIONS FOR SELECTING THE BEST BERNOULLI POPULATION

by

S. S. Gupta[1] and K. J. Miescke
Purdue University University of Illinois at Chicago

Let $\mathcal{P}_1, \ldots, \mathcal{P}_k$ be $k \geq 2$ independent Bernoulli populations with success probabilities $\theta_1, \ldots, \theta_k$, respectively. Suppose we want to find the population with the largest success probability, using a Bayes selection procedure based on a prior density $\pi(\underline{\theta})$, which is the product of k known Beta densities, and a linear loss function $L(\underline{\theta}, i)$, $\underline{\theta} = (\theta_1, \ldots, \theta_k)$, $i = 1, \ldots, k$. Assume that k independent samples of sizes $n_1, \ldots, n_k$, respectively, have been observed already at a first stage, and that m more observations are allowed to be taken at a future second stage. The problem considered is how to allocate these m observations in suitable manner among the k populations, given the information gathered so far. Several allocation schemes are examined and compared analytically as well as numerically. A simple look-ahead-one-observation-at-a-time allocation rule is shown to have good performance properties.

1. Introduction

Let $\mathcal{P}_1, \ldots, \mathcal{P}_k$ be $k \geq 2$ independent Bernoulli populations with success probabilities $\theta_1, \ldots, \theta_k \in [0,1]$, respectively. Suppose we want to find that population which has the largest θ-value, using a Bayes selection rule which is based on a known prior density $\pi(\underline{\theta})$, $\underline{\theta} = (\theta_1, \ldots, \theta_k) \in [0,1]^k$, and a given loss function $L(\underline{\theta}, i)$ for selecting population $\mathcal{P}_i$, $i \in \{1, 2, \ldots, k\}$ at $\underline{\theta} \in [0,1]^k$. Assume that k independent samples of sizes $n_1, \ldots, n_k$, respectively, have been observed already at a first stage, and that m additional observations are planned to be taken at a second stage. Several sampling designs for allocating these m observations to the k populations will be considered in this paper.

Multi-stage selection procedures have been studied extensively in the past. In their

[1]This research was supported in part by US Army Research Office Grant DAAH04-95-1-0165 at Purdue University.

pioneering monograph, Bechhofer, Kiefer, and Sobel (1968) derived, in the frequentist approach, optimum sequential rules based on a *vector-at-a-time* sampling without elimination of populations. Gupta and Miescke (1984) considered Bayes truncated sequential procedures with elimination of populations. Overviews in this direction can be found in Gupta and Panchapakesan (1979, 1991) and in Miescke (1984). Elimination of populations from further sampling *and* final selection can lead to a conflicting situation: the data collected from a population before it was eliminated may later make it look better again, after further sampling from other populations turn out to be less favorable for them. This suggests to use *adaptive sampling*, i.e. to keep all population in the selection pool, but to allow for temporary exclusions of suitable populations from sampling. Such a *soft elimination* can be incorporated conveniently in the Bayesian approach, since Bayes procedures utilize all the data that have been collected. This approach allows also the use of nonsymmetric priors, which occur in a natural way as updated priors at various intermediate sampling steps. Gupta and Miescke (1993) have derived in this way simultaneous selection and estimation procedures for Bernoulli populations, taking into consideration the cost of sampling. For simplicity of presentation, estimation and cost of sampling will not be considered here. The former would affect selection through the posterior expected loss due to estimation, whereas the latter would involve stopping rules. Modifications in this respect of the allocations considered later on are straightforward.

Looking ahead m allocated but not yet drawn observations, using the expected posterior Bayes risk, given the prior and all the observations collected so far, and then minimizing it across all possible allocations of m observations, does not only provide an optimum allocation of the next m observations. It also allows to assess how much better the final decision can be expected to be after further sampling has been done, following this optimum allocation. Finding the minimum expected posterior risk is intimately connected with finding the single-stage Bayes selection rule for each of the allocations considered. This problem for binomial populations with Beta priors, in a nonsymmetric setting, has been treated in Abughalous and Miescke (1989).

Once the optimum sampling allocation of m observations has been found, the following natural question arises: Why allocate all m observations at once, rather than allocate only a few observations (or just one), learn more through them (it), and then continue with

more appropriate allocations of the remaining observations?

Apparently, the best allocation scheme can be found, in principle, in an m-truncated sequential approach in which one observation at a time is observed. After finding the Bayes terminal selection rule for every possible allocation and realization of m observations, backward optimization could be used to optimize successively every single allocation before. Although the former are not too hard to find, the latter appears too cumbersome to be carried out in practice. Therefore, one reasonable procedure proposed in this paper is to allocate in an optimum way one observation at a time, pretending that it is the last one to be drawn before the final selection, and then to iterate this process until m observations have been drawn. Allocating more than one observation at a time appears to be less appealing, since with each new observation more is learned about the unknown parameters. In some practical situations it may be required to always choose a next observation according to the *state-of-the-art*, i.e. from that population which appears to be the best at present. This policy, which is different from the one proposed here, will also be considered later in making comparisons.

Bayes look ahead techniques, which have been used by Dunnett (1960) for selecting the largest of k normal population means, and by Govindarajulu and Katehakis (1991) in survey sampling, are discussed in various settings in Berger (1985), including also the relevant work by Amster (1963). Work for k normal populations, which is closely related to the present approach, can be found in Gupta and Miescke (1994,95). One particular result found there is that just for $k = 2$, but not for $k > 2$ populations, the outcome of the observations from the first stage are irrelevant for any further optimum allocations. As will be seen below, in the Bernoulli case, first stage observations will always be relevant for the same purpose.

Selecting the population with the largest (overall) sample mean, with ties broken at random, is usually called the *natural selection rule*, since it is the uniformly best permutation invariant selection procedure, in the frequentist sense, for a general class of loss functions, provided that the sample sizes are equal. For unequal sample sizes, however, the natural selection rule loses much of its quality, as it has been shown in Risko (1985) for $k = 2$, and in Abughalous and Miescke (1989) for $k \geq 2$. Similar results for comparisons

of binomial populations can be found in Bratcher and Bland (1975). Nevertheless, for the binomial, normal, and other models, the natural selection rule apparently continues to enjoy rather unquestioned popularity.

Finally, it should be noted that adaptive sampling from k Bernoulli populations has been considered in the frequentist approach, with objectives different from the present, by many authors. One particular approach by Bechhofer and Kulkarni (1982), which has been followed up in several papers, is to save observations in the natural single stage procedure for equal sample sizes without any loss of performance power. The results in this respect, many references, and a thorough overview of the related literature can be found in the recent monograph by Bechhofer, Santner, and Goldsman (1995).

In Section 2, a general outline of the proposed Bayes look ahead sampling approach is presented. An application of these general results to the case of k Bernoulli populations with independent Beta priors and a linear loss is worked out in details in Section 3. Finally, in Section 4, numerical results from computer evaluations and simulations, at suitable parameter settings, are presented which provide the basis for a critical comparison of all sampling allocation schemes considered. It is found that the Bayes look-ahead-one-observation-at-a-time allocation scheme performs similarly to the Bayes allocation of all m observations at once, and better than three others also considered.

2. Bayes Look Ahead Sampling Allocations

In this section, a general outline of the proposed Bayes look ahead sampling approach is presented. After a standard reduction of the data by sufficiency, the model assumptions can be summarized as follows: At $\underline{\theta} = (\theta_1, \ldots, \theta_k) \in [0,1]^k$, the parameter space, let X_i and Y_i be real valued sufficient statistics of the samples from population $\mathcal{P}_i$ at Stage 1 and Stage 2, resp., $i = 1, \ldots, k$, which altogether are assumed to be independent. A priori, the parameters are considered as realizations of a random variable $\underline{\Theta} = (\Theta_1, \ldots, \Theta_k)$ which follows a given prior distribution. Let the loss for selecting $\mathcal{P}_i$, at both stages, be $L(\underline{\theta}, i)$, at $\underline{\theta} \in [0,1]^k$, $i = 1, \ldots, k$. Cost of sampling, which would require the incorporation of a stopping rule, is not included in the loss in order to simplify the presentation of basic ideas. Modification of the allocations discussed below to this more general setting are straightforward.

Having observed $\underline{X} = \underline{x}$ at Stage 1, based on samples of sizes $n_1, \ldots, n_k$, every (it may not be unique) nonrandomized Bayes selection rule $d_1^*(\underline{x})$ satisfies

$$E\{L(\underline{\Theta}, d_1^*(\underline{x}))|\underline{X} = \underline{x}\} = \min_{i=1,\ldots,k} E\{L(\underline{\Theta}, i)|\underline{X} = \underline{x}\}. \tag{1}$$

Likewise, after $\underline{Y} = \underline{y}$ has been observed at Stage 2, every nonrandomized Bayes rule $d_2^*(\underline{x}, \underline{y})$ satisfies

$$E\{L(\underline{\Theta}, d_2^*(\underline{x}, \underline{y}))|\underline{X} = \underline{x}, \underline{Y} = \underline{y}\} = \min_{i=1,\ldots,k} E\{L(\underline{\Theta}, i)|\underline{X} = \underline{x}, \underline{Y} = \underline{y}\}. \tag{2}$$

After the observations have been drawn at Stage 1 or Stage 2, the set of Bayes selections consists of all possible random choices among the respective nonrandomized Bayes selections, which in the following is assumed to be made with equal probabilities.

Many results for Bayes selection rules in symmetric models can be found in the literature. An overview of the earlier literature is provided by Gupta and Panchapakesan (1979). Only recently, however, attention has been given also to nonsymmetric models. The binomial case has been treated in Abughalous and Miescke (1989), the normal case in Gupta and Miescke (1988), and more involved models have been considered in Berger and Deely(1988) and in Fong and Berger (1993). Let us assume now that the selection rules of type d_1^* and d_2^* have been derived already, are ready to be used by the experimenter, and that all that is left to do is to allocate sample sizes in an optimum manner. Then the Bayes allocation of a fixed number of observations can be determined as follows.

Before entering Stage 1, by looking ahead one stage, one has to minimize the expected posterior risk subject to $n_1 + \ldots + n_k = n$, where n is the total number of observations allowed to be taken at Stage 1. Thus, one has to find $n_1, \ldots, n_k$ which yield

$$\min_{\substack{n_1,\ldots,n_k \\ n_1+\ldots+n_k=n}} E(\min_{i=1,\ldots,k} E\{L(\underline{\Theta}, i)|\underline{X}\}). \tag{3}$$

Likewise, before entering Stage 2, at which m observations are to be drawn, one has to find $m_1, \ldots, m_k$ which yield

$$\min_{\substack{m_1,\ldots,m_k \\ m_1+\ldots+m_k=m}} E\{\min_{i=1,\ldots,k} E\{L(\underline{\Theta}, i)|\underline{X} = \underline{x}, \underline{Y}\}|\underline{X} = \underline{x}\}, \tag{4}$$

where the outer expectation is with respect to the conditional distribution of $\underline{Y}$, given $\underline{X} = \underline{x}$. In case of tied allocations, it is assumed that one of them is chosen at random with equal probabilities.

These two minimization problems under side conditions can be seen to be equivalent within the present framework. (3) is of the type (4) if one assumes that some Stage 0, say, with no observations at all has proceeded Stage 1 and $\underline{X}$ is taking over the role of $\underline{Y}$ in (4). On the other hand, (4) is of the type (3) if the information of $\underline{X} = \underline{x}$ is incorporated into an updated prior (cf. Berger (1985), p. 445) and $\underline{Y}$ is taking over the role of $\underline{X}$ in (3). The latter approach will be adopted in Section 3 where it turns out to be the more convenient tool.

As mentioned already in Section 1, allocating all m observations at once for Stage 2 misses the chance of learning more about the unknown parameters at some intermediate sampling, which in turn would provide a better basis for further sampling. Breaking down the allocation of m observations into more than one step can be done in various fashions. Let $\mathcal{R}_t$ denote the optimum allocation of $t \leq m$ observations determined by (4), with m replaced by t there, where ties among several possible allocations are broken at random with equal probabilities. The latter will be seen, later in Section 4, to be not only natural and convenient but relevant if more sampling is planned to be done afterwards. Allocating only one of the t observations determined by $\mathcal{R}_t$, at random and with equal probabilities $1/t$, results in yet another type of allocation, which will be denoted by $\mathcal{R}_{t,1}$, say. Finally, let $\mathcal{B}_1$ denote the optimum allocation of one new observation, knowing all future allocation and selection rules, whatever their choice may be in a particular situation.

These three basic allocations can be put together in various ways to build complete allocation rules for allocating all m observations at Stage 2. For example, $(\mathcal{B}_1, \mathcal{R}_{m-1})$ allocates the last $m-1$ observations through (4) and allocates the first observation by backward optimization. Similarly, $(\mathcal{R}_{m-1}, \mathcal{B}_1)$ allocates the first $m-1$ observations through (4) and then uses $\mathcal{R}_1$ to allocate the last observation, i.e. this allocation scheme is equal to $(\mathcal{R}_{m-1}, \mathcal{R}_1)$. It should be noted that $\mathcal{R}_t$ and $\mathcal{R}_{t,1}$ are stand-alone procedures, i.e. procedures which can be used directly without knowing future actions, whereas $\mathcal{B}_1$ is only meaningful in connection with completely specified future actions.

Using the basic fact that every relevant (conditional) expectation of a minimum is smaller than the minimum of the respective (conditional) expectations, except for trivial cases where equality may hold, a partial ordering of allocation rules can be established with respect to their expected posterior Bayes risks, and thus with respect to their overall Bayes risks. In this respect, one can see that $(\mathcal{B}_1, \mathcal{B}_1, \ldots, \mathcal{B}_1, \mathcal{R}_1)$, with $m-1$ repetitions of $\mathcal{B}_1$, is the best possible allocation of m observations. Since it is hard to implement in practice, as it requires an extensive backward optimization programming, other allocation rules will be considered in the following.

Looking ahead one observation at a time in m consecutive steps, and allocating all m observations at least as good as $(\mathcal{R}_m)$ in terms of the Bayes risk, can be achieved by using $(\mathcal{R}_{m,1}, \mathcal{R}_{m-1,1}, \ldots, \mathcal{R}_{2,1}, \mathcal{R}_1)$, where in fact any type of randomization for breaking ties may be used. However, each application of an allocation $\mathcal{R}_{t,1}$ is numerically quite involved, whereas the allocation rule $(\mathcal{R}_1, \mathcal{R}_1, \ldots, \mathcal{R}_1)$ is much simpler to use and appears to be a reasonable approximation. Whenever in an allocation rule a certain step $\mathcal{R}_{t,1}$ is replaced by a step $\mathcal{R}_1$, one may be tempted to assume that the latter picks one population from those from which $\mathcal{R}_{t,1}$ made its choice, i.e. from those given by $\mathcal{R}_t$. Although this is not always true, it seems to occur quite frequently. To determine, whether $(\mathcal{R}_{m,1}, \mathcal{R}_{m-1,1}, \ldots, \mathcal{R}_{2,1}, \mathcal{R}_1)$, perhaps with a properly determined choice for each single allocation instead of pure randomization, is much better than $(\mathcal{R}_1, \mathcal{R}_1, \ldots, \mathcal{R}_1)$, and thus preferable, requires further research in this direction. It will be shown below for the Bernoulli case that the latter allocation rule is not only simple to use, but also performs about as well as $(\mathcal{R}_m)$ and favorably to three other allocation rules, in terms of the Bayes risk.

The proposed allocation rule $(\mathcal{R}_1, \mathcal{R}_1, \ldots, \mathcal{R}_1)$ allocates at each step one more observation in an optimum manner, pretending that it would be the last one before final selection. Thus for every single observation, it allocates it to that population which appears to be the most promising to improve the pretended final selection decision after it has been drawn. As it has been mentioned already in Section 1, this allocation policy is different from the *state-of-the-art* allocation, which would at each of the m steps find those populations which appear at that moment to be the best, i.e. having the largest posterior means of $\Theta_1, \ldots, \Theta_k$, and then allocate the next observation to one of them at random.

3. Allocations for Bernoulli Populations

The Bayes look ahead sampling allocation approach, which has been introduced in a general framework in the previous section, will now be applied to independent Bernoulli sequences with independent Beta priors under the linear loss $L(\underline{\theta}, i) = \theta_{[k]} - \theta_i,\ i = 1, \ldots, k,\ \underline{\theta} \in [0,1]^k$, where $\theta_{[k]} = max\{\theta_1, \ldots, \theta_k\}$. Justifications for adopting this type of loss in the present setting, and reasons for not using the $0-1$ loss, which is another standard loss for selections, are provided in Abughalous and Miescke (1989).

Given $\underline{\Theta} = \underline{\theta}$, using sufficiency, the observations at Stage 1 can be combined into $\underline{X} = (X_1, \ldots, X_k)$ with $X_i \sim \mathcal{B}(n_i, \theta_i),\ i = 1, \ldots, k$, and the observations at Stage 2 can be combined into $\underline{Y} = (Y_1, \ldots, Y_k)$ with $Y_i \sim \mathcal{B}(m_i, \theta_i),\ i = 1, \ldots, k$, where all these $2k$ binomial random variables are independent. A priori, the k parameters $\underline{\Theta} = (\Theta_1, \ldots, \Theta_k)$ are assumed to be independent Beta random variables with $\Theta_i \sim \mathcal{BE}(\alpha_i, \beta_i)$, where $\alpha_i > 0$, and $\beta_i > 0, i = 1, \ldots, k$, are known.

In what follows, it is assumed that Stage 1 has been completed already, i.e. that $\underline{X} = \underline{x}$ has been observed, and that all further considerations are focused on Stage 2. In this setting, it proves convenient to update the prior with the information gained through $\underline{X} = \underline{x}$, resulting in $\Theta_i \sim \mathcal{BE}(\alpha_i + x_i, \beta_i + n_i - x_i),\ i = 1, \ldots, k$, which in turn are independent. The optimum allocation of all m observations at Stage 2, i.e. $(\mathcal{R}_m)$, can then be found by using criterion (3), with the prior, $n_1, \ldots, n_k$, and $\underline{X}$ replaced by this updated prior, $m_1, \ldots, m_k$, and $\underline{Y}$, respectively. For simplicity of notation, let $a_i = \alpha_i + x_i$ and $b_i = \beta_i + n_i - x_i,\ i = 1, \ldots, k$, in the sequel. Later on, we will return to the original notation to allow for an interpretation of the numerical findings.

To determine now $(\mathcal{R}_m)$, one has to find sample sizes $m_1, \ldots, m_k$, subject to $m_1 + \ldots + m_k = m$, which yield the smallest expected posterior loss

$$\begin{aligned} &\min_{\substack{m_1,\ldots,m_k \\ m_1+\ldots+m_k=m}} E_{\underline{x}}(\min_{i=1,\ldots,k} E_{\underline{x}}\{L(\underline{\Theta}, i)|\underline{Y}\}) \qquad (5) \\ &= E_{\underline{x}}(\Theta_{[k]}) - \max_{\substack{m_1,\ldots,m_k \\ m_1+\ldots+m_k=m}} E_{\underline{x}}(\max_{i=1,\ldots,k} E_{\underline{x}}\{\Theta_i|\underline{Y}\}) \\ &= E_{\underline{x}}(\Theta_{[k]}) - \max_{\substack{m_1,\ldots,m_k \\ m_1+\ldots+m_k=m}} E_{\underline{x}}(\max_{i=1,\ldots,k} \frac{a_i + Y_i}{a_i + b_i + m_i}), \end{aligned}$$

where the subscript at expectations, and below at probabilities, indicates the use of the

updated prior, based on the observations $\underline{X} = \underline{x}$, before entering Stage 2. Thus $E_{\underline{x}}(\Theta_{[k]})$ is the expectation of the maximum of the k independent Beta random variables $\Theta_i \sim \mathcal{BE}(a_i, b_i)$, $i = 1, \ldots, k$, and the last expectation in (5) is with respect to the marginal distribution of $\underline{Y}$ at the end of Stage 1, which in Gupta and Miescke (1993) is shown to be

$$P_{\underline{x}}(\underline{Y} = \underline{y}) = \prod_{i=1}^{k} \binom{m_i}{y_i} \frac{\Gamma(a_i + b_i)\Gamma(a_i + y_i)\Gamma(b_i + m_i - y_i)}{\Gamma(a_i)\Gamma(b_i)\Gamma(a_i + b_i + m_i)} \ , \qquad (6)$$
$$y_i = 0, \ldots, m_i \ , \ i = 1, \ldots, k \ ,$$

i.e. it is the product of k Pólya-Eggenberger distributions, cf. Johnson and Kotz (1969), p. 230.

To summarize the findings about allocation procedure $(\mathcal{R}_m)$, one can state the following theorem.

Theorem 1. The allocation rule $(\mathcal{R}_m)$ chooses, with equal probabilities, one of the allocations $(m_1, \ldots, m_k)$ which satisfy $m_1 + \ldots + m_k = m$ and maximize

$$\sum_{\underline{y}} \max_{i=1,\ldots,k} \{ \frac{a_i + y_i}{a_i + b_i + m_i} \} \ P_{\underline{x}}(\underline{Y} = \underline{y}) \ , \qquad (7)$$

where the distribution of $\underline{Y}$ is given by (6).

For any other allocation rule for m observations, one can see that the expected posterior loss under the linear loss is, similar as for $(\mathcal{R}_m)$ in (5), equal to the difference of $E_{\underline{x}}(\Theta_{[k]})$ and, say, the expected posterior gain of that particular rule. Therefore, it suffices to deal only with the latter in the following, where besides $(\mathcal{R}_m)$, four other specific allocation rules will be considered. The first is $(\mathcal{R}_1, \mathcal{R}_1, \ldots, \mathcal{R}_1)$, where in each of the first $m - 1$ allocations $\mathcal{R}_1$ ties are broken at random with equal probabilities. The second rule allocates one observation at a time according to the, at each incidence, largest posterior expectation of $\Theta_1, \ldots, \Theta_k$, i.e. the *state-of-the-art*, with ties broken in the same fashion as before. The third rule allocates one observation at a time, each purely at random, and the fourth rule is the fixed sample size rule which allocates m/k observations, provided that m is divisible by k, to each of the k populations. In Section 4, numerical comparisons will be made between these five allocation rules, where the look ahead expected posterior gains of $(\mathcal{R}_m)$ and the fixed sample size rule are computed directly, whereas those of the

other three allocation rules are determined through computer simulations. As it will be seen there, the two allocation rules $(\mathcal{R}_m)$ and $(\mathcal{R}_1, \mathcal{R}_1, \ldots, \mathcal{R}_1)$ perform similarly well, but better that the other three rules.

Next we will examine in some detail how the allocation rule $(\mathcal{R}_1, \mathcal{R}_1, \ldots, \mathcal{R}_1)$ operates. Each of its m steps is of the type $\mathcal{R}_1$, based on the respective updated prior. Thus, without loss of generality, let us consider how it allocates the first of the m observations. This is done by choosing with equal probabilities one of those populations $\mathcal{P}_j$, $j \in \{1, \ldots, k\}$, for which the look ahead one observation expected posterior gain g_j, say, is maximized:

$$\begin{aligned} g_j := E_{\underline{x}}\big(\max\{\frac{a_j + Y_j}{a_j + b_j + 1}\ ,\ \max_{i \neq j}\{\mu_i\}\}\big) \\ = \max\{\frac{a_j + 1}{a_j + b_j + 1}\ ,\ \max_{i \neq j}\{\mu_i\}\}\ \mu_j \\ + \max\{\frac{a_j + 0}{a_j + b_j + 1}\ ,\ \max_{i \neq j}\{\mu_i\}\}\ (1 - \mu_j)\ , \end{aligned} \tag{8}$$

where $\mu_t = a_t/(a_t + b_t), t = 1, \ldots, k$ for brevity. As it will be seen below, this type of allocation is done in a simple and specific way, which also reveals an intuitively understandable mechanism.

In the sequel, let $\mu_{[1]} \leq \mu_{[2]} \leq \ldots \leq \mu_{[k]}$ denote the ordered posterior means of the parameters $\Theta_1, \ldots, \Theta_k$ at the end of Stage 1. Let $\mathcal{P}_{(t)}$ be any population associated with $\mu_{[t]}$, $t = 1, \ldots, k$, but in some specific way that all k populations are included. Moreover, let $a_{(t)}$, $b_{(t)}$, $m_{(t)}$, $g_{(t)}$, and $\epsilon_{(t)}$ in turn be associated with $\mathcal{P}_{(t)}$, where $\epsilon_{(t)} = 1/(a_{(t)} + b_{(t)})$ for brevity, $t = 1, \ldots, k$. Then $\mathcal{R}_1$ allocates the next observation,with equal probabilities, to one of those populations which are tied for a maximum expected posterior gain, i.e. for the maximum of the k quantities

$$\begin{aligned} g_{(k)} &= \mu_{(k)} + \max\Big\{\ 0\ ,\ (1 - \mu_{(k)})\mu_{(k-1)} - \frac{\mu_{(k)}(1 - \mu_{(k)})}{1 + \epsilon_{(k)}}\ \Big\} \\ g_{(j)} &= \mu_{(k)} + \max\Big\{\ 0\ ,\ (1 - \mu_{(k)})\ \mu_{(j)} \quad - \frac{\mu_{(j)}(1 - \mu_{(j)})}{1 + \epsilon_{(j)}}\ \Big\}\ , \\ & \qquad\qquad\qquad\qquad\qquad\qquad j = 1, \ldots, k-1\ , \end{aligned} \tag{9}$$

which can be derived from (8) in a straightforward manner. Facilitating this are the facts that $g_{(k)}$, the first summand in (8) turns out to be $(\mu_{(k)} + \epsilon_{(k)})/(1 + \epsilon_{(k)})\mu_{(k)}$, and that for

$g_{(k)}$ with $j < k$, the second summand in (8) turns out to be $\mu_{(k)}(1-\mu_{(j)})$. Of course, all of the gains in (9) have to be at least as large as $\mu_{(k)}$, since observing one more observation can never decrease the Bayes posterior risk achieved already, which in the present situation is equal to $\mu_{(k)}$. It is also interesting to find out under which conditions one more observation from any particular one of the k populations would yield an expected posterior gain larger that $\mu_{(k)}$. Although this is not relevant for the performance considerations of allocation rules later in Section 4, it will be done below of (10) for completeness.

To summarize the findings about allocation procedure $(\mathcal{R}_1, \mathcal{R}_1, \ldots, \mathcal{R}_1)$, one can state the following theorem.

Theorem 2. The first allocation $\mathcal{R}_1$ in $(\mathcal{R}_1, \mathcal{R}_1, \ldots, \mathcal{R}_1)$ is made, with equal probabilities, to one of those populations $\mathcal{P}_{(t)}$ with $g_{(t)} = \max\{g_{(1)}, \ldots, g_{(k)}\}$, $t = 1, \ldots, k$, where $g_{(1)}, \ldots g_{(k)}$ are given by (9). All consecutive allocations of the type $\mathcal{R}_1$ are made analogously, with $\mu_{(t)}$ and $\epsilon_{(t)}$, $t = 1, \ldots, k$, properly updated by all previous allocations and observations.

Finding all populations which are tied for the maximum value of the expected posterior gains given by (9) can be done through paired comparisons. One of these is seen to be different from all others: the comparison of $g_{(k-1)}$ and $g_{(k)}$ is made only through the respective fractions in (9). A similar phenomenon occurs in the normal case, as it is shown in Gupta and Miescke (1995), where the comparison of the *state-of-the-art* population and its *runner-up* is different from all other comparisons.

Another representation of (9), which allows further insight into the behavior of $\mathcal{R}_1$, also with respect to $\alpha_{(t)}, \beta_{(t)}, n_{(t)}$, and $x_{(t)}$ associated with $\mathcal{P}_{(t)}$, is the following.

$$g_{(k)} = \mu_{(k)} + \max\left\{ 0 \, , \, (1-\mu_{(k)})\big(\mu_{(k-1)} - \frac{\mu_{(k)}}{1+\epsilon_{(k)}}\big) \right\} \tag{10}$$

$$g_{(j)} = \mu_{(k)} + \max\left\{ 0 \, , \, \big(\frac{\mu_{(j)}+\epsilon_{(j)}}{1+\epsilon_{(j)}} - \mu_{(k)}\big)\, \mu_{(j)} \right\} , j = 1, \ldots, k-1 \, .$$

From this one can see that $g_{(k)}$ is bigger than $\mu_{(k)}$ if and only if allocating one more observation to $\mathcal{P}_{(k)}$ would lead, in case of a subsequent failure, to an updated posterior expectation $\mu_{(k)}/(1+\epsilon_{(k)})$ for this population which falls below of $\mu_{(k-1)}$. In such case, $g_{(k)}$ is decreasing in $\mu_{(k)}$, increasing in $\mu_{(k-1)}$ and $\epsilon_{(k)}$, and thus decreasing in $n_{(k)}$.

On the other hand, for $j \in \{1, \ldots, k-1\}$, $g_{(j)}$ is bigger than $\mu_{(k)}$ if and only if allocating one more observation to $\mathcal{P}_{(j)}$ with a subsequent success would lead to an updated posterior expectation $(\mu_{(j)} + \epsilon_{(j)})/(1 + \epsilon_{(j)})$ for this population which exceeds $\mu_{(k)}$. In such case, $g_{(j)}$ is decreasing in $\mu_{(k)}$, increasing in $\mu_{(j)}$ and $\epsilon_{(j)}$, and thus decreasing in $n_{(j)}$.

In conclusion of this section, it should be pointed out that while using $(\mathcal{R}_1, \mathcal{R}_1, \ldots, \mathcal{R}_1)$, at some intermediate steps equality may occur for all populations in the associated updated version of (9). Considering such a step solely on its own, this would suggest that it is not worth taking one more observation. However, in the bigger picture, where all m steps are considered acting one after another, it may be well worth continuing the sampling process, since at a later step the situation may turn around. In the present setting, however, where it is assumed that m observations are to be taken, this point is of no further concern.

4. Numerical Results and Comparisons

In this section five allocation rules, each designed for m additional observations at Stage 2, will be compared numerically in terms of their respective expected posterior gains. The smallest possible number of populations for which the particular features of selection procedures can be studied is $k = 3$, which is thus chosen in the following for simplicity. The five allocation rules considered are explained as follows.

RAN	Assign one observation at a time, each purely at random.
EQL	Assign $m/3$ observations to each population $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$.
SOA	Assign one observation at a time, following *state-of-the-art*.
LAH	Assign one observation at a time, using $(\mathcal{R}_1, \mathcal{R}_1, \ldots, \mathcal{R}_1)$.
OPT	Assign m_t observations to $\mathcal{P}_t, t = 1, 2, 3$, using $(\mathcal{R}_m)$.

Several clarifying comments should be made hereby. Below, m is chosen to be $1, 3, 9$, and 15. Thus for $m = 1$, to be specific, procedure **EQL** has been set to take that one observation from population $\mathcal{P}_1$, rather than leaving the respective spaces empty in the tables. As to procedure **SOA**, *state-of-the-art* means the largest posterior expectation of $\Theta_1, \ldots, \Theta_k$ at any present moment. Finally, all procedures but **EQL** are assumed to breaking ties at random, with equal probabilities, whenever they occur. Several other ways of breaking ties at random have been also tried out but proved to be less efficient.

The comparisons of these five allocation rules are made below by means of three examples, each with specifically chosen values for a_i and b_i, $i = 1, 2, 3$. The results are reported in Tables 1, 2, and 3, respectively. Although there are several ways of interpreting the values of $a_i = \alpha_i + x_i$ and $b_i = \beta_i + n_i - x_i$, $i = 1, 2, 3$, one appears to be the easiest to follow intuitively and is thus adopted for convenience. It is assumed from now on that $\alpha_i = \beta_i = 1$, $i = 1, 2, 3$, i.e. that the prior is noninformative (uniform), and thus that $n_i = a_i + b_i - 2$ observations with $x_i = a_i - 1$ successes and $n_i - x_i$ failures have been collected from population $\mathcal{P}_i$ previously at Stage 1, $i = 1, 2, 3$.

The three example share several common features which should be described first. Given the specific choice of $a_1, a_2, a_3, b_1, b_2, b_3$, allocation rules **RAN**, **EQL**, **SOA**, **LAH**, and **OPT** are compared, in terms of their expected posterior gains, for allocations of $m = 1, 3, 9, 15$ observations at Stage 2. The populations $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{P}_3$ are labelled in such a way that always $n_1 < n_2 < n_3$ holds. Within each case of m, eight configurations of $a_1, a_2, a_3, b_1, b_2, b_3$ are considered, where the consecutive seven configurations result from the first by exchanging the values of a_i and b_i, or equivalently the number of successes and failures, for one or more population $\mathcal{P}_i$, $i \in \{1, 2, 3\}$. The latter allow to cover various different situations which will be discussed later on in the examples in more details. Finally, for each particular configuration of $a_1, a_2, a_3, b_1, b_2, b_3$, one possible solution m_1, m_2, m_3 of allocation rule $(\mathcal{R}_m)$, with $m = m_1 + m_2 + m_3$, is reported in the tables. Since it is not always unique, the lexicographically smallest choice has been made throughout.

The numerical results reported in the tables have been calculated on an IBM-type Pentium 66MHz computer with Microsoft QuickBASIC Version 4.5 software. Calculations for allocation rules **EQL** and **OPT** have been performed directly, using routines BICO, FACTLN, and GAMMLN from Sprott (1991). Every single result in the tables for allocation rules **RAN**, **SOA**, and **LAH** is the average over 100,000 computer simulation runs, using the random number generator RND of QuickBASIC, where the seed has been reset with RANDOMIZE(TIMER) at the beginning of each of these 100,000 runs. As to the precision, each expected posterior gain reported in the tables is, due to rounding, accurate only up to $\pm$ .0001. All programs used for this purpose are available from the authors upon request.

EXAMPLE 1. In this example as well as in the other two examples, $\alpha_i = \beta_i = 1$, $i = 1,2,3$ are chosen for convenience of interpretation. This makes the original prior noninformative (uniform) and allows to represent all relevant information up to Stage 2 by the data observed at Stage 1.

To begin with, it is first assumed that at Stage 1 samples with the following sizes and numbers of successes have been observed: $n_1 = 10$ and $x_1 = 4$ from population $\mathcal{P}_1$, $n_2 = 15$ and $x_2 = 6$ from $\mathcal{P}_2$, and $n_3 = 20$ and $x_3 = 8$ from $\mathcal{P}_3$. The resulting parameter configuration $(a_1, a_2, a_3, b_1, b_2, b_3) = (5, 7, 9, 7, 10, 13)$ is shown in Table 1 as the first one for each of the cases $m = 1, 3, 9, 15$. The seven other parameter configurations are obtained by exchanging the number of successes and failures, i.e. the values of a and b, within one or more populations.

Each configuration can be interpreted conveniently through the sample size $n_i = a_i + b_i - 2$ and success rate $\bar{x}_i = x_i/n_i = (a_i - 1)/(a_i + b_i - 2)$ from population $\mathcal{P}_i$, $i = 1, 2, 3$. The parameter values have been chosen in such a way that situations of the type $\bar{x}_1 = \bar{x}_2 = \bar{x}_3$ are included. As one can see, this occurs at the first and last configuration with common success rates .4 and .6, respectively. The six configurations in between represent all other combinations of success rates .4 and .6 with sample sizes 10, 15, and 20.

Comparing now the expected posterior gains of the five allocation rules, one can see from Table 1 that overall, **LAH** and **OPT** are performing similarly well, but better than **RAN**, **EQL**, and **SOA**, and that the latter effect is increasing with m.

EXAMPLE 2. In this example again $\alpha_i = \beta_i = 1$, $i = 1, 2, 3$ are chosen for convenience of interpretation. To begin with, it is assumed that at Stage 1 samples with sizes and number of successes $n_1 = 5$ and $x_1 = 3$ from $\mathcal{P}_1$, $n_2 = 14$ and $x_2 = 9$ from $\mathcal{P}_2$, and $n_3 = 18$ and $x_3 = 11$ from $\mathcal{P}_3$ have been observed. The resulting parameter configuration $(a_1, a_2, a_3, b_1, b_2, b_3) = (4, 10, 12, 3, 6, 8)$ is shown in Table 2 as the first one for each of the cases $m = 1, 3, 9, 15$. The seven other parameter configurations are obtained, as in the previous example, by exchanging the number of successes and failures, i.e. the values of a and b, within one or more populations.

Interpretation of the parameter configurations can be done again conveniently through

the sample sizes $n_i = a_i + b_i - 2$ and success rates $\bar{x}_i = (a_i - 1)/(a_i + b_i - 2)$ from $\mathcal{P}_i$, $i = 1, 2, 3$. The parameter values have been chosen in such a way that situations are included where $\bar{x}_1$, $\bar{x}_2$, and $\bar{x}_3$ are close to each other, but based on different sample sizes. As one can see from Table 2, this occurs at the first parameter configuration with $\bar{x}_1 = .6$, $\bar{x}_2 = .6429$, $\bar{x}_3 = .6111$ and $n_1 = 5$, $n_2 = 14$, $n_3 = 18$, and at the last one with $\bar{x}_1 = .4$, $\bar{x}_2 = .3571$, $\bar{x}_3 = .3889$ and $n_1 = 5$, $n_2 = 14$, $n_3 = 18$. The six configurations in between represent all other combinations of success rates .6 or .4, .6429 or .3571, and .6111 or .3889 with sample sizes 5, 14, and 18, respectively.

Comparing this time the expected posterior gains of the five allocation rules, one can see from Table 2 that overall again, **LAH** and **OPT** are performing similarly well, but better than **RAN**, **EQL**, and **SOA**, and that the latter effect is increasing with m.

EXAMPLE 3. This example is similar to the previous example and thus it will be discussed only briefly. $\alpha_i = \beta_i = 1$, $i = 1, 2, 3$ are again chosen for convenience of interpretation. Now we assume first that at Stage 1 samples with $n_1 = 5$ and $x_1 = 3$ from $\mathcal{P}_1$, $n_2 = 12$ and $x_2 = 7$ from $\mathcal{P}_2$, and $n_3 = 17$ and $x_3 = 9$ from $\mathcal{P}_3$ have been observed with respective success rates $\bar{x}_1 = .6$, $\bar{x}_2 = .5833$, and $\bar{x}_1 = .5294$. The resulting parameter configuration $(a_1, a_2, a_3, b_1, b_2, b_3) = (4, 8, 10, 3, 6, 9)$ is shown in Table 3 as the first one for each of the cases $m = 1, 3, 9, 15$. The seven other parameter configurations are obtained in the same manner as described in Example 2.

The main difference to Example 2 is that this time in the first (last) parameter configuration, the second largest of the success rates $\bar{x}_1$, $\bar{x}_2$, and $\bar{x}_3$ is associated with the second largest of the sample sizes n_1, n_2, and n_3, rather than with the largest (smallest) sample size as it was done in Example 2. The present setting allows to study two situations where $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$ are close together and where extreme success rates and sample sizes are associated with each other. For completeness, yet another example could be considered where in the first parameter configuration, the second largest success rate is associated with the smallest sample size. But this is omitted for brevity.

Comparing now the performances of the five allocation rules shown in Table 3 is leading to the same conclusions as before in Example 1 and in Example 2.

Finally, after the results of the three examples have been discussed, some conclusions can be drawn. It has been shown that allocation rule **LAH** compares favorably with the rules **RAN**, **EQL**, and **SOA**, and that overall its performance is close to the one of **OPT**. That **LAH** is not always as good as **OPT** indicates clearly that it cannot be a version, i.e. with any type of tie breaking, of the rule $(\mathcal{R}_{m,1}, \mathcal{R}_{m-1,1}, \ldots, \mathcal{R}_{2,1}, \mathcal{R}_1)$ considered at the end of Section 2, since the latter is at least as good as **OPT** $= (\mathcal{R}_m)$. However, in comparison to these two allocation rules, the proposed rule **LAH** $= (\mathcal{R}_1, \mathcal{R}_1, \ldots, \mathcal{R}_1)$ is

more flexible in terms of the predetermined number of observations m at Stage 2,

Table 1	**Expected Posterior Gains in Example 1**							
a_1 a_2 a_3 b_1 b_2 b_3	*RAN*	*EQL*	*SOA*	*LAH*	*OPT*	m_1	m_2	m_3
m = 1								
5 7 9 7 10 13	.4282	.4325	.4325	.4325	.4325	1	0	0
5 7 13 7 10 9	.5909	.5909	.5908	.5909	.5909	0	1	0
5 10 9 7 7 13	.5882	.5882	.5882	.5882	.5882	1	0	0
5 10 13 7 7 9	.5981	.5909	.6004	.6028	.6028	0	1	0
7 7 9 5 10 13	.5834	.5833	.5834	.5833	.5833	1	0	0
7 7 13 5 10 9	.5982	.6052	.5982	.6052	.6052	1	0	0
7 10 9 5 7 13	.5972	.6041	.5996	.6041	.6041	1	0	0
7 10 13 5 7 9	.6028	.6052	.6004	.6052	.6052	1	0	0
m = 3								
5 7 9 7 10 13	.4384	.4394	.4410	.4432	.4435	2	1	0
5 7 13 7 10 9	.5910	.5909	.5910	.5908	.5909	0	3	0
5 10 9 7 7 13	.5882	.5882	.5882	.5882	.5882	2	1	0
5 10 13 7 7 9	.6034	.6028	.6068	.6077	.6081	0	2	1
7 7 9 5 10 13	.5835	.5833	.5831	.5834	.5833	1	2	0
7 7 13 5 10 9	.6047	.6052	.6063	.6112	.6119	3	0	0
7 10 9 5 7 13	.6045	.6041	.6084	.6105	.6109	3	0	0
7 10 13 5 7 9	.6119	.6101	.6111	.6184	.6172	2	1	0
m = 9								
5 7 9 7 10 13	.4540	.4543	.4554	.4582	.4575	5	3	1
5 7 13 7 10 9	.5910	.5910	.5910	.5912	.5918	9	0	0
5 10 9 7 7 13	.5885	.5885	.5882	.5886	.5892	6	3	0
5 10 13 7 7 9	.6125	.6122	.6159	.6159	.6176	0	9	0
7 7 9 5 10 13	.5836	.5837	.5847	.5839	.5845	8	1	0
7 7 13 5 10 9	.6153	.6157	.6173	.6212	.6241	9	0	0
7 10 9 5 7 13	.6161	.6161	.6196	.6207	.6226	9	0	0
7 10 13 5 7 9	.6274	.6272	.6273	.6322	.6305	6	3	0
m = 15								
5 7 9 7 10 13	.4631	.4638	.4628	.4663	.4660	9	5	1
5 7 13 7 10 9	.5915	.5917	.5912	.5919	.5930	12	0	3
5 10 9 7 7 13	.5892	.5896	.5895	.5904	.5909	9	6	0
5 10 13 7 7 9	.6178	.6181	.6210	.6208	.6231	0	12	3
7 7 9 5 10 13	.5847	.5851	.5864	.5859	.5864	11	4	0
7 7 13 5 10 9	.6216	.6222	.6232	.6262	.6292	13	0	2
7 10 9 5 7 13	.6224	.6231	.6260	.6258	.6291	11	4	0
7 10 13 5 7 9	.6364	.6369	.6354	.6399	.6390	9	6	0

Table 2	Expected Posterior Gains in Example 2							
a_1 a_2 a_3 b_1 b_2 b_3	*RAN*	*EQL*	*SOA*	*LAH*	*OPT*	m_1	m_2	m_3
m = 1								
4 10 12 3 6 8	.6270	.6250	.6294	.6294	.6294	0	1	0
4 10 8 3 6 12	.6250	.6250	.6248	.6250	.6250	0	0	1
4 6 12 3 10 8	.6048	.6143	.6000	.6143	.6143	1	0	0
4 6 8 3 10 12	.5714	.5714	.5713	.5714	.5714	0	0	1
3 10 12 4 6 8	.6271	.6250	.6294	.6294	.6294	0	1	0
3 10 8 4 6 12	.6251	.6250	.6252	.6250	.6250	0	0	1
3 6 12 4 10 8	.6000	.6000	.6000	.6000	.6000	0	0	1
3 6 8 4 10 12	.4333	.4429	.4428	.4429	.4429	1	0	0
m = 3								
4 10 12 3 6 8	.6359	.6359	.6345	.6418	.6429	3	0	0
4 10 8 3 6 12	.6322	.6329	.6311	.6377	.6429	3	0	0
4 6 12 3 10 8	.6135	.6143	.6115	.6238	.6238	2	0	1
4 6 8 3 10 12	.5715	.5714	.5713	.5716	.5714	0	1	2
3 10 12 4 6 8	.6307	.6319	.6345	.6338	.6355	0	3	0
3 10 8 4 6 12	.6250	.6250	.6249	.6251	.6250	0	1	2
3 6 12 4 10 8	.6000	.6000	.6000	.6001	.6000	1	0	2
3 6 8 4 10 12	.4437	.4468	.4453	.4539	.4540	2	1	0
m = 9								
4 10 12 3 6 8	.6531	.6551	.6478	.6586	.6591	6	3	0
4 10 8 3 6 12	.6453	.6466	.6419	.6512	.6554	8	1	0
4 6 12 3 10 8	.6281	.6295	.6240	.6369	.6410	9	0	0
4 6 8 3 10 12	.5744	.5742	.5772	.5748	.5780	9	0	0
3 10 12 4 6 8	.6391	.6400	.6434	.6423	.6440	0	7	2
3 10 8 4 6 12	.6260	.6263	.6251	.6270	.6290	8	1	0
3 6 12 4 10 8	.6019	.6019	.6000	.6034	.6066	9	0	0
3 6 8 4 10 12	.4624	.4642	.4624	.4691	.4701	7	2	0
m = 15								
4 10 12 3 6 8	.6636	.6646	.6559	.6680	.6676	8	6	1
4 10 8 3 6 12	.6527	.6534	.6481	.6577	.6612	13	2	0
4 6 12 3 10 8	.6356	.6371	.6296	.6421	.6458	15	0	0
4 6 8 3 10 12	.5779	.5781	.5801	.5788	.5812	13	1	1
3 10 12 4 6 8	.6458	.6462	.6483	.6480	.6494	0	10	5
3 10 8 4 6 12	.6281	.6284	.6252	.6295	.6316	11	4	0
3 6 12 4 10 8	.6046	.6049	.6004	.6065	.6091	12	0	3
3 6 8 4 10 12	.4730	.4741	.4703	.4775	.4780	10	4	1

Table 3	**Expected Posterior Gains in Example 3**							
a_1 a_2 a_3 b_1 b_2 b_3	*RAN*	*EQL*	*SOA*	*LAH*	*OPT*	m_1	m_2	m_3
m = 1								
4 8 10 3 6 9	.5870	.6020	.5949	.6020	.6020	1	0	0
4 8 9 3 6 10	.5871	.6020	.5949	.6020	.6020	1	0	0
4 6 10 3 8 9	.5750	.5827	.5827	.5827	.5827	1	0	0
4 6 9 3 8 10	.5715	.5714	.5714	.5714	.5714	0	0	1
3 8 10 4 6 9	.5713	.5714	.5714	.5714	.5714	0	0	1
3 8 9 4 6 10	.5715	.5714	.5714	.5714	.5714	0	0	1
3 6 10 4 8 9	.5264	.5263	.5263	.5263	.5263	0	0	1
3 6 9 4 8 10	.4775	.4850	.4737	.4850	.4850	1	0	0
m = 3								
4 8 10 3 6 9	.6000	.6037	.6071	.6113	.6122	3	0	0
4 8 9 3 6 10	.5991	.6020	.6048	.6107	.6122	3	0	0
4 6 10 3 8 9	.5814	.5827	.5892	.5924	.5940	3	0	0
4 6 9 3 8 10	.5738	.5714	.5776	.5759	.5802	3	0	0
3 8 10 4 6 9	.5745	.5752	.5767	.5764	.5770	0	3	0
3 8 9 4 6 10	.5717	.5714	.5717	.5726	.5748	3	0	0
3 6 10 4 8 9	.5286	.5263	.5263	.5310	.5351	3	0	0
3 6 9 4 8 10	.4861	.4871	.4793	.4964	.4962	3	0	0
m = 9								
4 8 10 3 6 9	.6185	.6203	.6222	.6260	.6271	7	2	0
4 8 9 3 6 10	.6158	.6173	.6195	.6233	.6254	9	0	0
4 6 10 3 8 9	.5951	.5963	.6001	.6030	.6068	9	0	0
4 6 9 3 8 10	.5829	.5838	.5871	.5875	.5908	9	0	0
3 8 10 4 6 9	.5852	.5856	.5861	.5878	.5876	5	4	0
3 8 9 4 6 10	.5782	.5787	.5763	.5806	.5815	5	4	0
3 6 10 4 8 9	.5379	.5386	.5302	.5425	.5457	9	0	0
3 6 9 4 8 10	.5037	.5050	.4946	.5110	.5096	7	2	0
m = 15								
4 8 10 3 6 9	.6296	.6303	.6300	.6348	.6358	9	6	0
4 8 9 3 6 10	.6251	.6261	.6265	.6305	.6327	9	6	0
4 6 10 3 8 9	.6033	.6043	.6054	.6092	.6119	14	0	1
4 6 9 3 8 10	.5904	.5911	.5936	.5941	.5955	12	0	3
3 8 10 4 6 9	.5932	.5935	.5914	.5955	.5949	7	8	0
3 8 9 4 6 10	.5841	.5849	.5799	.5868	.5875	9	6	0
3 6 10 4 8 9	.5453	.5462	.5349	.5492	.5500	13	0	2
3 6 9 4 8 10	.5146	.5153	.5035	.5191	.5179	9	4	2

since its individual steps are of type $\mathcal{R}_1$ which do not depend on m. Thus, if in an ongoing experiment the number m of observations at Stage 2 has to be changed at any time and by any reason, the allocation rule **LAH** can be adapted very easily to such a change by just reducing or extending the number of repetitions of the steps of type $\mathcal{R}_1$ accordingly.

References

Abughalous, M. M. and Miescke, K.J. (1989). On selecting the largest success probability under unequal sample sizes. *J. Statist. Plann. Inf.* **21** 53–68.

Amster, S. J. (1963). A modified Bayesian stopping rule. *Ann. Math. Statist.* **34** 1404–1413.

Bechhofer, R. E., Kiefer, J., and Sobel, M. (1968). Sequential Identification and Ranking Procedures. The University of Chicago Press, Chicago.

Bechhofer, R. E. and Kulkarni, R. V. (1982). Closed adaptive sequential procedures for selecting the best of $k \geq 2$ Bernoulli populations. *Statistical Decision Theory and Related Topics - III, Vol. 1, S. S. Gupta and J. O. Berger, eds.* Academic Press, New York 61–108.

Bechhofer, R. E., Santner, T. J., and Goldsman, D. M. (1995). Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons. J. Wiley, New York.

Berger, J. O. (1985). Statistical Decision Theory and Bayesian Analysis. 2nd ed. Springer Verlag, New York.

Berger, J. O. and Deely, J. (1988). A Bayesian approach to ranking and selection of related means with alternatives to AOV methodology. *J. Amer. Statist. Assoc.* **83** 364–373.

Bratcher, T. L. and Bland, R. P. (1975). On comparing binomial probabilities from a Bayesian viewpoint. *Commun. Statist.* 4 975–985.

Dunnett, C. W. (1960). On selecting the largest of k normal population means. — With Discussion. *J. Royal Statist. Soc.* **B-22** 1–40.

Fong, D. K. H. and Berger, J. O. (1993). Ranking, estimation, and hypothesis testing in unbalanced two-way additive models - A Bayesian approach. *Statist. Decisions* **11**

1–24.

Govindarajulu, Z. and Katehakis, M. N. (1991). Dynamic allocation in survey sampling. *Amer. J. Math. Manag. Sci.* **11** 199–221.

Gupta, S. S. and Miescke, K. J. (1984). Sequential selection procedures - A decision theoretic approach. *Ann. Statist.* **12** 336–350.

Gupta, S. S. and Miescke, K. J. (1988). On the problem of finding the largest normal mean under heteroscedasticity. *Statistical Decision Theory and Related Topics - IV , Vol. 2, S. S. Gupta and J. O. Berger, eds.* Springer Verlag, New York 37–49.

Gupta, S. S. and Miescke, K. J. (1993). On combining selection and estimation in the search of the largest binomial parameter. *J. Statist. Plann. Inf.* **36** 129–140.

Gupta, S. S. and Miescke, K. J. (1994). Bayesian look ahead one-stage sampling allocations for selecting the largest normal mean. *Statist. Papers* **35** 169–177.

Gupta, S. S. and Miescke, K. J. (1995). Bayesian look ahead one-stage sampling allocations for selection of the best population. *J. Statist. Plann. Inf.* 20 pp. in ms, to appear.

Gupta, S. S. and Panchapakesan, S. (1979). Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations. J. Wiley, New York.

Gupta, S. S. and Panchapakesan, S. (1991). Sequential ranking and selection procedures. *Handbook of Sequential Analysis, B.K. Ghosh and P.K. Sen, eds.* Marcel Dekker, New York 363–380

Johnson, N. L. and Kotz, S. (1969). Discrete Distributions. Houghton and Mifflin, Boston.

Miescke, K. J. (1984). Recent results on multi-stage selection procedures. *Proceedings of the Seventh Conference on Probability Theory, M. Iosifescu, ed.* Editura Academiei Republicii Socialiste România, Bucharest 259–268.

Risko, K. R. (1985). Selecting the better binomial population with unequal sample sizes. *Commun. Statist.-Theor. Meth.* **14** 123–158.

Sprott, J. C. (1991). Numerical Recipes: Routines and Examples in BASIC. Cambridge University Press, Cambridge.