BAYES SAMPLING DESIGNS FOR SELECTION PROCEDURES*

by

Klaus J. Miescke
University of Illinois at Chicago
Chicago, IL 60607-7045, U.S.A.

Technical Report #97-12C

Department of Statistics
Purdue University
West Lafayette, IN USA

August, 1997

*This research was supported in part by US Army Research Office Grant DAAH04-95-1-0165 at Purdue University under the direction of Professor Shanti S. Gupta.

BAYES SAMPLING DESIGNS FOR SELECTION PROCEDURES *

Klaus J. Miescke
Department of MSCS (M/C 249)
University of Illinois at Chicago
851 South Morgan Street
Chicago, IL 60607-7045, U.S.A.

ABSTRACT. From k independent populations $P_1,...,P_k$, which belong to a one parameter exponential family $\{F_\theta\}$, $\theta \in \Omega \subseteq \Re$, random samples of sizes $m_1,...,m_k$, respectively, are to be drawn. After the observations have been drawn, a selection procedure will be used to determine which of these k populations has the largest value of θ. Given a loss for selections at each parameter configuration, given n past observations, and given a prior for the k parameters, a Bayes selection procedure can be found and its Bayes risk can be determined, where both depend on $m_1,...,m_k$. Let the sample sizes be restricted by $m_1 + ... + m_k = m$, where m is fixed. The problem of how to find the optimum (minimum Bayes risk) sample design subject to this constraint is considered, as well as m-truncated sequential sampling allocations. Results for *normal* and *binomial* families, under the *"0-1" loss* and the *linear loss*, are presented and discussed. An introduction to Bayes selection procedures is included.

AMS(1980) Subject Classification. Primary 62F07; Secondary 62F15.
Key Words and Phrases. Selection procedures; Bayesian designs; optimum sampling allocations; binomial populations; normal populations.

* This research was supported in part by US Army Research Office Grant DAAH04-95-1-0165 at Purdue University under the direction of Professor Shanti S. Gupta.

1. INTRODUCTION

Let $P_1,...,P_k$ be k populations which belong to a given one parameter exponential family $\{F_\theta\}, \theta \in \Omega \subseteq \Re$, where P_i is associated with a certain parameter $\theta_i \in \Omega$, $i = 1,...,k$, but the values of $\theta_1,...,\theta_k$ are unknown. Suppose one wants to select, based on independent random samples of respective sizes $m_1,...,m_k$, that population which has the largest θ-value $\theta_{[k]}$, say. In the decision theoretic approach, let $L(\theta,i)$ be a given loss for selecting population P_i at any $\theta = (\theta_1,...,\theta_k) \in \Omega^k$, $i = 1,...,k$. Two special types of loss functions, which will be primarily considered later on, are the so-called *"0-1" loss* $L(\theta,i) = 0$ if $\theta_i = \theta_{[k]}$, and 1 otherwise, and the *linear loss* $L(\theta,i) = \theta_{[k]} - \theta_i$, $i = 1,...,k$. The performance of a selection rule d, i.e. a measurable function from the sampling space into the set $\{1,...,k\}$, can then be measured by its expected loss, i.e. its frequentist risk, at each parameter configuration $\theta \in \Omega^k$. Extending this framework to the Bayes approach, it is assumed that the parameters $\Theta = (\Theta_1,...,\Theta_k)$, say, are a priori random and follow a known prior density $\pi(\theta)$, $\theta \in \Omega^k$. The purpose of this paper is to provide an introduction to Bayes selection procedures, a brief review of multi-stage selection procedures, and a thorough discussion of recent results on Bayes look-ahead sampling designs for selection procedures.

The history of selection procedures dates back to the 1950s. The first such procedures considered have been based on k independent samples of equal sample sizes. An overview and thorough discussions of these early procedures, and of the development of numerous branches of the theory of selection thereafter until 1979, is provided in the by now classical monograph by Gupta and Panchapakesan (1979). In celebration of "40 Years of Statistical Selection

Theory", and its pioneers Robert E. Bechhofer, Shanti S. Gupta, and Milton Sobel, a conference has been held on September 5-10, 1993, at Bad Doberan in Germany, and its proceedings have been included in a special journal issue, edited by Miescke and Rasch (1996).

To reduce the effort or cost of sampling, without losing too much power in the decisions, selection procedures which incorporate combinations of various types of sampling, stopping, and selection components have been proposed and studied in the literature over the past three decades. In their fundamental monograph, Bechhofer, Kiefer, and Sobel (1968) have derived for exponential families, in the frequentist approach, optimum sequential selection rules. These are based on vector-at-a-time sampling, i.e. sampling of the same number of observations from each population at a time or stage, the natural terminal selection decision, and an optimum stopping rule. Elimination of certain populations from further sampling, which may emerge as apparently inferior populations during the sampling process, is not allowed there as an option at intermediate stages of the sampling process. At this point it should be mentioned that in case that an elimination of populations from further sampling would indeed be allowed, the option exists of extending this elimination to the pool of populations available for selection at the end, or not. An overview of sequential ranking and selection procedures is provided by Gupta and Panchapakesan (1991).

A simple selection procedure, based on vector-at-a-time sampling, which incorporates elimination from sampling and selection, is a two-stage selection procedure of the following type. After t_1 observations have been drawn from each population at Stage 1, a suitable set of populations is eliminated from further sampling. Then t_2 observations are drawn from each population that has not been screened out, and a final selection is then made from the latter, using all of the data observed from them. Here an option exists regarding the number of

populations retained for Stage 2: it can be chosen to be random or fixed pre-determined. The former case has been studied in Gupta and Miescke (1984a) in the Bayes approach. Both cases, and their extensions to multi-stage selection procedures have been treated, also in the Bayes approach, by Gupta and Miescke (1984b), using backward optimization. An overview of this class of procedures, which select the best population efficiently in terms of risk and sampling costs, is provided by Miescke (1984). The results derived in this respect depend heavily on the assumption that at each sampling stage an equal number of observations, which may, however, vary from stage to stage, is drawn from each population still in the running. This assumption allows to utilize permutation symmetry in the posterior risks in connection with permutation invariant priors, which simplifies the analysis of such procedures.

Whenever a population is eliminated from the pool of populations retained for a final selection, a conflict of the following type may arise. The data collected from such a population prior to its elimination could make it look better again, relatively to the other populations, after further sampling from the latter have turned out to be not so favorable for those. The Bayes approach clearly calls for utilizing the information from all observations that have been drawn, since more observations cannot decrease the Bayes risk. Using this approach, it may in fact occur that such an eliminated population emerges, in terms of the posterior risk, as the population that appears to be the best. Elimination of populations from final selections is thus unreasonable from a Bayesian point of view. Elimination, or just temporary elimination, of populations from sampling at some stages of the sampling process can be incorporated in a natural way into the Bayes approach. The advantage of inherent permutation symmetry, however, is lost here. In conclusion, allocation of a possibly unequal number of observations, where some of them may be actually zero, to the k populations at various stages seems to be

more appropriate. Temporarily drawing no new observation from certain populations, but retaining all k populations in the pool for the final selection decision, may be called a soft elimination. Such a soft elimination invites to use also priors which are not permutation symmetric, since updated priors of this type occur anyway in a natural manner at the various stages, due to soft elimination.

In many statistical experiments, the sampling process extends over a substantial length of time. One of the advantages of the Bayesian approach in a statistical analysis is that it allows to perform conclusions at intermediate time points. Especially, such conclusions can be made toward modifications of the future sampling and decision process. Such types of adaptive sampling or sampling allocation schemes will be the main topic of discussions later on, when Bayes look-ahead selection procedures are considered. Dealing with information from samples of possibly unequal sizes from the k populations may occur quite naturally. A test person may not always come on schedule, or drop out of the study, a test object may break under stress, a budget cut may force to abandon some test runs, or random time spans are under study, which are subject to some form of censoring.

The main type of problem considered in this paper is how to allocate observations to the k populations in a stepwise manner, where the goal is to select the best population at the end of the sampling process. More precisely, assume that k independent samples of respective sizes $n_1,\ldots,n_k$ have been observed already at a first stage from populations $P_1,\ldots,P_k$, which may be the combined outcomes of several previous stages, and that m additional observations are allowed to be taken at a future second stage. One interesting problem that is considered later on is how to allocate $m_1,\ldots,m_k$ observations, subject to $m_1+\ldots+m_k = m$, in an optimum way among the k populations, given all the information, prior and first stage observations, gathered

so far. Looking ahead with the expected posterior Bayes risk, given the information presently at hand, and then minimizing it, does not only provide an optimum allocation of observations in the future. 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. In marketing research such as direct marketing, medical research such as clinical trials (Whitehead, 1991), and social research such as survey sampling (Govindarajulu, Katehakis, 1991), very often interim analyses are performed at certain stages to decide if sampling should be continued, and if so, how to allocate new observations. Such Bayes designs have been studied in the binomial case, under various loss functions, by Gupta and Miescke (1993) for the more general problem of simultaneous selection and estimation, including cost of sampling. For the sake of simplicity of presentation, simultaneous estimation with selection and cost of sampling will only be considered briefly in the following. The former would require to use more involved loss functions, and the latter the incorporation of stopping rules. Modifications of the allocations considered later on to such extended features are straightforward, but technically more involved.

Allocating m new observations at a second stage, using the expected posterior risk, can only be done after the terminal selection rule is known. The latter is the Bayes selection which is based on all observations drawn in the complete sampling process. Thus, the first step toward a Bayes design for the second stage is to determine the optimum (Bayes) single-stage selection rules for various sample sizes $m_1,\ldots,m_k$. This has been done, under both the *"0-1" loss* and under the *linear loss*, for the binomial case in Abughalous and Miescke (1989), including extensions to a larger class of loss functions, and for the normal case in Gupta and Miescke (1988).

After the Bayes terminal selection decisions are known, one can proceed as described above, looking ahead for all possible sampling allocations, which are in the present setting restricted by $m_1+...+m_k = m$, compare the associated expected posterior risks, find its minimum value, and then implement a design associated with it. At this point one may wonder why all m observations are allocated at once, rather than allocating only a few (or just one), learning more through them (it), and then initiating a new allocation optimization process for the remaining allocations. As will be shown later, such a breakdown of allocation of m observation, if done properly, cannot increase the Bayes risk and may in fact be better than allocating all m observations at once. The best possible allocation scheme is to allocate one observation at a time, in m consecutive steps, which are altogether determined by backward optimization, starting at the end with the Bayes terminal selection for every possible allocation $m_1,...,m_k$ with $m_1+...+m_k = m$, and then optimizing successively every single allocation before. However, this appears to be only feasible for discrete distributions, and the only Bayes sequential design of this type that has been treated up to now is for the binomial case (Miescke, Park, 1997a). Alternative allocation schemes, which appear to perform close to the best based on backward optimization, are studied and discussed for the normal case in Gupta and Miescke (1994, 1996a), and for the binomial case in Gupta and Miescke (1996b) and Miescke and Park (1997a). Another type of adaptive sampling and selection for Bernoulli populations, which is in the frequentist approach, can be found in Bechhofer and Kulkarni (1982).

One reasonable procedure is to allocate in an optimum way one observation at a time, pretending that it is the last one to be drawn before final selection, and then to iterate this process until all m observations have been taken. This will be considered later on in this paper. Other procedures, which may allocate more than one observation at a time, will also be

considered. However, they appear to be less appealing since with each new observation more can be learnt about the unknown parameters, which in turn can improve the basis for further decisions. Look ahead procedures, which have been utilized previously by Govindarajulu and Katehakis (1991) in survey sampling, and which are described and discussed in various other settings in Berger (1985), will be discussed thoroughly later in this paper.

Selecting in terms of the largest sample mean is called in the literature the natural selection rule. It is the uniformly best permutation invariant selection procedure, in the frequentist sense, for a general class of loss functions, as long as the sample sizes are equal. However, for unequal sample sizes, the natural selection rule appears to be less powerful, although it still remains intuitively appealing. In view of this fact, optimum sample size allocations for the natural selection rule have been considered in the frequentist approach by Bechhofer (1969), Dudewicz and Dalal (1975), Bechhofer, Hayter, and Tamhane (1991), and Bechhofer, Santner, and Goldsman (1995). Bayes selection rules under unequal sample sizes can have complicated forms which may not be represented in closed form, as it has been shown in Abughalous and Miescke (1989) and Gupta and Miescke (1988). Bayes rules for more involved normal models have been studied by Berger and Deely (1988) and Fong and Berger (1993). Earlier ideas of and results on sampling allocations for Bayes rules under normality and the *linear loss* are due to Dunnett (1960).

An introduction to Bayes selection procedure is provided in Section 2. Here, as well as in the remaining sections, special emphasis is given to two specific models under the *"0-1" loss* and the *linear loss*. The first is the normal case with independent normal priors for the k parameters, which is called the *normal-normal* model, and the second is the binomial case with independent beta priors, which is called the *binomial-beta* model. As a first step toward Bayes

look-ahead sequential sampling designs, Bayes one- and two-stage sampling designs are studied in Section 3. Finally, Bayes look-ahead sequential sampling designs are treated in Section 4.

2. BAYES SELECTION PROCEDURES

Let $P_1,...,P_k$ belong to a one parameter exponential family $\{F_\theta\}$, $\theta \in \Omega \subseteq \Re$, where P_i is associated with a certain parameter $\theta_i \in \Omega$, $i = 1,...,k$, but where the values of $\theta_1,...,\theta_k$ are unknown. Let the goal be to find that population which has the largest parameter value. Special emphasis will be given to the normal family $\{N(\theta,\sigma^2)\}$, $\theta \in \Omega = \Re$, with $\sigma^2 > 0$ known, and to the binomial family $\{B(n,\theta)\}$, $\theta \in \Omega = [0,1]$. Let $X_1,...,X_k$ denote sufficient statistics from independent random samples of sizes $n_1,..., n_k$ from $P_1,...,P_k$, respectively. Since Bayes selection procedures are the topic, only non-randomized decision rules need to be considered in the following. These can be represented as measurable functions $d(\mathbf{x})$ with values in $\{1,...,k\}$, where $\mathbf{x} = (x_1,...,x_k)$ are the observed values of $\mathbf{X} = (X_1,...,X_k)$. In the decision theoretic approach, let $L(\theta,i)$ be the loss for selecting population P_i at $\theta = (\theta_1,...,\theta_k)$, with $L(\theta,i) \le L(\theta,j)$ for $\theta_i \ge \theta_j$, $i,j = 1,...,k$. Later on, emphasis will be given to two special loss functions, the *"0-1" loss* and the *linear loss*, which are defined by

$$\begin{aligned} L(\theta,i) &= 1 - I_{\{\theta_{[k]}\}}(\theta_i), && \textit{"0-1" loss,} \\ L(\theta,i) &= \theta_{[k]} - \theta_i, && \textit{linear loss,} \end{aligned} \tag{1}$$

where $\theta_{[k]} = max\{\theta_1,...,\theta_k\}$, $i = 1,...,k$.

Finally, the Bayesian component is added to the problem. Here the parameters $\Theta = (\Theta_1,...,\Theta_k)$ are assumed to be a priori independent random variables which follow a prior distribution with a known density $\pi(\theta) = \pi_1(\theta_1)\times...\times\pi_k(\theta_k)$. For the *normal* family, *normal* priors $\Theta_i \sim N(\mu_i, \nu_i^{-1})$ with $\mu_i \in \Re, \nu_i > 0$ will be assumed, and for the *binomial* family, *beta* priors $\Theta_i \sim \text{Beta}(\alpha_i,\beta_i)$, with $\alpha_i,\beta_i > 0$, $i = 1,...,k$. The frequentist risk of a selection rule d at $\Theta = \theta$ is given by $R(\theta,d) = E_\theta(L(\theta,d(\mathbf{X}))$, and its Bayes risk by $r(\pi,d) = E^\pi(R(\Theta,d))$. The latter is minimized by every Bayes rule, and this minimum $r(\pi)$, say, is called the Bayes risk of the problem. The Bayes risk of a rule d can be represented in two ways as follows.

$$\begin{aligned} r(\pi,d) &= E(L(\Theta,d(\mathbf{X}))) \\ &= E^\pi(E\{L(\Theta,d(\mathbf{X}))|\Theta\}) \\ &= E^m(E\{L(\Theta,d(\mathbf{X}))|\mathbf{X}\}), \end{aligned} \tag{2}$$

where m denotes the marginal density or discrete probability function of $\mathbf{X}$. The standard way of determining a Bayes rule d^B, say, is to minimize, at every $\mathbf{X} = \mathbf{x}$, the posterior expected loss, i.e. the posterior Bayes risk $E\{L(\Theta,d(\mathbf{x}))|\mathbf{X} = \mathbf{x}\}$. Depending on the type of loss function, one arrives at the following criteria.

$$\begin{aligned} E\{L(\Theta,d^B(\mathbf{x}))|\mathbf{X} = \mathbf{x}\} &= \min_{i=1,...,k} E\{L(\Theta,i)|\mathbf{X} = \mathbf{x}\} && \text{in general,} \\ &= 1 - \max_{i=1,...,k} P\{\Theta_i = \Theta_{[k]}|\mathbf{X} = \mathbf{x}\} && \text{for } \textit{"0-1" loss,} \\ &= E\{\Theta_{[k]}|\mathbf{X} = \mathbf{x}\} - \max_{i=1,...,k} E\{\Theta_i|\mathbf{X} = \mathbf{x}\} && \text{for } \textit{linear loss.} \end{aligned} \tag{3}$$

Apparently, under the *linear loss*, $E\{\Theta_{[k]}|\mathbf{X} = \mathbf{x}\}$ does not need to be considered for finding a Bayes rule. However, it is relevant for the evaluation of the Bayes risk $r(\pi)$, and thus it will be relevant for the Bayes designs to be considered in the subsequent sections. For the two special loss functions, the quantities to be minimized or maximized in (3) can be represented by

$$E\{L(\Theta,i)|\mathbf{X}=\mathbf{x}\} = \int_{\Re^k} L(\theta,i)\,\pi(\theta|\mathbf{x})\,d\theta\,, \qquad \text{in general,} \tag{4}$$

$$P\{\Theta_i=\Theta_{[k]}|\mathbf{X}=\mathbf{x}\} = \int_{\Re}\prod_{j\neq i}\left(\int_{-\infty}^{\theta_i}\pi_j(\theta_j|x_j)\,d\theta_j\right)\pi_i(\theta_i|x_i)\,d\theta_i\,, \quad \text{for } \textit{"0 - 1" loss,}$$

$$E\{\Theta_i|\mathbf{X}=\mathbf{x}\} = \int_{\Re}\theta_i\,\pi_i(\theta_i|x_i)\,d\theta_i\,, \qquad \text{for } \textit{linear loss,}$$

where $\pi(\theta|\mathbf{x})$ is the posterior density of Θ, given $\mathbf{X}=\mathbf{x}$, and $\pi_r(\theta_r|x_r)$ is the posterior marginal density of Θ_r, given $\mathbf{X}=\mathbf{x}$, $r=1,\ldots,k$. In the second and third equation of (4), the fact that $\pi(\theta|\mathbf{x}) = \pi_1(\theta_1|x_1)\times\ldots\times\pi_k(\theta_k|x_k)$ is utilized. This means that a posteriori, $\Theta_1,\ldots,\Theta_k$ are not only independent, but that the posterior distribution of Θ_r, given $\mathbf{X}=\mathbf{x}$, depends on $\mathbf{x}$ only through x_r, $r=1,\ldots,k$.

In the remainder of this section, Bayes rules for the *normal-normal* and the *binomial-beta* model will be studied under the *"0-1" loss* and the *linear loss*. Because of the inherent independence, only the distributions associated with each individual i-th of the k populations, $i\in\{1,\ldots,k\}$, have to be specified. For the *normal-normal* model one has the following.

$$\textit{Normal - Normal:}\quad X_i|\Theta=\theta \;\sim\; N\left(\theta_i, p_i^{-1}\right), \qquad \Theta_i \;\sim\; N\left(\mu_i, \nu_i^{-1}\right), \tag{5}$$

$$\Theta_i|\mathbf{X}=\mathbf{x} \;\sim\; N\left(\frac{p_i x_i+\nu_i\mu_i}{p_i+\nu_i}, (p_i+\nu_i)^{-1}\right), \; X_i \sim N\left(\mu_i, p_i^{-1}+\nu_i^{-1}\right),$$

where X_i is the sample mean of the i-th population, and $p_i = n_i\sigma^{-2}$ is its precision. The Bayes selection rules under *"0-1" loss* and *linear loss* can now be seen, in view of (3), (4), and (5), to be $d^B(\mathbf{x}) = i_0$, if for $i=1,\ldots,k$, the following respective quantity is maximized at $i=i_0$.

$$P\{\Theta_i=\Theta_{[k]}|\mathbf{X}=\mathbf{x}\} = \int_{\Re}\prod_{j\neq i}\Phi_j(\theta|x_j)\,d\Phi_i(\theta|x_i)\,, \qquad \text{for } \textit{"0 - 1" loss,} \tag{6}$$

$$E\{\Theta_i|\mathbf{X}=\mathbf{x}\} = \frac{p_i x_i+\nu_i\mu_i}{p_i+\nu_i}\,, \qquad \text{for } \textit{linear loss,}$$

where $\Phi_r(\cdot|x_r)$ is the c.d.f. of the conditional distribution of Θ_r, given $\mathbf{X} = \mathbf{x}$, $r = 1,\ldots,k$, which is represented in (5).

At this point it should be mentioned that there is a natural selection rule d^N, say, which selects in terms of the largest of the sample means $X_1,\ldots,X_k$. In terms of the frequentist risk, and for a large class of loss functions, including the *"0-1" loss* and the *linear loss*, it is the uniformly best permutation invariant selection procedure, if the sample sizes $n_1,\ldots,n_k$ are all equal. More generally, an analogous fact holds for monotone likelihood ratio families. The history of its proofs, one of which is in the Bayesian approach utilizing permutation invariant priors, can be found in Gupta and Miescke (1984b). For the present situation, where $n_1,\ldots,n_k$ may not be equal, no optimum properties of d^N under the *"0-1" loss* are known, except admissibility, which has been proved only recently in Miescke and Park (1997b). Gupta and Miescke (1988) have shown that d^N is minimax under the *"0-1" loss* if and only if $n_1 = \ldots = n_k$. Here the minimax value of the problem is $1 - 1/k$, which can be proved with a suitable sequence of independent normal priors.

An undesirable property of d^N under the *"0-1" loss* was first discovered (Lam and Chiu, 1976, Tong and Wetzell, 1979) by noting that the frequentist risk of d^N is not always increasing in each of the sample sizes $n_1,\ldots,n_k$. On the other hand, under the *linear loss*, d^N is a proper Bayes rule and, because of its uniqueness, also admissible (Berger, 1985). This can be readily seen (Gupta and Miescke, 1988) from (6), by letting $\mu_r = \mu$ and $\nu_r = c\,p_r$, $r = 1,\ldots,k$, for some fixed real μ and a positive c. Properties of Bayes rules for other priors are also discussed there, as well as in Berger and Deely (1988) and Fong and Berger (1993).

Turning now to the *binomial-beta* model, the situation regarding the distributions presents itself as follows:

$$\textit{Binomial - Beta:} \quad \begin{aligned} X_i|\Theta=\theta \;&\sim\; B(n_i,\theta_i), \qquad & \Theta_i \;&\sim\; \text{Beta}(\alpha_i,\beta_i), \\ \Theta_i|\mathbf{X}=\mathbf{x} \;&\sim\; \text{Beta}(\alpha_i+x_i,\beta_i+n_i-x_i), \quad & X_i \;&\sim\; \text{PE}(n_i,\alpha_i,\beta_i,1), \end{aligned} \tag{7}$$

where $\alpha_i,\beta_i>0$, $i=1,\ldots,k$. Here the unconditional marginal distribution of X_i, $i=1,\ldots,k$, is a Pólya-Eggenberger type distribution (Johnson and Kotz, 1969), sometimes called beta-binomial distribution, which in the present situation turns out to be

$$P\{X_i=x_i\} = \binom{n_i}{x_i} \frac{\Gamma(\alpha_i+\beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} \frac{\Gamma(\alpha_i+x_i)\Gamma(\beta_i+n_i-x_i)}{\Gamma(\alpha_i+\beta_i+n_i)}, \quad x_i=0,1,\ldots,n_i. \tag{8}$$

The Bayes selection rules under *"0-1" loss* and *linear loss* can be seen, in view of (3), (4), and (7), to be $d^B(\mathbf{x})=i_0$, if for $i=1,\ldots,k$, the following is maximized at $i=i_0$.

$$\begin{aligned} P\{\Theta_i=\Theta_{[k]}|\mathbf{X}=\mathbf{x}\} &= \int_{[0,1]} \prod_{j\neq i} F_j(\theta|x_j)\, dF_i(\theta|x_i), \qquad \text{for } \textit{"0-1" loss}, \\ E\{\Theta_i|\mathbf{X}=\mathbf{x}\} &= \frac{\alpha_i+x_i}{\alpha_i+\beta_i+n_i}, \qquad \text{for } \textit{linear loss}, \end{aligned} \tag{9}$$

where $F_r(\cdot|x_r)$ is the c.d.f. of the conditional distribution of Θ_r, given $\mathbf{X}=\mathbf{x}$, $r=1,\ldots,k$, which is represented in (7).

At this point, again, the natural selection rule d^N, which selects in terms of the largest of the sample means $X_1/n_1,\ldots,X_k/n_k$, has to be discussed. Since the binomial family is also an exponential family, d^N is also here uniformly best invariant selection rule if and only if the sample sizes are equal. Likewise, Abughalous and Miescke (1989) have shown that under the *"0-1" loss*, minimaxity of d^N holds if and only if $n_1=\ldots=n_k$. As in the normal case, the minimax value of the problem is $1-1/k$, which can be proved with a suitable sequence of

independent Beta priors. Other results regarding d^N turn out to be different from their counterparts in the normal case. For example, d^N is not a proper Bayes rule here. Properties of Bayes rules, for various priors, under the *"0-1" loss* and the *linear loss* have been studied in the same paper. Questions regarding the optimality of d^N for unequal sample sizes $n_1,\ldots,n_k$, in the present setting, have been addressed in Bratcher and Bland (1975), Risko (1985), and in Abughalous and Miescke (1989).

There are two other types of Bayes selection procedures, which are based on different goals or philosophies. These will be described briefly at the end of this section. The first is within the subset selection approach, which is due to Gupta (1956, 1965). Here the goal is to select a non-empty subset $s \subseteq \{1,\ldots,k\}$, of preferably small size, which contains the best population. This goal can be represented in various ways by means of the loss function. The first paper within this framework is due to Deely and Gupta (1968), which deals with the linear loss function $L(\theta,s) = \sum_{i\in s} \alpha_{s,i}(\theta_{[k]} - \theta_i)$. The loss function $L(\theta,s) = \sum_{i\in s}(a - b\, I_{\{\theta_{[k]}\}}(\theta_i))$ has been used by Bratcher and Bhalla (1974) and Gupta and Hsu (1977), the loss function $L(\theta,s) = c|s| + \theta_{[k]} - max_{i\in s}\theta_i$ has been used by Goel and Rubin (1977), and the additive loss function $L(\theta,s) = \sum_{i\in s} l_i(\theta)$, with emphasis on the special case $L(\theta,s) = \sum_{i\in s}(\theta_{[k]} - \theta_i - \varepsilon)$, has been used by Miescke (1979). Another, non-additive, loss function has been used in Chernoff and Yahav (1977). Further references in this regard can be found in Gupta and Panchapakesan (1979).

The other type of selection procedures combines selection of the best population with the estimation of the parameter of the selected population. The decision rules are now of the form $\delta(\mathbf{x}) = (d(\mathbf{x}), e_{d(\mathbf{x})}(\mathbf{x}))$, where $d(\mathbf{x}) \in \{1,\ldots,k\}$ is the selection rule, and $e_i(\mathbf{x})$, $i = 1,\ldots,k$, is a

collection of estimates for θ_i, $i = 1,\ldots,k$, which are available after selection. As it is shown in Cohen and Sackrowitz (1988) and Gupta and Miescke (1990), the decision theoretic treatment of the combined selection and estimation problem consists of two steps of optimization. First, the possible estimates are determined, which turn out to be the usual Bayes estimates of the related problem of estimation without considering selection. Then, after knowing the available estimates, the optimum selection is made. Detailed results for the normal case, under the additive loss function $L(\theta,\delta) = A(\theta,d) + B(\theta_d, e_d)$, and various special cases, are presented in Gupta and Miescke (1990). Here $A(\theta,d)$ is the loss due to selecting population P_d, and $B(\theta_d, e_d)$ is the loss of estimating θ_d with e_d, $d = 1,\ldots,k$. Similar work for the binomial case has been done by Gupta and Miescke (1993). In both papers, overviews of work in this direction and further references can be found.

3. Bayes One- and Two-Stage Sampling Designs

Starting with the situation described in the previous section up to (3), let us now consider a fixed total sample size allocation problem. Suppose that in the planning stage of the experiment, a total of $n_1+\ldots+n_k = n$ observations are planned to be drawn from the respective populations. Since after every observed $\mathbf{X} = \mathbf{x}$, the posterior Bayes risk will be equal to $min_{i=1,\ldots,k} E\{L(\Theta,i)|\mathbf{X} = \mathbf{x}\}$, the optimum allocation of $n_1,\ldots,n_k$, i.e. the Bayes design, is given by the following criterion:

$$\begin{aligned} &\min_{n_1+\ldots+n_k=n} E\Big(\min_{i=1,\ldots,k} E\{L(\Theta,i)|\mathbf{X}\}\Big), && \text{in general,} \\ &\max_{n_1+\ldots+n_k=n} E\Big(\max_{i=1,\ldots,k} P\{\Theta_i = \Theta_{[k]}|\mathbf{X}\}\Big), && \text{for } \textit{"0-1" loss,} \\ &\max_{n_1+\ldots+n_k=n} E\Big(\max_{i=1,\ldots,k} E\{\Theta_i|\mathbf{X}\}\Big), && \text{for } \textit{linear loss,} \end{aligned} \tag{10}$$

where the outer expectation is with respect to the unconditional marginal distribution of **X**. For the *normal-normal* model and the *binomial-beta* model, its representation is given in (5) and in (7) and (8), respectively. The inner conditional expectations and probabilities in (10) are represented in general form in (4), and in their special forms for the *normal-normal* model and the *binomial-beta* model in (6) and (9), respectively.

In the first part of this section, Bayes designs for the two special models will be studied under the *"0-1" loss* and the *linear loss*. For the *normal-normal* model, the Bayes designs consist of those sample sizes $n_1,\ldots,n_k$ which achieve the following maximum.

$$\max_{n_1+\ldots+n_k=n} E\left(\max_{i=1,\ldots,k} \int_{\Re} \prod_{j\neq i} \Phi\left(\frac{\delta_i\theta + \mu_i(X_i) - \mu_j(X_j)}{\delta_j} \right) \varphi(\theta)d\theta \right), \quad \text{for } \textit{"0-1" loss}, \qquad (11)$$

$$\max_{n_1+\ldots+n_k=n} E\left(\max_{i=1,\ldots,k} \frac{p_iX_i + \nu_i\mu_i}{p_i + \nu_i} \right), \quad \text{for } \textit{linear loss},$$

where in the first criterion, Φ and φ denote the c.d.f. and density, respectively, of N(0,1), and where $\delta_r = (p_r + \nu_r)^{-1/2}$, $\mu_r(X_r) = (p_rX_r + \nu_r\mu_r)/(p_r + \nu_r)$, and $p_r = n_r\sigma^{-2}$, $r = 1,\ldots k$, for brevity. The outer expectations are with respect to the marginally independent random variable $X_r \sim N(\mu_r, p_r^{-1} + \nu_r^{-1})$, $r = 1,\ldots k$.

The special case of $k = 2$ populations has been completely analyzed in Gupta and Miescke (1994). For both loss functions it has been shown there, using different techniques, that the Bayes design is determined by minimizing $|p_1 + \nu_1 - p_2 - \nu_2|$, the absolute difference between the two posterior precisions. It is zero if and only if the joint posterior distribution of Θ_1 and Θ_2 is decreasing in transposition, a situation in which the Bayes selection rules are of simple forms (Gupta and Miescke, 1988). The results for $k = 2$ mentioned above carry over to $k \geq 3$ populations to some extend. This has been shown in Gupta and Miescke (1996a) for the *linear*

loss, mainly for the special case of n = 1. The latter plays an important role in optimum sequential allocations, which will be discussed in more details in the next section.

For the *binomial-beta* model, the Bayes designs consists of those sample sizes $n_1,\ldots,n_k$ which achieve the following maximum.

$$\max_{n_1+\ldots+n_k=n} E\left(\max_{i=1,\ldots,k} \int_{\Re} \prod_{j\neq i} H_{j,X_j}(\theta)\, h_{i,X_i}(\theta)\, d\theta \right), \quad \text{for } \textit{"0-1" loss}, \tag{12}$$

$$\max_{n_1+\ldots+n_k=n} E\left(\max_{i=1,\ldots,k} \frac{\alpha_i + X_i}{\alpha_i+\beta_i+n_i} \right), \quad \text{for } \textit{linear loss},$$

where in the first criterion, H_{r,x_r} and h_{r,x_r} denote the c.d.f. and the density, respectively, of $\text{Beta}(\alpha_r + x_r, \beta_r + n_r - x_r)$, $r = 1,\ldots,k$, for brevity. The outer expectations are with respect to the marginally independent random variables $X_r \sim \text{PE}(n_r,\alpha_r,\beta_r,1)$, the Pólya-Eggenberger distribution given by (8), $r = 1,\ldots,k$.

In the second part of this section, the one-stage model considered above will be extended to a two-stage model, which can be summarized, after a standard reduction by sufficiency, as follows. At $\theta \in \Omega^k$, let X_i and Y_i be sufficient statistics of samples of sizes n_i and m_i from population P_i at Stage 1 and Stage 2, respectively, which altogether are independent. It is assumed that sampling at Stage 1 has been completed already, where $\mathbf{X} = (x_1,\ldots,x_k)$ has been observed, and that it is planned to allocate observations $\mathbf{Y} = (Y_1,\ldots,Y_k)$ for Stage 2, subject to $m_1+\ldots+m_k = m$, where m is fixed given.

First, let us consider the situation at the end of Stage 2, where both, $\mathbf{X} = \mathbf{x}$ and $\mathbf{Y} = \mathbf{y}$ have been observed. From (2) and (3) it follows that every Bayes selection rule $d^B(\mathbf{x},\mathbf{y})$, say, is determined by

$$E\{L(\Theta, d^B(\mathbf{x},\mathbf{y}))|\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y}\} = \min_{i=1,\ldots,k} E\{L(\Theta,i)|\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y}\}. \quad (13)$$

Here $\mathbf{X}$ and $\mathbf{Y}$ are not combined into an overall sufficient statistic, since the situation at the end of Stage 1 will be studied now. The criterion for allocating observations $\mathbf{Y}$ for Stage 2, after having observed $\mathbf{X}=\mathbf{x}$ at Stage 1, is to find $m_1,\ldots,m_k$, subject to the side condition $m_1+\ldots+m_k = m$, for which the following minimum is achieved.

$$\begin{aligned} &\min_{m_1+\ldots+m_k=m} E\{\, E\{L(\Theta, d^B(\mathbf{x},\mathbf{Y}))|\mathbf{X}=\mathbf{x},\mathbf{Y}\} \mid \mathbf{X}=\mathbf{x}\} \quad (14)\\ &= \min_{m_1+\ldots+m_k=m} E\{\, \min_{i=1,\ldots,k} E\{L(\Theta,i)|\mathbf{X}=\mathbf{x},\mathbf{Y}\} \mid \mathbf{X}=\mathbf{x}\}, \end{aligned}$$

where the outer expectation is with respect to the conditional distribution of $\mathbf{Y}$, given $\mathbf{X}=\mathbf{x}$. It should be pointed out that in (13) and (14), d^B does not only depend on $n_1,\ldots,n_k$, which are fixed here, but also on $m_1,\ldots,m_k$, which are varying, since every design of specific n's and m's has its own Bayes selection rules.

From now on it is assumed that Stage 1 has been completed already, i.e. that $\mathbf{X}=\mathbf{x}$ has been observed, and that a Bayes design for Stage 2 with $m_1+\ldots+m_k = m$ has to be determined. In this situation, it proves to be convenient to update the prior with the information provided by $\mathbf{X}=\mathbf{x}$ (Berger, 1985), i.e. to treat Stage 2 with observations $\mathbf{Y}$ as a first stage and to use the updated prior density $\pi(\theta \mid \mathbf{x})$ as a prior density. The Bayes designs are then all sampling allocations $m_1,\ldots,m_k$ for which the following minimum or maximum, respectively, is achieved.

$$\begin{aligned} &\min_{m_1+\ldots+m_k=m} E_{\mathbf{x}}(\min_{i=1,\ldots,k} E_{\mathbf{x}}\{L(\Theta,i)|\mathbf{Y}\}) && \text{in general,} \quad (15)\\ &\max_{m_1+\ldots+m_k=m} E_{\mathbf{x}}(\max_{i=1,\ldots,k} P_{\mathbf{x}}\{\Theta_i=\Theta_{[k]}|\mathbf{Y}\}) && \text{for } \textit{"0-1" loss,}\\ &\max_{m_1+\ldots+m_k=m} E_{\mathbf{x}}(\max_{i=1,\ldots,k} E_{\mathbf{x}}\{\Theta_i|\mathbf{Y}\}) && \text{for } \textit{linear loss,} \end{aligned}$$

where here and in the following, the subscript $\mathbf{x}$ at all probabilities and expectations indicates that the updated prior, based on $\mathbf{X}=\mathbf{x}$, is used as prior. Comparing now (15) with (10), one

can see that the design problem for Stage 2, posed at the end of Stage 1, can be considered as a one-stage design problem and treated as such.

On the other hand, if one prefers not to update the prior, then in (15), the inner operations E_x and P_x are with respect to the conditional distribution of Θ, given $\mathbf{X} = \mathbf{x}$ and $\mathbf{Y}$, and the outer operations E_x are with respect to the conditional distribution of $\mathbf{Y}$, given $\mathbf{X} = \mathbf{x}$. Both approaches are valid, equivalent, and lead to the same results.

To conclude this section, Bayes designs for the two special models will be studied under the *"0-1" loss* and the *linear loss*. For the *normal-normal* model, it can be shown (Gupta and Miescke 1994) that the Bayes designs consist of those sample sizes $m_1,\ldots,m_k$ which achieve the following respective maximum.

$$\max_{m_1+\ldots+m_k=m} E\left(\max_{i=1,\ldots,k} \int_{\Re} \prod_{j\neq i} \Phi\left(\vartheta_j^{-1}\left[\vartheta_i z+\mu_i(x_i)-\mu_j(x_j)+\gamma_i N_i-\gamma_j N_j\right]\right)\varphi(z)dz\right), \qquad (16)$$

for *"0-1" loss,*

$$\max_{m_1+\ldots+m_k=m} E\left(\max_{i=1,\ldots,k}\left[\mu_i(x_i)+\gamma_i N_i\right]\right),$$

for *linear loss,*

where $\vartheta_r = (p_r+q_r+\nu_r)^{-1/2}$, $\gamma_r = \vartheta_r(p_r+\nu_r)^{-1/2} q_r^{1/2}$, $\mu_r(x_r) = (p_r x_r+\nu_r\mu_r)/(p_r+\nu_r)$, with $p_r = n_r\sigma^{-2}$ and $q_r = m_r\sigma^{-2}$, $r = 1,\ldots,k$, and $N_1,\ldots,N_k$ are generic independent N(0,1) random variables.

The results for the special case of $k = 2$ are analogous to the one-stage Bayes design mentioned earlier. Results for the case of $k \geq 3$ are only known for the *linear loss*. They are based, in view of (16), on the properties of $E(max_{i=1,\ldots,k}\{\lambda_i+\tau_i N_i\})$ as a function of $\lambda_i \in \Re$ and $\tau_i > 0$, $i = 1,\ldots,k$. These properties have been derived in Gupta and Miescke (1996a) using the auxiliary function

$$T(w) = w\,\Phi(w) + \varphi(w) = \int_{-\infty}^{w} \Phi(v)\,dv\,, \quad w \in \Re\,. \tag{17}$$

Here the special case of $m = 1$ has been worked out in details, which is relevant for the Bayes sequential allocations to be considered in Section 4. Discussion of further results in this respect will thus be postponed until Section 4.

For the *binomial-beta* model, the Bayes designs for Stage 2 consist of those sample sizes $m_1,\ldots,m_k$ for which, subject to $m_1+\ldots+m_k = m$, the following maximums are achieved.

$$\max_{m_1+\ldots+m_k=m} \sum_{\mathbf{y}} \left(\max_{i=1,\ldots,k} \int_{\Re} \prod_{j\neq i} H_{j,x_j y_j}(\theta)\, h_{i,x_i y_i}(\theta)\, d\theta \right) P_{\mathbf{x}}\{\mathbf{Y} = \mathbf{y}\}, \quad \text{for } \textit{"0-1" loss}, \tag{18}$$

$$\max_{m_1+\ldots+m_k=m} \sum_{\mathbf{y}} \left(\max_{i=1,\ldots,k} \frac{a_i + y_i}{a_i + b_i + m_i} \right) P_{\mathbf{x}}\{\mathbf{Y} = \mathbf{y}\}\,, \qquad \text{for } \textit{linear loss},$$

where in the first criterion, H_{r,x_r,y_r} and h_{r,x_r,y_r} denote the c.d.f. and the density, respectively, of $\text{Beta}(a_r + y_r, b_r + m_r - y_r)$, with $a_r = \alpha_r + x_r$ and $b_r = \beta_r + n_r - x_r$, $r = 1,\ldots,k$, for brevity. The sums in (18) are expectations with respect to the conditional distribution of $\mathbf{Y}$, given $\mathbf{X} = \mathbf{x}$, which, analogously to (8), are given by

$$P_{\mathbf{x}}\{\mathbf{Y} = \mathbf{y}\} = \prod_{i=1}^{k} \binom{m_i}{y_i} \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} \frac{\Gamma(a_i + y_i)\Gamma(b_i + m_i - y_i)}{\Gamma(a_i + b_i + m_i)}\,, \tag{19}$$

where $y_i = 0,1,\ldots,m_i$, $i = 1,\ldots,k$.

No further results are known in this situation under the *"0-1" loss*. However, under the *linear loss*, interesting theoretical as well as numerical results have been found by Gupta and Miescke (1996b) and Miescke and Park (1997a). These are relevant for the Bayes sequential allocations and will be discussed in the next section.

To conclude this section, some comments will be made regarding cost of sampling. Suppose that every single observation costs a certain amount λ, say. Then (10) and (15) have

to be compared with the respective cost of sampling $n\lambda$ and $m\lambda$. If the cost of sampling turns out to be larger that the minimum posterior expectation given in (10) or (15), respectively, then apparently it is not worth taking all of these observations. This approach has been treated by Gupta and Miescke (1993), within the problem of combined selection and estimation, in the *binomial-beta* model. It leads, among other considerations, to finding the largest sample size, subject to its given upper bound, which is worth allocating to incur a gain. In the next section, where observations are allocated and taken in a sequential fashion, cost of sampling would require the incorporation of a stopping rule. However, this will not be done there to keep the presentation of basic ideas simple. Modifications of these sequential allocation rules to this more general setting are straightforward, but more involved. Therefore, cost of sampling will not be considered any further from now on.

4. BAYES LOOK-AHEAD SEQUENTIAL SAMPLING DESIGNS

From now on it is assumed that Stage 1 has been completed, i.e. that $\mathbf{X} = \mathbf{x}$ has been observed already, and that m additional observations $\mathbf{Y} = (Y_1,\ldots,Y_k)$ are planned to be drawn at Stage 2. The optimum allocations of sample sizes $m_1,\ldots,m_k$, i.e. the Bayes designs, are determined by criterion (15), i.e. by

$$\min_{m_1+\ldots+m_k=m} E_{\mathbf{x}}\Big(\min_{i=1,\ldots,k} E_{\mathbf{x}}\{L(\Theta,i)\,|\,\mathbf{Y}\}\Big)\,. \tag{20}$$

The first step toward sequential Bayes designs is to consider an intermediate step of Stage 2 sampling, where so far only $\tilde{\mathbf{Y}} = \tilde{\mathbf{y}}$ has been observed, with $\tilde{m}_i$ observations from population P_i, $i = 1,\ldots,k$, where $\tilde{m}_1 + \ldots + \tilde{m}_k = \tilde{m}$, and where $\tilde{m}$ with $1 \le \tilde{m} < m$ is fixed. The

best allocation of the remaining $\hat{m} = m - \tilde{m}$ observations with $\hat{m}_i = m_i - \tilde{m}_i \geq 0$, $i = 1,\ldots,k$, achieves

$$\min_{\hat{m}_1+\ldots+\hat{m}_k=\hat{m}} E_{\mathbf{x}}\left\{ \min_{i=1,\ldots,k} E_{\mathbf{x}}\{L(\Theta,i) \mid \tilde{\mathbf{Y}} = \tilde{\mathbf{y}}, \hat{\mathbf{Y}}\} \,\middle|\, \tilde{\mathbf{Y}} = \tilde{\mathbf{y}} \right\}, \tag{21}$$

where the outer expectation is with respect to the conditional distribution of the new observations $\hat{\mathbf{Y}}$, say, given $\tilde{\mathbf{Y}} = \tilde{\mathbf{y}}$ (and $\mathbf{X} = \mathbf{x}$).

Returning now to the end of Stage 1, the optimum two-step allocation for drawing first $\tilde{m}$ and then $\hat{m} = m - \tilde{m}$ observations at Stage 2 is found by backward optimization. First one has to consider every possible sample size configuration $\tilde{m}_1,\ldots,\tilde{m}_k$ and every possible outcome $\tilde{\mathbf{Y}} = \tilde{\mathbf{y}}$. For each such setting, one allocation $\hat{m}_i(\tilde{\mathbf{y}},\tilde{m}_1,\ldots,\tilde{m}_k)$, $i = 1,\ldots,k$, has to be found which achieves (21). Then one has to find an allocation $\tilde{m}_1,\ldots,\tilde{m}_k$ which achieves

$$\min_{\tilde{m}_1+\ldots+\tilde{m}_i=\tilde{m}} E_{\mathbf{x}}\left(\min_{\hat{m}_1+\ldots+\hat{m}_i=\hat{m}} E_{\mathbf{x}}\left\{ \min_{i=1,\ldots,k} E_{\mathbf{x}}\{L(\Theta,i) \mid \tilde{\mathbf{Y}}, \hat{\mathbf{Y}}\} \,\middle|\, \tilde{\mathbf{Y}} \right\} \right). \tag{22}$$

Here one should be aware of the fact that the information contained in $\mathbf{Y}$ is the combined information gained from $\tilde{\mathbf{Y}}$ and $\hat{\mathbf{Y}}$. It should also be pointed out clearly that in the middle minimization operation of (22), $\hat{m}_1,\ldots,\hat{m}_k$ depend on $\tilde{m}_1,\ldots,\tilde{m}_k$ and on $\tilde{\mathbf{Y}}$, and thus they are random variables themselves! This is the very reason why (22) can be handled numerically and in computer simulations in the *binomial-beta* model, where $\mathbf{Y}$ is discrete and assumes only finitely many values, but not in the *normal-normal* model.

Comparing now (20) with (22), one can show that the latter must be less than or equal to the former. If one deletes in (22) the minimum to the right of the first expectation, and inserts a

minimum to the left of the same, subject to $\hat{m}_i = m_i - \tilde{m}_k \geq 0$, and subject to $\hat{m}_1 + \ldots + \hat{m}_k = \hat{m}$, then the resulting value cannot be smaller. Combining now the two iterated minimization operations into one leads to (20). To summarize, one can state (Miescke and Park, 1997a) the following result.

Theorem 1. *For fixed* m *and* $\tilde{m} < m$, *the best allocation for drawing first* $\tilde{m}$ *and then* $\hat{m} = m - \tilde{m}$ *observations at Stage 2 is at least as good as the best allocation of all* m *observations in one step, in the sense that the posterior Bayes risk* (22) *of the former is not larger than that one of the latter, given by* (20). *This process of stepwise optimum allocation can be iterated for further improvements. The overall best allocation scheme is to draw, in* m *steps, one observation at a time, which are determined by backward optimization.*

In view of this theorem, several reasonable sampling allocation schemes can be constructed which utilize information gained from observations at previous steps. Let R_t, for $t \leq m$, denote the allocation of t observations determined by (15), with m replaced by t there. Moreover, let $R_{t,1}$ allocate any single observation to one of the populations sampled by R_t. In a similar way let $R^*_{t,1}$ allocate one observation to one of the populations to which R_t assigns the largest allocation. Finally, denote by B_1 the optimum allocation of one observation, knowing all future allocation strategies. It should be pointed out that, unlike the other allocations considered above, B_1 is not a stand-alone procedure, since it requires the knowledge of what will be done after its has been applied.

Using these three types of intermediate allocation rules, the following schemes of allocating m observations are possible. (R_m) allocates all m observations at once, using (15). This fixed sample size m Bayes design will be denoted by **OPT** in the following. A better

allocation scheme, in terms of the Bayes risk, is $(R_{m,1},R_{m-1})$, which uses $R_{m,1}$ for the first allocation, and then R_{m-1} for the rest. Better than $(R_{m,1},R_{m-1})$, of course, is (B_1,R_{m-1}), which uses backward optimization B_1 for the first allocation, knowing that R_{m-1} will be used for allocating the remaining $m-1$ observations in one step. In this fashion, similar and also more complicated allocation schemes can be constructed (Gupta and Miescke 1996a), which are linked through a partial ordering in terms of their Bayes risks. Such constructions are motivated by the fact that the overall optimum allocation scheme $(B_1,B_1,...,B_1,R_1)$, denoted by **BCK**, is not practicable, except for small m and k, up to about $m=20$ for $k=3$, in the *binomial-beta* model. For this model, the allocation scheme **APP**, say, which is $(R^*_{m,1},R^*_{m-1,1},...,R^*_{2,1}R_1)$, appears to be a very good approximation to **BCK** under the *linear loss*. This will be justified at the end of this section.

The allocation scheme $(R_1,R_1,...,R_1)$ allocates in m steps one observation at a time, using R_1, pretending that it would be the last one before making the final (selection) decision. It looks ahead one observation at a time (Berger, 1985, Amster, 1963) and will be henceforth denoted by **LAH**. It should not be confused with the allocation scheme **SOA**, say, which allocates in m steps one observation at a time, using the "state of the art". To be more specific, suppose that $\tilde{\mathbf{Y}}=\tilde{\mathbf{y}}$ has been drawn so far. Then **SOA** allocates the next observation to any one of those populations which are associated with the minimum of the k values of $E_x\{L(\Theta,i)\,|\,\tilde{\mathbf{Y}}=\tilde{\mathbf{y}}\}$, $i=1,...,k$. Two other allocation schemes, which will be considered later in the simulation study of the *binomial-beta* model, should be mentioned here. The first assigns one observation at a time, each purely at random, regardless of the previous observations, and

is denoted by **RAN**. The second assigns m/k observations to each populations, provided that m is divisible by k, and is denoted by **EQL**.

Theoretical results for allocation scheme **LAH** in the *normal-normal* model and under the *linear loss*, which are presented in Gupta and Miescke (1994,1996a), will now be discussed. Here, it is sufficient to consider the first allocation R_1 in $(R_1, R_1, \ldots, R_1)$, which is based on $\mathbf{X} = \mathbf{x}$. All consecutive allocations R_1 are decided analogously, based on $\mathbf{X} = \mathbf{x}$ and the observations $\tilde{\mathbf{Y}} = \tilde{\mathbf{y}}$ that have been taken so far at Stage 2. Starting with criterion (16) for the *linear loss* with $m = 1$, where exactly one of the sample sizes $m_1, \ldots, m_k$ is equal to one, and all others are zero, i.e. where exactly one of $q_1, \ldots, q_k$ is equal to σ^{-2} and all others are zero, this first observation is taken from one of the populations which yield

$$\max_{i=1,\ldots,k} E\left(\max\left\{ \mu_i(x_i) + \sigma_i N_i, \max_{j \neq i}\{\mu_j(x_j)\} \right\} \right), \tag{23}$$

$$= \max_{i=1,\ldots,k} \left\{ \mu_i(x_i) + \sigma_i T([\max_{j \neq i}\{\mu_j(x_j)\} - \mu_i(x_i)]/\sigma_i) \right\},$$

where $\sigma_r = (p_r + \sigma^{-2} + \nu_r)^{-1/2}(p_r + \nu_r)^{-1/2}\sigma^{-1}$, and $\mu_r(x_r)$, $r = 1, \ldots, k$, are defined below of (16), and the function T is given by (17).

To describe the properties of the first allocation R_1 in $(R_1, R_1, \ldots, R_1)$, it proves useful to consider the ordered values $\mu_{[1]}(\mathbf{x}) < \mu_{[2]}(\mathbf{x}) < \ldots < \mu_{[k]}(\mathbf{x})$ of $\mu_r(\mathbf{x}) = \mu_r(x_r)$, $r = 1, \ldots, k$. Let $P_{(i)}$ be the population, and let $\sigma_{(i)}$ be the standard deviation, which is associated with $\mu_{[i]}(\mathbf{x})$, $i = 1, \ldots, k$. Then one can state the following result.

Theorem 2. *After* $\mathbf{X}=\mathbf{x}$ *has been observed, the preferences of the first allocation* R_1 *in* $(R_1,R_1,\ldots,R_1)$ *are as follows.*

(I) *If* $\sigma_{(k-1)} < (=,>)\, \sigma_{(k)}$, *then allocating to* $P_{(k-1)}$ *is worse than (equivalent to, better than) allocating to* $P_{(k)}$.

(II) *If for* $1 \le i < j \le k-2$, $\sigma_{(i)} \le \sigma_{(j)}$, *then allocating to* $P_{(i)}$ *is worse than allocating to* $P_{(j)}$.

(III) *If for* $1 \le i < j \le k-2$, $\sigma_{(i)} > \sigma_{(j)}$, *and* $\sigma_{(i)} < (=,>)\, \sigma_{ij}$, *then allocating to* $P_{(i)}$ *is worse than (equivalent to, better than) allocating to* $P_{(j)}$, where σ_{ij} *is determined by*

$$\mu_{[j]}(\mathbf{x}) + \sigma_{(j)} T\left((\mu_{[k]}(\mathbf{x}) - \mu_{[j]}(\mathbf{x}))/\sigma_{(j)}\right) = \mu_{[i]}(\mathbf{x}) + \sigma_{ij} T\left((\mu_{[k]}(\mathbf{x}) - \mu_{[i]}(\mathbf{x}))/\sigma_{ij}\right). \qquad (24)$$

(IV) *Let* $P_{(*)}$ *be a best allocation to either* $P_{(k-1)}$ *or* $P_{(k)}$ *according to* (I). *Likewise, let* $P_{(\bullet)}$ *be a best allocation to* $P_{(1)},\ldots,P_{(k-2)}$ *according to* (II) *and* (III). *Then an overall best allocation is found by using* (II) *and* (III) *with* (i), (j), σ_{ij}, $\mu_{[i]}(\mathbf{x})$, *and* $\mu_{[j]}(\mathbf{x})$ *replaced by* $(\bullet)$, $(*)$, $\sigma_{\bullet *}$, $\mu_{[\bullet]}(\mathbf{x})$, *and* $\mu_{[k-1]}(\mathbf{x})$.

The proof can be found in Gupta and Miescke (1996a), along with further comments. Moreover, a numerical example with real life data is presented there, in which also comparisons are made with respect to standard multiple comparison procedures, such as the Scheffé's and Tukey's methods.

Theoretical and numerical simulation results for the *binomial-beta* model under the *linear loss* are presented in Gupta and Miescke (1996b) and Miescke and Park (1997a). These will be discussed in the remainder of this section. First, properties of **LAH**, the allocation scheme $(R_1,R_1,\ldots,R_1)$, will be studied. As it has been justified above, it suffices to consider the first allocation R_1 of it. Starting with criterion (18) for the *linear loss* with $m=1$, where exactly

one of the sample sizes $m_1,\ldots,m_k$ is equal to one, and all others are zero, this first observation is taken from one of the populations which yield the following maximum.

$$\max_{i=1,\ldots,k} E_x\left\{\max\left\{\frac{a_i+Y_i}{a_i+b_i+1}, \max_{j\neq i}\left\{\frac{a_j}{a_j+b_j}\right\}\right\}\right\} = \tag{25}$$

$$\max_{i=1,\ldots,k}\left\{\max\left\{\frac{a_i+1}{a_i+b_i+1}, \max_{j\neq i}\left\{\frac{a_j}{a_j+b_j}\right\}\right\}\frac{a_i}{a_i+b_i} + \max\left\{\frac{a_i+0}{a_i+b_i+1}, \max_{j\neq i}\left\{\frac{a_j}{a_j+b_j}\right\}\right\}\frac{b_i}{a_i+b_i}\right\},$$

where $a_r=\alpha_r+x_r$ and $b_r=\beta_r+n_r-x_r$, $r=1,\ldots,k$, for brevity. To summarize, one can state here the following result.

Theorem 3. *The first allocation R_1 in $(R_1,R_1,\ldots,R_1)$ is made with respect to one of the populations P_i, $i=1,\ldots,k$, for which the maximum in* (25) *is achieved. All consecutive allocations of the type R_1 are made analogously, with a_r and b_r updated to $\alpha_r+x_r+\tilde{y}_r$ and $\beta_r+n_r-x_r+\tilde{m}_r-\tilde{y}_r$, respectively, with respect to the $\tilde{m}_r$ observations, represented by $\tilde{y}_r$, $r=1,\ldots,k$, which have been made so far at Stage 2.*

The proof is given in Gupta and Miescke (1996b), along with further details of the behavior of this allocation scheme. One interesting point, worth to be mentioned, is related to the fact that under the *linear loss* (with no cost of sampling), the Bayes look ahead risk cannot increase when more future observations are included. This fact implies that the maximum in (25) is always greater than or equal to $\max_{i=1,\ldots,k}\{a_i/(a_i+b_i)\}$. However, equality may occur, in which case one additional observation from any of the populations would not be worth to be taken, from the Bayesian point of view. A similar situation may arise at any allocation of the type R_1 in $(R_1,R_1,\ldots,R_1)$, and especially at the last allocation. In the latter case, one can accelerate the process by stopping one observation short of m. In the context of sequentially

allocating m observations at Stage 2, however, this is of minor concern and will not be considered any further.

Numerical results for allocation schemes **EQL** and **OPT**, and computer simulation results for **RAN**, **SOA**, and **LAH** have been presented for $k = 3$ populations in Gupta and Miescke (1996b). The computer programs were written in Microsoft Quick Basic Version 4.5, using subroutines from Sprott (1991). These results have been extended, using Microsoft Visual Basic Version 4.0, to numerical results for the overall optimum allocation scheme **BCK** in Miescke and Park (1997a), where it has been also found that **APP** appears to be a very good approximation to **BCK**. In summary, the performances of the following allocation schemes have been studied, which have been explained in more details earlier in this section.

RAN	Assign one observation at a time, each purely at random.
EQL	Assign $m/3$ observations to each population P_1, P_2, P_3.
SOA	Assign one observation at a time, following the state of the art.
LAH	Assign one observation at a time, using $(R_1, R_1, \ldots, R_1)$.
OPT	Assign m_t observations to P_t, $t = 1,2,3$, using (R_m).
APP	Assign one observation at a time, using $(R^*_{m,1}, R^*_{m-1,1}, \ldots, R^*_{2,1}, R_1)$.
BCK	Assign one observation at a time, using backward optimization.

In three examples, with suitably chosen values for $a_r = \alpha_r + x_r$ and $b_r = \beta_r + n_r - x_r$, $r = 1, 2, 3$, to cover various interesting settings, the performances of these allocation schemes have been compared. The values for m considered have been 1, 3, 9, and 15. For $m = 1$, **EQL** has been set to take its observation from population P_1, rather than leaving the respective spaces empty in the tables. As to ties, **RAN**, **SOA**, **LAH**, and **APP** have been used, and are recommended to be used, with ties broken purely at random, with equal probabilities,

whenever they occur. This recommendation is corroborated by findings in the numerical studies.

Comparing the expected posterior gains of the first five allocation schemes, it turns out that overall, **LAH** and **OPT** are performing similarly well, each sometimes better than the other, but clearly better than **RAN**, **EQL**, and **SOA**. The latter effect is found to be increasing in m. That **LAH** is not always as good as **OPT** proves that it cannot be any version, i.e. with any type of breaking ties, of the allocation scheme $(R_{m,1}, R_{m-1,1}, \ldots, R_{2,1} R_1)$, and thus in particular it cannot be equal to $(R^*_{m,1}, R^*_{m-1,1}, \ldots, R^*_{2,1} R_1)$, i.e. **APP**, since the latter two are always at least as good as **OPT**. One advantage of **LAH**, besides its easy implementation, is that each of its individual allocations is self-contained. Thus, if in an ongoing experiment the total number m of observations has to be changed, this has only minor effects on its usage.

The numerical results for **BCK** in Miescke and Park (1997a) became feasible with the release of Microsoft Visual Basic Version 4.0 , which allows to handle, on a typical IBM type Pentium Computer, a 6-dimensional array (for the a_i's and b_i's) with a common subscript range of $1,2,\ldots,15$ (for the $\tilde{m}_i$'s), i.e. more than 10^7 variables. As anticipated, **LAH** and **OPT** turn out to be good approximations to **BCK**.

One striking fact has been observed by comparing the first allocation of **BCK** with the allocation m_1, m_2, m_3 of **OPT**. In all but one of the 96 parameter settings considered in the three examples, the population to which **OPT** allocates the largest sample size is one of those which **BCK** would allow to start with. This indicates clearly that the allocation scheme $(R^*_{m,1}, R^*_{m-1,1}, \ldots, R^*_{2,1} R_1)$, i.e. **APP**, should be considered as a good approximation to **BCK**, and thus be used in practice. A study of the performance of **APP** within the framework of the three

examples does not appear to be feasible at this time because of the length of the computing time required for such a task. It would be a combination of calculating the individual steps R_{t1}^{*}, randomizing tied populations, and simulating the m outcomes.

ACKNOWLEDGMENTS

The author is grateful to a Referee for carefully reading the manuscript and for suggested changes.

REFERENCES

MM Abughalous, KJ Miescke. On selecting the largest success probability under unequal sample sizes. J Statist Plann Inf 21:53-68, 1989.

SJ Amster. A modified Bayesian stopping rule. Ann Math Statist 34:1404-1413, 1963.

RE Bechhofer. Optimal allocation of observations when comparing several treatments with a control. In: Krishnaiah PR, ed. Multivariate Analysis. New York, Academic Press, 1969, pp. 465-473.

RE Bechhofer, AJ Hayter, AC Tamhane. Designing experiments for selecting the largest normal mean when the variances are known and unequal: optimal sample size allocation. J Statist Plann Inf 28:271-289, 1991.

RE Bechhofer, J Kiefer, M Sobel. Sequential Identification and Ranking Procedures. Chicago, IL: The University of Chicago Press, 1968.

RE Bechhofer, RV Kulkarni. Closed adaptive sequential procedures for selecting the best of $k \geq 2$ Bernoulli populations. In: Gupta SS, Berger JO, eds. Statistical Decision Theory and Related Topics - III, Vol. 1. New York, NY: Academic Press 1982, pp. 61-108.

RE Bechhofer, TJ Santner, DM Goldsman. Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons. New York, NY: John Wiley, 1995.

JO Berger. Statistical Decision Theory and Bayesian Analysis. 2nd ed. New York, NY: Springer, 1985.

JO Berger, J Deely. A Bayesian approach to ranking and selection of related means with alternatives to AOV methodology. J Amer Statist Assoc 83:364-373, 1988.

TL Bratcher, RP Bland. On comparing binomial probabilities from a Bayesian viewpoint. Commun Statist 4:975-985, 1975.

TL Bratcher, P. Bhalla. On the properties of an optimal selection procedure. Commun Statist 3:191-196, 1974.

H Chernoff, JA Yahav. A subset selection problem employing a new criterion. In: Gupta SS, Moore DS, eds. Statistical Decision Theory and Related Topics - II, New York, NY: Academic Press 1977, pp. 93-119.

A Cohen, HB Sackrowitz. A decision theoretic formulation for population selection followed by estimating the mean of the selected population. In: Gupta SS, Berger JO, eds. Statistical Decision Theory and Related Topics - IV, Vol.2, New York, Springer, 1988, pp. 33-36.

JJ Deely, SS Gupta. On the properties of subset selection procedures. Sankhya A-30:37-50, 1968.

EJ Dudewicz, SR Dalal. Allocation of observations in ranking and selection with unequal variances. Sankhya B-37:28-78, 1975.

CW Dunnett. On selecting the largest of k normal population means. - With Discussion. J Royal Statist Soc B-22:1-40, 1960.

DKH Fong, JO Berger. Ranking, estimation, and hypothesis testing in unbalanced two-way additive models - A Bayesian approach. Statistics and Decisions 11:1-24, 1993.

PK Goel, H Rubin. On selecting a subset containing the best population - a Bayesian approach. Ann Statist 5:969-983, 1977.

Z Govindarajulu, MN Katehakis. Dynamic allocation in survey sampling. Amer J of Math and Managem Sci 11:199-221, 1991.

SS Gupta. On a decision rule for a problem in ranking means. Mimeo.Ser.No. 150, Inst. of Statist., Univ. of North Carolina, Chapel Hill, NC, 1956.

SS Gupta. On some multiple decision (selection and ranking) rules. Technometrics 7:225-245, 1965.

SS Gupta, JC Hsu. On the monotonicity of Bayes subset selection procedures. In: Proceedings of the 41st Session of the ISI, Book 4:208-211, 1977.

SS Gupta, KJ Miescke. On two-stage Bayes selection procedures. Sankhya B-46, Pt.2:123-134, 1984a.

SS Gupta, KJ Miescke. Sequential selection procedures - A decision theoretic approach. Ann Statist 12:336-350, 1984b.

SS Gupta, KJ Miescke. On the problem of finding the largest normal mean under heteroscedasticity. In: Gupta SS, Berger JO eds. Statistical Decision Theory and Related Topics - IV, Vol. 2. New York, NY: Springer Verlag, 1988, pp 37-49.

SS Gupta, KJ Miescke. On finding the largest normal mean and estimating the selected mean. Sankhya B-52, Pt.2:144-157, 1990.

SS Gupta, KJ Miescke. On combining selection and estimation in the search of the largest binomial parameter. J Statist Plann Inf 36:129-140, 1993.

SS Gupta, KJ Miescke. Bayesian look ahead one-stage sampling allocations for selecting the largest normal mean. Statistical Papers 35:169-177, 1994.

SS Gupta, KJ Miescke. Bayesian look ahead one-stage sampling allocations for selection of the best population. J Statist Plann Inf. 54:229-244, 1996a.

SS Gupta, KJ Miescke. Bayesian look ahead one-stage sampling allocations for selection of the best Bernoulli population. In: Eds. Brunner E, Denker M, eds. Research Developments in Probability and Statistics, Festschrift in Honor of ML Puri., Zeist: VSP International Publishers, 1996b, pp 353-369.

SS Gupta, S Panchapakesan. Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations, New York, NY: John Wiley, 1979.

SS Gupta, S Panchapakesan. Sequential ranking and selection procedures. In: Ghosh BK, Sen, PK, eds. Handbook of Sequential Analysis. New York, NY: M. Dekker, 1991, pp 363-380.

NL Johnson, S Kotz. Discrete Distributions. Boston, MA: Houghton and Mifflin, 1969.

K Lam, WK Chiu. On the probability of correctly selecting the best of several normal populations. Biometrika 63:410-411, 1976.

KJ Miescke. Bayesian subset selection for additive and linear loss functions. Commun Statist A8-12:1205-1226,1979.

KJ Miescke. Recent results on multi-stage selection procedures. In: Iosifescu M, ed. Proceedings of the Seventh Conference on Probability Theory. Bucharest, Romania: Editura Academiei Republicii Socialiste Romania, 1984, pp 259-268.

KJ Miescke, H Park. Bayes m-truncated sampling allocations for selecting the best Bernoulli population. In: Balakrishnan N, Panchapakesan S, eds., Festschrift for Shanti S. Gupta. 15 pp. in LaTeX, to appear, 1997a.

KJ Miescke, H Park. On the admissibility of the natural selection rule. Tech. Report Series, Fachbereich Statistik, Universität Dortmund, 14 pp. in MS Word 7, 1997b.

KJ Miescke, D Rasch. (eds.). 40 Years of Statistical Selection Theory", Part I & II. Special Issue, J Statist Plann Inf 54-No.2 & 3, 1996.

KR Risko. Selecting the better binomial population with unequal sample sizes. Comm Statist - Theory Meth 14:123-158, 1985.

JC Sprott. Numerical Recipes: Routines and Examples in BASIC, Cambridge, MA: Cambridge University Press, 1991.

YL Tong, DE Wetzell. On the behavior of the probability function for selecting the best normal population. Biometrika 66:174-176, 1979.

J Whitehead. Sequential methods in clinical trials. In: Ghosh BK, Sen PK, eds., Handbook of Sequential Analysis, New York, Dekker, 1991, pp 593-611.

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August, 1997	Technical Report; August 1997

4. TITLE AND SUBTITLE
Bayes Sampling Designs for Selection Procedures

5. FUNDING NUMBERS
DAAH04-95-0165

6. AUTHOR(S)
Klause J. Miescke

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
Purdue University
Department of Statistics
West Lafayette, IN 47907-1399

8. PERFORMING ORGANIZATION REPORT NUMBER
Technical Report #97-12C

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

From k independent populations $P_1, \ldots, P_k$, which belong to one parameter exponential family $\{F_\theta\}$, $\theta \in \Omega \subseteq \mathcal{R}$, random samples of sizes $m_1, \ldots, m_k$, respectively, are to be drawn. After the observations have been drawn, a selection procedure will be used to determine which of these k populations has the largest value of θ. Given a loss for selections at each parameter configuration, given n past observations, and given a prior for the k parameters, a Bayes selection procedure can be found and its Bayes risk can be determined, where both depend on $m_1, \ldots, m_k$. Let the sample sizes be restricted by $m_1 + \ldots + m_k = m$, where m is fixed. The problem of how to find the optimum (minimum Bayes risk) sample design subject to this constraint is considered, as well as m-truncated sequential sampling allocations. Results for *normal* and *binomial* families, under the *"0-1" loss* and the *linear loss*, are presented and discussed. An introduction to Bayes selection procedures in included.

14. SUBJECT TERMS	15. NUMBER IF PAGES
Selection procedures; Bayesian designs; optimum sampling allocations; binomial populations; normal populations	34
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL