HIERARCHICAL BAYESIAN SELECTION PROCEDURES
FOR THE BEST BINOMIAL POPULATION *

by

John J. Deely
University of Canterbury

Shanti S. Gupta
Purdue University

Technical Report # 88-21C

Department of Statistics
Purdue University

May 1988
Revised November 1994
Revised February 2000

* This research was supported in part by NSF Grants DMS-8606964, DMS-8702620, and DMS-8923071 at Purdue University.

△

HIERARCHICAL BAYESIAN SELECTION PROCEDURES FOR THE BEST BINOMIAL POPULATION

by

John J. Deely
University of Canterbury
and
Shanti S. Gupta
Purdue University

Abstract

In this paper a hierarchical Bayesian (HB) model is adopted to derive selection procedures for selecting the best of k binomial parameters, say the probability of success corresponding to k different suppliers. This model facilitates the use of prior information in the analysis for both small and large sample sizes. In addition to computing posterior probabilities that the i^{th} supplier is best, this paper presents expressions for deciding how much better a given supplier is relative to the others. Prior information is assumed to begin with exchangeability and can be more informative if the experimenter has other knowledge about the suppliers as a group. A robust Bayesian approach is also incorporated. A numerical example is provided to illustrate the techniques proposed.

AMS 1980 Subject Classification: 62F07, 62C12.

Key words and phrases: binomial, Bayes, hierarchical, selection procedures, exchangeable, hyperpriors, empirical Bayes

Hierarchical Bayesian Selection Procedures

for the Best Binomial Population

1. Introduction

Suppose there are k suppliers of a particular item and a sample of n_i items is taken from the i^{th} supplier yielding x_i the number of successes (or defectives) in the sample. Then X_i has a binomial distribution with parameter θ_i, which denotes the true unknown probability of a success (failure) from the i^{th} supplier. The *best* supplier is defined to be the one with largest (smallest) θ_i. Based on the observed data and prior information available, we seek procedures which will select a non-empty subset of the k suppliers and assert with some confidence that the best supplier is amongst those in the selected subset. The approach in this paper is to present a model which has the capability of incorporating prior information concerning suppliers as a group into the derivation of selection procedures. The HB model is one way this can be done easily and with useful results. This application to binomial data parallels and further extends the ideas contained in the normal means problem considered by Berger and Deely (1988) and the Poisson rates described in Smith and Deely (1997). In particular we show how this model can

- accommodate *unequal* and small sample sizes;
- incorporate the practitioner's knowledge of *suppliers as a group* into the selection procedures;
- provide easily computable answers to the practical question of which supplier is best by *how much.*

The general problem of selecting the best binomial population has received considerable attention in the literature but mainly from a non-Bayesian approach. Pioneering papers by Sobel and Huyett (1957) and Gupta and Sobel (1960) dealt with selecting the best and selecting a subset containing the best binomial population respectively. Later, Gupta, Huang and Huang (1976) studied a conditional subset selection rule and a related test of homogeneity. Further treatment with new results were also presented in Gupta and McDonald (1986). A good discussion of these and other non-Bayesian papers can be found in books by Gibbons, Olkin and Sobel (1977) and Gupta and Panchapakesan (1979).

Empirical Bayes approach to statistical inference was started by Robbins (1956). Later Deely (1965) and Deely and Gupta (1968) developed empirical Bayes procedures for general selection problems including among them the binomial case with independent θ_i's, each with a beta prior with unknown parameters. Bratcher and Bland (1975) considered a naive Bayesian approach in which the θ_i's are independent with known but perhaps different beta priors. They considered various multiple comparisons based on computing the posterior probabilities of each population being best and used numerical integration to calculate these. Gupta and Liang (1986) and Gupta, Liang and Rau (1994) derived non-parametric empirical Bayes procedures for selecting the best binomial population under the assumption that $\theta_1, \ldots, \theta_h$ are independent each with an unknown non-parametric prior distribution. Also later Yang (1987) applied their model but adopted the so called **PP*** criterion, which had been previously introduced for a general selection problem by Gupta and Yang (1985). This criterion, in an effort to relate the Bayesian criterion to the classical **P*** condition, states that the Bayes **P*** procedure selects the smallest subset for which the posterior probability that the subset contains the best is at least **P***. Subset selection procedures for binomial models based on a class of priors were obtained by Gupta and Liao (1993).

In addition there have been other relevant papers dealing with *estimation* as opposed to *selection* for the binomial case. Albert (1984) considers the simultaneous estimation of

k binomial probabilities and develops empirical Bayes estimators under an exchangeable hierarchical model. Leonard (1972) also considers this problem but uses a logit transformation to bring the problem into a multivariate normal context. A lot acceptance problem was considered by Eaves (1980) in which n items are drawn from each of k lots under binomial sampling. An exchangeable hierarchical model is assumed and the predictive distribution for the next lot is computed when all items from all lots are good.

A related problem, that of allocating the observations to the various suppliers constrained by the fact that the total is fixed, has also received some attention in the literature. Brooks (1987) deals with a Bayesian approach for $k = 2$, while Brittain and Schlesselman (1982) discuss this case from a frequentist viewpoint when trying to estimate $p_1 - p_2$ or p_1/p_2. These problems will not be discussed in any detail but some suggestions about allocation will be made in Section 5.

Thus it appears that the literature to date has largely ignored the situation in which prior information about the *group* of suppliers as opposed to suppliers *individually* is available. Whereas it has been recognized that such prior information is useful in various other problems (see for example Berger (1985), Chapter 3), the application to the important practical binomial selection problem has been ignored. Furthermore the selection procedures derived so far for binomial populations have not addressed the issue of "by how much" is the best supplier "best" (see Smith and Deely (1997) for a description of this concept in Poisson context). It is the purpose of this paper to address both of these deficiencies. A more thorough discussion is contained in Section 3 after having presented the mathematical details of the model in Section 2. Examples illustrating various aspects of the model are given in Section 4 with concluding remarks and suggestions for further work given in Section 5.

2. Mathematical details, the prior distribution and selection criteria

Let $\underset{\sim}{x} = (x_1, \ldots, x_k)$ be the vector of observations from the k suppliers, x_i conditional on θ_i having the binomial distribution

$$f(x_i|\theta_i) = \binom{n_i}{x_i} \theta_i^{x_i}(1-\theta_i)^{n_i - x_i},$$

and let $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ be the vector of unknown parameters for which we want to select that supplier with largest θ_i. The prior distribution $\pi(\underline{\theta})$ on $\underline{\theta}$ will be obtained via the hierarchical Bayesian structure (see Berger (1985), Section 4.6) in which $\pi(\underline{\theta})$ is given as a mixture of a conditional prior with hyperparameters β and η and a hyperprior distribution on these parameters; that is,

$$\pi(\underline{\theta}) = \int\int \pi(\underline{\theta}|\beta, \eta) h(\beta, \eta) d\beta d\eta.$$

Conditional upon the hyperparameters β, η, the components of $\underline{\theta}$ are assumed to be i.i.d. with a common beta distribution given by

$$\pi(\theta_i|\beta, \eta) = \frac{\Gamma(1/\eta)}{\Gamma(\beta/\eta)\Gamma((1-\beta)/\eta)} \theta_i^{\frac{\beta}{\eta}-1}(1-\theta_i)^{\frac{1-\beta}{\eta}-1} \tag{2.1}$$

where $0 < \beta < 1, \eta > 0$; thus

$$\pi(\underline{\theta}|\beta, \eta) = \prod_{i=1}^{k} \pi(\theta_i|\beta, \eta).$$

This particular form of the beta distribution will be convenient for the numerical computations and elicitation of prior information. These topics will be discussed more fully in the next section. Special note is taken here that

$$\beta = E(\theta_i|\beta, \eta) \text{ and } (1-\beta)\beta(\eta/(\eta+1)) = \text{ Var } (\theta_i|\beta, \eta) = \sigma^2 \tag{2.2}$$

We will use the following notation for the beta distribution:

$$g(y|a,b) = B(a,b)y^{a-1}(1-y)^{b-1}$$

$$B(a,b) = \Gamma(a+b)/(\Gamma(a)\Gamma(b))$$

$$G(t|a,b) = \int_0^t g(y|a,b)dy.$$

The form of hyperpriors on β and η will be discussed in Section 3 in the context of various types of prior information.

Under the notation and assumptions above, it follows that the conditional distribution of $\underline{\theta}$ given $\underline{x}, \beta$ and η is given by

$$\pi(\underline{\theta}|\underline{x},\beta,\eta) = \prod_{i=1}^{k} \pi(\theta_i|x_i,\beta,\eta) \tag{2.4}$$

where

$$\pi(\theta_i|x_i,\beta,\eta) = \frac{f(x_i|\theta_i)\,\pi(\theta_i|\beta,\eta)}{f(x_i|\beta,\eta)} = g(\theta_i|a_i,b_i),$$

$$f(x_i|\beta,\eta) = \int_0^1 f(x_i|\theta_i)\,\pi(\theta_i|\beta,\eta)\,d\theta_i = \binom{n_i}{x_i} B(a,b)/B(a_i,b_i)$$

and $a = \beta/\eta, b = (1-\beta)/\eta, a_i = a + x_i, b_i = b + n_i - x_i$. Let

$$f(\underline{x}|\beta,\eta) = \prod_{i=1}^{k} f(x_i|\beta,\eta) \text{ and } f(\underline{x}) = \int_0^\infty \int_0^1 f(\underline{x}|\beta,\eta)h_1(\beta)h_2(\eta)d\beta d\eta. \tag{2.5}$$

Then the posterior distribution of $\underline{\theta}$ given the data $\underline{x}$ can be written as

$$\pi(\underline{\theta}|\underline{x}) = \int_0^\infty \int_0^1 \pi(\underline{\theta}|\underline{x},\beta,\eta)\frac{f(\underline{x}|\beta,\eta)}{f(\underline{x})}h_1(\beta)h_2(\eta)d\beta d\eta. \tag{2.6}$$

In fact we will not require the precise form of $\pi(\underline{\theta}|\underline{x})$ since decisions about which supplier or subset of suppliers should be selected will be based on easily computed expectations taken with respect to this posterior. We now develop two such criteria.

C1. Posterior probability of selecting the best.

Let

$$P_j(b) = P(\theta_j > b\theta_i \text{ for all } i \neq j | \underline{x}),$$

where $b \geq 1$. It will be noted that $P_j(1)$ is just the posterior probability that θ_j is largest and the usual PP^* selection criterion of Gupta and Yang is obtained by putting in the selected subset the smallest number of suppliers for which the sum of their corresponding $P_j(1)$'s is at least P^*. We have suggested here a stronger criterion for selection purposes; one that allows the practitioner to express a quantity b, i.e. how superior does the best have to be, and a probability P^* to be attained by the selected subset. Of course for $b > 1$ it is no longer true that $\Sigma P_j(b) = 1$ and in fact it may be that for a given $b > 1$ no supplier is better than the others by amount b with sufficiently high probability. The experimenter can easily take another look and perhaps lower b or the probability requirement. In any case we believe the $P_j(b)$'s provide a useful criterion for selecting one or more suppliers and gives the experimenter the interpretation which relates to the practical problem. In particular the quantity $P_j(b)$ can be most useful in the case in which two suppliers seem to be better than the rest and a comparison of just the two is required. Here the calculation of, say $P_{12}(b)$, as a function of "b" can be very useful in deciding the extent to which supplier 1 is really better than supplier 2.

Using (2.1) and letting $A_j(b) = \{\underline{\theta} : \theta_j > b\theta_i \text{ for all } i \neq j\}$, the expression for $P_j(b)$ can be derived as follows:

(2.7) $$P_j(b) = \int_{A_j(b)} \pi(\underline{\theta} | \underline{x})\, d\theta$$

$$= \int_0^\infty \int_0^1 \left[\int_0^1 \prod_{\substack{i=1 \\ i \neq j}}^{k} G(\theta_j / b | a_i, b_i) g(\theta_j | a_j, b_j) d\theta_j \right] \frac{f(\underline{x}|\beta, \eta)}{f(\underline{x})} h_1(\beta) h_2(\eta) d\beta d\eta$$

noting that the terms in brackets are equal, a_i and b_i being defined earlier. Thus evaluation of each $P_j(b)$ requires only a three dimensional numerical integration for all choices of h_1 and h_2, provided the incomplete beta function is available.

C2. Expected number of future successes.

Another useful criterion for selection purposes is obtained by considering future observations. Suppose n future observations are to be taken from any one of the suppliers. Let the total number of successes be denoted by Y and compute $E(Y_i|\underline{x})$ for each supplier $i = 1, \ldots, k$, where E is the expectation taken with respect to the distribution of Y conditional on $\underline{x}$, i.e. the Bayesian predictive distribution. Then the supplier with largest $E(Y_i)$ is called best. Calculation for $E(Y_i|\underline{x})$ is easily obtained as

$$E\left(Y_i|\underline{x}\right) = \int_0^1 E(Y_i|\underline{x}, \theta_i) \pi(\theta_i|\underline{x}) d\theta_i = nE(\theta_i|\underline{x}).$$

Thus ranking suppliers on the basis of largest expectation of the number of successes in n future items is equivalent to ranking them on the basis of their posterior means based on the present data $\underline{x}$. An expression for $E(\theta_i|\underline{x})$ involves only a two dimensional integration and is given by:

$$(2.8) \qquad E\left(\theta_i|\underline{x}\right) = \int_0^\infty \int_0^1 \left[\frac{\beta}{1+\eta n_i} + \frac{\eta n_i}{1+\eta n_i} \left(\frac{x_i}{n_i} \right) \right] \frac{f(\underline{x}|\beta, \eta)}{f(\underline{x})} h_1(\beta) h_2(\eta) d\beta d\eta$$

using (2.4), (2.5) and noting that the mean of $g(\theta_i|a_i, b_i)$ is $a_i/(a_i + b_i)$.

Using this Criterion to rank the suppliers, various selection procedures can be defined. Here we discuss two as follows:

(i) put the i^{th} supplier in the selected subset if and only if $E(\theta_i|\underline{x}) \geq c$, or

(ii) select a subset of "r" suppliers corresponding to the "r" largest $E(\theta_i|\underline{x})$ values;

the values "c" and "r" being specified by the decision maker.

The procedure in (i) assures the decision maker that ALL of the suppliers thus selected will have the expected number of successes at least as large as "$n\ c$" whereas the procedure in (ii) assures the decision maker that the expected number of successes for that group of selected suppliers is larger than that for any other subset of "r" suppliers. Further amplification of this point will be made in Section 5. We now turn our attention to hyperpriors h_1 and h_2 and discuss how they are influenced by the various types of prior information available.

3. Prior information and elicitation for h_1 and h_2

There are two main advantages of the hierarchical structure. Firstly, it provides a realistic Bayesian model which can easily accommodate the type of prior information which is likely to be available; secondly, it provides a more precise model for what is commonly called "parametric empirical Bayes" (see Morris (1983)). In the particular application we make to supplier's data, it is clear that there is some prior information concerning the suppliers as a group, e.g. approximately where their quality is likely to be and what sort of variability amongst the θ_i's can be expected. But if this kind of information is unavailable, then it is still realistic to treat the θ_i's as exchangeable with non-informative hyperpriors. Both of these ideas are covered in the HB model. This type of prior information is to be contrasted to those Bayesian models which assume the θ_i's are independent with known but perhaps different distributions. This approach is generally quite unrealistic and therefore has limited application. On the other hand it is sometimes argued that a prior distribution on the θ_i's exists but is *unknown*. When this prior is assumed to be in some parametric family, it is then suggested that repetitions of the process may yield estimates of these parameters. Acting as though these estimates were the *true* unknown parameter values, one can then use the Bayes procedures, since the prior is then "known" and hence the expression "parametric empirical Bayes". What estimators are sensible in this context is generally answered by embedding the unknowns in a larger truly Bayesian model, hence the incorporation of hyperpriors and the expression Bayes empirical Bayes, (see Deely and Lindley (1981)). Often such truly Bayesian models yield complicated numerical problems related to the form of the posterior distribution, but recent developments of the Gibbs sampler and other related Markov chain Monte Carlo methods (see Smith and Roberts (1993)) have generally overcome these problems. In any case the particular problem treated in this paper does not involve such complications since $P_j(b)$ in (2.7) and $E(\theta_i|\underline{x})$ in (2.8) are easily computed.

We now discuss the choices for h_1 and h_2 and their relationship to the form of the prior information available.

Case 1 : Hierarchial Bayesian

It is quite possible that in some cases decision makers will have enough prior information to specify *precise* values for β and η in $\pi(\theta|\beta,\eta)$; that is, a particular beta distribution as a prior distribution for $\theta_1,\ldots,\theta_k$ can be determined. However we believe a more realistic and more frequently encountered situation involves prior information about the group of suppliers as opposed to individual suppliers. The hierarchical model facilitates this kind of prior information via the introduction of the hyperpriors. Specifically we consider how to use prior information which arises from eliciting answers from the practitioner to the following questions:

(1) Where do you expect the average of the θ_i's to lie; i.e. can you specify an interval, say (β_L, β_U), within which you are confident that the average of the θ_i's will lie?

(2) How variable do you consider the θ_i's to be; that is, can you specify a value, say θ_0 such that all of the θ_i's will be at least this large with some specified confidence.

Answering the first question will determine $h_1(\beta)$ as a member of the beta distribution whose mean is taken as the midpoint of the interval $(\beta_L, \beta_v$. This choice is influenced by convenience but it is consistent with the elicited information while also allowing a small probability that the mean of the θ_i's is outside the interval specified by the experimenter. Computation of $h_1(\beta|\underline{x})$ could be used to assess the experimenters original judgment.

We will use the answer to the second question in conjunction with the answer to the first question to determine a region in the (β, η) plane which satisfies both. This will take the form of computing an upper boundary for η , say $u(\beta)$, as a function of β. The hyperprior on η conditional on β will be taken as uniform on the interval $(0, u(\beta))$.

Case 2 : Robust Bayesian

In this situation we want to use prior information that may not be good enough to precisely specify the hyperprior distributions. In particular suppose that the elicited information about the θ_i's provides only intervals on β and η with respective probabilities. This means that we know only that the hyperpriors belong to a certain "quantile" family of distributions denoted by H. Our technique in this context is to compute maximum and minimum values of the given criterion function over H. If the difference between these two values is small, we can then confidently make a "robust" decision whereas if the difference is large we will know that the elicited information is too weak and hence contributes nothing beyond a purely noninformative analysis. The numerical technique used to perform these calculations are given in Section 4 in the context of a numerical example. Further details of this approach can be found in Deely and Johnson (1997) where the normal means problem in analyzed using this Robust Bayesian technique.

To obtain this type of information, consider the questions asked in Case 2 above. However, instead of using a specific hyperprior over the whole interval $[0, 1]$ for β, we can elicit intervals for β, say I_1, I_2 and I_3, with respective probabilities p_1, p_2, p_3. Similarly we can elicit intervals J_1 and J_2 for η with probabilities q_1 and q_2. Thus we have obtained six rectangles in the (β, η) plane in which rectangle $I_1 \times J_j$ has probability $p_i q_j$. This in turn defines the quantile family H which contains the hyperprior $h(\beta, \eta)$ and the calculations of the minimum and the maximum of the criterion function over H can proceed.

Case 3 : Non informative Hierarchical Bayesian

In the situation in which practically nothing is known a priori about the suppliers with respect to $\underline{\theta}$, one reasonable starting point is to assume that $\theta_1, \ldots, \theta_k$ are exchangeable but not independent random variables. The reason being that knowledge of the one θ_i helps in determining another. Of course any prior distribution obtained via a mixture implies exchangeability; in particular the structure given in Section 2 insures exchangeability for any choice of $h(\beta, \eta)$. Consistent with the absence of prior information is the assumption of non-informative hyperpriors. Since $0 < \beta < 1$ we can take a non-informative choice for h_1 as the uniform distribution.

For h_2 we use the fact that conditional on β the variance as given in (2.2) is bounded by $\beta(1-\beta)$ and take the hyperprior for the variance conditional on β as uniform on the interval $[0, \beta(1-\beta)]$. This then induces the hyperprior on η as

$$h_2(\eta) = \frac{1}{(\eta+1)^2}, 0 < \eta < \infty,$$

i.e. η and β are independent.

It could be suggested that a simple non-informative choice for h_2 would be $h_2(\eta) \equiv 1$. However this distribution induces an improper posterior distribution in the special case in which the n components of the data vector $\underline{x}$ are either all 0 or all n_i for all $i = 1, 2, \ldots, k$.

4. Numerical Examples

In this Section we illustrate the suggested techniques as they apply to the criteria functions. Herein we use just one hypothetical set of binomial data but analyze it for each of the three cases for prior information described in Section 3 and for the two criteria functions defined in Section 2.

Suppose then that the data given below has been observed.

X	21	19	14	12
n	23	21	18	19
sample p	0.913	0.905	0.778	0.632

For this data it is clear that Supplier 1 seems to be best, so we will compute the relevant values of the criteria functions for this supplier; namely we will compute $P_1(b)$ and $P_{14}(b)$ and $E[\theta_i|\underline{X}]$ as defined in Section 2. Clearly, there are many other computations that could be made but we will focus on just these to illustrate the main ideas.The Tables below summarize the various scenarios with Tables 1 and 2 dealing with Criterion 1 and Criterion 2 respectively.

Table 1

Probability Supplier 1 is best by an amount "b"

			b	
		1.000	**1.050**	**1.100**
	Case 1	0.510	0.220	0.070
$P_1(b)$	Case 2	0.497 (0.411)	0.294 (0.057)	0.155 (0.003)
	Case 3	0.498	0.269	0.137
	Case 1	0.960	0.870	0.740
$P_{14}(b)$	Case 2	0.973 (0.814)	0.935 (0.525)	0.855 (0.195)
	Case 3	0.989	0.978	0.946

Table 2

Predictive values and Posterior Means

	Case 1	**Case 2**	**Case 3**
$E[Y_1\|\underline{X}]$	20.7	19.1 (15.6)	20.9
$E(\theta_1\|\underline{X})$	.0900	0.908(.742)	.0910
$E[Y_2\|\underline{X}]$	18.7	19 (15.5)	19.0
$E(\theta_2\|\underline{X})$	0.892	0.903 (.739)	0.904
$E[Y_3\|\underline{X}]$	14.9	16 (12.8)	14.0
$E(\theta_3\|\underline{X})$	0.828	0.877 (.711)	0.779
$E[Y_4\|\underline{X}]$	14.4	16 (12.4)	12.1
$E(\theta_4\|\underline{X})$	0.757	0.852 (.655)	0.635

For Case 1 we imagined that the elicited answers to the two questions posed in Section 3 were: (1) an interval for β being [0.7, 0.9] with 95% confidence and from (2) the boundary for η being a line from "0" at β = 0.65 to "0.3" for β =0.95. Furthermore the answer in (1) fixed the hyperprior on β as a beta with parameters "a" =48 and "b" =12 and the hyperprior on η conditional on β was taken to be a restricted beta on the calculated interval with mean at the midpoint of the interval.

For Case 2 we used the answers above but interpreted them less rigidly and came up with the regions (as defined in section 3) and respective probabilities as: for β, intervals [0.65, 0.7] , [0.7, 0.9], [0.9, 0.95] with respective probabilities of 0.1, 0.8, 0.1; and for η, intervals [0, 0.07], [0.07, 1] with respective probabilities of 0.8 and 0.2. The noninformative Case 3 is as given in Section 3. and P_{14} (b) and $E[\theta_i|\underline{X}]$

Several features of these Tables should be noted. Firstly, by looking at the data it is clear that Supplier 1 is not very much different from Supplier 2 and hence the value of P_1 (b) is close to 0.5. But when comparing Supplier 1 to Supplier 4 a very different result obtains, namely we are quite certain that supplier 1 is better than Supplier 4 and in fact we are 74% certain (Case 1) that Supplier 1 is 10% better (b= 1.1) than Supplier 4. The noninformative analysis gives an even better picture.

These ideas are supported by the results in Table 2. Here the predictive values for Suppliers 1 and 2 are very close whereas there is considerable difference between Suppliers 1 and 4. It can also be observed from Table 2 that the hierarchical model has brought about the expected "shrinkage" effect. This can be seen vividly by examining the posterior means reported in Table 2.

It does appear from the data presented here as a hypothetical example and the imagined elicited prior information necessary for Case 2 does not indicate that very strong statements can be made in the Robust Bayesian scenario. This situation can arise however if the elicited information is not good enough when compared to the data obtained. In any practical situation, this would result in further elicitation or an acceptance that strong inferences are not possible.

5. Remarks, discussion and conclusions

(i) Test of hypothesis

There may be some situations in which a decision maker is concerned in the first instance about testing the equality of the supplier's quality, i.e. test $H_0 : \theta_1 = \ldots = \theta_k$. Whereas we feel that this is not in general the ultimate goal of the experimenter, it is quite easy to incorporate this situation into the model by simply incorporating a prior probability γ that H_0 is true (i.e. $P(H_0 \text{ is true}) = P(\eta = 0) = \gamma$) and then computing the posterior probability of H_0 which is given by:

$$\gamma^* = \left[1 + \frac{1-\gamma}{\gamma} \frac{f(\underline{x})}{f(\underline{x}|0)}\right]^{-1}$$

where $f(\underline{x})$ and $f(\underline{x})|\beta, \eta)$ are given in (2.5) and

$$f(\underline{x}|0) = \int_0^1 f(\underline{x}|\beta, 0) h_1(\beta) d\beta = \int_0^1 \prod_{i=1}^{k} \binom{n_i}{x_i} \beta^{\Sigma x_i} (1-\beta)^{N - \Sigma x_i} h_1(\beta) d\beta.$$

Then each P_j should be multiplied by $(1-\gamma^*)$ to obtain the posterior probability that θ_j is largest since p_j as given in (2.7) is conditional upon H_0 false, i.e. $\eta > 0$. One could simply compute the Bayes factor, $BF = f(\underline{x}|0)/f(\underline{x})$, as evidence for believing H_0. We point out however that the model of Deely and Zimmer (1987) seems more appropriate for testing the equality of supplier's quality.

(ii) Comparisons and possible extensions

It is clear that the HB model offers a much wider class of models than the naive Bayesian or the empirical Bayesian approaches which have been reported in the literature thus far. In the first instance, the HB model allows through the hyperpriors h_1 and h_2 the facility to use prior information about the suppliers as a group whereas the naive models have no place for such information. We believe that this prior information begins with an assumption of at least exchangeability, but more informative models are also possible as we have shown in the examples in Section 4. One could argue that some approximations of the P_j's or $E(\theta_i|\underline{x})$'s might be close enough and not require numerical integration. There has been some work in this direction (see Albert and Gupta (1985) and Leonard (1972)) but since the numerical integrations required herein are relatively easy, such approximations would appear to be unnecessary. Secondly, we point out that only a very simple hierarchical model was used in this paper. It is clear that there is scope for richer models. For example, one could replace β in (2.1) with $y_{i1}\beta_1+y_{i2}\beta_2$ where y_{i1}, y_{i2} are known "regressors" for $i = 1, \ldots, k$ and $\underline{\beta} = (\beta_1, \beta_2)$ is a vector of unknown "regression" coefficients with hyperprior $h_1(\underline{\beta})$. This model would incorporate various descriptions of changes in θ_i as well as the naive Bayesian model in which each θ_i is assumed independent with a known beta distribution possibly with different parameters. This latter case would be modeled by taking h_1 and h_2 as point distributions at (1, 1) and 1 respectively and then solving for y_{i1} and y_{i2} to obtain the given known beta parameters.

Another possible extension of the HB model would involve covering partial exchangeability particularly relevant when k is large. In this paper we have discussed analysis when k is small and have tacitly assumed all k binomial probabilities are exchangeable. It may be the case that, in a large group of suppliers, exchangeability is only tenable within subgroups and from subgroup to subgroup there may be exchangeability only in their means. Of course this fact may not be recognizable until after observing the data. The HB model should be enriched to allow the possibility of partial exchangeability being indicated by the data and then proceeding with the selection problem.

Finally it should be noted that the HB model has no difficulty with either small or variable sample sizes whereas naive empirical Bayes procedures require large sample sizes to imply their optimality properties. In addition these models cannot give practical answers to allocation of small samples amongst suppliers. In contrast the formulas for $P_j(b)$ or $E(\theta_j|\underline{x})$ developed herein can be used to generate a matrix of possibilities over a grid of varying small samples. The experimenter is then given tangible information by which a satisfactory design can be selected. There has been very little work done in this area. Recently, Yang (1988), has given sufficient conditions for $P_i(1) \leq P_j(1)$ as a function of x_i and x_j. He showed that if $x_j - x_i \geq \max(0, n_j - n_i)$ then $P_j(1) \geq P_i(1)$. Although this condition is useful, it does not completely partition the (x_i, x_j) space and in fact when $n_j - n_i$ is large there are many possibilities for x_i and x_j which do not satisfy Yang's condition. In particular the region where $(x_i/n_i) = (x_j/n_j)$ (or nearly so) does not in general satisfy this condition. Our numerical results seem to indicate that over this region the smaller sample size gives the larger $P_j(1)$; but this remains to be demonstrated theoretically.

(iii) Differences in selection criteria

It has been proposed in this paper that either the $P_j(b)$'s or the $E(\theta_j|\underline{x})$'s be used for selection purposes. Which to use will depend on the requirements of the practical situation. If a decision is to be made, say contracting with the selected suppliers for delivery of items over a period of time, then $P_j(b)$ should be used for either selecting the best or selecting the smallest subset for which the posterior probability that the largest (by amount b) θ_j is in the selected subset is at least P^*, i.e. the PP^* rule. If, however, a decision for the short term is to be made, say which machine to use for the next n items, then $E(\theta_j|\underline{x})$ is more appropriate. To select a subset using this criterion, the requirement could be either to insure that the expected number of successes is at least N^* (i.e. take $c = N^*/n$ in Section 2 (ii)) or to maximize the expected number of successes from a fixed number r of the k suppliers, $r < k$.

$$L(S_r, \underline{\theta}) = k\theta_{[k]} - \sum_{i \in S_r} \theta_i$$

where S_r ranges over subsets of size r. However the procedure which insures the expected number of successes is at least N^* has not yet been shown to be a Bayes procedure in the decision theoretic sense.

References

Albert, J.H. (1984), "Empirical Bayes estimation of a set of binomial probabilities", *Journal of Statistical Computation and Simulation*, 20, 129–144.

Albert, J.A. and Gupta, A.K. (1985), "Bayesian methods for binomial data with applications to a nonresponse problem", *Journal of the American Statistical Association*, 80, 167-174.

Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag.

Berger, J.O. and Deely, J.J. (1988), "A Bayesian approach to ranking and selection of related means with alternatives to AOV methodology", *Journal of the American Statistical Association* (in press).

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

Brooks, R.J. (1987), "Optimal allocation for Bayesian inference about an odds ratio", *Biometrika*, 74, 196–199.

Brittain, E. and Schlesselman, J.J. (1982), "Optimal allocation for the comparison of proportions", *Biometrics*, 38, 1003–1009.

Deely, J.J. (1965), Multiple decision procedures from an empirical Bayes approach", Ph.D. (Mimeo. Series No. 45), Dept. of Statistics, Purdue University, West Lafayette, Indiana.

Deely, J.J. and Gupta, S.S. (1968), "On the property of subset selection procedures," *Sankhya*, Series A, 3037–50.

Deely, J.J. and Johnson, W.O. (1997), "Normal means revisited" *Advances in Statistical Decision Theory and Applications*, edited by S. Panchepakesan and N. Balakrishman, Boston, Birkhouser, 19–30.

Deely, J.J. and Lindley, D.V. (1981), "Bayes empirical Bayes", *Journal of the American Statistical Association*, 76, 833–841.

Deely, J.J. and Smith, A.F.M. (1998), "Quantitiative requirements for comparison of institutional performance," *J. Roy Stat. Soc. A*, 161, Part 1, 5–12.

Deely, J.J. and Zimmer, W.J. (1988), "Choosing a quality supplier – a Bayesian approach", *Bayesian Statistics III*, (Eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith), In press.

Eaves, D.M. (1980), "On exchangeable priors in lot acceptance", *J. Roy. Statist. Soc.* B, 42, 88–93.

Gibbons, J.D., Olkin, I. and Sobel, M. (1977), *Selecting and Ordering Populations*, New York: Wiley.

Gupta, S.S., Huang, D.-Y. and Huang, W.-T. (1976), "On ranking and selection procedures and tests of homogeneity for binomial populations", *Essays in Probability and Statistics* (Eds. S. Ikeda, et al.), Shinko Tsusho Co. Ltd., Tokyo, Japan, pp. 501–533.

Gupta, S.S. and Liang, T. (1987), "Empirical rules for selecting the best binomial population", *Statistical Decision Theory and Related Topics – IV* (Eds. S.S. Gupta and J.O. Berger), Vol. 1, Springer-Verlag, New York, pp. 213–224.

Gupta, S.S., Liang, T. and Rau, Re-Bin, (1994), "Empirical Bayes two-stage procedures for selecting the best Bernoulli population compared with a control," *In Statistical Decision Theory and Related Topics - V*, (Eds. S.S. Gupta and J.O. Berger), Springer, 277–292.

Gutpa, S.S., Liao, Y. (1993), "Subset selection procedures for binomial models based on a class of priors and some applications," *In Multiple Comparisons, Selection, and Applications in Biometry*, Marcel Dekker, Inc., New York, 331–351.

Gupta, S.S. and McDonald, G.C. (1986), "A statistical selection approach to binomial models", *Journal of Quality Technology*, 18, 103–115.

Gupta, S.S. and Sobel, M. (1960), "Selecting a subset containing the best of several binomial parameters", *Contributions to Probability and Statistics*, (Eds. I. Olkin et al.), Stanford University Press, Stanford, California, pp. 224–248.

Gupta, S.S. and Panchapakesan, S. (1979), *Multiple Decision Procedures*, New York: Wiley.

Gupta, S.S. and Yang, H.M. (1985), "Bayes-P^* subset selection procedures for the best population", *Journal of Statistical Planning and Inference*, 12, 213–233.

Leonard, T. (1972), "Bayesian methods for binomial data", *Biometrika*, 59, 581–589.

Morris, C.N. (1983), "Parametric empirical Bayes inference: theory and applications", *Journal of the American Statistical Association*, 78, 47–65.

Robbins, H. (1956), "An empirical Bayes approach to statistics", *Proc. Third Berkeley Symp. Math. Statist. and Probability*, Vol. 1, 157–163, University of California Press.

Smith, A.F.M. and Roberts, G.O. (1993), "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods", *J. Roy. Statist. Soc.* B, 55, (3–23) (with discussion).

Sobel, M. and Huyett, M.J. (1957), "Selecting the best one of several binomial populations", *Bell System Technical Journal*, 36, 537–576.

Yang, H.M. (1987), *"On selecting the treatment with largest probability of survival"*, submitted for publication.