INVESTIGATION OF THE RATE OF CONVERGENCE IN THE TWO SAMPLE NONPARAMETRIC EMPIRICAL BAYES APPROACH TO AN ESTIMATION PROBLEM

by

Alan Then Kang Wang

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

STATISTICS

May 1965

Blacksburg, Virginia

TABLE OF CONTENTS

I INTRODUCTION

In the general estimation problem there is a parameter which one desires to estimate. One usually bases his estimate on an observation of a random variable or random variables whose distribution or distributions are completely specified up to but not including the value of the parameter to be estimated. In the Bayesian decision problem the parameter is assumed to be a random variable itself with completely specified distribution. With certain loss functions, among these the squared error loss function, the Bayes solution becomes particularly simple.

The empirical Bayesians, in particular H. Robbins {3} do not assume knowledge of the distribution of the parameter but only that it belongs to some wide class of distribution. They assume further that this problem has presented itself many times in the past and that sample observations from the past are available to them when they make their present decision. With the aid of this empirical information, the empirical Bayesians are able to obtain an expected loss which converges to the expected loss which would be obtained if the distribution of the parameter was known. The difference between their true expected loss and the minimum (Bayes) expected loss is called the regret. It is how much they can expect to lose over and above the minimum because of their lack of knowledge of the distribution of the parameter.

In {1}M.V. Johns Jr. eliminates the knowledge of the distribution of the observations thereby inventing a nonparametric empirical Bayes approach. He assumes that he has a vector of observations where the random variables in the vector are independent and identically distributed with the same parameter. He also assumes that these random variables are all unbiased estimates of the parameter. Assuming that this problem has occurred many times in the past he too is able, in a restricted sense, to obtain a regret which converges to zero.

The two sample nonparametric empirical Bayes approach was first introduced by Krutchkoff in {2}. There the estimate must be made on the basis of an observation whose distribution is completely unknown. The observation is not necessarily an unbiased estimate of the parameter nor is its expectation known (it may not even exist). In order to do this Krutchkoff assumes the existence of a secondary observation. This secondary observation can not be obtained before the estimate is to be made and therefore can not be used in making the estimate. Of this secondary or so called detailed random variable he assumes only that it be an unbiased estimate of the parameter. He does, as all empirical Bayesians, assume that the problem has occurred often and independently in the past. He also assumes that although the secondary observation of the present problem is not available, the secondary observations of the previous problems are available. With this empirical information he is able to obtain an estimate whose regret converges to zero without knowledge of any distributions at all.

This convergence of the regret to zero is called asymptotic optimality of the decision procedure. An obvious question one can ask is: How many past occurrences of this problem are actually necessary for this convergence? Clearly the answer depends on the distributions involved.

In this paper we do two things:

1) We solve a hypothetical problem by means of the two sample nonparametric empirical Bayes approach.

2) We find the regret as a function of n by assuming knowledge of all the distributions.

As a result of our efforts in 2) a modification to the two sample nonparametric empirical Bayes decision procedure suggested itself. This modification increased the rate of convergence considerably.

II A HYPOTHETICAL PROBLEM

Let us consider the following hypothetical situation. After several years of testing their atomic bombs, the United States and Russia decide to call a moratorium and promise (without inspection) to halt all atmospheric tests.

One day our instruments record an atmospheric nuclear explosion in Siberia. In order to confront the Russian delegation on the floor of the United Nations we need to know how large the explosion was and if it was in fact a nuclear explosion, we need about two weeks in order to obtain fallout samples as a function of both time and distance from Siberia. Needless to say, this will not do in this particular situation.

In order to obtain an immediate estimate of the size of the nuclear blast, we decide to use the radiation increase at our Alaskan test site one hour after explosion. Clearly this method is not a very good procedure when compared to the two week tests. If we assume that in the past when a nuclear explosion took place, we took data at our Alaskan test site as well as conducting the two week tests, then we can use the two sample nonparametric empirical Bayes approach to estimate the size of the bomb.

It should be remarked that it is not at all apparent why the underlying parameter, i.e. the size of the bomb is itself a random variable but we will not labor this point here.

The approach introduced by Krutchkoff and summarized in Chapter III does not assume any particular distributions. We will, however, in Chapter IV choose certain reasonable distributions for:

a) the size of the bomb,

b) the preliminary test in our Alaskan site for a given bomb size, and

c) the secondary two week test for a given bomb size.

With these assumed distributions we will check the rate of convergence of the regret.

III OUTLINE OF KRUTCHKOFF'S TWO SAMPLE NONPARAMETRIC EMPIRICAL BAYES APPROACH.

The following is a brief outline of the formalism and results of {2} as they pertain to our problem. The details of the general formalism and the proofs are omitted. Various simplifications of the notation have been made to facilitate ease in reading this section. In particular we have assumed that the size of the bomb is a discrete parameter, the increase in radiation in our Alaskan site is a discrete variable and the secondary two week test result is a continuous variable. It should be noted that although in {2} these variables are general and can be of any form, we have assumed them of this form since this is the form in which they will be used in Chapter IV.

Let Θ, with generic element θ, be called the parameter space. (θ represents the size of the explosion in megatons).

Let D, with generic element d, be called the estimation space. (d represents an estimate of the size of the explosion in megatons.)

Let $L(d,\theta) = (d - \theta)^2$ be the loss function defined for all $d \in D$ and $\theta \in \Theta$

This loss function, commonly called the squared error loss function, is the one most used in estimation problems in both decision theory and the classical estimation procedures.

Let $P(\theta)$ be the a priori probability on the parameter space Θ, i.e. $P(\theta)$ is the probability that the bomb is of size θ before we make any tests or observations at all.

The two sample nonparametric empirical Bayes approach does not assume knowledge of $P(\theta)$ at all except that it belong to the class of probabilities for which

$$E\theta^2 = \sum_{\theta} \theta^2 P(\theta) < \infty.$$

Let X be a random variable with conditional probability density function $P(x|\theta)$. x represents a preliminary observation (which in our example is the increase in radiation at the Alaskan site). In this approach $P(x|\theta)$ is assumed completely unknown. It is on this x that we must base our estimate of the size of the bomb.

Let Y be a random variable with the conditional density given θ of $f(y|\theta)$. y is the secondary random variable which in our example represents the two week tests. Krutchkoff's approach does not assume knowledge of $f(y|\theta)$ but only that

$$E(Y|\theta) = \int y\, f(y|\theta)dy = \theta \qquad \text{for all } \theta \in \Theta$$

and

$$EY^2 = \int y^2 \sum_{\theta} f(y|\theta)P(\theta)\ dy < \infty$$

that is to say, the results of the two week test is an unbiased estimate of the size of the bomb and the variance of the test is finite.

The approach assumes that information in the form of (x,y) are available from say n-1 previous occurrences of the problem and that for this problem we only have x. Then on the basis of the Alaskan test and past experience of Alaskan tests and two week tests, we must estimate θ with a squared error loss.

If the results of the i-th occurrence of the problem is x_i, y_i and we set z equal to the vector of observations

$$(x_1, y_1, x_2, y_2, \ldots x_{n-1}, y_{n-1})$$

then all our information can be written simply as Z, x where x is the present X value.

It has been shown by Krutchkoff in {2} that the risk in this situation is given by

(3.1) $$R(d) = E\{(d(Z,X) - E(\theta|X))^2\} + E\{Var(\theta|X)\},$$

where d(Z,x) is the decision based on the information Z, x, and $E(\theta|x)$ is the conditional expectation of θ for given X = x and is given by

(3.2) $$E(\theta|x) = \frac{\sum_\theta \theta P(x|\theta)P(\theta)}{\sum_\theta P(x|\theta)P(\theta)} .$$

$Var(\theta|x)$ is the variance of θ given X = x, and is given by

$$Var(\theta|x) = \frac{\sum_\theta \theta^2 P(x|\theta)P(\theta)}{\sum_\theta P(x|\theta)P(\theta)} - \{E(\theta|x)\}^2$$

The Bayes risk is then given by

$$(3.3) \qquad R(d^B) = E\{Var(\theta|x)\},$$

and the regret by

$$(3.4) \qquad r(d) = R(d) - R(d^B)$$
$$= E\{(d(Z,X) - E(\theta|X))^2\}.$$

We now solve the estimation problem in the following manner. Let

$$\delta(x_i x) = 1 \quad \text{if} \quad x_i = x$$
$$= 0 \quad \text{if} \quad x_i \neq x$$

and $m = \sum_{i=1}^{n-1} \delta(x_i, x).$

Now define

$$(3.5) \qquad \phi_n = \frac{1}{m} \sum_{i=1}^{n-1} \delta(x_i, x) y_i \qquad m>0$$
$$= 0 \qquad m=0 .$$

ϕ_n is called "a two sample nonparametric empirical Bayes procedure". Essentially what ϕ_n says is:

a) If the increase of radiation at the Alaskan site is x then look at all your past occurrences of the problem and pull out the data for all those with the same x. The procedure is then simply to use as your estimate of the bomb size, the average of your estimates obtained from the two week tests for these occurrences of the problem.

b) If x has occurred for the first time then use zero as your estimate. (This part is the one which causes trouble in the convergence and is the part we alter in our work.)

The regret for any n is then shown to be:

(3.6) $$r(\phi_n) = \sum_x \mathrm{Var}(Y|x) \sum_{m=1}^{n-1} \frac{1}{m}\binom{n-1}{m} P(x)^{m+1}\{1-P(x)\}^{n-m-1}$$
$$+\sum_x \{E(\theta|x)\}^2 \{1-P(x)\}^{n-1} P(x)$$

where P(x) is the marginal probability that X = x and is given by

(3.7) $$P(x) = \sum_\theta P(x|\theta)P(\theta).$$

Var(Y|x) is the variance of y given that X = x. It is given by

(3.8) $$\mathrm{Var}(Y|x) = E(Y^2|x) - \{E(y|x)\}^2$$
$$= \int y^2 \sum_\theta f(y|\theta)P(x|\theta)P(\theta) \frac{1}{P(x)}\, dy$$
$$-\{\int y \sum_\theta f(y|\theta)P(x|\theta)P(\theta) \frac{1}{P(x)}\, dy\}^2.$$

It is then proven in {2} that

$$\lim_{n\to\infty} r(\phi_n) = 0,$$

and ϕ_n is then said to be asymptotically optimal.

IV THE PARAMETRIC ASSUMPTIONS

The problem is solved by giving the estimate ϕ_n. That is we have in ϕ_n an asymptotically optimal estimate of the bomb size θ. We would like to investigate now the rate at which this convergence takes place. In order to do this we now invent some reasonable distributions and use them to obtain the regret of ϕ_n (i.e. $r(\phi_n)$) as a function of n.

It would seem reasonable that the two week test be fairly accurate and to this end we will assume that it is normally distributed with mean θ and variance 1.

We will assume that the preliminary test, the increase in radiation in our Alaskan site is a discrete variable with, as is not unusual in radiations problems, a Poisson distributed with mean θ.

Since it is doubtful that the bomb size is really a random variable and since we do not want to aid in the convergence by assuming a convenient prior distribution, we will choose the prior usually considered least informative, the uniform distribution.

Our parametric assumptions can therefore be summarized by;

A) $P(\theta) = 1/20 \quad \theta = 1,2,\ldots..20$

(We also used $P(\theta) = 1/10 \quad \theta = 1,2,\ldots..10$. We will however carry through the discussion with the first $P(\theta)$

and then simply state the changes necessary in order to use the second $P(\theta)$).

B) $P(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}$ $x = 0,1,2,\ldots\ldots$

and

C) $f(y|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\theta)^2}$ $-\infty<y<\infty$.

In order to find the regret we see from equation (3.6) that we will need

1) $P(x)$, the marginal of x,

2) $E(\theta|x)$, the expected value of θ in the posterior distribution of θ given x, and

3) $Var(Y|x)$, the variance of y in the conditional distribution of y given x.

By (3.7), A), and B) we see that

(4.1) $$P(x) = \sum_{\theta=1}^{20} P(x|\theta)P(\theta)$$

$$= \frac{1}{20} \sum_{\theta=1}^{20} \frac{e^{-\theta}\theta^x}{x!} \qquad x = 0,1,2,\ldots\ldots$$

By (3.2), A) and B) we see that

(4.2) $$E(\theta|x) = \frac{\sum_{\theta=1}^{20} P(x|\theta)P(\theta)}{\sum_{\theta=1}^{20} P(x|\theta)P(\theta)}$$

$$= \frac{\sum_{\theta=1}^{20} e^{-\theta}\theta^{x+1}}{\sum_{\theta=1}^{20} e^{-\theta}\theta^{x}} \qquad x = 0,1,2,\ldots\ldots$$

By (3.8), A), B), and C) above we see that

$$(4.3) \qquad Var(Y|x) = \int y^2 \sum_{\theta=1}^{20} f(y|\theta)P(x|\theta)P(\theta) \frac{1}{P(x)}\, dy$$

$$- \{\int y \sum_{\theta=1}^{20} f(y|\theta)P(x|\theta)P(\theta) \frac{1}{P(x)}\, dy\}^2$$

$$= \frac{1}{\sqrt{2\pi}} \int y^2 \frac{\sum_{\theta=1}^{20} e^{-\frac{1}{2}(y-\theta)^2}\theta^x}{\sum_{\theta=1}^{20} e^{-\theta}\theta^x}\, dy$$

$$- \frac{1}{\sqrt{2\pi}} \{\int y \frac{\sum_{\theta=1}^{20} e^{-\frac{1}{2}(y-\theta)^2-\theta}}{\sum_{\theta=1}^{20} e^{-\theta}\theta^x}\, dy\}^2 .$$

These formulas do not lead themselves to simple analysis and we will therefore use computer methods as described in Chapter V to solve them.

V COMPUTATIONAL METHODS

The first and simplest step was computing

$$P(x) = \frac{1}{20} \sum_{\theta=1}^{20} \frac{e^{-\theta}\theta^{x}}{x!} \qquad x = 0,1,2,\ldots\ldots$$

In order to avoid overflowing the machine on computing x! we computed the term $\frac{\theta^{x}}{x!}$ successively by storing $\frac{\theta^{0}}{0!} = 1$ and then multiplying by θ/x successively for $x = 1,2,3,\ldots\ldots$ up to as many as we needed which was only 43. This had to be done for each of the twenty values of θ.

$\sum_{x=0}^{43} P(x) = 1$ to eight decimal places. It was felt that this was sufficient accuracy.

The second step was to compute

$$E(\theta|x) = \frac{\sum_{\theta=1}^{20} e^{-\theta}\theta^{x+1}}{\sum_{\theta=1}^{20} e^{-\theta}\theta^{x}} \qquad x = 0,1,\ldots\ldots 43.$$

It turned out that it was easier to use the values obtained for P(x) and the procedure developed there to obtain

$$E(\theta|x) = \frac{1}{20P(x)} \sum_{\theta=1}^{20} e^{-\theta} \frac{\theta^{x+1}}{x!} \qquad x = 0,1,\ldots\ldots 43.$$

Actually these computations were done in two steps. First we found

$$P(\theta|x) = \frac{1}{20P(x)} \frac{e^{-\theta}\theta^{x}}{x!} \qquad x = 0,1,\ldots\ldots 43,$$

and then

$$E(\theta|x) = \sum_{\theta=1}^{20} \theta P(\theta|x) \qquad x = 0,1,\ldots\ldots 43 .$$

The most difficult part indeed was the computation of Var(y|x). From equation (4.3) we see that

$$\mathrm{Var}(Y|x) = \int y^2 \sum_{\theta=1}^{20} \frac{1}{P(x)} f(y|\theta)P(x|\theta)P(\theta)dy$$
$$-\{\int y \sum_{\theta=1}^{20} \frac{1}{P(x)} f(y|\theta)P(x|\theta)P(\theta)dy\}^2 .$$

In order to avoid large numbers and because the Var(Y|x) term only occurs where it is multiplied by $P^2(x)$ we decided to calculate

$$\mathrm{Var}(Y|x)P^2(x) = P(x)\int y^2 \sum_{\theta=1}^{20} f(y|\theta)P(x|\theta)P(\theta)dy$$
$$-\{\int y\sum_{\theta=1}^{20} f(y|\theta)P(x|\theta)P(\theta)dy\}^2$$
$$= P(x) \frac{1}{20\sqrt{2\pi}} \int y^2 \sum_{\theta=1}^{20} e^{-\frac{1}{2}(y-\theta)^2-\theta} \frac{\theta^x}{x!} dy$$
$$- \frac{1}{400(2\pi)} \{\int y \sum_{\theta=1}^{20} \frac{\theta^x}{x!} e^{-\frac{1}{2}(y-\theta)^2-\theta} dy\}^2 .$$

The next step is to interchange the order of summation and integration giving

$$(5.1) \qquad \mathrm{Var}(Y|x)P^2(x) = \frac{P(x)}{20\sqrt{2\pi}} \sum_{\theta=1}^{20} \frac{e^{-\theta}\theta^x}{x!} \int y^2 e^{-\frac{1}{2}(y-\theta)^2} dy$$
$$- \frac{1}{400(2\pi)} \{\sum_{\theta=1}^{20} \frac{e^{-\theta}\theta^x}{x!} \int y e^{-\frac{1}{2}(y-\theta)^2} dy\}^2 .$$

We now perform the integration by the rectangular method. By well known properties of the normal distribution we see that the integration need be done only for

$\theta - 5 < y < \theta + 5$.

We decided to use widths of 0.1 for our rectangles since a preliminary calculation indicated that this would be sufficient for our purposes. The y ordinate was chosen at the center of the rectangle leading to the formula

(5.2) $$\mathrm{Var}(Y|x)P^2(x) = \frac{P(x)}{2\sqrt{2\pi}} \sum_{\theta=1}^{20} \frac{e^{-\theta}\theta^x}{x!} \sum_{y=\theta-5+.05}^{y=\theta+5+.05} y^2 e^{-\frac{1}{2}(y-\theta)^2} - \frac{1}{4(2\pi)} \left\{\sum_{\theta=1}^{20} \frac{e^{-\theta}\theta^x}{x!} \sum_{y=\theta-5+.05}^{y=\theta+5+.05} y e^{-\frac{1}{2}(y-\theta)^2}\right\}^2$$

These values could now be computed for $x = 0,1,.....43$.

We then had left the rest of the binomial terms, in particular

(5.3) $$B_n(x) = \sum_{m=1}^{n-1} \frac{1}{m} \binom{n-1}{m} P(x)^{m-1}\{1 - P(x)\}^{n-m-1}$$

The only problem here was the factorial terms which we obtained as before by computing the ratios sequentially. The exact procedure can be found in the appendix.

With all these computations made, we could obtain the regret in using ϕ_n by

(5.4) $$r(\phi_n) = \sum_{x=0}^{43} \mathrm{Var}(Y|x)P^2(x)B_n(x) + \sum_{x=0}^{43} \{E(\theta|x)\}^2\{1 - P(x)\}^{n-1}P(x).$$

VI PRELIMINARY RESULTS AND DISCUSSION

Table I gives the results of the computations given thus far. The column marked F.T. is the first term of equation (3.6) i.e.

$$F.T. = \sum_x Var(Y|x) \sum_{m=1}^{n-1} \frac{1}{m}\binom{n-1}{m} P(x)^{m+1}\{1 - P(x)\}^{n-m-1},$$

and is the contribution when the present x has been observed before.

The column marked S.T. is the second term of equation (3.6) i.e.

$$S.T. = \sum_x \{E(\theta|x)\}^2\{1 - P(x)\}^{n-1}P(x),$$

and is the contribution to the regret when x has not been observed before.

$R(d^B)$ is the minimum Bayes risk and is computed by

$$(6.1) \quad R(d^B) = E\{Var(\theta|x)\}$$

$$= \sum_x \{\frac{1}{20} \sum_{\theta=1}^{20} \theta^2 e^{-\theta} \frac{\theta^x}{x!} \frac{1}{P(x)} - \{E(\theta|x)\}^2\}P(x)$$

$$= \sum_x \{\frac{1}{20} \sum_{\theta=1}^{20} \theta^2 e^{-\theta} \frac{\theta^x}{x!}\} - \sum_x \{E(\theta|x)\}^2 P(x).$$

Table I

n	F.T.	S.T.	$r(\emptyset_n)$	$R(d^B)$
10	0.005	98.495	98.501	6.833
50	0.008	29.451	29.459	6.833
100	0.004	11.485	11.489	6.833
200	0.001	4.675	4.676	6.833
300	0.000	2.890	2.890	6.833

It is clear from Table I that the first term or what one does when x has been observed in the past converges rapidly. The second term, however, what to do when x occurs for the first time, converges very slowly. In order to see if the number of θ values would make a great difference in this result, we decided to use

$P(\theta) = 1/10 \qquad \theta = 1,2,.....10.$

The computations are similar for this prior except that we needed only $x = 0, 1, 2,.....31$ rather than $x = 0,1,.....43$.

Table II shows the results using this new prior distribution with only 10 values of θ.

Table II

n	F.T.	S.T.	$r(\phi_n)$	$R(d^B)$
10	0.013	19.857	19.870	2.891
50	0.008	3.901	3.909	2.891
100	0.002	1.669	1.671	2.891
200	0.000	0.754	0.755	2.891
300	0.000	0.480	0.480	2.891

The first term converged rapidly again as expected. The second term converged somewhat more rapidly than before. This is deceiving however since the minimum risk also became smaller. As a matter of fact, the increase in rate of convergence is very small indeed considering that we halved the possible number of values for θ.

It became apparent that it was the decision function which had to be altered in order to increase the rate of convergence. The part when x occurred more than once was fine, we had to change what to do when x occurred for the first time.

Several procedures suggest themselves. Of these we enumerate:

1) use as our estimate the average of the previous y values,

2) guess at a value using your intuition,

3) choose the midrange of θ values.

The first suggestion is very likely the best. It does however require involved mathematical calculations.

The second suggestion at first seems arbitrary but it is no more arbitrary than always choosing zero as our estimate which is what is done in ϕ_n. At least your guess would be in the range of θ values whereas zero is not.

The third suggestion seemed more appropriate for our problem. The numerical changes involved are very few and simple to calculate. Using this, the new decision function or estimate became

$$(6.2) \qquad \phi_n^* = \frac{1}{m} \sum_{i=1}^{n-1} \delta(x_i, x) y_i \qquad \text{if} \qquad m>0$$

$$= \text{mid range of } \theta \qquad \text{if} \qquad m=0 .$$

The new regret is

$$(6.3) \qquad r(\phi_n^*) = \sum_x \text{Var}(Y|x) \sum_{m=1}^{n-1} \frac{1}{m} \binom{n-1}{m} P(x)^{m+1} \{1-P(x)\}^{n-m-1}$$

$$+ \sum_x \{\text{mid range of } \theta - E(\theta|x)\}^2 P(x) \{1-P(x)\}^{n-1}$$

The first term remains the same and the new second term is very simple to calculate.

For $P(\theta) = 1/20 \qquad \theta = 1,2,\ldots\ldots 20$

The mid range of $\theta = 10.5$

for $P(\theta) = 1/10 \qquad \theta = 1,2,\ldots\ldots 10$

the mid range of $\theta = 5.5$.

The results for ϕ_n^* are recorded in Table Ia and IIa.

Table Ia

n	F.T.	S.T.	$r(\phi_n^*)$	$R(d^B)$
10	0.005	17.161	17.167	6.833
50	0.008	5.022	5.029	6.833
100	0.004	2.027	2.031	6.833
200	0.001	0.878	0.879	6.833
300	0.000	0.556	0.556	6.833

Table IIa

n	F.T.	S.T.	$r(\phi_n^*)$	$R(d^B)$
10	0.013	2.480	2.493	2.891
50	0.008	0.559	0.567	2.891
100	0.002	0.265	0.267	2.891
200	0.000	0.126	0.127	2.891
300	0.000	0.082	0.082	2.891

From these tables we see that without reducing the number of values of θ (or the minimum risk) we can increase the rate at which the second term converges.

VII CONCLUSIONS

We found a two sample nonparametric empirical Bayes estimate of the bomb size in our hypothetical example. We investigated the rate of convergence of the risk when using this estimate by assuming certain reasonable parametric distributions. The convergence was much slower than desirable.

In order to increase the rate of convergence, we altered the estimate in that we used the mid range of the possible parameter values as an estimate of the parameter when we had insufficient information for a sensible estimate by the first procedure. The rate of convergence was increased by this method. It would seem reasonable however that other alterations of the procedure could increase this rate still further.

VIII ACKNOWLEDGEMENTS

The author wishes to extend sincere gratitude to Professor W.L. Johnson and Dr. Raymond Myers for reading the manuscripts.

He wishes to thank Dr. Richard G. Krutchkoff, his advisor, for suggesting this problem and for his advice and guidance throughout this project. Thanks are also extended to Dr. B. Harshbarger and other members of his staff for their interest and help.

Special thanks are extended to for his help in the computer programming.

He wishes to express his great appreciation to for her continuous encouragement, and to his parents for their assistance and financial help.

IX BIBLIOGRAPHY

{1} Johns Jr., M.V. (1957) Nonparametric Empirical Bayes Procedures. The Annals of Mathematical Statistics Vol. 28, pp. 649-669.

{2} Krutchkoff, R.G. (1964) A Two Sample Nonparametric Empirical Bayes Approach to Some Problems in Decision Theory. Ph.D. thesis. Columbia University, New York, N.Y.

{3} Robbins, H. (1955) An Empirical Bayes Approach to Statistics. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.

The vita has been removed from the scanned document

APPENDIX 1

```
$JOB            717540005 A WANG 3-19-65 REGRET ANALYSIS
$IBJOB
$IBFTC D1
C      THESIS EMPERICAL BAYES DECISION THEORY
       EXTERNAL MOLMUN
       DIMENSION A(44),B(20),C(290),PAGB(43,20),PX(43),
      1DCATUR(43),EBGA(43),PBGA(20,43),SMARK(43),AJOINT(43,20)
       IRENE=1
   904 READ(5,905) M1, M2, M3, PBINT, PB, AVEB
   905 FORMAT(1X, 3I3, F4.3, F3.2, F4.1)
       WRITE(6,33)
    33 FORMAT(1X,3HX+1,4X11HP**2*V(Y/X),6X8HBINOMIAL,
      17X8HPRODUCTS)
       M1P1=M1+1
       A(2)=1.
       B(1)=1.
       DO 20 I=3,M1P1
    20 A(I)=A(I-1)+1.
       DO 21 J=2,M2
    21 B(J)=B(J-1)+1.
       A(1)=0.
       DO 14 J=1,M2
    14 PAGB(1,J)=B(J)*EXP(-B(J))
       DO 5 J=1,M2
```

```
      DO 6 I=2,M1
    6 PAGB(I,J)=PAGB(I-1,J)*B(J)/A(I+1)
    5 CONTINUE
C     FIND EBGA(I)
      DO 15 I=1,M1
      SUM=0.
      DO 16 J=1,M2
   16 SUM=SUM+PAGB(I,J)
      DO 17 J=1,M2
   17 PBGA(J,I)=PAGB(I,J)/SUM
      DEL=0.
      DO 18 J=1,M2
   18 DEL=DEL+B(J)*PBGA(J,I)
      EBGA(I)=DEL
   15 CONTINUE
      DO 9 I=1,M1
      SUM=0.
      DO 10 J=1,M2
   10 SUM=SUM+PAGB(I,J)
    9 PX(I)=SUM*PB
      C(1)=-4.
      DO 22 K=2,M3
   22 C(K)=C(K-1)+.1
      DO 3 I=1,M1
      W=0.
```

```
      Z=0.
      DO 2 K=1,M3
      T=0.
      DO 1 J=1,M2
    1 T=T+1./SQRT(6.28318)*EXP(-((C(K)-B(J))**2/2.))
     1*PAGB(I,J)*PBINT
      V=C(K)*T
      W=W+V*C(K)*PX(I)
      Z=Z+V
    2 CONTINUE
      DELFIN=W-Z**2
    3 DCATUR(I)=PX(I)**2*DELFIN
  200 READ(5,500) N
  500 FORMAT(1X, I3)
      DO 300 I=1,M1
      LILIAN=N-1
      LILIN=N-2
      ALLEN=LILIAN
      SMARK(I)=ALLEN*(1.-PX(I))**LILIN
      DO 100 L=2,LILIAN
      LILIAN=LILIAN*(N-L)/L
      XC=LILIAN
      XL=L
      XD=XC/XL
      SMARK(I)=SMARK(I)+XD*PX(I)**(L-1)*(1.-PX(I))**(N-L-1)
```

```
  100 CONTINUE
      C(I)=DCATUR(I)*SMARK(I)
      WRITE(6,7) I,DCATUR(I),SMARK(I),C(I)
    7 FORMAT(1X,I3,3F15.8)
  300 CONTINUE
      ALAN=0.
      DO 4 I=1,M1
    4 ALAN=ALAN+C(I)
      WRITE(6,55) ALAN
   55 FORMAT(1X,9H1ST TERM=F12.8)
      DANLIN=0.
      DO 19 I=1,M1
   19 DANLIN=DANLIN+(AVEB-EBGA(I))**2*PX(I)*(1.-PX(I))**(N-1)
      WRITE(6,23)DANLIN
   23 FORMAT(1X,9H2ND TERM=F12.8)
      REGRET=ALAN+DANLIN
      WRITE(6,24) N,REGRET
   24 FORMAT(1X,9HREGRET(N=I3,2H)=F12.8//)
      IF(N-300)200,444,444
C     FIND MINIMUM BAYES RISK
  444 DO 134 I=1,M1
      DO 135 J=1,M2
  135 AJOINT(I,J)=PAGB(I,J)*PB
  134 CONTINUE
      SUE=0.
```

```
      DO 36 I=1,M1
      DO 37 J=1,M2
   37 SUE=SUE+(B(J)-EBGA(I))**2*AJOINT(I,J)
   36 CONTINUE
      WRITE (6,38) SUE
   38 FORMAT(1X,13HMINIMUM RISK=F12.8)
      IRENE=IRENE+1
      IF(IRENE-2)904,904,907
  907 STOP
      END
$ENTRY D1
  43 20290.005.0510.5
  10
  50
 100
 200
 300
  31 10190.010.10 5.5
  10
  50
 100
 200
 300
$IBSYS
```

APPENDIX 2

```
$JOB            717540005 A WANG 4-26-65  REGRET ANALYSIS
$IBJOB          NODECK
$IBFTC D1
      EXTERNAL NOLMUN
      DIMENSION A(44),B(20),PAGB(43,20),PX(43),
     1PBGA(20,43),EBGA(43)
      WRITE (6,333)
  333 FORMAT(2X,17HN      SECOND TERM)
      LILIAN=1
  111 READ(5,1) PB,M1,M2
    1 FORMAT(1X,F5.2,2I3)
      M1P1=M1+1
      A(2)=1.
      B(1)=1.
      DO 3 I=3,M1P1
    3 A(I)=A(I-1)+1.
      DO 4 J=2,M2
    4 B(J)=B(J-1)+1.
      A(1)=0.
      DO 5 J=1,M2
    5 PAGB(1,J)=B(J)*EXP(-B(J))
      DO 7 J=1,M2
      DO 6 I=2,M1
    6 PAGB(I,J)=PAGB(I-1,J)*B(J)/A(I+1)
```

```
    7 CONTINUE
      DO 11 I=1,M1
      SUM=0.
      DO 8 J=1,M2
    8 SUM=SUM+PAGB(I,J)
      PX(I)=SUM*PB
      DO 9 J=1,M2
    9 PBGA(J,I)=PAGB(I,J)/SUM
      DTUR=0.
      DO 10 J=1,M2
   10 DTUR=DTUR+B(J)*PBGA(J,I)
      EBGA(I)=DTUR
   11 CONTINUE
  222 READ(5,12) N
   12 FORMAT(1X,I3)
      SECOND=0.
      DO 13 I=1,M1
   13 SECOND=SECOND+((EBGA(I))**2)*(1.-PX(I))**(N-1)*PX(I)
      WRITE(6,14) N,SECOND
   14 FORMAT(1X,I3,F15.8)
      READ (5,114) FIRST
  114 FORMAT(1X,F12.8)
      RABBIT=FIRST+SECOND
      WRITE(6,15) N,RABBIT
   15 FORMAT(1X,9HREGRET(N=I3,2H)=F12.8//)
```

```
      IF (N-300) 222,20,20
   20 LILIAN=LILIAN+1
      IF (LILIAN-2) 111,111,21
   21 STOP
      END
$ENTRY D1
   .05.43 20
  10
    .00519065
  50
    .00769250
 100
    .00428472
 200
    .00051099
 300
    .00005007
   .10 31 10
  10
    .01343349
  50
    .00819894
 100
    .00186210
 200
```

.00004350

300

.00000796

$IBSYS

ABSTRACT

In this thesis we consider the following. We choose the random variable θ, which has some fixed but unknown distribution with a finite second moment. We observe the value x, of a preliminary random variable X, which has an unknown distribution which is conditional on θ. Using x and our past experience we are asked to estimate the value of θ with a squared error loss function. After we have made our decision we are given the value, y, of a detailed random variable Y, which has an unknown distribution conditional on θ. The random variable X and Y are assumed independent given a particular θ. Our past experience is made up of the values of preliminary and detailed random variables from previous decision problems which are independent of but similar to the present one.

With the risk defined in the usual way the Bayes decision function is the expected value of θ given that $X = x$. Since the distributions are unknown, the use of the two sample nonparametric empirical Bayes decision function is proposed. With the regret defined in the usual way it can be shown that the two sample nonparametric empirical Bayes decision function is asymptotically optimal, i.e. for a large number of past decision problems, the regret in using the two sample nonparametric empirical

Bayes decision function tends to zero, and it is the main purpose of this thesis to verify this property by using a hypothetical numerical example.