

**An Empirical Bayes Approach to Estimating
the Probability of Correct Selection**

BU-842-M

Revised October, 1984

by

Charles E. McCulloch and Aimee Dechter

AN EMPIRICAL BAYES APPROACH TO
ESTIMATING THE PROBABILITY OF
CORRECT SELECTION

Charles E. McCulloch and Aimee Dechter

Cornell University and Census Bureau
Ithaca, N.Y. Washington, D.C.

Key Words and Phrases: ranking and selection; shrinkage estimators.

ABSTRACT

In the problem of selecting the best of k populations, Olkin, Sobel, and Tong (1976) have introduced the idea of estimating the probability of correct selection. In an attempt to improve on their estimator we consider an empirical Bayes approach. We compare the two estimators via analytic results and a simulation study.

1. INTRODUCTION

In the problem of selecting the best of k populations, Olkin, Sobel and Tong (1976) have introduced (see also Gibbons, Olkin and Sobel, 1977) the idea of *a posteriori* analysis of the data. They suggested forming an estimator of the probability of correct selection ($P\{CS\}$). However, this estimator was found to possess some shortcomings (Faltin and McCulloch, 1983). Specifically, when the means are close together, their estimator greatly over-

estimates $P\{CS\}$ and when the means are far apart, it tends to underestimate $P\{CS\}$.

In an attempt to improve on the Olkin-Sobel-Tong estimator, we consider an empirical Bayes approach to estimation of the $P\{CS\}$. The motivation for this approach was to use shrinkage estimators to improve performance, especially when the true means are close together. Our estimator is constructed in Section 2 and its properties are evaluated in Section 3. As well as information on this new estimator, the simulation results in Section 3 give additional information as to the performance of the Olkin-Sobel-Tong estimator.

2. AN EMPIRICAL BAYES APPROACH

Let X_{ij} , $1 \leq i \leq k$, $1 \leq j \leq n$ be independent observations from k populations with cdf's which are normal with mean μ_i and variance σ^2 (σ^2 assumed known) and denote the sample means by $\bar{X}_i = \sum_{j=1}^n X_{ij}/n$. Let $\mu_{[i]}$ denote the ranked parameter values $\mu_{[1]} \leq \mu_{[2]} \leq \dots \leq \mu_{[k]}$ and let $\bar{X}_{[i]}$ denote the ranked sample means, $\bar{X}_{[1]} \leq \bar{X}_{[2]} \leq \dots \leq \bar{X}_{[k]}$. Also, let $\bar{X}_{(i)}$ be defined as the sample mean from the population with mean $\mu_{[i]}$ ($1 \leq i \leq k$). It is assumed that there is no *a priori* knowledge of the pairings of the \bar{X}_i and the $\mu_{[j]}$, ($1 \leq i \leq k$, $1 \leq j \leq k$).

We wish to obtain an estimator of the probability of correct selection ($P\{CS\}$) which is given by the formula (Bechhofer, 1954)

$$P\{CS\} = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi(y + n^{1/2} \delta_i / \sigma) d\Phi(y), \quad (2.1)$$

where $\Phi(\cdot)$ = standard normal c.d.f.

and $\delta_i = \mu_{[k]} - \mu_{[i]}$.

The estimator proposed by Olkin, Sobel and Tong (1977) for the normal means case is given by the formula

$$\hat{P}\{CS\} = \hat{P} = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi(y + n^{1/2} \hat{\delta}_i / \sigma) d\Phi(y), \quad (2.2)$$

where

$$\hat{\delta}_i = \bar{X}_{[k]} - \bar{X}_{[1]} .$$

In other words, the $\mu_{[i]}$ in (2.1) are replaced by the ranked sample means $\bar{X}_{[i]}$ to obtain (2.2)

In an attempt to improve on (2.2), a different estimator for δ_i was used. Since \hat{P} tends to overestimate $P\{CS\}$ when the μ_i are close together, a natural thing to try is a shrinkage estimator. This will pull the sample means together, especially when the true means are nearly equal, thus reducing the average value of the estimator. An empirical Bayes estimator of μ_i in the normal means case is given by

$$\tilde{\mu}_i = \frac{(k-3)\sigma^2/n}{\sum_{i=1}^k (\bar{X}_i - \bar{X}_.)^2} \bar{X}_i + \left(1 - \frac{(k-3)\sigma^2/n}{\sum_{i=1}^k (\bar{X}_i - \bar{X}_.)^2}\right) \bar{X}_1, \quad (2.3)$$

where

$$\bar{X}_. = \sum_{i=1}^k \bar{X}_i / k .$$

The estimator in (2.3) can be motivated by considering a Bayesian framework in which $(k-3)\sigma^2/n \sum_{i=1}^k (\bar{X}_i - \bar{X}_.)^2$ is used as an unbiased estimator of $\frac{\sigma^2/n}{\sigma^2/n + \tau^2}$ ($k > 3$) where $\tau^2 = \text{var}(\mu_i)$ (see Casella, 1982). Using the $\tilde{\mu}_i$, we can form estimators of the δ_i from $\tilde{\mu}_{[k]} - \tilde{\mu}_{[1]}$, where $\tilde{\mu}_{[i]}$ represents the ranked $\tilde{\mu}_i$, $\tilde{\mu}_{[1]} \leq \tilde{\mu}_{[2]} \leq \dots \leq \tilde{\mu}_{[k]}$. This simplifies to

$$\tilde{\mu}_{[k]} - \tilde{\mu}_{[1]} = \left(1 - \frac{(k-3)\sigma^2/n}{\sum_{i=1}^k (\bar{X}_i - \bar{X}_.)^2}\right) \hat{\delta}_i .$$

Often, the shrinkage factor is not allowed to be negative and the positive part (denoted by a +) is taken. To obtain a positive part estimator that also works for $k = 2$ and 3 , define

$$\tilde{\delta}_i = \begin{cases} \left(1 - \frac{(k-3)\sigma^2/n}{\sum_{i=1}^k (\bar{X}_i - \bar{X}_.)^2}\right)^+ \hat{\delta}_i & k > 3 \\ \left(1 - \frac{\sigma^2/n}{\sum_{i=1}^k (\bar{X}_i - \bar{X}_.)^2}\right)^+ \hat{\delta}_i & k = 2, 3 \end{cases} \quad (2.4)$$

We can then form an estimator of $P\{CS\}$ as

$$\tilde{P}\{CS\} = \tilde{P} = \int_{-\infty}^{\infty} \prod_{i=1}^{k-1} \Phi(y + n^{1/2} \tilde{\delta}_i / \sigma) d\Phi(y) .$$

3. PROPERTIES OF THE ESTIMATORS

Analytic Results The positive part estimator defined in (2.4) has intuitive appeal for estimating the $P\{CS\}$. When the sample means are close enough together $\left(1 - (k-3)\sigma^2/n \sum_{i=1}^k (\bar{X}_i - \bar{X}_.)^2 < 0\right)$, the shrinkage factor is zero and all the $\tilde{\delta}_i$ are zero. This yields a value of $1/k$ for \tilde{P} . This could be interpreted as saying that the sample means are not spread out enough to conclude that the μ_i are different from each other and therefore the best estimate of the $P\{CS\}$ is $1/k$. This partially alleviates the problem that when the μ_i are equal, the $\bar{X}_{[i]}$ are order statistics from a $N(\mu, \sigma^2/n)$ distribution and lead to overestimates of $P\{CS\}$.

If the means are actually equal, then $P\{\tilde{P} = 1/k\} = P\{1 < (k-3)\sigma^2/n \sum_{i=1}^k (\bar{X}_i - \bar{X}_.)^2\} = P\{\chi_{k-1}^2 < k-3\}$ for $k > 3$. The values are given in Table I for various values of k .

TABLE I
Probability of $\tilde{P} = 1/k$ in the Equal Means Configuration

k	$P\{\tilde{P} = 1/k\}$
2	.304
3	.199
4	.091
5	.151
6	.192
7	.221
8	.243
9	.261
10	.275
15	.321

The probabilities $P\{\tilde{P} = 1/k\}$ slowly increase to .5 as $k \rightarrow \infty$.

For $k = 2$, \tilde{P} can be simplified as follows:

$$\begin{aligned} \tilde{P} &= \phi \left(\left[\frac{n}{2} \right]^{\frac{1}{2}} \frac{\tilde{\delta}_1}{\sigma} \right) \\ &= \phi \left(\left[\frac{n}{2} \right]^{\frac{1}{2}} \left(1 - \frac{\sigma^2/n}{\sum_{i=1}^2 (\bar{X}_1 - \bar{X}_i)^2} \right)^+ \frac{|\bar{X}_1 - \bar{X}_2|}{\sigma} \right) \\ &= \phi \left(\left[\frac{n}{2\sigma^2} \right]^{\frac{1}{2}} \left[|\bar{X}_1 - \bar{X}_2| - \frac{2\sigma^2/n}{|\bar{X}_1 - \bar{X}_2|} \right]^+ \right) . \end{aligned}$$

Thus $E[\tilde{P}] = E \left[\phi \left(\left[|Z| - \frac{1}{|Z|} \right]^+ \right) \right]$, where $Z \sim N(\sqrt{n/2}\delta_1/\sigma, 1)$.

This can be written as

$$E[\tilde{P}] = \int_{-\infty}^{\infty} \phi \left(\left[|y + n^{\frac{1}{2}}\delta_1/\sigma| - \frac{1}{|y + n^{\frac{1}{2}}\delta_1/\sigma|} \right]^+ \right) d\phi(y) . \quad (3.1)$$

Via numerical integration, using (3.1), we evaluated $E[\tilde{P}]$ for $n^{\frac{1}{2}}\delta_1/\sigma = 0, 1, 2$ and 3 . The bias of \tilde{P} and \hat{P} are given in Table 2.

TABLE 2
Biases of \tilde{P} and \hat{P}

$n^{\frac{1}{2}}\delta_1/\sigma$	Bias	
	$E[\tilde{P}] - P\{CS\}$	$E[\hat{P}] - P\{CS\}$
0	.08	.25
1	-.13	-.03
2	-.18	-.06
3	-.13	-.15

Simulations For the normal means case, for $k = 2, 4, 10$ and 15 and for various parameter configurations (given in Table 3) the performance of \tilde{P} and \hat{P} was simulated. Two types of parameter configurations were simulated, a slippage configuration:

$\mu_{[1]} = \mu_{[2]} = \dots = \mu_{[k-1]} = \mu_{[k]} - \delta$ and an equally spaced configuration $\mu_{[i]} - \mu_{[i-1]} = \Delta$ ($i=k, k-1, \dots, 2$). For all cases σ^2/n was set equal to 1. For details of the computational techniques see Section 5.

TABLE 3
Parameter Configurations Simulated

Number of Populations, k	Configurations
2	Slippage: $\delta = 0, 1, 2, 3$
4	Slippage: $\delta = 0(.5)5$ Equally spaced: $\Delta = .5, 1$
10	Slippage: $\delta = 0(.5)5$ Equally spaced: $\Delta = .5, 1$
15	Slippage: $\delta = 0(.5)5$ Equally spaced: $\Delta = .5, 1$

From the simulation, estimates were derived for the expected values and mean squared errors of \tilde{P} and \hat{P} . Figures 1 through 6 are plots of the bias and root mean square error as a function of the parameter values for $k = 4, 10$ and 15 .

4. DISCUSSION

As expected, \tilde{P} performed better than \hat{P} both in terms of root mean square error and bias when the means were close together. In addition, for $k = 2$ and 4 , considering overall performance, \tilde{P} performed about as well as \hat{P} . Unfortunately, for cases when the means were moderately far apart \tilde{P} performed poorly. This was especially so for $k = 10$ and 15 . Thus \tilde{P} cannot be recommended over \hat{P} in all circumstances.

This poor performance of \tilde{P} for large k was somewhat unexpected. Investigation of the histograms of \tilde{P} from the simulations showed a large incidence of values equal to $1/k$, even when the means were not equal. For example, in the slippage configuration for $k = 10$, the proportion of estimates, \tilde{P} , between .1 and .125 was .35 ($\delta=0$), .326 ($\delta=1$), .142 ($\delta=2$), .026 ($\delta=3$) and .004 ($\delta=4$). This could perhaps have been predicted from Table I, which shows the very high incidence of $\tilde{P} = 1/k$ for large k when

the means are exactly equal. This suggests that the shrinkage factor is too severe, especially for large k . Modification of the shrinkage factor may hold out hope for a way to improve the estimator.

5. Computational Details

Many of the simulations were run on both a mainframe (Cornell's IBM 3081D) and on an IBM personal computer for comparison's sake. On the 3081D all computations were written in double precision FORTRAN and used the IMSL routines DCADRE (for numerical integration) and GGNML (for normal, pseudo-random number generation). The simulations were replicated 1000 times each for the parameter configurations: slippage $\delta = 0, 1, 2, 3$ and equally spaced $\Delta = .5$ for $k = 2, 4, 10$, and 15. Common random numbers were used to compare \hat{P} and \tilde{P} .

On the IBM personal computer, all computations were written in double precision BASIC and used a Romberg extrapolation technique for numerical integration (adapted from Forsythe and Brown, 1977). To generate normal, pseudo-random numbers the built-in BASIC uniform random number generator and a technique due to Marsaglia and Bray (1964) was used. The simulations were replicated at least 500 times for the parameter configurations: slippage $\delta = 0(.5)5$ and $\Delta = .5, 1$ for $k = 4, 10, 15$. Common random numbers were used to compare \hat{P} and \tilde{P} .

In Table 2 some comparisons between the two machines are made. It also gives an idea of the precision of estimates.

TABLE 4

Comparison of some Simulation Results
for $k=4$ Populations for Estimating $E[\tilde{P}]$.

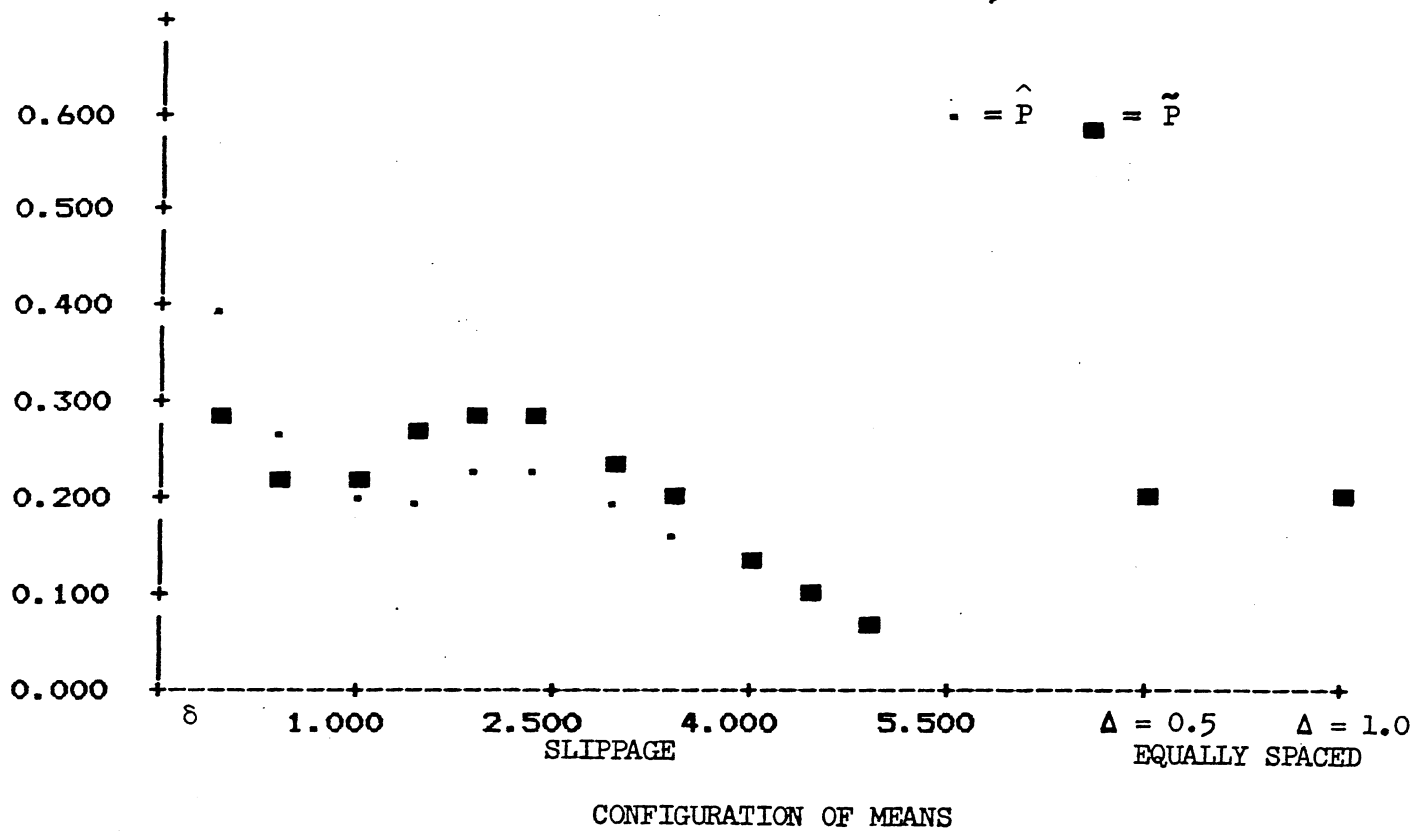
Parameter Configuration	3081D			Personal Computer		
	Average	SE	Reps	Average	SE	Reps
slippage $\delta = 0$.459	.006	1000	.461	.006	1000
slippage $\delta = 1$.513	.006	1000	.505	.006	1000
slippage $\delta = 2$.649	.007	1000	.648	.007	1000
slippage $\delta = 3$.810	.006	1000	.814	.006	1000
equally spaced $\Delta = .5$.529	.006	1000	.523	.008	500

BIBLIOGRAPHY

- Bechhofer, R.E. (1954). A single sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.*, 25: 16-39.
- Casella, G. (1982). An introduction to empirical Bayes data analysis. BU-787-M in the Biometrics Unit Mimeo Series, Cornell University, Ithaca, New York.
- Faltin, F.W. and McCulloch, C.E. (1983). On the small sample properties of the Olkin-Sobel-Tong estimator of the probability of correct selection. *J. Amer. Statist. Assoc.*, 78: 464-7.
- Forsythe, G.E. and Brown, M.B. (1977). *Computer Methods for Mathematical Computations*. Wiley, New York.
- Gibbons, J.D., Olkin, I., and Sobel, M. (1977). *Selecting and Ordering Populations*. Wiley, New York.
- Marsaglia, G., and Bray, T.A. (1964). A convenient method for generating normal variables. *SIAM Review*, 6: 260-4.
- Olkin, I., Sobel, M., and Tong, Y.L. (1976). Estimating the true probability of a correct selection for location and scale parameter families. Technical Report 110, Stanford University, Department of Statistics.

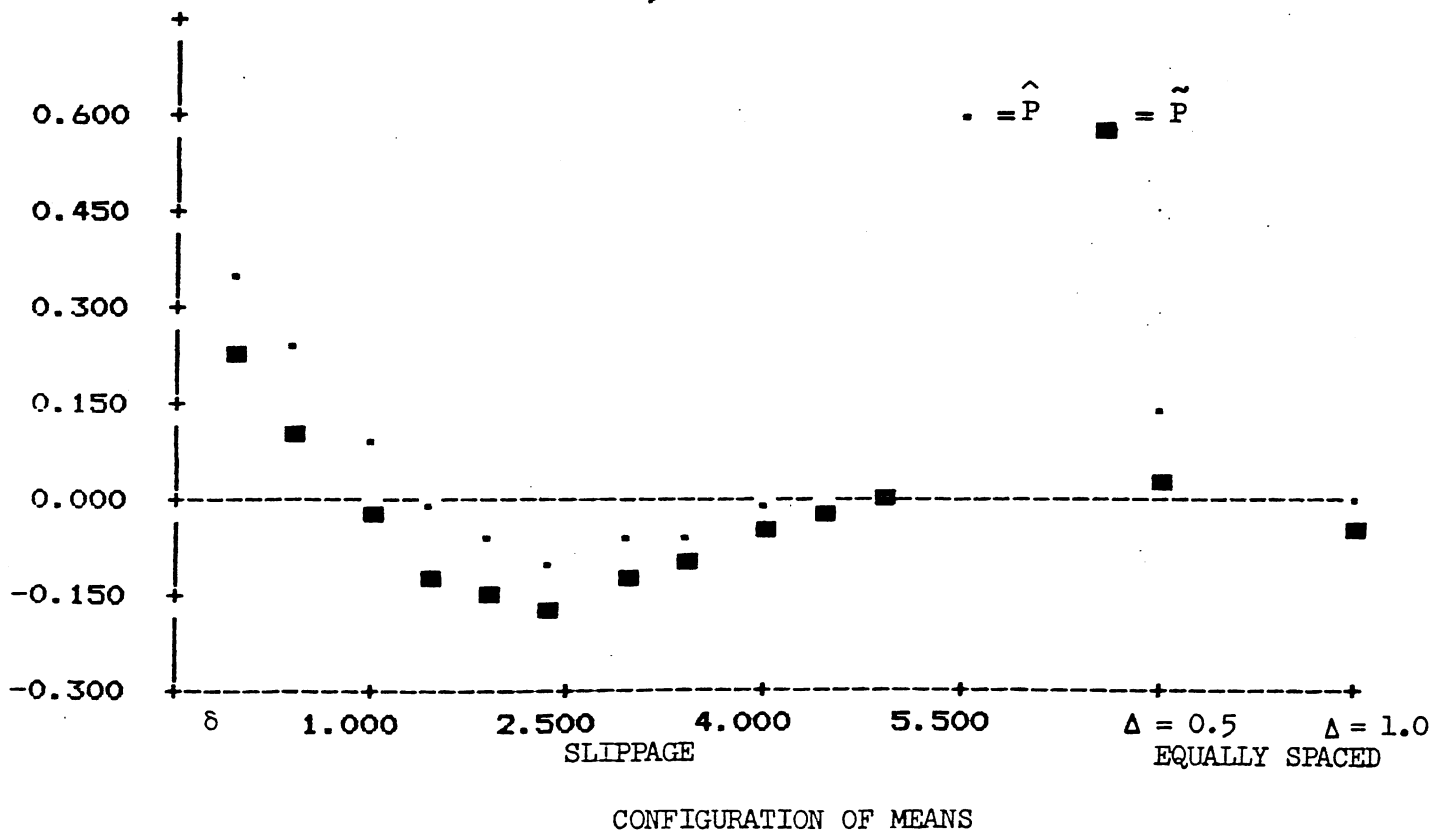
RMSE

FIGURE 1: ROOT MEAN SQUARE ERROR, K=4 POPULATIONS



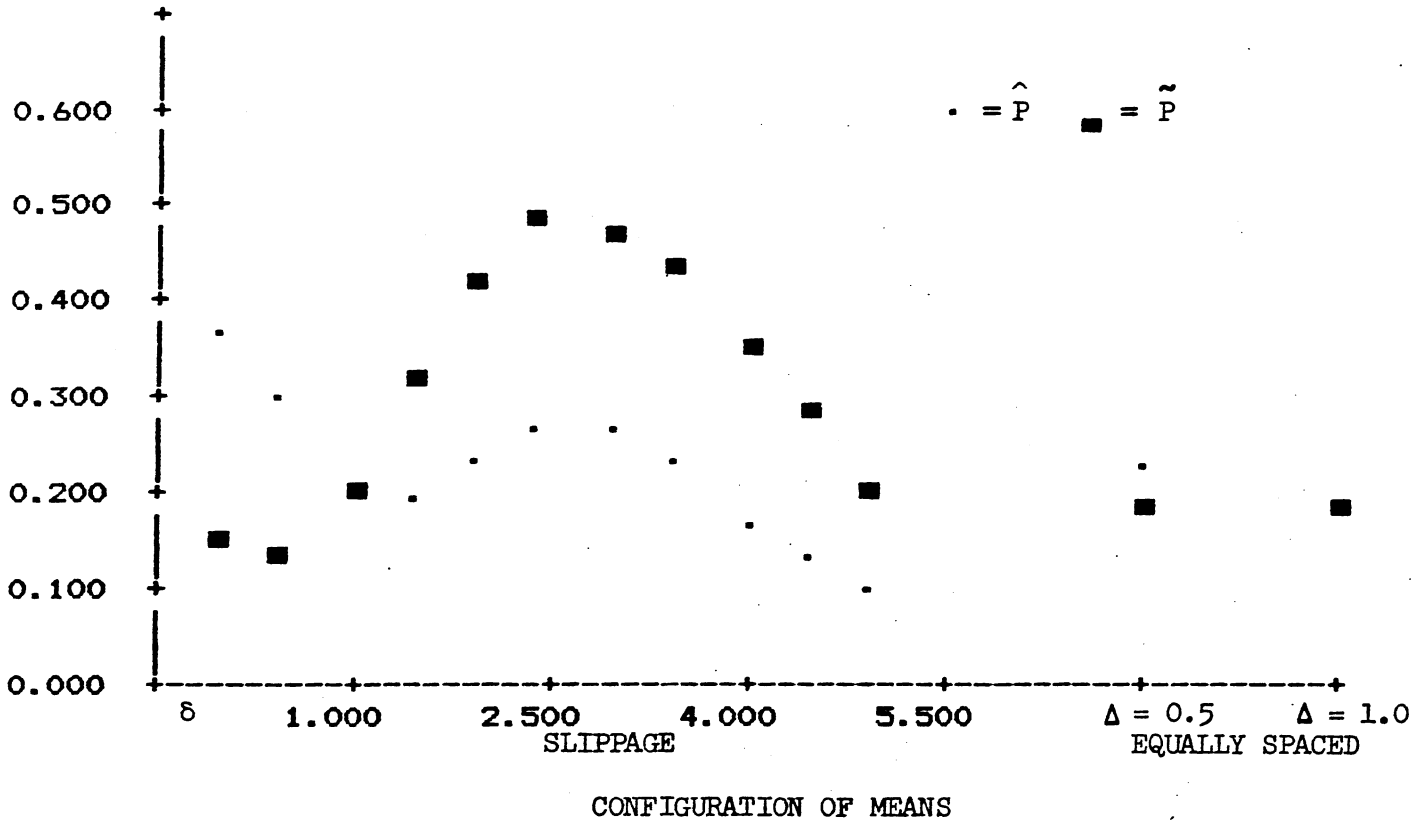
BIAS

FIGURE 2: BIAS, K=4 POPULATIONS



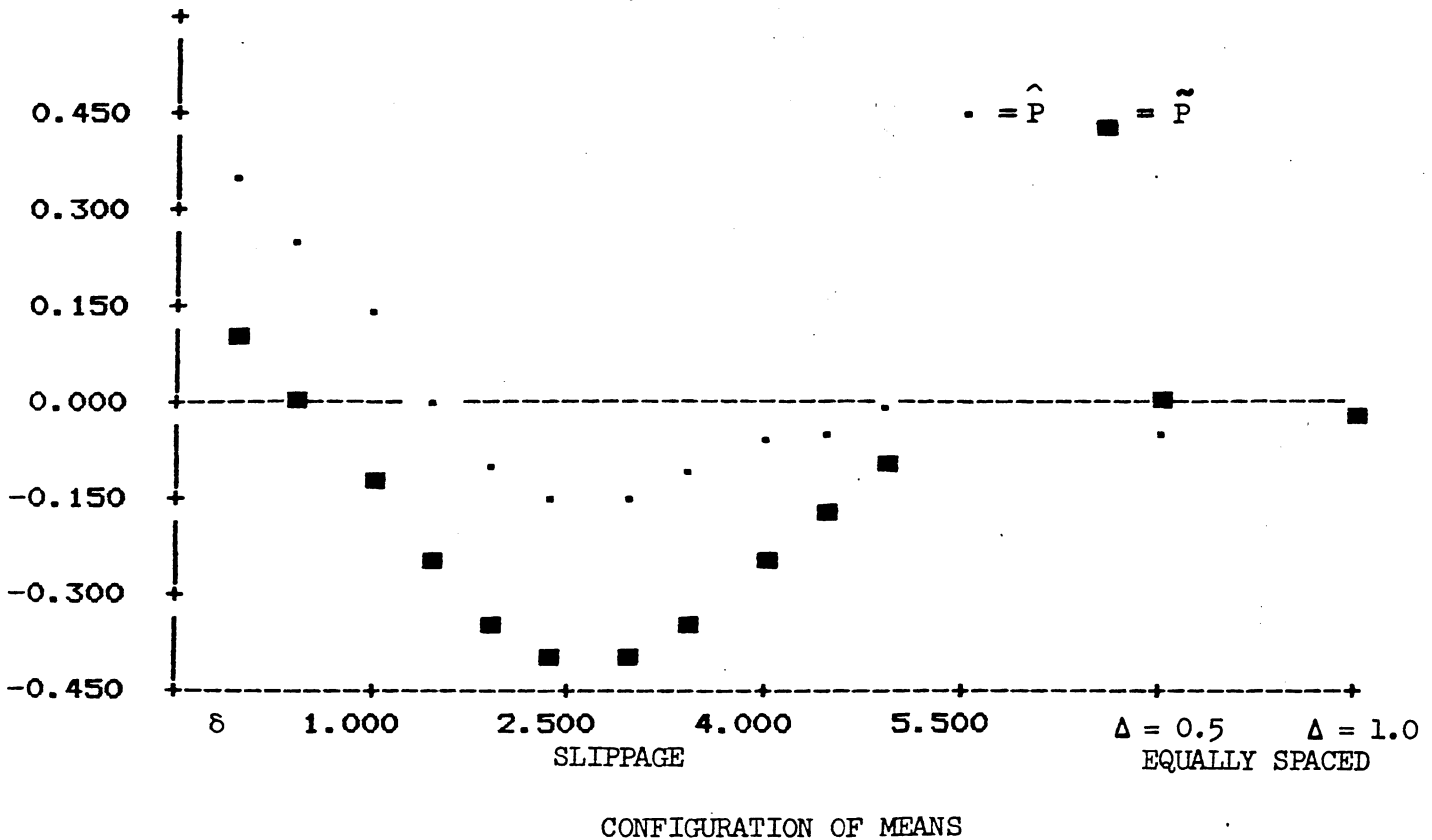
RMSE

FIGURE 3: ROOT MEAN SQUARE ERROR, K=10 POPULATIONS



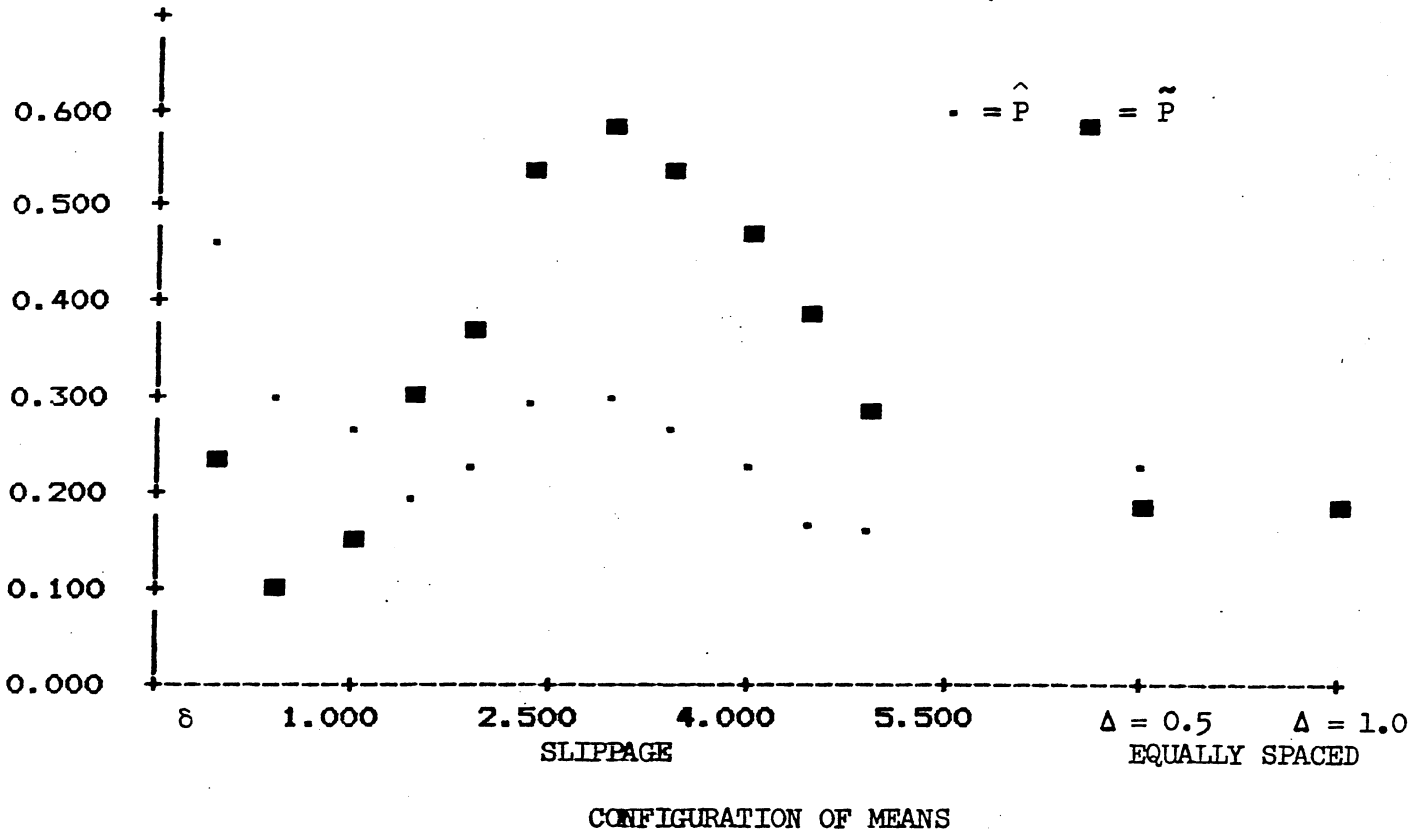
BIAS

FIGURE 4: BIAS, K=10 POPULATIONS



RMSE

FIGURE 5: ROOT MEAN SQUARE ERROR, K=15 POPULATIONS



BIAS

FIGURE 6: BIAS, K=15 POPULATIONS

