THE UNIVERSITY OF MANITOBA

POOLING OF SAMPLE MEANS
FROM TWO POISSON POPULATIONS

by

Shambhu Nath

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

Department of Statistics

WINNIPEG, MANITOBA

October 1972

ABSTRACT

In this thesis an attempt has been made to investigate the problem of pooling of means of two independent random samples from Poisson populations with parameters λ_1 and λ_2 respectively, where it is suspected, but not known with certainty, that $\lambda_2 = \lambda_1$. Estimators of λ_1 have been obtained using the following three approaches:

1. Non-Bayesian Approach
2. Semi-Bayesian Approach
3. Empirical Bayes Approach.

Restricting to the case of equal sample sizes we first consider the non-Bayesian approach which is a generalization of the preliminary test of significance (PTS) procedure first suggested by Huntsberger (1955). Following Huntsberger we use a weight function $\phi(S_1, S_2)$ to obtain an estimator of λ_1 given by

$$T(S_1, S_2) = 0 \qquad \text{if } S_1 = 0,\ S_2 = 0$$

$$= \frac{S_1}{m}\left[\frac{S_1^2 + 3S_2^2}{(S_1 + S_2)^2}\right] \qquad \text{otherwise}$$

where S_1 and S_2 are the respective sample totals, and m the common sample size.

The expected value and variance of $T(S_1, S_2)$, both exact and asymptotic, have been evaluated. The computer results on the asymptotic relative bias of $T(S_1, S_2)$ indicated that there is close agreement between the exact and asymptotic formulae. The asymptotic relative efficiency (ARE) of $T(S_1, S_2)$ with respect to $\bar{x} = \frac{S_1}{m}$, defined as

$$e = \frac{\lambda_1/m}{\text{Asymptotic MSE}[T(S_1, S_2)|\lambda_1, \lambda_2]} \cdot 100\%$$

has been computed for the values of $\lambda_1 = 0.5\ (0.1)\ 1.0$, $\lambda_2 = .5\ (.1)\ 1.0$ and $m = 10, 12, 14, 16, 18, 20, 25, 30$. It is observed that except for a small region of the values of λ_1 and λ_2, there is a gain in efficiency. For all the values of λ_1 and λ_2 where $\lambda_1 = \lambda_2$ the ARE is 200%.

Next we discuss the semi-Bayesian approach to the problem. A prior gamma distribution $G(\frac{\alpha}{\lambda_1}, \alpha)$ where α is known, is assumed for the parameter of the second population. Two different models with examples have been considered. It is found that the asymptotic variance of the maximum likelihood estimator in both the cases is smaller than the exact variance of $\bar{x}$, the mean of the first sample. Here we also derive an estimator on the basis of the best unbiased linear combination where the weights are estimated from the first sample, as it has smaller variance than that of the second sample. However, we find that in neither of the two situations considered does it give any improvement over the maximum likelihood estimator.

Finally we deal with the case where α is assumed unknown. Following empirical Bayes approach it is assumed that past experience for estimating α is available. Because of the various complexities involved with this procedure, only a partial solution has been given. The procedure for estimating α has been discussed and the empirical Bayes estimators of λ_1 for the two cases have been obtained.

ACKNOWLEDGEMENTS

I feel deeply indebted to Professor B. K. Kale, Head, Department of Statistics, for suggesting to me the problem and for his encouraging guidance and readiness to assist me at all times. I am also grateful to Dr. (Mrs.) K. Subrahmaniam, Department of Statistics, and Dr. C. R. Bector, Department of Actuarial and Business Mathematics, for their helpful suggestions in preparing the final draft of this thesis.

My sincere thanks are also due to Mrs. A. Loewen for her diligent typing.

TO MY PARENTS AND SISTER

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION AND SUMMARY

1.1 Review of Literature.

The problem of pooling of means of two independent random samples from normal populations has been studied in the past by various authors, namely Mosteller (1948), Bennett (1952), Graybill and Deal (1959), Kitagawa (1963), Zacks (1966) and Kale and Bancroft (1967), to mention a few. The first attempt to explore the same problem for discrete populations(in particular Poisson and Binomial) was made by Kale and Bancroft (1967). They used the square root transformation for the Poisson case and the arcsine transformation for the Binomial case and transformed the corresponding problem for the discrete distributions to the problem of pooling of two sample means from the normal populations with known variances.

Kale and Bancroft developed the theory for the following problem: There are two independent random samples $(Y_{11}, Y_{12}, \ldots, Y_{1n_1})$ and $(Y_{21}, Y_{22}, \ldots, Y_{2n_2})$ available from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ respectively, σ^2 being known. It is suspected, but not known with certainty, that $\mu_1 = \mu_2$. The problem is how to use this prior information in estimating μ_1.

They approached this problem through the "Theory of Incompletely Specified Models" as outlined by Bancroft (1964). [An extensive bibliography of the papers in this area is given in the paper by Bancroft (1972)].

Using a preliminary test of significance of size α to test $\mu_1 = \mu_2$, against $\mu_1 \neq \mu_2$, they proposed the following estimator for μ_1:

$$\bar{x}^* = \bar{x}_1 \qquad \text{if } |Z| \geq \xi_\alpha \sigma_Z$$

$$= \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \qquad \text{if } |Z| < \xi_\alpha \sigma_Z$$

where $Z = \bar{x}_1 - \bar{x}_2$, $\sigma_Z^2 = \sigma^2(\frac{1}{n_1} + \frac{1}{n_2})$ and ξ_α is given by $1 - \phi(\xi_\alpha) = \alpha/2$, ϕ being the cumulative distribution function of a $N(0, 1)$ variable.

The bias and mean squared error (MSE) of $\bar{x}^*$ were studied and the regions in the parameter space in which $\bar{x}^*$ has smaller mean squared error than the usual estimator $\bar{x}_1$, the mean of the first sample, were investigated. They also discussed the test of the hypothesis $\mu_1 = \mu_0$ subsequent to the preliminary test of significance (PTS) and studied its size and power.

Mosteller (1948) in his paper on "Pooling Data", which seems to be the first significant work on the subject of pooling of means as such, has also discussed this problem. Besides making a brief study of this problem, based on the test of significance of the null hypothesis $\mu_1 = \mu_2$, he suggested a Bayes approach to the problem. He assumed a prior distribution, namely, $N(0, a^2\sigma^2)$, $a^2\sigma^2$ known, for the difference $d = \mu_1 - \mu_2$ which is equivalent to assuming that μ_2 has a $N(\mu_1, a^2\sigma^2)$ distribution. Using the method of maximum likelihood, he derived the estimator:

$$\hat{\mu}_1 = \frac{\bar{x}(na^2 + 1) + \bar{y}}{na^2 + 2}$$

for the case where the two samples are of equal size. The mean squared error (MSE) of this estimator is:

$$D^2(\hat{\mu}_1) = \frac{\sigma^2}{n} \frac{1 + na^2}{2 + na^2} .$$

In fact, $\hat{\mu}_1$ is the best linear unbiased estimator of μ_1. These results can be generalized to the case of unequal sample sizes. Compared to the estimator $\bar{x}^*$ of Kale and Bancroft (1967), Mosteller's has uniformly smaller MSE than that of $\bar{x}$.

Zacks (1966) considered the problem of pooling of two sample means in a slightly different situation. There are two independent random samples of equal size from $N(\mu, \sigma_i^2)$, $(i = 1, 2)$. The problem is to estimate the common mean μ, the variance ratio $\rho = \sigma_2^2/\sigma_1^2$, being unknown. As in most of the works by various authors in this area, Zacks also used a preliminary test of significance to derive his "Class of Estimators" but in the closing section of his paper he suggests the use of Bayesian approach, i.e., assuming a prior distribution of ρ values, to investigate the problem. It may be noted that Graybill and Deal (1959) have also proposed a solution to the same problem by estimating ρ.

A very detailed investigation of the problem of pooling of sample means from two normal populations with same variance but different means, suspected to be close to each other, was made by Bruner (1967). Instead of using a preliminary test of significance, an empirical Bayes approach was used. Bruner studied the problem under the following two models:

Model I: $(X_1, X_2,\ldots,X_{n_1})$ and $(Y_1, Y_2,\ldots,Y_{n_2})$ are two independent random samples from normal populations $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ respectively, σ^2 being known. The mean of the first population, μ_1 is taken as fixed but unknown. The mean of the second population, μ_2 itself is assumed to be normally distributed with mean μ_1 and variance $a^2\sigma^2$ where a^2 is unknown.

Model II: In this model, the first sample is the same as in the first model but each member Y_j of the set of observations $(Y_1, Y_2,\ldots,Y_{n_2})$ from the second population is assumed to be normally distributed with mean μ_{2j} and known variance σ^2, where μ_{2j}, $j = 1, 2,\ldots,n_2$, is a random sample from a normal population with mean μ_1 and variance $a^2\sigma^2$ where a^2 is unknown.

Using the empirical Bayes method and thus assuming that there is some past experience available, the prior distribution of μ_2, or more specifically the parameter a^2, was estimated. Bruner derived the estimators for these two models and obtained exact expressions for their MSE.

These MSE's were then compared with that of $\bar{x}$, the mean of the first sample and under the first model it was shown that if the past experience consists of more than 10 samples from the second population then the empirical Bayes approach produces a better estimator (smaller MSE) than $\bar{x}$ and the same is true for model II if the size of the second sample $n_2 > 10$.

1.2 Problem and Summary of Results

In this thesis we will consider the problem of pooling of sample means from two Poisson populations with parameters λ_1 and λ_2 respectively. It is suspected, but not known with certainty, that $\lambda_1 = \lambda_2$. The problem is to estimate λ_1 using this prior information on λ_2. Solutions to this problem have been obtained using the following three approaches:

1. Non-Bayesian Approach
2. Semi-Bayesian Approach
3. Empirical Bayes Approach.

The models for which these three approaches have been used are described below. A brief summary of the results obtained is also presented.

Non-Bayesian Approach: We restrict ourselves to the case when two sample sizes are equal and consider the following situation:

There are two independent random samples $(X_1, X_2, \ldots, X_m)$ and $(Y_1, Y_2, \ldots, Y_m)$ available from Poisson populations $P(\lambda_1)$ and $P(\lambda_2)$ respectively, the problem being that of estimating λ_1 when it is suspected that $\lambda_2 = \lambda_1$.

Following Huntsberger (1955) we consider a weight function $\phi(S_1, S_2)$ such that $0 \leq \phi(S_1, S_2) \leq 1$ and construct estimator of the type

$$T = [1 - \phi(S_1, S_2)] \frac{S_1}{m} + \phi(S_1, S_2) \frac{S_1+S_2}{2m}$$

where $S_1 = \sum_{i=1}^{m} X_i$ and $S_2 = \sum_{i=1}^{m} Y_i$.

For particular choice of

$$\phi(S_1, S_2) = 1 \quad \text{if } \left|\frac{S_1}{S_1 + S_2} - \frac{1}{2}\right| < C_\alpha$$

$$= 0 \quad \text{otherwise}$$

where C_α is a constant, α being the level of significance of the PTS, we get estimators obtained by using the PTS approach. However, in this thesis we consider only the "continuous" weight functions which do not correspond to any estimator based on PTS.

The estimator derived using this approach is:

$$T(S_1, S_2) = 0 \qquad \text{if } S_1 = 0 \text{ and } S_2 = 0$$

$$= \frac{S_1}{m}\left[\frac{S_1^2 + 3S_2^2}{(S_1+S_2)^2}\right] \qquad \text{otherwise.}$$

The exact and asymptotic expected value and variance of $T(S_1, S_2)$ have also been obtained in Chapter II.

Some computer results on the asymptotic relative efficiency of $T(S_1, S_2)$ with respect to $\bar{x}$, the mean of the first sample, for sample sizes 10 (2) 20 (5) 30 and for pairs (λ_1, λ_2) where $\lambda_1 = .5\ (.1)\ 1.0$ and $\lambda_2 = .5\ (.1)\ 1.0$ have been obtained. It is observed that except for a small region of the values of λ_1 and λ_2, there is a gain in efficiency. It is also found that for all those pairs of values of λ_1 and λ_2 in which $\lambda_1 = \lambda_2$, the asymptotic relative efficiency (ARE) is 200%. We note that $T(S_1, S_2)$ is as efficient as $\frac{S_1+S_2}{2m}$, which is the UMVUE of the common mean $\lambda_1 = \lambda_2$.

Keeping in view the close theoretical connection between the Bayesian and empirical Bayes procedures we present a combined description of the models for which these have been used and give examples in each case.

Bayesian and Empirical Bayes Approach

Model I $X_1, X_2, \ldots, X_m$ is a random sample from a Poisson population with parameter λ_1 which is taken as fixed but unknown. $Y_1, Y_2, \ldots, Y_n$ is a second random sample from a Poisson population with parameter λ_2. We formalize the prior information "λ_2 is suspected to be close to λ_1" by assuming that the ratio $\theta = \lambda_2/\lambda_1$ has a gamma distribution $G(\alpha, p)$ defined by

$$g(\theta) = \frac{\alpha^p}{\Gamma(p)} e^{-\theta\alpha} \theta^{p-1}, \; \alpha > 0$$

such that $E(\theta) = 1$. This is equivalent to assuming that the parameter λ_2 of the second population follows a gamma distribution $G(\frac{\alpha}{\lambda_1}, \alpha)$.

We discuss the problem of estimating the mean of the first population λ_1, in both the cases (i) α known and (ii) α unknown. Assuming α to be known is similar to following Mosteller's approach (Bayesian) for normal populations where $\mu_2 - \mu_1$ is assigned a known prior distribution. In the case when α is assumed unknown, we use empirical Bayes approach according to which it is assumed that there is some past experience on the basis of which the prior distribution of λ_2 can be estimated. This past experience comprises p samples

$Y_{11}, Y_{12}, \ldots, Y_{1n}$

$Y_{21}, Y_{22}, \ldots, Y_{2n}$

$Y_{p1}, Y_{2p}, \ldots, Y_{pn}$

of size n each.

As an example of this model we may consider the following:

A random sample $X_1, X_2, \ldots, X_m$ from a population of bacterial colonies is available from an experiment. The number of bacterial colonies is assumed to follow a Poisson distribution with parameter λ_1. Another sample $Y_1, Y_2, \ldots, Y_n$ is available from a second population similar to the first one from another experiment. We assume that the second population of bacterial colonies follows a Poisson distributuion with parameter λ_2 where it is suspected that λ_2 is equal to λ_1. We further assume that the ratio $\theta = \lambda_2/\lambda_1$ follows a gamma distribution $G(\alpha, p)$ such that $E(\theta) = 1$. This is equivalent to assuming that λ_2 has a $G(\frac{\alpha}{\lambda_1}, \alpha)$. Previous p samples of size n each constitute the past experience for the estimation of the prior distribution of λ_2 or more specifically of α.

Model II $X_1, X_2, \ldots, X_m$ is a random sample from a Poisson distribution with parameter λ_1 where λ_1 is fixed but unknown – similar to the first model. We have another sample $Y_1, Y_2, \ldots, Y_n$ from a second population but in this model, instead of assuming a Poisson distribution with parameter λ_2 for each member of the second sample as is the case in the first model, we assume that each Y_j follows a Poisson distribution with parameter λ_{2j}, $(j = 1, 2, \ldots, n)$. It is further assumed that $\theta_{2j} = \lambda_{2j}/\lambda_1$, $(j = 1, 2, \ldots, n)$ are i.i.d. variates $G(\alpha, p)$ such that $E(\theta_{2j}) = 1$. This assumption amounts to saying that $\{\lambda_{2j}\}_{j=1}^{n}$ are i.i.d. $G(\frac{\alpha}{\lambda_1}, \alpha)$.

It may be noted here that a situation very similar to the structure of the second sample for this model has been considered by Bates and

Neyman (1952).

As in the first model here also we consider both cases: α known and α unknown. It may be noted that this model is the same as the model I with the past experience. Here the past experience is constituted by the second sample itself. Thus we may consider $\{Y_{2j}\}_{j=1}^{n}$ as n samples of size one each. Since $\{Y_{2j}\}_{j=1}^{n}$ are i.i.d. $G(\frac{\alpha}{\lambda_1}, \alpha)$ we can estimate α from this set of observations alone.

As an example of this model we consider the following case which has been described by Arbous and Kerrich (1951).

We have for the current year a random sample $X_1, X_2, \ldots, X_m$ from a population of accidents assumed to follow a Poisson distribution with parameter λ_1. Also available is another sample from a similar population for the preceding year. We can conceive that the second population is non-homogeneous with regard to the accident proneness of its members. Thus it can be assumed that each Y_j of the second sample comes from a Poisson population with parameter λ_{2j}, $(j = 1, 2, \ldots, n)$. It is further assumed that $\theta_{2j} = \lambda_{2j}/\lambda_1$ follows a gamma distribution $G(\alpha, p)$ such that $E(\theta_{2j}) = 1$, or equivalently, λ_{2j} has a $G(\frac{\alpha}{\lambda_1}, \alpha)$. We note that here as contrasted with model I, we can estimate α from the second sample itself and therefore no past experience is necessary.

The case of α known for both the models I and II has been discussed in Chapter III. It is shown here that the maximum likelihood estimator of λ_1 in both the cases has asymptotically smaller MSE than that of either $\bar{x}$ or $\bar{y}$ alone, where $\bar{x}$ and $\bar{y}$ denote the mean of the first and second samples respectively. We also show that a weighted combination $u = w_1\bar{x} + w_2\bar{y}$ does not lead to a better (smaller MSE) estimation procedure. This should be contrasted with the results of Graybill and Deal (1959) for the normal populations with common mean but unequal variances where they proved that the estimator of the common mean obtained on the basis of the weighted combination is uniformally better than either $\bar{x}$ or $\bar{y}$ if n_1 and n_2 are both larger than 10.

The case of α unknown for both the models, where we use the empirical Bayes procedure, is presented in Chapter IV. Because of various complexities involved with this procedure, only a partial solution to the problem has been given. We obtain an estimate of α for model I with past experience and for model II without any additional past experience. We use this estimator of α to obtain empirical Bayes estimator of λ_1.

CHAPTER II

NONBAYESIAN APPROACH

2.1 Introduction

In this chapter we consider the problem of pooling of means from two samples of equal size from the Poisson populations with parameters λ_1 and λ_2 and it is suspected that $\lambda_1 = \lambda_2$. We derive an estimator of λ_1 by using the approach of Huntsberger (1955). Huntsberger's approach is a generalization of preliminary test of significance (PTS) approach and does not use any prior distribution for the parameter λ_2. If one were to use a PTS approach the new estimator T^* would be

$$T^* = \frac{S_1}{m}, \qquad \frac{S_1}{S_1 + S_2} < C_1(\alpha) \text{ or } \frac{S_1}{S_1 + S_2} > C_2(\alpha)$$
$$= \frac{S_1 + S_2}{2m}, \qquad C_1(\alpha) < \frac{S_1}{S_1 + S_2} < C_2(\alpha)$$

where $C_1(\alpha)$ and $C_2(\alpha)$ are constants and α is the level of PTS. S_1 and S_2 are defined as before. In view of the symmetry of the problem $C_1(\alpha)$ and $C_2(\alpha)$ are symmetric around $\frac{1}{2}$, i.e., $C_1(\alpha) + C_2(\alpha) = 1$.

We now consider the generalization of T^* by a weight function $\phi(\frac{S_1}{S_1+S_2})$

$$T_\phi(S_1, S_2) = [1 - \phi(\frac{S_1}{S_1 + S_2})] \frac{S_1}{m} + [\phi(\frac{S_1}{S_1 + S_2})] \frac{S_1 + S_2}{2m}$$

where $0 \leq \phi(u) \leq 1$. Note that $\phi(u) = 0$ corresponds to never pool procedure while $\phi(u) = 1$ corresponds to always pool procedure and

$\phi(u) = 1$ for $C_1(\alpha) \le u \le C_2(\alpha)$ and zero otherwise corresponds to the sometimes pool procedure where the PTS is carried at level α.

2.2 Selection of Weight Function ϕ

We want to select a weight function ϕ defined over $0 \le u \le 1$ such that

(a) $0 \le \phi(u) \le 1$

(b) $\phi(\frac{1}{2}) = 1$ and $\phi(u) \to 0$ as $u \to 0$ and $u \to 1$

(c) $\phi(u)$ is symmetric around $u = \frac{1}{2}$

(d) $\phi(u)$ is differentiable everywhere.

We further want ϕ to be a fairly simple function. (d) rules out PTS estimators or linear functions of u. The simplest quadratic function of u that satisfies the four conditions mentioned above is $\phi(u) = 4u(1 - u)$, $0 \le u \le 1$. For this weight function, the resulting estimator is

$$T(S_1, S_2) = 0 \qquad \text{if } S_1 = 0 \text{ and } S_2 = 0$$

$$= \frac{S_1}{m}\left[\frac{S_1^2 + 3S_2^2}{(S_1+S_2)^2}\right] \qquad \text{otherwise} \tag{2.2.1}$$

2.3 Exact Bias and MSE of $T(S_1, S_2)$

We have from (2.2.1)

$$E[T(S_1, S_2)] = E\left[\frac{S_1}{m} - \frac{2S_1S_2(S_1 - S_2)}{m(S_1 + S_2)^2}\right]$$

$$= \lambda_1 - E\left[\frac{2S_1S_2(S_1 - S_2)}{m(S_1 + S_2)^2}\right]$$

as S_1 and S_2 have Poisson distributions $P(m\lambda_1)$ and $P(m\lambda_2)$ respectively. Thus we can rewrite as

$$E[T(S_1, S_2)] = \lambda_1 - \underset{t}{E}\left[\underset{S_1}{E}\left\{\left[\frac{2S_1S_2(S_1 - S_2)}{m(S_1 + S_2)^2}\right] \middle| S_1 + S_2 = t\right\}\right]$$

$$= \lambda_1 - \frac{2}{m}\underset{t}{E}\left[\underset{S_1}{E}\left\{\left[\frac{S_1(t - S_1)(2S_1 - t)}{t^2}\right] \middle| S_1 + S_2 = t\right\}\right].$$

(2.3.1)

Since the conditional distribution of S_1 given $S_1 + S_2 = t$ is a binomial distribution with $n = t$ and $p = \frac{m\lambda_1}{m\lambda_1 + m\lambda_2} = \frac{\lambda_1}{\lambda_1 + \lambda_2}$, (2.3.1) is:

$$E[T(S_1, S_2)] = \lambda_1 - \frac{2}{m}\underset{t}{E}\left[\sum_{S_1=0}^{t} \frac{S_1(t - S_1)(2S_1 - t)}{t^2}\binom{t}{S_1}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{S_1}\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{t-S_1}\right]$$

$$= \lambda_1 - \frac{2}{m}\underset{t}{E}\left[\frac{1}{t^2}\left\{\sum_{S_1=0}^{t}(3S_1^2 - t^2S_1 - 2S_1^3)\binom{t}{S_1}p^{S_1}q^{t-S_1}\right\}\right]$$

(2.3.2)

On substitution of the first three raw moments of the binomial distribution B(t, p) in (2.3.2), we get

$$E[T(S_1, S_2)] = \lambda_1 - \frac{2}{m}\underset{t}{E}\left[\frac{1}{t^2}\{t^3(-p + 3p^2 - 2p^3) + t^2(3p - 9p^2 + 6p^3) + t(-2p + 6p^2 - 4p^3)\}\right]$$

$$= \lambda_1 - \frac{6A}{m}\sum_{t=1}^{\infty} p(t) + \frac{2A}{m}\sum_{t=1}^{\infty} tp(t) + \frac{4A}{m}\sum_{t=1}^{\infty}\frac{1}{t}p(t), \quad (2.3.3)$$

where

$$A = p - 3p^2 + 2p^3 = \frac{\lambda_1\lambda_2(\lambda_2 - \lambda_1)}{(\lambda_1 + \lambda_2)^3}$$

and

$$p(t) = \frac{e^{-m(\lambda_1+\lambda_2)}\{m(\lambda_1 + \lambda_2)\}^t}{t!}, \quad t = 0, 1, 2, \ldots, \infty.$$

After simplifying (2.3.3) a little further we find that

$$E[T(S_1, S_2)] = \lambda_1 - \frac{2\lambda_1\lambda_2(\lambda_1 - \lambda_2)}{m(\lambda_1 + \lambda_2)^3}$$
$$\{m(\lambda_1 + \lambda_2) - 3[1 - e^{-m(\lambda_1+\lambda_2)}]\}$$
$$- \frac{4\lambda_1\lambda_2(\lambda_1 - \lambda_2)}{m(\lambda_1 + \lambda_2)^3}\{1 - e^{-m(\lambda_1+\lambda_2)}\}E(X^{-1}) \qquad (2.3.4)$$

where X has a Poisson distribution $P(m\lambda_1 + m\lambda_2)$ truncated at zero.

Thus the bias of the estimator $T(S_1, S_2)$ is

$$B(T) = - \frac{2\lambda_1\lambda_2(\lambda_1 - \lambda_2)}{m(\lambda_1 + \lambda_2)^3}\Big[\{m(\lambda_1 + \lambda_2) - 3[1 - e^{-m(\lambda_1+\lambda_2)}]\}$$
$$+ 2\{1 - e^{-m(\lambda_1+\lambda_2)}\}E(X^{-1})\Big]. \qquad (2.3.5)$$

Now for variance of $T(S_1, S_2)$, we evaluate first

$$E[T(S_1, S_2)]^2 = E\left[\frac{S_1}{m} - \frac{2S_1S_2(S_1 - S_2)}{m(S_1 + S_2)^2}\right]^2$$
$$= E\left[\frac{S_1^2}{m^2} + \frac{4S_1^2S_2^2(S_1 - S_2)^2}{m^2(S_1 + S_2)^4} - \frac{4S_1^2S_2^2(S_1 - S_2)}{m^2(S_1 + S_2)^2}\right]$$

$$= \lambda_1^2 + \frac{\lambda_1}{m} + \frac{4}{m^2} \mathop{E}_{t} \mathop{E}_{S_1} \left[\left\{ \frac{S_1^2 S_2^2 (S_1 - S_2)^2}{(S_1 + S_2)^4} \right\} \Big| S_1 + S_2 = t \right]$$

$$- \frac{4}{m^2} \mathop{E}_{t} \mathop{E}_{S_1} \left[\left\{ \frac{S_1^2 S_2 (S_1 - S_2)}{(S_1 + S_2)^4} \right\} \Big| S_1 + S_2 = t \right]$$

$$= \lambda_1^2 + \frac{\lambda_1}{m} + \frac{4}{m^2} \mathop{E}_{t} \mathop{E}_{S_1} [\{t^4 S_1^2 - 6t^3 S_1^3 + 13t^2 S_1^4 - 12t S_1^5 + 4S_1^6\} | S_1 + S_t = t]$$

$$- \frac{4}{m^2} \mathop{E}_{t} \mathop{E}_{S_1} \left[\left\{ - \frac{t^4 S_1^2 + 3t^3 S_1^3 - 2t^2 S_1^4}{t^4} \right\} \Big| S_1 + S_2 = t \right]$$

$$= \lambda_1^2 + \frac{\lambda_1}{m} + \frac{4}{m^2} \mathop{E}_{t} \left[A_1 t^2 + A_2 t + A_3 + \frac{A_4}{t} + \frac{A_5}{t^2} + \frac{A_6}{t^3} \right] \qquad (2.3.6)$$

where t has $P(m\lambda_1 + m\lambda_2)$ truncated at zero and

$$A_1 = \frac{\lambda_1^2}{(\lambda_1 + \lambda_2)^6} [21\lambda_1^4 + 26\lambda_1^3\lambda_2 + 27\lambda_1^2\lambda_2^2 + 8\lambda_1\lambda_2^3 + 2\lambda_2^4],$$

$$A_2 = \frac{\lambda_1}{(\lambda_1 + \lambda_2)^6} [5\lambda_1^4\lambda_2 - 13\lambda_1^3\lambda_2^2 + 21\lambda_1^2\lambda_2^3 - 19\lambda_1\lambda_2^4 + 2\lambda_2^5],$$

$$A_3 = -\frac{\lambda_1}{(\lambda_1 + \lambda_2)^6} [56\lambda_1^4\lambda_2 - 83\lambda_1^3\lambda_2^2 + 150\lambda_1^2\lambda_2^3 - 87\lambda_1\lambda_2^4 + 9\lambda_2^5],$$

$$A_4 = \frac{\lambda_1}{(\lambda_1 + \lambda_2)^6} [15\lambda_1^4\lambda_2 - 210\lambda_1^3\lambda_2^2 + 450\lambda_1^2\lambda_2^3 - 210\lambda_1\lambda_2^4 + 15\lambda_2^5],$$

$$A_5 = -\frac{\lambda_1}{(\lambda_1 + \lambda_2)^6} [12\lambda_1^4\lambda_2 - 244\lambda_1^3\lambda_2^2 + 584\lambda_1^2\lambda_2^3 - 244\lambda_1\lambda_2^4 + 12\lambda_2^5],$$

and

$$A_6 = \frac{4\lambda_1}{(\lambda_1 + \lambda_2)^6} [\lambda_1^4\lambda_2 - 26\lambda_1^3\lambda_2^2 + 66\lambda_1^2\lambda_2^3 - 26\lambda_1\lambda_2^4 + \lambda_2^5].$$

Further calculations give

$$E[T(S_1, S_2)]^2 = \lambda_1^2 + \frac{\lambda_1}{m}$$

$$+ \frac{4}{m^2\{1-e^{-m(\lambda_1+\lambda_2)}\}}\Big[A_1\{1 + m^2(\lambda_1+\lambda_2)^2\} + mA_2(\lambda_1+\lambda_2) + A_3$$

$$- e^{-m(\lambda_1+\lambda_2)}\{A_1\{1 + m(\lambda_1+\lambda_2)\} + mA_2(\lambda_1+\lambda_2) + 2A_3\}$$

$$+ A_3 e^{-2m(\lambda_1+\lambda_2)}\Big]$$

$$+ \frac{4}{m^2}\{1 - e^{-m(\lambda_1+\lambda_2)}\}[A_4E(X^{-1}) + A_5E(X^{-2}) + A_6E(X^{-3})]. \quad (2.3.7)$$

Hence, from (2.3.7) and (2.3.4) we get

$$\mathrm{Var}[T(S_1, S_2)] = \frac{\lambda_1}{m}$$

$$+ \frac{4\lambda_1^2\lambda_2(\lambda_1-\lambda_2)}{m(\lambda_1+\lambda_2)^3}\{m(\lambda_1+\lambda_2) - 3[1 - e^{-m(\lambda_1+\lambda_2)}]\}$$

$$- \frac{4\lambda_1^2\lambda_2^2(\lambda_1-\lambda_2)^2}{m^2(\lambda_1+\lambda_2)^6}\{m(\lambda_1+\lambda_2) - 3[1 - e^{-m(\lambda_1+\lambda_2)}]\}^2$$

$$+ \frac{4}{m^2\{1 - e^{-m(\lambda_1+\lambda_2)}\}}\Big[A_1\{1 + m^2(\lambda_1+\lambda_2)^2\} + mA_2(\lambda_1+\lambda_2) + A_3$$

$$- e^{-m(\lambda_1+\lambda_2)}\{A_1[1 + m(\lambda_1+\lambda_2)] + mA_2(\lambda_1+\lambda_2) + A_3\}$$

$$+ A_3 e^{-2m(\lambda_1+\lambda_2)}\Big]$$

$$+ \left[\frac{4}{m^2}\{1 - e^{-m(\lambda_1+\lambda_2)}\}A_4\right.$$

$$+ \frac{8\lambda_1\lambda_2(\lambda_1-\lambda_2)}{m(\lambda_1+\lambda_2)^3}\{1 - e^{-m(\lambda_1+\lambda_2)}\}$$

$$\left.\cdot\{\lambda_1 - \frac{2\lambda_1\lambda_2(\lambda_1-\lambda_2)}{m(\lambda_1+\lambda_2)^3}[m(\lambda_1+\lambda_2) - 3\{1 - e^{-m(\lambda_1+\lambda_2)}\}]\}\right]E(X^{-1})$$

$$- \frac{16\lambda_1^2\lambda_2^2(\lambda_1-\lambda_2)^2}{m^2(\lambda_1+\lambda_2)^6}\{1 - e^{-m(\lambda_1+\lambda_2)}\}^2\{E(X^{-1})\}^2$$

$$+ \frac{4}{m^2}\{1 - e^{-m(\lambda_1+\lambda_2)}\}[A_5E(X^{-2}) + A_6E(X^{-3})] \qquad (2.3.8)$$

2.4 Asymptotic Bias and MSE

Now $T(S_1, S_2)$ is differentiable and the conditions given by Kendall and Stuart (1958) for the variance of the estimators $\bar{x}$ and $\bar{y}$ are satisfied. Therefore, in this case, the Taylor series expansion can be used to obtain the asymptotic mean and variance of the estimator $T(S_1, S_2)$.

We have

$$T(S_1, S_2) = \frac{\bar{x}^3 + 3\bar{x}\bar{y}^2}{(\bar{x} + \bar{y})^2} .$$

Expanding $T(S_1, S_2)$ in a Taylor series at $\bar{x} = \lambda_1$ and $\bar{y} = \lambda_2$, we can write $T(S_1, S_2)$ as:

$$T(S_1, S_2) = T(S_1, S_2)\Big|_{\substack{\bar{x} = \lambda_1 \\ \bar{y} = \lambda_2}}$$

$$+ \frac{1}{1!} \{(\bar{x} - \lambda_1) \frac{\partial}{\partial \bar{x}} + (\bar{y} - \lambda_2) \frac{\partial}{\partial \bar{y}}\} T(S_1, S_2) \Big|_{\substack{\bar{x} = \lambda_1 \\ \bar{y} = \lambda_2}}$$

$$+ \frac{1}{2!} \{(\bar{x} - \lambda_1)^2 \frac{\partial^2}{\partial \bar{x}^2} + 2(\bar{x} - \lambda_1)(\bar{y} - \lambda_2) \frac{\partial^2}{\partial \bar{y} \partial \bar{x}}$$

$$+ (\bar{y} - \lambda_2)^2 \frac{\partial^2}{\partial \bar{y}^2}\} T(S_1, S_2) \Big|_{\substack{\bar{x} = \lambda_1 \\ \bar{y} = \lambda_2}}$$

ignoring terms of order $\frac{1}{m^{2+\delta}}$, $\delta > 0$. Using the formula of Kendall and Stuart (1958) and after some algebra, we obtain

$$E[T(S_1, S_2)] = \lambda_1 - \frac{2\lambda_1 \lambda_2 (\lambda_1 - \lambda_2)}{(\lambda_1 + \lambda_2)^2} \{1 - \frac{3}{m(\lambda_1 + \lambda_2)}\}. \tag{2.4.1}$$

Thus the asymptotic bias of $T(S_1, S_2)$ is given by

$$Ba(T) = - \frac{2\lambda_1 \lambda_2 (\lambda_1 - \lambda_2)}{m(\lambda_1 + \lambda_2)^3} \{m(\lambda_1 + \lambda_2) - 3\}. \tag{2.4.2}$$

Similarly, the asymptotic variance is given by

$$Var_a[T(S_1, S_2)] = \frac{\lambda_1}{m(\lambda_1 + \lambda_2)^6} [\lambda_1^6 + 10\lambda_1^5 \lambda_2 - 21\lambda_1^4 \lambda_2^2 + 24\lambda_1^3 \lambda_2^3$$

$$+ 27\lambda_1^2 \lambda_2^4 - 18\lambda_1 \lambda_2^5 + 9\lambda_2^6]. \tag{2.4.3}$$

2.5 Some Asymptotic Results on Bias and MSE

The asymptotic relative bias B is defined as

$$B = \frac{|\text{asymptotic bias} - \text{exact bias}|}{\text{exact bias}} . \tag{2.5.1}$$

100B% values for sample sizes 10 (2) 20 (5) 30 and for pairs (λ_1, λ_2) where $\lambda_1 = .5\ (.1)\ 1.0$ and $\lambda_2 = .5\ (.1)\ 1.0$ were computed and it was observed that as the sample size increased, B decreased and even for the smallest sample size m = 10, we found that the maximum value of B was 0.5219×10^{-4}. This indicated that the asymptotic formula provided very good approximations. For sample of size 16 the maximum of B was of order 10^{-7} and for sample of size 30 the maximum of B was of order 10^{-13}. It is for this reason and the fact that the computation of the exact variance of $T(S_1, S_2)$ is too involved, we computed the asymptotic variance of $T(S_1, S_2)$ only. The comparison of $T(S_1, S_2)$ and $\bar{x}$ was based on the asymptotic MSE.

The asymptotic relative efficiency (ARE) of $T(S_1, S_2)$ with respect to $\bar{x}$ is defined as

$$e = \frac{\lambda_1/m}{\text{Asymptotic MSE}[T(S_1, S_2)/\lambda_1, \lambda_2]} \times 100\%.$$

Tables 1 through VIII give the values of e for each sample size and combination of values of (λ_1, λ_2) mentioned above.

For sample size 10, the ARE is less than 100% for the combination of values: (.5, .9), (.5, 1.0), (.6, 1.0), the minimum being 70% for (.5, 1.0). For all other values it is greater than 100%, the maximum being 208% which occurs at $\lambda_1 = .6$, $\lambda_2 = .5$.

For sample size 12, there is a loss in efficiency for the same combination of values of λ_1 and λ_2 as those for sample size 10. In this case also, the minimum and maximum ARE's which are respectively 63% and 205% occur at the same points as for sample size 10.

The ARE for sample size 14 is less than 100% for the following values of (λ_1, λ_2): (.5, .8), (.5, .9), (.5, 1.0), and (.6, 1.0), the minimum being 56% for (.5, 1.0). For all other combinations the ARE is greater than 100% with a maximum of 201% for $\lambda_1 = .6$, $\lambda_2 = .5$.

The loss in efficiency for sample sizes 16 and 18 occurs at the following common points (.5, .8), (.5, .9), (.5, 1.0), (.6, .9), (.6, 1.0), and (1.0, .5) with an additional point (.7, 1.0) for the sample size 18. The minimum ARE occurs at (.5, 1.0) in both the cases, which is 50% for sample size 16 and 46% for sample size 18. The maximum ARE is 200% for both the sample sizes and this occurs along the diagonal $\lambda_1 = \lambda_2$.

For sample sizes 20, 25 and 30 the ARE is less than 100% for the following combinations: (.5, .8), (.5, .9), (.5, 1.0), (.6, .9), (.6, 1.0), (.7, 1.0), (.9, .5), (1.0, .5), (1.0, .6), with three extra points (.5, .7), (.8, .5), and (.9, .6) for sample size 30. The minimum ARE's for the three cases are respectively 42%, 35%, 30% which occur at the same point (.5, 1.0). For all other combinations the ARE is more than 100% with a maximum of 200% which occurs along the diagonal $\lambda_1 = \lambda_2$ in each of the three cases.

We note that when $\lambda_1 = \lambda_2$, the ARE is 200%. This implies that when $\lambda_1 = \lambda_2$

$$MSE[T(S_1, S_2)|\lambda_1, \lambda_2] = \frac{\lambda_1}{2m}$$

which is the MSE of the minimum variance unbiased estimator $\frac{\bar{x} + \bar{y}}{2}$.

TABLE I

VALUES OF e

Sample Size 10

λ_1 \ λ_2	.5	.6	.7	.8	.9	1.0
.5	200	174	142	113	89	71
.6	208	200	177	147	118	95
.7	196	206	200	179	151	123
.8	174	194	205	200	181	155
.9	151	173	193	204	200	183
1.0	132	152	173	192	203	200

TABLE II

VALUES OF e

Sample Size 12

λ_1 \ λ_2	.5	.6	.7	.8	.9	1.0
.5	200	171	135	104	80	63
.6	205	200	174	141	110	86
.7	186	203	200	177	145	115
.8	160	186	202	200	179	149
.9	137	161	185	201	200	181
1.0	117	138	162	186	201	200

TABLE III

VALUES OF e

Sample Size 14

λ_1 \ λ_2	.5	.6	.7	.8	.9	1.0
.5	200	168	129	96	72	56
.6	201	200	172	135	102	78
.7	178	200	200	175	140	108
.8	149	178	200	200	177	144
.9	124	150	179	199	200	179
1.0	105	126	152	179	199	200

TABLE IV

VALUES OF e

Sample Size 16

λ_1 \ λ_2	.5	.6	.7	.8	.9	1.0
.5	200	166	123	89	66	50
.6	198	200	170	129	96	72
.7	170	198	200	173	135	101
.8	139	171	197	200	175	139
.9	114	141	172	197	200	177
1.0	96	116	143	173	197	200

TABLE V

VALUES OF e

Sample Size 18

λ_1 \ λ_2	.5	.6	.7	.8	.9	1.0
.5	200	163	118	83	61	46
.6	195	200	167	124	90	66
.7	162	195	200	171	130	96
.8	130	164	195	200	173	135
.9	105	132	166	195	200	176
1.0	86	107	135	168	195	200

TABLE VI

VALUES OF e

Sample Size 20

λ_1 \ λ_2	.5	.6	.7	.8	.9	1.0
.5	200	161	113	78	56	42
.6	192	200	165	119	85	61
.7	156	192	200	169	125	90
.8	122	158	193	200	172	130
.9	97	125	160	193	200	174
1.0	81	99	127	162	193	200

TABLE VII

VALUES OF e

Sample Size 25

λ_1 \ λ_2	.5	.6	.7	.8	.9	1.0
.5	200	155	102	68	47	35
.6	185	200	160	109	74	52
.7	141	186	200	164	115	80
.8	106	144	187	200	167	121
.9	83	109	148	188	200	170
1.0	67	85	112	151	189	200

TABLE VIII

VALUES OF e

Sample Size 30

λ_1 \ λ_2	.5	.6	.7	.8	.9	1.0
.5	200	149	93	60	41	30
.6	178	200	155	100	66	45
.7	129	180	200	159	107	71
.8	93	133	182	200	163	113
.9	72	97	137	183	200	166
1.0	58	74	101	140	184	200

CHAPTER III

SEMI-BAYESIAN APPROACH

In this chapter we derive the estimators of λ_1 for the two models when α is assumed known. We consider the model I first.

3.1 Model I

Here we have $(X_1, X_2, \ldots, X_m)$ and $(Y_1, Y_2, \ldots, Y_n)$ independent random samples from two Poisson populations with parameters λ_1 and λ_2 respectively where λ_1 is taken as fixed but unknown. It is assumed that the ratio of the two parameters $\theta = \lambda_2/\lambda_1$ is distributed as $G(\alpha, p)$ such that $E(\theta) = 1$. It can then be easily shown that this is equivalent to assuming that λ_2 has a distribution $G(\frac{\alpha}{\lambda_1}, \alpha)$.

Let $\underline{X} = (X_1, X_2, \ldots, X_m)$

$\underline{Y} = (Y_1, Y_2, \ldots, Y_n)$.

The joint distribution of $\underline{X}$ and $\underline{Y}$ given λ_1 and λ_2 is

$$L^*(\underline{X}, \underline{Y}|\lambda_1, \lambda_2) = \left[\prod_{i=1}^{m} \frac{e^{-\lambda_1}\lambda_1^{X_i}}{X_i!}\right]\left[\prod_{j=1}^{n} \frac{e^{-\lambda_2}\lambda_2^{Y_j}}{Y_j!}\right] \tag{3.1.1}$$

$$= e^{-m\lambda_1}\lambda_1^{S_1} e^{-n\lambda_2}\lambda_2^{S_2} H_1(\underline{X}, \underline{Y}) \tag{3.1.2}$$

where $S_1 = \sum_{i=1}^{m} X_i$, $S_2 = \sum_{j=1}^{n} Y_j$ and $H_1(\underline{X}, \underline{Y})$ is a function of sample observations alone. Now since λ_2 is a random variable with distribution $G(\frac{\alpha}{\lambda_1}, \alpha)$, we have the marginal distribution of $\underline{X}$ and $\underline{Y}$ given by:

$$L(\underline{X}, \underline{Y}|\lambda_1) = e^{-m\lambda_1}\lambda_1^{S_1} H_1(\underline{X}, \underline{Y}) \int_0^\infty e^{-n\lambda_2}\lambda_2^{S_2}$$

$$\cdot \frac{(\alpha/\lambda_1)^\alpha}{\Gamma(\alpha)} e^{-\lambda_2(\alpha/\lambda_1)} \lambda_2^{\alpha-1} d\lambda_2$$

$$= H_1(\underline{X}, \underline{Y}) e^{-m\lambda_1}\lambda_1^{S_1} \frac{(\alpha/\lambda_1)^\alpha}{\Gamma(\alpha)} \frac{\Gamma(S_2 + \alpha)}{(n+\alpha/\lambda_1)^{S_2+\alpha}}$$

$$= H_1(\underline{X}, \underline{Y}) \lambda_1^{S_1} \left(\frac{\alpha}{\alpha+n\lambda_1}\right)^\alpha \left(\frac{\alpha}{\alpha+n\lambda_1}\right)^{S_2} \frac{\Gamma(S_2+\alpha)}{\Gamma(\alpha)} \text{ .} \qquad (3.1.3)$$

After some suitable adjustments (3.1.3) can be written as

$$L(\underline{X}, \underline{Y}|\lambda_1) = p_1(S_1)\cdot p_2(S_2)\cdot H(\underline{X}, \underline{Y}) \qquad (3.1.4)$$

where

$$p_1(S_1) = \frac{e^{-m\lambda_1}(m\lambda_1)^{S_1}}{S_1!}$$

$$p_2(S_2) = \frac{\Gamma(S_2+\alpha)}{\Gamma(\alpha)S_2!} \frac{1}{(1+\frac{n\lambda_1}{\alpha})^\alpha} \left[\frac{n\lambda_1/\alpha}{1 + \frac{n\lambda_1}{\alpha}}\right]^{S_2} \qquad (3.1.5)$$

and $H(\underline{X}, \underline{Y})$ is a function of sample observations alone. This shows that S_1 is $P(m\lambda_1)$ and S_2 is $NB(\alpha, \frac{n\lambda_1}{\alpha})$. The general form of the negative binomial distribution that has been used throughout is given by:

$$NB(N, P) = \frac{\Gamma(N + x)}{\Gamma(N)x!} Q^{-N}(P/Q)^x \qquad x = 0, 1, 2, \ldots, \qquad (3.1.6)$$

where

$$Q - P = 1.$$

From (3.1.3) it is easily verified that the usual regularity conditions such as given by Cramer (1946) and Huzurbazar (1948) are satisfied and the maximum likelihood estimate (MLE) λ_1 is the unique solution of the likelihood equation $\frac{\partial \log L}{\partial \lambda_1} = 0$. Further the asymptotic variance of $\hat{\lambda}_1$ is given by

$$\left\{-E\left[\frac{\partial^2 \log L}{\partial \lambda_1^2}\right]\right\}^{-1}.$$

Taking logrithm of both sides of (3.1.3), differentiating it partially with respect to λ_1 and equating the derivative to zero, we get

$$\frac{\partial \log L}{\partial \lambda_1} = -m + \frac{S_1 + S_2}{\lambda_1} - \frac{(S_2 + \alpha)n}{\alpha + n\lambda_1} = 0 \qquad (3.1.7)$$

or

$$\frac{-m\lambda_1(\alpha + n\lambda_1) + (S_1 + S_2)(\alpha + n\lambda_1) - n\lambda_1(S_2 + \alpha)}{\lambda_1(\alpha + n\lambda_1)} = 0 \qquad (3.1.8)$$

or

$$\lambda_1^2 mn - \lambda_1\{nS_1 - \alpha(m + n)\} - \alpha(S_1 + S_2) = 0,$$

the roots of which are

$$\hat{\lambda}_1 = \frac{nS_1 - \alpha(m + n) \pm \sqrt{\{nS_1 - \alpha(m + n)\}^2 + 4\alpha mn(S_1 + S_2)}}{2mn}.$$

Since $\{nS_1 - \alpha(m + n)\}^2 + 4\alpha mn(S_1 + S_2) > \{nS_1 - \alpha(m + n)\}^2$ for all values of $S_1 \geq 0$, $S_2 \geq 0$, $\hat{\lambda}_1$ is given by

$$\hat{\lambda}_1 = \frac{nS_1 - \alpha(m + n) + \sqrt{\{nS_1 - \alpha(m + n)\}^2 + 4\alpha mn(S_1 + S_2)}}{2mn}. \qquad (3.1.9)$$

We also show that $\hat{\lambda}_1$ provides the maximum of the likelihood by proving that $\frac{\partial^2 \log L}{\partial \lambda_1^2} < 0$ at $\lambda_1 = \hat{\lambda}_1$.

We have from (3.1.8)

$$\frac{\partial^2 \log L}{\partial \lambda_1^2} = N \frac{d}{d\lambda_1}\left\{\frac{1}{\lambda_1(\alpha + n\lambda_1)}\right\} + \frac{1}{\lambda_1(\alpha + n\lambda_1)} \frac{d}{d\lambda_1}(N) \qquad (3.1.10)$$

where

$$N = -m\lambda_1(\alpha + n\lambda_1) + (S_1 + S_2)(\alpha + n\lambda_1) - n\lambda_1(S_2 + \alpha).$$

At $\hat{\lambda}_1$, $N(\hat{\lambda}_1) = 0$. Also $\hat{\lambda}_1(\alpha + n\hat{\lambda}_1) > 0$. Therefore, to show that $\frac{\partial^2 \log L}{\partial \lambda_1^2} < 0$, we need to investigate the sign of $\frac{dN}{d\lambda_1}$ at $\hat{\lambda}_1$.

$$\frac{dN}{d\lambda_1} = -m(\alpha + n\lambda_1) - mn\lambda_1 + n(S_1 + S_2) - n(S_2 + \alpha)$$

$$= -\alpha(m + n) + nS_1 - 2mn\lambda_1$$

so that

$$\frac{dN}{d\hat{\lambda}_1} = -\alpha(m + n) + nS_1 - 2mn\hat{\lambda}_1. \qquad (3.1.11)$$

But from (3.1.9)

$$2mn\hat{\lambda}_1 = nS_1 - \alpha(m + n) + D \qquad (3.1.12)$$

where

$$D = \sqrt{\{nS_1 - \alpha(m + n)\}^2 + 4\alpha mn(S_1 + S_2)} \quad , \qquad D > 0.$$

Hence on substitution of (3.1.12) in (3.1.11), we get

$$\frac{dN}{d\hat{\lambda}_1} = -\alpha(m + n) + nS_1 - nS_1 + \alpha(m + n) - D$$

$$= -D < 0.$$

Thus

$$\frac{\partial^2 \log L}{\partial \lambda_1^2} < 0 \text{ at } \lambda_1 = \hat{\lambda}_1.$$

Therefore, $\hat{\lambda}_1$ given by (3.1.9) is the maximum likelihood estimate of λ_1.

The asymptotic variance of $\hat{\lambda}_1$ is given by

$$\text{Var}(\hat{\lambda}_1) = -\left| E \frac{\partial^2 \log L}{\partial \lambda_1^2} \right|^{-1} .$$

From (3.1.7) we have

$$\frac{\partial^2 \log L}{\partial \lambda_1^2} = -\frac{(S_1 + S_2)}{\lambda_1^2} + \frac{(S_2 + \alpha)n^2}{(\alpha + n\lambda_1)^2}$$

$$E(\frac{\partial^2 \log L}{\partial \lambda_1^2}) = -\frac{(m + n)}{\lambda_1} + \frac{n^2}{\alpha + n\lambda_1}$$

$$= \frac{-m(\alpha + n\lambda_1) - n\alpha}{\lambda_1(\alpha + n\lambda_1)} .$$

Hence

$$\text{Var}(\hat{\lambda}_1) = \frac{\lambda_1}{m} \frac{(1 + \frac{n\lambda_1}{\alpha})}{(1 + \frac{n\lambda_1}{\alpha}) + \frac{n}{m}} . \tag{3.1.13}$$

It is easily verified that $\text{Var}(\hat{\lambda}_1) \leq V(\bar{x})$. It may be remarked that $\text{Var}(\hat{\lambda}_1)$ is the asymptotic variance of $\hat{\lambda}_1$, the maximum likelihood estimate whereas $V(\bar{x})$ is the exact variance of $\bar{x}$.

Next we consider model II.

3.2 Model II

A random sample $(X_1, X_2, \ldots, X_m)$ from a Poisson distribution with parameter λ_1 is available where λ_1 is taken as fixed but unknown. We have another sample $(Y_1, Y_2, \ldots, Y_n)$ from a second population where it is assumed that each Y_j has a Poisson distribution with parameter λ_{2j}, $(j = 1, 2, \ldots, n)$. It is further assumed that $\{\theta_{2j}\}_{j=1}^n$, where $\theta_{2j} = \lambda_{2j}/\lambda_1$ are i.i.d. random variables $G(\alpha, p)$ such that $E(\theta_{2j}) = 1$, which is equivalent to assuming that $\{\lambda_{2j}\}_{j=1}^n$ are i.i.d. $G(\frac{\alpha}{\lambda_1}, \alpha)$.

Since Y_j, $(j = 1, 2, \ldots, n)$ is assumed to follow a Poisson distribution $P(\lambda_{2j})$, where λ_{2j} has a $G(\frac{\alpha}{\lambda_1}, \alpha)$, it follows that the marginal distribution of Y_j is a $NB(\alpha, \frac{\lambda_1}{\alpha})$.

Therefore, the marginal distribution of $\underline{X}$ and $\underline{Y}$ is

$$L(\underline{X}, \underline{Y}|\lambda_1) = \left(\prod_{i=1}^{m} \frac{1}{X_i!}\right)\left(\prod_{j=1}^{n} \frac{\Gamma(Y_j + \alpha)}{Y_j!}\right)$$

$$\cdot \left(\frac{\alpha}{\alpha + \lambda_1}\right)^{\alpha n} \frac{1}{\{\Gamma(\alpha)\}^n} \left(\frac{\lambda_1}{\alpha + \lambda_1}\right)^{S_2} e^{-m\lambda_1} \lambda_1^{S_1} \qquad (3.2.1)$$

where S_1 and S_2 are defined as before.

Here again it is easy to verify that the regularity conditions assumed by Cramer (1946) and Huzurbazar (1948) are satisfied and $\tilde{\lambda}_1$, the MLE, is the unique solution of $\frac{\partial \log L}{\partial \lambda_1} = 0$ and the asymptotic variance of $\tilde{\lambda}_1$ is given by

$$\left\{-E\left[\frac{\partial^2 \log L}{\partial \lambda_1^2}\right]\right\}^{-1}.$$

Now from (3.2.1), the likelihood equation is

$$\frac{\partial \log L}{\partial \lambda_1} = -\frac{\alpha n}{\alpha + \lambda_1} + \frac{S_2}{\lambda_1} - \frac{S_2}{\alpha + \lambda_1} - n + \frac{S_2}{\lambda_1} = 0. \qquad (3.2.2)$$

The roots of (3.2.2) are

$$\tilde{\lambda}_1 = \frac{S_1 - \alpha(m + n) \pm \sqrt{\{S_1 - \alpha(m + n)\}^2 + 4\alpha m(S_1 + S_2)}}{2m}. \qquad (3.2.3)$$

Following very similar arguments as those given in the previous section, we find that the maximum likelihood estimate of λ_1 for the second model, is

$$\tilde{\lambda}_1 = \frac{S_1 - \alpha(m + n) + \sqrt{\{S_1 - \alpha(m + n)\}^2 + 4\alpha m(S_1 + S_2)}}{2m} \qquad (3.2.4)$$

and it provides the unique maximum of the likelihood since it can be shown that $\frac{\partial^2 \log L}{\partial \lambda_1^2} < 0$ at $\lambda_1 = \tilde{\lambda}_1$.

To obtain the asymptotic variance of $\tilde{\lambda}_1$, the maximum likelihood estimate, we evaluate first

$$\frac{\partial^2 \log L}{\partial \lambda_1^2} = \frac{\alpha n}{(\alpha + \lambda_1)^2} - \frac{\alpha S_2(\alpha + 2\lambda_1)}{\{\lambda_1(\alpha + \lambda_1)\}^2} - \frac{S_1}{\lambda_1^2}$$

from (3.2.2), so that

$$E\left(\frac{\partial^2 \log L}{\partial \lambda_1^2}\right) = \frac{\alpha n \lambda_1^2 - \alpha n \lambda_1(\alpha + 2\lambda_1) - m\lambda_1(\alpha + \lambda_1)^2}{\{\lambda_1(\alpha + \lambda_1)\}^2} .$$

Hence, after some simplification,

$$\begin{aligned} \mathrm{Var}(\tilde{\lambda}_1) &= -\left\{ E\left[\frac{\partial^2 \log L}{\partial \lambda_1^2}\right]\right\}^{-1} \\ &= \frac{\lambda_1}{m} \quad \frac{1}{1 + \frac{n}{m}\left(\frac{\alpha}{\alpha + \lambda_1}\right)} . \end{aligned} \tag{3.2.5}$$

As in model I, here also we find that the asymptotic variance of the maximum likelihood estimate $\tilde{\lambda}_1$, is smaller than the exact variance of the first sample mean $\bar{x}$. It may also be noted that the asymptotic variance of the maximum likelihood estimate for the first model is larger than that for the second model.

3.3 Combining Unbiased Estimators

Graybill and Deal (1959) considered the same problem for two samples $(X_1, X_2, \ldots, X_{n_1})$ and $(Y_1, Y_2, \ldots, Y_{n_2})$ from normal populations with common mean μ and unknown variances σ_1^2 and σ_2^2 respectively. They considered the best unbiased linear combination of $\bar{x}$ and $\bar{y}$ and estimated weights by using estimators s_1^2, s_2^2 of σ_1^2 and σ_2^2 respectively and obtained the estimator

$$\hat{\mu} = (n_1 s_2^2 \bar{x} + n_2 s_1^2 \bar{y})/(n_1 s_2^2 + n_2 s_1^2)$$

where $s_1^2 = \Sigma(x_i - \bar{x})^2/(n_1 - 1)$ and $s_2^2 = \Sigma(y_i - \bar{y})^2/(n_2 - 1)$. They proved that $\hat{\mu}$ is an unbiased estimator of μ and that it is uniformally better than either $\bar{x}$ or $\bar{y}$ if n_1 and n_2 are both larger than 10. We note that the estimators of the weights are stochastically independent of $\bar{x}$ and $\bar{y}$.

We consider model I first and assume that the sample sizes are equal. The best unbiased linear combination is now given by

$$T = w_1 \bar{x} + w_2 \bar{y}$$

where

$$w_1 = \frac{1 + \frac{m\lambda_1}{\alpha}}{2 + \frac{m\lambda_1}{\alpha}} \quad \text{and} \quad w_2 = \frac{1}{2 + \frac{m\lambda_1}{\alpha}}$$

and m denotes the common sample size so that

$$T = \frac{\bar{x} + \bar{y} + \frac{m\bar{x}\lambda_1}{\alpha}}{2 + \frac{m\lambda_1}{\alpha}} . \tag{3.3.1}$$

We find that the weights w_1 and w_2 are functions of λ_1 which is being estimated by $\bar{x}$ and $\bar{y}$ both. It is not possible to obtain an estimator of λ_1 which is independent of $\bar{x}$ and $\bar{y}$, that can be used in the weights w_1 and w_2. Thus we have three alternatives:

(1) Estimate w_1 and w_2 by estimating λ_1 by $\bar{x}$;

(2) Estimate w_1 and w_2 by estimating λ_1 by $\bar{y}$;

(3) Estimate w_1 and w_2 by estimating λ_1 from both the samples.

If we use (3), the evaluation of MSE of the final estimator would be too complex. Between (1) and (2) we prefer (1), because we note that $V(\bar{x}) = \frac{\lambda_1}{m}$, while $V(\bar{y}) = \frac{\lambda_1}{m}(1 + \frac{m\lambda_1}{\alpha})$ and $V(\bar{y})$ does not tend to zero even for large samples. It is for this reason it is expected that combining $\bar{x}$ and $\bar{y}$ may not lead generally to an improved estimator. This must be contrasted with the MLE of λ_1 obtained in the previous section, which has asymptotic variance smaller than that of $\bar{x}$ or $\bar{y}$. Choosing the first alternative we get from (3.3.1) the estimator

$$u^* = \frac{1}{m} \frac{S_1 + S_2 + S_1^2/\alpha}{2 + \frac{S_1}{\alpha}} \tag{3.3.2}$$

we obtain $E(u^*)$ and $Var(u^*)$ and show that u^* does not give any improvement (i.e., smaller MSE) over $\bar{x}$. Now

$$E(u^*) = \frac{1}{m} \sum_{S_1=0}^{\infty} \sum_{S_2=0}^{\infty} \frac{S_1 + S_2 + S_1^2/\alpha}{2 + \frac{S_1}{\alpha}} p_1(S_1)p_2(S_2)$$

$$= \frac{1}{m} \sum_{S_1=0}^{\infty} \left[S_1 - \frac{\alpha S_1}{2\alpha + S_1} + \frac{\alpha m\lambda_1}{2\alpha + S_1}\right] p_1(S_1)$$

which after some further simplification is

$$E(u^*) = \lambda_1 - \frac{\alpha}{m} + \frac{1}{m} \sum_{S_1=0}^{\infty} \frac{2\alpha^2 + \alpha m\lambda_1}{2\alpha + S_1} p_1(S_1).$$

After some further algebra we get

$$E(u^*) = \lambda_1 - \frac{\alpha}{m} + (\frac{\alpha}{m} + \frac{1}{2})e^{-m\lambda_1} {}_1F_1\left[\begin{matrix}2\alpha; \\ 2\alpha + 1;\end{matrix} \; m\lambda_1\right] \tag{3.3.3}$$

where ${}_pF_q$ is the generalized hypergeometric function defined as

$$ {}_pF_q\left[\begin{matrix}\alpha_1, \alpha_2, \ldots, \alpha_p; \\ \beta_1, \beta_2, \ldots, \beta_q;\end{matrix}\; Z\right] = \sum_{n=0}^{\infty} \frac{\prod_{i=1}^{p} (\alpha_i)_n}{\prod_{j=1}^{q} (\beta_j)_n} \frac{Z^n}{n!} $$

and

$$ (a)_n = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)} . $$

The variance of u^* is

$$ \mathrm{Var}(u^*) = E[(u^*)^2] - [E(u^*)]^2. $$

We have

$$ E(u^*)^2 = \frac{1}{m^2} E\left[\frac{(S_1 + S_2 + S_1^2/\alpha)}{(2 + S_1/\alpha)}\right]^2 $$

$$ = \frac{1}{m^2} \sum_{S_1=0}^{\infty} \sum_{S_2=0}^{\infty} \left[S_1^2 - \frac{2\alpha S_1^2}{2\alpha+S_1} + \frac{\alpha^2 S_1^2}{(2\alpha+S_1)^2} + \frac{2\alpha S_1 S_2}{2\alpha+S_1} \right. $$

$$ \left. + \frac{\alpha^2 S_2^2}{(2\alpha+S_1)^2} - \frac{2\alpha^2 S_1 S_2}{(2\alpha+S_1)^2}\right] $$

$$ \cdot \frac{e^{-m\lambda_1}(m\lambda_1)^{S_1}}{S_1!} \frac{\Gamma(S_2 + \alpha)}{\Gamma(\alpha) S_2!} \frac{1}{(1 + \frac{m\lambda_1}{\alpha})} \left(\frac{\frac{m\lambda_1}{\alpha}}{1 + \frac{m\lambda_1}{\alpha}}\right)^{S_2} \qquad (3.3.4) $$

$$ = \frac{1}{m^2} \sum_{S_1=0}^{\infty} \left[S_1^2 - \frac{2\alpha S_1^2 - 2\alpha S_1 m\lambda_1}{2\alpha + S_1} \right. $$

$$ \left. + \frac{\alpha^2 S_1^2 - 2\alpha^2 m\lambda_1 S_1 + \alpha^2\{(m\lambda_1)^2 + m\lambda_1(1 + \frac{m\lambda_1}{\alpha})\}}{(2\alpha + S_1)^2}\right] $$

$$\cdot \frac{e^{-m\lambda_1}(m\lambda_1)^{S_1}}{S_1!} . \tag{3.3.5}$$

Rearranging the terms in (3.3.5) and simplifying it further, we get

$$E[(u^*)^2] = \frac{1}{m^2}\Big[(m\lambda_1)^2 + m\lambda_1 + 5\alpha^2 - (12\alpha^3 + 6\alpha^2 m\lambda_1)E(\frac{1}{2\alpha + S_1})$$

$$+ \{4\alpha^4 + 4\alpha^3 m\lambda_1 + \alpha^2[(m\lambda_1)^2 + m\lambda_1(1 + \frac{m\lambda_1}{\alpha})]\}E(\frac{1}{(2\alpha + S_1)^2})\Big] \tag{3.3.6}$$

$$= \frac{1}{m^2}\Big[(m\lambda_1)^2 + m\lambda_1 + 5\alpha^2 - (12\alpha^3 + 6\alpha^2 m\lambda_1) \sum_{S_1=0}^{\infty} \frac{1}{2\alpha + S_1} \frac{e^{-m\lambda_1}(m\lambda_1)^{S_1}}{S_1!}$$

$$+ \{4\alpha^4 + 4\alpha^3 m\lambda_1 + \alpha^2[(m\lambda_1)^2 + m\lambda_1(1 + \frac{m\lambda_1}{\alpha})]\}$$

$$\cdot \sum_{S_1=0}^{\infty} \frac{1}{(2\alpha + S_1)^2} \frac{e^{-m\lambda_1}(m\lambda_1)^{S_1}}{S_1!}\Big] . \tag{3.3.7}$$

After suitable adjustment of the terms in (3.3.7), we obtain finally

$$E(u^*)^2 = \frac{1}{m^2}\Big[(m\lambda_1)^2 + m\lambda_1 + 5\alpha^2 - (6\alpha^2 + 3\alpha m\lambda_1)e^{-m\lambda_1}$$

$${}_1F_1\left[\begin{matrix}2\alpha & ; \\ 2\alpha + 1; & m\lambda_1\end{matrix}\right] + \{\alpha^2 + \alpha m\lambda_1 + \frac{1}{4}[(m\lambda_1)^2 + m\lambda_1(1 + \frac{m\lambda_1}{\alpha})]\}$$

$$\cdot e^{-m\lambda_1} {}_2F_2\left[\begin{matrix}2\alpha, 2\alpha & ; \\ 2\alpha + 1, 2\alpha + 1; & m\lambda_1\end{matrix}\right]\Big] . \tag{3.3.8}$$

Therefore, from (3.3.3) and (3.3.8)

$$V(u^*) = \frac{\lambda_1}{m} + \frac{2\alpha}{m}(\frac{2\alpha}{m} + \lambda_1) - (\frac{2\alpha}{m} + \lambda_1)^2 e^{-m\lambda_1} {}_1F_1\left[\begin{matrix}2\alpha & ; \\ 2\alpha + 1; & m\lambda_1\end{matrix}\right]$$

$$- \left[\left(\frac{\alpha}{m} + \frac{\lambda_1}{2}\right) e^{-m\lambda_1} \; {}_1F_1\left[\begin{matrix} 2\alpha & ; \\ 2\alpha + 1; & \end{matrix} m\lambda_1\right] \right]^2$$

$$+ \left[\left(\frac{\alpha}{m}\right)^2 + \frac{\alpha\lambda_1}{m} + \frac{1}{4}\left\{\lambda_1^2 + \frac{\lambda_1}{m}\left(1 + \frac{m\lambda_1}{\alpha}\right)\right\}\right] e^{-m\lambda_1} \; {}_2F_2\left[\begin{matrix} 2\alpha, \; 2\alpha & ; \\ 2\alpha + 1, \; 2\alpha + 1; & \end{matrix} m\lambda_1\right]. \tag{3.3.9}$$

We now discuss the asymptotic behaviour of the estimator u^*.

It is known that asymptotically the following formula is true: [Refer to "Handbook of Mathematical Functions", edited by M. Abramowitz and I. A. Stegun (1964)].

$${}_1F_1\left[\begin{matrix} 2\alpha & ; \\ 2\alpha + 1; & \end{matrix} m\lambda_1\right] = \frac{\Gamma(2\alpha + 1)}{\Gamma(2\alpha)} e^{m\lambda_1} (m\lambda_1)^{2\alpha - \overline{2\alpha+1}} \left[1 + 0\left(\frac{1}{m\lambda_1}\right)\right]$$

$$= 2\alpha \, e^{m\lambda_1} (m\lambda_1)^{-1} \left[1 + 0\left(\frac{1}{m\lambda_1}\right)\right].$$

Hence, from (3.3.3),

$$E(u^*) \approx \lambda_1 + \frac{2\alpha^2}{m\lambda_1} + 0\left(\frac{1}{m^2}\right) \tag{3.3.10}$$

which shows that for large m the bias is negligible. Now

$$E\left(\frac{1}{2\alpha + S_1}\right) = \sum_{S_1=0}^{\infty} \frac{e^{-m\lambda_1} (m\lambda_1)^{S_1}}{S_1!} \frac{1}{(2\alpha + S_1)} \tag{3.3.11}$$

$$= \frac{e^{-m\lambda_1}}{2} \; {}_1F_1\left[\begin{matrix} 2\alpha & ; \\ 2\alpha + 1; & \end{matrix} m\lambda_1\right]$$

$$\approx \frac{1}{m\lambda_1} + 0\left(\frac{1}{m^2}\right).$$

As the series in (3.3.11) is uniformally convergent differentiation with respect to α on both sides can be performed [see Cramer (1966)].

Thus we get

$$E \frac{1}{(2\alpha + S_1)^2} \approx O(\frac{1}{m^2}).$$

Hence from (3.3.6)

$$E[(u^*)^2] \approx \lambda_1^2 + \frac{\lambda_1}{m} + O(\frac{1}{m^2}) \tag{3.3.12}$$

and from (3.3.10) and (3.3.12) we have

$$\text{Var}(u^*) \approx \frac{\lambda_1}{m} + O(\frac{1}{m^2}).$$

Thus asymptotically u^* as an estimate of λ_1 would be as good as $\bar{x}$ only. We have already seen that the MLE $\hat{\lambda}_1$ has asymptotic variance smaller than that of $\bar{x}$. Therefore, we prefer $\hat{\lambda}_1$ to $\bar{x}$ and would also prefer $\hat{\lambda}_1$ to u^*.

Similarly for the model II we know that $E(\bar{x}) = \lambda_1$, $E(\bar{y}) = \lambda_1$, $V(\bar{x}) = \frac{\lambda_1}{m}$ and $V(\bar{y}) = \frac{\lambda_1}{m}(1 + \frac{\lambda_1}{\alpha})$. Thus the optimum weights are

$$\tilde{w}_1 = \frac{1 + \frac{\lambda_1}{\alpha}}{2 + \frac{\lambda_1}{\alpha}} \quad \text{and} \quad \tilde{w}_2 = \frac{1}{2 + \frac{\lambda_1}{\alpha}}.$$

Here also we note that $V(\bar{y}) > V(\bar{x})$. Hence, following the arguments given earlier, we estimate λ_1 occurring in $(\tilde{w}_1, \tilde{w}_2)$ by $\bar{x}$. This gives us the estimator

$$u^{**} = \frac{1}{m}\left\{\frac{S_1 + S_2 + S_1^2/\alpha m}{2 + S_1/\alpha m}\right\}. \tag{3.3.13}$$

After a long algebra and following precisely the same methods as were used for evaluating $E(u^*)$, we find that

$$E(u^{**}) = \lambda_1 + \frac{2\alpha^2}{\lambda_1} + 0(\frac{1}{m\lambda_1}). \qquad (3.3.14)$$

Thus we note that the bias of u^{**} has a constant term $\frac{2\alpha^2}{\lambda_1}$ which is independent of m, the sample size. This shows that u^{**} would not be as good as $\tilde{\lambda}_1$, which being MLE, has bias tending to zero as $m \to \infty$. Therefore, in model II also we prefer the MLE $\tilde{\lambda}_1$ to u^{**}.

CHAPTER IV

EMPIRICAL BAYES APPROACH

Here we consider the problem of estimating λ_1 for the two models when α is assumed unknown. First we consider model I.

4.1 Empirical Bayes Estimator for Model I

Following empirical Bayes approach we assume that past experience for estimating α in the form of p previous samples

$Y_{11}, Y_{12},\ldots,Y_{1n}$

$Y_{21}, Y_{22},\ldots,Y_{2n}$

$Y_{p1}, Y_{p2},\ldots,Y_{pn}$

each of size n is available. It may be noted that α and λ_1 both cannot be estimated from the second sample. It is also asusmed that each sample $(Y_{j1}, Y_{j2},\ldots,Y_{jn})$ $(j = 1, 2,\ldots,p)$ comes from a Poisson distribution with parameter λ_{2j}. It is further assumed that $\{\lambda_{2j}\}_{j=1}^{n}$ are i.i.d. $G(\frac{\alpha}{\lambda_1}, \alpha)$.

Let

$$S_{2j} = \sum_{k=1}^{n} Y_{kj}, \; j = 1, 2,\ldots,p.$$

The conditional distribution of S_{2j}, given λ_{2j}, is Poisson $P(n\lambda_{2j})$. As λ_{2j} has been assumed to follow a $G(\frac{\alpha}{\lambda_1}, \alpha)$, it is easily verified that the unconditional distribution of the j^{th} sample total S_{2j}, has a negative binomial distribution $NB(\alpha, \frac{n\lambda_1}{\alpha})$.

The joint distribution function for the sample totals

$S_{21}, S_{22}, \ldots, S_{2p}$ is given by

$$L(S_{21}, S_{22}, \ldots, S_{2p}) = \prod_{j=1}^{p} \frac{\Gamma(S_{2j} + \alpha)}{S_{2j}!\Gamma(\alpha)} \frac{1}{(1 + \frac{n\lambda_1}{\alpha})^{\alpha}} \left\{ \frac{n\lambda_1/\alpha}{1 + \frac{n\lambda_1}{\alpha}} \right\}^{S_{2j}}$$

$$= \frac{1}{\{\Gamma(\alpha)\}^p (1 + \frac{n\lambda_1}{\alpha})^{\alpha p}} \left\{ \frac{n\lambda_1/\alpha}{1 + \frac{n\lambda_1}{\alpha}} \right\}^{S} \prod_{j=1}^{p} \left\{ \frac{\Gamma(S_{2j} + \alpha)}{S_{2j}!} \right\} \qquad (4.1.1)$$

where

$$S = \sum_{j=1}^{p} S_{2j} = \sum_{j=1}^{p} \sum_{k=1}^{n} Y_{kj} .$$

We need to estimate α. Using the method of maximum likelihood estimate, we have from (4.1.1)

$$\frac{\partial \log L}{\partial \lambda_1} = - \frac{\alpha pn}{\alpha + n\lambda_1} + \frac{S}{\lambda_1} - \frac{nS}{\alpha + n\lambda_1} = 0 \qquad (4.1.2)$$

which when solved for λ_1 gives

$$\hat{\lambda}_1 = \frac{S}{np} . \qquad (4.1.3)$$

Next taking the partial derivative of the logrithm of (4.1.1) with respect to α and equating it to zero, we get

$$\frac{\partial \log L}{\partial \alpha} = -p\psi(\alpha) + \sum_{j=1}^{p} \psi(S_{2j} + \alpha) + p \log \frac{\alpha}{\alpha + n\lambda_1} - \frac{S - np\lambda_1}{\alpha + n\lambda_1} = 0 \qquad (4.1.4)$$

where

$$\psi(x) = \frac{d}{dx} [\log \Gamma(x)] = \frac{\Gamma'(x)}{\Gamma(x)}$$

is the diagamma function.

From (4.1.3) and (4.1.4) we get

$$\frac{1}{p}\left[\sum_{j=1}^{p}\{\psi(S_{2j}+\hat{\alpha})-\psi(\hat{\alpha})\}\right]-\log(1+\frac{n\hat{\lambda}_1}{\hat{\alpha}})=0 \qquad (4.1.5)$$

where $\hat{\alpha}$ denotes the maximum likelihood estimator of α.

The numerical solution of the equation (4.1.5) is facilitated by the use of the tables of the function

$$\lambda(r,\hat{p})=\psi(\hat{p}+r)-\psi(\hat{p}) \qquad (4.1.6)$$

for various values of $\hat{p}$ and $r = 0, 1, 2,\ldots,35$. These tables are given in the paper by Sichel (1951). Sichel has illustrated the procedure and suggested the use of the approximation

$$\lambda(r,\hat{p})\approx\log(\hat{p}+r-1)+\frac{1}{2(\hat{p}+r-1)}-\frac{1}{12(\hat{p}+r-1)^2}-\psi(\hat{p}) \qquad (4.1.7)$$

in case values of $\lambda(r, \hat{p})$ for $r > 35$ are required. It may be noted that S_{2j} and $\hat{\alpha}$ in (4.1.5) are the counterparts of r and $\hat{p}$ respectively.

Substituting the MLE $\hat{\alpha}$ of α thus obtained in (3.1.9), we obtain the estimator

$$\hat{\hat{\lambda}}_1=\frac{nS_1-\hat{\alpha}(m+n)\pm\sqrt{\{nS_1-\hat{\alpha}(m+n)\}^2+4\hat{\alpha}mn(S_1+S_2)}}{2mn}. \qquad (4.1.8)$$

4.2 Empirical Bayes Estimator for Model II

In this model the past experience is replaced by the second sample. The n observations from the second population may be treated as n samples of size one each.

According to our assumptions in this model $\{Y_j\}_{j=1}^{n}$ are i.i.d.

$NB(\alpha, \frac{\lambda_1}{\alpha})$. Hence, the joint likelihood of $\underline{Y}$ is

$$L(Y_1, Y_2,\ldots,Y_n) = \prod_{j=1}^{n} \left[\frac{\Gamma(Y_j + \alpha)}{\Gamma(\alpha)\Gamma(Y_j)} \frac{1}{(1 + \frac{\lambda_1}{\alpha})^{\alpha}} \left\{ \frac{\lambda_1/\alpha}{1 + \frac{\lambda_1}{\alpha}} \right\}^{Y_j} \right]$$

$$= \left\{ \prod_{j=1}^{n} \frac{\Gamma(Y_j + \alpha)}{Y_j!} \right\} \frac{1}{\{\Gamma(\alpha)\}^n (1 + \frac{\lambda_1}{\alpha})^{n\alpha}} \left\{ \frac{\lambda_1/\alpha}{1 + \frac{\lambda_1}{\alpha}} \right\}^{S_2} . \qquad (4.2.1)$$

Using the method of maximum likelihood we find that the equation giving the maximum likelihood estimator $\hat{\alpha}$ of α in this case is

$$\frac{1}{n} \left[\sum_{j=1}^{n} \{\psi(Y_j + \hat{\alpha}) - \psi(\hat{\alpha})\} \right] - \log(1 + \frac{\hat{\lambda}_1}{\hat{\alpha}}) = 0 \qquad (4.2.2)$$

where

$$\hat{\lambda}_1 = S_2/n.$$

Here again the method of Sichel as explained for the first model can be used to obtain a solution of α. It may be noted that Y_j and $\hat{\alpha}$ in (4.2.2) are the counterparts of r and $\hat{p}$ respectively.

The ML estimator of α thus obtained from (4.2.2) which we denote by $\tilde{\tilde{\alpha}}$, can be substituted in (3.2.4). Thus the MLE when α is assumed unknown for the second model becomes

$$\tilde{\tilde{\lambda}}_1 = \frac{S_1 - \tilde{\tilde{\alpha}}(m + n) + \sqrt{\{S_1 - \tilde{\alpha}(m + n)\}^2 + 4\tilde{\alpha}m(S_1 + S_2)}}{2m} . \qquad (4.2.3)$$

In view of the rather complicated expressions for the two estimators $\hat{\hat{\lambda}}_1$ and $\tilde{\tilde{\lambda}}_1$, the evaluation of the MSE of these two estimators proved to be too complex. The general theory of empirical Bayes approach, however, would guarantee that the MSE of $\hat{\hat{\lambda}}_1$ and $\tilde{\tilde{\lambda}}_1$ would converge to the corresponding minimum Bayes risk, as the past experience tends to infinity.

BIBLIOGRAPHY

Abramowitz, M. and Stegun, I. A. (1964). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards. Applied Mathematics Series 55.

Arbous, A. G. and Kerrich, J. E. (1951). Accident statistics and the concept of accident-proneness. Biometrics, Vol. 7: 340-432.

Bancroft, T. A. (1964). Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance. Biometrics, Vol. 20: 427-442.

Bancroft, T. A. (1972). Paper circulated at the session on "Inference procedures incorporating preliminary tests of significance" held on April 26, 1972, at the 133rd meeting of the Institute of Mathematical Statistics, (joint with ENAR), at Ames, Iowa.

Bates, G. E. and Neyman, J. (1952). Contributions to the Theory of Accident Proneness. I: An optimistic model of the correlation between light and severe accidents. University of California publication.

Bennett, B. M. (1952). Estimation of means on the basis of preliminary tests of significance. Institute of Statistical Mathematics Annals, Vol. 4: 31-43.

Bruijn, N. G. De (1958). Asymptotic methods in analysis. InterScience Publishers, Inc., New York.

Bruner, N. K. (1967). Pooling of Sample Means and Empirical Bayes Approach. Unpublished M.Sc. Thesis. Iowa State University, Ames, Iowa.

Cramer, H. (1966). Mathematical Methods of Statistics, Princeton University Press.

Erdelyi, A. (1953). Higher Transcendental Functions. Vol. 1, McGraw-Hill Book Company, Inc.

Feller, W. (1968). An Introduction to Probability Thoery and Its Applications, Vol. 1, 3rd ed., John Wiley and Sons, Inc.

Graybill, F. A. and Deal, R. B. (1959). Combining unbiased estimators. Biometrics, Vol. 15: 541-550.

Greenwood, M. and Yule, G. V. (1920). An inquiry into the nature of frequency distributions of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. J. Roy. Statist. Soc., Vol. 83: 255-279.

Hogg, R. V. (1960). On Conditional Expectations of Location Statistics. American Statistical Association Journal, Vol. 55: 714-17.

Huntsberger, D. V. (1955). A generalization of a preliminary testing procedure for pooling data. Ann. Math. Statist., Vol. 26: 734-43.

Huzurbazar, V. S. (1948). The likelihood equation, consistency and the maxima of the likelihood function. Ann. Eugen., Vol. 14: 185-200.

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

Kale, B. K. and Bancroft, T. A. (1967). Inference for some incompletely specified models involving normal approximations to discrete data. Biometrics, Vol. 23: 335-348.

Kendall, M. G. and Stuart, A. (1963). The Advanced Thoery of Statistics, Vol. 1. Hafner Publishing Company, New York.

Kitagawa, T. (1963). Estimation after preliminary tests of significance. University of California Publication in Statistics, Vol. 3: 147-86.

Mehta, J. S. and Gurland, J. (1969). Combinations of unbiased estimators of the mean which consider inequality of unknown variances. American Statistical Association Journal, Vol. 64: 1042-1055.

Mosteller, F. (1948). On pooling data. American Statistical Association Journal, Vol. 43: 231-242.

Rainville, E. D. (1960). Special Functions. The MacMillan Company, New York.

Rao, C. R. (1965). Linear Inference and Its Applications. New York, John Wiley and Sons, Inc.

Seshadri, V. (1963). Constructing uniformally better estimators. American Statistical Association Journal, Vol. 58: 172-175.

Sichel, H. S. (1951). The estimation of the parameters of a negative binomial distribution with special reference to psychological data. Psychometrica, Vol. 16: 107-127.

Zacks, S. (1966). Unbiased estimation of the common mean of two normal distributions based on small samples of equal size. American Statistical Association Journal, Vol. 61: 467-476.