GROUPING IN ITEM DEMAND PROBLEMS

by

Walter Hansbrough Carter, Jr.

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in partial fulfillment for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED: R. H. Myers
Chairman Dr. Raymond H. Myers

Boyd Harshbarger — Dr. Boyd Harshbarger

Brian Conolly — Prof. Brian W. Conolly

Donald R. Jensen — Dr. Donald R. Jensen

H. Grady Callison, Jr. — Dr. H. Grady Callison, Jr.

March, 1968

Blacksburg, Virginia

TABLE OF CONTENTS

Chapter		Page
	ACKNOWLEDGEMENTS	iv
	LIST OF TABLES	v
	LIST OF FIGURES	vi
I.	INTRODUCTION	1
	1.1 Single Distributions	1
	1.2 Mixtures of Distributions	2
	1.3 The Combination of Distributions Diff kb	5
II.	GROUPING IN A SINGLE POISSON DISTRIBUTION.	10
	2.1 Hartley's Iterative Procedure	10
	2.2 Grouping In the Poisson Distribution.	14
	2.3 Comparison of the Monte Carlo and Asymptotic Variances of the Estimator	44
III.	COMBINATIONS OF DISTRIBUTIONS DIFF kb	56
	3.1 Properties of Combinations Diff kb...	56
	3.2 Full Data Case	60
	3.3 Missing Partial Frequencies	64
	3.4 Discussion of the Asymptotic Properties of the Estimators	84
IV.	ESTIMATION OF THE PARAMETERS IN THE DISTRIBUTION OF ITEM DEMANDS	86
	4.1 The Poisson Demand Problem	86

Chapter Page

4.2 Maximum Likelihood Estimation In The Full Data Poisson Demand Problem. 89

4.3 Estimation In The Poisson Demand Problem When Only The Total Frequencies are Known 101

4.4 Monte Carlo Study of the Variances of the Maximum Likelihood Estimators 110

4.5 General Comments Concerning The Demand Problem 131

4.6 A Further Application of Diff kb Combinations of Distributions 132

BIBLIOGRAPHY 134

VITA 136

ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Dr. Raymond H. Myers for suggesting this research and for his guidance during the course of this investigation. Gratitude is also extended to Dr. Boyd Harshbarger, Head of the Statistics Department, for his assistance and encouragement. The author acknowledges with gratitude the service of Professor Brian Conolly as co-reader of this manuscript.

This research was carried out through fellowship funds available from the Department of Health, Education, and Welfare. For this support the author is indebted.

The author is also grateful for the help and cooperation he received from the staff of the Computing Center at Virginia Polytechnic Institute.

Finally, the author is indebted to Mrs. Claude Boyd Loadholt for her excellent work in typing this thesis.

The author was supported from HEW on grant EES 33-04 and EES 33-05.

LIST OF TABLES

Table		Page
4.1	Variance Comparison, b=6 (6 parameters) ..	112
4.2	Variance Comparison, b=6 (10 parameters)..	113
4.3	Variance Comparison, b=7 (10 parameters)..	114
4.4	Variance Comparison, b=8 (10 parameters)..	115
4.5	Variance Comparison, b=9 (10 parameters)..	116
4.6	Variance Comparison, b=10 (10 parameters).	117
4.7	Varying Bulk Size (10 parameters)	119
4.8	Varying Bulk Size (20 parameters)	120
4.9	Varying Bulk Size when $\underline{\theta=1.0}$ (10 parameters)	123
4.10	Varying Bulk Size when $\underline{\theta=2.0}$ (10 parameters)	124
4.11	Varying Bulk Size when $\underline{\theta=3.0}$ (10 parameters)	125
4.12	Varying Bulk Size when $\underline{\theta=4.0}$ (10 parameters)	126
4.13	Varying Bulk Size when $\underline{\theta=5.0}$ (10 parameters)	127
4.14	Bulk Size vs. Generalized Variance (10 parameters)	129
4.15	Bulk Size vs. Generalized Variance (20 parameters)	130

LIST OF FIGURES

Figure		Page
2.1	Asymptotic Per Cent Increase In Variance	48
2.2	Small Sample Per Cent Increase In Variance Due To Grouping θ=1.0	49
2.3	Small Sample Per Cent Increase In Variance Due To Grouping θ=2.0	50
2.4	Small Sample Per Cent Increase In Variance Due To Grouping θ=3.0	51
2.5	Small Sample Per Cent Increase In Variance Due To Grouping θ=4.0	52
2.6	Small Sample Per Cent Increase In Variance Due To Grouping θ=5.0	53
2.7	Small Sample Per Cent Increase In Variance Due To Grouping θ=6.0	54

CHAPTER I

INTRODUCTION

There are two ways in which the problem of missing data may arise when a random sample is observed: grouping, of which censoring is a special case, and truncation. The purpose of this dissertation is to study the estimation procedures which are appropriate when a sample from certain unique types of discrete distributions is defective due to grouping. First, we shall be concerned with grouping in a single distribution and then in a combination of distributions which resembles a mixture of these particular distributions, but which contains some important differences.

1.1 Single Distributions

The usual situation in which censoring arises occurs when a random sample has been drawn from a population and for all observations in the region, or regions, of censorship only the total frequency is known while the exact values of the censored frequencies is unknown. For this particular situation, it is assumed that in repeated sampling the total sample size is fixed whereas the frequency of observations belonging to the censored regions is an observable variable. In this dissertation we consider the more general situation of censoring which we shall designate as grouping of the sample frequencies. Here, it is assumed that certain points in the sample space are grouped in such a manner that, when sampling is performed, only the group totals of the frequencies are known.

To illustrate grouping in the single distribution case, consider the following example given by Hartley [9]. Assuming that the number of males per litter of pigs follows a binomial distribution, it is desired to estimate the sex ratio, i.e. the probability of a male pig being born. From a random sample of 106 litters of 8 pigs each, the data obtained were grouped in the following manner:

14 litters with 0, 1, or 2 males

73 litters with 3, 4, or 5 males

19 litters with 6, 7, or 8 males.

It is easily seen from this example that cases may arise in which a considerable amount of detailed information is lost due to grouping. Maximum likelihood estimation of the parameters of the assumed distributions in such cases will be considered in Chapter II.

1.2 Mixtures of Distributions

A mixture is a weighted sum of probability functions. That is, if

1) $f_i(x|\underset{\sim}{\theta}_i)$ is a probability function for all $i = 1,2,\ldots,n$

and

$$2)\quad 0 < \alpha_i \leq 1 \quad \text{for } i = 1,2,\ldots,n$$

and

$$\sum_{i=1}^{n} \alpha_i = 1,$$

then

$$G(x) = \sum_{i=1}^{n} \alpha_i f_i(x|\underline{\theta}_i) \tag{1.2.1}$$

is a mixture of the probability functions, $f_i(x|\underline{\theta}_i)$.

There are many practical applications for mixtures of distributions. Cohen [4], [5], and [6] discusses the analysis of atmospheric data and the distribution of the physical dimensions of various mass-produced items as two examples in which mixtures of distributions must be dealt with. Weiner [18] has also used them to fit the reliability curve for electronic equipment.

In some of the practical problems where mixtures have to be considered, the observed frequencies actually occur as sums of "natural" partial frequencies which are not observed and frequently it is the case that they are impossible to observe due to the experimenter's inability to classify them. More will be said concerning the classification of these partial frequencies in the following section. We will conclude this section with a practical example in which it is impossible to classify the partial frequencies.

Suppose it is desired to test the ability of two boar hogs to produce offspring possessing a certain desirable trait. For the sake of completeness, assume that the number of pigs per litter with this trait follows a binomial distribution and further that the parameter in the distribution remains constant as the litter size varies. A given number, n_1, of sows is bred to the first boar and placed in a pen and n_2 different sows are bred to the second boar and placed in a separate pen. During the course of the gestation period, the sows become inadvertently mixed. Prior to the mixing of the sows, the procedure would have been the rather simple estimation of the parameter in two separate binomial distributions. However, due to the mixing, the distribution from which one is now sampling is a mixture of the two original binomial distributions given by

$$P(x=n) = \alpha\binom{N}{n}p_1^n(1-p_1)^{N-n} + (1-\alpha)\binom{N}{n}p_2^n(1-p_2)^{N-n}, \quad (1.2.2)$$

where N is the litter size and α is the probability of a sow being bred to the first boar. A sample from this distribution would be of the following form:

Number of pigs per litter of size N possessing the desirable trait	frequency
0	n_o
1	n_1
⋮	
N	n_N

Notice that the frequencies are actually made up of the sum of two partial frequencies, one of which would correspond to the number of times a given number of pigs with the trait appeared in a litter of size N due to the first boar and the remainder being due to the second boar. It would be impossible to classify these partial frequencies with no more information than is given.

1.3 The Combination of Distributions Diff kb

To facilitate the study of the material described in this section, we make the following definition.

Definition: n Diff kb = n-kb for $n \geq kb$ where $k = 0,1,\ldots,[\frac{n}{b}]$, and where $[\frac{n}{b}]$ represents the largest integer contained in the fraction $\frac{n}{b}$.

For $n < kb$, n Diff kb is undefined.

In this section we shall introduce the combination of distributions mentioned earlier. These have properties similar to those of mixtures, but they are not a mixture as defined in the usual sense.

Consider the problem faced by the manufacturer of a consumer product when trying to decide upon the "optimum" bulk size in which to package his product. A possible solution to this problem might be to present several different bulk sizes to the public and see which it prefers by estimating the mean number of single items demanded, given that a particular bulk size or multiple of that bulk size was also demanded, the smaller mean indicating that that particular bulk size is preferable to all bulk sizes with larger conditional means. Regardless of the form of the underlying distribution, the distribution of the number of items demanded would be combinations of this distribution diff γ, where γ takes on all possible integer multiples of the bulk size which are less than or equal to the number of items demanded. In order to clarify the preceding statement, consider what happens when the bulk size is six and twenty-three items are demanded in a given period of time. There are four possible ways in which such a demand could be made:

1) twenty-three single items
2) one bulk and seventeen single items
3) two bulk and eleven single items
4) three bulk and five single items.

If each of the four cases listed above follows a different probability distribution, then the probability that twenty-three items are demanded can be expressed in the following manner:

$$P_{\underline{\theta}}(x=23) = \alpha_1 P_{\underline{\theta}_1}(x=23 \text{ diff } 0) + \alpha_2 P_{\underline{\theta}_2}(x=23 \text{ diff } 6)$$

$$+ \alpha_3 P_{\underline{\theta}_3}(x=23 \text{ diff } 12) + \alpha_4 P_{\underline{\theta}_4}(x=23 \text{ diff } 18),$$

where $\underline{\theta}_i$ is a vector of parameters associated with the probability of twenty three items being demanded by means of (i-1) bulk and α_i is the probability that a demand is made by means of (i-1) bulk.

Thus far it appears as if the distribution of the number of items demanded is a mixture of distributions as defined in section 1.2 of this chapter. However, consider what would happen if there were also a demand for twenty-four items. In addition to the four ways in which a demand for twenty-three items could be made we now have the possibility of four bulks being demanded and hence

$$P_{\underline{\theta}}(x=24) = \alpha_1 P_{\underline{\theta}_1}(x=24 \text{ diff } 0) + \alpha_2 P_{\underline{\theta}_2}(x=24 \text{ diff } 6)$$

$$+ \alpha_3 P_{\underline{\theta}_3}(x=24 \text{ diff } 12) + \alpha_4 P_{\underline{\theta}_4}(x=24 \text{ diff } 18)$$

$$+ \alpha_5 P_{\underline{\theta}_5}(x=24 \text{ diff } 24).$$

In this situation we see that a new distribution is combined with the existing ones everytime that a demand is made for a number of items which is greater than or equal to the appropriate integer multiple of the bulk size. This property makes such a combination of distributions different from a mixture of distributions as defined in section 1.2

Here it should be obvious that when sampling from this distribtuion the observed frequencies are quite naturally sums of partial frequencies. For example, suppose that a demand for seven items was made N_7 times when the bulk size is six. This frequency could be divided further into the number of times, n_7^0, which a demand for seven single items was made and the number of times, n_7^1, for which there was a demand for a bulk of six and a single item. It should also be noted that it is possible to observe these partial frequencies.

So far, in this section we have discussed the combination of distributions diff kb in the light of an example concerning item demands and we have also pointed out the difference between such a combination of distributions and a mixture of distributions as defined in section 1.2. Now we shall present the probability model for the diff kb combination of distributions in general; it is given by

$$P_{\underline{\theta}}(x=j) = \sum_{k=0}^{[\frac{j}{b}]\leq\beta<\infty} P_{\underline{\theta}_k}(x=j \text{ diff } kb|\Omega=k)P(\Omega=k), \qquad (1.3.1)$$

where Ω is a random variable whose interpretation in a practical situation will be made clear later, and b is a known constant.

It is with this distribution that we will be concerned in Chapter III and IV. In Chapter III, we shall develop procedures for obtaining the maximum likelihood estimators for the parameters involved when there are varying degrees of completeness in the observed samples. Chapter IV will be concerned with the estimators of the parameters and their properties when the conditional distributions in the model are Poisson. There will also be a discussion of the applications of this probability model when various distributions are used.

CHAPTER II

GROUPING IN A SINGLE POISSON DISTRIBUTION

As was indicated in Chapter I, a random sample may be incomplete due to grouping of the observations. In this chapter we shall review an iterative method, due to Hartley [9], for obtaining the maximum likelihood (M.L.) estimator of the parameter when grouping occurs in a one-parameter discrete distribution, give a proof that this method converges for the Poisson distribution for all starting values of the parameter, and then give the results of a study of the effects of grouping on both the small sample variance, which was obtained by means of Monte Carlo simulation, and the asymptotic variance of the estimator. An extension of Hartley's iterative method will provide us with the iterative procedure which will be used in later chapters to obtain the M.L. estimators of the parameters in the combination of distributions diff kb.

2.1 Hartley's Iterative Procedure

Various conditions may arise which make the observation of a complete sample impractical or even impossible, but permit the observance of grouped samples. For this reason, it is desirable to have a method for obtaining the M.L. estimator of the parameter in the underlying probability distribution. Given that our random sample is partitioned into G groups, the following notation is necessary to develop the iterative procedure used in obtaining the M.L. estimator of the parameter in question.

X is a discrete random variable which takes on the integral value i in the g-th group with probability $f(i,g,\theta)$.

The probability that an observation belongs to the g-th group is written as

$$F(g,\theta) = \sum_{i\varepsilon g} f(i,g,\theta), \qquad (2.1.1)$$

and N_g denotes the total observed frequency of the g-th group.

The likelihood of such a sample is

$$L = \prod_{g=1}^{G} F(g,\theta)^{N_g} . \qquad (2.1.2)$$

The maximum likelihood estimator of θ, $\hat{\theta}$, is that value of θ which maximizes (2.1.2). However, since (2.1.2) and

$$\log L = \sum_{g=1}^{G} N_g \log F(g,\theta) \qquad (2.1.3)$$

are maximum for the same value of θ, it is convenient to call log L the 'likelihood function'. If the range of the distribution is independent of the parameter, the M.L. estimator, $\hat{\theta}$, will be a solution, if any exist, of the 'likelihood equation' given by

$$\frac{d \log L}{d\theta} = 0. \tag{2.1.4}$$

For our case, the likelihood equation is

$$\frac{d \log L}{d\theta} = \sum_{g=1}^{G} \frac{N_g \frac{d}{d\theta} F(g,\theta)}{F(g,\theta)}$$

$$= \sum_{g=1}^{G} \sum_{i\epsilon g} \frac{N_g \frac{d}{d\theta} f(i,g,\theta)}{F(g,\theta)} = 0. \tag{2.1.5}$$

It is usually the case that (2.1.5) cannot be solved for θ in closed form or if a solution is obtained it is gotten by use of 'special aid tables', see Cohen [3], which would have to be compiled for every different grouping situation. The method presented here is a general one in the sense that it is applicable to all discrete grouping situations.

Proceeding with Hartley's method, we make the following definition:

$$\tilde{n}_i^g = \frac{N_g f(i,g,\theta)}{F(g,\theta)} . \tag{2.1.6}$$

$\tilde{n}_i^g$ can actually be considered as an estimator of the i-th frequency in the g-th group. Upon rearranging (2.1.6), we obtain

$$\frac{\tilde{n}_i^g}{N_g} = \frac{f(i,g,\theta)}{F(g,\theta)} \qquad (2.1.7)$$

which, when substituted into (2.1.5), yields

$$\frac{d \log L}{d\theta} = \sum_{g=1}^{G} \sum_{i \epsilon g} \frac{\tilde{n}_i^g \frac{d}{d\theta} f(i,g,\theta)}{f(i,g,\theta)} = 0 \qquad (2.1.8)$$

which is exactly the form of the likelihood equation to be solved in the full data case.

The method for obtaining the solution to the likelihood equation (2.1.8) is as follows:

1) Partition the observed total group frequencies into individual frequencies. The only restriction placed on these individual initial values is that they sum to the observed group frequencies.

2) With the estimates of the missing frequencies, solve equation (2.1.8) for θ.

3) Using the new value of θ, calculate new values of the missing frequencies by use of equation (2.1.6).

4) Repeat steps (2) and (3) until there is no change in the θ values.

The value of θ to which this process converges is the M.L. estimate. This follows when one recognizes that equation (2.1.8), by means of (2.1.6), is identical to equation (2.1.5). Hence, any solution of (2.1.8) must be a solution to (2.1.5).

The point should be made here that, assuming convergence of the iterative process, the exact likelihood equations are being solved, not an approximation to them. Very often, when sampling from a continuous distribution, the observations are grouped but the likelihood equations solved are the same as those that would be solved when no grouping is present and are, therefore, approximations to the correct equations. It is for this case that Lindley [12] and Tallis [15] have discussed grouping corrections to the solution of the approximate likelihood equations.

2.2 Grouping In The Poisson Distribution

Now we illustrate the methods of the last section with the Poisson distribution. Equation (2.1.5) becomes

$$\sum_{g=1}^{G} \sum_{i \in g} \frac{N_g \, f(i,g,\theta)[\frac{i}{\theta}-1]}{F(g,\theta)} = 0, \qquad (2.2.1)$$

where

$$f(i,g,\theta) = \frac{e^{-\theta}\theta^{i}}{i!} .$$

Upon substitution of (2.1.6), we have

$$\sum_{g=1}^{G} \sum_{i \in g} \tilde{n}_i^g \, [\frac{i}{\theta} - 1] = 0 \qquad (2.2.2)$$

which yields the solution

$$\hat{\theta} = \frac{\sum_{g=1}^{G} \sum_{i \in g} i \, \tilde{n}_i^g}{N}, \qquad (2.2.3)$$

where

$$N = \sum_{g=1}^{G} N_g .$$

Equation (2.2.3) is solved with the various estimates of the $\tilde{n}_i^g$ obtained from (2.1.6) until convergence occurs.

Clearly, this iterative procedure is one of successive approximations. We shall now state and prove a theorem due to Ford [8] which gives sufficient conditions for convergence of this iterative procedure and then shows that it converges for the Poisson case regardless of the initial guesses of the missing individual frequencies.

Theorem 2.2.1: Let Θ be a solution of the equation $\theta=s(\theta)$ and let $|s'(\theta)| < M < 1$ in the interval $|R:\Theta-h\leq\theta\leq\Theta+h|$. If θ_o is in the interval R and if $\theta_1,\theta_2,\ldots$ are found successively from the equations

$$\theta_1 = s(\theta_o),\ \theta_2 = s(\theta_1),\ \ldots,\ \theta_n = s(\theta_{n-1}) \qquad (2.2.4)$$

then

$$\lim_{n\to\infty} \theta_n = \Theta . \qquad (2.2.5)$$

<u>Proof</u>: First we show that if θ_{n-1} is in R then θ_n is in R. Since Θ is a solution to $\theta = s(\theta)$, we have

$$\Theta = s(\Theta) \; .$$

Applying the Mean Value Theorem, Taylor [16, page 70], to the second member of the following equation

$$\theta_n - \Theta = s(\theta_{n-1}) - s(\Theta) \qquad (2.2.6)$$

we have

$$s(\theta_{n-1}) - s(\Theta) = s'(\lambda_n)(\theta_{n-1} - \Theta), \qquad (2.2.7)$$

where

$$\theta_{n-1} \leq \lambda_n \leq \Theta$$

and therefore λ_n lies in R, since θ_{n-1} is assumed to be in R. From equations (2.2.6) and (2.2.7), we have

$$\begin{aligned} \left|\theta_n - \Theta\right| &= \left|s'(\lambda_n)\right| \left|\theta_{n-1} - \Theta\right| \\ &< M \left|\theta_{n-1} - \Theta\right| \\ &\leq Mh \qquad (2.2.8) \\ &< h \; . \end{aligned}$$

Hence, $\theta_n \in R$. Since we have chosen θ_o, our initial guess, to be in R, we have, by the inductive principle, that θ_1, θ_2, ... are in R.

By a repeated application of (2.2.8), we can write

$$\left|\theta_n - \Theta\right| < M\left|\theta_{n-1} - \Theta\right| < M^2\left|\theta_{n-2} - \Theta\right| < \ldots < M^n\left|\theta_o - \Theta\right| \qquad (2.2.9)$$

and if our initial guess is such that $\left|\theta_o - \Theta\right|$ is finite, we have

$$\lim_{n \to \infty} \left|\theta_n - \Theta\right| = 0 , \qquad (2.2.10)$$

since $M < 1$.

Now that we have a sufficient condition for the convergence of the process of successive approximations, let us see if Hartley's procedure for Poisson grouping satisfies it. For Hartley's procedure

$$s(\theta) = \sum_{g=1}^{G} \sum_{i \epsilon g} \frac{i \tilde{n}_i^g}{N} .$$

Now consider

$$s'(\theta) = \sum_{g=1}^{G} \sum_{i \epsilon g} \frac{i \frac{d}{d\theta} \tilde{n}_i^g}{N} . \qquad (2.2.11)$$

When the above derivative is taken, (2.2.11), as a result of (2.1.6), becomes

$$s'(\theta) = \left\{ \sum_{g=1}^{G} \sum_{i\epsilon g} i \left[\frac{N_g \frac{d}{d\theta} f(i,\theta)}{F(g,\theta)} - \frac{N_g f(i,\theta)}{F(g,\theta)} \frac{\frac{d}{d\theta} F(g,\theta)}{F(g,\theta)} \right] \right\} \div N$$

$$= \left\{ \sum_{g=1}^{G} \frac{N_g}{F(g,\theta)} \sum_{i\epsilon g} \left[i \frac{d}{d\theta} f(i,\theta) - \frac{i f(i,\theta)}{F(g,\theta)} \frac{d}{d\theta} F(g,\theta) \right] \right\} \div N. \qquad (2.2.12)$$

We define

$$\begin{aligned} {}_g\mu_1 &= \sum_{i\epsilon g} \frac{i\ f(i,\theta)}{F(g,\theta)} \\ {}_g\mu_2' &= \sum_{i\epsilon g} \frac{i^2 f(i,\theta)}{F(g,\theta)} \end{aligned} \qquad (2.2.13)$$

Now we can write

$$s'(\theta) = \left\{ \sum_g \frac{N_g}{F(g,\theta)} \sum_{i\epsilon g} i \frac{d}{d\theta} f(i,\theta) - \sum_g \frac{N_g}{F(g,\theta)} \sum_{i\epsilon g} {}_g\mu_1 \frac{d}{d\theta} f(i,\theta) \right\} \div N$$

$$= \left\{ \sum_g \frac{N_g}{F(g,\theta)} \sum_{i\epsilon g} (i - {}_g\mu_1) \frac{d}{d\theta} f(i,\theta) \right\} \div N \qquad (2.2.14)$$

but, for the case where $f(i,\theta)$ is the Poisson mass function, we have

$$\frac{d}{d\theta}f(i,\theta) = f(i,\theta)\left(\frac{i}{\theta}-1\right). \tag{2.2.15}$$

Making use of (2.2.15), we are finally able to write

$$s'(\theta) = \left\{\sum_g N_g \sum_{i\epsilon g}\left(\frac{i^2 f(i,\theta)}{\theta F(g,\theta)} - \frac{if(i,\theta)}{F(g,\theta)} - \frac{{}_g\mu_1 if(i,\theta)}{\theta F(g,\theta)} + \frac{{}_g\mu_1 f(i,\theta)}{F(g,\theta)}\right)\right\} \div N$$

$$= \left\{\sum_g \frac{N_g\ {}_g\mu_2'}{\theta} - \sum_g N_g\ {}_g\mu_1 - \sum_g \frac{N_g\ {}_g\mu_1^2}{\theta} + \sum_g N_g\ {}_g\mu_1\right\} \div N$$

$$= \sum_g \frac{N_g}{\theta}({}_g\mu_2' - {}_g\mu_1^2) \div N$$

$$= \sum_g N_g \text{ Var } g \div N\theta\ , \tag{2.2.16}$$

where Var $g = {}_g\mu_2' - {}_g\mu_1^2$ is the variance of the "sub-distribution" defined on the g-th group. From (2.2.16) we see that if Var $g < \theta$, the variance of the complete, ungrouped Poisson distribution, then $s'(\theta)$ will be less than unity and Hartley's iterative procedure will always converge for the grouped Poisson case. In the remainder of this section, we shall show that the variance of the "sub-distribution" defined on the g-th group, for certain types of groups to be defined later, is less than the variance of the total Poisson distribution. This problem, in itself, has generated considerable interest. Bowen [2] and Hayles [10] have considered it in work involving various conditional distributions. As a result of this property of the Poisson distribution, we will be able to show that Hartley's procedure, when applied to the grouped Poisson distribution, converges for any starting value of the unknown parameter.

It almost appears obvious that the variance of the distribution defined on a group would always be less than the variance of the complete distribution. Indeed, it is easy to construct counter-examples to the above statement. Consider the random variable X which takes on the three values 1,2, and 24, each with probability 1/3. The variance of this distribution is found to be 112 and 2/3. Now we consider the group which contains 1 and 24. In this group X assumes the two values equiprobably. Hence, the variance of the grouped distribution is 132 and 1/4.

We define a connected group of integers to be such that it contains every integer between and including its end points. Thus $\{3,4,6\}$ is not connected while $\{2,3,4\}$ is a connected group. In what follows, it will be shown that for the sub-distribution defined on a connected group of (k+1) Poisson variates the variance is less than the variance of the complete distribution.

To simplify some of the algebraic expressions in what follows, the following notation is introduced:

$$P_i^{i+k} = (i+1)(i+2)\ldots(i+k) = \frac{(i+k)!}{i!} . \qquad (2.2.17)$$

Hence, the probability of belonging to the g-th connected group, whose initial element is i, containing (k+1) elements can be written as

$$F(g,\theta) = P(i,\theta)\left[1+\frac{\theta}{P_i^{i+1}}+\frac{\theta^2}{P_i^{i+2}}+\ldots+\frac{\theta^k}{P_i^{i+k}}\right], \qquad (2.2.18)$$

where $P(i,\theta) = \frac{e^{-\theta}\theta^i}{i!}$. Within this group we define the following distribution:

$$P(i,\theta|g) = \frac{P(i,\theta)}{F(g,\theta)}$$

$$P(i+1,\theta|g) = \frac{\theta P(i,\theta)/P_i^{i+1}}{F(g,\theta)} \qquad (2.2.19)$$

$$\vdots$$

$$P(i+k,\theta|g) = \frac{\theta^k P(i,\theta)/P_i^{i+k}}{F(g,\theta)}$$

where $P(i+j,\theta|g)$ is the conditional probability that the random variable X takes on the value i+j in the g-th group. By use of (2.2.17), we can simplify (2.2.19) and write

$$P(i+j,\theta|g) = \frac{\theta^j P^{i+k}_{i+j}}{\sum_{h=0}^{k} \theta^h P^{i+k}_{i+h}}, \quad j=0,1,\ldots,k. \tag{2.2.20}$$

The first two non-central moments of this distribution are

$$\mu_1 = \frac{\sum_{h=0}^{k} \theta^h (i+h) P^{i+k}_{i+h}}{\sum_{h=0}^{k} \theta^h P^{i+k}_{i+h}}$$

$$\mu_2' = \frac{\sum_{h=0}^{k} \theta^h (i+h)^2 P^{i+k}_{i+h}}{\sum_{h=0}^{k} \theta^h P^{i+k}_{i+h}}$$

and hence the variance becomes

$$\mu_2 = \frac{\left(\sum_{h=0}^{k} \theta^h P^{i+k}_{i+h}\right)\left(\sum_{h=0}^{k} \theta^h (i+h)^2 P^{i+k}_{i+h}\right) - \left(\sum_{h=0}^{k} \theta^h (i+h) P^{i+k}_{i+h}\right)^2}{\left(\sum_{h=0}^{k} \theta^h P^{i+k}_{i+h}\right)^2}$$

$$= \frac{\left(\sum_{h=0}^{k} \theta^h P^{i+k}_{i+h}\right)\left(\sum_{h=0}^{k} \theta^h (i+h)^2 P^{i+k}_{i+h}\right) - \left(\sum_{h=0}^{k} \theta^{2h} (i+h)^2 (P^{i+k}_{i+h})^2\right)}{\left(\sum_{h=0}^{k} \theta^h P^{i+k}_{i+h}\right)^2}$$

$$- \frac{2 \sum\sum_{h<l=0}^{k} \theta^{h+l} (i+h)(i+l) P^{i+k}_{i+h} P^{i+k}_{i+l}}{\left(\sum_{h=0}^{k} \theta^h P^{i+k}_{i+h}\right)^2}. \tag{2.2.21}$$

Working with the numerator of (2.2.21), the coefficient of θ^j can be found for the following situations:

1) j odd and $j \leq k$

2) j even and $j \leq k$

3) j even and $k < j \leq 2(k-1)$

4) j odd and $k < j \leq 2(k-1)$.

Situation 1:

From (2.2.21), we see that the desired coefficient can be expressed as

$$\sum_{h=0}^{j} P_{i+h}^{i+k} P_{i+j-h}^{i+k} (i+j-h)^2 - 2 \sum_{h=0}^{((j-1)/2)} (i+h)(i+j-h) P_{i+h}^{i+k} P_{i+j-h}^{i+k}$$

$$= P_{i+j}^{i+k} P_{i}^{i+k} ((i+j)^2 + i^2 - 2(i)(i+j))$$

$$+ P_{i+j-1}^{i+k} P_{i+1}^{i+k} ((i+j-1)^2 + (i+1)^2 - 2(i+1)(i+j-1))$$

$$+ \ldots$$

$$+ P_{i+((j+1)/2)}^{i+k} P_{i+((j-1)/2)}^{i+k} ((i+((j+1)/2))^2$$

$$+ (i+((j-1)/2))^2 - 2(i+((j+1)/2))(i+((j-1)/2)))$$

$$= j^2 P_{i+j}^{i+k} P_{i}^{i+k} + (j-2)^2 P_{i+j-1}^{i+k} P_{i+1}^{i+k} + \ldots$$

$$+ P_{i+((j+1)/2)}^{i+k} P_{i+((j-1)/2)}^{i+k}. \qquad (2.2.22)$$

Situation 2:

Here we can express the coefficient as

$$\sum_{h=0}^{j}(i+j-h)^2 P_{i+h}^{i+k}P_{i+j-h}^{i+k} - (i+(j/2))^2(P_{i+(j/2)}^{i+k})^2$$

$$- 2\sum_{h=0}^{(j/2)-1}(i+j-h)(i+h)P_{i+h}^{i+k}P_{i+j-h}^{i+k}$$

$$= P_{i}^{i+k}P_{i+j}^{i+k}((i+j)^2+i^2-2i(i+j))$$

$$+ P_{i+1}^{i+k}P_{i+j-1}^{i+k}((i+j-1)^2+(i+1)^2-2(i+1)(i+j-1))$$

$$+ \cdots$$

$$+ P_{i+(j/2)-1}^{i+k}P_{i+(j/2)+1}^{i+k}((i+(j/2)-1)^2+(i+(j/2)+1)^2$$

$$-2(i+(j/2)-1)(i+(j/2)+1))$$

$$+ (P_{i+(j/2)}^{i+k})^2(i+(j/2))^2-(P_{i+(j/2)}^{i+k})^2(i+(j/2))^2$$

$$= j^2P_{i}^{i+k}P_{i+j}^{i+k} + (j-2)^2P_{i+1}^{i+k}P_{i+j-1}^{i+k} + \cdots$$

$$+ 4P_{i+(j/2)-1}^{i+k}P_{i+(j/2)+1}^{i+k} . \qquad (2.2.23)$$

Situation 3:

In a similar manner, the desired coefficient is found to be

$$\sum_{h=j-k}^{(j/2)-1} (j-2h)^2 P_{i+h}^{i+k} P_{i+j-h}^{i+k} \quad . \tag{2.2.24}$$

Situation 4:

Using the same methods that were employed in the previous situations, we find the desired coefficient to be

$$\sum_{h=j-k}^{((j-1)/2)} (j-2h)^2 P_{i+h}^{i+k} P_{i+j-h}^{i+k} \quad . \tag{2.2.25}$$

Now that we have the coefficient of θ^j in the numerator of (2.2.21) for all the desired combinations of relationships between j and k, we obtain the corresponding coefficients of θ^j in the denominator in a similar manner. After considerable algebra, one arrives at the following results:

j even and $j \leq k$

$$(P_{i+(j/2)}^{i+k})^2 + 2 \sum_{h=0}^{(j/2)-1} P_{i+h}^{i+k} P_{i+j-h}^{i+k} \tag{2.2.26}$$

j odd and $j \leq k$

$$2 \sum_{h=0}^{((j-1)/2)} P_{i+h}^{i+k} P_{i+j-h}^{i+k} \tag{2.2.27}$$

j even and j > k

$$(p^{i+k}_{i+(j/2)})^2 + 2\sum_{h=j-k}^{(j/2)-1} p^{i+k}_{i+h}\,p^{i+k}_{i+j-h} \tag{2.2.28}$$

j odd and j > k

$$2\sum_{h=j-k}^{((j-1)/2)} p^{i+k}_{i+h}\,p^{i+k}_{i+j-h} \quad . \tag{2.2.29}$$

It can easily be seen that the coefficients of θ^{o} and θ^{2k} in the numerator of (2.2.21) are both zero. We notice that θ^{j}, in the numerator, has a non-zero coefficient as j ranges from one to 2k-1, so we are able to factor a θ in the numerator and have the variance of the g-th group of size k+1 expressed as

$$\frac{\theta a(\theta)}{b(\theta)} , \tag{2.2.30}$$

where the exponents of θ in $a(\theta)$ range from zero to 2(k-1) and the exponents of θ in $b(\theta)$ range from zero to 2k. We shall see that these extra two positive terms in the denominator will be of no consequence in the proof that follows. Now that we have the variance expressed in the form of (2.2.30), we are able to compare the coefficient of θ^{j}, for all j, in the numerator with the corresponding coefficient in the denominator. It will be shown that the coefficient of θ^{j}, for all j, in the numerator is less than the corresponding coefficient of θ^j in the denominator and hence (2.2.21) becomes

$$\text{Var} = \theta\rho \quad \text{where } \rho < 1 , \tag{2.2.31}$$

since the sum of terms in the numerator would be less than the sum of terms in the denominator. Thus, as mentioned earlier, the extra two positive terms can be ignored since the desired result can be obtained without considering them.

In what follows, we shall consider, without loss of generality, the coefficient of θ^j where j is even and merely write down the corresponding result when j is odd without showing the algebra involved. Perhaps it is best to begin with an example in order to become acquainted with the problem and to illustrate the method involved. The coefficient of θ^6, $6 \leq k$, in the numerator of (2.2.30) is

$$49P_i^{i+k}P_{i+7}^{i+k}+25P_{i+1}^{i+k}P_{i+6}^{i+k}+9P_{i+2}^{i+k}P_{i+5}^{i+k}+P_{i+3}^{i+k}P_{i+4}^{i+k} \tag{2.2.32}$$

and in the denominator it is

$$(P_{i+3}^{i+k})^2+2P_i^{i+k}P_{i+6}^{i+k}+2P_{i+1}^{i+k}P_{i+5}^{i+k}+2P_{i+2}^{i+k}P_{i+4}^{i+k} . \tag{2.2.33}$$

We have to show that

$$(P_{i+3}^{i+k})^2+2P_i^{i+k}P_{i+6}^{i+k}+2P_{i+1}^{i+k}P_{i+5}^{i+k}+2P_{i+2}^{i+k}P_{i+4}^{i+k}$$

$$\geq 49P_i^{i+k}P_{i+7}^{i+k}+25P_{i+1}^{i+k}P_{i+6}^{i+k}+9P_{i+2}^{i+k}P_{i+5}^{i+k}+P_{i+3}^{i+k}P_{i+4}^{i+k}. \quad (2.2.34)$$

Cancelling $((i+k)!)^2$, (2.2.34) becomes

$$\frac{1}{(i+3)!(i+3)!}+\frac{2}{i!(i+6)!}+\frac{2}{(i+1)!(i+5)!}+\frac{2}{(i+2)!(i+4)!}$$

$$\geq \frac{49}{i!(i+7)!}+\frac{25}{(i+1)!(i+6)!}+\frac{9}{(i+2)!(i+5)!}$$

$$+\frac{1}{(i+3)!(i+4)!} \quad (2.2.35)$$

which when multiplied by $i!(i+6)!$ simplifies to

$$\frac{(i+6)(i+5)(i+4)}{(i+3)(i+2)(i+1)}+2+\frac{2(i+6)}{(i+1)}+\frac{2(i+6)(i+5)}{(i+2)(i+1)}$$

$$\geq \frac{49}{i+7}+\frac{25}{i+1}+\frac{9(i+6)}{(i+2)(i+1)}+\frac{(i+6)(i+5)}{(i+3)(i+2)(i+1)}. \quad (2.2.36)$$

Equation (2.2.36) reduces to

$$(i+7)(i+6)(i+5)(i+4)+2(i+7)(i+3)(i+2)(i+1)+2(i+7)(i+6)(i+3)(i+2)$$

$$+2(i+7)(i+6)(i+3) \geq 49(i+3)(i+2)(i+1)+25(i+7)(i+3)(i+2)$$

$$+9(i+7)(i+6)(i+3)+(i+7)(i+6)(i+5). \quad (2.2.37)$$

Note carefully the manner in which these terms are factored. Upon factoring (2.2.37), we have

$$(i+7)(i+6)(i+5)[\overbrace{(i+4)+2(i+3)-1}^{3(i+3)}]+(i+7)(i+6)(i+3)[2(i+2)-9]$$

$$+(i+7)(i+3)(i+2)[2(i+1)-25]-49(i+3)(i+2)(i+4) \geq 0 \qquad (2.2.38)$$

which, in turn, can be written as

$$(i+7)(i+6)(i+3)[\overbrace{3(i+5)+2i-5}^{5(i+2)}]+(i+7)(i+3)(i+2)[2(i+1)-25]$$

$$-49(i+3)(i+2)(i+1) \geq 0. \qquad (2.2.39)$$

Equation (2.2.39)factors to

$$(i+7)(i+3)(i+2)[\overbrace{5(i+6)+2(i+1)-25}^{7(i+1)}]-49(i+3)(i+2)(i+1)$$

$$\geq 0. \qquad (2.2.40)$$

Finally, we can write

$$7i(i+1)(i+2)(i+3) \geq 0. \qquad (2.2.41)$$

The above inequality, (2.2.41) is seen to be true for all $i \geq 0$. Hence, the initial inequality, (2.2.34), is true and the coefficient of θ^6 in the numerator of (2.2.30) is indeed less than the corresponding coefficient in the denominator.

Now we show that for $j < k$ and j even the coefficient of θ^j in $a(\theta)$ is less than or equal to that in $b(\theta)$, i.e.

$$P^{i+k}_{i+(j/2)}P^{i+k}_{i+(j/2)}+2\sum_{h=0}^{(j/2)-1}P^{i+k}_{i+h}P^{i+k}_{i+j-h}$$

$$-\sum_{h=0}^{j/2}((j+1)-2h)^2 P^{i+k}_{i+h}P^{i+k}_{i+(j+1)-h} \geq 0. \qquad (2.2.42)$$

Notice that each term in the above expression has $[(i+k)!]^2$ in it, so we can write

$$P^{1}_{i+(j/2)}P^{1}_{i+(j/2)}+2\sum_{h=0}^{(j/2)-1}P^{1}_{i+h}P^{1}_{i+j-h}$$

$$-\sum_{h=0}^{j/2}[(j+1)-2h]^2 P^{1}_{i+h}P^{1}_{i+(j+1)-h} \geq 0. \qquad (2.2.43)$$

After multiplying (2.2.43) by $i!(i+j)!$, we have

$$\frac{(i+j)(i+j-1)\ldots(i+(j/2)+1)}{(i+(j/2)(i+(j/2)-1)\ldots(i+1)} + 2 + \frac{2(i+j)}{i+1} + \frac{2(i+j)(i+j-1)}{(i+2)(i+1)} + \ldots$$

$$+ \frac{2(i+j)(i+j-1)(i+j-2)\ldots(i+j-p+1)}{(i+p)(i+p-1)\ldots(i+1)} + \ldots$$

$$+ \frac{2(i+j)(i+j-1)\ldots(i+(j/2)+2)}{(i+(j/2)-1)(i+(j/2)-2)\ldots(i+1)} - \frac{(j+1)^2}{i+j+1} - \frac{(j-1)^2}{i+1}$$

$$- \frac{(j-3)^2(i+j)}{(i+2)(i+1)} - \ldots - \frac{(j+1-2s)^2(i+j)(i+j-1)\ldots(i+j+2-s)}{(i+s)(i+s-1)\ldots(i+1)}$$

$$- \ldots - \frac{(i+j)(i+j-1)(i+j-2)\ldots(i+(j/2)+2)}{(i+(j/2))(i+(j/2)-1)\ldots(i+1)}, \qquad (2.2.44)$$

where

$$0 \leq p \leq (j/2)-1$$

and

$$0 \leq s \leq (j/2) .$$

Now multiplying thru by

$$(i+j+1)(i+(j/2))(i+(j/2)-1)\dots(i+1),$$

we obtain

$$(i+j+1)(i+j)(i+j-1)\dots(i+(j/2)+1)+2(i+j+1)(i+(j/2))(i+(j/2)-1)$$

$$\dots(i+1)+2(i+j+1)(i+j)(i+(j/2))(i+(j/2)-1)\dots(i+2)$$

$$+2(i+j+1)(i+j)(i+j-1)(i+(j/2))(i+(j/2)-1)\dots(i+3)+\dots$$

$$+2(i+j+1)(i+j)\dots(i+j-p+1)(i+(j/2))(i+(j/2)-1)\dots(i+p+1)$$

$$+\dots+2(i+j+1)(i+j)\dots(i+(j/2)+2)(i+(j/2))$$

$$-(j+1)^2(i+(j/2))(i+(j/2)-1)\dots(i+1)$$

$$-(j-1)^2(i+j+1)(i+(j/2))(i+(j/2)-1)\dots(i+2)$$

$$-(j-3)^2(i+j+1)(i+j)(i+(j/2))(i+(j/2)-1)\dots(i+3)-\dots$$

$$-(j+1-2s)^2(i+j+1)(i+j)\dots(i+j+2-s)(i+(j/2))(i+(j/2)-1)$$

$$\dots(i+s+1)-\dots-(i+j+1)(i+j)(i+j-1)\dots(i+(j/2)+2). \quad (2.2.45)$$

To factor (2.2.45), we combine the last negative term with the first and last positive terms, the next to last negative term and the next to last positive term, etc., until the only term left is the first negative term by itself. Hence, (2.2.45) becomes

$$(i+j+1)(i+j)(i+j-1)\ldots(i+(j/2)+2)(\overbrace{i+(j/2)+1+2i+j-1}^{3(i+(j/2))})$$

$$+(i+j+1)(i+j)\ldots(i+(j/2)+3)(i+(j/2))(2i+j-2-9)$$

$$+(i+j+1)(i+j)\ldots(i+(j/2)+4)(i+(j/2))(i+(j/2)-1)(2i+j-4-25)$$

$$+\ldots+(i+j+1)(i+j)\ldots(i+(j/2)+r+2)(i+(j/2))(i+(j/2)-1)$$

$$\ldots(i+(j/2)-r+1)(2(i+(j/2)-r)-(2r+1)^2)+\ldots$$

$$+(i+j+1)(i+j)(i+(j/2))(i+(j/2)-1)\ldots(i+3)(2i+4-(j-3)^2)$$

$$+(i+j+1)(i+(j/2))(i+(j/2)-1)\ldots(i+2)(2i+2-(j-1)^2)$$

$$-(j+1)^2(i+(j/2))(i+(j/2)-1)\ldots(i+1). \qquad (2.2.46)$$

The general term is obtained by letting $p=(j/2)-(r+1)$ and $s=(j/2)-r$ in (2.2.45). The variable r is used to count back from the last positive and negative terms in (2.2.45), e.g. when $r=2$ we are combining the second from the last negative term in (2.2.45) to the second from last positive term.

We are now able to combine the first two terms of (2.2.46). Doing so, we have

$$(i+j+1)(i+j)(i+j-1)$$

$$\ldots(i+(j/2)+3)(i+(j/2))(3i+\overbrace{(3j/2)+6+2i+j-1}^{5(i+(j/2)-1)}1)$$

$$+(i+j+1)(i+j)\ldots(i+(j/2)+4)(i+(j/2)(i+(j/2)-1)(2i+j-29)$$

$$+\ldots+(i+j+1)(i+j)\ldots(i+(j/2)+r+2)(i+(j/2))(i+(j/2)-1)$$

$$\ldots(i+(j/2)-r+1)(2(i+(j/2)-r)-(2r+1)^2)+\ldots$$

$$+(i+j+1)(i+j)(i+(j/2)(i+(j/2)-1)\ldots(i+3)(2i+4-(j-3)^2)$$

$$+(i+j+1)(i+(j/2))(i+(j/2)-1)\ldots(i+2)(2i+2-(j-1)^2)$$

$$-(j+1)^2(i+(j/2))(i+(j/2)-1)\ldots(i+1). \qquad (2.2.47)$$

Combing the first two terms of (2.2.47) yields

$$(i+j+1)(i+j)$$

$$\ldots(i+(j/2)+4)(i+(j/2))(i+(j/2)-1)(5i+\overbrace{(5j/2)+15+2i+j}^{7(i+(j/2)-2)}-29)$$

$$+\ldots+(i+j+1)(i+j)\ldots(i+(j/2)+r+2)(i+(j/2))(i+(j/2)-1)$$

$$\ldots(i+(j/2)-r+1)(2(i+(j/2)-r)-(2r+1)^2)+\ldots$$

$$+(i+j+1)(i+j)(i+(j/2))(i+(j/2)-1)\ldots(i+3)(2i+4-(j-3)^2)$$

$$+(i+j+1)(i+(j/2))(i+(j/2)-1)\ldots(i+2)(2i+2-(j-1)^2)$$

$$-(j+1)^2(i+(j/2))(i+(j/2)-1)\ldots(i+1). \qquad (2.2.48)$$

In equations (2.2.46), (2.2.47), and (2.2.48), we see that the coefficient of the first term was initially equal to 3 and has increased by 2 after each step. We notice that after the initial combining of terms there will be $((j/2)-1)$ others each adding a factor of 2 to the previous one. Hence, the final coefficient will be

$$3+2((j/2)-1) = j+1. \qquad (2.2.49)$$

In general, if $r=n$ the coefficient of the first term will be $2n+1$.

To complete the proof of (2.2.42) we must obtain a general term in the above factoring. It is, where $r=n$,

$$(2n+1)(i+j+1)(i+j)\ldots(i+(j/2)+n+1)(i+(j/2))(i+(j/2)-1)$$
$$\ldots(i+(j/2)-(n-1)). \quad (2.2.50)$$

That (2.2.50) is the general term follows from the following inductive proof. When $n=1$ the term is

$$3(i+j+1)(i+j)\ldots(i+(j/2)+2)(i+(j/2)). \qquad (2.2.51)$$

This is exactly the term that was calculated in equation (2.2.46). Now we assume (2.2.50) to be true when $n=c$ and write (2.2.50) as

$$(i+j+1)(i+j)\ldots(i+(j/2)+c+1)(i+(j/2))(i+(j/2)-1)\,(2c+1)$$
$$\ldots(i+(j/2)-(c-1)). \qquad (2.2.52)$$

Now, following the factoring procedure, we see that (2.2.52) will be combined with

$$(i+j+1)(i+j)\ldots(i+(j/2)+c+2)(i+(j/2))(i+(j/2)-1)$$

$$\ldots(i+(j/2)-c+1)(2(i+(j/2)-c)-(2c+1)^2); \qquad (2.2.53)$$

the expression is obtained from (2.2.46) by letting $r=c$. Upon combining (2.2.52) and (2.2.53), we are able to write that the resulting term is

$$(2c+3)(i+j+1)(i+j)(i+j-1)\ldots(i+(j/2)+c+2)(i+(j/2))(i+(j/2)-1)$$

$$\ldots(i+(j/2)-(c-1))(i+(j/2)-c), \qquad (2.2.54)$$

which is of the desired form. Hence, we are able to conclude that (2.2.50) is the correct general term.

From equation (2.2.49), we have that the coefficient of the last term is $j+1$. Looking at (2.2.50) we see that n must be $(j/2)$ so the result of combining the last two positive terms is

$$(j+1)(i+j+1)(i+(j/2))(i+(j/2)-1)\ldots(i+1), \qquad (2.2.55)$$

which must now be combined with the negative term

$$-(j+1)^2(i+(j/2))(i+(j/2)-1)\ldots(i+1). \qquad (2.2.56)$$

The final result is

$$(j+1)i(i+(j/2))(i+(j/2)-1)\ldots(i+1) \geq 0, \qquad (2.2.57)$$

which is seen to be true for all $i \geq 0$. Hence, (2.2.42) is true and the coefficient of θ^j in $a(\theta)$ of (2.2.30) is less than the corresponding coefficient in $b(\theta)$.

Now we consider the case where $j=k$ and j is even. We would like to show that

$$(P^{i+k}_{i+(k/2)})^2 + 2\sum_{h=0}^{(k/2)-1} P^{i+k}_{i+h} P^{i+k}_{i+k-h}$$

$$- \sum_{h=1}^{(k/2)} (k+1-2h)^2 P^{i+k}_{i+h} P^{i+k}_{i+(k+1)-h} \geq 0. \qquad (2.2.58)$$

Recall that the coefficient of θ^k in $a(\theta)$ is the coefficient of θ^{k+1} in the numerator of the expression for the variance; hence the $(k+1)$ appears in the negative term above. Cancelling $((i+k)!)^2$ and multiplying (2.2.58) by $i!(i+k)!$ leaves

$$\frac{(i+k)(i+k-1)\ldots(i+(k/2)+1)}{(i+(k/2))(i+(k/2)-1)\ldots(i+1)} + 2 + \frac{2(i+k)}{i+1} + \ldots$$

$$+ \frac{2(i+k)(i+k-1)\ldots(i+k-p+1)}{(i+p)(i+p-1)\ldots(i+1)} - \frac{(k-1)^2}{i+1} - \frac{(i+k)(k-3)^2}{(i+2)(i+1)}$$

$$- \frac{(i+k)(i+k-1)(k-5)^2}{(i+3)(i+2)(i+1)} - \ldots - \frac{(k+1-2s)^2(i+k)(i+k-1)\ldots(i+k-s+2)}{(i+s)(i+s-1)\ldots(i+1)}$$

$$- \ldots - \frac{(i+k)(i+k-1)\ldots(i+(k/2)+2)}{(i+(k/2))(i+(k/2)-1)\ldots(i+1)}, \qquad (2.2.59)$$

where

$$0 \le p \le (k/2)-1$$

and

$$1 \le s \le (k/2).$$

After multiplying (2.2.59) by $(i+(k/2))(i+(k/2)-1)$ $\ldots(i+1)$ we have

$$\begin{aligned}
&(i+k)(i+k-1)\ldots(i+(k/2)+1)+2(i+(k/2))(i+(k/2)-1)\ldots(i+1)\\
&+2(i+k)(i+(k/2))(i+(k/2)-1)\ldots(i+2)\\
&+2(i+k)(i+k-1)(i+(k/2))(i+(k/2)-1)\ldots(i+3)\\
&+2(i+k)(i+k-1)(i+k-2)(i+(k/2))(i+(k/2)-1)\ldots(i+4)+\ldots\\
&+2(i+k)(i+k-1)\ldots(i+k-p+1)(i+(k/2))(i+(k/2)-1)\ldots(i+p+1)\\
&+\ldots+2(i+k)(i+k-1)\ldots(i+(k/2)+2)(i+(k/2))\\
&-(k-1)^2(i+(k/2))(i+(k/2)-1)\ldots(i+2)\\
&-(k-3)^2(i+k)(i+(k/2))(i+(k/2)-1)\ldots(i+3)\\
&-(k-5)^2(i+k)(i+k-1)(i+(k/2))(i+(k/2)-1)\ldots(i+4)-\ldots\\
&-(k+1-2s)^2(i+k)(i+k-1)\ldots(i+k-s+2)(i+(k/2))(i+(k/2)-1)\\
&\ldots(i+s-1)-\ldots-(i+k)(i+k-1)\ldots(i+(k/2)+2).
\end{aligned} \tag{2.2.60}$$

To factor (2.2.60), we use the same method that was employed previously. We combine the last negative term with the first and last positive terms, the next to last negative term with the next to last positive term, etc., until the first negative term is combined with the second positive term. In this case, unlike the case where $j < k$ and j even, there is no term left by itself. Factoring yields

$$(i+k)(i+k-1)\dots(i+(k/2)+2)[\overbrace{i+(k/2)+1+2i+k-1}^{3(i+(k/2))}]$$

$$+(i+k)(i+k-1)\dots(i+(k/2)+3)(i+(k/2))[2i+k-2-9]$$

$$+(i+k)(i+k-1)\dots(i+(k/2)+4)(i+(k/2))(i+(k/2)-1)[2i+k-4-25]$$

$$+(i+k)(i+k-1)\dots(i+(k/2)+r+2)(i+(k/2))(i+(k/2)-1)$$

$$\dots(i+(k/2)-r+1)[2(i+(k/2)-r)-(2r+1)^2]+\dots$$

$$+(i+k)(i+(k/2))(i+(k/2)-1)\dots(i+3)[2i+4-(k-3)^2]$$

$$+(i+(k/2))(i+(k/2)-1)\dots(i+2)[2i+2-(k-1)^2]. \qquad (2.2.61)$$

With the exception of the term $(i+k+1)$, we notice that (2.2.61) is identical to the corresponding equations for the case j even and $j < k$. By the same inductive proof we have that the general term in the factoring of (2.2.61) will be

$$(2n+1)(i+k)(i+k-1)\dots(i+(k/2)+n+1)(i+(k/2))(i+(k/2)-1)$$

$$\dots(i+(k/2)-(n-1)), \qquad (2.2.62)$$

where

$$1 \leq n \leq (k/2)-1.$$

Letting $n = (k/2)-1$ in equation (2.2.62) yields

$$(k-1)(i+k)(i+(k/2))(i+(k/2)-1)\ldots(i+2)\cdot \qquad (2.2.63)$$

This term must be combined with the last term appearing in (2.2.61). This final factoring yields

$$(k+1)(i+(k/2))(i+(k/2)-1)\ldots(i+1), \qquad (2.2.64)$$

which we see is greater than zero for all non-negative i.

Finally we consider the case for which j is even and $j > k$. We have to show that

$$(P^{i+k}_{i+(j/2)})^2 + 2\sum_{h=j-k}^{(j/2)-1} P^{i+k}_{i+h}P^{i+k}_{i+j-h}$$

$$- \sum_{h=j+1-k}^{(j/2)} (j-2h+1)^2 P^{i+k}_{i+h}P^{i+k}_{i+j+1-h} \geq 0. \qquad (2.2.65)$$

Many of the steps taken here are identical with those taken in the previous two cases; whenever this is the case only the results will be presented.

After dividing by $((i+k)!)^2$ and multiplying by $(i+k)!(i+j-k)!$ we are left with

$$\frac{(i+k)(i+k-1)\ldots(i+(j/2)+1)}{(i+(j/2))(i+(j/2)-1)\ldots(i+j-k+1)} + 2 + \frac{2(i+k)}{i+j-k+1} + \ldots$$

$$+ \frac{2(i+k)(i+k-1)\ldots(i+k-p+1)}{(i+j-k+p)(i+j-k+p-1)\ldots(i+j-k+1)} + \ldots$$

$$+ \frac{2(i+k)(i+k-1)\ldots(i+(j/2)+4)}{(i+(j/2)-3)(i+(j/2)-4)\ldots(i+j+1-k)}$$

$$+ \frac{2(i+k)(i+k-1)\ldots(i+(j/2)+3)}{(i+(j/2)-2)(i+(j/2)-3)\ldots(i+j+1-k)}$$

$$+ \frac{2(i+k)(i+k-1)\ldots(i+(j/2)+2)}{(i+(j/2)-1)(i+(j/2)-2)\ldots(i+j-k+1)} - \frac{(2k-j-1)^2}{(i+j+1-k)}$$

$$- \frac{(2k-j-3)^2(i+k)}{(i+j+2-k)(i+j+1-k)} - \ldots - \frac{(2k-j-2s-1)(i+k)(i+k-1)\ldots(i+k-s+1)}{(i+j+1-k+s)(i+j-k+s)\ldots(i+j-k+1)}$$

$$- \ldots - \frac{(i+k)(i+k-1)\ldots(i+(j/2)+2)}{(i+(j/2))(i+(j/2)-1)\ldots(i+j+1-k)}, \qquad (2.2.66)$$

where

$$1 \leq p \leq k-(j/2)-1$$

and

$$1 \leq s \leq k-(j/2)-1.$$

Multiplying (2.2.66) by $(i+(j/2))(i+(j/2)-1)\ldots(i+j+1-k)$ we obtain

$$(i+k)(i+k-1)\ldots(i+(j/2)+1)+2(i+(j/2))(i+(j/2)-1)\ldots(i+j+1-k)$$

$$+2(i+k)(i+(j/2))(i+(j/2)-1)\ldots(i+(j/2)+k+2)+\ldots$$

$$+2(i+k)(i+k-1)\ldots(i+k-p+1)(i+(j/2))(i+(j/2)-1)\ldots(i+j-k+p+1)$$

$$+\ldots+2(i+k)(i+k-1)\ldots(i+(j/2)+3)(i+(j/2))(i+(j/2)-1)$$

$$+2(i+k)(i+k-1)\ldots(i+(j/2)+2)(i+(j/2))$$

$$-(2k-j-1)^2(i+(j/2))(i+(j/2)-1)\ldots(i+j-k+2)$$

$$-(2k-j-3)^2(i+k)(i+(j/2))(i+(j/2)-1)\ldots(i+j-k+3)-\ldots$$

$$-(2k-j-2s-1)^2(i+k)(i+k-1)\ldots(i+k-s+1)(i+(j/2))(i+(j/2)-1)$$

$$\ldots(i+j+2-k+s)-\ldots-9(i+k)(i+k-1)\ldots(i+(j/2)+3)(i+(j/2))$$

$$-(i+k)(i+k-1)\ldots(i+(j/2)+2). \qquad (2.2.67)$$

We now factor (2.2.67) precisely as we have done in the previous two cases and obtain

$$(i+k)(i+k-1)\ldots(i+(j/2)+2)(\overbrace{i+(j/2)+1+2i+j-1}^{3(i+(j/2))})+(i+k)(i+k-1)$$

$$\ldots(i+(j/2)+3)(i+(j/2))(2i+j-2-9)+(i+k)(i+k-1)$$

$$\ldots(i+(j/2)+4)(i+(j/2))(i+(j/2)-1)(2i+j-4-25)+\ldots$$

$$+(i+k)(i+k-1)\ldots(i+(j/2)+r+2)(i+(j/2))(i+(j/2)-1)$$

$$\ldots(i+(j/2)-r+1)(2(i+(j/2)-r)-(2r+1)^2)+\ldots$$

$$+(i+k)(i+(j/2))(i+(j/2)-1)\dots(i+j+3-k)(2i+j+4-2k+j-(2k-j-3)^2)$$

$$+(i+(j/2))(i+(j/2)-1)\dots(i+j+2-k)(2i+2j+2-2k-(2k-j-1)^2), \quad (2.2.68)$$

where r is a counting variable as explained in connection with equation (2.2.46). As a result of the inductive proof given earlier, we have that when r=n the first term in the resulting combination is given by

$$(2n+1)(i+k)(i+k-1)\dots(i+(j/2)+n+1)(i+(j/2))(i+(j/2)-1)\dots(i+(j/2)-(n-1)). \quad (2.2.69)$$

We notice that after the first combination, the coefficient of the leading term is 3 and there will be (k-(j/2)-1) additional combinations each contributing a factor of 2 before (2.2.65) is completely factored. Hence, the coefficient of the final term will be

$$2(k-(j/2)-1)+3 = 2k-j+1. \quad (2.2.70)$$

The final step in the factoring of (2.2.65) is

$$(2k-j-1)(i+k)(i+(j/2))(i+(j/2)-1)\dots(i+j-k+2)$$

$$+(i+(j/2))(i+(j/2)-1)\dots(i+j+2-k)(2i+2j+2-2k-(2k-j-1)^2), \quad (2.2.71)$$

which factors to

$$(i+(j/2))(i+(j/2)-1)$$

$$\ldots(i+j+2-k)[(2k-j-1)(i+k)+2(i+j-k+1)-(2k-j-1)^2]$$

$$= (2k-j+1)(i+(j/2))(i+(j/2)-1)(i+(j/2)-2)$$

$$\ldots(i+j-k+2)(i+j-k+1) \qquad (2.2.72)$$

and we see that this quantity is greater than zero for all non-negative i.

The differences between the coefficients of θ^j, j odd, in $a(\theta)$ and $b(\theta)$ are

for <u>$j < k$</u>

$$(k+1)(i)(i+1)\ldots(i+((j-1)/2)) \qquad (2.2.73)$$

for <u>$j = k$</u>

$$(k+1)(i+1)(i+2)\ldots(i+((k-1)/2)) \qquad (2.2.74)$$

for <u>$j > k$</u>

$$(2k-j+1)(i+((j-1)/2))(i+((j-3)/2))$$

$$\ldots(i+j+2-k)(i+j+1-k). \qquad (2.2.75)$$

These terms are all seen to be greater than or equal to zero for all non-negative i. Hence, we have that for each term in $a(\theta)$ of (2.2.30) the corresponding term in $b(\theta)$ is at least as large ($j<k$) or larger ($j\geq k$) and therefore ρ in

equation (2.2.31) is indeed less than unity. Thus we have shown that for a connected group of Poisson variates the variance of the sub-distribution defined on the group is less than the variance of the complete distribution.

As a result of this property of the Poisson distribution, equation (2.2.16) becomes

$$s'(\theta) = \frac{\rho\theta\sum_g N_g}{N\theta}$$

$$= \rho < 1 \qquad (2.2.76)$$

and for the grouped (or censored) Poisson distribution, Hartley's iterative procedure always converges regardless of the initial values of the missing frequencies. The only restriction on the initial guesses of the missing partial frequencies is that these guesses sum to the observed total frequencies for each group.

2.3 Comparison of the Monte Carlo and Asymptotic Variances of the Estimator

In this section a comparison is made of the asymptotic and Monte Carlo variances of the estimator, $\hat{\theta}$, of the parameter, θ, in the grouped Poisson distribution. For the ungrouped Poisson, the variance of $\hat{\theta}$ is given by

$$\frac{\theta}{N} \qquad (2.3.1)$$

and for the grouped case, the asymptotic variance is given by

$$-\left\{E\frac{d^2\log L}{d\theta^2}\right\}^{-1} = -\left\{N\left(\sum_g \frac{d^2F(g,\theta)}{d\theta^2} - \sum_g \frac{(\frac{d}{d\theta}F(g,\theta))^2}{F(g,\theta)}\right)\right\}^{-1}, \quad (2.3.2)$$

where

$$F(g,\theta) = \sum_{i\epsilon g} \frac{e^{-\theta}\theta^i}{i!} .$$

To determine how the size of the group affects the variance of $\hat{\theta}$, we calculate the per cent increase in variance due to grouping; this is given by

$$\left[\frac{\dfrac{-\left\{\sum_g \frac{d^2}{d\theta^2}F(g,\theta) - \sum_g \frac{(\frac{d}{d\theta}F(g,\theta))^2}{F(g,\theta)}\right\}^{-1} - \theta}{N}}{\dfrac{\theta}{N}}\right] \times 100 \quad (2.3.3)$$

but

$$\sum_g \frac{d^2}{d\theta^2}F(g,\theta) = \frac{d^2}{d\theta^2}\sum_g F(g,\theta) = \frac{d^2}{d\theta^2}(1) = 0. \quad (2.3.4)$$

Upon simplification (2.3.3) becomes

$$\text{Per Cent Increase} = \left[\frac{\left\{\sum_g \frac{(\frac{d}{d\theta}F(g,\theta))^2}{F(g,\theta)}\right\}^{-1} - \theta}{\theta}\right] \times 100. \quad (2.3.5)$$

Note that the above expression for per cent increase in variance is independent of sample size.

If we define the efficiency (Lindley [12]) of the estimator of a parameter from grouped data as

$$\text{Eff} = \frac{\text{Var } \hat{\theta} \text{ ungrouped}}{\text{Var } \hat{\theta} \text{ grouped}}, \qquad (2.3.6)$$

we have that

$$\text{Per Cent Increase} = \left(\frac{1}{\text{Eff}} - 1\right) \times 100. \qquad (2.3.7)$$

The variance of $\hat{\theta}$ and the per cent increase in the variance of $\hat{\theta}$ due to grouping were investigated for small samples by Monte Carlo simulation on an IBM 7040 computer. The purpose of this study was threefold:

1) To learn how changes in the group size affect the variance of $\hat{\theta}$.

2) To learn how changes in θ affect the variance of $\hat{\theta}$.

3) To see how large the sample size must be in order that the asymptotic variance of $\hat{\theta}$ be adequate.

For a given value of θ and a given group size, 1000 random samples of a pre-determined size were generated from the grouped Poisson distribution in ten groups of one hundred samples. After each sample was generated, the estimate of θ, $\hat{\theta}$, was obtained by Hartley's method and stored in the computer. From each group of one hundred values of $\hat{\theta}$, the variance was calculated and then these ten values of the variance were averaged and a standard error of this mean was obtained. It should be noted here that whenever the sample size was large enough to expect agreement between the Monte Carlo and asymptotic values of the variance of $\hat{\theta}$, the Monte Carlo value was within at most two standard errors of the asymptotic value. After the Monte Carlo value for the variance of $\hat{\theta}$ was calculated, it was compared to the asymptotic value, to the value of var $\hat{\theta}$ when there was no grouping, and finally the Monte Carlo value of the per cent increase in the variance of $\hat{\theta}$ due to grouping was compared to the asymptotic value. This study was made using various values of the parameter with different group sizes and sample sizes -- two of these variables being held fixed while the third was allowed to vary. The results are presented in figures (2.1)-(2.7). It can be seen from the graphs that as the parameter increases, the asymptotic and Monte Carlo values of the per cent increase in variance due to grouping agree for smaller values of the sample size. From these figures it also becomes evident that the per cent increase in the variance due to grouping varies directly with some function of the group size and inversely with a function of the parameter size, a result which is explained by the fact that for the Poisson distribution the variance also equals θ. If θ is large, the

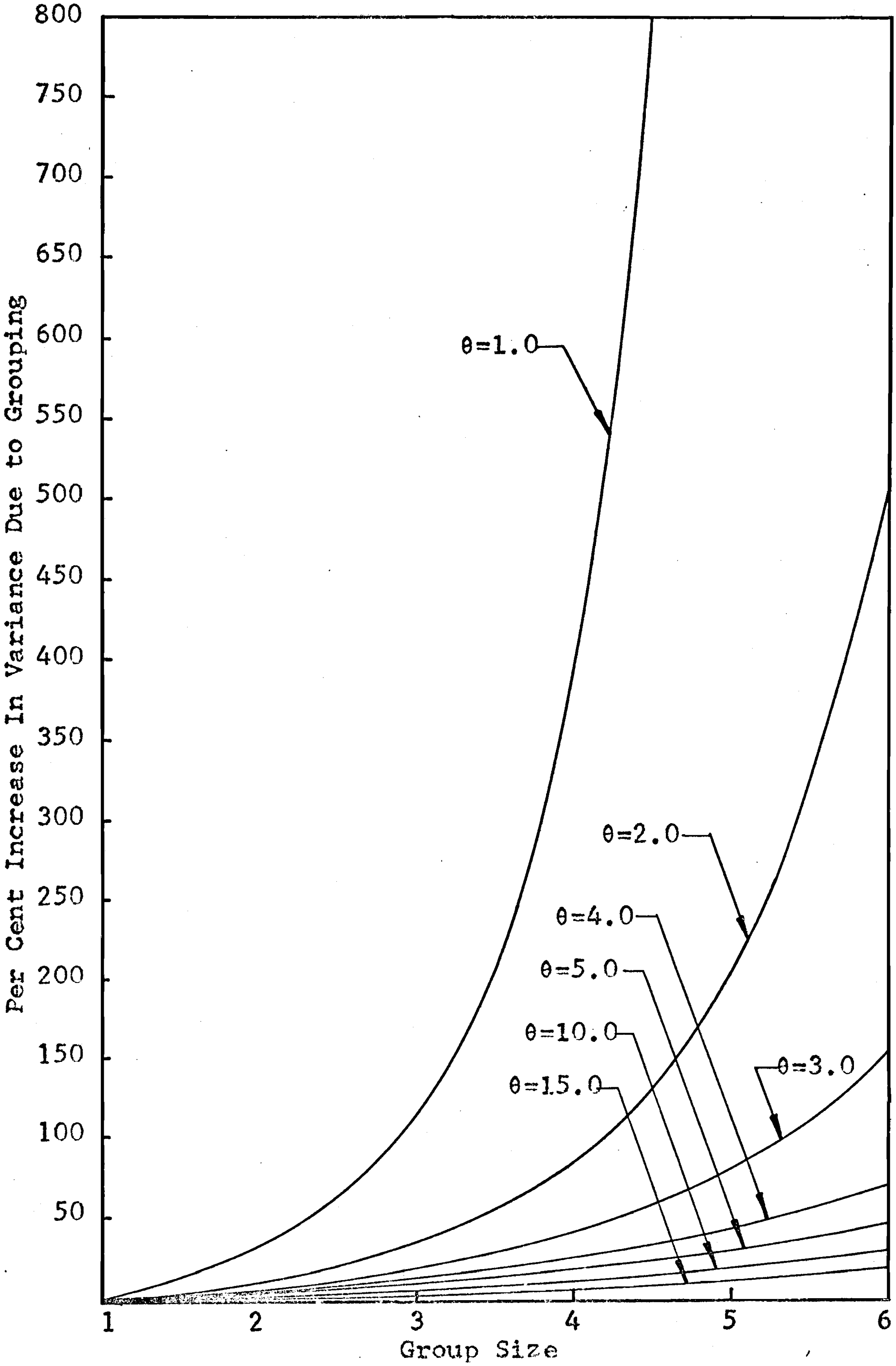

FIGURE 2.1. Asymptotic Per Cent Increase In Variance

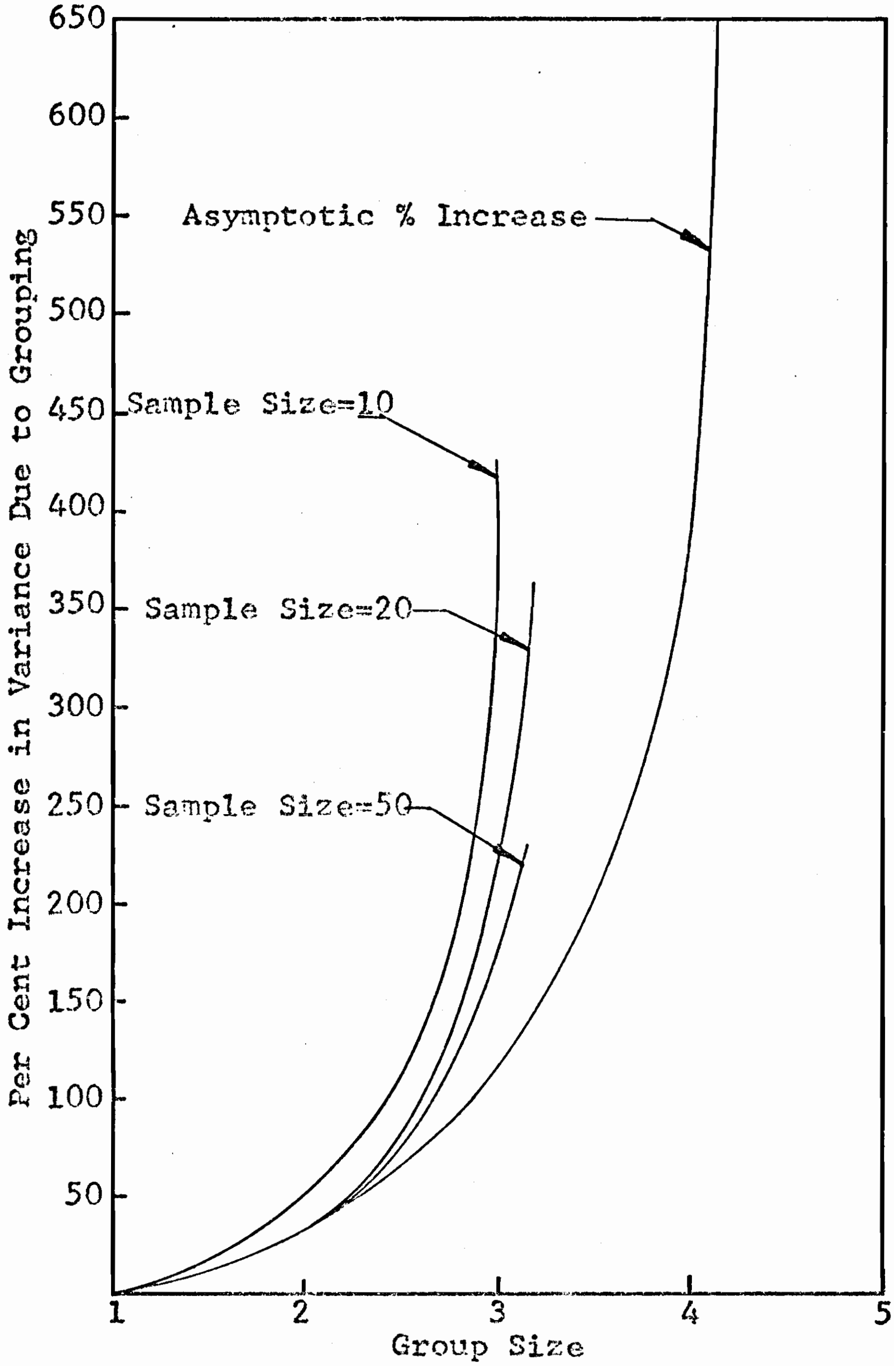

FIGURE 2.2. Small Sample Per Cent Increase In Variance Due To Grouping $\theta=1.0$

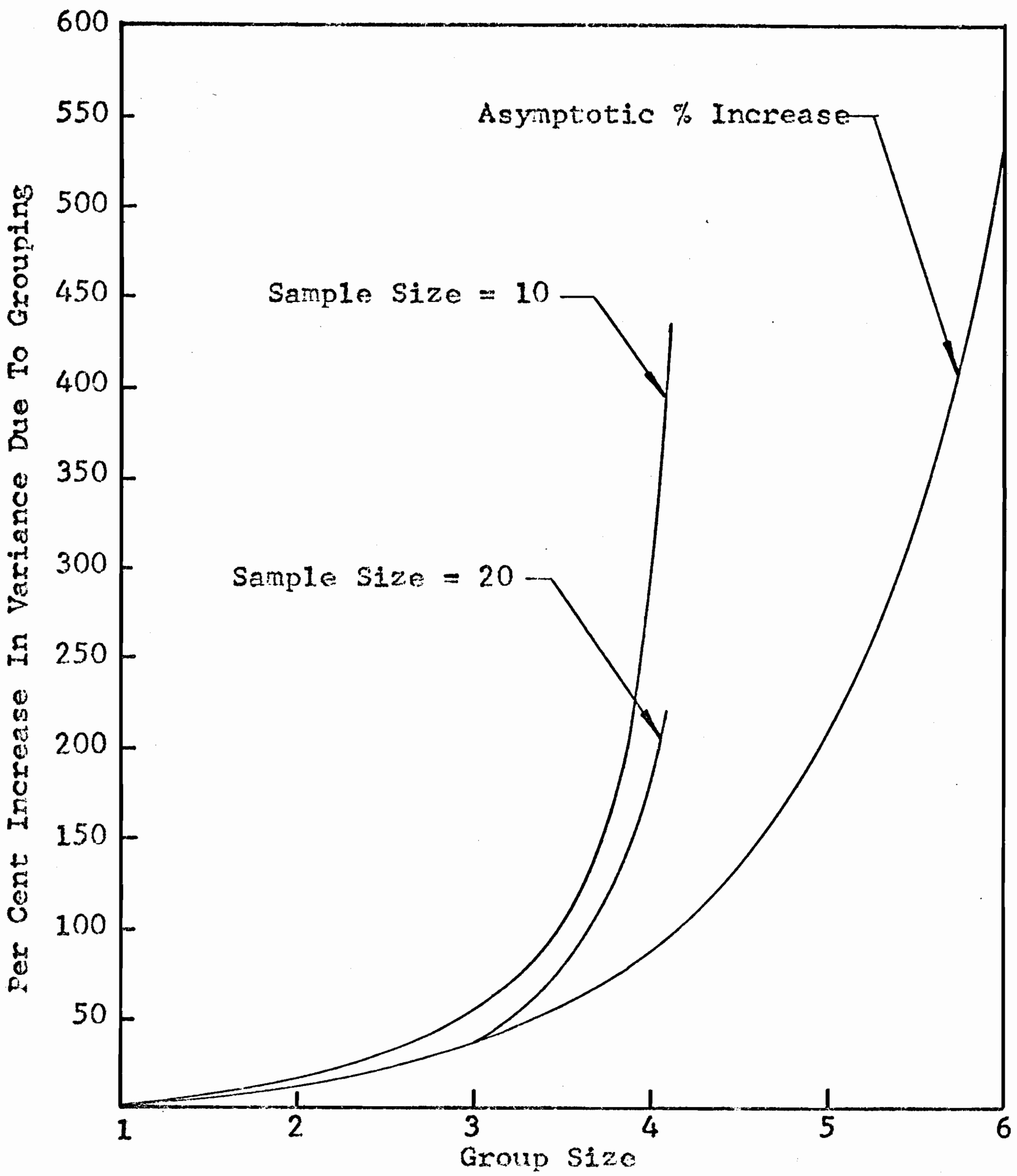

FIGURE 2.3. Small Sample Per Cent Increase In Variance Due To Grouping $\theta=2.0$

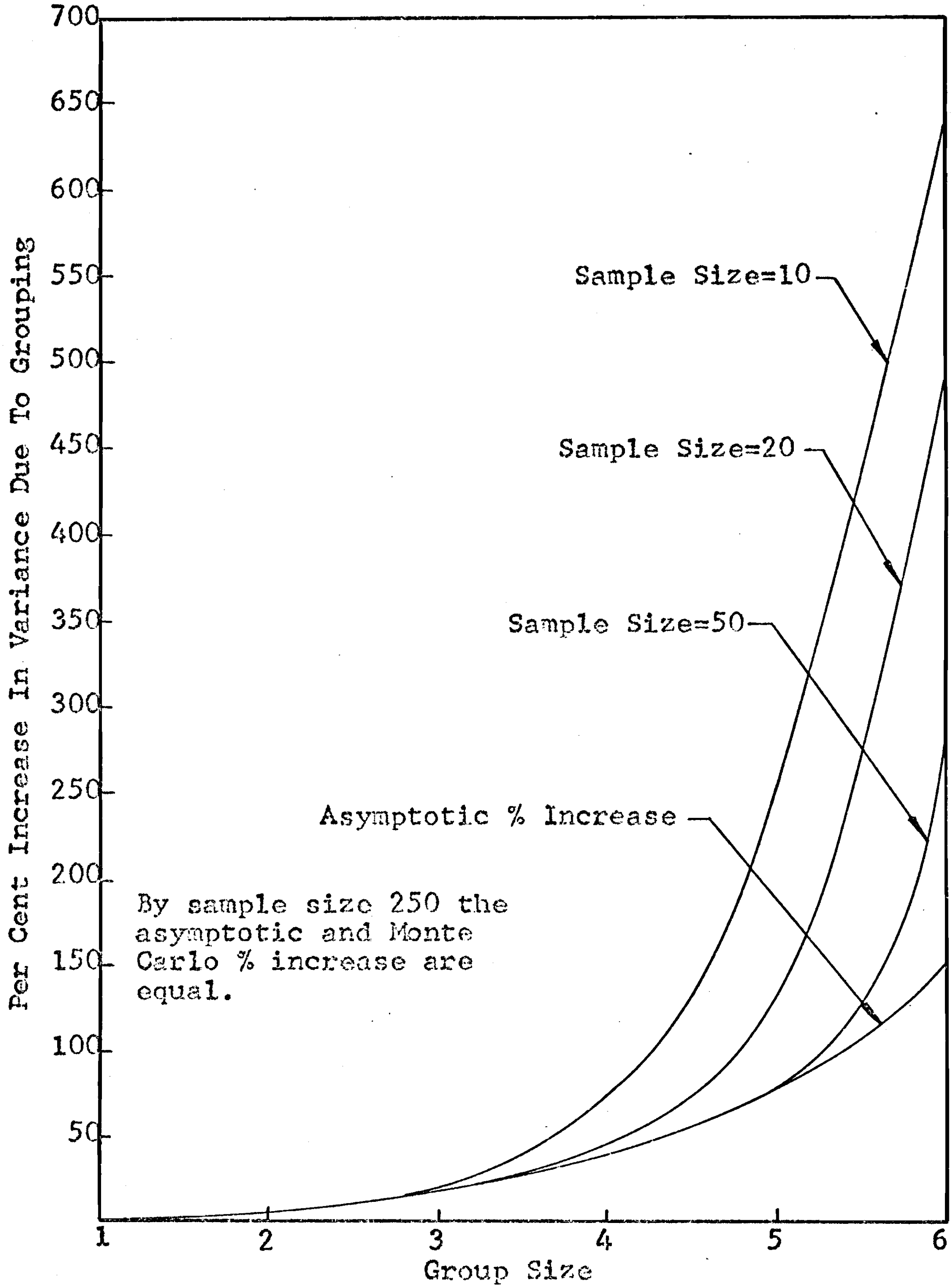

FIGURE 2.4. Small Sample Per Cent Increase In Variance Due To Grouping $\theta=3.0$

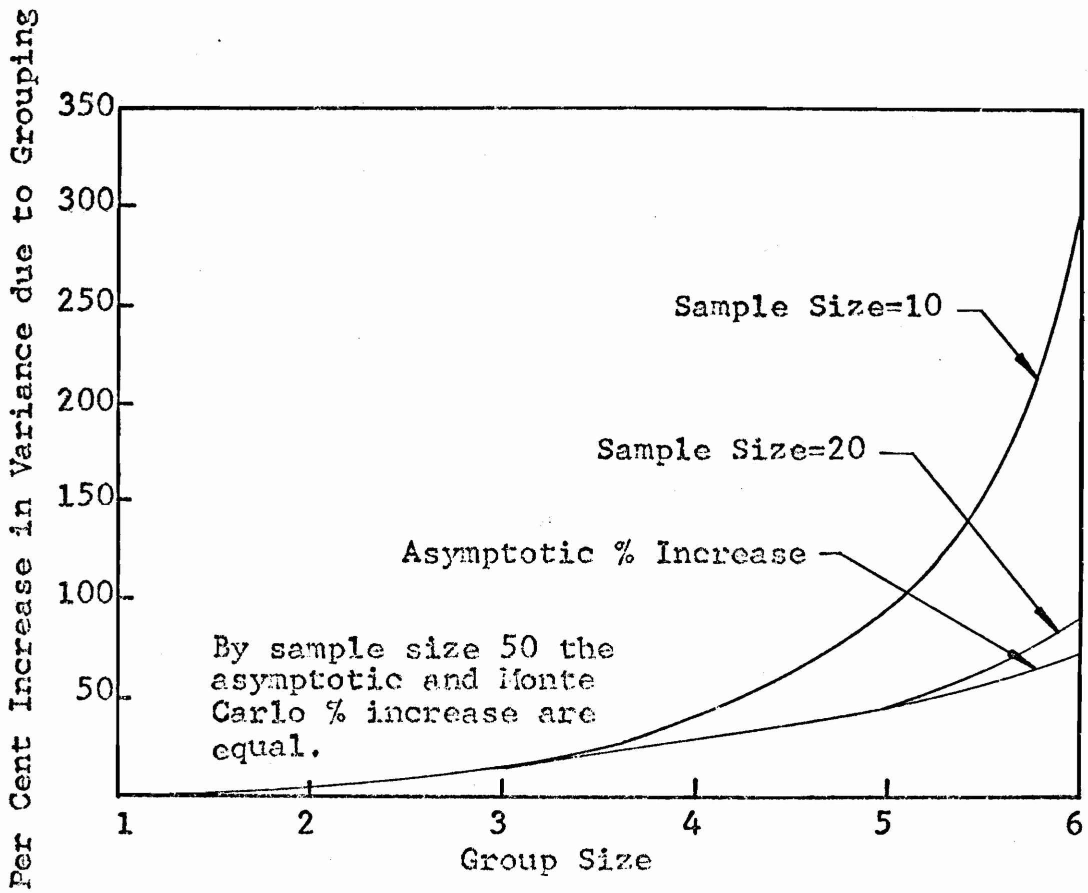

FIGURE 2.5. Small Sample Per Cent Increase In Variance Due To Grouping θ=4.0

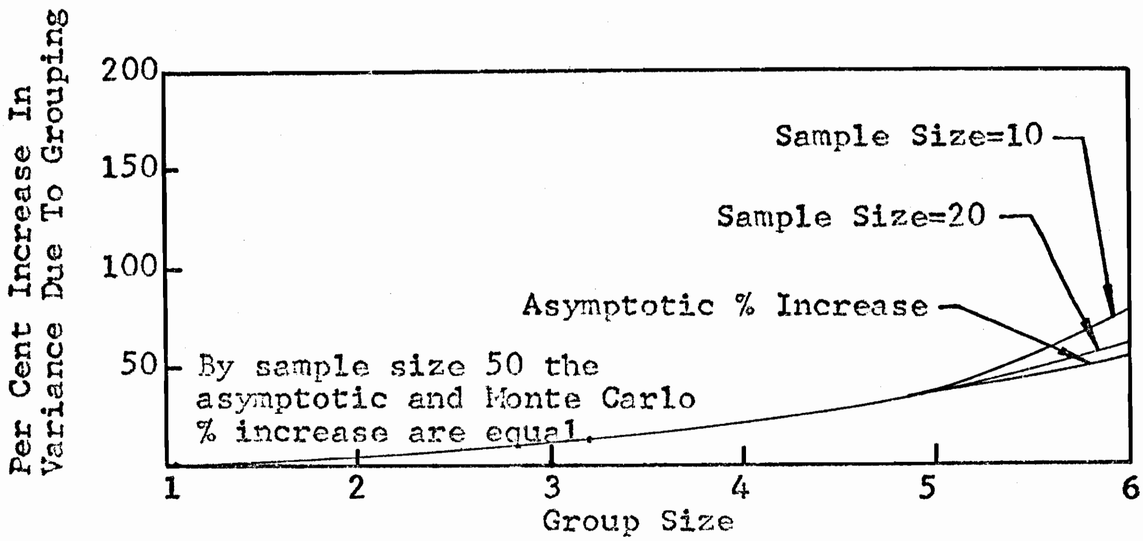

FIGURE 2.6. Small Sample Per Cent Increase In Variance Due To Grouping θ=5.0

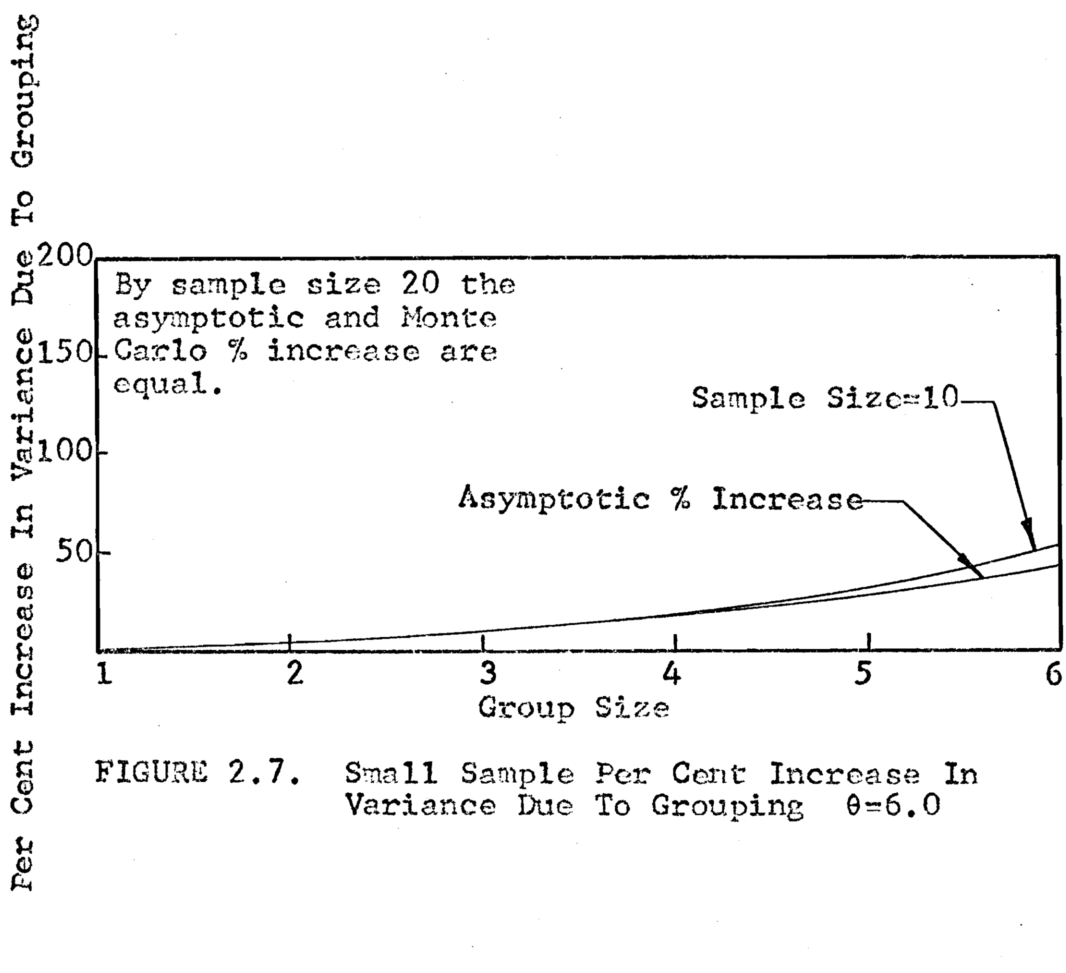

FIGURE 2.7. Small Sample Per Cent Increase In Variance Due To Grouping θ=6.0

sample is more likely to be spread over several groups; whereas if the group size is large or the parameter small, it becomes very likely that the majority of the observations lie in a single group which would have the same result as placing the entire sample space in a single group.

CHAPTER III

COMBINATIONS OF DISTRIBUTIONS DIFF kb

In this chapter we shall derive some basic properties of the combination of distributions diff kb and then develope the procedures necessary to obtain the Maximum Likelihood estimators of the parameters involved under various degrees of completeness of the observed sample. A set of very general regularity conditions will then be given which, when satisfied, will insure that the Maximum Likelihood estimators are consistent and asymptotically normally distributed.

3.1 Properties of Combinations Diff kb

Recall that in Chapter I we defined the probability that a random variable X takes on the non-negative integral value j in the combination of distributions diff kb to be

$$P(X=j) = \sum_{k=0}^{[j/b]\leq\beta<\infty} P(X=j \text{ diff } kb|\Omega=k)P(\Omega=k), \qquad (3.1.1)$$

where

$$P(X=j \text{ diff } kb|\Omega=k)$$

and

$$P(\Omega=k)$$

are valid probability functions defined on the non-negative integers and β is a finite bound on the number of terms in the model. Here we show that (3.1.1) sums to unity and obtain the moment generating function for diff kb combinations, assuming that such a function exists for the distribution $P(X=j|\Omega=k)$.

$$\sum_{j=0}^{\infty} P(X=j) = \sum_{j=kb}^{\infty} \sum_{k=0}^{[j/b]\leq\beta<\infty} P(X=j \text{ diff } kb|\Omega=k)P(\Omega=k)$$

$$= \sum_{k=0}^{[j/b]\leq\beta<\infty} \sum_{j=kb}^{\infty} P(X=j \text{ diff } kb|\Omega=k)P(\Omega=k).$$

Letting

$$a = j-kb$$

and recalling the definition of j diff kb we can write the preceding summation as

$$\sum_{k=0}^{[j/b]\leq\beta<\infty} \sum_{a=0}^{\infty} P(X=a|\Omega=k)P(\Omega=k);$$

but we know that

$$\sum_{a} P(X=a|\Omega=k) = 1$$

and

$$\sum_{k=0}^{\beta} P(\Omega=k) = 1,$$

so we have that

$$\sum_{j=0}^{\infty} P(X=j) = 1. \tag{3.1.2}$$

In order to obtain the moment generating function (m.g.f.) of (3.1.1), it is necessary to find the m.g.f., $\Psi_\alpha(t)$, of X diff α, where X is a non-negative random variable with probability function P(X) and m.g.f. $\Psi(t)$. Recalling the definition of a m.g.f. we have

$$\Psi_\alpha(t) = \sum_{j=\alpha}^{\infty} e^{tj} P(X=j \text{ diff } \alpha). \tag{3.1.3}$$

If we let

$$l = j \text{ diff } \alpha,$$

then

$$j = l + \alpha$$

and

$$\Psi_\alpha(t) = \sum_{l=0}^{\infty} e^{t(l+\alpha)} P(X=l)$$

$$= e^{t\alpha} \sum_{l=0}^{\infty} e^{tl} P(X=l);$$

but

$$\Psi(t) = \sum_{l=0}^{\infty} e^{tl} P(X=l).$$

Hence,

$$\Psi_{\alpha}(t) = e^{t\alpha}\Psi(t). \qquad (3.1.4)$$

We can use (3.1.4) to obtain the m.g.f. of the combination of distributions diff kb. Assuming that each component distribution in the diff kb combination has a m.g.f., denoted by $\Psi_{kb}(t)$, $k=0,1,\ldots,\beta<\infty$, the m.g.f. of the combination is obtained by evaluating

$$\sum_{j=0}^{\infty} e^{tj}P(X=j)$$

$$= \sum_{k=0}^{[j/b]\leq\beta<\infty} \sum_{j=kb}^{\infty} e^{tj}P(X=j \text{ diff } kb \,|\, \Omega=k)P(\Omega=k). \qquad (3.1.5)$$

Letting

$$l = j-kb,$$

(3.1.5) becomes

$$\sum_{j=0}^{\infty} e^{tj}P(X=j)$$

$$= \sum_{k=0}^{[j/b]\leq\beta<\infty} \sum_{l=0}^{\infty} e^{t(l+kb)}P(X=l \,|\, \Omega=k)P(\Omega=k)$$

$$= \sum_{k=0}^{\beta} e^{tkb}\, \Psi_{kb}(t)P(\Omega=k). \qquad (3.1.6)$$

Since much emphasis is placed on applications where the component probabilities are Poisson, in the next chapter the m.g.f. and the first two moments are found from (3.1.6) for the case where $P(X=1|\Omega=k)$ is the Poisson probability function.

We now turn our attention toward the estimation problem in the combination of distributions diff kb.

3.2 Full Data Case

In the introduction, we mentioned the fact that when sampling from the type of distribution defined by (3.1.1) the observed "total" frequencies can be partitioned into partial frequencies which might be, but often in practical situations are not, observed. In this section we deal with the problem of obtaining the maximum likelihood estimators of the parameters of the underlying distributions, and the "mixing parameters", designated by $P(\Omega=k)$ in (3.1.1), when all the partial frequencies have been observed and are available.

Before continuing with the estimation problem, we introduce the following notation:

N_g is the total observed frequency representing the number of times the random variable X takes on the value g.

n_g^j is the j-th observed partition of the total observed frequency, N_g, where $j=0,1,\ldots,[g/b]$ and the n_g^j are such that

$$\sum_{j=0}^{[g/b]} n_g^j = N_g \ . \qquad (3.2.1)$$

$f(g \text{ diff } bk, \underline{\theta}_k, a_k) = a_k f(g \text{ diff } bk, \underline{\theta}_k)$ is the actual probability of g diff bk. (3.2.2)

$$a_k = P(\Omega=k) \text{ and } \sum_{k=0}^{\beta} a_k = 1. \qquad (3.2.3)$$

$\underline{\theta}_k$ is the parameter vector of the k-th component distribution.

Assuming that the observations are independent, identically distributed random variables, the likelihood of the sample is

$$L = \prod_g \prod_{k=0}^{[g/b]\leq\beta<\infty} f(g \text{ diff } bk, \underline{\theta}_k, a_k)^{n_g^k} \qquad (3.2.4)$$

We once again define log L to be the likelihood function. The maximum likelihood estimators of $\underline{\theta}_k$ and a_k will be solutions of the likelihood equations

$$\frac{\partial \log L}{\partial \underline{\theta}_k} = \underline{0} \qquad k=0,1,\ldots,\beta \qquad (3.2.5)$$

and

$$\frac{\partial \log L}{\partial a_k} = 0 \qquad k=0,1,\ldots,\beta-1 \qquad (3.2.6)$$

Writing the likelihood equations for the $\underline{\theta}_k$, we see that, for every k, each equation can be solved independently of the others, that is,

$$\frac{\partial \log L}{\partial \underline{\theta}_k} = \sum_{g=kb}^{\infty} \frac{n_g^k \frac{\partial}{\partial \underline{\theta}_k} f(g \text{ diff } bk, \underline{\theta}_k, a_k)}{f(g \text{ diff } bk, \underline{\theta}_k, a_k)} = \underline{0}. \qquad (3.2.7)$$

However, to obtain the M.L. estimators for the a_k, we have to solve a system of β simultaneous equations since

$$a_\beta = 1 - \sum_{j=0}^{\beta-1} a_j \; .$$

The system of likelihood equations is

$$\frac{\partial \log L}{\partial a_o} = \sum_{g=0} \frac{n_g^o \frac{\partial}{\partial a_o} f(g \text{ diff } 0, \underline{\theta}_o, a_o)}{f(g \text{ diff } 0, \underline{\theta}_o, a_o)}$$

$$+ \sum_{g=\beta b} \frac{n_g^\beta \frac{\partial}{\partial a_o} f(g \text{ diff } \beta b, \underline{\theta}_\beta, a_\beta)}{f(g \text{ diff } \beta b, \underline{\theta}_\beta, a_\beta)} = 0$$

$$\vdots$$

$$\frac{\partial \log L}{\partial a_k} = \sum_{g=bk} \frac{n_g^k \frac{\partial}{\partial a_k} f(g \text{ diff } bk, \underline{\theta}_k, a_k)}{f(g \text{ diff } bk, \underline{\theta}_k, a_k)}$$

$$+ \sum_{g=\beta b} \frac{n_g^\beta \frac{\partial}{\partial a_k} f(g \text{ diff } \beta b, \underline{\theta}_\beta, a_\beta)}{f(g \text{ diff } \beta b, \underline{\theta}_\beta, a_\beta)} \qquad (3.2.8)$$

$$\vdots$$

$$\frac{\partial \log L}{\partial a_{\beta-1}} = \sum_{g=b(\beta-1)} \frac{n_g^{\beta-1} \frac{\partial}{\partial a_{\beta-1}} f(g \text{ diff } b(\beta-1), \underline{\theta}_{\beta-1}, a_{\beta-1})}{f(g \text{ diff } b(\beta-1), \underline{\theta}_{\beta-1}, a_{\beta-1})}$$

$$+ \sum_{g=\beta b} \frac{n_g^{\beta} \frac{\partial}{\partial a_{\beta-1}} f(g \text{ diff } \beta b, \underline{\theta}_{\beta}, a_{\beta})}{f(g \text{ diff } \beta b, \underline{\theta}_{\beta}, a_{\beta})} .$$

As a result of (3.2.2) and (3.2.3), (3.2.8) becomes

$$\left.\begin{aligned}
&\sum_{g=0} \frac{n_g^o}{a_o} - \sum_{g=\beta b} \frac{n_g^\beta}{1 - \sum_{j=0}^{\beta-1} a_j} = 0 \\
&\sum_{g=b} \frac{n_g^1}{a_1} - \sum_{g=\beta b} \frac{n_g^\beta}{1 - \sum_{j=0}^{\beta-1} a_j} = 0 \\
&\vdots \\
&\sum_{g=b(\beta-1)} \frac{n_g^{\beta-1}}{a_{\beta-1}} - \sum_{g=\beta b} \frac{n_g^\beta}{1 - \sum_{j=0}^{\beta-1} a_j} = 0 .
\end{aligned}\right\} \quad (3.2.9)$$

The solution to (3.2.9) is

$$\hat{a}_k = \frac{\sum\limits_{g=bk} n_g^k}{N} \qquad k=0,1,\ldots,\beta-1$$

and (3.2.10)

$$\hat{a}_\beta = 1 - \sum_{k=0}^{\beta-1} \hat{a}_k \ .$$

The properties of these estimators will be discussed later in the chapter.

We have just obtained the maximum likelihood estimators of the parameters in the combination of distributions diff kb when all the partial frequencies were observed and available to the experimenter. Notice that the estimation, in this case, is similar to the estimation of the parameters in $(\beta+1)$ independent probability functions. We shall see that this is not the case when some or, especially, all of the partial frequencies are missing and only the total frequencies are available.

3.3 Missing Partial Frequencies

Here we consider the problem of estimating the parameters in (3.1.1) when any combination of the partial frequencies is unknown to the experimenter. As always, each of the total frequencies has been observed and is available to the experimenter. Perhaps it should be mentioned here that while the methods developed in this section are general and will apply to any combination of missing partial frequencies, particular emphasis will be given later to the case where all of the partial frequencies are missing with only the total frequencies being known. This is due to the fact that in many practical situations if any of the partial frequencies are known there usually is no reason why they should not all be known. However, there are cases in which only the total frequencies are recorded.

Before proceeding with the estimation problem, we introduce the following notation in addition to that developed in the preceding section.

$F(g,\underline{\theta},\underline{a})$ is the total probability of the random variable X taking on the value g, i.e.,

$$F(g,\underline{\theta},\underline{a}) = \sum_{k=0}^{[g/b]\leq\beta<\infty} f(g \text{ diff } bk,\underline{\theta}_k,a_k). \qquad (3.3.1)$$

Once we have observed a sample, we partition it into the following sub-classes

C_1 = (g:some or all of the $n_g^j, j=0,1,\ldots,[g/b]$, are observed)

C_2 = (g:only N_g, the total frequency, is observed)

The remaining sub-classes are defined to be such that they contain those g which do not belong to C_1 or C_2 and have the same partial frequencies unobserved. For example, when b=6 the following data arrangement is possible:

$$N_{18} = (n^o_{18}) + (n^1_{18}) + n^2_{18} + n^3_{18}$$

$$N_{19} = (n^o_{19}) + (n^1_{19}) + (n^2_{19}) + (n^3_{19})$$

$$\vdots$$

$$N_{23} = (n^0_{23}) + (n^1_{23}) + (n^2_{23}) + (n^3_{23})$$

$$N_{24} = n^o_{24} + n^1_{24} + n^2_{24} + n^3_{24} + n^4_{24}$$

$$N_{25} = (n^o_{25}) + (n^1_{25}) + n^2_{25} + n^3_{25} + n^4_{25}$$

$$\vdots$$

$$N_{29} = (n^o_{29}) + (n^1_{29}) + (n^2_{29}) + (n^3_{29}) + (n^4_{29})$$

where the partial frequencies in parentheses are unobserved. Then

$$C_1 = \{24\}$$

and

$$C_2 = \{19,23,29\} .$$

Another of the subsets would be

$$C = \{18,\ 25\} ,$$

since n^o_{18}, n^1_{18}, n^o_{25}, and n^1_{25} are the only missing partial frequencies for 18 and 25, respectively.

We also define for each $g \notin C_1 + C_2$

$$H(g,\underline{\theta},\underline{a}) = F(g,\underline{\theta},\underline{a}) - \sum_{\substack{\text{observed} \\ j}} f(g \text{ diff } bj,\underline{\theta}_j,a_j) \quad (3.3.2)$$

and

$$N_g^* = N_g - \sum_{\substack{\text{observed} \\ j}} n_g^j \; . \quad (3.3.3)$$

The likelihood of an incomplete sample can be expressed as

$$L = \prod_{g \in C_1} \prod_{j=0}^{[g/b] \leq \beta < \infty} f(g \text{ diff } bj,\underline{\theta}_j,a_j)^{n_g^j} \prod_{g \in C_2} F(g,\underline{\theta},\underline{a})^{N_g}$$

$$\prod_{g \notin C_1+C_2} H(g,\underline{\theta},\underline{a})^{N_g^*} \quad (3.3.4)$$

and hence the likelihood function becomes

$$\log L = \sum_{g \in C_1} \sum_{j=0}^{[g/b] \leq \beta < \infty} n_g^j \log f(g \text{ diff } bj,\underline{\theta}_j,a_j)$$

$$+ \sum_{g \in C_2} N_g \log F(g,\underline{\theta},\underline{a}) + \sum_{g \notin C_1+C_2} N_g^* \log H(g,\underline{\theta},\underline{a}). \quad (3.3.5)$$

The likelihood equations for the $\underline{\theta}_j$ are

$$\frac{\partial \log L}{\partial \underline{\theta}_j} = \sum_{\substack{g=jb \\ g\varepsilon C_1}} \frac{n_g^j \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{f(g \text{ diff } bj, \underline{\theta}_j, a_j)}$$

$$+ \sum_{g\varepsilon C_2} \frac{N_g \frac{\partial}{\partial \underline{\theta}_j} F(g,\underline{\theta},\underline{a})}{F(g,\underline{\theta},\underline{a})} + \sum_{g\not\varepsilon C_1+C_2} \frac{N_g^* \frac{\partial}{\partial \underline{\theta}_j} H(g,\underline{\theta},\underline{a})}{H(g,\underline{\theta},\underline{a})}$$

$$= \underline{0} \,, \qquad (3.3.6)$$

where $j = 0,1,\ldots,\beta$.

Notice that if n_g^j is observed then $f(g \text{ diff } bj,\underline{\theta}_j,a_j)$ is subtracted from $F(g,\underline{\theta},\underline{a})$ in forming $H(g,\underline{\theta},\underline{a})$ and

$$\frac{\partial}{\partial \underline{\theta}_j} H(g,\underline{\theta},\underline{a}) = 0. \qquad (3.3.7)$$

As (3.3.6) stands, it can't be solved since it contains more than one of the unknowns. It must be solved simultaneously with the other likelihood equations for $\underline{\theta}_k$, $k \neq j$ and a_j, $j=0,1,\ldots,\beta-1$. This would involve solving a system of $2\beta+1$ non-linear simultaneous equations - a task we would like to avoid if at all possible. By applying an iterative procedure similar to the one described in Chapter II we are hopefully able to solve this system of simultaneous equations. The definitions, corresponding to those in Chapter II, which must be made here are

$$\tilde{n}_g^j = \frac{N_g\ f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{F(g, \underline{\theta}, \underline{a})} \quad g \in C_2 \qquad (3.3.8)$$

and

$$_*\tilde{n}_g^j = \frac{N_g^*\ f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{H(g, \underline{\theta}, \underline{a})} \quad g \notin C_1 + C_2\ . \qquad (3.3.9)$$

Applying these to (3.3.6), we have to solve

$$\frac{\partial \log L}{\partial \underline{\theta}_j} = \sum_{g \in C_1} \frac{n_g^j \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{f(g \text{ diff } bj, \underline{\theta}_j, a_j)}$$

$$+ \sum_{g \in C_2} \frac{\tilde{n}_g^j \frac{\partial}{\partial \underline{\theta}_j} F(g, \theta, a)}{f(g \text{ diff } bj, \underline{\theta}_j, a_j)} + \sum_{g \notin C_1 + C_2} \frac{_*\tilde{n}_g^j \frac{\partial}{\partial \underline{\theta}_j} H(g, \underline{\theta}, \underline{a})}{f(g \text{ diff } bj, \underline{\theta}_j, a_j)}$$

$$= 0. \qquad (3.3.10)$$

By the definitions of $F(g, \underline{\theta}, \underline{a})$ and $H(g, \underline{\theta}, \underline{a})$ we have

$$\frac{\partial}{\partial \underline{\theta}_j} F(g, \underline{\theta}, \underline{a}) = \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj,\ \underline{\theta}_j,\ a_j)$$

and

$$\frac{\partial}{\partial \underline{\theta}_j} H(g, \underline{\theta}, \underline{a}) = \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj,\ \underline{\theta}_j,\ a_j)$$

so, finally, (3.3.10) becomes

$$\frac{\partial \log L}{\partial \underline{\theta}_j} = \sum_{g \epsilon C_1} \frac{n_g^j \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{f(g \text{ diff } bj, \underline{\theta}_j, a_j)}$$

$$+ \sum_{g \epsilon C_2} \frac{\tilde{n}_g^j \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{f(g \text{ diff } bj, \underline{\theta}_j, a_j)}$$

$$+ \sum_{g \not\epsilon C_1 + C_2} \frac{{}^{*}\tilde{n}_g^j \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{f(g \text{ diff } bj, \underline{\theta}_j, a_j)}$$

$$= 0 \qquad (3.3.11)$$

and we see now that (3.3.11) involves only $\underline{\theta}_j$ and hence can be solved for $\underline{\theta}_j$ itself without resorting to the simultaneous solution of the system of $2\beta+1$ equations. We now have the likelihood equations to be solved <u>iteratively</u> for the $\hat{\underline{\theta}}_j$. A step-by-step description of the iteration procedure will be given later in this chapter.

We can obtain similar equations for the $\hat{a}_j$, $j=0,1,\ldots,\beta-1$. By the definitions of $f(g \text{ diff } bj, \underline{\theta}_j, a_j)$, $F(g, \underline{\theta}, \underline{a})$, and $H(g, \underline{\theta}, \underline{a})$ we can write

$$\frac{\partial}{\partial a_j} F(g, \underline{\theta}, \underline{a}) = \frac{1}{a_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)$$

$$- \frac{1}{a_\beta} f(g \text{ diff } \beta b, \underline{\theta}_\beta, a_\beta), \qquad (3.3.12)$$

since

$$a_\beta = 1 - \sum_{j=0}^{\beta-1} a_j .$$

There are several combinations of missing data which lead to different values of $\frac{\partial}{\partial a_j} H(g,\underline{\theta},\underline{a})$. They will be listed and considered separately.

I. n_g^j and n_g^β are both observed

a) $$\frac{\partial}{\partial a_j} H(g,\underline{\theta},\underline{a}) = \frac{\partial}{\partial a_j}\Big[F(g,\underline{\theta},\underline{a}) - f(g \text{ diff } bj,\underline{\theta}_j,a_j) - f(g \text{ diff } \beta b,\underline{\theta}_\beta,a_\beta) - \sum_{\substack{\text{remaining}\\ \text{observed } i}} f(g \text{ diff } bi,\underline{\theta}_i,a_i)\Big]$$

$$= \frac{\partial}{\partial a_j} f(g \text{ diff } bj,\underline{\theta}_j,a_j) + \frac{\partial}{\partial a_j} f(g \text{ diff } \beta b,\underline{\theta}_\beta,a_\beta) - \frac{\partial}{\partial a_j} f(g \text{ diff } bj,\underline{\theta}_j,a_j) - \frac{\partial}{\partial a_j} f(g \text{ diff } \beta b,\underline{\theta}_\beta,a_\beta)$$

$$= 0 \qquad \text{for } g \geq \beta b. \tag{3.3.13}$$

b) $$\frac{\partial}{\partial a_j} H(g,\underline{\theta},\underline{a}) = \frac{\partial}{\partial a_j}\Big[F(g,\underline{\theta},\underline{a}) - f(g \text{ diff } bj,\underline{\theta}_j,a_j)\Big]$$

$$= 0 \qquad \text{for } g < \beta b. \tag{3.3.14}$$

II. n_g^j observed, n_g^β not observed

a) $$\frac{\partial}{\partial a_j} H(g,\underline{\theta},\underline{a}) = \frac{\partial}{\partial a_j}\Big[F(g,\underline{\theta},\underline{a}) - f(g \text{ diff } bj,\underline{\theta}_j,a_j) - \sum_{\substack{\text{remaining}\\ \text{observed } i}} f(g \text{ diff } bi,\underline{\theta}_i,a_i)\Big]$$

$$= \frac{1}{a_j} f(g \text{ diff } bj,\underline{\theta}_j,a_j) - \frac{1}{a_\beta} f(g \text{ diff } \beta b,\underline{\theta}_\beta,a_\beta) - \frac{1}{a_j} f(g \text{ diff } bj,\underline{\theta}_j,a_j)$$

$$= -\frac{1}{a_\beta} f(g \text{ diff } \beta b,\underline{\theta}_\beta,a_\beta) \quad g \geq \beta b. \qquad (3.3.15)$$

b) $$\frac{\partial}{\partial a_j} H(g,\underline{\theta},\underline{a}) = 0 \qquad g < \beta b. \qquad (3.3.16)$$

III. n_g^j not observed, n_g^β observed

a) $$\frac{\partial}{\partial a_j} H(g,\underline{\theta},\underline{a}) = \frac{\partial}{\partial a_j}\Big[F(g,\underline{\theta},\underline{a}) - f(g \text{ diff } \beta b,\underline{\theta}_\beta,a_\beta) - \sum_{\substack{\text{remaining}\\ \text{observed } i}} f(g \text{ diff } bi,\underline{\theta}_i,a_i)\Big]$$

$$= \frac{1}{a_j} f(g \text{ diff } bj,\underline{\theta}_j a_j)$$

$$- \frac{1}{a_\beta} f(g \text{ diff } \beta b, \underline{\theta}_\beta, a_\beta)$$

$$+ \frac{1}{a_\beta} f(g \text{ diff } \beta b, \underline{\theta}_\beta, a_\beta)$$

$$= \frac{1}{a_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j) \quad j \geq \beta b. \qquad (3.3.17)$$

b) $$\frac{\partial}{\partial a_j} H(g,\underline{\theta},\underline{a}) = \frac{\partial}{\partial a_j}\Big[F(g,\underline{\theta},\underline{a}) - \sum_{\substack{\text{remaining} \\ \text{observed } i}} f(g \text{ diff } bi, \underline{\theta}_i, a_i)\Big]$$

$$= \frac{1}{a_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j) \quad j < \beta b. \qquad (3.3.18)$$

IV. n_g^j not observed, n_g^β not observed

a) $$\frac{\partial}{\partial a_j} H(g,\underline{\theta},\underline{a}) = \frac{\partial}{\partial a_j}\Big[F(g,\underline{\theta},\underline{a}) - \sum_{\substack{\text{remaining} \\ \text{observed } i}} f(g \text{ diff } bi, \underline{\theta}_i, a_i)\Big]$$

$$= \frac{1}{a_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)$$

$$- \frac{1}{a_\beta} f(g \text{ diff } \beta b, \underline{\theta}_\beta, a_\beta) \quad g \geq \beta b. \qquad (3.3.19)$$

b) $$\frac{\partial}{\partial a_j} H(g,\underline{\theta},\underline{a}) = \frac{1}{a_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j) \quad g < \beta b. \qquad (3.3.20)$$

Hence, we can write

$$\frac{\partial \log L}{\partial a_j} = \sum_{g\in C_1} \frac{n_g^j\, f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{a_j\, f(g \text{ diff } bj, \underline{\theta}_j, a_j)}$$

$$+ \sum_{\substack{g\in C_2 \\ g<\beta b}} \frac{N_g}{F(g,\underline{\theta},\underline{a})} \left(\frac{f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{a_j}\right)$$

$$+ \sum_{\substack{g\in C_2 \\ g\geq\beta b}} \frac{N_g}{F(g,\underline{\theta},\underline{a})} \left(\frac{f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{a_j} - \frac{f(g \text{ diff } \beta b, \underline{\theta}_\beta, a_\beta)}{1 - \sum_{k=0}^{\beta-1} a_k}\right)$$

$$+ \sum_{g\in IIa} \frac{N_g^*}{H(g,\underline{\theta},\underline{a})} \left(- \frac{1}{a_\beta} f(g \text{ diff } \beta b, \underline{\theta}_\beta, a_\beta)\right)$$

$$+ \sum_{g\in IIIa} \frac{N_g^*}{H(g,\underline{\theta},\underline{a})} \left(\frac{1}{a_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)\right)$$

$$+ \sum_{g\in IIIb} \frac{N_g^*}{H(g,\underline{\theta},\underline{a})} \left(\frac{1}{a_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)\right)$$

$$+ \sum_{g\in IVa} \frac{N_g^*}{H(g,\underline{\theta},\underline{a})} \left(\frac{1}{a_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j) - \frac{1}{a_\beta} f(g \text{ diff } \beta b, \underline{\theta}_\beta, a_\beta)\right)$$

$$+ \sum_{g\in IVb} \frac{N_g^*}{H(g,\underline{\theta},\underline{a})} \left(\frac{1}{a_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)\right)$$

$$= 0. \qquad (3.3.21)$$

Notice that the last five terms of (3.3.21) are due to the partitioning of $\left| g : g \notin C_1 + C_2 \right|$.

By an application of (3.3.8) and (3.3.9) we can simplify (3.3.21) and obtain

$$\frac{\partial \log L}{\partial a_j} = \sum_{g \in C_1} \frac{n_g^j}{a_j} + \sum_{\substack{g \in C_2 \\ g < \beta b}} \frac{\tilde{n}_g^j}{a_j} + \sum_{\substack{g \in C_2 \\ g \geq \beta b}} \left(\frac{\tilde{n}_g^j}{a_j} - \frac{\tilde{n}_g^\beta}{1 - \sum_{k=0}^{\beta-1} a_k} \right)$$

$$+ \sum_{g \in IIa} \left(- \frac{{}_*\tilde{n}_g^\beta}{1 - \sum_{k=0}^{\beta-1} a_k} \right) + \sum_{g \in IIIa} \frac{{}_*\tilde{n}_g^j}{a_j} + \sum_{g \in IIIb} \frac{{}_*\tilde{n}_g^j}{a_j}$$

$$+ \sum_{g \in IVa} \left(\frac{{}_*\tilde{n}_g^j}{a_j} - \frac{{}_*\tilde{n}_g^\beta}{1 - \sum_{k=0}^{\beta-1} a_k} \right) + \sum_{g \in IVb} \frac{{}_*\tilde{n}_g^j}{a_j}$$

$$= 0, \qquad (3.3.22)$$

where $j=0,1,\ldots,\beta-1$.

As with the full data case, in order to obtain the solution, $\hat{a}_j$, $j=0,1,\ldots,\beta-1$, to (3.3.22), we must solve a system of simultaneous equations. The solution is

$$\hat{a}_j = \frac{\sum_{g \in C_1} n_g^j + \sum_{g \in C_2} \tilde{n}_g^j + \sum_{g \notin C_1 + C_2} {}_*\tilde{n}_g^j}{N}, \qquad (3.3.23)$$

where

$$N = \sum_g N_g$$

and

$$\hat{a}_\beta = 1 - \sum_{k=0}^{\beta-1} \hat{a}_k \ . \qquad (3.3.24)$$

In Chapter II, some general remarks concerning M.L. estimation in incomplete samples due to grouping were made. Perhaps the reader has noticed the similarity between grouping as discussed in Chapter II for classical probability models and the natural grouping we have here when some or all the partial frequencies are missing for a given set of N_g. In the former case the grouping is vertical whereas in the latter we have horizontal grouping. Therefore, when dealing with combinations of distributions diff kb, it is possible to have two dimensional grouping - horizontally and vertically. In what follows, we show how to obtain the M.L. estimators when we do have this two-dimensional grouping.

To avoid repetition and remain general, we deal only with the case in which we have a subset of our data grouped horizontally and vertically, i.e. for a given subset of the sample space we do not know any of the total frequencies or any of the partial frequencies, but we know the sum of the missing total frequencies. For example, if b=6, the following data arrangement is possible:

$$N_o = n_o^o$$

$$N_1 = n_1^o$$

$$\vdots$$

$$N_6 = n_6^o \ (+) \ n_6^1$$

$$N_7 = n_7^o \ (+) \ n_7^1$$

$$\vdots$$

$$\boxed{\begin{array}{l} N_{11} = n_{11}^o \ (+) \ n_{11}^1 \\ N_{12} = n_{12}^o \ (+) \ n_{12}^1 \ (+) \ n_{12}^2 \\ \vdots \\ N_{17} = n_{17}^o \ (+) \ n_{17}^1 \ (+) \ n_{17}^2 \end{array}}$$

where the "boxed in" frequencies are all missing but their total

$$\tau = \sum_{g=11}^{17} N_g$$

is known.

Before we obtain the M.L. estimators of the parameters, we define

$F(g,i,\underline{\theta},\underline{a})$ = Pr(X=g-th integer in i-th group)

$\sum_{g\epsilon i} F(g,i,\underline{\theta},\underline{a}) = G(i,\underline{\theta},\underline{a})$ = Pr(X ϵ i-th group)

τ_i = total observed frequency for i-th group.

The term in the general likelihood for a sample (3.3.4) due to such two-dimensional grouping is

$$\prod_i G(i,\underline{\theta},\underline{a})^{\tau_i} \qquad (3.3.25)$$

and hence the term in the likelihood function becomes

$$\sum_i \tau_i \log G(i,\underline{\theta},\underline{a}). \qquad (3.3.26)$$

Taking the derivative of (3.3.26) with respect to $\underline{\theta}_j$ yields the following term in the general likelihood:

$$\sum_i \frac{\tau_i \frac{\partial}{\partial \underline{\theta}_j} G(i,\underline{\theta},\underline{a})}{G(i,\underline{\theta},\underline{a})}. \qquad (3.3.27)$$

By an application of the chain rule for partial derivatives and the definition of $G(i,\underline{\theta},\underline{a})$, (3.3.27) reduces to

$$\frac{\partial \log L}{\partial \underline{\theta}_j} = \sum_i \sum_{g\epsilon i} \frac{\tau_i}{G(i,\underline{\theta},\underline{a})} \frac{\partial F(g,i,\underline{\theta},\underline{a})}{\partial \underline{\theta}_j}, \; j=0,1,\ldots,\beta. \quad (3.3.28)$$

To obtain the M.L. estimator for $\underline{\theta}_j$, (3.3.28) must be equated to zero and solved, but more than one of the unknown parameters appears in (3.3.28) and therefore a direct solution is not possible. So, once again, we are forced into solving a system of $2\beta+1$ simultaneous equations. However, the procedure which permitted us to by-pass this problem earlier in this chapter also applies in this case if we make the following definition:

$$\tilde{\tilde{n}}_g^j = \frac{\tau_i \, f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{G(i, \underline{\theta}, \underline{a})} . \tag{3.3.29}$$

Using (3.3.29), (3.3.28) reduces to

$$\frac{\partial \log L}{\partial \underline{\theta}_j} = \sum_i \sum_{g \epsilon i} \frac{\tilde{\tilde{n}}_g^j \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{f(g \text{ diff } bj, \underline{\theta}_j, a_j)} . \tag{3.3.30}$$

Notice that we have reduced the term in the likelihood equation due to two-dimensional grouping to the same form as if g were a member of C_1 in equation (3.3.11).

It should be remembered that (3.3.30) is an equation that will be solved for $\hat{\theta}_j$ many times, with the aid of (3.3.29), until the iterative process converges to a single value of $\hat{\theta}_j$.

We obtain the estimators for the a_j, $j=0,1,\ldots,\beta$ in a similar manner:

$$\frac{\partial \log L}{\partial a_j} = \sum_i \frac{\tau_i \frac{\partial}{\partial a_j} G(i, \underline{\theta}, \underline{a})}{G(i, \underline{\theta}, \underline{a})} . \tag{3.3.31}$$

After applying the chain rule for partial derivatives and the definition of $G(i,\underline{\theta},\underline{a})$, (3.3.31) becomes

$$\frac{\partial \log L}{\partial a_j} = \sum_i \sum_{g\epsilon i} \frac{\tau_i \frac{\partial}{\partial a_j} F(g,i,\underline{\theta},\underline{a})}{G(i,\underline{\theta},\underline{a})}. \qquad (3.3.32)$$

Recalling (3.3.29) and the definition of $F(g,i,\underline{\theta},\underline{a})$ we have

$$\frac{\partial \log L}{\partial a_j} = 0 \qquad 0 \leq g < jb \qquad (3.3.33)$$

$$= \sum_i \sum_{g\epsilon i} \frac{\tilde{\tilde{n}}^j_g}{a_j} \qquad jb \leq g < \beta b \qquad (3.3.34)$$

$$= \sum_i \sum_{g\epsilon i} \frac{\tilde{\tilde{n}}^j_g}{a_j} - \frac{\tilde{\tilde{n}}^\beta_g}{a_\beta} \qquad g \geq \beta b. \qquad (3.3.35)$$

Including these terms in the general likelihood function (3.3.4) and then writing the likelihood equations for $\underline{\theta}_j$ and the solutions of the likelihood equations for the a_j, $\hat{a}_j$, we have

$$\frac{\partial \log L}{\partial \underline{\theta}_j} = \sum_{g\epsilon C_1} \frac{n^j_g \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj,\underline{\theta}_j,a_j)}{f(g \text{ diff } bj,\underline{\theta}_j,a_j)}$$

$$+ \sum_{g\epsilon C_2} \frac{\tilde{n}^j_g \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj,\underline{\theta}_j,a_j)}{f(g \text{ diff } bj,\underline{\theta}_j,a_j)}$$

$$+ \sum_{g \notin C_1 + C_2} \frac{{}_*\tilde{n}_g^j \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{f(g \text{ diff } bj, \underline{\theta}_j, a_j)}$$

$$+ \sum_i \sum_{g \in i} \frac{\tilde{\tilde{n}}_g^j \frac{\partial}{\partial \underline{\theta}_j} f(g \text{ diff } bj, \underline{\theta}_j, a_j)}{f(g \text{ diff } bj, \underline{\theta}_j, a_j)}$$

$$= 0 \,, \qquad j = 0, 1, \ldots, \beta. \qquad (3.3.36)$$

$$\hat{a}_j = \frac{\sum_{g \in C_1} n_g^j + \sum_{g \in C_2} \tilde{n}_g^j + \sum_{\substack{g \notin C_1 + C_2 \\ g \notin \text{vertical groups}}} {}_*\tilde{n}_g^j + \sum_{g \in \text{vertical groups}} \tilde{\tilde{n}}_g^j}{N} \,,$$

$$\text{for } j = 0, 1, \ldots, \beta - 1 \qquad (3.3.37)$$

and

$$\hat{a}_\beta = 1 - \sum_{j=0}^{\beta-1} \hat{a}_j \,.$$

Except for the full data case, the method for obtaining the estimators is an iterative one. The steps to follow in order to obtain the roots of the likelihood equation are

1. Make an initial guess at the missing partial frequencies such that for each g <u>not</u> in the vertically grouped section of the sample, these partial frequencies sum to N_g; for those g in the vertically grouped section of the

sample, make an initial guess at the missing partial frequencies such that for each group they sum to the group total.

2. With these values of the missing frequencies, solve equations (3.3.36) and (3.3.37) for $\hat{\underline{\theta}}_j$ and $\hat{a}_j$, $j=0,1,\ldots,\beta$.

3. Using these new values of the θ_j and a_j, calculate new values of the missing frequencies by use of equations (3.3.8), (3.3.9), and (3.3.29).

4. Using the new values of the missing partial frequencies, solve (3.3.36) and (3.3.37) for different values of $\hat{\theta}_j$ and $\hat{a}_j$.

5. Continue this process until the sequences of $\hat{\theta}_j$ and $\hat{a}_j$ have converged.

It is readily seen that if this iterative process converges, it must converge to the M.L. solution. This follows since equation (3.3.36), by means of (3.3.8),(3.3.9), and (3.3.29), is identical to equation (3.3.6) including the term due to vertical grouping and equation (3.3.22) with the extra term due to vertical grouping included is identical, by using (3.3.8), (3.3.9), and (3.3.29), to (3.3.21); therefore, a solution of (3.3.36) must be a solution to (3.3.6) and a solution to (3.3.22) must be a solution to (3.3.21).

We shall now state a theorem due to Ford [8] which gives sufficient conditions for the convergence of the preceding iterative process. The proof is omitted due to its similarity to the corresponding proof in Chapter II.

Theorem: Let $(\Theta_o,\Theta_1,\ldots,\Theta_\beta)$ be a solution to the equations

$$\theta_i = s_i(\theta_o,\theta_1,\ldots,\theta_\beta) \qquad i = 0,1,\ldots,\beta$$

and in the region

$$R:(\Theta_i - h \le \theta_i \le \Theta_i + h) \qquad i = 0,1,\ldots,\beta$$

let

$$\left|\frac{\partial s_i}{\partial \theta_j}\right| < M_{ij} \qquad i,j = 0,1,\ldots,\beta$$

where

$$\sum_{j=0}^{\beta} M_{ij} < r < 1.$$

Let $({}_o\theta_o, {}_o\theta_1,\ldots,{}_o\theta_\beta)$ be in the region R and let

$$({}_1\theta_o, {}_1\theta_1,\ldots,{}_1\theta_\beta)$$

$$({}_2\theta_o, {}_2\theta_1,\ldots,{}_2\theta_\beta)$$

$$\vdots$$

be found successively from the equations

$$_1\theta_i = s_i({}_o\theta_o, {}_o\theta_1, \ldots, {}_o\theta_\beta)$$

$$_2\theta_i = s_i({}_1\theta_o, {}_1\theta_1, \ldots, {}_1\theta_\beta)$$

$$\vdots$$

$$_n\theta_i = s_i({}_{n-1}\theta_o, {}_{n-1}\theta_1, \ldots, {}_{n-1}\theta_\beta)$$

then

$$\lim_{n \to \infty} {}_n\theta_i = \Theta_i \qquad i=0,1,\ldots,\beta.$$

In the following chapter there will be a discussion of these conditions when working with a practical situation in which the Poisson distribution is used.

3.4 Discussion of The Asymptotic Properties of the Estimators

It is well known that under certain regularity conditions the maximum likelihood estimators are consistent, asymptotically normal and unbiased with dispersion matrix calculated as the inverse of the information matrix, a matrix whose elements are expectations of the second derivatives of the likelihood function. In this section, we shall list these regularity conditions and discuss them with reference to our model.

The regularity conditions which insure the desirable properties of a set of maximum likelihood estimators are given by Rao [14]. They are, for our type of model

1. $\sum_{g=0}^{\infty} F(g,\underline{\theta}_o,\underline{a}_o) \log F(g,\underline{\theta}_o\underline{a}_o) > -\infty$ where the zero subscript denotes the true value of the parameter vectors.

2. $F(g,\underline{\theta},\underline{a}) \neq F(g,\underline{\beta},\underline{b})$ for at least one g when $[\underline{\theta},\underline{a}] \neq [\underline{\beta},\underline{b}]$, which is an identifiability condition.

3. $F(g,\underline{\theta},\underline{a})$ admits first order partial derivatives which are continuous at the true value of the parameter vector, $[\underline{\theta}_o,\underline{a}_o]$.

4. The information matrix is non-singular at $[\underline{\theta}_o,\underline{a}_o]$.

Since the form of the underlying distributions in the combination of distributions diff kb partially determines whether or not these regularity conditions will be satisfied, we cannot make the statement that for any such combinations of distributions the estimators will be consistent and asymptotically normally distributed. For this reason, the satisfaction of conditions 1-4 must be discussed in the presence of a given family or families of component distributions. In Chapter IV, we shall show that when the component distributions are members of the Poisson family the resulting estimators are consistent and asymptotically normally distributed.

CHAPTER IV

ESTIMATION OF THE PARAMETERS IN THE DISTRIBUTION OF ITEM DEMANDS

In this chapter we shall apply the M.L. estimation procedure developed in the preceding chapter to the demand problem which was mentioned briefly in Chapter I.

4.1 The Poisson Demand Problem

As the title of the present section indicates, the Poisson probability function will be the underlying distribution in the combination model. The Poisson model is developed through the following:

Theorem 4.1: Assuming that the conditional probability of a single item being demanded in a time interval of length Δt, given that a non-negative integer multiple of bulks has also been demanded in the time interval, can be expressed as

$$\theta_k \Delta t + o(\Delta t) \qquad (4.1.1)$$

and that the probability of two or more such demands in the same time interval can be expressed as $o(\Delta t)$, then

$$Pr(X(t)=n \text{ diff } kb) = \frac{e^{-\theta_k t}(\theta_k t)^{n \text{ diff } kb}}{(n \text{ diff } kb)!}, \qquad (4.1.2)$$

where $Pr(X(t)=n \text{ diff } kb)$ is the probability that after a period of time, t, n diff kb single items will have been demanded given that k bulks have also been demanded. The proof of this theorem is given in many textbooks on stochastic processes. See, for example, Bailey [1,page 67]. Hence, for the demand model we have

$$P_{\underline{\theta}}(X=j) = \sum_{k=0}^{[j/b]\leq\beta<\infty} a_k \frac{e^{-\theta_k}\,\theta_k^{(j \text{ diff } kb)}}{(j \text{ diff } kb)!}, \quad j=0,1,\ldots \quad (4.1.3)$$

Comparing the terms in (4.1.3) with those in (1.3.1), we see that the Poisson portion of (4.1.3) corresponds to

$$P_{\underline{\theta}_k}(j \text{ diff } kb \mid \Omega=k),$$

i.e., the conditional probability that (j-kb) single items (singles) are demanded given that a demand for j items was made by means of k bulks. The term $P(\Omega=k)$ corresponds to a_k which is the probability of k bulks being demanded.

Perhaps mention should be made here of the physical significance of the parameters in the present form of the demand model. The a_k have already been explained. The θ_k are the mean number of singles demanded given that a demand has been made by k bulks. If one knew the values of these parameters, it would then be possible to determine how frequently multiples of bulks were demanded and the efficiency of the bulk size. The efficiency of a particular bulk size would be inversely proportional to the average number of singles demanded along with a bulk or a multiple of bulks. For these reasons, we shall be interested in estimating the parameters in the model determined by (4.1.3).

The Poisson is certainly not the only distribution that can be applied to the demand problem. In fact, in certain cases, it is likely that it would not fit the data well. In section 4.5 of this chapter, the negative binomial distribution is discussed in connection with this problem.

Before beginning with the estimation problem, consider the moment generating function of this Poisson model. In the preceding chapter we saw that the moment generating function of such a combination of distributions could be written as

$$\Psi(t) = \sum_{k=0}^{\beta} e^{tkb}\Psi_{kb}(t)P(\Omega=k),$$

where $\Psi_{kb}(t)$ represents the m.g.f. of the underlying distribution. For the Poisson distribution, we have

$$\Psi(t) = e^{\theta(e^t-1)}. \tag{4.1.4}$$

Hence, the m.g.f. of the Poisson combination model becomes

$$\Psi(t) = \sum_{k=0}^{\beta} a_k e^{tkb} e^{\theta_{kb}(e^t-1)}. \tag{4.1.5}$$

As is well known, the p-th non-central moment can be calculated from (4.1.5) by taking the p-th derivative with respect to t and evaluating it at t=0. In this manner, the first two non-central moments of the Poisson demand model were calculated and found to be

$$\mu_1 = \sum_{k=0}^{\beta} a_k(\theta_k+bk) \qquad (4.1.6)$$

and

$$\mu_2' = \sum_{k=0}^{\beta} a_k[(\theta_k+bk)^2+\theta_k]. \qquad (4.1.7)$$

4.2 Maximum Likelihood Estimation In The Full Data Poisson Demand Problem

In this section we will obtain the Maximum Likelihood estimators of the parameters and discuss their properties. From equation (3.2.7) we obtain

$$\hat{\theta}_j = \frac{\sum_{g=jb}^{\infty}(g-jb)\, n_g^j}{\sum_{g=jb}^{\infty} n_g^j} \quad j=0,1,\ldots,\beta. \qquad (4.2.1)$$

Notice that this is exactly the estimator arrived at when there is but one Poisson parameter to be estimated and all the data are available to the experimenter. The estimators for the a_j were established earlier as

$$\hat{a}_j = \frac{\sum_{g=jb}^{\infty} n_g^j}{N} \qquad j=0,1,\ldots,\beta. \tag{4.2.2}$$

As we shall see, unlike the case of missing partial frequencies, in the full data situation it is very simple to show that the estimators given by (4.2.1) and (4.2.2) have the usual asymptotic properties for maximum likelihood estimators. In fact it is trivial to show that the full data distribution admits a set of jointly sufficient statistics for the parameters. The likelihood for the full data sample is written as

$$L = \prod_{i=0}^{k\leq\beta<\infty} \frac{a_i^{N^{(i)}} \; e^{-N^{(i)}\theta_i} \; \theta_i^{\sum_{g=bi}^{\infty}(g-bi)n_g^i}}{\prod_g (g-bi)!} \tag{4.2.3}$$

$$= h(\underline{\theta}, \underline{a}, \sum_{g=bi}^{\infty} n_g^i (g-bi), N^{(i)}) \; f(x),$$

where

$$N^{(i)} = \sum_{g=bi} n_g^i$$

and $f(x)$ denotes a function of the observations only. Thus the statistics $\sum_{g=bi}^{\infty}(g-bi)n_g^i$ and $N^{(i)}$ are jointly sufficient for a_i and θ_i, $i=1,2,\ldots,\beta$. (See Mood and Graybill [13]). Since it is well known that in large samples any function of the sufficient statistics will estimate its expected value with variance covariance matrix given by the inverse of the information matrix, (See Kendall and Stuart [11], vol. 2, page 27), we have the desired result. It can also be shown that sufficiency insures that the likelihood equations have a unique solution, and the solution occurs at a maximum of the likelihood function. Asymptotical normality and consistency of the estimators given by (4.2.1) and (4.2.2) will follow as a result of an argument which is presented in section 4.3, as applied to the incomplete data situation.

Since, for the full data case, the dispersion matrix of the estimators is given by the inverse of the information matrix, we shall proceed with the derivation of the elements of the information matrix. We have

$$\log L = \sum_{k=0}^{\beta} \sum_{g=bk}^{\infty} n_g^k \log f(g \text{ diff } bk, \theta_k, a_k) \qquad (4.2.4)$$

and

$$\frac{\partial^2 \log L}{\partial \theta_k^2} = \sum_{g=bk}^{\infty} n_g^k \left[\frac{f(g \text{ diff } bk, \theta_k, a_k) \frac{\partial^2}{\partial \theta_k^2} f(g \text{ diff } bk, \theta_k, a_k)}{f^2(g \text{ diff } bk, \theta_k, a_k)} \right.$$

$$\left. - \frac{\left(\frac{\partial}{\partial \theta_k} f(g \text{ diff } bk, \theta_k, a_k)\right)^2}{f^2(g \text{ diff } bk, \theta_k, a_k)} \right], \quad (4.2.5)$$

where the random variable n_g^k is multinomially distributed. Hence,

$$E\, n_g^k = N\, f(g \text{ diff } bk, \theta_k, a_k). \quad (4.2.6)$$

Therefore,

$$E \frac{\partial^2 \log L}{\partial \theta_k^2} = N \frac{\partial^2}{\partial \theta_k^2} \sum_{g=bk}^{\infty} f(g \text{ diff } bk, \theta_k, a_k)$$

$$- N \sum_{g=bk}^{\infty} \frac{\left(\frac{\partial}{\partial \theta_k} f(g \text{ diff } bk, \theta_k, a_k)\right)^2}{f(g \text{ diff } bk, \theta_k, a_k)}$$

$$= -N \sum_{g=bk}^{\infty} f(g \text{ diff } bk, \theta_k, a_k) \left(\frac{g \text{ diff } bk - \theta_k}{\theta_k}\right)^2,$$

since $f(g \text{ diff } bk, \theta_k, a_k)$ is the Poisson probability function.

$$E \frac{\partial^2 \log L}{\partial \theta_k^2} = -Na_k \sum_{g=bk}^{\infty} f(g \text{ diff } bk, \theta_k) \left(\frac{g \text{ diff } bk - \theta_k}{\theta_k}\right)^2,$$

as a result of

$$f(g\ \text{diff}\ bk, \theta_k, a_k) = a_k f(g\ \text{diff}\ bk, \theta_k).$$

Thus, we can write

$$E\,\frac{\partial^2 \log L}{\partial \theta_k^2} = -\frac{Na_k}{\theta_k^2}\,E(g\ \text{diff}\ bk - \theta_k)^2,$$

but

$$E(g\ \text{diff}\ bk) = \theta_k\ ,$$

which follows since $f(g\ \text{diff}\ bk, \theta_k)$ is the Poisson frequency function. Therefore,

$$E(g\ \text{diff}\ bk - \theta_k)^2 = \theta_k\ .$$

Finally,

$$E\,\frac{\partial^2 \log L}{\partial \theta_k^2} = -\frac{Na_k}{\theta_k} \qquad k=0,1,\ldots,\beta. \tag{4.2.7}$$

The negative of (4.2.7) gives the first $(\beta+1)$ diagonal elements of the inverse of the dispersion matrix (information matrix).

To obtain the terms of the information matrix which will lead to the covariances between the θ_k, we have to evaluate

$$E\,\frac{\partial^2 \log L}{\partial \theta_i\,\partial \theta_k}$$

where, without loss of generality, k is assumed to be greater than i.

$$E \frac{\partial^2 \log L}{\partial\theta_i\,\partial\theta_k} = E \sum_{g=bk}^{\infty} n_g^k \left[\frac{\frac{\partial^2}{\partial\theta_i\partial\theta_k} f(g \text{ diff } bk,\theta_k,a_k)}{f^2(g \text{ diff } bk,\theta_k,a_k)}\right]$$

$$- E \sum_{g=bk}^{\infty} n_g^k \left[\frac{\frac{\partial}{\partial\theta_i} f(g \text{ diff } bi,\theta_i,a_i)\ \frac{\partial}{\partial\theta_k} f(g \text{ diff } bk,\theta_k,a_k)}{f^2(g \text{ diff } bk,\theta_k,a_k)}\right]$$

but,

$$\frac{\partial}{\partial\theta_i} f(g \text{ diff } bk,\theta_k,a_k) = 0.$$

Hence,

$$E \frac{\partial^2 \log L}{\partial\theta_i\,\partial\theta_k} = 0 \quad \text{for all } i,j=0,1,\ldots,\beta,\ i\neq j. \tag{4.2.8}$$

Since

$$a_\beta = 1 - \sum_{j=0}^{\beta-1} a_j, \tag{4.2.9}$$

the variance of $\hat{a}_\beta$ and covariance between $\hat{a}_i$ and $\hat{a}_\beta$ can be calculated from the variances and covariances of the remaining β independent $\hat{a}_i$ using the following formulae:

$$\text{Var } \hat{a}_\beta = \sum_{i=0}^{\beta-1} \text{Var } \hat{a}_i + 2\sum_{i\neq j=0}^{\beta-1} \text{Cov } (\hat{a}_i,\hat{a}_j) \tag{4.2.10}$$

$$\text{Cov}\ (\hat{a}_i, \hat{a}_\beta) = -(\text{Var}\ \hat{a}_i + \sum_{\substack{j=0 \\ j \neq i}}^{\beta-1} \text{Cov}\ (\hat{a}_i, \hat{a}_j)\ . \qquad (4.2.11)$$

Continuing with the development of the information matrix, we determine

$$\frac{\partial^2 \log L}{\partial a_j^2} = \sum_{g=bj}^{\infty} n_g^j \Big[\frac{f(g\ \text{diff}\ bj, \theta_j, a_j)\ \frac{\partial^2}{\partial a_j^2} f(g\ \text{diff}\ bj, \theta_j, a_j)}{f^2(g\ \text{diff}\ bj, \theta_j, a_j)}$$

$$- \frac{(\frac{\partial}{\partial a_j} f(g\ \text{diff}\ bj, \theta_j, a_j))^2}{f^2(g\ \text{diff}\ bj, \theta_j, a_j)} \Big]$$

$$+ \sum_{g=\beta b}^{\infty} n_g^\beta \Big[\frac{f(g\ \text{diff}\ \beta b, \theta_\beta, a_\beta)\ \frac{\partial^2}{\partial a_j^2} f(g\ \text{diff}\ \beta b, \theta_\beta, a_\beta)}{f^2(g\ \text{diff}\ \beta b, \theta_\beta, a_\beta)}$$

$$- \frac{(\frac{\partial}{\partial a_j} f(g\ \text{diff}\ \beta b, \theta_\beta, a_\beta))^2}{f^2(g\ \text{diff}\ \beta b, \theta_\beta, a_\beta)} \Big]. \qquad (4.2.12)$$

That (4.2.12) must be partitioned is a result of (4.2.9).

By the definition of $f(g\ \text{diff}\ bj, \theta_j, a_j)$, (3.2.2), we have

$$\frac{\partial}{\partial a_j} f(g\ \text{diff}\ bj, \theta_j, a_j) = \frac{1}{a_j} f(g\ \text{diff}\ bj, \theta_j, a_j) \qquad (4.2.13)$$

and

$$\frac{\partial}{\partial a_j} f(g \text{ diff } \beta b, \theta_\beta, a_\beta) = -\frac{1}{a_\beta} f(g \text{ diff } \beta b, \theta_\beta, a_\beta). \quad (4.2.14)$$

Using (4.2.13) and (4.2.14) in (4.2.12) and taking the necessary expectations results in

$$E \frac{\partial^2 \log L}{\partial a_j^2} = -\frac{N}{a_j^2} \sum_{g=bj}^{\infty} f(g \text{ diff } bj, \theta_j, a_j)$$

$$- \frac{N}{a_\beta^2} \sum_{g=\beta b}^{\infty} f(g \text{ diff } \beta b, \theta_\beta, a_\beta)$$

$$= -N\left(\frac{1}{a_j} + \frac{1}{a_\beta}\right) . \quad (4.2.15)$$

The negative of (4.2.15), for all $j \neq \beta$, yields the remainder of the diagonal elements of the information matrix.

Consider the case for which <u>$i \neq \beta$ and $i \neq j$</u> (without loss of generality $i > j$)

$$\frac{\partial^2 \log L}{\partial \theta_i \, \partial a_j} = \sum_{g=bi}^{\infty} n_g^i \left[\frac{\frac{\partial^2}{\partial \theta_i \partial a_i} f(g \text{ diff } bi, \theta_i, a_i)}{f^2(g \text{ diff } bi, \theta_i, a_i)} \right]$$

$$- \sum_{g=bi}^{\infty} n_g^i \left[\frac{\frac{\partial}{\partial \theta_i} f(g \text{ diff } bi, \theta_i, a_i) \; \frac{\partial}{\partial a_j} f(g \text{ diff } bi, \theta_i, a_i)}{f^2(g \text{ diff } bi, \theta_i, a_i)} \right]$$

$$= - \sum_{g=bi}^{\infty} n_g^i \frac{\frac{\partial}{\partial \theta_i} f(g \text{ diff } bi, \theta_i, a_i) \frac{\partial}{\partial a_j} f(g \text{ diff } bi, \theta_i, a_i)}{f^2(g \text{ diff } bi, \theta_i, a_i)}$$

but, for $i \neq \beta$

$$\frac{\partial}{\partial a_j} f(g \text{ diff } bi, \theta_i, a_i) = 0$$

and

$$E \frac{\partial^2 \log L}{\partial \theta_i \, \partial a_j} = 0 \, . \tag{4.2.16}$$

If we change the conditions which led to (4.2.16) by letting $i=\beta$ and $i \neq j$, we obtain

$$E \frac{\partial^2 \log L}{\partial \theta_\beta \partial a_j} = E \sum_{g=\beta b}^{\infty} n_g^\beta \left[\frac{\frac{\partial^2}{\partial \theta_\beta \partial a_j} f(g \text{ diff } \beta b, \theta_\beta, a_\beta)}{f^2(g \text{ diff } \beta b, \theta_\beta, a_\beta)} \right]$$

$$-E \sum_{g=\beta b}^{\infty} n_g^\beta \left[\frac{\frac{\partial}{\partial \theta_\beta} f(g \text{ diff } \beta b, \theta_\beta, a_\beta) \frac{\partial}{\partial a_j} f(g \text{ diff } \beta b, \theta_\beta, a_\beta)}{f^2(g \text{ diff } \beta b, \theta_\beta, a_\beta)} \right.$$

$$= -N \sum_{g=\beta b}^{\infty} \frac{\frac{\partial}{\partial \theta_\beta} f(g \text{ diff } \beta b, \theta_\beta, a_\beta) \frac{\partial}{\partial a_j} f(g \text{ diff } \beta b, \theta_\beta, a_\beta)}{f(g \text{ diff } \beta b, \theta_\beta, a_\beta)} \, .$$

Making use of the fact that $f(g \text{ diff } \beta b, \theta_\beta, a_\beta)$ is nothing more than the Poisson probability function multiplied by

a_β, we can write

$$E\frac{\partial^2 \log L}{\partial\theta_\beta\,\partial a_j} = N\sum_{g=\beta b}^{\infty}\frac{f(g\text{ diff }\beta b,\theta_\beta,a_\beta)}{a_\beta}\left(\frac{g\text{ diff }\beta b-\theta_\beta}{\theta_\beta}\right)$$

$$= \frac{N}{\theta_\beta}\sum_{g=\beta b}^{\infty} f(g\text{ diff }\beta b,\theta_\beta)(g\text{ diff }\beta b-\theta_\beta).$$

Following the same argument used to obtain (4.2.7), we arrive at

$$E\frac{\partial^2 \log L}{\partial\theta_\beta\partial a_j} = \frac{N}{\theta_\beta}E(g\text{ diff }\beta b-\theta_\beta)$$

but,

$$E\ g\text{ diff }\beta b = \theta_\beta \qquad (4.2.17)$$

so, we have

$$E\frac{\partial^2 \log L}{\partial\theta_\beta\,\partial a_j} = 0 \qquad j=0,1,\ldots,\beta-1\ . \qquad (4.2.18)$$

In a similar manner, we are able to show that

$$E\frac{\partial^2 \log L}{\partial\theta_j\,\partial a_j} = 0\ . \qquad (4.2.19)$$

Finally, we consider the case where $i\neq j$ ($i>j$ without loss of generality)

$$E\,\frac{\partial^2 \log L}{\partial a_i\,\partial a_j}$$

$$= E \sum_{g=bi}^{\infty} n_g^i \left[\frac{f(g \text{ diff } bi,\theta_i,a_i)\,\frac{\partial^2}{\partial a_i \partial a_j} f(g \text{ diff } bi,\theta_i,a_i)}{f^2(g \text{ diff } bi,\theta_i,a_i)}\right]$$

$$- E \sum_{g=bi}^{\infty} n_g^i \left[\frac{\frac{\partial}{\partial a_i} f(g \text{ diff } bi,\theta_i,a_i)\,\frac{\partial}{\partial a_j} f(g \text{ diff } bi,\theta_i,a_i)}{f^2(g \text{ diff } bi,\theta_i,a_i)}\right]$$

$$+ E \sum_{g=\beta b}^{\infty} n_g^\beta \left[\frac{f(g \text{ diff } \beta b,\theta_\beta,a_\beta)\,\frac{\partial^2}{\partial a_i \partial a_j} f(g \text{ diff } \beta b,\theta_\beta,a_\beta)}{f^2(g \text{ diff } \beta b,\theta_\beta,a_\beta)}\right]$$

$$- E \sum_{g=\beta b}^{\infty} n_g^\beta \left[\frac{\frac{\partial}{\partial a_i} f(g \text{ diff } \beta b,\theta_\beta,a_\beta)\,\frac{\partial}{\partial a_j} f(g \text{ diff } \beta b,\theta_\beta,a_\beta)}{f^2(g \text{ diff } \beta b,\theta_\beta,a_\beta)}\right].$$

Upon taking the necessary expectations and recalling that for $i \neq \beta$

$$\frac{\partial}{\partial a_j} f(g \text{ diff } bi,\theta_i,a_i) = 0 ,$$

we are left with

$$E\,\frac{\partial^2 \log L}{\partial a_i\,\partial a_j} = -N \sum_{g=\beta b}^{\infty} \frac{1}{a_\beta^2} f(g \text{ diff } \beta b,\theta_\beta,a_\beta)$$

$$= -\frac{N}{a_\beta} . \qquad (4.2.20)$$

Once again, the proper term to place in the information matrix is the negative of (4.2.20). This completes the information matrix.

Although the asymptotic variances and covariances of the estimators of the parameters must be obtained from the inverse of the information matrix, the information matrix for the full data Poisson demand problem is such that it permits one to obtain the variances of the $\hat{\theta}_j$, $j = 0,1,\ldots,\beta$, without inverting the entire matrix. These variances are merely

$$\frac{\theta_j}{Na_j} \ . \tag{4.2.21}$$

Notice the similarity between (4.2.21) and the variance of $\hat{\theta}$ in the ordinary one-parameter Poisson problem which is given by

$$\frac{\theta}{N} \ . \tag{4.2.22}$$

The denominator of (4.2.21) is nothing more than the expected frequency of demands made by means of j bulks.

4.3 Estimation in The Poisson Demand Problem When Only The Total Frequencies are Known

As a result of the large number of different combinations of incomplete data for the demand problem, we shall consider only the case in which the total frequencies, N_g, of item demands are observed with all partial frequencies missing.

The demand problem itself first arose out of a need by the Logistics Branch of The Office of Naval Research. Navy administrative personnel have been concerned for some time with the problem of fitting distributions of items demanded by Navy personnel on submarines and other ships that remain at sea for long periods of time. The eventual goal, of course, was one of an inventory nature. For certain types of items, the Poisson or Negative Binomial Distribution normally seemed to fit the data extremely well. However, for other types of items, there seemed to be "surges" of frequencies at X=2,6,or 10, etc., depending on the item, and multiples of these numbers. The result is, of course, that for these items, no classical type of probability model could be used to fit the data. After a close scrutiny of the item types, it was determined that the items whose demand frequencies had the strange surges at X=kb, k=1,2,... , were those that were packed in bulks of size b.

As a result of this, it was felt that a different type of probability model was needed, one which takes into account the bulk problem. The combination of distributions diff kb probability structure seemed like a possible solution to the problem. To make the problem more difficult, the only data that is available to the logistics people are total frequencies, i.e., the number of times that X items were demanded, without reference to whether the demand was satisfied by bulks, singles, or a combination of bulks and singles.

Recall that in Chapter III we outlined the iterative procedure which would yield the maximum likelihood estimators of the unknown parameters, if the method itself converged. The iterative solutions to the Poisson likelihood equations are simply

$$\hat{\theta}_j = \frac{\sum_{g=bj}^{\infty} (g-bj)\, \tilde{n}_g^j}{\sum_{g=bj}^{\infty} n_g^j} \tag{4.3.1}$$

and

$$\hat{a}_j = \frac{\sum_{g=bj}^{\infty} \tilde{n}_g^j}{N} \tag{4.3.2}$$

where, after the initial guess at the missing partial frequencies, the $\tilde{n}_g^j$ are obtained from

$$\tilde{n}_g^j = \frac{N_g \hat{a}_j e^{-\hat{\theta}_j} \hat{\theta}_j^{g \text{ diff } bj}}{(g \text{ diff } bj)! \sum_{i=0}^{[g/b]} \frac{\hat{a}_i e^{-\hat{\theta}_i} \hat{\theta}_i^{g \text{ diff } bi}}{(g \text{ diff } bi)!}} \qquad (4.3.3)$$

Briefly, the iterative procedure is the following: With an initial guess at the missing partial frequencies, obtain values of $\hat{\theta}_j$ and $\hat{a}_j$ from (4.3.1) and (4.3.2) respectively. Using these new values of $\hat{\theta}_j$ and $\hat{a}_j$ in equation (4.3.3) one can obtain new values of the $\tilde{n}_g^j$ which will lead to a different value of $\hat{\theta}_j$ and $\hat{a}_j$. Continue this process until the sequences $\{\hat{\theta}_j\}$ and $\{\hat{a}_j\}$ converge.

A considerable amount of time and effort was spent unsuccessfully trying to obtain a proof that the above mentioned iterative procedure always converges. However, in performing the Monte Carlo work to be presented later in this chapter, a rather large body of empirical evidence was compiled which would support such a proposition. The iterative procedure was performed approximately 11,500 times on different bulk sizes, different numbers of parameters involved, and different sample sizes, and it never failed to converge. As the number of parameters was increased, the rates of convergence, of course, decreased, but convergence was always achieved.

Now that we have the iterative procedure which, if the process converges, yields the maximum likelihood estimators, we shall show that when only the total frequencies are known the regularity conditions listed in 3.4 are satisfied. Hence, we will have established the desirable asymptotic properties of the estimators.

Condition 1: For our model, this condition is written

$$\sum_g F(g,\underline{\theta},\underline{a}) \log F(g,\underline{\theta},\underline{a}) > -\infty \quad . \tag{4.3.4}$$

By the definition of $F(g,\underline{\theta},\underline{a})$,

$$\sum_g F(g,\underline{\theta},\underline{a}) \log F(g,\underline{\theta},\underline{a}) > \Sigma f(g,\theta_o,a_o) \log f(g,\theta_o,a_o),$$

where

$$f(g,\theta_o,a_o) = \frac{a_o e^{-\theta_o}\theta_o^g}{g!} \tag{4.3.5}$$

and, therefore,

$$\log f(g,\theta_o,a_o) = \log a_o - \theta_o + g \log \theta_o - \log g! \;.$$

So we have

$$\sum_g F(g,\underline{\theta},\underline{a}) \log F(g,\underline{\theta},\underline{a}) > \sum_g f(g,\theta_o,a_o) \log f(g,\theta_o,a_o)$$

$$= a_o E\log a_o - a_o\theta_o - a_o \log \theta_o \, Eg - a_o E \log g!$$

$$= a_o \log a_o - a_o\theta_o - a_o^2\theta_o \log \theta_o - a_o E \log g! \tag{4.3.6}$$

and, for non-zero a_o, we have all the terms in (4.3.6) finite except for the possible exception of E log g!.

Consider E log g!. From Feller [7], we have

$$g! < (2\pi)^{\frac{1}{2}} (g+\tfrac{1}{2})^{(g+\frac{1}{2})} e^{-(g+\frac{1}{2})} \tag{4.3.7}$$

which implies

$$\log g! < \tfrac{1}{2} \log 2\pi + (g+\tfrac{1}{2}) \log (g+\tfrac{1}{2}) - (g+\tfrac{1}{2}).$$

Hence,

$$E \log g! < \tfrac{1}{2} \log 2\pi + E\ (g+\tfrac{1}{2}) \log (g+\tfrac{1}{2}) - E\ (g+\tfrac{1}{2}).$$

However, since

$$(g+\tfrac{1}{2}) > \log (g+\tfrac{1}{2}) \quad \text{for all integers,}$$

we have

$$E \log g! < \tfrac{1}{2} \log 2\pi + E\ (g+\tfrac{1}{2})^2 - E\ (g+\tfrac{1}{2}) < \infty \quad .$$

So, finally, we have

$$\sum_g F(g,\underline{\theta},\underline{a}) \log F(g,\underline{\theta},\underline{a}) > \sum f(g,\theta_o,a_o) \log f(g,\theta_o,a_o)$$
$$> -\infty \tag{4.3.8}$$

and condition 1 is satisfied.

<u>Condition 2</u>: Since linear combinations of Poisson distributions are identifiable, Teicher [17], condition 2 is

satisfied.

Condition 3: Since $F(g,\underline{\theta}_o,a_o)$ is nothing more than the weighted sum of Poisson probability functions, each admitting continuous first derivatives, the condition is satisfied.

Thus, for the Poisson demand problem with only the total frequencies known, we have insured the existence of a consistent, asymptotically unbiased and normally distributed solution to the likelihood equations with the asymptotic variance-covariance matrix given by the inverse, assuming it exists, of the information matrix. Note that this implies, for the full data case, that the unique solution to the likelihood equations is consistent, asymptotically unbiased and normally distributed with covariance matrix equal to the inverse of the information matrix, the elements of which were developed earlier in this chapter.

As we have just seen, for the case in which only the total frequencies are known, the elements of the inverse of the dispersion matrix of the estimators are calculated from the various second derivatives of the likelihood function. In this case

$$\text{Log } L = \sum_g N_g \log F(g,\underline{\theta},\underline{a}). \tag{4.3.9}$$

The second derivatives are

$$\frac{\partial^2 \log L}{\partial \theta_j^2} = \sum_{g=bj}^{\infty} N_g \left[\frac{F(g,\underline{\theta},\underline{a}) \frac{\partial^2}{\partial \theta_j^2} F(g,\underline{\theta},\underline{a}) - \left(\frac{\partial}{\partial \theta_j} F(g,\underline{\theta},\underline{a})\right)^2}{F^2(g,\underline{\theta},\underline{a})} \right] \quad (4.3.10)$$

$$\frac{\partial^2 \log L}{\partial a_j^2} = \sum_{g=bj}^{\infty} N_g \left[\frac{F(g,\underline{\theta},\underline{a}) \frac{\partial^2}{\partial a_j^2} F(g,\underline{\theta},\underline{a}) - \left(\frac{\partial}{\partial a_j} F(g,\underline{\theta},\underline{a})\right)^2}{F^2(g,\underline{\theta},\underline{a})} \right] \quad (4.3.11)$$

$$\frac{\partial^2 \log L}{\partial \theta_i \, \partial \theta_j} = \sum_{g=bj}^{\infty} N_g \left[\frac{F(g,\underline{\theta},\underline{a}) \frac{\partial^2}{\partial \theta_i \partial \theta_j} F(g,\underline{\theta},\underline{a})}{F^2(g,\underline{\theta},\underline{a})} - \frac{\frac{\partial}{\partial \theta_i} F(g,\underline{\theta},\underline{a}) \frac{\partial}{\partial \theta_j} F(g,\underline{\theta},\underline{a})}{F^2(g,\underline{\theta},\underline{a})} \right] \quad (4.3.12)$$

$$\frac{\partial^2 \log L}{\partial a_i \, \partial a_j} = \sum_{g=bj}^{\infty} N_g \left[\frac{F(g,\underline{\theta},\underline{a}) \frac{\partial^2}{\partial a_i \partial a_j} F(g,\underline{\theta},\underline{a})}{F^2(g,\underline{\theta},\underline{a})} - \frac{\frac{\partial}{\partial a_i} F(g,\underline{\theta},\underline{a}) \frac{\partial}{\partial a_j} F(g,\underline{\theta},\underline{a})}{F^2(g,\underline{\theta},\underline{a})} \right] \quad (4.3.13)$$

$$\frac{\partial^2 \log L}{\partial \theta_j \, \partial a_i} = \sum_{g=bj}^{\infty} N_g \left[\frac{F(g,\underline{\theta},\underline{a}) \frac{\partial^2}{\partial \theta_j \partial a_i} F(g,\underline{\theta},\underline{a})}{F^2(g,\underline{\theta},\underline{a})} - \frac{\frac{\partial}{\partial \theta_j} F(g,\underline{\theta},\underline{a}) \frac{\partial}{\partial a_i} F(g,\underline{\theta},\underline{a})}{F^2(g,\underline{\theta},\underline{a})} \right] \quad (4.3.14)$$

$$\frac{\partial^2 \log L}{\partial a_j\, \partial \theta_j} = \sum_{g=bj}^{\infty} N_g\Bigg[\frac{F(g,\underline{\theta},\underline{a})\, \frac{\partial^2}{\partial a_j \partial \theta_j} F(g,\underline{\theta},\underline{a})}{F^2(g,\underline{\theta},\underline{a})}$$

$$- \frac{\frac{\partial}{\partial a_j} F(g,\underline{\theta},\underline{a})\, \frac{\partial}{\partial \theta_j} F(g,\underline{\theta},\underline{a})}{F^2(g,\underline{\theta},\underline{a})}\Bigg] \qquad (4.3.15)$$

Notice that in equations (4.3.10)-(4.3.15) whenever i and j appeared together j was assumed to be larger.

The N_g are multinomially distributed with

$$E\ N_g = N\ F(g,\underline{\theta},\underline{a})\ . \qquad (4.3.16)$$

Upon taking the expectations and simplifying, equations (4.3.10)-(4.3.15) become

$$E\ \frac{\partial^2 \log L}{\partial \theta_j^2} = -N \sum_{g=bj}^{\infty} \frac{(f(g \text{ diff } bj-1,\theta_j,a_j)-f(g \text{ diff } bj,\theta_j,a_j))^2}{F(g,\underline{\theta},\underline{a})} \qquad (4.3.17)$$

$$E\ \frac{\partial^2 \log L}{\partial a_j^2} = -N \sum_{g=bj}^{\infty} \frac{(\frac{1}{a_j} f(g \text{ diff } bj,\theta_j,a_j)-\frac{1}{a_\beta} f(g \text{ diff } \beta b,\theta_\beta,a_\beta))^2}{F(g,\underline{\theta},\underline{a})} \qquad (4.3.18)$$

where the second term in the numerator does not contribute until $g \geq \beta b$. For example, let β=2 and b=6.

$$F(11,\underline{\theta},\underline{a}) = \sum_{k=0}^{1} f(11 \text{ diff } 6k,\theta_k,a_k). \qquad (4.3.19)$$

We see that a_2 does not appear in (4.3.19) and therefore the derivative with respect to a_2 of (4.3.19) is zero. However, when g=14, we have

$$F(14,\underline{\theta},\underline{a}) = \sum_{k=0}^{2} f(14 \text{ diff } 6k,\theta_k,a_k), \qquad (4.3.20)$$

with a_2 appearing in (4.3.20) and hence, the derivative of (4.3.20) with respect to a_2 is non-zero.

$$E\,\frac{\partial^2 \log L}{\partial\theta_i\,\partial\theta_j} = -N\sum_{g=bj}^{\infty}\Big[\Big(\frac{f(g \text{ diff } bj-1,\theta_j,a_j) - f(g \text{ diff } bj,\theta_j,a_j)}{F(g,\underline{\theta},\underline{a})}\Big)$$

$$\cdot(f(g \text{ diff } bi-1,\theta_i,a_i)-f(g \text{ diff } bi,\theta_i,a_i))\Big]. \qquad (4.3.21)$$

Again, the terms preceded by a minus sign in the numerator do not contribute until $g \geq \beta b$.

$$E\,\frac{\partial^2 \log L}{\partial\theta_j\,\partial a_i} = -N\sum_{g=bj}^{\infty}\Big[\Big(\frac{f(g \text{ diff } bj-1,\theta_j,a_j)-f(g \text{ diff } bj,\theta_j,a_j)}{F(g,\underline{\theta},\underline{a})}$$

$$\cdot\Big(\frac{1}{a_i}f(g \text{ diff } bi,\theta_i,a_i)-\frac{1}{a_\beta}f(g \text{ diff } \beta b,\theta_\beta,a_\beta)\Big)\Big]. \qquad (4.3.22)$$

The term that contains θ_β and a_β does not contribute until $g \geq \beta b$.

$$E\,\frac{\partial^2 \log L}{\partial a_j\,\partial\theta_j} = -N \sum_{g=bj}^{\infty} \Big[\Big(\frac{f(g \text{ diff } bj-1,\theta_j,a_j)-f(g \text{ diff } bj,\theta_j,a_j)}{F(g,\underline{\theta},\underline{a})}\Big)$$

$$\cdot\Big(\frac{1}{a_j}f(g \text{ diff } bj,\theta_j,a_j)-\frac{1}{a_\beta}f(g \text{ diff } \beta b,\theta_\beta,a_\beta)\Big)\Big]. \quad (4.3.23)$$

As before, the term involving θ_β and a_β does not contribute until $g \geq \beta b$.

The matrix whose elements are the negative of the terms just derived will be the information matrix. The variances and covariance of the estimators will be the elements of the inverse of the information matrix.

4.4 Monte Carlo Study of the Variances of the Maximum Likelihood Estimators

It was decided to compare the large sample variances of the estimators when all the data are available with the variances of the estimators when all the partial frequencies are missing with only the total frequencies known, in order to gain an insight on how the horizontal grouping affects the variances. Consideration of any of the cases of data arrangement between full data and totally, horizontally grouped data was impractical due to the large number of possible combinations of partially complete data.

The variances of the estimators obtained in the totally, horizontally grouped demand problem were investigated by Monte Carlo simulation on an IBM 7040 computer.

For given known values of the parameters involved, 1000 random samples of a predetermined size were generated from the demand distribution. The estimates of the parameters were obtained by using the iterative procedure discussed in Chapter III and section 4.3 of the present chapter. From these 1000 values of the estimates of the parameters, the variance of each estimator, including its standard error, was calculated. The results of the Monte Carlo study are given in tables (4.1)-(4.6). In each of the six tables, it will be noticed that the Monte Carlo variances are close to the large sample variances obtained from the diagonal of the dispersion matrices.

As a result of the demand model, in any practical application, the sample space of the component distribution containing θ_j is smaller by a bulk size than the sample space of the distribution containing θ_{j-1}. Hence, the distribution containing θ_β has the smallest sample space of any of the distributions considered and thus it would appear that the variances of $\hat{\theta}_\beta$ and $\hat{a}_\beta$ would be much larger than the variances of some of the other estimators. However, such is not the case as can be seen in the six tables.

From studying the elements of the information matrix in both the full data and horizontally grouped data cases, equations (4.2.7)-(4.2.20) and (4.3.17)-(4.3.23) respectively,

TABLE 4.1

Variance Comparison, b=6 (6 parameters)

Monte Carlo	Large Sample
Var θ_o = 0.071329 (.009048)*	Var θ_o = 0.065670
Var θ_1 = 0.018792 (.002851)	Var θ_1 = 0.019046
Var θ_2 = 0.002361 (.000173)	Var θ_2 = 0.002473
Var a_o = 0.000938 (.000106)	Var a_o = 0.000860
Var a_1 = 0.000870 (.000094)	Var a_1 = 0.000861
Var a_2 = 0.000428 (.000060)	Var a_2 = 0.000457

$\theta_o = 4.0$ $\theta_1 = 2.0$ $\theta_2 = 0.4$

$a_o = 0.33$ $a_1 = 0.34$ $a_2 = 0.33$

Bulk Size = 6

Sample Size = 500

*The quantities in parentheses are the standard errors associated with each variance.

TABLE 4.2

Variance Comparison, b=6 (10 parameters)

Monte Carlo	Large Sample
Var θ_o = 0.026074 (.003105)	Var θ_o = 0.025403
Var θ_1 = 0.050141 (.006417)	Var θ_1 = 0.053807
Var θ_2 = 0.127383 (.019195)	Var θ_2 = 0.153354
Var θ_3 = 0.315606 (.105881)	Var θ_3 = 0.147088
Var θ_4 = 0.043542 (.006186)	Var θ_4 = 0.046030
Var a_o = 0.000147 (.000018)	Var a_o = 0.000147
Var a_1 = 0.000307 (.000054)	Var a_1 = 0.000308
Var a_2 = 0.003716 (.000407)	Var a_2 = 0.004054
Var a_3 = 0.003022 (.000323)	Var a_3 = 0.003689
Var a_4 = 0.000217 (.000035)	Var a_4 = 0.000211

$\theta_o = 1.0$ $\theta_1 = 3.2$ $\theta_2 = 5.6$ $\theta_3 = 2.7$ $\theta_4 = 1.4$

$a_o = 0.08$ $a_1 = 0.17$ $a_2 = 0.50$ $a_3 = 0.17$ $a_4 = 0.08$

Bulk Size = 6

Sample Size = 500

TABLE 4.3

Variance Comparison, b=7 (10 parameters)

Monte Carlo	Large Sample
Var θ_o = 0.024794 (.002177)	Var θ_o = 0.025079
Var θ_1 = 0.046325 (.005708)	Var θ_1 = 0.046315
Var θ_2 = 0.075092 (.011769)	Var θ_2 = 0.078534
Var θ_3 = 0.105330 (.025509)	Var θ_3 = 0.089345
Var θ_4 = 0.041953 (.003821)	Var θ_4 = 0.037903
Var a_o = 0.000150 (.000021)	Var a_o = 0.000147
Var a_1 = 0.000287 (.000047)	Var a_1 = 0.000293
Var a_2 = 0.001408 (.000216)	Var a_2 = 0.001566
Var a_3 = 0.001225 (.000136)	Var a_3 = 0.001335
Var a_4 = 0.000162 (.000011)	Var a_4 = 0.000159

θ_o = 1.0 θ_1 = 3.2 θ_2 = 5.6 θ_3 = 2.7 θ_4 = 1.4

a_o = 0.08 a_1 = 0.17 a_2 = 0.50 a_3 = 0.17 a_4 = 0.08

Bulk Size = 7

Sample Size = 500

TABLE 4.4

Variance Comparison, b=8 (10 parameters)

Monte Carlo	Large Sample
Var θ_o = 0.024896 (.003254)	Var θ_o = 0.025013
Var θ_1 = 0.043160 (.004627)	Var θ_1 = 0.042364
Var θ_2 = 0.049552 (.006665)	Var θ_2 = 0.050635
Var θ_3 = 0.061692 (.010593)	Var θ_3 = 0.061218
Var θ_4 = 0.039943 (.004462)	Var θ_4 = 0.035752
Var a_o = 0.000149 (.000025)	Var a_o = 0.000147
Var a_1 = 0.000299 (.000061)	Var a_1 = 0.000287
Var a_2 = 0.000880 (.000117)	Var a_2 = 0.000887
Var a_3 = 0.000634 (.000068)	Var a_3 = 0.000667
Var a_4 = 0.000150 (.000020)	Var a_4 = 0.000150

$\theta_o = 1.0$ $\theta_1 = 3.2$ $\theta_2 = 5.6$ $\theta_3 = 2.7$ $\theta_4 = 1.4$

$a_o = 0.08$ $a_1 = 0.17$ $a_2 = 0.50$ $a_3 = 0.17$ $a_4 = 0.08$

Bulk Size = 8

Sample Size = 500

TABLE 4.5

Variance Comparison, b=9 (10 parameters)

Monte Carlo	Large Sample
Var θ_o = 0.022857 (.003318)	Var θ_o = 0.025002
Var θ_1 = 0.038778 (.005350)	Var θ_1 = 0.040036
Var θ_2 = 0.036120 (.005893)	Var θ_2 = 0.037797
Var θ_3 = 0.052841 (.005934)	Var θ_3 = 0.047380
Var θ_4 = 0.033679 (.005632)	Var θ_4 = 0.035192
Var a_o = 0.000151 (.000018)	Var a_o = 0.000147
Var a_1 = 0.000259 (.000037)	Var a_1 = 0.000283
Var a_2 = 0.000667 (.000086)	Var a_2 = 0.000658
Var a_3 = 0.000417 (.000060)	Var a_3 = 0.000439
Var a_4 = 0.000144 (.000023)	Var a_4 = 0.000148

$\theta_o = 1.0$ $\theta_1 = 3.2$ $\theta_2 = 5.6$ $\theta_3 = 2.7$ $\theta_4 = 1.4$

$a_o = 0.08$ $a_1 = 0.17$ $a_2 = 0.50$ $a_3 = 0.17$ $a_4 = 0.08$

Bulk Size = 9

Sample Size = 500

TABLE 4.6

Variance Comparison, b=10 (10 parameters)

Monte Carlo	Large Sample
Var θ_o = 0.027557 (.004236)	Var θ_o = 0.025002
Var θ_1 = 0.038934 (.004815)	Var θ_1 = 0.038685
Var θ_2 = 0.031666 (.005242)	Var θ_2 = 0.031103
Var θ_3 = 0.039453 (.006476)	Var θ_3 = 0.040159
Var θ_4 = 0.036126 (.003044)	Var θ_4 = 0.035047
Var a_o = 0.000158 (.000018)	Var a_o = 0.000147
Var a_1 = 0.000277 (.000034)	Var a_1 = 0.000283
Var a_2 = 0.000564 (.000094)	Var a_2 = 0.000568
Var a_3 = 0.000350 (.000033)	Var a_3 = 0.000350
Var a_4 = 0.000145 (.000022)	Var a_4 = 0.000147

$\theta_o = 1.0$ $\theta_1 = 3.2$ $\theta_2 = 5.6$ $\theta_3 = 2.7$ $\theta_4 = 1.4$

$a_o = 0.08$ $a_1 = 0.17$ $a_2 = 0.50$ $a_3 = 0.17$ $a_4 = 0.08$

Bulk Size = 10

Sample Size = 500

it can be seen that the information matrix is multiplied by the scalar N and hence the dispersion matrix is multiplied by 1/N. Therefore, the effect of sample size is such that the variances of the estimators when the sample size is N are k times greater than those when the sample size is kN. For this reason, only one sample size, N=500, will be considered in the discussion of the trends observed in the asymptotic variances of the estimators.

In order to indicate what effect bulk size has on the variance of the estimators when different numbers of parameters are present in the model, the dispersion matrix was calculated under a variety of conditions and the variances were studied and compared with the full data variances. Thus we are able to see how a complete sample increases the precision of the estimator as compared to the totally, horizontally grouped sample. The results are tabulated and can be found in tables (4.7) and (4.8).

From studying tables (4.7) and (4.8), the effect of changing bulk size is very evident. As the bulk size increases, the variances of the estimators in the totally, horizontally grouped situation approach the variances of the estimators in the full data case. This can be explained in the following manner. For the Poisson distribution, the variance equals the mean and therefore if the mean is small, the "spread" of the distribution is also small. Hence, the larger the bulk size the more nearly the total observed

TABLE 4.7

Varying Bulk Size (10 parameters)

	Bulk Size					
	6	7	8	9	10	Full Data
Var $\hat{\theta}_o$	.025403	.025079	.025013	.025002	.025002	.025000
Var $\hat{\theta}_1$	.053807	.046315	.042364	.040036	.038685	.037647
Var $\hat{\theta}_2$	.153354	.078534	.050635	.037797	.031103	.022400
Var $\hat{\theta}_3$	.147088	.089345	.061218	.047380	.040159	.031765
Var $\hat{\theta}_4$	.046030	.037903	.035752	.035192	.035047	.035000
Var $\hat{a}_o$	.000147	.000147	.000147	.000147	.000147	.000147
Var $\hat{a}_1$	.000308	.000293	.000286	.000284	.000283	.000282
Var $\hat{a}_2$	.004050	.001566	.000888	.000658	.000568	.000500
Var $\hat{a}_3$	.003639	.001335	.000668	.000439	.000350	.000282
Var $\hat{a}_4$	.000206	.000160	.000152	.000151	.000147	.000147

$\theta_o = 1.0$ $\theta_1 = 3.2$ $\theta_2 = 5.6$ $\theta_3 = 2.7$ $\theta_4 = 1.4$

$a_o = 0.08$ $a_1 = 0.17$ $a_2 = 0.50$ $a_3 = 0.17$ $a_4 = 0.08$

Sample Size = 500

TABLE 4.8 - Varying Bulk Size (20 parameters).

	Bulk Size					
	6	7	8	9	10	Full Data
Var $\hat{\theta}_o$	.070265	.070036	.070004	.070000	.070000	.070000
Var $\hat{\theta}_1$	.030321	.030056	.030008	.030001	.030001	.030000
Var $\hat{\theta}_2$	.025022	.024568	.024467	.024448	.024445	.024444
Var $\hat{\theta}_3$	.019391	.018776	.018615	.018580	.018573	.018571
Var $\hat{\theta}_4$	.038201	.033352	.030690	.029318	.028729	.028421
Var $\hat{\theta}_5$	.072133	.055486	.047933	.044165	.042188	.040000
Var $\hat{\theta}_6$	.357607	.188688	.130616	.104798	.091426	.074286
Var $\hat{\theta}_7$	1.13275	.443029	.245794	.170089	.135526	.102222
Var $\hat{\theta}_8$	.189935	.127062	.099215	.086585	.080812	.076667
Var $\hat{\theta}_9$	.182814	.158770	.152226	.150521	.150111	.15000
Var a_o	.000039	.000039	.000039	.000039	.000039	.000039
Var a_1	.000113	.000113	.000113	.000113	.000113	.000113
Var a_2	.000164	.000164	.000164	.000164	.000164	.000164
Var a_3	.000242	.000241	.000241	.000241	.000241	.000241
Var a_4	.000328	.000315	.000310	.000309	.000308	.000308
Var a_5	.000363	.000331	.000318	.000312	.000310	.000308
Var a_6	.000514	.000334	.000281	.000259	.000250	.000241
Var a_7	.000498	.000280	.000213	.000186	.000174	.000164
Var a_8	.000473	.000204	.000143	.000124	.000117	.000113
Var a_9	.000045	.000041	.000041	.000039	.000039	.000039

θ_o=0.7 θ_1=0.9 θ_2=1.1 θ_3=1.3 θ_4=2.7 θ_5=3.8 θ_6=5.2 θ_7=4.6 θ_8=2.3 θ_9=1.5 a_o=.02 a_1=.06 a_2=.09 a_3=.14 a_4=.19 a_5=.19 a_6=.14 a_7=.09 a_8=.06 a_9=.02 Sample Size = 500 frequencies are equal to the partial frequencies of the dominant component distribution, e.g. suppose that the bulk size is six, then from a table of Poisson probabilities with $\theta_1 = 4.2$, $\theta_2 = 1.1$, and $a_1 = a_2 = 0.5$, the probability of having a demand for 7 single items is 0.0343 and the probability of having a demand for 7 items by means of a bulk of six and a single is 0.1831. While the probability of a demand for 7 items by means of a bulk and a single item is approximately six times larger than that for having a demand for 7 single items, it is still reasonably likely that a demand for 7 singles would be made. However, consider the same probelm when the bulk size is 12. The probability of a demand for 13 single items (one more than the bulk size - as considered when the bulk size was 6) is 0.00015 whereas the probability of having a demand for 13 items by means of a bulk of twelve and a single is about 0.1831. In this case, with a large probability, the total frequency of 13 items equals the partial frequency of a bulk and a single. Hence, we are in the full data case for all practical purposes and for this reason as the bulk size increases the variances of the estimators approach those of the full data case. Following the same argument, increasing the size of the parameters would have the same effect as decreasing the bulk size. It is rather interesting to note the inverse similarity between bulk size in this chapter and group size in Chapter II. Here, for larger bulk sizes the precision is good, while for the situation discussed in Chapter II, large group sizes result in less precision.

Recall that for the Poisson distribution the variance of the maximum likelihood estimator of the parameter is directly proportional to the size of the parameter. Since we are dealing with a combination of Poisson distributions, it was thought that the size of the parameter might affect the variance of the estimators. In tables (4.7) and (4.8) the parameters are allowed to vary and it is impossible to determine whether or not the parameters are being estimated with equal precision. In order to study the precision with which the various parameters are estimated, all the θ values were made equal as were the a values. The variance-covariance matrix of the estimators of the estimators was calculated and the diagonal elements of this matrix appear in tables (4.9) - (4.13). It can be seen from these tables that the initial and latter values of the parameter vectors are estimated with more precision than those toward the center of the vectors. The other trends explained earlier in this chapter can also be observed and they are unchanged.

In the preceding work, we have only considered the diagonal elements of the dispersion matrix of the estimators of the parameters for the demand problem. While these individual elements are important, it is difficult to obtain

TABLE 4.9

Varying Bulk Size when $\underline{\theta=1.0}$ (10 parameters)

	Bulk Size								
	3	4	5	6	7	8	9	10	Full Data
Var θ_o	.019362	.012599	.010697	.010162	.010032	.010005	.010001	.010000	.010000
Var θ_1	.021547	.012859	.010734	.010167	.010032	.010005	.010001	.010000	.010000
Var θ_2	.021586	.012859	.010734	.010167	.010032	.010005	.010001	.010000	.010000
Var θ_3	.021496	.012857	.010734	.010167	.010032	.010005	.010001	.010000	.010000
~~Var θ_4~~	~~.010943~~	~~.010187~~	~~.010033~~	~~.010005~~	~~.010001~~	~~.010000~~	~~.010000~~	.010000	.010000
Var a_o	.000385	.000330	.000322	.000320	.000320	.000320	.000320	.000320	.000320
Var a_1	.000439	.000339	.000323	.000320	.000320	.000320	.000320	.000320	.000320
Var a_2	.000440	.000339	.000323	.000320	.000320	.000320	.000320	.000320	.000320
Var a_3	.000440	.000339	.000323	.000320	.000320	.000320	.000320	.000320	.000320
Var a_4	.000387	.000330	.000322	.000320	.000320	.000320	.000320	.000320	.000320

$\theta_o = \theta_1 = \theta_2 = \theta_3 = \theta_4 = 1.0$ $a_o = a_1 = a_2 = a_3 = a_4 = 0.20$

Sample Size = 500

TABLE 4.10

Varying Bulk Size when $\underline{\theta=2.0}$ (10 parameters)

	Bulk Size								
	3	4	5	6	7	8	9	10	Full Data
Var θ_o	.098389	.045083	.029768	.023822	.021379	.020444	.020126	.020032	.020000
Var θ_1	.224443	.063188	.033274	.024581	.021551	.020482	.020134	.020033	.020000
Var θ_2	.251102	.064345	.033326	.024583	.021551	.020482	.020134	.020033	.020000
Var θ_3	.217124	.062223	.033177	.024574	.021550	.020482	.020134	.020033	.020000
Var θ_4	.039477	.026100	.021963	.020579	.020154	.020037	.020008	.020002	.020000
Var a_o	.001078	.000451	.000351	.000328	.000322	.000320	.000320	.000320	.000320
Var a_1	.001451	.000526	.000375	.000335	.000324	.000321	.000320	.000320	.000320
Var a_2	.001611	.000536	.000375	.000335	.000324	.000321	.000320	.000320	.000320
Var a_3	.001498	.000529	.000375	.000335	.000324	.000321	.000320	.000320	.000320
Var a_4	.001241	.000456	.000351	.000328	.000322	.000321	.000320	.000320	.000320

$\theta_o = \theta_1 = \theta_2 = \theta_3 = \theta_4 = 2.0$ $\qquad$ $a_o = a_1 = a_2 = a_3 = a_4 = 0.20$

Sample Size = 500

TABLE 4.11
Varying Bulk Size when $\underline{\theta}$= $\underline{3.0}$ (10 parameters)

	Bulk Size								
	3	4	5	6	7	8	9	10	Full Data
Var θ_o	.348797	.117325	.065415	.046951	.038373	.033890	.031627	.030608	.030000
Var θ_1	1.87009	.290784	.102155	.057625	.041770	.034940	.031935	.030694	.030000
Var θ_2	2.74077	.329594	.105391	.057962	.041802	.034942	.031936	.030694	.030000
Var θ_3	1.78283	.278258	.100221	.057283	.041718	.034934	.031935	.030694	.030000
Var θ_4	.130977	.060546	.042327	.035460	.032309	.030862	.030282	.030083	.030000
Var a_o	.004110	.000871	.000456	.000364	.000335	.000325	.000322	.000320	.000320
Var a_1	.007289	.001071	.000522	.000393	.000347	.000330	.000323	.000321	.000320
Var a_2	.007907	.001159	.000533	.000394	.000348	.000330	.000323	.000321	.000320
Var a_3	.007383	.001088	.000524	.000393	.000347	.000330	.000323	.000321	.000320
Var a_4	.005289	.000937	.000461	.000364	.000335	.000326	.000322	.000321	.000320

$\theta_o = \theta_1 = \theta_2 = \theta_3 = \theta_4 = 3.0$ $a_o = a_1 = a_2 = a_3 = a_4 = 0.20$

Sample Size = 500

TABLE 4.12

Varying Bulk Size when $\underline{\theta=4.0}$ (10 parameters)

	Bulk Size								
	3	4	5	6	7	8	9	10	Full Data
Var θ_o	1.05819	.264744	.123498	.079876	.061338	.051924	.046611	.043421	.040000
Var θ_1	11.4803	1.13841	.279724	.124023	.077335	.058502	.049385	.044513	.040000
Var θ_2	21.9395	1.51413	.309631	.127763	.077905	.058592	.049398	.044514	.040000
Var θ_3	11.0808	1.07133	.269924	.122064	.076889	.058401	.049365	.044510	.040000
Var θ_4	.400020	.129708	.074608	.056312	.048351	.044349	.042160	.040951	.040000
Var a_o	.014428	.001946	.000677	.000433	.000364	.000338	.000328	.000323	.000320
Var a_1	.045165	.002681	.000798	.000498	.000391	.000352	.000334	.000326	.000320
Var a_2	.037429	.002884	.000845	.000496	.000393	.000353	.000334	.000326	.000320
Var a_3	.045614	.002699	.000805	.000489	.000392	.000352	.000334	.000326	.000320
Var a_4	.027301	.002333	.000703	.000436	.000364	.000339	.000328	.000323	.000320

$\theta_o = \theta_1 = \theta_2 = \theta_3 = \theta_4 = 4.0$ $a_o = a_1 = a_2 = a_3 = a_4 = 0.20$

Sample Size = 500

TABLE 4.13

Varying Bulk Size when θ=5.0 (10 parameters)

	Bulk Size								
	3	4	5	6	7	8	9	10	Full Data
Var θ_o	2.89176	.556606	.218556	.127130	.091406	.073990	.064296	.058515	.050000
Var θ_1	53.8806	3.93862	.724865	.256556	.136879	.093093	.073117	.062791	.050000
Var θ_2	126.079	6.19998	.886534	.276642	.140263	.093746	.073246	.062816	.050000
Var θ_3	52.5947	3.68867	.690470	.249838	.135214	.092643	.072997	.062761	.050000
Var θ_4	1.12224	.267494	.126717	.085479	.068720	.060450	.055905	.053302	.050000
Var a_o	.044479	.004456	.001113	.000556	.000410	.000359	.000338	.000329	.000320
Var a_1	.244066	.008052	.001359	.000640	.000455	.000384	.000352	.000336	.000320
Var a_2	.153278	.007531	.001467	.000666	.000461	.000386	.000352	.000336	.000320
Var a_3	.258935	.007949	.001371	.000644	.000456	.000385	.000352	.000336	.000320
Var a_4	.104093	.006036	.001213	.000566	.000412	.000359	.000338	.000329	.000320

$\theta_o = \theta_1 = \theta_2 = \theta_3 = \theta_4 = 5.0$ $a_o = a_1 = a_2 = a_3 = a_4 = 0.20$

Sample Size = 500

a measure of the over-all variability in the estimators of these parameters from them. The determinant of the dispersion matrix, called the generalized variance, is a scalar value which does give us an insight into this overall variability. Tables (4.14) and (4.15) illustrate how increasing the bulk size affects the generalized variance.

TABLE 4.14

Bulk Size vs. Generalized Variance (10 parameters)

Bulk Size	Gen. Var.
6	73.21608×10^{-13}
7	9.17677×10^{-13}
8	2.57541×10^{-13}
9	1.17095×10^{-13}
10	0.71404×10^{-13}
Full Data	0.34682×10^{-13}

$\theta_o = 1.0$ $\theta_1 = 3.2$ $\theta_2 = 5.6$

$\theta_3 = 2.7$ $\theta_4 = 1.4$ $a_o = 0.08$

$a_1 = 0.17$ $a_2 = 0.50$ $a_3 = 0.17$

$a_4 = 0.08$

Sample Size = 50

TABLE 4.15

Bulk Size vs. Generalized Variance (20 parameters)

Bulk Size	Gen. Var.
4	$905,603,340 \times 10^{-29}$
5	$62,397 \times 10^{-29}$
6	353.21×10^{-29}
7	19.84×10^{-29}
8	3.77×10^{-29}
9	1.42×10^{-29}
10	0.8008×10^{-29}
Full Data	0.3999×10^{-29}

$\theta_o=0.7$ $\theta_1=0.9$ $\theta_2=1.1$ $\theta_3=1.3$

$\theta_4=2.7$ $\theta_5=3.8$ $\theta_6=5.2$ $\theta_7=4.6$

$\theta_8=2.3$ $\theta_9=1.5$ $a_o=.02$ $a_1=.06$

$a_2=.09$ $a_3=.14$ $a_4=.19$ $a_5=.19$

$a_6=.14$ $a_7=.09$ $a_8=.06$ $a_9=.02$

Sample Size = 50

4.5 General Comments Concerning the Demand Problem

In Chapter III, the M.L. estimation procedure for combinations of distributions diff kb was developed without regard to specific component distributions. In the present chapter, we have applied these methods to the demand problem in which the underlying distribution was taken to be Poisson. The Negative Binomial Distribution can also be applied to problems concerning item demands. Its use is justified in either of two ways.

1) Suppose that the number of demands per unit time follows the Poisson distribution whereas the number of items per demand follows the Logarithmic Distribution, then the number of items demanded will be distributed as a Poisson compounded with the logarithmic distribution. It is well known that the compounding of these two distributions results in the Negative Binomial Distribution. See, for instance, Feller [7], page 271.

2) If one is willing to assume a prior distribution on the parameter in each one of the Poisson distributions in the demand model and, as a result of some further knowledge, one knows that these priors are of the Gamma family, then one can also arrive at the Negative Binomial as the underlying distribution for the demand problem. See Kendall and Stuart [11], Vol 1, pages 129-130.

The only method of estimation studied in this dissertation has been that of maximum likelihood. The method of moments is another possibility. However, from the difficulty of the moment generating function and the fact that to obtain these estimators a system of $2\beta+1$ simultaneous equations must be solved for the Poisson demand problem, it hardly seems worth the effort considering the ease with which the maximum likelihood estimators are obtained.

4.6 A Further Application of Diff kb Combinations of Distributions

The combination of distributions diff kb probability model has possible applications in evaluating the performance of anti-missile missile systems for use against missiles which are capable of carrying multiple warheads. The random variable which would be of interest here would be the number of warheads destroyed by the anti-ballistic missile before they reach their targets. Since the warheads are carried in bulks of size b on an offensive missile, it is assumed that by destroying one missile carrying the multiple warheads, b warheads would be destroyed simultaneously; otherwise a single defensive missile could only destroy one warhead. Hence, the probability model

$$P(X=n) = \sum_{k=0}^{[n/b]} P(X=n \text{ diff } kb \mid \Omega=k)P(\Omega=k)$$

where $P(X=n)$ is the total probability of destroying n warheads; $P(X=n \text{ diff } kb \mid \Omega=k)$ is the probability distribution of (n-kb) single warheads being destroyed given that k missiles carrying bulks of warheads has been destroyed; and $P(\Omega=k)$ is the probability of destroying k missiles carrying multiple warheads.

It is possible that the component probability functions do not all belong to the same family of distributions since there could be more than one type of defensive missile being used. One system could be developed to give distant protection, i.e. to destroy the incoming missiles at great distances from the target while another system could be used to back up the first system by destroying the warheads that penetrated the initial defensive system. From this model one could study the effects of bulk size on the probabilities of destruction of warheads and perhaps arrive at some optimum bulk size which minimizes the probability of destruction.

BIBLIOGRAPHY

[1] Bailey, N. T. J. (1964). The Elements of Stochastic Processes With Applications to the Natural Sciences. John Wiley and Sons, Inc., New York.

[2] Bowen, J. V., Jr. (1966). Some Properties of Conditional Distributions of A Special Type, M.S. Thesis. Virginia Polytechnic Institute, Blacksburg, Virginia.

[3] Cohen, A. C., Jr. (1960). "Estimation In the Truncated Poisson Distribution When Zeros and Some Ones Are Missing," Journal of the American Statistical Association, 55, pp. 342-348.

[4] Cohen, A. C., Jr. (1965). "Estimation In Mixtures of Poisson and Mixtures of Exponential Distributions," NASA Technical Memorandum, NASA TM X-53245, George C. Marshall Space Flight Center, Huntsville, Alabama.

[5] Cohen, A. C., Jr. (1965). "Estimation in Mixtures of Two Normal Distributions," University of Georgia, Department of Statistics, T. R. No. 13.

[6] Cohen, A. C., Jr. (1965). "Estimation In A Mixture of A Poisson With A Negative Binomial Distribution," University of Georgia, Department of Statistics, T. R. No. 15.

[7] Feller, W. (1957). An Introduction to Probability Theory and Its Applications (2nd edition). John Wiley and Sons, Inc., New York.

[8] Ford, L. R. (1925). "The Solution of Equations By the Method of Successive Approximations," American Mathematical Monthly, 32, pp. 272-287.

[9] Hartley, H. O. (1958). "Maximum Likelihood Estimation From Incomplete Data," Biometrics, 14, pp. 174-194.

[10] Hayles, G. C. (1967). Variances of Some Truncated Distributions, M. S. Thesis. Virginia Polytechnic Institute, Blacksburg, Virginia.

[11] Kendall, M. G. and Stuart, A. (1963). The Advanced Theory of Statistics, Vols. 1 and 2 (2nd edition). Hafner Publishing Company, New York.

[12] Lindley, D. V. (1950). "Grouping Corrections and Maximum Likelihood Equations," Proceedings Cambridge Philosophical Society, 46, pp. 106-110.

[13] Mood, A. M. and Graybill, F. A. (1963). Introduction to the Theory of Statistics (2nd edition). McGraw-Hill Book Company, Inc., New York.

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

[15] Tallis, G. M. (1967). "Approximate Maximum Likelihood Estimates From Grouped Data," Technometrics, 9, pp. 599-606.

[16] Taylor, A. E. (1959). Calculus with Analytic Geometry. Prentice Hall, Inc., New Jersey.

[17] Teicher, H. (1960). "On the Mixture of Distributions," Ann. Math. Stat., 31, pp. 55-73.

[18] Weiner, S. (1962). "Samples From Mixed-Exponential Populations," Mimeographed paper, ARINC Research Corporation, Washington, D. C., Research under Contract No. bsr_64508.

VITA

Walter H. Carter, Jr. was born March 20, 1941, in Winchester, Virginia. He attended Warren County High School and, in June 1959, was graduated from Clarke County High School in Berryville, Virginia. The following fall he enrolled at Richmond College of the University of Richmond. In June, 1963, he completed the requirements for a Bachelor of Science degree in Mathematics and Physics.

The author enrolled in graduate school at Virginia Polytechnic Institute in the Department of Mathematics in September, 1963, as a graduate assistant and obtained a Master of Science degree in Mathematics in June, 1966. In March, 1965, he entered the Department of Statistics of Virginia Polytechnic Institute. During the course of his graduate study in statistics, the author was supported by fellowship grants from the National Science Foundation, The National Aeronautics and Space Administration, and The Department of Health, Education, and Welfare.

He was elected into Phi Beta Kappa, Pi Mu Epsilon for Mathematics, and Sigma Pi Sigma for Physics.

The author is married to the former Judy Gott and they have two children, Bart and Kelly. Upon graduation, the author and his family will locate in Richmond, Virginia.

Walter H. Carter, Jr.

<u>GROUPING IN ITEM DEMAND PROBLEMS</u>

by

Walter H. Carter, Jr.

<u>ABSTRACT</u>

In this dissertation an iterative procedure, due to Hartley [9], for obtaining the maximum likelihood estimators of the parameters from underlying discrete distributions is studied for the case of grouped random samples. It is shown that when the underlying distribution is Poisson the process always converges and does so regardless of the initial values taken for the unknown parameter. In showing this, a rather interesting property of the Poisson distribution was derived. If one defines a connected group of integers to be such that it contains all the integers between and including its end points, it is shown that the variance of the subdistribution defined on this connected set is strictly less than the variance of the complete Poisson distribution. A Monte Carlo study was performed to indicate how increasing group sizes affected the variances of the maximum likelihood estimators.

As a result of a problem encountered by the Office of Naval Research, combinations of distributions diff kb were introduced. The difference between such combinations and the classical mixtures of distributions is that a new distribution must be considered whenever the random variable

in question increases by an integral multiple of a known integer constant, b. When all the data are present, the estimation problem is no more complicated than when estimating the individual parameters from the component distributions. However, it is pointed out that very frequently the observed samples are defective in the fact that none of the component frequencies are observed. Hence, horizontal grouping of the sample values occurs as opposed to the vertical grouping encountered previously in the one parameter Poisson case. An extension of the iterative procedure used to obtain the maximum likelihood estimator of the single parameter grouped Poisson distribution is made to obtain the estimators of the parameters in a horizontally grouped sample.

As a practical example, the component distributions were all taken to be from the Poisson family. The estimators were obtained and their properties were studied. The regularity conditions which are sufficient to show that a consistent and asymptotically normally distributed solution to the likelihood equations exist are seen to be satisfied for such combinations of Poisson distributions. Further, in the full data case, a set of jointly sufficient statistics is exhibited and since, in the presence of sufficient statistics, the solutions to the likelihood equations are unique, the estimators are consistent and asymptotically normal.

It is seen that such combinations of distributions can be applied to problems in item demands. A justification of the Poisson distribution is given for such applications, but it is also pointed out that the Negative Binomial distribution might be applicable. It is also shown that such a probability model might have an application in testing the efficiency of an anti-ballistic missile system when under attack by missiles which carry multiple warheads. However, no data were available and hence the study of this application could be carried no further.