THE UNIVERSITY OF CALGARY

INFERENCE THEORY FOR SOME GENERALIZED DISCRETE PROBABILITY MODELS

by

KIYA FELIX FAMOYE

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

AUGUST, 1986

ISBN Ø-315-32658-1

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled, "Inference Theory for Some Generalized Discrete Probability Models" submitted by Felix K. Famoye in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

P.C. Consul

P.C. Consul, supervisor,
Dept. of Maths. & Stats.

E.G. Enns, Dept. of Maths. & Stats.

W.W. Zwirner, Dept. of Educ. Psych.

Dudley Paul Johnson

D.P. Johnson, Dept. Maths. & Stats.

J.K. Wani, Dept. of Maths. & Stats.

M. Srivastava

M.S. Srivastava, external examiner,
University of Toronto

DATE: September 5, 1986

ABSTRACT

In this thesis, we studied estimation and hypothesis testing in some generalized discrete probability models.

The family of generalized Poisson distribution (GPD) characterized by two parameters, was defined by Consul and Jain (1973). The GPD models have been found useful in many different areas like queueing theory, branching process, genetics and ecology. The family of GPD models belongs to the class of Lagrangian probability distributions [Consul and Shenton, 1973]. The restricted GPD is a member of the class of modified power series distributions (MPSD), [Gupta, 1974]. Both the class of Lagrangian probability distributions and the class of MPSD also contain the families of the generalized negative binomial distribution (GNBD) and the generalized logarithmic series distribution (GLSD) among many others.

Some properties and applications of the GPD family and those of restricted GPD, the GNBD and the GLSD as members of the class of MPSD are reviewed in Chapter I.

In Chapter II, we have proved that the GLSD, GNBD and GPD are unimodal.

We investigated the problem of interval estimation in the class of MPSD in Chapter III. Both the cases of small and large samples are considered in setting two-sided $100(1-\alpha)\%$ confidence bounds for the parameters. By using the critical region for the

uniformly most powerful test, we have also obtained a uniformly most accurate one-sided confidence bound.

Chapters IV and V contain estimation in small and large samples for GPD model. Confidence intervals, likelihood intervals as well as likelihood regions are obtained for the parameters of GPD when the sample size is small. When the sample size is large, expressions for finding confidence intervals for each of the two parameters when one is unknown and when both are unknown are derived. Furthermore, we have discussed the problem of setting confidence regions for the two parameters.

From recent research work, meaningful interpretations have been given to the parameters of GPD when the model is used to describe a natural phenomenon. Quite often, a user of the model formulates hypotheses about these parameters. Accordingly, Chapter VI is devoted to tests of hypotheses on the parameters of the GPD. The cases in which one parameter is unknown and in which both parameters are unknown are separately considered.

Some estimation problems for the GNBD are discussed in Chapter VII. In conclusion, we briefly outline some unsolved problems which need further research work.

ACKNOWLEDGEMENTS

I wish to express my sincerest gratitude to my supervisor Professor P.C. Consul, for his supervision, encouragement, invaluable guidance and the time he spent with me during the preparation of this thesis.

I would like to thank my wife Busola, for her understanding and the much needed family cooperation she gave during this work.

I seize this opportunity to thank the Canadian Commonwealth Scholarship and Fellowship Committee that made it possible for me to study in Canada. The supplementary awards of the University of Calgary through GA(T) is gratefully acknowledged.

Finally, I would like to express my thanks to Maria Dourado for painstakingly and speedily typing the manuscript.

Dedicated to my parents

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 SOME CLASSES OF DISCRETE DISTRIBUTIONS

Let X be a random variable (r.v.) and let $F(x) = P(X \le x)$ be its distribution function. If $F(x)$ is a step function with only an enumerable number of steps and if the height of the step at $X = x_j$ is P_j, then

$$P(X = x_j) = P_j$$

and the r.v. X is said to be a discrete r.v. and also if $\sum_j P_j = 1$ then X is said to have a discrete probability distribution.

A variety of the earthly phenomena deal with random counts. Some examples are the number of a particular plant species per quadrant in an ecological habitat, the number of girls in a family of six in the city of Calgary, the number of bacteria per colony, the number of traffic accidents incurred over a period of time by the bus drivers in a city and the number of deaths due to epileptic disease in a city.

Any set of data which conforms with the above different types or some other form of random counts is adequately analysed by using a discrete probability distribution. However, it is sometimes possible to consider the problem by using a continuous distribution which is an approximation but it may not lead to a satisfactory result in all cases. In fact, one uses approximations, either when an exact method is not available or when the exact methods are too laborious to use. In effect, one should fit a continuous data with a continuous probability distribution and a discrete data with a discrete probability distribution. When the samples are very large, one may have to appeal to the central limit theorem in order to apply an appropriate continuous distribution.

It is very difficult to classify the various types of discrete distributions. We will not hesitate to state here that there is no hard and fast rule for this classification. Each author comes up with one or the other type of classification. The field of discrete probability distributions is so wide and diversified that it is hard to provide the definition of classes. Patel, Kapadia and Owen (1976) considered exponential family, Pearson distributions (in continuous case) and generalized power series distributions as families of probability distributions. They also considered among others, the binomial, the Poisson and the negative binomial distributions as different classes. On the other hand, Johnson and Kotz (1969) considered the generalized power series distributions, the systems defined by difference equations, mixture/compound and generalized distributions and the contagious distributions among others as major classes of discrete distributions. Also, they gave a number of discrete probability distributions which do not fall into any of the major classes. In this work, we shall follow the latter classification and refer to a member of a class as a family.

It must be pointed out that sometimes there may exist some relationships between any two or more members of a class with some members of another class. We shall now discuss the following four major classes: the mixture/compound and generalized distributions, the generalized power series distributions, the modified power series distributions and the Lagrangian probability distributions. The models we shall be considering in subsequent chapters are associated with these different classes.

1.1.1 THE MIXTURE/COMPOUND AND GENERALIZED DISTRIBUTIONS

When an applied statistician or a researcher in any subject area obtains a set of data, usually he determines the mean and the variance of the data and on the basis of the property that the mean is greater than the variance, the mean equals the variance or the mean is less than the variance he will try the binomial, the Poisson or the negative binomial distribution. It was discovered in many situations that none of these three distributions was the approporiate model for the data. In view of this, statisticians began to look for other types of distributions. This had in no small measure resulted in the development of various 'modified' forms of the classical distributions. The modification is generally in the form of mixing, compounding and generalization. Further with this modification, statisticians went as far as possible to obtain the discrete analogues of some well known continuous distributions. Cassie (1962), for example, obtained the discrete lognormal distribution by considering the log counts of sample data as normally distributed. Other forms of such modifications appear as truncated and modified distributions. The basic motivation was to obtain an appropriate discrete probability distribution to fit the observed data.

If $j \in T$ where T is any subset of the set of non-negative integers and if $\{P_j(x_1,x_2,\ldots,x_n)\}$ represent different cumulative distribution functions and if $w_j \geq 0$ such that

$$\sum_{j\in T} w_j = 1, \qquad \text{then}$$

$$P(x_1,x_2,\ldots,x_n) = \sum_{j\in T} w_j\, P_j(x_1,x_2,\ldots,x_n) \tag{1.1.1}$$

is also a proper cumulative distribution function. The distribution (1.1.1) is called a mixture of the distributions $\{P_j(x_1,x_2,\ldots,x_n)\}$.

If X is a r.v. whose probability mass function is

$$P(X=x \mid \boldsymbol{\Theta}=\theta) \tag{1.1.2}$$

for x in the domain of the distribution and if $\boldsymbol{\Theta}$ is now regarded as a new r.v. with its own probability distribution $f(\theta)$, say, then the given probability in (1.1.2) becomes a conditional probability and one can obtain the unconditional probability P(X=x) as follows:

$$P(X=x) = \begin{cases} \displaystyle\int P(X=x|\boldsymbol{\Theta}=\theta)f(\theta)d\theta, & \text{if } \theta \text{ is continuous} \\ \displaystyle\sum_{\theta} P(X=x|\boldsymbol{\Theta}=\theta)f(\theta), & \text{if } \theta \text{ is discrete} \end{cases} \tag{1.1.3}$$

This is written in a symbolic form as

$$X \wedge \boldsymbol{\Theta}.$$

P(X=x) in (1.1.3) is referred to as compound distribution. From the above definitions of mixture and compounding one can see that there is a relationship between the two terms. The term 'compound' distributions is usually used synonymously with the term 'mixture' distributions.

Suppose $g_1(t)$ and $g_2(t)$ are the probability generating functions of the random variables X and Y respectively. A new probability generating function for a different r.v. Z can be easily formulated as

$$H(t) = g_1(g_2(t)). \tag{1.1.4}$$

The new r.v. Z is said to be a generalized distribution of the previous random variable X. In (1.1.4), the distribution of Y is called the 'generalizer'. Symbolically, it is denoted

$$Z = X \vee Y.$$

A good exposition of the relationship between mixture, compounding and generalization is given by Blischke (1963). Some well known members of this class of discrete distributions are Poisson-Poisson, Poisson-Binomial, Neyman Types A, B and C, Thomas and Polya-Eggenberger.

1.1.2 THE GENERALIZED POWER SERIES DISTRIBUTIONS

Let $f(\theta)$ be an analytic function of θ such that

$$f(\theta) = \sum_{x=0}^{\infty} a(x)\, \theta^{x},$$

where $a(x) \geq 0$ for all x, then

$$P(X=x) = \begin{cases} a(x)\, \theta^{x}/f(\theta); & x = 0,1,2,\ldots,\ \theta > 0 \\ 0 \quad ; & \text{otherwise} \end{cases} \qquad (1.1.5)$$

is a power series distribution (PSD). This class of discrete distributions was introduced by Kosambi (1949) and Noack (1950) independently. Patil (1962) considered the domain of the distribution (1.1.5) to be the set T where T is a subset of the set of non-negative integers. This extended class is referred to as generalized power series distributions (GPSD).

The class of GPSD has got many interesting properties. The probability generating function of the distribution (1.1.5) is

$$g(t) = f(\theta t)/f(\theta).$$

It is observed that the truncated distributions, excluded in the class of PSD, are included in the class of GPSD. Thus, if a GPSD is truncated, the truncated distribution is also a GPSD. If each of $X_1, X_2, \ldots, X_n$ has a GPSD, then the sum $Y = \sum_1^n X_i$ also belongs to the same class with series function

$$\{f(\theta)\}^n.$$

A lot of work has been done on the GPSD. Some references are Tweedie and Veevers (1968), Patil (1962) and Khatri (1959) on properties, Roy and Mitra (1957) on estimation and Patil (1963) on applications. Some members of the GPSD class are the binomial distribution, the negative binomial distribution, the Poisson distribution and the logarithmic series distribution.

1.1.3 THE MODIFIED POWER SERIES DISTRIBUTIONS

A discrete r.v. X is said to have a modified power series distribution (MPSD), [Gupta, 1974], if its probability function is given by

$$P(X=x) = \begin{cases} a(x)\{g(\theta)\}^x/f(\theta); & x \in T \\ 0 \quad ; & \text{otherwise} \end{cases} \qquad (1.1.6)$$

where T is a subset of the set of non-negative integers, $a(x) \geq 0$, $g(\theta)$ and $f(\theta)$ are positive, finite and differentiable functions of θ and $f(\theta)$ is such that

$$f(\theta) = \sum_{x \in T} a(x) \{g(\theta)\}^x.$$

If $g(\theta)$ equals θ or is invertible, (1.1.6) reduces to the GPSD and if T is the set of non-negative integers, (1.1.6) becomes the PSD.

It is obvious that the PSD and the GPSD are special cases of the MPSD. A truncated MPSD is also an MPSD. Some members of the MPSD class are the binomial distribution, the negative binomial distribution, the generalized negative binomial distribution, the logarithmic series distribution, the generalized logarithmic series distribution, the Poisson distribution and the restricted generalized Poisson distribution.

1.1.4 THE LAGRANGIAN PROBABILITY DISTRIBUTIONS

The class of Lagrangian probability distributions (LPD) was introduced into the statistical literature by Consul and Shenton (1972). These probability distributions are generated by the well known Lagrange's expansion. Before the introduction of this class, Otter (1948) pointed out the applicability of Lagrange's expansion to branching processes in univariate situation. His result was later extended to the multivariate case by Good (1960). However, these authors seemed not to realize the usefulness of the Lagrange's expansion in generating discrete probability distributions.

Let $g(t)$ and $f(t)$ be two analytic functions of t which are successively differentiable and such that $g(1) = f(1) = 1$, $g(0) \neq 0$, then under the transformation

$$u = t/g(t) \tag{1.1.7}$$

one can consider f(t) as an implicit function of u, say h(u) and expand it as a power series in u within its circle of convergence by Lagrange's expansion. It can be easily shown that the function h(u) is a probability generating function in u and gives birth to a new r.v. X whose probability distribution is given by

$$P(X=x) = \begin{cases} \dfrac{1}{x!}\left[\dfrac{\partial^{x-1}}{\partial t^{x-1}}\left\{\left[g(t)\right]^{x} f'(t)\right\}\right]_{t=0} & ; \quad x \in T \\ f(0) & ; \quad x = 0 \end{cases} \tag{1.1.8}$$

where T is a subset of the set of non-negative integers. The probability generating function h(u) of the distribution in (1.1.8) is given by f(t) = h(u), where t = u·g(t), or by

$$h(u) = f(t) = f(0) + \sum_{x=1}^{\infty} \frac{u^x}{x!}\left[\frac{\partial^{x-1}}{\partial t^{x-1}}\left\{\left[g(t)\right]^{x} f'(t)\right\}\right]_{t=0} .$$

Consul and Shenton (1972) took f(t) and g(t) as probability generating functions defined on non-negative integers such that $g(0) \neq 0$. Consul (1981) pointed out that one can obtain a true probability distribution by removing the restriction that f(t) and g(t) need be probability generating functions. Thus, the class of Lagrangian probability distributions is widened. Consul further showed that the class of MPSD is a subclass of the extended class of LPD.

Some of the members of this class are the generalized negative binomial distribution, the generalized Poisson distribution, the generalized logarithmic series distribution, the Haight distribution and the Borel-Tanner distribution. Some papers have been published on the characteristics of this class. Among them are Consul and Shenton (1973) on interesting properties and Pakes and Speed (1977) on limiting theorems.

1.2 THE GENERALIZED NEGATIVE BINOMIAL DISTRIBUTION

Jain and Consul (1971) obtained the generalized negative binomial distribution (GNBD) by using the Lagrange's expansion. It is not a generalization of the negative binomial distribution in the sense of subsection 1.1.1 but it is to be remembered that the distribution encompasses not only the negative binomial distribution but also the binomial distribution among many other distributions as special cases.

Since the GNBD is a member of the class of LPD and coupled with the fact that the binomial distribution is a special case, some of the earlier writers referred to this distribution as the Lagrangian binomial distribution.

A random variable X is said to have a GNBD if its probability distribution is given by

$$P(X=x) = \begin{cases} \dfrac{m}{m+\beta x} \begin{bmatrix} m+\beta x \\ x \end{bmatrix} \theta^{x} (1-\theta)^{m+\beta x-x}; & x = 0,1,2,\ldots \\ 0 \quad ; & \text{otherwise} \end{cases} \tag{1.2.1}$$

where $0 < \theta < 1$, $m > 0$ and $\beta = 0$ or $1 \le \beta \le \theta^{-1}$. It reduces to the binomial distribution when $\beta = 0$ and m is an integer and to the negative binomial distribution when $\beta = 1$. For values of $\beta > 1$, it represents many other distributions which are very useful in problems of random walk, Mohanty (1977).

All the moments of this distribution exist if $\beta < \theta^{-1}$. The characterization of GNBD by zero regression was considered by Consul and Gupta (1980).

1.2.1 PROPERTIES OF GNBD

THE GENERATING FUNCTIONS: The probability generating function (p.g.f.) of the GNB distribution (1.2.1) is given by either

$$h(u) = f(t_1) = (1-\theta)^n (1-\theta t_1)^{-n} \qquad (1.2.2)$$

under the transformation

$$t_1 = u(1-\theta)^{\beta-1} (1-\theta t_1)^{-\beta+1}, \qquad (1.2.3)$$

where $n > 0$ and $\beta \ge 1$, or by

$$h(u) = f(t) = (1-\theta+\theta t)^n \qquad (1.2.4)$$

under the transformation

$$t = u(1-\theta+\theta t)^{\beta}, \qquad (1.2.5)$$

when n and β are positive integers.

Though, the two forms of the p.g.f. seem to be different from each other and makes one wonder as to what has happened to the uniqueness property of the probability generating functions, but really both forms are equivalent to each other. If we put

$$t_1 = t/(1-\theta+\theta t)$$

in the first form of the p.g.f., it reduces to the second form of the p.g.f.. Thus, the two forms are identical. Also, we note that the p.g.f. is a function of u and not of t_1 or t as it seems to be. In view of these remarks, any of the two forms may be used in future.

The moment generating function (m.g.f.) of any r.v. X is given by

$$m(T) = E\left[e^{TX}\right].$$

Accordingly, the m.g.f. of the GNB distribution (1.2.1) is given by replacing u either in (1.2.2) and (1.2.3) or in (1.2.4) and (1.2.5) by e^U. Sometimes t may also be replaced by e^T for convenience. Thus

$$m(U) = f(e^T) = (1-\theta+\theta e^T)^n$$

where

$$e^T = e^U(1-\theta+\theta e^T)^{\beta}.$$

Jain and Consul (1971) obtained the first four moments about the mean by using a recurrence relation. Dyczka (1978) showed that the GNBD belongs to the PSD class. He obtained the first few moments by using the recurrence relation among the PSD moments. Ali-Amidi (1978) obtained a recurrence relation among the central moments. His recurrence relation is similar to that of Shoukri (1980).

CONVOLUTION PROPERTY: Let X_1 and X_2 be two independent random variables. If X_1 and X_2 have GNB distributions with parameters (m_1, β, θ) and (m_2, β, θ) respectively, then the sum $X_1 + X_2 = X$ has a GNB distribution with parameters (m_1+m_2, β, θ), [Jain and Consul (1971)]. In general, if we have $X_1, X_2, \ldots, X_n$ GNB random variables with parameters (m_i, β, θ), $i = 1,2,\ldots,n$, then the sample sum

$$Y = X_1 + X_2 + \cdots + X_n$$

has a GNB distribution with parameters $\left[\sum_{i=1}^{n} m_i, \beta, \theta\right]$. If $m_i = m$ for all i, $i = 1,2,\ldots,n$, then the probability function of Y is given by

$$P(Y=y) = \frac{nm}{nm+\beta y} \binom{nm+\beta y}{y} \theta^y (1-\theta)^{nm+\beta y-y} \qquad (1.2.6)$$

RELATED DISTRIBUTIONS: The binomial distribution and the negative binomial distribution have both been mentioned earlier to be special cases of the GNBD, given by $\beta = 0$ and $\beta = 1$ respectively.

The lost games distribution defined by Kemp and Kemp (1968) and subsequently considered by Gupta and Singh (1982) is a special case of the GNBD and it is given by $\beta = 2$, $m = a$ together with a shift in the origin to 0.

The generalized factorial distribution (or the GNB beta distribution) with parameters k, β and λ is obtained from the GNB distribution with parameters ($m = \lambda-k+1$, β, θ) through compounding, Jain and Consul (1971).

The above authors also showed that if X_1 and X_2 are independent GNB variates with parameters (m_1, β, θ) and ($m_2 = m-m_1$, β, θ), then the conditional distribution of $X_1 = x$ given that $X_1 + X_2 = z$ is a generalized negative hypergeometric distribution (GNHD) with parameters m, m_1, z and β.

If $m \to 0$, the zero-truncated (decapitated) GNB distribution with parameters (m, β, θ) tends to the generalized logarithmic series distribution with parameters β and θ.

Let X be a GNB r.v. with parameters (m, β, θ), mean μ and variance σ^2, the standardized r.v.

$$Z = \frac{X-\mu}{\sigma}$$

approaches the normal form as $m \to \infty$. This result was obtained by Consul and Shenton (1973).

1.2.2 APPLICATIONS OF GNBD

The GNBD has many interesting applications in various fields of study. Since the binomial and the negative binomial distributions are two of its special cases, one can easily visualize that the GNBD will be applicable to those physical situations which are being modelled not only by the binomial and negative binomial distributions but also by many other distributions.

Univariate distributions associated with Lagrange's expansion have been considered by Consul and Shenton (1972) and Jain (1975) and the multivariate cases by Consul and Shenton (1973), Jain and Singh (1975) and Good (1975). It has been shown that these distributions arise as the distributions of the number of customers served in a busy period of a single server queueing system under different conditions. In particular, suppose m customers are waiting for service in a queue at a counter when the service is initially started. Suppose further that customers arrive at the service point in batches of size $\beta-1$, in accordance with a Poisson process with traffic intensity λ. The customers are served individually by a single server. The service times are assumed to be independent and identically distributed random variables and have exponential probability distributions with parameter μ. Service times are independent of the arrival times. If

$$\theta = \frac{\lambda}{\lambda+\mu}$$

is the probability of arrival of a batch, then the probability distribution of the number of customers served before the queue first vanishes conforms to the GNB distribution (1.2.1).

The GNBD has an important use in chemistry in the reaction called polymerization where the substance formed are generally classified into unbranched linear chains and the branched chains. A chemist is usually interested in finding the size and weight of the formed substance after polymerization has taken place. The molecular sizes and weights distributions can be suitably represented by the GNB distributions.

Let $X(t)$ denote the total number of infected anywhere in a habitat, starting from those initially infected at t and up to the time of extinction of an epidemic. Kumar (1981) showed that the Lagrangian probability distributions are useful in the theory of epidemics and that the distribution of the r.v. $X(t)$ belongs to the family of the GNB distributions.

Good (1960, 1965) has shown that the class of LPD which contains the family of the GNB distributions and also the distributions of the sizes of trees in a branching process are likely to be important in the analysis of biological data and other areas where a branching mechanism is involved. In particular, the size distribution of the whole tree including the original individual is that of the GNB distribution. This multiplicative process has various applications especially in the study of population growth, the spread of rumours and the nuclear chain reactions.

1.3 THE GENERALIZED POISSON DISTRIBUTION

Let X be a GNB r.v. with parameters (m, β, p). If $p \to 0$, $m \to \infty$ and $\beta \to \infty$ such that $mp = \theta$ and $\beta p = \lambda$, Consul and Jain (1973) showed that the distribution of X tends to a generalized Poisson distribution (GPD) with parameters θ and λ. The probability function of a GP random variate is given by

$$P(X=x)=P_x(\theta,\lambda) = \begin{cases} \theta(\theta+\lambda x)^{x-1} e^{-\theta-\lambda x}/x!; & x = 0,1,2,\ldots \\ 0 & ; \text{ for } x > k \text{ when } \lambda < 0 \end{cases} \tag{1.3.1}$$

and zero otherwise where $\theta > 0$, $-\theta/4 \le \lambda \le 1$ and k is that largest positive integer for which $\theta + \lambda k > 0$ when λ is negative. The variance of the above generalized Poisson distribution is greater than, equal to or less than the mean according as the second parameter λ is positive, zero or negative. Both the mean and the variance tend to increase or decrease in value as λ increases or decreases. Moments of all order exist if $\lambda < 1$.

Recently, Consul (1986) has shown that the GPD is generated by two different physical models. He gave a number of axioms for a steady state point process which produces the generalized Poisson process.

By using the parametric transformation $\lambda = \varphi\theta$, the distribution (1.3.1) reduces to the restricted model

$$P(X=x) = P_x(\theta,\varphi\theta) = \begin{cases} (1+\varphi x)^{x-1} \theta^x e^{-\theta(1+\varphi x)}/x!; & x = 0,1,2,\ldots \\ 0 & ; \text{ for } x > k \text{ when } \varphi < 0 \end{cases} \tag{1.3.2}$$

and zero otherwise where $\theta > 0$, $-\frac{1}{4} \le \varphi < \theta^{-1}$ and k is defined in

(1.3.1). The case $\lambda = 0$ in (1.3.1) or $\gamma = 0$ in (1.3.2) corresponds to the Poisson distribution.

1.3.1 PROPERTIES OF GPD

Consul and Shenton (1972, 1973) obtained a number of properties of the GPD and have shown that it belongs to the class of Lagrangian probability distributions.

GENERATING FUNCTIONS: The p.g.f. of the distribution (1.3.1) is given by

$$h(u) = f(t) = e^{\theta(t-1)} \tag{1.3.3}$$

where

$$t = u\, e^{\lambda(t-1)}. \tag{1.3.4}$$

The m.g.f. is obtained from above by replacing u by e^U. Thus, we have the m.g.f. as

$$m(U) = f(t) = e^{\theta(t-1)}$$

where

$$t = e^U\, e^{\lambda(t-1)}.$$

CONVOLUTION PROPERTY: The GPD has been shown to possess the important convolution property. Thus, the sum of two independent GP random variates with parameters (θ_1, λ) and (θ_2, λ) is itself a GP random variate with parameters $(\theta_1 + \theta_2, \lambda)$. In general, the sum

$$Y = X_1 + X_2 + \cdots + X_n$$

of n independent GP random variates with parameters (θ_i, λ), $i = 1,2,\ldots,n$ is also a GP variate with parameters $\left[\sum^n \theta_i, \lambda\right]$. In particular, if all the θ_i's $i = 1,2,\ldots,n$ are equal to θ we obtain a GP variate with parameters $(n\theta, \lambda)$ and its probability function is given by

$$P(Y=y) = n\theta(n\theta+\lambda y)^{y-1} e^{-n\theta-\lambda y}/y! \qquad (1.3.5)$$

RELATED DISTRIBUTIONS: When $\lambda = 0$ in (1.3.1) or $\varphi = 0$ in (1.3.2), these distributions reduce to the Poisson distribution. By putting $\varphi = \lambda^{-1}$, $\theta = \lambda\phi$ in (1.3.2), we obtain the Borel-Tanner distribution.

Let X be a GP variate with mean μ and variance σ^2, Consul and Shenton (1973) showed that the standardized variate

$$Z = \frac{X-\mu}{\sigma}$$

tends to a standard normal form as θ increases without limit and λ takes a specified value in the interval (0.0, 0.5). These authors also showed that if $\theta \to \infty$, $\lambda \to 1$ such that $\theta(1-\lambda) = C^2\lambda$, where C is a constant, then the standardized random variable Z tends to inverse Gaussian distribution with parameter C.

Consul (1986) showed that the limiting distribution of a quasi-binomial distribution based on urn models is the GPD.

1.3.2 APPLICATIONS OF GPD

The GPD models have been used to describe natural phenomena where the parameter of the Poisson distribution is taken to be a linear function of the number of occurrences. Thus, the conditions of the experiment do not remain constant in time and the number of events in any interval is a function of the number of events which have already taken place.

The GPD is used to model the number of customers served in the busy periods of some queueing systems. If the initial number k of customers is a Poisson r.v. with mean θ per unit service interval and the subsequent arrivals are also Poissonian with mean λ per unit service interval, then the probability distribution of the number of customers served in the first busy period [Consul and Shenton, 1973] is given by the GPD in (1.3.1).

The GPD model is useful in the theory of branching process. As a member of the LPD class, it is the distribution of the total progeny in a Galton-Watson branching process.

Janardan and Schaeffer (1977) have used the GPD models for the analysis of chromosomal aberrations in human leukocytes. They assumed that the number of aberrations per cell follows a Poisson distribution with mean rate θ and the number of aberrations undergoing healing is random and also Poissonian with parameter λ. According to them, some of the induced aberrations may not be healed immediately and may form a queue of aberrations awaiting restitution.

Therefore, the frequencies of induced aberrations will be modified by restitution and the probability distribution of the number of aberrations awaiting restitution is given by the GPD in (1.3.1). Their statements were strongly supported by the results of fitting the GPD model (1.3.1) to about 90 different sets of experimental data. They discovered that the fits by the GPD model were better than any other known distributions considered in the literature. Also, they were able to give reasonable interpretations to the two parameters θ and λ.

1.4 THE GENERALIZED LOGARITHMIC SERIES DISTRIBUTION

Jain and Gupta (1973) considered a generalization of the logarithmic series distribution through Lagrange's expansion and obtained the probability function of the generalized logarithmic series distribution (GLSD) as

$$P(X=x) = \frac{1}{\beta x} \binom{\beta x}{x} \theta^{x} (1-\theta)^{\beta x-x}/\{-\ln(1-\theta)\} \qquad (1.4.1)$$

for $x = 1,2,3,\ldots$ and zero otherwise where $\beta \geq 1$ and $0 < \theta < \beta^{-1}$. The logarithmic series distribution is a special case of the GLSD and it is obtained when $\beta = 1$ in (1.4.1).

As a matter of fact, little has been done on this family of discrete distributions. We now give some properties and applications of the GLSD model.

The p.g.f. of the distribution (1.4.1) is given by

$$h(u) = f(t) = \ell n(1-\theta t)/\ell n(1-\theta) \tag{1.4.2}$$

where

$$t = u(1-\theta)^{\beta-1}\ (1-\theta t)^{-\beta+1}. \tag{1.4.3}$$

From (1.4.2) and (1.4.3), we obtain the m.g.f. as

$$f(e^T) = \ell n(1-\theta e^T)/\ell n(1-\theta)$$

where

$$e^T = e^U(1-\theta)^{\beta-1}\ (1-\theta e^T)^{-\beta+1}.$$

Gupta (1976) showed that if each X_i, $i = 1,2,\ldots,n$ is a GLS variate with parameters (θ,β), then the distribution of the sample sum

$$Y = X_1 + X_2 + \cdots + X_n$$

is given by

$$P(Y=y) = \frac{n!}{y} \sum_{k=n-1}^{y} (-1)^{y-k-1} \frac{|S_k^{n-1}|}{k!} \binom{-\beta y+y-1}{y-k-1} \frac{\theta^y(1-\theta)^{\beta y-y}}{\{-\ell n(1-\theta)\}^n} \tag{1.4.4}$$

where S_k^n is defined as Stirling number of the first kind and it is given by

$$S_k^n = \left[\frac{1}{n!} D^n\ (x)_k\right]_{x=0}$$

$$= \frac{1}{n!}\left[\frac{d^n}{dx^n}\left\{\prod_{i=1}^{k} (x-i+1)\right\}\right]_{x=0}. \tag{1.4.5}$$

Among the related distributions of the GLSD is the GNBD which is mentioned in section 1.2. Under certain conditions, the GLSD is generated by the zero-truncated GNBD.

CHAPTER II

UNIMODALITY OF GENERALIZED DISCRETE DISTRIBUTIONS

2.1 INTRODUCTION

The property of unimodality plays an important role in statistical estimation. The problem of density estimation has been considered by many authors. Notably among them are Robertson (1967), Prakasa Rao (1969) and Wegman (1970a, '70b, '72). It has been shown that the method of maximum likelihood estimation can be used to estimate a unimodal density. This maximum likelihood estimator has been shown to be consistent. Although these results were proved for the field of continuous distributions, the maximum likelihood estimation method can also be used to estimate a discrete probability function.

Cryer and Robertson (1975) considered the isotonized estimate of the probability of extinction of a branching process. They obtained the offspring distribution estimate by assuming that the distribution was unimodal. This estimate was also shown to be consistent.

Barndorff-Nielsen (1976) applied the property of unimodality to the theory of plausibility inference.

The property of discrete unimodality can also be of interest in connection with optimization. This property is very important for many decomposition problems of probabilistic and statistical nature as indicated in the well known book by Medgyessy (1977).

A discrete probability distribution $\{P_x\}$ is said to be unimodal if there exists at least an integer M such that

$$P_x \geq P_{x-1} \qquad \text{for all} \qquad x \leq M \tag{2.1.1}$$

and

$$P_{x+1} \leq P_x \qquad \text{for all} \qquad x \geq M. \tag{2.1.2}$$

Keilson and Gerber (1971) have defined the strong unimodality of discrete probability distributions, have proved a number of results on the strong unimodality of discrete distributions and have shown that the binomial distribution, the negative binomial distribution and the Poisson distribution are all strongly unimodal. A necessary and sufficient condition that the sequence $\{P_x\}$ be strongly unimodal is

$$P_x^2 \,/\, \left[P_{x-1}\, P_{x+1}\right] \geq 1 \tag{2.1.3}$$

for all values of x.

The GLSD model given by (1.4.1), the GNBD model given by (1.2.1) and the GPD model given by (1.3.1) do not satisfy the property (2.1.3). For the GLSD, we have

$$\frac{P_2^2}{P_1\, P_3} = \frac{3}{2}\, \frac{\Gamma(2\beta)\ \Gamma(2\beta)\ \Gamma(\beta)\ \Gamma(3\beta-2)}{\Gamma(2\beta-1)\ \Gamma(2\beta-1)\ \Gamma(\beta)\ \Gamma(3\beta)}$$

$$= \frac{3}{2}\, \frac{(2\beta-1)(2\beta-1)}{(3\beta-1)(3\beta-2)}$$

which can be easily shown to be a decreasing function of β.

Hence

$$\frac{P_2^2}{P_1\, P_3} < \lim_{\beta\to 1} \frac{3}{2}\, \frac{(2\beta-1)^2}{(3\beta-1)(3\beta-2)}$$

$$= \frac{3}{4}$$

$$< 1.$$

For the GNBD, if $0 < m < \beta$, we have

$$\frac{P_1^2}{P_0\, P_2} = \frac{2m}{m+2\beta-1}$$

$$< 1.$$

For the case of the GPD, it can be easily seen that if $\theta < 2\lambda$,

$$\frac{P_1^2}{P_0\, P_2} = \frac{2\theta^2}{\theta(\theta+2\lambda)}$$

$$= \frac{2\theta}{\theta+2\lambda}$$

$$< 1$$

and accordingly the condition (2.1.3) for strong unimodality does not hold even for $x = 2$ in the case of GLSD and $x = 1$ for the GNBD and the GPD.

In section (2.3), we shall prove a theorem on the unimodality of the GLSD.

In sections (2.4) and (2.5), we shall show that the GNBD and the GPD belong to the class of discrete self-decomposable distributions studied by Steutel and van Harn (1979) and then prove that they are unimodal.

2.2 SOME PRELIMINARY RESULTS

We shall state two important results here for ready reference whenever necessary.

Lagrange's expansion: Under mild conditions of successive differentiability of the functions $\varphi(t)$ and $f(t)$, and when $\varphi(0) \neq 0$, Whittaker and Watson (1927) have given the Lagrange's expansion [see page 133] as

$$f(t) = f(0) + \sum_{k=1}^{\infty} \frac{z^k}{k!} \left\{ \left[\frac{d}{dx}\right]^{k-1} \left[\varphi^k(x) f'(x)\right] \right\}_{x=0} \qquad (2.2.1)$$

where t and z are related by

$$t = z\,\varphi(t) .$$

Steutel and van Harn's results.

Result 1: A p.g.f. is discrete self-decomposable iff it has the form

$$f(t) = \exp\left\{-\theta \int_t^1 \frac{1-G(u)}{1-u}\, du\right\} \tag{2.2.2}$$

where $\theta > 0$ and G is a p.g.f. with $G(0) = 0$. Equivalently, $f(t)$ is discrete self-decomposable iff

$$f(t) = \exp\left\{-\int_t^1 R(u)\, du\right\} \tag{2.2.3}$$

where $R(u) = \sum_{i=1}^{\infty} r_i u^i$ with $r_i \geq 0$ and r_i is non-increasing.

Also, $\sum_{i=1}^{\infty} r_i (i+1)^{-1} < \infty$.

Result 2: Let $\{P_x\}_0^{\infty}$ be a probability distribution on the non-negative integers with p.g.f. $G(z)$ satisfying,

$$\frac{d}{dz} \log G(z) = R(z) = \sum_{k=0}^{\infty} r_k z^k \tag{2.2.4}$$

where the r_k, $k = 0,1,2,\ldots$ are all non-negative. Then $\{P_x\}_0^{\infty}$ is unimodal if $\{r_k\}_0^\infty$ is non-increasing, and $\{P_x\}_0^\infty$ is non-increasing if and only if in addition $r_0 \le 1$.

2.3 UNIMODALITY OF GLSD

THEOREM: The GLSD defined in (1.4.1), is unimodal for all values of θ in $0 < \theta < \beta^{-1}$ and of $\beta \ge 1$ and the mode is at the point $x = 1$.

PROOF: Since the unimodality of the logarithmic series distribution (β=1) is well established [Johnson and Kotz, 1969], we shall consider the unimodality of GLSD for $\beta > 1$. Let the mode be at the point $x = M$. For the mode of the GLSD to be at point $M = 1$, we must show that

$$P_{x+1} < P_x \qquad \text{for all} \qquad x = 1,2,3,\ldots.$$

Now,

$$\frac{P_{x+1}}{P_x} = \frac{\beta x}{\beta(x+1)} \frac{\Gamma(\beta x+\beta+1)x!\ \Gamma(\beta x-x+1)\ \theta(1-\theta)^{\beta-1}}{(x+1)!\ \Gamma[(\beta-1)(x+1)+1]\ \Gamma(\beta x+1)}$$

$$= \frac{x}{(x+1)^2} \frac{\Gamma(\beta x+\beta+1)\ \Gamma(\beta x-x+1)\ \theta(1-\theta)^{\beta-1}}{\Gamma[(\beta-1)(x+1)+1]\ \Gamma(\beta x+1)}. \tag{2.3.1}$$

Since $\theta(1-\theta)^{\beta-1}$ is an increasing function of θ and $0 < \theta < \beta^{-1}$, we have

$$\frac{P_{x+1}}{P_x} < \frac{x}{(x+1)^2} \frac{\Gamma(\beta x+\beta+1)\ \Gamma(\beta x-x+1)}{\Gamma[(\beta-1)(x+1)+1]\ \Gamma(\beta x+1)} \cdot \frac{1}{\beta}\left[1-\frac{1}{\beta}\right]^{\beta-1}. \tag{2.3.2}$$

When $x = 1$,

$$\frac{P_2}{P_1} < \frac{1}{4} \frac{\Gamma(2\beta+1)\ \Gamma(\beta)}{\Gamma(2\beta-1)\ \Gamma(\beta+1)} \cdot \frac{1}{\beta} \left[1 - \frac{1}{\beta}\right]^{\beta-1}$$

$$= \frac{1}{4} \cdot \frac{2\beta\ (2\beta-1)}{\beta} \cdot \frac{1}{\beta} \left[1 - \frac{1}{\beta}\right]^{\beta-1}$$

$$= \frac{2\beta-1}{2\beta} \left[1 - \frac{1}{\beta}\right]^{\beta-1}$$

$$< 1 \qquad\qquad \text{for all} \quad \beta \geq 1.$$

From the above, we observed that $P_{x+1}/P_x < 1$ for $x = 1$. We now consider the ratio in general.

From (2.3.2), we obtain

$$\frac{P_{x+1}}{P_x} < \frac{x}{(x+1)^2} \cdot \frac{1}{\beta} \cdot \left[\frac{\beta-1}{\beta}\right]^{\beta-1} \frac{(\beta x+\beta)!\ (\beta x-x)!}{(\beta x)!\ (\beta x+\beta-x-1)!}$$

$$= \frac{x}{(x+1)^2} \frac{1}{\beta} \cdot \left[\frac{\beta-1}{\beta}\right]^{\beta-1} \frac{(\beta x+\beta)(\beta x+\beta-1)(\beta x+\beta-2)\cdots(\beta x+1)}{(\beta x-x+\beta-1)(\beta x-x+\beta-2)\cdots(\beta x-x+1)}$$

$$= \frac{x}{x+1} \left[\frac{\beta-1}{\beta}\right]^{\beta-1} \prod_{i=1}^{\beta-1} \left[\frac{\beta x+\beta-i}{\beta x-x+\beta-i}\right]$$

$$= \frac{x}{x+1} \left[\frac{\beta-1}{\beta}\right]^{\beta-1} \prod_{i=1}^{\beta-1} \left[1 + \frac{x}{\beta x+\beta-x-i}\right]$$

$$< \frac{x}{x+1} \left[\frac{\beta-1}{\beta}\right]^{\beta-1} \left[1 + \frac{x}{\beta x+\beta-x-\beta+1}\right]^{\beta-1}$$

$$= \frac{x}{x+1}\left[\frac{\beta-1}{\beta}\right]^{\beta-1}\left[\frac{\beta x+1}{\beta x+1-x}\right]^{\beta-1}$$

$$= \frac{x}{x+1}\left\{\frac{(\beta-1)(\beta x+1)}{\beta(\beta x+1-x)}\right\}^{\beta-1}$$

$$= \frac{x}{x+1}\left\{\frac{\beta(\beta x+1)-\beta x-1}{\beta(\beta x+1)-\beta x}\right\}^{\beta-1}$$

$$< 1.$$

Therefore, the GLSD is unimodal with its mode at the point $M = 1$.

2.4 UNIMODALITY OF GNBD

THEOREM: The GNBD, defined in (1.2.1), is unimodal for all values of m, θ and β given in section 1.2.

PROOF: Since the unimodality of the binomial distribution and of the negative binomial distribution is well established [Keilson and Gerber, 1971], we shall consider the unimodality of GNBD for $\beta > 1$ only.

The p.g.f. of the GNBD is given by (1.2.2) and (1.2.3) as

$$G(z) = \sum_{x=0}^{\infty} P(X=x)\, z^x = (1-\theta)^m (1-\theta t)^{-m} \qquad (2.4.1)$$

where

$$t = z(1-\theta)^{\beta-1} (1-\theta t)^{-\beta+1}. \qquad (2.4.2)$$

Now,

$$G(z) = \exp\left\{\ln[(1-\theta)^m \cdot (1-\theta t)^{-m}]\right\}$$

$$= \exp\left\{m\ell n[(1-\theta) - m\ell n(1-\theta t)]\right\}$$

$$= \exp\left\{m[\ell n(1-\theta) - \ell n(1-\theta t)]\right\}$$

$$= \exp\left\{- \int_t^1 \frac{m\theta}{1-\theta u}\, du\right\}$$

which is equivalent to (2.2.3) with

$$R(u) = \frac{m\theta}{1-\theta u}$$

$$= \sum_{i=0}^{\infty} m\, \theta^{i+1}\, u^i.$$

Thus $r_i = m\, \theta^{i+1}$.

Clearly, $r_i \geq 0$ and also

$$\frac{r_{i-1}}{r_i} = \frac{m\theta^i}{m\theta^{i+1}} = \frac{1}{\theta} \geq 1, \quad \text{since} \quad 0 < \theta < 1.$$

That is, $r_{i-1} \geq r_i$ for $i = 1,2,3,\ldots$, therefore the r_i's are non-increasing and by using Result 1, the GNBD belongs to a self-decomposable class.

From (2.2.4) and (2.4.1), we obtain

$$\frac{d}{dz} \log G(z) = \frac{m\theta}{1-\theta t} \frac{dt}{dz}. \tag{2.4.3}$$

In order to obtain (2.4.3), we use the Lagrange's expansion on

$$\log G(z) = m \log(1-\theta) - m \log(1-\theta t)$$

by taking $f(t) = \log G(z)$ and

$$\varphi(t) = (1-\theta)^{\beta-1} (1-\theta t)^{-\beta+1}.$$

Thus, we get

$$\log G(z) = m\log(1-\theta)-m\left[0+\sum_{k=1}^{\infty} \frac{z^k}{k!}\left\{\left[\frac{d}{dt}\right]^{k-1}(1-\theta)^{(\beta-1)k}(1-\theta t)^{-\beta k+k}\frac{(-\theta)}{1-\theta t}\right\}_{t=0}\right]$$

$$= m\log(1-\theta)+m\theta \sum_{k=1}^{\infty} \frac{z^k}{k!} (1-\theta)^{(\beta-1)k} \left\{\left[\frac{d}{dt}\right]^{k-1} (1-\theta t)^{-\beta k+k-1}\right\}_{t=0}$$

$$= m\log(1-\theta)+m\theta \sum_{k=1}^{\infty} \frac{z^k}{k!} (1-\theta)^{(\beta-1)k} (\beta k-k+1)(\beta k-k+2)\cdots(\beta k-k+k-1)\ \theta^{k-1}$$

$$= m\log(1-\theta)+m \sum_{k=1}^{\infty} \frac{z^k}{k!} \theta^k (1-\theta)^{\beta k-k} (\beta k-1)(\beta k-2)\cdots(\beta k-k+1)$$

$$= m\log(1-\theta)+m \sum_{k=1}^{\infty} \frac{z^k}{k!} \theta^k (1-\theta)^{\beta k-k} \frac{(\beta k-1)!}{(\beta k-k)!}. \tag{2.4.4}$$

On differentiating (2.4.4) with respect to z, we have

$$\frac{d}{dz} \log G(z) = m \sum_{k=1}^{\infty} \frac{z^{k-1}}{(k-1)!} \theta^k (1-\theta)^{\beta k-k} \frac{(\beta k-1)!}{(\beta k-k)!}$$

$$= m \sum_{k=0}^{\infty} \frac{z^k}{k!} \theta^{k+1} (1-\theta)^{(\beta-1)(k+1)} \frac{(\beta k+\beta-1)!}{[(\beta-1)(k+1)]!}$$

$$= m \sum_{k=0}^{\infty} z^k \theta^{k+1} (1-\theta)^{(\beta-1)(k+1)} \begin{bmatrix} \beta k+\beta-1 \\ k \end{bmatrix}$$

from which one obtains r_k, the coefficient of z^k in (2.2.4) as

$$r_k = m\, \theta^{k+1}(1-\theta)^{(\beta-1)(k+1)} \begin{bmatrix} \beta k+\beta-1 \\ k \end{bmatrix}.$$

Clearly r_k, $k = 0,1,2,\ldots$ are non-negative. We now need to verify that the r_k's are non-increasing.

$$\frac{r_k}{r_{k-1}} = \frac{\theta(1-\theta)^{\beta-1}\, \Gamma(\beta k+\beta)\, \Gamma(\beta k-k+1)}{k\, \Gamma(\beta k)\, \Gamma(\beta k+\beta-k)}. \qquad (2.4.5)$$

When $k = 1$, $\dfrac{r_1}{r_0} = \theta(1-\theta)^{\beta-1}(2\beta-1)$

$$< \left[1 - \frac{1}{\beta}\right]^{\beta-1} \left[2 - \frac{1}{\beta}\right]$$

as $0 < \theta < \beta^{-1}$ and as the maximum of $\theta(1-\theta)^{\beta-1}$ occurs at $\theta = \beta^{-1}$.

By logarithmic differentiation, one can show that the right hand side of the above is a decreasing function of β and it is less than 1 for all values of $\beta > 1$.

In general, we consider the ratio given by (2.4.5).

Now,

$$\frac{r_k}{r_{k-1}} < \frac{1}{\beta k}\left[\frac{\beta-1}{\beta}\right]^{\beta-1} \frac{(\beta k+\beta-1)!\ (\beta k-k)!}{(\beta k-1)!\ (\beta k+\beta-k-1)!}$$

$$= \frac{1}{\beta k}\left[\frac{\beta-1}{\beta}\right]^{\beta-1} \frac{(\beta k+\beta-1)(\beta k+\beta-2)\cdots(\beta k+1)(\beta k)}{(\beta k+\beta-k-1)(\beta k+\beta-k-2)\cdots(\beta k-k+1)}$$

$$= \left[\frac{\beta-1}{\beta}\right]^{\beta-1} \prod_{i=1}^{\beta-1}\left[\frac{\beta k+\beta-i}{\beta k+\beta-k-i}\right]$$

$$= \left[\frac{\beta-1}{\beta}\right]^{\beta-1} \prod_{i=1}^{\beta-1}\left[1+\frac{k}{\beta k+\beta-k-i}\right]$$

$$< \left[\frac{\beta-1}{\beta}\right]^{\beta-1} \left[1+\frac{k}{\beta k+\beta-k-\beta+1}\right]^{\beta-1}$$

$$= \left[\frac{\beta-1}{\beta}\right]^{\beta-1} \left[\frac{\beta k+1}{\beta k+1-k}\right]^{\beta-1}$$

$$= \left[\frac{\beta(\beta k+1)-\beta k-1}{\beta(\beta k+1)-\beta k}\right]^{\beta-1}$$

$$< 1.$$

So $\dfrac{r_k}{r_{k-1}} < 1$ for all $k = 1,2,3,\ldots$

and it therefore follows that the sequence $\{r_k\}_0^\infty$ is non-increasing.

Thus, the GNBD is unimodal.

Now, the value of r_0 is found to be

$$r_0 = m\theta(1-\theta)^{\beta-1}.$$

If $r_0 = m\theta(1-\theta)^{\beta-1} < 1$, the GNB distribution given by (1.2.1) is non-increasing and so the mode is at the point $x = 0$. If $r_0 = m\theta(1-\theta)^{\beta-1} = 1$, the mode is at the dual points $x = 0$ and $x = 1$ as both have the same probability mass.

2.4.1 BOUNDS FOR THE MODE OF GNBD

For $m\theta(1-\theta)^{\beta-1} = m\varphi > 1$, the mode is at some point $x = M$ such that

$$\frac{m\varphi-1}{1-(2\beta-1)\varphi} < M < u \tag{2.4.6}$$

where u is the value of M given by the inequality

$$M^2.\beta(\beta-1)\varphi + M[(2\beta m-m+1)\varphi - (m+2\beta-1)] + (m^2-1)\varphi > 0. \tag{2.4.7}$$

The ratio of any two consecutive probabilities of the GNBD is given by

$$\frac{P_{x+1}}{P_x} = \frac{(m+\beta x+\beta-1)\cdots(m+\beta x+\beta-x+1)(m+\beta x+\beta-x)\varphi}{(m+\beta x-1)\cdots(m+\beta x-x+1)(x+1)}. \tag{2.4.8}$$

Since $x = M$ is the mode, by using (2.1.1) and (2.1.2), the above relation gives

$$\frac{P_{M+1}}{P_M} = \frac{(m+\beta M+\beta-1)\cdots(m+\beta M+\beta-M+1)(m+\beta M+\beta-M)\varphi}{(m+\beta M-1)\cdots(m+\beta M-M+1)(M+1)} < 1 \quad (2.4.9)$$

and

$$\frac{P_M}{P_{M-1}} = \frac{(m+\beta M-1)\cdots(m+\beta M-M+2)(m+\beta M-M+1)\varphi}{(m+\beta M-\beta-1)\cdots(m+\beta M-\beta-M+2)\ M} > 1. \quad (2.4.10)$$

From the inequality (2.4.9),

$$M + 1 > (m+\beta M+\beta-1) \prod_{i=1}^{M-1} \left[\frac{m+\beta M+\beta-1-i}{m+\beta M-i}\right] \varphi$$

$$> (m+\beta M+\beta-1) \prod_{i=1}^{M-1} \left[\frac{m+\beta M+(\beta-1)(i+1)}{m+\beta M+(\beta-1)i}\right] \varphi$$

$$= [m+\beta M+(\beta-1)M]\ \varphi$$

which gives

$$M > \frac{m\varphi-1}{1-(2\beta-1)\varphi}.$$

From the inequality (2.4.10),

$$1 < \frac{m+\beta M-1}{M} \cdot \frac{(m+\beta M-2)\cdots(m+\beta M-M+2)(m+\beta M-M+1)}{(m+\beta M-\beta-1)\cdots(m+\beta M-\beta-M+2)} \varphi$$

which can be expressed as

$$\frac{(m+\beta M-1)\ \varphi}{M} > \frac{(m+\beta M-\beta-1)(m+\beta M-\beta-2)\cdots(m+\beta M-\beta-M+2)}{(m+\beta M-2)(m+\beta M-3)\cdots(m+\beta M-M+1)}$$

$$= \prod_{i=2}^{M-1} \left[\frac{m+\beta M-i-\beta+1}{m+\beta M-i}\right]$$

$$= \prod_{i=2}^{M-1} \left[1 - \frac{\beta-1}{m+\beta M-i}\right]$$

$$> \left[1 - \frac{\beta-1}{m+\beta M-M+1}\right]^{M-2}$$

$$> 1 - \frac{(\beta-1)(M-2)}{m+\beta M-M+1}$$

$$= \frac{m+2\beta-1}{m+\beta M-M+1}.$$

Hence, we have

$$(m+\beta M-1)\varphi > \frac{M(m+2\beta-1)}{m+(\beta-1)M+1}$$

i.e. $$(m+\beta M-1)[m+(\beta-1)M+1]\varphi > (m+2\beta-1)M. \qquad (2.4.11)$$

The value of M satisfying this inequality will be an upper bound to the mode of the distribution (1.2.1). On simplifying the inequality (2.4.11), we obtain the result in (2.4.7).

If $P_{M+1}/P_M = 1$ and $P_{x+1}/P_x < 1$ for all $x > M + 1$, the GNB distribution has its mode at the two consecutive points $x = M$ and $x = M+1$.

Particular cases (i) When $\beta = 0$, the bounds of the mode is given by

$$\theta(m+1) - 1 < M < \theta(m+1)$$

which is the result for the binomial distribution given by Johnson and

Kotz (1969, page 53).

(ii) When $\beta = 1$, the bound in (2.4.6) yields

$$\frac{m\theta-1}{1-\theta} < M < \frac{(m-1)\theta}{1-\theta}$$

which is the result for the negative binomial distribution.

2.5 UNIMODALITY OF GPD

THEOREM: The GPD is unimodal for all values of $\theta > 0$ and of λ in $0 \le \lambda < 1$.

PROOF: The ratio of any two successive probabilities of the GPD is

$$\frac{P_{x+1}}{P_x} = \frac{(\theta+\lambda x+\lambda)^x}{(\theta+\lambda x)^{x-1}} \cdot \frac{e^{-\lambda}}{x+1}. \tag{2.5.1}$$

It may be noted that $P_1/P_0 = \theta\ e^{-\lambda}$ which can be $\le$ or ≥ 1 depending upon the value of θ, as $0 \le \lambda < 1$ and $e^{-\lambda} \le 1$. The p.g.f. of the GPD is given by (1.3.3) and (1.3.4) as

$$G(z) = \sum_{x=0}^{\infty} P(X=x)\ z^x = e^{\theta(t-1)}. \tag{2.5.2}$$

where

$$t = z\ e^{\lambda(t-1)}. \tag{2.5.3}$$

By using Lagrange's expansion on (2.5.3) with $f(t) = t$ and $\varphi(t) = e^{\lambda(t-1)}$, we obtain

$$t = z\, e^{\lambda(t-1)}$$

$$= f(0) + \sum_{k=1}^{\infty} \frac{z^k}{k!} \left\{ \left[\frac{d}{dt}\right]^{k-1} \left[e^{\lambda k(t-1)}\right] \right\}_{t=0}$$

$$= \sum_{k=1}^{\infty} e^{-\lambda k} \frac{(\lambda k)^{k-1}}{k!} z^k = Q(z). \tag{2.5.4}$$

Now,

$$G(z) = e^{\theta(t-1)}$$

$$= \exp\{\theta(t-1)\}$$

$$= \exp\left\{ - \theta \int_t^1 \frac{1-G(u)}{1-u}\, du \right\}$$

where $G(u) = u$.

The above is equivalent to (2.2.2) and therefore, the GPD is a discrete self-decomposable distribution and Result 2 can be applied to show that the distribution is unimodal.

From (2.2.4) and (2.5.2), we have that

$$\frac{d}{dz} \log G(z) = \theta \frac{d}{dz} Q(z)$$

$$= \sum_{k=0}^{\infty} \theta\, e^{-\lambda(k+1)} \frac{[\lambda(k+1)]^k}{k!} z^k$$

where r_k, the coefficient of z^k in (2.2.4) is given by

$$r_k = \theta\ e^{-\lambda(k+1)}\ \frac{[\lambda(k+1)]^k}{k!}.$$

We now show that the r_k's are non-increasing. Obviously, r_k, $k = 0,1,2,\ldots$ are non-negative. Now,

$$\frac{r_k}{r_{k-1}} = \frac{e^{-\lambda(k+1)}[\lambda(k+1)]^k/k!}{e^{-\lambda k}\ [\lambda k]^{k-1}/(k-1)!}$$

$$= \lambda\ e^{-\lambda}\ \left[\frac{k+1}{k}\right]^k \tag{2.5.5}$$

$$= \lambda\ e^{-\lambda}\left[1 + \frac{1}{k}\right]^k$$

$$\leq \lambda\ e^{-\lambda+1}$$

$$\leq 1$$

since $\left[1 + \frac{1}{k}\right]^k \leq e^1$ and $1 + x < e^x$ for all $x \in \mathbb{R}$.

Hence $\left\{r_k\right\}_0^\infty$ is non-increasing and so the GPD is unimodal for all values of θ and λ by using result 2.

When $r_0 = \theta\ e^{-\lambda} \leq 1$, $\left\{P_x\right\}_0^\infty$ becomes non-increasing.

Accordingly, the mode will be at $x = 0$ if $\theta\ e^{-\lambda} < 1$ and at the dual points $x = 0$ and $x = 1$ if $\theta\ e^{-\lambda} = 1$.

2.5.1 BOUNDS FOR THE MODE OF GPD

For $\theta\ e^{-\lambda} > 1$, the mode of the GPD is at some point $x = M$ such that

$$(\theta - e^{\lambda})(e^{\lambda} - 2\lambda)^{-1} < M < u \qquad (2.5.6)$$

where u is the value of M given by the inequality

$$\lambda^2 M^2 + M\left[2\lambda\theta - (\theta + 2\lambda)e^{\lambda}\right] + \theta^2 > 0. \qquad (2.5.7)$$

Since M is the mode, we use (2.1.1), (2.1.2) and the relation (2.5.1) to obtain

$$\frac{P_{M+1}}{P_M} = \frac{(\theta + M\lambda + \lambda)^M\ e^{-\lambda}}{(M+1)(\theta + M\lambda)^{M-1}} < 1 \qquad (2.5.8)$$

and

$$\frac{P_M}{P_{M-1}} = \frac{(\theta + M\lambda)^{M-1}\ e^{-\lambda}}{M(\theta + M\lambda - \lambda)^{M-2}} > 1. \qquad (2.5.9)$$

From the inequality (2.5.8), we have

$$(M+1)(\theta + M\lambda)^{M-1} > e^{-\lambda}\ (\theta + M\lambda + \lambda)^M$$

i.e. $$(M+1)\ e^{\lambda} > (\theta + M\lambda)\ \left[\frac{\theta + M\lambda + \lambda}{\theta + M\lambda}\right]^M$$

$$= (\theta + M\lambda)\ \left[1 + \frac{\lambda}{\theta + M\lambda}\right]^M$$

$$> (\theta + M\lambda)\ \left[1 + \frac{M\lambda}{\theta + M\lambda}\right]$$

$$= \theta + 2M\lambda.$$

The above inequality gives

$$M > (\theta - e^{\lambda})(e^{\lambda} - 2\lambda)^{-1}. \qquad (2.5.10)$$

By using the inequality (2.5.9), we get

$$\frac{(\theta+M\lambda)e^{-\lambda}}{M} > \left[1 - \frac{\lambda}{\theta+M\lambda}\right]^{M-2}$$

$$> 1 - \frac{(M-2)\lambda}{\theta+M\lambda}$$

$$= \frac{\theta+2\lambda}{\theta+M\lambda}.$$

Thus

$$(\theta+M\lambda)^2 > M(\theta+2\lambda)\ e^{\lambda}.$$

On simplifying this inequality, we obtain (2.5.7) and so the upper bound of the mode M is given by the inequality (2.5.7).

Particular Case: When $\lambda = 0$, the bounds for the mode are given by

$$\theta - 1 < M < \theta$$

which is the result for the poisson distribution.

CHAPTER III

INTERVAL ESTIMATION IN THE CLASS OF MODIFIED POWER SERIES DISTRIBUTIONS

3.1 INTRODUCTION

A discrete random variable X is said to have an MPSD if its probability function is given by (1.1.6). The mean μ and variance σ^2 of the distribution (1.1.6) are given by

$$\mu = \frac{g(\theta)f'(\theta)}{g'(\theta)f(\theta)}, \qquad \sigma^2 = \frac{g(\theta)}{g'(\theta)}\frac{d\mu}{d\theta}. \tag{3.1.1}$$

The problem of point estimation in the MPSD class has been considered by many researchers. Gupta (1975) considered the maximum likelihood (ML) estimation for the class of MPSD. Kumar and Consul (1979) obtained a recurrence relation for the negative moments and used these results to find the exact amount of bias and the variance of the ML estimators for some members of the MPSD class. Kumar and Consul (1980) also considered the minimum variance unbiased estimation for the MPSD class. The estimation of probabilities in this class was given by Gupta and Singh (1982).

The restricted GPD (1.3.2), the GNBD (1.2.1) and the GLSD (1.4.1) are three important families of the MPSD class. The particular values of $f(\theta)$, $g(\theta)$, $a(x)$ and means for the MPSD class which provide these three families are given in the following table.

TABLE 3.1

$f(\theta)$, $g(\theta)$, $a(x)$ and the mean of some modified power series distributions

Distribution	$f(\theta)$	$g(\theta)$	$a(x)$	μ
GPD	e^{θ}	$\theta\, e^{-\varphi\theta}$	$\frac{(1+x\varphi)^{x-1}}{x!}$	$\theta(1-\varphi\theta)^{-1}$
GNBD	$(1-\theta)^{-m}$	$\theta(1-\theta)^{\beta-1}$	$\frac{m}{m+\beta x}\binom{m+\beta x}{x}$	$m\theta(1-\theta\beta)^{-1}$
GLSD	$-\ln(1-\theta)$	$\theta(1-\theta)^{\beta-1}$	$\frac{1}{\beta x}\binom{\beta x}{x}$	$\frac{\theta(1-\theta\beta)^{-1}}{\{-\ln(1-\theta)\}}$

Let a random sample $X_1, X_2, \ldots, X_n$ of size n be taken from the MPSD given in (1.1.6). Its likelihood function is given by

$$L(\theta) = \frac{\{g(\theta)\}^{\sum_{i}^{n} x_i}}{\{f(\theta)\}^n} \prod_{i=1}^{n} a(x_i) \,. \tag{3.1.2}$$

The statistic $Y = \sum^{n} X_i$ is a complete and sufficient statistic for the parameter θ and its probability function is also an MPSD which is given by

$$P_y(\theta) = b(y)\, \frac{\{g(\theta)\}^y}{\{f(\theta)\}^n}, \qquad y \in R_n \tag{3.1.3}$$

where $R_n = \left\{y \mid y = \sum^{n} x_i,\ x_i \in T\right\}$, a subset of non-negative integers and where

$$b(y) = \Sigma\ a(x_1)\ a(x_2)\ \cdots\ a(x_n),$$

the summation extending over all order n-tuples $(x_1, x_2, \ldots, x_n)$ of integers $x_i \in T$ with the condition that $\sum^{n} x_i = y$ is satisfied.

3.2 CONFIDENCE INTERVALS FOR θ IN SMALL SAMPLES

Since the statistic Y is sufficient for the parameter θ in (1.1.6), one can equate the sum of the probabilities of the distribution of Y on each tail with $\frac{1}{2}\alpha$. Let θ_ℓ be the lower bound and θ_u be the upper bound of the $100(1-\alpha)\%$ confidence interval (C.I.) for θ. Thus, we get the equations

$$\sum_{x=y}^{\infty} b(x)\ \{g(\theta_\ell)\}^x\ \{f(\theta_\ell)\}^{-n} = \frac{1}{2}\alpha \qquad (3.2.1)$$

and

$$\sum_{x=0}^{y} b(x)\ \{g(\theta_u)\}^x\ \{f(\theta_u)\}^{-n} = \frac{1}{2}\alpha. \qquad (3.2.2)$$

One might feel that the above sums may not exactly equal $\frac{1}{2}\alpha$ as the sample sum Y is a discrete variate. However, for any given value of y, the expressions on the left hand side of (3.2.1) and (3.2.2) are continuous functions of θ and accordingly, there must exist some values of θ for which the equations hold exactly. In general, the values of θ_ℓ and θ_u will be difficult to get algebraically and one may have to use an algorithm on a computer to get them numerically.

Of course, the above results are very general for the MPSD class and one has to put specific values of $b(x)$, $f(\theta)$ and $g(\theta)$ to get the C.Is. We shall now apply the above results to the three important families of the MPSD class.

3.2.1 APPLICATIONS

(i) GENERALIZED POISSON DISTRIBUTION:

The probability distribution of Y for the model (1.3.2) is given by

$$P_y(\theta,\theta\varphi) = n(n+\varphi y)^{y-1}\,\theta^y\, e^{-\theta(n+\varphi y)}/y! \qquad y = 0,1,2,\ldots \tag{3.2.3}$$

By using the above distribution of Y, we obtain from (3.2.1) and (3.2.2) the following equations for finding the lower bound θ_ℓ and upper bound θ_u for the parameter θ in the GPD model (1.3.2).

$$\sum_{k=y}^{\infty} \frac{n(n+\varphi k)^{k-1}}{k!} \frac{\left\{\theta_\ell\, e^{-\theta_\ell \varphi}\right\}^k}{e^{n\theta_\ell}} = \frac{1}{2}\alpha \tag{3.2.4}$$

and

$$\sum_{k=0}^{y} \frac{n(n+\varphi k)^{k-1}}{k!} \frac{\left\{\theta_u\, e^{-\theta_u \varphi}\right\}^k}{e^{n\theta_u}} = \frac{1}{2}\alpha. \tag{3.2.5}$$

The above equations, for known values of α, y, n and φ, cannot be solved algebraically. A computer programme will have to be used to solve them numerically and to find the values of θ_ℓ and θ_u as solutions of (3.2.4) and (3.2.5) respectively. Thus, $100(1-\alpha)\%$ confidence interval for θ becomes (θ_ℓ, θ_u).

As a particular case when $\varphi = 0$ and a sample of size 1 is taken, we get the results in Johnson and Kotz (1969, page 96) for the Poisson distribution.

EXAMPLE 3.1

We consider the following data given by Lüders (1934) which contains the frequencies of the number of deaths in 60 rail and road fatal accidents. The GPD model fits this data very well. The expected GPD frequencies by ML method are also given.

TABLE 3.2

Number of deaths in fatal rail and road accidents
in Saargebiet

No. of deaths	0	1	2	3	4	5	≥ 6	Total
Observed frequency	20	17	11	8	2	0	2	60
Expected GPD freq.	19.40	18.22	11.35	5.94	2.84	1.28	0.97	60.00

$\bar{x} = 1.38333, \qquad s^2 = 2.07090$

The method in this subsection is applied to find C.I. for the parameter θ assuming that the quantity φ is known. In real life, it is hard to know the actual value of φ except in some very specific cases. Accordingly, to illustrate the method we shall use the ML estimate of φ as the known value of φ. The ML estimates of θ and φ in the GPD model (1.3.2) are 1.128826 and 0.162984 respectively.

Now, by putting $\varphi = 0.162984$ in equations (3.2.4) and (3.2.5) and by solving them for $\alpha = 0.05$ with the help of a computer, we obtain the 95% C.I. for θ as (0.90, 1.40).

(ii) **GENERALIZED NEGATIVE BINOMIAL DISTRIBUTION:**

The distribution of the sample sum Y is given by (1.2.6). By using the distribution of Y in (3.2.1) and (3.2.2), we obtain

$$\sum_{k=y}^{\infty} \frac{nm}{nm+\beta k} \binom{nm+\beta k}{k} \frac{\left\{\theta_\ell (1-\theta_\ell)^{\beta-1}\right\}^k}{(1-\theta_\ell)^{-nm}} = \frac{1}{2}\alpha \tag{3.2.6}$$

and

$$\sum_{k=0}^{y} \frac{nm}{nm+\beta k} \binom{nm+\beta k}{k} \frac{\left\{\theta_u (1-\theta_u)^{\beta-1}\right\}^k}{(1-\theta_u)^{-nm}} = \frac{1}{2}\alpha. \tag{3.2.7}$$

Again, it may not be possible to solve the above two equations algebraically. One will have to solve them iteratively by using a computer programme. Thus one obtains the $100(1-\alpha)\%$ confidence bounds θ_ℓ from equation (3.2.6) and θ_u from equation (3.2.7).

When $\beta = 0$, the above equations correspond to the results in Johnson and Kotz (1969, page 58) for the binomial distribution which is a particular case of the GNBD. For $\beta = 1$, equations (3.2.6) and (3.2.7) reduce to

$$\sum_{k=y}^{\infty} \begin{bmatrix} nm+k-1 \\ k \end{bmatrix} \theta_{\ell}^{k} (1-\theta_{\ell})^{nm} = \frac{1}{2}\alpha$$

and

$$\sum_{k=0}^{y} \begin{bmatrix} nm+k-1 \\ k \end{bmatrix} \theta_{u}^{k} (1-\theta_{u})^{nm} = \frac{1}{2}\alpha$$

respectively. These are the corresponding results for the negative binomial distribution.

(iii) **GENERALIZED LOGARITHMIC SERIES DISTRIBUTION:**

The distribution of the sample sum is given by (1.4.4). By applying this distribution to equation (3.2.1) and (3.2.2), we get

$$\sum_{x=y}^{\infty} \frac{n!}{x} \sum_{k=n-1}^{x} (-1)^{x-k-1} \frac{|S_k^{n-1}|}{k!} \begin{bmatrix} -\beta x+x-1 \\ x-k-1 \end{bmatrix} \frac{\left\{\theta_{\ell}(1-\theta_{\ell})^{\beta-1}\right\}^{x}}{\{-\ell n(1-\theta_{\ell})\}^{n}} = \frac{1}{2}\alpha \quad (3.2.8)$$

and

$$\sum_{x=0}^{y} \frac{n!}{x} \sum_{k=n-1}^{x} (-1)^{x-k-1} \frac{|S_k^{n-1}|}{k!} \begin{bmatrix} -\beta x+x-1 \\ x-k-1 \end{bmatrix} \frac{\left\{\theta_{u}(1-\theta_{u})^{\beta-1}\right\}^{x}}{\{-\ell n(1-\theta_{u})\}^{n}} = \frac{1}{2}\alpha. \quad (3.2.9)$$

As in the previous cases, equations (3.2.8) and (3.2.9) are solved iteratively with the help of a computer programme to obtain θ_{ℓ} and θ_u respectively. The quantities θ_{ℓ} and θ_u are the respective lower and upper $100(1-\alpha)\%$ confidence bounds for the parameter θ in the GLSD.

When $\beta = 1$, equations (3.2.8) and (3.2.9) reduce to

$$\sum_{x=y}^{\infty} \frac{n!}{x!} \frac{|S_x^n| \, \theta_\ell^x}{\{-\ln(1-\theta_\ell)\}^n} = \frac{1}{2}\alpha$$

and

$$\sum_{x=0}^{y} \frac{n!}{x!} \frac{|S_x^n| \, \theta_u^x}{\{-\ln(1-\theta_u)\}^n} = \frac{1}{2}\alpha$$

respectively. These are the corresponding results for the logarithmic series distribution which is a particular case of the GLSD.

3.3 CONFIDENCE INTERVALS IN LARGE SAMPLES

When n is large the distribution of the sample sum Y, which is also a modified power series variate as indicated in (3.1.3), converges stochastically to a normal distribution with mean μ_Y and variance σ_Y^2 by the central limit theorem. The mean and variance of Y are given by

$$\mu_Y = n\mu \qquad \text{and} \qquad \sigma_Y^2 = n\sigma^2$$

where μ and σ^2 are defined in (3.1.1). Therefore,

$$W = \frac{Y-n\mu}{\sqrt{n}\,\sigma} = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$$

has a limiting distribution that is normal with mean zero and variance unity. We also know that S^2, the variance of a random sample of size $n \geq 2$, given by

$$S^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

converges stochastically to σ^2. Hence,

$$Z = \frac{W}{(S^2/\sigma^2)^{1/2}} = \frac{\sqrt{n}(\bar{X}-\mu)}{S}$$

has a limiting distribution which is standard normal. Therefore, the $100(1-\alpha)\%$ C.I. for θ can be obtained by using the above. Thus we have

$$1-\alpha = P\left\{- z_{\alpha/2} < \frac{\sqrt{n}(\bar{X}-\mu)}{S} < z_{\alpha/2}\right\}$$

$$= P\left\{\bar{X} - z_{\alpha/2}\, S/\sqrt{n} < \mu < \bar{X} + z_{\alpha/2}\, S/\sqrt{n}\right\}.$$

By using the value of μ given by (3.1.1) in the above result, we obtain

$$1-\alpha = P\left\{\bar{X} - z_{\alpha/2}\, S/\sqrt{n} < \frac{g(\theta)f'(\theta)}{g'(\theta)f(\theta)} < \bar{X} + z_{\alpha/2}\, S/\sqrt{n}\right\}. \quad (3.3.1)$$

Quite often when specific values of $g(\theta)$ and $f(\theta)$ are available, it is possible to solve the inequalities in the above probability for θ

and to re-express equation (3.3.1) in the form

$$1-\alpha = P\left\{\theta_\ell < \theta < \theta_u\right\} \tag{3.3.2}$$

where θ_ℓ and θ_u are the lower and the upper $100(1-\alpha)\%$ confidence bounds for θ respectively. When one cannot do this simplification, one has to get the values of θ_ℓ and θ_u numerically with the help of a computer programme.

3.3.1 APPLICATIONS

We now apply the above result to the three important families of GPD, GNBD and GLSD.

(i) **GENERALIZED POISSON DISTRIBUTION:**

We substitute the mean of GPD from Table 3.1 in equation (3.3.1). Thus we have

$$1-\alpha = P\left\{\bar{X} - z_{\alpha/2}\, S/\sqrt{n} < \frac{\theta}{1-\theta\gamma} < \bar{X} + z_{\alpha/2}\, S/\sqrt{n}\right\}$$

$$= P\left\{\frac{\bar{X}-z_{\alpha/2}S/\sqrt{n}}{1+\gamma(\bar{X}-z_{\alpha/2}S/\sqrt{n})} < \theta < \frac{\bar{X}+z_{\alpha/2}S/\sqrt{n}}{1+\gamma(\bar{X}+z_{\alpha/2}S/\sqrt{n})}\right\}.$$

Hence, the $100(1-\alpha)\%$ C.I. for θ in GPD is

$$\left[\frac{\bar{x}-z_{\alpha/2}s/\sqrt{n}}{1+\varphi(\bar{x}-z_{\alpha/2}s/\sqrt{n})}\ ,\ \frac{\bar{x}+z_{\alpha/2}s/\sqrt{n}}{1+\varphi(\bar{x}+z_{\alpha/2}s/\sqrt{n})}\right]. \qquad (3.3.3)$$

EXAMPLE 3.1 CONTINUED:

We now use the data in example 3.1 to find the approximate 95% C.I. for the paramater θ. The value of φ is taken as 0.162984 which is the ML estimate. The mean and variance of the data are given by 1.38333 and 2.07090 respectively. From (3.3.3), the lower bound for θ is

$$\theta_{\ell} = \frac{\bar{x}-z_{\alpha/2}s/\sqrt{n}}{1+\varphi(\bar{x}-z_{\alpha/2}s/\sqrt{n})}$$

$$= \frac{1.38333 - 1.96\ (2.0709/60)^{1/2}}{1 + 0.162984\ [1.38333-1.96(2.0709/60)^{1/2}]}$$

$$= \frac{1.019197}{1.166113}$$

$$= 0.87.$$

In the same way, the upper bound for θ is

$$\theta_u = 1.36.$$

Hence the approximate 95% C.I. is (0.87, 1.36).

(ii) **GENERALIZED NEGATIVE BINOMIAL DISTRIBUTION:**

For the GNBD, we apply its mean given in Table 3.1 to equation (3.3.1). This gives

$$1-\alpha = P\left\{\bar{X} - z_{\alpha/2}\, S/\sqrt{n} < \frac{\theta m}{1-\theta\beta} < \bar{X} + z_{\alpha/2}\, S/\sqrt{n}\right\}$$

$$= P\left\{\frac{\bar{X}-z_{\alpha/2}S/\sqrt{n}}{m+\beta(\bar{X}-z_{\alpha/2}S/\sqrt{n})} < \theta < \frac{\bar{X}+z_{\alpha/2}S/\sqrt{n}}{m+\beta(\bar{X}+z_{\alpha/2}S/\sqrt{n})}\right\}.$$

Therefore, the 100(1−α)% C.I. for θ in GNBD is given by

$$\left[\frac{\bar{x}-z_{\alpha/2}s/\sqrt{n}}{m+\beta(\bar{x}-z_{\alpha/2}s/\sqrt{n})}\ ,\ \frac{\bar{x}+z_{\alpha/2}s/\sqrt{n}}{m+\beta(\bar{x}+z_{\alpha/2}s/\sqrt{n})}\right]. \qquad (3.3.4)$$

(iii) **GENERALIZED LOGARITHMIC SERIES DISTRIBUTION:**

We now consider the case of GLSD. By using the value of μ from Table 3.1 in equation (3.3.1), we obtain

$$1-\alpha = P\left\{\bar{X} - z_{\alpha/2}\, S/\sqrt{n} < \frac{\theta(1-\theta\beta)^{-1}}{\{-\ell n(1-\theta)\}} < \bar{X} + z_{\alpha/2}\, S/\sqrt{n}\right\}. \qquad (3.3.5)$$

It is not an easy task to express (3.3.5) in the form of (3.3.2). However, one can obtain θ_ℓ and θ_u by solving the following equations.

$$\frac{\theta(1-\theta\beta)^{-1}}{\{-\ell n(1-\theta)\}} = \bar{x} - z_{\alpha/2}\ s/\sqrt{n} \tag{3.3.6}$$

and

$$\frac{\theta(1-\theta\beta)^{-1}}{\{-\ell n(1-\theta)\}} = \bar{x} + z_{\alpha/2}\ s/\sqrt{n}. \tag{3.3.7}$$

Equation (3.3.6) and (3.3.7) are solved numerically by using iterative procedure to yield θ_ℓ and θ_u respectively.

Consider the left hand side of the above equations and let

$$R = \frac{\theta(1-\theta\beta)^{-1}}{\{-\ell n(1-\theta)\}}$$

$$\frac{dR}{d\theta} = \frac{(1-\theta\beta)\{-\ell n(1-\theta)\} - \theta\left\{\frac{1-\theta\beta}{1-\theta} + \beta\ell n(1-\theta)\right\}}{(1-\theta\beta)^2\ \{-\ell n(1-\theta)\}^2}$$

$$= \frac{-\ell n(1-\theta) - \theta(1-\theta\beta)(1-\theta)^{-1}}{(1-\theta\beta)^2\ \{\ell n(1-\theta)\}^2}$$

$$> 0, \qquad\qquad \text{as} \quad 1-\theta\beta < 1-\theta.$$

Therefore, R is an increasing function of θ. As $\lim_{\theta\to 0} R = 1$, the lowest value of R is 1. Also $R \to \infty$ when $\theta \to \beta^{-1}$. Thus, the left hand side of equations (3.3.6) and (3.3.7) can take all values from 1 to ∞. Accordingly, the equation (3.3.6) will have no solution if $\bar{x} - z_{\alpha/2}\ s/\sqrt{n} < 1$, will have $\theta_\ell = 0$ if $\bar{x} - z_{\alpha/2}\ s/\sqrt{n} = 1$ and a unique value of θ_ℓ when $\bar{x} - z_{\alpha/2}\, s/\sqrt{n} > 1$. Similarly, θ_u will exist as a unique value when $\bar{x} + z_{\alpha/2}\, s/\sqrt{n} > 1$.

3.4 UNIFORMLY MOST ACCURATE CONFIDENCE BOUND

We assume that the function $g(\theta)$ is a monotone function of θ. This assumption does not create any handicap as all the members of MPSD class, considered in section 3.1, satisfy this assumption. In fact, $g(\theta)$ is an increasing function of θ in all the three families of GPD, GNBD and GLSD. It can be shown that there exists a uniformly most powerful (UMP) level α test for hypotheses of the form

$$H_0:\ \theta = \theta_0 \text{ (known)} \qquad \text{against}$$
$$H_a:\ \theta = \theta_1 > \theta_0$$

in the class of MPSD. We obtain a uniformly most accurate (UMA) confidence bound for the parameter θ by using the critical region for the UMP test.

By using Neyman-Pearson theorem, C is the best critical region of size α if

$$\frac{L(\theta_1)}{L(\theta_0)} = \left\{\frac{g(\theta_1)}{g(\theta_0)}\right\}^{\sum^n x_i} \left\{\frac{f(\theta_0)}{f(\theta_1)}\right\}^n > k$$

for $(x_1, x_2, \ldots, x_n) \in C$ where k is a constant. Thus, on taking logarithms of the above

$$\sum_{i=1}^{n} X_i . \log\left\{\frac{g(\theta_1)}{g(\theta_0)}\right\} > \log\left\{k\left[\frac{f(\theta_1)}{f(\theta_0)}\right]^n\right\}.$$

Since $g(\theta)$ is an increasing function of θ in all the families considered we have that

$$\sum_{i=1}^{n} X_i > \log\left\{k\left[\frac{f(\theta_1)}{f(\theta_0)}\right]^n\right\} \Big/ \log\left\{\frac{g(\theta_1)}{g(\theta_0)}\right\} = k^*$$

and so

$$Y = \sum_{i=1}^{n} X_i > k^*.$$

The critical region C is given by

$$C = \left\{(x_1, x_2, \ldots, x_n) : Y > k^*\right\}$$

where k^* is determined from

$$\alpha = P\left\{Y > k^* \mid H_0\right\} \tag{3.4.1}$$

if randomization is ignored. By using the probability function of Y in (3.1.3), we obtain

$$\alpha = \sum_{x=k^*+1}^{\infty} b(x) \left\{g(\theta_0)\right\}^x \left\{f(\theta_0)\right\}^{-n}. \tag{3.4.2}$$

Let the value of k^* obtained from (3.4.2) be k_0. When n is small, k_0 can be determined from equation (3.4.2) by using a computer programme after the values of $b(x)$, $g(\theta_0)$ and $f(\theta_0)$ are substituted from Table 3.1. Since randomization is ignored, we choose that smallest value of k^* as k_0 for which the right hand side of (3.4.2) is $\leq \alpha$. However, if n is large, normal approximation can be used to determine k^* from (3.4.1).

The UMP size α test is to reject H_0 if $Y > k_0$. On taking its expectation we have

$$\frac{ng(\theta)f'(\theta)}{g'(\theta)f(\theta)} > k_0 \ .$$

Hence the corresponding 100(1-α)% UMA upper confidence bound for θ is given by

$$1-\alpha = P\left\{\theta \leq \theta_u\right\} \tag{3.4.3}$$

where θ_u is the solution of equation

$$\frac{ng(\theta)f'(\theta)}{g'(\theta)f(\theta)} = k_0 . \tag{3.4.4}$$

Sometimes one may not be able to solve equation (3.4.4) through ordinary algebraic manipulations. In such circumstances, numerical solutions can be obtained through iteration with the help of a computer programme.

3.4.1 APPLICATIONS

We now apply the above procedure to the GPD, GNBD and GLSD families.

(i) GENERALIZED POISSON DISTRIBUTION:

By using the mean of GPD from Table 3.1 in (3.4.4), we have

$$\frac{n\theta}{1-\theta\gamma} = k_0$$

which gives $\theta = \dfrac{k_0}{n+\gamma k_0}$.

Thus, the $100(1-\alpha)\%$ UMA upper confidence bound for the parameter θ is

$$1-\alpha = P\left\{\theta \le \frac{k_0}{n+\gamma k_0}\right\}. \tag{3.4.5}$$

(ii) GENERALIZED NEGATIVE BINOMIAL DISTRIBUTION:

For the GNBD, we use μ in Table 3.1 to obtain from (3.4.4) the equation

$$\frac{nm\theta}{1-\theta\beta} = k_0.$$

On solving for θ in the above equation, we get

$$\theta = \frac{k_0}{nm+\beta k_0}$$

and so the 100(1-α)% UMA upper confidence bound for the parameter θ in GNBD is given by

$$1-\alpha = P\left\{\theta \le \frac{k_0}{nm+\beta k_0}\right\}. \tag{3.4.6}$$

(iii) **GENERALIZED LOGARITHMIC SERIES DISTRIBUTION:**

In the case of GLSD, the value of μ from Table 3.1 is used in equation (3.4.4) to obtain

$$\frac{n\theta(1-\theta\beta)^{-1}}{\{-\ln(1-\theta)\}} = k_0.$$

The above equation is solved numerically for θ. As shown in section 3.3, the left hand side of the above equation is an increasing function of θ and the equation will have a unique solution if $n - k_0 < 0$. Thus, the 100(1-α)% UMA upper confidence bound for the parameter θ in GLSD is given by (3.4.3) where θ_u is the solution of the above equation.

By considering the dual hypotheses problem H_0: $\theta = \theta_0$ against H_a: $\theta = \theta_1 < \theta_0$, one can obtain a UMA lower confidence bound for the parameter θ in each of the above three families of the MPSD class. This can be done by reversing all inequalities in the procedure for UMA upper confidence bound.

CHAPTER IV

ESTIMATION IN SMALL SAMPLES FOR GENERALIZED POISSON DISTRIBUTION

4.1 INTRODUCTION

The GPD model having two parameters θ and λ is given by (1.3.1). The restricted model is also provided by (1.3.2). Since the introduction of the GPD model, a large body of literature, mainly on its interesting properties, point estimators and applications has developed. [See the work of Charalambides (1974), Gupta (1977), Kumar and Consul (1980) and Consul and Shoukri (1984, '85)].

There is no two dimensional sufficient statistics for either of the two parameters (θ,λ) in the GPD model (1.3.1) or (θ,φ) in the restricted GPD model (1.3.2). It is only in the model (1.3.2) that the sample sum $Y = \sum_{i=1}^{n} X_i$ is a complete and sufficient statistic for the parameter θ when φ is assumed known.

We consider the problem of interval estimation of the GPD model under the following headings:

(i) when one parameter is assumed known

(ii) inference on a single parameter when both parameters are unknown and

(iii) inference on both parameters.

When none of the parameters is known, it is not an easy task to obtain a confidence interval (C.I.) or a confidence region (C.R.) for any or both of the parameters in a small sample. This problem will be considered under plausibility inference.

We assume that a random sample $X_1, X_2, \ldots, X_n$ of size n is taken from a discrete population with generalized Poisson distribution. The likelihood function is

$$L(\theta,\lambda) = \prod_{i=1}^{n} P_{x_i}(\theta,\lambda)$$

$$= \theta^n e^{-n(\theta+\lambda\bar{x})} \prod_{i=1}^{n} \left[\left[\theta+\lambda x_i\right]^{x_i-1}/(x_i)!\right] \qquad (4.1.1)$$

for the general model (1.3.1) and

$$L(\theta,\theta\varphi) = \theta^{n\bar{x}} e^{-n\theta(1+\varphi\bar{x})} \prod_{i=1}^{n} \left[\left[1+\varphi x_i\right]^{x_i-1}/(x_i)!\right] \quad (4.1.2)$$

for the restricted model (1.3.2). Through out this chapter, the parameter space for λ in the GPD model (1.3.1) and for φ in the GPD model (1.3.2) will be restricted to $(0,1)$ and $(0,\theta^{-1})$ respectively for the sake of convenience. Thus, the negative parts of the domains of λ in (1.3.1) and of φ in (1.3.2) are ignored. We obtain two sided confidence intervals and likelihood intervals as well as the likelihood regions for the parameters of the GPD model.

4.2 CONFIDENCE INTERVALS FOR A SINGLE PARAMETER

Suppose one of the parameters of the GPD model is known. Given a small sample, we use $1-\alpha$ confidence level in such a way that the total probabilities on each tail of the GPD model are $\le \alpha/2$.

CONFIDENCE INTERVAL FOR θ OR γ IN MODEL (1.3.2):

Suppose that the parameter λ in GPD model (1.3.1) is a known multiple γ_0 of θ. The parametric transformation in model (1.3.1) yields the model (1.3.2) with a known parameter $\gamma = \gamma_0$. To set C.I. for the unknown parameter θ, we make use of the sample sum $Y = \sum^{n} X_i$, which is a complete and sufficient statistic for θ. This problem is similar to the one considered in subsection 3.2.1 and will therefore not be repeated here.

Suppose we want to set C.I. for γ when θ is known to be θ_0, we use the sample sum distribution given by (3.2.3). The $100(1-\alpha)\%$ confidence bounds for γ can be obtained by solving the following equations for γ_ℓ and γ_u.

$$\sum_{x=y}^{\infty} n(n+\gamma_\ell x)^{x-1}\,\theta_0^x\, e^{-\theta_0(n+\gamma_\ell x)}/x! = \frac{1}{2}\alpha \tag{4.2.1}$$

and

$$\sum_{x=0}^{y} n(n+\gamma_u x)^{x-1}\,\theta_0^x\, e^{-\theta_0(n+\gamma_u x)}/x! = \frac{1}{2}\alpha. \tag{4.2.2}$$

It is very hard to get specific values of γ_ℓ in (4.2.1) and γ_u in (4.2.2) algebraically but the equations can be solved numerically on a computer. We have seen through many examples that these equations give unique values every time. In general, for a given value of γ, the sum of the probabilities in (4.2.1) and (4.2.2) need not add up to $\frac{1}{2}\alpha$, however, for a fixed $Y = y$, the expressions in (4.2.1) and (4.2.2) contain continuous functions of γ and accordingly exact values of γ can be obtained to satisfy these equations. Let γ_ℓ and γ_u be the numerical values of γ given by equations (4.2.1) and (4.2.2) respectively. The C.I. for γ is given by (γ_ℓ, γ_u).

CONFIDENCE INTERVAL FOR θ OR λ IN MODEL (1.3.1):

If λ is not a function of θ and it is a known quantity in the GPD model (1.3.1) and we want to find C.I. for the parameter θ, we may still use the above procedure. The main difference will be that the expressions will be slightly more complicated and the C.Is. will be generally wider than those obtained from (4.2.1) and (4.2.2) because Y is not a sufficient statistic for θ any more. For this case, we solve the following equations for θ_ℓ and θ_u.

$$\sum_{x=y}^{\infty} n\theta_\ell(n\theta_\ell+\lambda_0 x)^{x-1} e^{-(n\theta_\ell+\lambda_0 x)}/x! = \frac{1}{2}\alpha \qquad (4.2.3)$$

and

$$\sum_{x=0}^{y} n\theta_u(n\theta_u+\lambda_0 x)^{x-1} e^{-(n\theta_u+\lambda_0 x)}/x! = \frac{1}{2}\alpha. \qquad (4.2.4)$$

The solutions θ_ℓ of equation (4.2.3) and θ_u of equation (4.2.4) will give the $100(1-\alpha)\%$ C.I. for θ as (θ_ℓ, θ_u).

The previous claim of wider C.I. was verified by simulating a number of pseudo-random samples, each of size 40, from the GPD models. Both methods were used to obtain C.Is. for θ and, in general, the use of equations (3.2.4) and (3.2.5) gave shorter C.Is. than the formulae (4.2.3) and (4.2.4).

When the point ML estimate of θ is more than 10, we can obtain a sharper C.I. for θ by using the property of normal approximation. This property was given by Consul and Shenton (1973) who showed that the GP variate approaches a normal variate if θ is very large. Thus we have that

$$Z = \frac{X-\mu}{\sigma}$$

is approximately normally distributed with mean zero and variance unity. The approximate $100(1-\alpha)\%$ C.I. for θ is obtained by assuming that the sample mean $\bar{X}$ is normally distributed. We note here that this assumption of normality does not depend on the sample size n.

We now assume that the parameter θ is known and it is equal to θ_0. Then θ is replaced by θ_0 in the equations (4.2.3) and (4.2.4) and λ_0 in (4.2.3) and (4.2.4) is replaced by λ. The two equations are solved numerically on a computer to obtain the values of λ_ℓ and λ_u from (4.2.3) and (4.2.4) respectively. These two values become the boundaries of $100(1-\alpha)\%$ C.I. for the parameter λ.

EXAMPLE 4.1

We consider the data for example 3.1 and for convenience, the table is partially reproduced here.

No. of deaths	0	1	2	3	4	5	≥ 6	TOTAL
Observed freq.	20	17	11	8	2	0	2	60

The method in this section is applied to find C.I. for the parameter θ when the parameter λ is known. Since it is hard to know the actual value of λ in a real life data, we shall use the ML estimate of λ as the known value λ_0 of λ. The ML estimates of θ and λ in the GPD model (1.3.1) are 1.128826 and 0.183981 respectively.

Now, by putting $\lambda_0 = 0.183981$ in equations (4.2.3) and (4.2.4) and solving these equations numerically, the 95% C.I. for θ is found to be (0.85, 1.46). The two bounds are quite close to the value 1.129. In the same way, we use equations (4.2.3) and (4.2.4) to determine the 95% C.I. for λ, this yields (0.0, 0.43). From example 3.1, the ML estimates of θ and γ in the model (1.3.2) are 1.128826 and 0.162984 respectively. By putting $\theta_0 = 1.128826$ in equations (4.2.1) and (4.2.2) and solving them for $\alpha = 0.05$ with the help of a computer we obtain the 95% C.I. for γ as (0.0, 0.38).

By comparing the C.I. for θ obtained here with the one obtained in example 3.1, it is easy to note that the C.I. from the GPD model (1.3.2) is shorter. From this result and our empirical simulation results, it seems as if an interval based on a sufficient statistic is shorter than the one without a sufficient statistic.

4.3 PLAUSIBILITY INFERENCE

Plausibility inference is the use of likelihood functions to examine the parameter space and to determine which values of the parameters are likely (plausible) and which are implausible on the basis of the given data [See Kalbfleisch, 1979 page 20]. It is an exact statistical inference since it does not require mathematical approximation on the basis of large sample size n. By this procedure, likelihood statements can be made on the parameters of a distribution.

By using the likelihood function in (4.1.1) and with parameter λ known, the likelihood ratio for θ_0 versus θ_1, where θ_0 and θ_1 are distinct, is

$$\ell = \frac{L(\theta_0, \lambda)}{L(\theta_1, \lambda)}.$$

If $\ell > 1$, one can say that θ_0 is a more plausible value of the parameter θ than the value θ_1 since the data are more probable for $\theta = \theta_0$ than they are for $\theta = \theta_1$. For instance, if $\ell = 3$, then θ_0 is thrice as plausible as θ_1 in the sense that the data are thrice as probable if θ_0 is true than if θ_1 is true.

When both parameters θ and λ are unknown, the likelihood function is standardized with respect to its maximum to obain the relative likelihood function as

$$R(\theta,\lambda) = \frac{L(\theta,\lambda)}{L(\hat{\theta},\hat{\lambda})} \qquad (4.3.1)$$

where $\hat{\theta}$ and $\hat{\lambda}$ are the ML estimates of θ and λ respectively.

Since $R(\theta,\lambda)$ ranks the plausibilities of the values of θ and λ with respect to the most plausible values $\hat{\theta}$ and $\hat{\lambda}$, we infer that $0 \leq R(\theta,\lambda) \leq 1$ for all (θ,λ) in the parametric space. The likelihood ratio for (θ_0, λ_0) versus $(\hat{\theta},\hat{\lambda})$ becomes

$$R(\theta_0, \lambda_0) = \frac{L(\theta_0, \lambda_0)}{L(\hat{\theta}, \hat{\lambda})}.$$

If $R(\theta_0, \lambda_0)$ is small (e.g. $R(\theta_0, \lambda_0) \leq 0.1$), it implies that the pair (θ_0, λ_0) provides simultaneously implausible parametric values and there are other values of (θ,λ) for which the data are ten times more probable. If $R(\theta_0, \lambda_0)$ is large (e.g. $R(\theta_0, \lambda_0) \geq 0.6$), the pair (θ_0, λ_0) provides simultaneously more plausible parametric values since (θ_0, λ_0) gives to the data at least 60% of the maximum probability which is possible under the model.

The set of parameter values for which $R(\theta,\lambda) \geq \nu$ is called a $100\nu\%$ likelihood region. If one of the parameters is known or estimated out, the set of values given by $R(\theta,\lambda) \geq \nu$ will be called the likelihood interval (L.I.).

Kalbfleisch (1979) has suggested 50%, 10% and 1% likelihood intervals (regions). Values inside the 10% L.I. are said to be "plausible" and values outside this as "implausible". Values inside the 50% L.I. as "very plausible" and values outside the 1% as "very implausible". The choice of these division points by him is somewhat arbitrary, though they seem to be intuitively justified. Instead of considering the same values, we would assign slightly different values which seem to be more logical.

Let $r(\theta,\lambda) = \log R(\theta,\lambda)$ where $-\infty < r(\theta,\lambda) \le 0$, and $R(\theta,\lambda) \ge \nu$ implies that $r(\theta,\lambda) \ge \log \nu$.

We note that 50% likelihood region ($\nu = 0.5$) means that using the available data, any value of (θ,λ) in the region is a reasonable guess at the values of θ and λ. This does not mean 50% confidence region for (θ,λ). However, if the sample size n is large, approximate probability statement may be attached to the likelihood region. On the basis of this, we select the division points which correspond to 100(1-α)% confidence regions.

For a large sample

$$\ell(\theta,\lambda) = -2 \log R(\theta,\lambda) = -2\, r(\theta,\lambda) \qquad (4.3.2)$$

is approximately chi-square distributed with 2 d.f.. Therefore, the set of values of (θ,λ) for which $\ell(\theta,\lambda) \le \chi^2_{\alpha,2}$ gives an approximate 100(1-α)% confidence region for (θ,λ) where $\chi^2_{\alpha,2}$ is the upper percent point of the chi-square distribution with 2 d.f.

Therefore,

$$\ell(\theta,\lambda) \le \chi^2_{\alpha,2} \quad \text{iff} \quad r(\theta,\lambda) \ge -\frac{1}{2}\chi^2_{\alpha,2}.$$

From the above, $R(\theta,\lambda) \ge \exp\left[-\frac{1}{2}\chi^2_{\alpha,2}\right] = \nu$. For 95% confidence region, we have $\nu = 0.05$ which is equivalent to saying 5% likelihood region. Corresponding to the 99% and 90% confidence regions are the 1% and 10% likelihood regions. Values inside the 5% likelihood region will be considered to be "plausible", those outside "implausible". Also, values outside the 1% likelihood region will be called "very implausible".

If inference is required on only one of the parameters of the GPD model (1.3.1), we shall be setting likelihood intervals and the function in (4.3.2) will thus be chi-square distributed with 1 d.f.. Corresponding to the 90%, 95% and 99% confidence intervals are the 25.85%, 14.65% and 3.62% likelihood intervals.

We make use of large sample property in order to select our division points. We do not imply that plausibility inference is applicable to only large samples. Of course, the division points do not depend on the sample size. It may not be out of place to state here that the procedure is applicable to small samples as well as large samples.

4.3.1 PLAUSIBILITY INFERENCE ON ONE PARAMETER

In this subsection, we consider the likelihood interval for a single parameter when the other parameter is unknown. This unknown parameter, called a nuisance parameter, will be eliminated. Since we consider only the model (1.3.1), we apply the 'method of maximization' for the elimination of the nuisance parameter.

Suppose the parameter of interest is λ. Accordingly, θ becomes a nuisance parameter. We eliminate θ by maximizing $R(\theta,\lambda)$ over θ. The value of θ that will maximize $R(\theta,\lambda)$ is the same as that θ which will maximize $L(\theta,\lambda)$ with a specified value of λ.

By partial logarithmic differentiation of (4.1.1) with respect to θ and equating to zero, we have

$$0 = \frac{\partial}{\partial\theta} \ln L(\theta,\lambda)$$

$$= \frac{\partial}{\partial\theta} \left\{ n\ln\theta - [\theta+\lambda\bar{x}]n + \sum_{i=1}^{n} [(x_i-1)\ln(\theta+\lambda x_i) - \ln(x_i!)] \right\}$$

$$= \frac{n}{\theta} - n + \sum_{i=1}^{n} \frac{x_i-1}{\theta+\lambda x_i}.$$

Thus

$$\sum_{i=1}^{n} \frac{x_i-1}{\theta+\lambda x_i} + \frac{n(1-\theta)}{\theta} = 0. \qquad (4.3.3)$$

Let $\tilde{\theta}(\lambda)$ denote the value of θ that satisfies (4.3.3) for any given value of λ. Since the ML estimates of the GPD model (1.3.1) are unique [See Consul and Shoukri, 1984] the solution $\tilde{\theta}(\lambda)$ will be unique too.

From the above, we get

$$\max_{\theta} R(\theta,\lambda) = R(\tilde{\theta}(\lambda), \lambda)$$

$$= R_m(\lambda), \qquad \text{(say)}.$$

Thus the relative likelihood function is given as

$$R_m(\lambda) = \left[\frac{\tilde{\theta}(\lambda)}{\hat{\theta}}\right]^n e^{-n[\tilde{\theta}(\lambda)-\hat{\theta}+(\lambda-\hat{\lambda})\bar{x}]} \cdot \prod_{i=1}^{n} \left[\frac{\tilde{\theta}(\lambda)+\lambda x_i}{\hat{\theta}+\hat{\lambda}x_i}\right]^{x_i-1}. \qquad (4.3.4)$$

Hence,

$$r_m(\lambda) = \ln R_m(\lambda)$$

$$= n\left[\ln\left[\frac{\tilde{\theta}(\lambda)}{\hat{\theta}}\right] - (\tilde{\theta}(\lambda)-\hat{\theta})-(\lambda-\hat{\lambda})\bar{x}\right]$$

$$+ \sum_{i=1}^{n} (x_i-1) \ln\left[\frac{\tilde{\theta}(\lambda)+\lambda x_i}{\hat{\theta}+\hat{\lambda}x_i}\right]. \tag{4.3.5}$$

For each specified value of λ, we solve equation (4.3.3) iteratively by the help of a computer programme to obtain a numerical value of $\tilde{\theta}(\lambda)$ which is subsequently used in (4.3.5). The value of $r_m(\lambda)$ is computed by another computer programme. Thus, we plot numerous values of $r_m(\lambda)$ and trace the graph of the function in (4.3.5). This graph of $r_m(\lambda)$ for $\lambda \in (0,1)$ will be used to determine the L.I.. The $100\nu\%$ L.I. is the set of points for which $r_m(\lambda) \geq \log \nu$. The L.I. bounds are obtained by finding the two points of intersection of the graph of the function (4.3.5) and the straight line

$$r_m(\lambda) = \log \nu.$$

The graph of $-r_m(\lambda)$ is U-shaped or concave up.

A similar result can be obtained for the L.I. for the parameter θ. In this case, we maximize $R(\theta,\lambda)$ over λ. The value of λ that will maximize $R(\theta,\lambda)$ is the same as that λ which will maximize $L(\theta,\lambda)$ with a specified value of θ. By partial logarithmic differentiation of (4.1.1) with respect to λ and equating to zero we obtain

$$0 = \frac{\partial}{\partial\lambda} \ln L(\theta,\lambda)$$

$$= \frac{\partial}{\partial\lambda}\left\{n\ell n\theta - (\theta+\lambda\bar{x})n + \sum_{i=1}^{n}\left[(x_i-1)\ell n(\theta+\lambda x_i)-\ell n(x_i!)\right]\right\}$$

$$= -n\bar{x} + \sum_{i=1}^{n}\frac{x_i(x_i-1)}{\theta+\lambda x_i}$$

which gives

$$\sum_{i=1}^{n}\frac{x_i(x_i-1)}{\theta+\lambda x_i} - n\bar{x} = 0. \qquad (4.3.6)$$

Suppose $\tilde{\lambda}(\theta)$ is the value of λ that satisfies (4.3.6) for any given value of θ. Then

$$\max_{\lambda} R(\theta,\lambda) = R(\theta, \tilde{\lambda}(\theta))$$

$$= R_m(\theta).$$

Hence,

$$R_m(\theta) = \left[\frac{\theta}{\hat{\theta}}\right]^n e^{-[\theta-\hat{\theta}+(\tilde{\lambda}(\theta)-\hat{\lambda})\bar{x}]n} \cdot \prod_{i=1}^{n}\left[\frac{\theta+\tilde{\lambda}(\theta)x_i}{\hat{\theta}+\hat{\lambda}x_i}\right]^{x_i-1}. \qquad (4.3.7)$$

By taking the logarithm of the above relation, it yields

$$r_m(\theta) = n\left[\ell n\left[\frac{\theta}{\hat{\theta}}\right] - (\theta-\hat{\theta}) - (\tilde{\lambda}(\theta)-\hat{\lambda})\bar{x}\right]$$

$$+ \sum_{i=1}^{n}(x_i-1)\,\ell n\left[\frac{\theta+\tilde{\lambda}(\theta)x_i}{\hat{\theta}+\hat{\lambda}x_i}\right]. \qquad (4.3.8)$$

To compute the values of $r_m(\theta)$ in (4.3.8), we first obtain $\tilde{\lambda}(\theta)$ from equation (4.3.6). For each specified value of θ and with the help of a computer programme, we solve equation (4.3.6) iteratively to obtain the value of $\tilde{\lambda}(\theta)$. On substituting $\tilde{\lambda}(\theta)$ into (4.3.8) and by using another computer programme, we obtain the values of $r_m(\theta)$. These values are subsequently used to trace the graph of the function in (4.3.8). The graph of $-r_m(\theta)$ is U-shaped and this will be used to obtain the L.I. for the parameter θ.

EXAMPLE 4.2

The following data is on the distribution of sow bugs (Trachelipus rathkei) under boards taken from Janardan et. al. (1979).

TABLE 4.1

The distribution of sow bugs (Trachelipus rathkei)

No. per board	0	1	2	3	4	5-6	7-9	10-12	≥ 13	Total
Observed freq.	28	28	14	11	8	13	9	5	6	122
Expected GPD freq.	28.77	23.70	17.36	12.58	9.21	11.97	9.24	4.38	4.79	122.00

$$\bar{x} = 3.29508, \qquad s^2 = 15.05270$$

The ML estimation method was used to fit the GPD model to the observed data. It was found that the GPD model fitted the data very well.

The ML estimates of θ and λ are 1.444620 and 0.561583 respectively. We apply the data in Table 4.1 to equation (4.3.3) to obtain

$$0 = -\frac{28}{\theta} + \frac{14}{\theta+2\lambda} + \frac{22}{\theta+3\lambda} + \frac{24}{\theta+4\lambda}$$

$$+ \frac{44}{\theta+5\lambda} + \frac{10}{\theta+6\lambda} + \frac{18}{\theta+7\lambda} + \frac{21}{\theta+8\lambda}$$

$$+ \frac{24}{\theta+9\lambda} + \frac{27}{\theta+10\lambda} + \frac{20}{\theta+11\lambda} + \frac{12}{\theta+13\lambda}$$

$$+ \frac{26}{\theta+14\lambda} + \frac{14}{\theta+15\lambda} + \frac{32}{\theta+17\lambda} + \frac{122(1-\theta)}{\theta}. \qquad (4.3.9)$$

Equation (4.3.9) is used to obtain $\tilde{\theta}(\lambda)$ for each specified value of λ. In the same manner, by substituting the data into (4.3.5), we get

$$r_m(\lambda) = 122\left\{\ell n\left[\frac{\tilde{\theta}(\lambda)}{1.4446}\right] - (\tilde{\theta}(\lambda)-1.4446) - (\lambda-.5616)(3.2951)\right\}$$

$$- 28\ \ell n\left[\frac{\tilde{\theta}(\lambda)}{1.4446}\right] + 14\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 2\lambda}{1.4446+2(.5616)}\right]$$

$$+ 22\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 3\lambda}{1.4446+3(.5616)}\right] + 24\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 4\lambda}{1.4446+4(.5616)}\right]$$

$$+ 44\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 5\lambda}{1.4446+5(.5616)}\right] + 10\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 6\lambda}{1.4446+6(.5616)}\right]$$

$$+ 18\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 7\lambda}{1.4446+7(.5616)}\right] + 21\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 8\lambda}{1.4446+8(.5616)}\right]$$

$$+ 24\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 9\lambda}{1.4446+9(.5616)}\right] + 27\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 10\lambda}{1.4446+10(.5616)}\right]$$

$$+ 20\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 11\lambda}{1.4446+11(.5616)}\right] + 12\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 13\lambda}{1.4446+13(.5616)}\right]$$

$$+ 26\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 14\lambda}{1.4446+14(.5616)}\right] + 14\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 15\lambda}{1.4446+15(.5616)}\right]$$

$$+ 32\ \ell n\left[\frac{\tilde{\theta}(\lambda) + 17\lambda}{1.4446+17(.5616)}\right]$$

$$= 94\ \ell n\left[\frac{\tilde{\theta}(\lambda)}{1.4446}\right] - 122[\tilde{\theta}(\lambda)+\lambda-3.2951]$$

$$+ 14\ \ell n\left[\frac{\tilde{\theta}(\lambda)+2\lambda}{2.5678}\right] + 22\ \ell n\left[\frac{\tilde{\theta}(\lambda)+3\lambda}{3.1294}\right] + 24\ \ell n\left[\frac{\tilde{\theta}(\lambda)+4\lambda}{3.6910}\right]$$

$$+ 44\ \ell n\left[\frac{\tilde{\theta}(\lambda)+5\lambda}{4.2526}\right] + 10\ \ell n\left[\frac{\tilde{\theta}(\lambda)+6\lambda}{4.8142}\right] + 18\ \ell n\left[\frac{\tilde{\theta}(\lambda)+7\lambda}{5.3758}\right]$$

$$+ 21\ \ell n\left[\frac{\tilde{\theta}(\lambda)+8\lambda}{5.9374}\right] + 24\ \ell n\left[\frac{\tilde{\theta}(\lambda)+9\lambda}{6.4990}\right] + 27\ \ell n\left[\frac{\tilde{\theta}(\lambda)+10\lambda}{7.0606}\right]$$

$$+ 20\ \ell n\left[\frac{\tilde{\theta}(\lambda)+11\lambda}{7.6222}\right] + 12\ \ell n\left[\frac{\tilde{\theta}(\lambda)+13\lambda}{8.7454}\right] + 26\ \ell n\left[\frac{\tilde{\theta}(\lambda)+14\lambda}{9.3070}\right]$$

$$+ 14\ \ell n\left[\frac{\tilde{\theta}(\lambda)+15\lambda}{9.8686}\right] + 32\ \ell n\left[\frac{\tilde{\theta}(\lambda)+17\lambda}{10.9918}\right]. \qquad (4.3.10)$$

We used the expression in (4.3.10) to draw the graph of $-r_m(\lambda)$. For each value of λ, we first calculated $\tilde{\theta}(\lambda)$ from equation (4.3.9). All these were done on the computer for this example. The graph was used to obtain the 25.85% , 14.65% and 3.62% likelihood intervals for the parameter λ. For the case of 14.65% L.I. we obtained the interval (0.47, 0.66).

In the same way, we applied the data to equation (4.3.6) and expression (4.3.8). The equation yielded the value of $\tilde{\lambda}(\theta)$ which was subsequently used in the expression (4.3.8). The graph of the resulting expression was used to obtain the L.Is. for the parameter θ. We obtained the interval (1.24, 1.75) as the 14.65% L.I. for the parameter θ when λ is eliminated.

The U-shaped graphs of $-r_m(\lambda)$ and $-r_m(\theta)$ are shown in Figure 4.1 and Figure 4.2 respectively. Values of λ in the interval (0.47, 0.66) are said to be plausible as well as those values of θ in the interval (1.24, 1.75). By considering the data in Table 4.1 as a large sample, these intervals will correspond to approximate 95% confidence intervals.

FIGURE 4.1

3.62, 14.65 and 25.85% likelihood intervals for λ

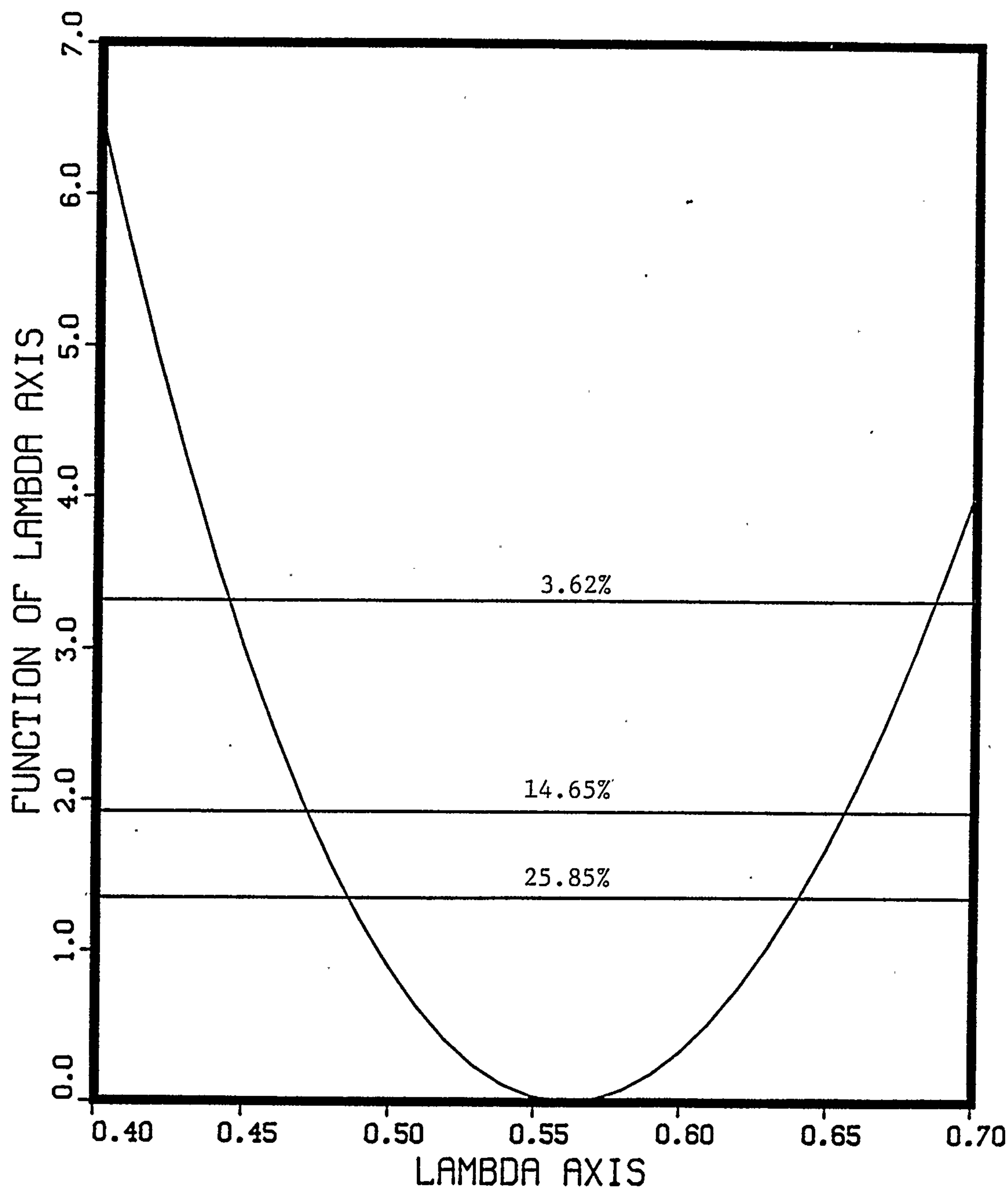
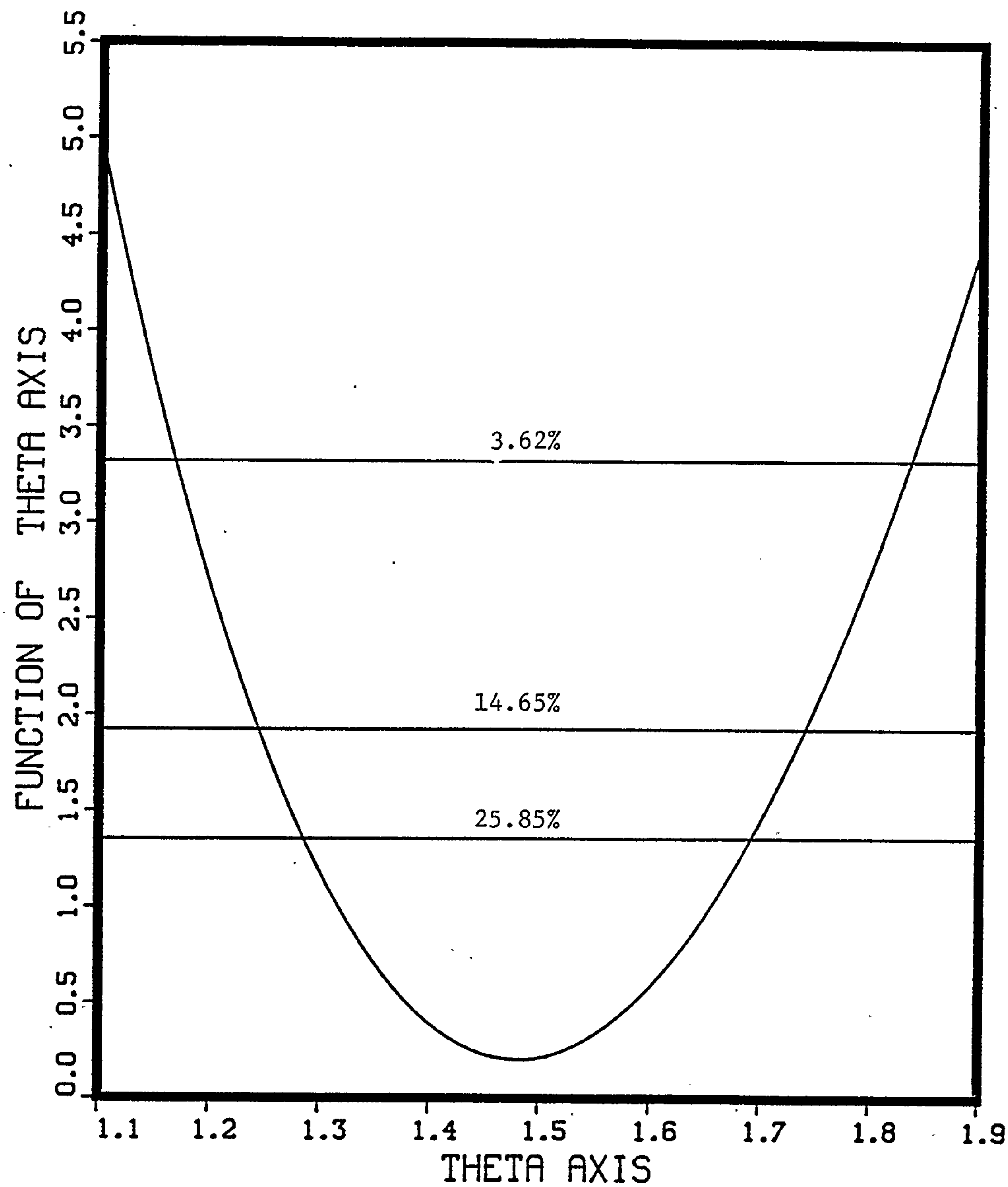


FIGURE 4.2

3.62, 14.65 and 25.85% likelihood intervals for θ



4.3.2 INFERENCE ON BOTH PARAMETERS

To determine the likelihood regions for both parameters of the GPD model (1.3.1), we use the function $R(\theta,\lambda)$ given by (4.3.1). By taking the logarithm of this function, we obtain

$$r(\theta,\lambda) = n\left[\ln\left[\frac{\theta}{\hat{\theta}}\right] - (\theta-\hat{\theta}) - \bar{x}(\lambda-\hat{\lambda})\right]$$

$$+ \sum_{i=1}^{n} (x_i-1)\ \ln\left[\frac{\theta+\lambda x_i}{\hat{\theta}+\hat{\lambda} x_i}\right]. \qquad (4.3.11)$$

The function $r(\theta,\lambda)$ is a bivariate function of θ and λ. Accordingly, contours can be drawn for each specific value of $r(\theta,\lambda)$. We compute the values of the function in (4.3.11) and use these to draw the contour lines corresponding to

$$r(\theta,\lambda) = \log \nu$$

in the (θ,λ) plane for different values of ν. Points inside the ν-contour line form the $100\nu\%$ likelihood region for the parameter θ and λ.

EXAMPLE 4.2 CONTINUED

The data in Table 4.1 was used to set likelihood region for the two parameters θ and λ.
By applying the data to the expression in (4.3.11), we get the following.

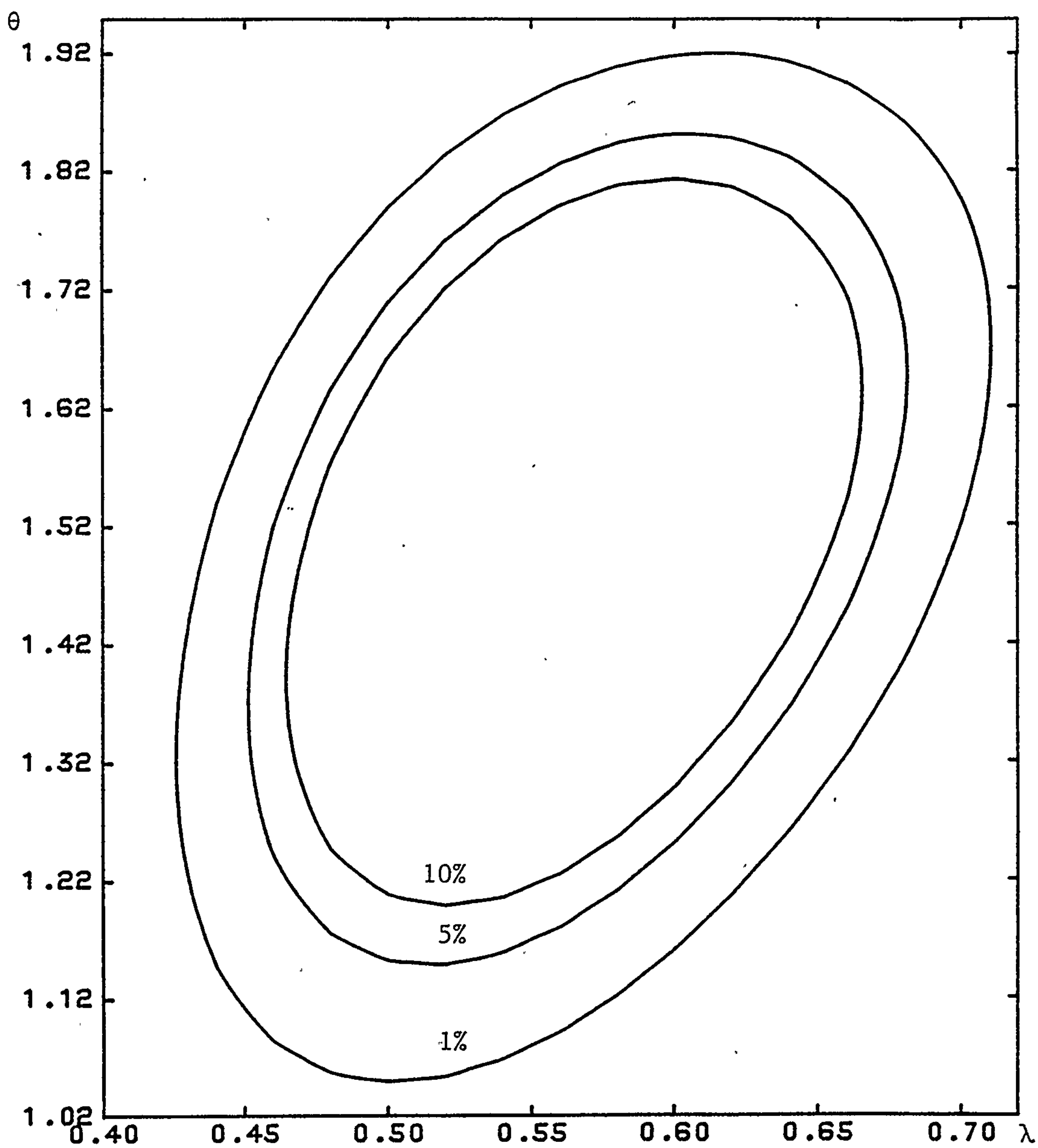
$$\begin{aligned} r(\theta,\lambda) = {} & 94\ \ell n\left[\frac{\theta}{1.4446}\right] - 122[\theta + \lambda - 3.2951] \\ & + 14\ \ell n\left[\frac{\theta+2\lambda}{2.5678}\right] + 22\ \ell n\left[\frac{\theta+3\lambda}{3.1294}\right] + 24\ \ell n\left[\frac{\theta+4\lambda}{3.6910}\right] \\ & + 44\ \ell n\left[\frac{\theta+5\lambda}{4.2526}\right] + 10\ \ell n\left[\frac{\theta+6\lambda}{4.8142}\right] + 18\ \ell n\left[\frac{\theta+7\lambda}{5.3758}\right] \\ & + 21\ \ell n\left[\frac{\theta+8\lambda}{5.9374}\right] + 24\ \ell n\left[\frac{\theta+9\lambda}{6.4990}\right] + 27\ \ell n\left[\frac{\theta+10\lambda}{7.0606}\right] \\ & + 20\ \ell n\left[\frac{\theta+11\lambda}{7.6222}\right] + 12\ \ell n\left[\frac{\theta+13\lambda}{8.7454}\right] + 26\ \ell n\left[\frac{\theta+14\lambda}{9.3070}\right] \\ & + 14\ \ell n\left[\frac{\theta+15\lambda}{9.8686}\right] + 32\ \ell n\left[\frac{\theta+17\lambda}{10.9918}\right]. \end{aligned} \tag{4.3.12}$$

Values of the expression $r(\theta,\lambda)$ in (4.3.12) were computed and these values were used to obtain the 10%, 5% and 1% likelihood regions. These are given in Figure 4.3.

Every one of the three contours in Figure 4.3 is roughly elliptical in form with its major axis towards the increasing values of θ and λ. It seems that there is a greater variation in the values of θ than in the values of λ. Values of (θ,λ) within the 5% likelihood contour are plausible and values outside this line are implausible. Also, values of (θ,λ) inside the 1% likelihood contour are very plausible. The 10%, 5% and 1% likelihood regions compare very well with the corresponding 90%, 95% and 99% confidence regions for the same example.

FIGURE 4.3

1, 5 AND 10 PERCENT LIKELIHOOD REGIONS



CHAPTER V

ESTIMATION IN LARGE SAMPLES FOR GENERALIZED POISSON DISTRIBUTION

5.1 INTRODUCTION

In the previous chapter, we considered the problem of estimation in small samples for the GPD model. In the present chapter we shall examine the same problem but in large samples.

We shall first derive some large sample properties of the maximum likelihood estimates (mle), the conditional maximum likelihood estimates, the likelihood ratio statistic and the conditional likelihood ratio statistic. As in Chapter IV, we assume that a random sample of size n is taken from a GPD model. Further, we assume that the parameter space for λ and φ in GPD models (1.3.1) and (1.3.2) are restricted to (0,1) and $(0, \theta^{-1})$ respectively.

5.2 LARGE SAMPLE PROPERTIES OF MLE AND LIKELIHOOD RATIO STATISTIC

Suppose the parameter θ is fixed at θ_0 in the GPD model (1.3.1). The likelihood function of a random sample of size n is given by (4.1.1) with θ replaced by θ_0. Denote $L(\theta_0, \lambda)$ by $L(\lambda)$.

Now,

$$\ln L(\lambda) = n\ln\theta_0 - \theta_0 n - \lambda \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \left[(x_i-1)\ln(\theta_0+\lambda x_i) - \ln(x_i!)\right]. \quad (5.2.1)$$

On differentiating (5.2.1) w.r.t. λ, we obtain

$$\frac{\partial \ln L(\lambda)}{\partial\lambda} = H(\lambda) = -\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{x_i(x_i-1)}{\theta_0+\lambda x_i} \quad (5.2.2)$$

and

$$\frac{\partial^2 \ln L(\lambda)}{\partial\lambda^2} = -\sum_{i=1}^{n} \frac{x_i^2(x_i-1)}{(\theta_0+\lambda x_i)^2}. \quad (5.2.3)$$

The mle $\hat{\lambda}$ of λ is obtained by solving equation

$$H(\lambda) = 0$$

i.e.

$$\sum_{i=1}^{n} \frac{x_i(x_i-1)}{\theta_0+\hat{\lambda} x_i} - \sum_{i=1}^{n} x_i = 0. \quad (5.2.4)$$

Since $\frac{\partial^2 \ln L(\lambda)}{\partial\lambda^2} = \frac{\partial H(\lambda)}{\partial\lambda}$ is less than zero, the function $H(\lambda)$ is a decreasing function of λ and hence it can have at most one solution.

But

$$H(0) = \sum_{i=1}^{n} \frac{x_i(x_i-1)}{\theta_0} - \sum_{i=1}^{n} x_i$$

$$= n\,(\theta_0)^{-1}[s^2 + \bar{x}^2 - \bar{x}(\theta_0+1)].$$

Also,

$$H(1) = \sum_{i=1}^{n} \frac{x_i(x_i-1)}{\theta_0+x_i} - \sum_{i=1}^{n} x_i$$

$$< -n.$$

Provided $H(0) > 0$, i.e. if $s^2 + \bar{x}^2 > \bar{x}(\theta_0+1)$, there exists a unique solution $\hat{\lambda}$ which satisfies (5.2.4).

$$E\left[\frac{\partial \ell n L(\lambda)}{\partial\lambda}\right] = E\left[\sum_{i=1}^{n} \frac{X_i(X_i-1)}{(\theta_0+\lambda X_i)} - \sum_{i=1}^{n} X_i\right]$$

$$= \sum_{i=1}^{n} \left\{E\left[\frac{X_i(X_i-1)}{\theta_0+\lambda X_i}\right] - E\,X_i\right\}$$

$$= n\left\{E\left[\frac{X(X-1)}{\theta_0+\lambda X}\right] - E\,X\right\}$$

since the X_i's are i.i.d. random variables. By using the probability function (1.3.1), we obtain

$$EX = \frac{\theta_0}{1-\lambda}$$

and

$$E\left[\frac{X(X-1)}{\theta_0+\lambda X}\right] = \sum_{x=0}^{\infty} \frac{x(x-1)}{\theta_0+\lambda x}\,\theta_0\,\frac{(\theta_0+\lambda x)^{x-1}}{x!}\,e^{-\theta_0-\lambda x}$$

$$= \sum_{x=2}^{\infty} \theta_0\,\frac{(\theta_0+\lambda x)^{x-2}}{(x-2)!}\,e^{-\theta_0-\lambda x}$$

$$= \sum_{x=0}^{\infty} \theta_0 \frac{(\theta_0+2\lambda+\lambda x)^x}{x!} e^{-\theta_0-2\lambda-\lambda x}$$

$$= \theta_0 \sum_{x=0}^{\infty} [\theta_0+2\lambda+\lambda x] \frac{(\theta_0+2\lambda+\lambda x)^{x-1}}{x!} e^{-\theta_0-2\lambda-\lambda x}$$

$$= \theta_0 \left\{1 + \frac{\lambda}{1-\lambda}\right\}$$

$$= \frac{\theta_0}{1-\lambda}.$$

Hence,

$$E\left[\frac{\partial \ell n L(\lambda)}{\partial \lambda}\right] = n\left\{\frac{\theta_0}{1-\lambda} - E\ X\right\}$$

$$= n\left\{\frac{\theta_0}{1-\lambda} - \frac{\theta_0}{1-\lambda}\right\}$$

i.e.

$$E\left[\frac{\partial \ell n L(\lambda)}{\partial \lambda}\right] = 0. \tag{5.2.5}$$

Also, we let

$$I^2(\lambda) = E\left[-\frac{\partial^2 \ell n L(\lambda)}{\partial \lambda^2}\right]$$

$$= \sum_{i=1}^{n} E\left[\frac{X_i^2(X_i-1)}{(\theta_0+\lambda X_i)^2}\right] > 0. \tag{5.2.6}$$

THEOREM 5.1. The mle $\hat{\lambda}$ is asymptotically normally distributed with variance $[I^2(\lambda)]^{-1}$. If λ_0 is the true parameter value of λ, the expression $(\hat{\lambda}-\lambda_0)\ I(\lambda_0)$ converges in distribution to the standard normal form.

PROOF:

Let $\left.\frac{\partial \ell nL(\lambda)}{\partial\lambda}\right|_{\lambda=\hat{\lambda}}$ be written as $\frac{\partial \ell nL(\hat{\lambda})}{\partial\lambda}$.

By Taylor's theorem, $\frac{\partial \ell nL(\hat{\lambda})}{\partial\lambda}$ can be expanded about the point λ_0 and written in the form

$$\frac{\partial \ell nL(\hat{\lambda})}{\partial\lambda} = \frac{\partial \ell nL(\lambda_0)}{\partial\lambda} + (\hat{\lambda}-\lambda_0)\ \frac{\partial^2 \ell nL(\lambda^*)}{\partial\lambda^2} \qquad (5.2.7)$$

where λ^* is a value between λ_0 and $\hat{\lambda}$ and thus $|\lambda^*-\hat{\lambda}| \le |\lambda_0-\hat{\lambda}|$. By using (5.2.2) and (5.2.3), we obtain from (5.2.7)

$$-\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{x_i(x_i-1)}{\theta_0+\hat{\lambda}x_i} = -\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{x_i(x_i-1)}{\theta_0+\lambda_0 x_i}$$

$$+ (\hat{\lambda}-\lambda_0)\left[-\sum_{i=1}^{n} \frac{x_i^2(x_i-1)}{(\theta_0+\lambda^* x_i)^2}\right]. \qquad (5.2.8)$$

By equation (5.2.4), the left hand side of (5.2.8) vanishes. Accordingly,

$$(\hat{\lambda}-\lambda_0)\left[-\sum_{i=1}^{n}\frac{x_i^2(x_i-1)}{(\theta_0+\lambda^* x_i)^2}\right] = -\left\{-\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\frac{x_i(x_i-1)}{\theta_0+\lambda_0 x_i}\right\}. \qquad (5.2.9)$$

From (5.2.9), we obtain

$$(\hat{\lambda}-\lambda_0)\ I(\lambda_0) = \frac{\left[-\sum\limits_{i=1}^{n} x_i+\sum\limits_{i=1}^{n}\frac{x_i(x_i-1)}{\theta_0+\lambda_0 x_i}\right]/I(\lambda_0)}{\left[-\sum\limits_{i=1}^{n}\frac{x_i^2(x_i-1)}{(\theta_0+\lambda^* x_i)^2}\right]/[-I^2(\lambda_0)]}. \qquad (5.2.10)$$

But the mle $\hat{\lambda}$ is consistent. That is, $\hat{\lambda}$ converges in probability to λ_0.

Therefore, as $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{x_i^2(x_i-1)}{(\theta_0+\hat{\lambda}x_i)^2} \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n}\frac{x_i^2(x_i-1)}{(\theta_0+\lambda_0 x_i)^2}. \qquad (5.2.11)$$

Now,

$$\frac{1}{n}\frac{\partial^2 \ell nL(\lambda)}{\partial\lambda^2} = -\frac{1}{n}\sum_{i=1}^{n}\frac{x_i^2(x_i-1)}{(\theta_0+\lambda x_i)^2}$$

is the mean of n i.i.d. random variables. By the strong law of large numbers

$$\frac{1}{n}\frac{\partial^2 \ell nL(\lambda)}{\partial\lambda^2}$$

will converge to its mean with probability 1. Thus, from (5.2.11), we obtain

$$\lim_{n\to\infty} P\left\{\sum_{i=1}^{n} \frac{X_i^2(X_i-1)}{(\theta_0+\hat{\lambda}X_i)^2} = \sum_{i=1}^{n} E\left[\frac{X_i^2(X_i-1)}{(\theta_0+\lambda_0X_i)^2}\right]\right\} = 1. \qquad (5.2.12)$$

Since $\hat{\lambda}$ converges in probability to λ_0 and λ^* lies between $\hat{\lambda}$ and λ_0, we must have that $\lambda^* \longrightarrow \lambda_0$ in probability.

Hence, from (5.2.12) we get

$$\lim_{n\to\infty} P\left\{\sum_{i=1}^{n} \frac{X_i^2(X_i-1)}{(\theta_0+\lambda^*X_i)^2} = \sum_{i=1}^{n} E\left[\frac{X_i^2(X_i-1)}{(\theta_0+\lambda_0X_i)^2}\right]\right\} = 1$$

which is the same as

$$\lim_{n\to\infty} P\left\{\sum_{i=1}^{n} \frac{X_i^2(X_i-1)}{(\theta_0+\lambda^*X_i)^2} = I^2(\lambda_0)\right\} = 1. \qquad (5.2.13)$$

Because of (5.2.13), the denominator of the right hand expression in (5.2.10) converges to 1 with probability 1.

The quantity in the numerator of the right hand expression in (5.2.10) is

$$H(\lambda_0) = -\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{x_i(x_i-1)}{\theta_0+\lambda_0x_i}.$$

According to the results (5.2.5) and (5.2.6) the quantity $H(\lambda_0)$ has mean zero and variance $I^2(\lambda_0)$. By applying the central limit theorem to $H(\lambda_0)/I(\lambda_0)$, we obtain a standard normal variate. Since the denominator of the right hand expression in (5.2.10) converges to 1 with probability 1, the whole right hand side expression converges to standard normal. Therefore, the left hand quantity in (5.2.10) is asymptotically standard normal.

Hence $\hat{\lambda}$ is asymptotically normal with mean λ_0 and variance $[I^2(\lambda_0)]^{-1}$. This completes the proof of theorem 5.1.

REMARK: If the parameter λ is initially fixed at λ_0, the mle $\hat{\theta}$ of θ can be similarly shown to be asymptotically normally distributed with mean θ_0 and variance $[I^2(\theta_0)]^{-1}$ where

$$I^2(\theta) = E\left[-\frac{\partial^2 \ell nL(\theta,\lambda_0)}{\partial\theta^2}\right].$$

COROLLAY 5.1: The conditional maximum likelihood estimate $\hat{\varphi}$ from the GPD model (1.3.2) is asymptotically normally distributed with variance $\left[I_c^2(\varphi)\right]^{-1}$. If φ_0 is the true value of φ, then $(\hat{\varphi}-\varphi_0)\ I_c(\varphi_0)$ converges in distribution to standard normal.

The proof of the above corollary follows from the proof of Theorem 5.1. If $L(\varphi)$ is the conditional likelihood function, then

$$I_c^2(\varphi) = E\left[-\frac{\partial^2 \ell nL(\varphi)}{\partial\varphi^2}\right] > 0.$$

By differentiating (5.2.3) one more time, we obtain

$$\frac{\partial^3 \ell nL(\lambda)}{\partial\lambda^3} = \sum_{i=1}^{n} \frac{2x_i^3(x_i-1)}{(\theta_0+\lambda x_i)^3}. \qquad (5.2.14)$$

But

$$\left|\frac{\partial^3 \ell n L(\lambda)}{\partial \lambda^3}\right| = \sum_{i=1}^{n} \frac{2x_i^3(x_i-1)}{(\theta_0+\lambda x_i)^3}$$

$$= \sum_{i=1}^{n} \left[\frac{x_i}{\theta_0+\lambda x_i}\right]^3 \cdot 2(x_i-1)$$

$$= \sum_{i=1}^{n} \left[\frac{\lambda x_i}{\theta_0+\lambda x_i}\right]^3 \cdot \frac{2(x_i-1)}{\lambda^3}$$

$$\leq \sum_{i=1}^{n} \frac{2x_i}{\lambda^3} = h(x), \qquad \text{(say)}. \qquad (5.2.15)$$

From above, $E[h(X)]$ and $\text{var}[h(X)]$ are finite.

THEOREM 5.2. If $\lambda = \lambda_0$ is the true value of the parameter λ, the likelihood ratio statistic $\ell(\lambda_0)$ is such that $-2\ell n\, \ell(\lambda_0)$ converges in distribution to a chi-square distributed random variable with 1 d.f. as $n \to \infty$.

PROOF:

The likelihood ratio is given by

$$\ell(\lambda_0) = \frac{L(\lambda_0)}{L(\hat{\lambda})}$$

$$= \frac{e^{-\lambda_0 \sum^n x_i} \prod_{i=1}^{n} (\theta_0+\lambda_0\, x_i)^{x_i-1}}{e^{-\hat{\lambda} \sum^n x_i} \prod_{i=1}^{n} (\theta_0+\hat{\lambda}\, x_i)^{x_i-1}}$$

$$= e^{(\hat{\lambda}-\lambda_0)\sum^n x_i} \prod_{i=1}^{n} \left[\frac{\theta_0+\lambda_0\, x_i}{\theta_0+\hat{\lambda}\, x_i}\right]^{x_i-1} . \qquad (5.2.16)$$

From the above, we get

$$-2\ell n\, \ell(\lambda_0) = -2\ell n\, L(\lambda_0) + 2\ell n\, L(\hat{\lambda}).$$

We use the Taylor's theorem to expand $-2\ell n\, \ell(\lambda_0)$ about $\hat{\lambda}$.

Thus, we obtain

$$-2\ell n\, \ell(\lambda_0) = -2\ell n\, L(\hat{\lambda}) + (\lambda_0-\hat{\lambda}) \frac{\partial}{\partial\lambda} [-2\ell n\, L(\lambda)]_{\lambda=\hat{\lambda}}$$

$$+ \frac{(\lambda_0-\hat{\lambda})^2}{2!} \frac{\partial^2}{\partial\lambda^2} [-2\ell n\, L(\lambda)]_{\lambda=\hat{\lambda}}$$

$$+ \frac{(\lambda_0-\hat{\lambda})^3}{3!} \frac{\partial^3}{\partial\lambda^3} [-2\ell n\, L(\lambda)]_{\lambda=\lambda^*}$$

$$+ 2\ell n\, L(\hat{\lambda})$$

$$= -\, 2(\lambda_0-\hat{\lambda}) \frac{\partial \ell n L(\hat{\lambda})}{\partial\lambda} - (\lambda_0-\hat{\lambda})^2 \frac{\partial^2 \ell n L(\hat{\lambda})}{\partial\lambda^2}$$

$$- \frac{1}{3} (\lambda_0-\hat{\lambda})^3 \frac{\partial^3 \ell n L(\lambda^*)}{\partial\lambda^3}$$

where λ^* lies between $\hat{\lambda}$ and λ_0. By using (5.2.4), the first term on the right hand side vanishes and we have

$$-2\ln \ell(\lambda_0) = -(\lambda_0-\hat{\lambda})^2 \frac{\partial^2 \ln L(\hat{\lambda})}{\partial\lambda^2} - \frac{1}{3}(\lambda_0-\hat{\lambda})^3 \frac{\partial^3 \ln L(\lambda^*)}{\partial\lambda^3}.$$

By expanding $\dfrac{\partial^2 \ln L(\hat{\lambda})}{\partial\lambda^2}$ about the true value λ_0 again, we obtain

$$\frac{\partial^2 \ln L(\hat{\lambda})}{\partial\lambda^2} = \frac{\partial^2 \ln L(\lambda_0)}{\partial\lambda^2} + (\hat{\lambda}-\lambda_0)\frac{\partial^3 \ln L(\lambda^{**})}{\partial\lambda^3}$$

where λ^{**} lies between $\hat{\lambda}$ and λ_0.

Now,

$$-2\ln \ell(\lambda_0) = -(\lambda_0-\hat{\lambda})^2 \left\{\frac{\partial^2 \ln L(\lambda_0)}{\partial\lambda^2} + (\hat{\lambda}-\lambda_0)\frac{\partial^3 \ln L(\lambda^{**})}{\partial\lambda^3}\right\}$$

$$- \frac{1}{3}(\lambda_0-\hat{\lambda})^3 \frac{\partial^3 \ln L(\lambda^*)}{\partial\lambda^3}$$

$$= (\hat{\lambda}-\lambda_0)^2 \left\{-\frac{\partial^2 \ln L(\lambda_0)}{\partial\lambda^2} + (\lambda_0-\hat{\lambda})\frac{\partial^3 \ln L(\lambda^{**})}{\partial\lambda^3}\right.$$

$$\left. - \frac{1}{3}(\lambda_0-\hat{\lambda})\frac{\partial^3 \ln L(\lambda^*)}{\partial\lambda^3}\right\}$$

$$= (\hat{\lambda}-\lambda_0)^2\, I^2(\lambda_0) \left\{-\frac{\partial^2 \ln L(\lambda_0)}{\partial\lambda^2}[I^2(\lambda_0)]^{-1}\right.$$

$$\left. + (\lambda_0-\hat{\lambda})[I^2(\lambda_0)]^{-1}\left[\frac{\partial^3 \ln L(\lambda^{**})}{\partial\lambda^3} - \frac{1}{3}\frac{\partial^3 \ln L(\lambda^*)}{\partial\lambda^3}\right]\right\}.$$

From (5.2.12), we obtain

$$\lim_{n\to\infty} P\left\{\frac{\partial^2 \ln L(\lambda_0)}{\partial\lambda^2} = -I^2(\lambda_0)\right\} = 1$$

and so
$$-\frac{\partial^2 \ln L(\lambda_0)}{\partial\lambda^2}\,[I^2(\lambda_0)]^{-1}$$
converges to 1 with probability 1.

But $\hat{\lambda} \longrightarrow \lambda_0$ in probability, so we have that $\lambda^* \longrightarrow \lambda_0$ and $\lambda^{**} \longrightarrow \lambda_0$ in probability as well. By using (5.2.15),

$$-\frac{\partial^3 \ln L(\lambda^*)}{\partial\lambda^3}\,[I^2(\lambda_0)]^{-1}$$
is bounded with probability 1 for large n. Since $\hat{\lambda} \longrightarrow \lambda_0$ in probability, we have that

$$(\lambda_0-\hat{\lambda})\,[I^2(\lambda_0)]^{-1}\left[\frac{\partial^3 \ln L(\lambda^{**})}{\partial\lambda^3} - \frac{1}{3}\,\frac{\partial^3 \ln L(\lambda^*)}{\partial\lambda^3}\right] \longrightarrow 0$$

in probability as $n \longrightarrow \infty$.

Hence, $-2\ln \ell(\lambda_0)$ converges to the distribution of the random variable $(\hat{\lambda}-\lambda_0)^2\, I^2(\lambda_0)$. From Theorem 5.1, $(\hat{\lambda}-\lambda_0)\, I(\lambda_0)$ has an asymptotic standard normal distribution. Therefore,

$$(\hat{\lambda}-\lambda_0)^2\, I^2(\lambda_0)$$

is a chi-square distributed random variable with 1 d.f.. Thus

$-2\ln \ell(\lambda_0)$ has an asymptotic chi-square distribution with 1 d.f..

COROLLARY 5.2: The conditional likelihood ratio statistic $\ell_c(\varphi_0)$ is such that $-2\ln \ell_c(\varphi_0)$ has an approximate chi-square distribution with 1 d.f..

The proof of this corollary follows from the proof of theorem 5.2 and will not be given here.

5.3 CONFIDENCE INTERVALS FOR A SINGLE PARAMETER

The mean and variance of the GPD model (1.3.1) are given by

$$\text{mean} = \theta(1-\lambda)^{-1} \qquad \text{and} \qquad \text{var} = \theta(1-\lambda)^{-3}.$$

The corresponding values for the GPD model (1.3.2) are $\theta(1-\theta\varphi)^{-1}$ and $\theta(1-\theta\varphi)^{-3}$.

By using the central limit theorem as in section 3.3, the distribution of the sample sum Y, which is a GP variate, converges stochastically to that of normal distribution. By following the explanation in section 3.3,

$$Z = \frac{\sqrt{n}(\bar{X}-\mu)}{S}$$

has a limiting distribution which is normal with mean zero and variance unity.

5.3.1 ESTIMATION FOR θ OR λ IN GPD MODEL (1.3.1):

When the parameter θ is known and fixed at θ_0, the approximate 100(1-α)% C.I. for λ can be obtained as follows:

$$1-\alpha = P\left\{- z_{\alpha/2} < \frac{\sqrt{n}(\bar{X}-\mu)}{S} < z_{\alpha/2}\right\}$$

$$= P\left\{\bar{X} - z_{\alpha/2}\, S/\sqrt{n} < \mu < \bar{X} + z_{\alpha/2}\, S/\sqrt{n}\right\}$$

$$= P\left\{\bar{X} - z_{\alpha/2}\, S/\sqrt{n} < \frac{\theta_0}{1-\lambda} < \bar{X} + z_{\alpha/2}\, S/\sqrt{n}\right\}$$

$$= P\left\{1 - \frac{\theta_0}{\bar{X}-z_{\alpha/2}S/\sqrt{n}} < \lambda < 1 - \frac{\theta_0}{\bar{X}+z_{\alpha/2}S/\sqrt{n}}\right\}.$$

Thus, the approximate 100(1-α)% C.I. for the parameter λ, when θ is known to be equal to θ_0, is

$$\left[1 - \frac{\theta_0}{\bar{x}-z_{\alpha/2}s/\sqrt{n}}, \quad 1 - \frac{\theta_0}{\bar{x}+z_{\alpha/2}s/\sqrt{n}}\right]. \tag{5.3.1}$$

In a similar way, we assume that λ is known to be λ_0. The standardized normal tables can be used to get $z_{\alpha/2}$ and then the approximate 100(1-α)% C.I. for θ can be obtained as follows:

$$1-\alpha = P\left\{\bar{X} - z_{\alpha/2}\, S/\sqrt{n} < \frac{\theta}{1-\lambda_0} < \bar{X} + z_{\alpha/2}\, S/\sqrt{n}\right\}$$

$$= P\left\{(1-\lambda_0)\ [\bar{X} - z_{\alpha/2}\ S/\sqrt{n}] < \theta < (1-\lambda_0)\ [\bar{X} + z_{\alpha/2}\ S/\sqrt{n}]\right\}.$$

Hence, the C.I. for the parameter θ becomes

$$\left[(1-\lambda_0)\ [\bar{x} - z_{\alpha/2}\ s/\sqrt{n}]\ ,\ (1-\lambda_0)\ [\bar{x} + z_{\alpha/2}\ s/\sqrt{n}]\right]. \quad (5.3.2)$$

5.3.2 **ESTIMATION FOR φ IN RESTRICTED GPD MODEL (1.3.2):**

The case of parameter θ was considered in section 3.3 and will not be repeated here. Suppose θ is known to be θ_0, the approximate $100(1-\alpha)\%$ C.I. for φ can be obtained from

$$1-\alpha = P\left\{\bar{X} - z_{\alpha/2}\ S/\sqrt{n} < \frac{\theta_0}{1-\varphi\theta_0} < \bar{X} + z_{\alpha/2}\ S/\sqrt{n}\right\}$$

$$= P\left\{\theta_0^{-1} - [\bar{X} - z_{\alpha/2}\ S/\sqrt{n}]^{-1} < \varphi < \theta_0^{-1} - [\bar{X} + z_{\alpha/2}\ S/\sqrt{n}]^{-1}\right\}.$$

Thus, the approximate C.I. for the parameter φ is

$$\left[\theta_0^{-1} - (\bar{x} - z_{\alpha/2}\ s/\sqrt{n})^{-1}\ ,\ \theta_0^{-1} - (\bar{x} + z_{\alpha/2}\ s/\sqrt{n})^{-1}\right]. \quad (5.3.3)$$

5.4 **CONFIDENCE INTERVALS WHEN BOTH PARAMETERS ARE UNKNOWN**

Since both parameters are unknown, the parameter which is not of interest becomes a nuisance parameter and will have to be eliminated before inference can be made on the parameter of real interest. Among various methods of parameter elimination, we consider the methods of 'maximization of likelihood' and of 'conditioning of likelihood'.

5.4.1 METHOD OF MAXIMIZATION

The likelihood function of the GPD model (1.3.1) is given by (4.1.1). If $\hat{\theta}$ and $\hat{\lambda}$ are the ML estimates of θ and λ respectively, the likelihood ratio function becomes

$$\ell(\theta,\lambda) = \frac{L(\theta,\lambda)}{L(\hat{\theta},\hat{\lambda})}$$

$$= \frac{\theta^n e^{-n(\theta+\lambda\bar{x})} \prod_{i=1}^{n} [(\theta+\lambda x_i)^{x_i-1}/(x_i)!]}{\hat{\theta}^n e^{-n(\hat{\theta}+\hat{\lambda}\bar{x})} \prod_{i=1}^{n} [(\hat{\theta}+\hat{\lambda} x_i)^{x_i-1}/(x_i)!]}$$

$$= \left[\frac{\theta}{\hat{\theta}}\right]^n e^{n[\hat{\theta}-\theta+(\hat{\lambda}-\lambda)\bar{x}]} \prod_{i=1}^{n} \left[\frac{\theta+\lambda x_i}{\hat{\theta}+\hat{\lambda} x_i}\right]^{x_i-1} . \quad (5.4.1)$$

Suppose the parameter of interest is θ. By following the technique in section 4.3 of chapter IV under the method of maximization, the maximum likelihood ratio function is given by

$$\ell_m(\theta) = \ell(\theta, \tilde{\lambda}(\theta))$$

$$= \left[\frac{\theta}{\hat{\theta}}\right]^n e^{-n[\theta-\hat{\theta}+\bar{x}(\tilde{\lambda}(\theta)-\hat{\lambda})]} \prod_{i=1}^{n} \left[\frac{\theta+\tilde{\lambda}(\theta)x_i}{\hat{\theta}+\hat{\lambda}\, x_i}\right]^{x_i-1} . \qquad (5.4.2)$$

According to Theorem 5.2 when the sample size n is large, for a specified value $\theta = \theta_0$, the statistic

$$T_m(\theta_0) = -2\ell n\ \ell_m(\theta_0)$$

is approximately distributed as a chi-square random variable with 1 degree of freedom. Now,

$$T_m(\theta) = -2n\left[\ell n \left[\frac{\theta}{\hat{\theta}}\right] - (\theta-\hat{\theta}) - \bar{x}(\tilde{\lambda}(\theta)-\hat{\lambda})\right]$$

$$- 2 \sum_{i=1}^{n} (x_i-1)\ \ell n\left[\frac{\theta+\tilde{\lambda}(\theta)x_i}{\hat{\theta}+\hat{\lambda}x_i}\right]. \qquad (5.4.3)$$

The graph of $T_m(\theta)$ against θ is concave up (i.e. U-shaped). To set $100(1-\alpha)\%$ C.I. for θ, we find two values θ_ℓ and θ_u of θ at which the straight line

$$T_m(\theta) = \chi^2_{\alpha,1}$$

intersects the graph of the function $T_m(\theta)$ in (5.4.3). It may be noted that while drawing the graph of (5.4.3), the value of $\tilde{\lambda}(\theta)$ will have to be computed first from equation (4.3.6) for each value of θ.

The values θ_ℓ and θ_u are respectively the lower and upper confidence bounds for θ and $\chi^2_{\alpha,1}$ denotes the upper percent point of the chi-square distribution with 1 d.f..

Since the domain of the parameter θ is unbounded on the right hand side, the tracing of the graph of the function in (5.4.3) by the computer is somewhat tricky. Whenever the GPD model is used to describe a natural phenomenon it has been observed that the values of θ generally lie below 15.0. Thus, $T_m(\theta)$ can, in general, be drawn for values of $\theta \le 20.0$.

An advantage of the method of maximization is that the parameters θ and λ can be easily interchanged in the elimination procedure.

To set C.I. for the parameter λ, the above procedure is repeated with the roles of λ and θ interchanged. In this case, we obtain

$$\ell_m(\lambda) = \ell(\tilde{\theta}(\lambda), \lambda)$$

$$= \left[\frac{\tilde{\theta}(\lambda)}{\hat{\theta}}\right]^n e^{-[\tilde{\theta}(\lambda)-\hat{\theta}+\bar{x}(\lambda-\hat{\lambda})]n} \prod_{i=1}^{n} \left[\frac{\tilde{\theta}(\lambda)+\lambda x_i}{\hat{\theta}+\hat{\lambda}x_i}\right]^{x_i-1} \qquad (5.4.4)$$

where $\tilde{\theta}(\lambda)$, the solution of equation (4.3.3), is the value of θ that maximizes $L(\theta,\lambda)$ with a specified value of λ.

An analogue of relation (5.4.3) becomes

$$T_m(\lambda) = -2n\left[\ell n\left[\frac{\tilde{\theta}(\lambda)}{\hat{\theta}}\right] - (\tilde{\theta}(\lambda)-\hat{\theta}) - \bar{x}(\lambda-\hat{\lambda})\right]$$

$$-2\sum_{i=1}^{n}(x_i-1)\ \ell n\left[\frac{\tilde{\theta}(\lambda)+\lambda x_i}{\hat{\theta}+\hat{\lambda} x_i}\right]. \qquad (5.4.5)$$

So, the approximate $100(1-\alpha)\%$ C.I. for λ is obtained from the intersection of the graph of $T_m(\lambda)$ in (5.4.5) with the line

$$T_m(\lambda) = \chi^2_{\alpha,1}.$$

5.4.2 **METHOD OF CONDITIONING**

This method of nuisance parameter elimination is applicable if a sufficient statistic can be found for the nuisance parameter. Since $Y = \sum^{n} X_i$ is a sufficient statistic for θ in the GPD model (1.3.2), the likelihood function can be factorized as follows:

$$L(\theta,\theta\gamma) = L_1(\gamma)\ P_y(\theta,\ \theta\gamma)$$

where $P_y(\theta,\ \theta\gamma)$ is given by (3.2.3). Hence,

$$L_1(\gamma) = \frac{L(\theta,\theta\gamma)}{P_y(\theta,\theta\gamma)}$$

$$= \frac{\theta^y e^{-\theta(n+\varphi y)} \prod_{i=1}^{n} \left[(1+\varphi\, x_i)^{x_i-1} / x_i!\right]}{n(n+\varphi y)^{y-1}\ \theta^y e^{-\theta(n+\varphi y)}/y!}$$

$$= \frac{y!}{n(n+\varphi y)^{y-1}} \prod_{i=1}^{n} \left[\frac{(1+\varphi\, x_i)^{x_i-1}}{x_i!}\right]. \tag{5.4.6}$$

By logarithmic differentiation of (5.4.6) with respect to φ and by equating to zero, we have

$$0 = \frac{d}{d\varphi} \ln L_1(\varphi)$$

$$= \frac{d}{d\varphi} \left\{ \ln\left[\frac{y!}{n}\right] - (y-1)\ \ln(n+\varphi y) + \sum_{i=1}^{n} \left[(x_i-1)\ \ln(1+\varphi x_i) - \ln(x_i!)\right]\right\}$$

$$= \frac{-y(y-1)}{n+\varphi y} + \sum_{i=1}^{n} \frac{x_i(x_i-1)}{1+\varphi x_i}.$$

Therefore,

$$\sum_{i=1}^{n} \frac{x_i(x_i-1)}{1+\varphi x_i} - \frac{\bar{x}(n\bar{x}-1)}{1+\varphi\bar{x}} = 0. \tag{5.4.7}$$

We have not been able to prove the uniqueness of the solution $\hat{\varphi}$ that satisfies (5.4.7). However, by using the moment estimate of φ in the GPD model (1.3.2) as the initial value of φ, the iterative procedure to solve (5.4.7) takes very few steps (about 5 steps) before convergence.

We consider the conditional likelihood function in (5.4.6) and form the conditional likelihood ratio function

$$\ell_c(\varphi) = \frac{L_1(\varphi)}{L_1(\hat{\varphi})}$$

$$= \frac{\dfrac{y!}{n(n+\varphi y)^{y-1}} \prod_{i=1}^{n} \left[(1+\varphi x_i)^{x_i-1}/x_i!\right]}{\dfrac{y!}{n(n+\hat{\varphi} y)^{y-1}} \prod_{i=1}^{n} \left[(1+\hat{\varphi} x_i)^{x_i-1}/x_i!\right]}$$

$$= \left[\frac{1+\hat{\varphi}\bar{x}}{1+\varphi\bar{x}}\right]^{n\bar{x}-1} \prod_{i=1}^{n} \left[\frac{1+\varphi x_i}{1+\hat{\varphi} x_i}\right]^{x_i-1} . \qquad (5.4.8)$$

By using Corollary 5.2, $-2 \times$ the logarithm of conditional likelihood ratio statistic has an approximate chi-square distribution with 1 d.f. when the sample size n is large. Therefore, for a specified value $\varphi = \varphi_0$ and for a large sample, the statistic

$$T_c(\varphi_0) = -2\ln \ell_c(\varphi_0)$$

is approximately chi-square distributed with 1 d.f.. Hence,

$$T_c(\varphi) = -2\left\{(n\bar{x}-1)\ \ln\left[\frac{1+\hat{\varphi}\bar{x}}{1+\varphi\bar{x}}\right] + \sum_{i=1}^{n} (x_i-1)\ \ln\left[\frac{1+\varphi x_i}{1+\hat{\varphi} x_i}\right]\right\} \qquad (5.4.9)$$

is similar to $T_m(\theta)$ in (5.4.3) and so (5.4.9) is used in the same way as $T_m(\hat{\theta})$ to obtain the approximate $100(1-\alpha)\%$ C.I. for the parameter φ.

Since the above method requires the presence of a sufficient statistic for the nuisance parameter, it can only be used to eliminate θ in the GPD model (1.3.2). A number of GPD pseudo-random samples of sizes 200, 500 and 1,000 were generated, the methods of conditioning and of maximization were applied to eliminate θ in the model (1.3.2) and C.Is. for ν were obtained from these samples. It was observed that there is not much difference in C.Is. obtained by the two methods. This was due to the fact that the sample size was large. A notable difference was observed for corresponding analyses in small samples.

EXAMPLE 5.1

The data in the following table is the distribution obtained from Skellam (1952) of the number of plants of "Plantago major" present in quadrants of area 100 sq-cm laid down in grassland.

TABLE 5.1

The distribution of plants (Plantago major)

Plants per quadrant	0	1	2	3	4	5
Observed frequency	235	81	43	18	9	6
Expected GPD freq.	233.08	87.26	38.50	18.78	9.78	5.32

Plants per quadrant	6	7	8	9	Total
Observed frequency	4	3	0	1	400
Expected GPD freq.	2.99	1.72	1.01	1.56	400.00

$$\bar{x} = 0.8525, \qquad s^2 = 1.9807$$

We used the ML estimation method to fit the GPD model to this data and found that the observations are described very well by the GPD model. The ML estimates of the parameters θ and λ in the GPD model (1.3.1) are 0.540090 and 0.366463 respectively.

The method of maximization is used in order to set C.I. for each of the parameters θ and λ in the GPD model (1.3.1). We use the data in Table 5.1 in equation (4.3.6) to obtain

$$\begin{aligned} 0 &= \frac{2(43)}{\theta+2\lambda} + \frac{6(18)}{\theta+3\lambda} + \frac{12(9)}{\theta+4\lambda} + \frac{20(6)}{\theta+5\lambda} \\ &\quad + \frac{30(4)}{\theta+6\lambda} + \frac{42(3)}{\theta+7\lambda} + \frac{72}{\theta+9\lambda} - 400(.8525) \\ &= \frac{86}{\theta+2\lambda} + \frac{108}{\theta+3\lambda} + \frac{108}{\theta+4\lambda} + \frac{120}{\theta+5\lambda} \\ &\quad + \frac{120}{\theta+6\lambda} + \frac{126}{\theta+7\lambda} + \frac{72}{\theta+9\lambda} - 341. \end{aligned} \tag{5.4.10}$$

Equation (5.4.10) is used for finding $\tilde{\lambda}(\theta)$ for each value of θ. In the same way, we substitute the data into (5.4.3) to get

$$\begin{aligned} T_m(\theta) &= -800\left\{\ell n\left[\frac{\theta}{.5401}\right] - \theta + .5401 - .8525(\tilde{\lambda}(\theta) - .3665)\right\} \\ &\quad - 2\left\{-235\ \ell n\left[\frac{\theta}{.5401}\right] + 43\ \ell n\left[\frac{\theta + 2\tilde{\lambda}(\theta)}{.5401+2(.3665)}\right]\right. \\ &\quad + 36\ \ell n\left[\frac{\theta + 3\tilde{\lambda}(\theta)}{.5401+3(.3665)}\right] + 27\ \ell n\left[\frac{\theta + 4\tilde{\lambda}(\theta)}{.5401+4(.3665)}\right] \end{aligned}$$

$$+ 24 \ln\left[\frac{\theta + 5\tilde{\lambda}(\theta)}{.5401+5(.3665)}\right] + 20 \ln\left[\frac{\theta + 6\tilde{\lambda}(\theta)}{.5401+6(.3665)}\right]$$

$$+ 18 \ln\left[\frac{\theta + 7\tilde{\lambda}(\theta)}{.5401+7(.3665)}\right] + 8 \ln\left[\frac{\theta + 9\tilde{\lambda}(\theta)}{.5401+9(.3665)}\right]\Bigg\}$$

$$= - 330 \ln\left[\frac{\theta}{.5401}\right] + 800[\theta + .8525 \tilde{\lambda}(\theta) - .8525]$$

$$- 2 \Bigg\{43 \ln\left[\frac{\theta + 2\tilde{\lambda}(\theta)}{1.2731}\right] + 36 \ln\left[\frac{\theta + 3\tilde{\lambda}(\theta)}{1.6396}\right]$$

$$+ 27 \ln\left[\frac{\theta + 4\tilde{\lambda}(\theta)}{2.0061}\right] + 24 \ln\left[\frac{\theta + 5\tilde{\lambda}(\theta)}{2.3726}\right]$$

$$+ 20 \ln\left[\frac{\theta + 6\tilde{\lambda}(\theta)}{2.7391}\right] + 18 \ln\left[\frac{\theta + 7\tilde{\lambda}(\theta)}{3.1056}\right]$$

$$+ 8 \ln\left[\frac{\theta + 9\tilde{\lambda}(\theta)}{3.8386}\right]\Bigg\}. \tag{5.4.11}$$

By using the same data in equation (4.3.3), we obtain

$$0 = \frac{165}{\theta} + \frac{43}{\theta+2\lambda} + \frac{36}{\theta+3\lambda} + \frac{27}{\theta+4\lambda} + \frac{24}{\theta+5\lambda}$$

$$+ \frac{20}{\theta+6\lambda} + \frac{18}{\theta+7\lambda} + \frac{8}{\theta+9\lambda} - 400. \tag{5.4.12}$$

Equation (5.4.12) is used for finding $\tilde{\theta}(\lambda)$ for each value of λ. In the same manner, by substituting the data into (5.4.5) we obtain

$$T_m(\lambda) = - 330 \ln\left[\frac{\tilde{\theta}(\lambda)}{.5401}\right] + 800 \, [\tilde{\theta}(\lambda) + .8525\lambda - .8525]$$

$$- 2\left\{43 \ \ell n\left[\frac{\tilde{\theta}(\lambda) + 2\lambda}{1.2731}\right] + 36 \ \ell n\left[\frac{\tilde{\theta}(\lambda) + 3\lambda}{1.6396}\right]\right.$$

$$+ 27 \ \ell n\left[\frac{\tilde{\theta}(\lambda) + 4\lambda}{2.0061}\right] + 24 \ \ell n\left[\frac{\tilde{\theta}(\lambda) + 5\lambda}{2.3726}\right] + 20 \ \ell n\left[\frac{\tilde{\theta}(\lambda) + 6\lambda}{2.7391}\right]$$

$$\left. + 18 \ \ell n\left[\frac{\tilde{\theta}(\lambda) + 7\lambda}{3.1056}\right] + 8 \ \ell n\left[\frac{\tilde{\theta}(\lambda) + 9\lambda}{3.8386}\right]\right\}. \qquad (5.4.13)$$

We use the expression in (5.4.11) to draw the graph of $T_m(\theta)$. For each value of θ, we first obtain $\tilde{\lambda}(\theta)$ from equation (5.4.10). Both equation (5.4.10) and the expression (5.4.11) are evaluated by the use of a computer programme. By doing these, we obtain the 95% C.I. for the parameter θ as (0.46, 0.63). Similarly, we use the graph of the function of λ in (5.4.13) to obtain the 95% C.I. for λ. For each specified value of λ, we use equation (5.4.12) to obtain $\tilde{\theta}(\lambda)$ which is subsequently used to evaluate the expression (5.4.13). These values are used to draw the U-shaped graph and the 95% C.I. for λ is found to be (0.29, 0.45). The graphs of the functions of θ and λ are shown in Figure 5.1 and Figure 5.2 respectively. The graphs also indicate the 90% and 99% C.Is.

FIGURE 5.1

90, 95 and 99% confidence intervals for θ

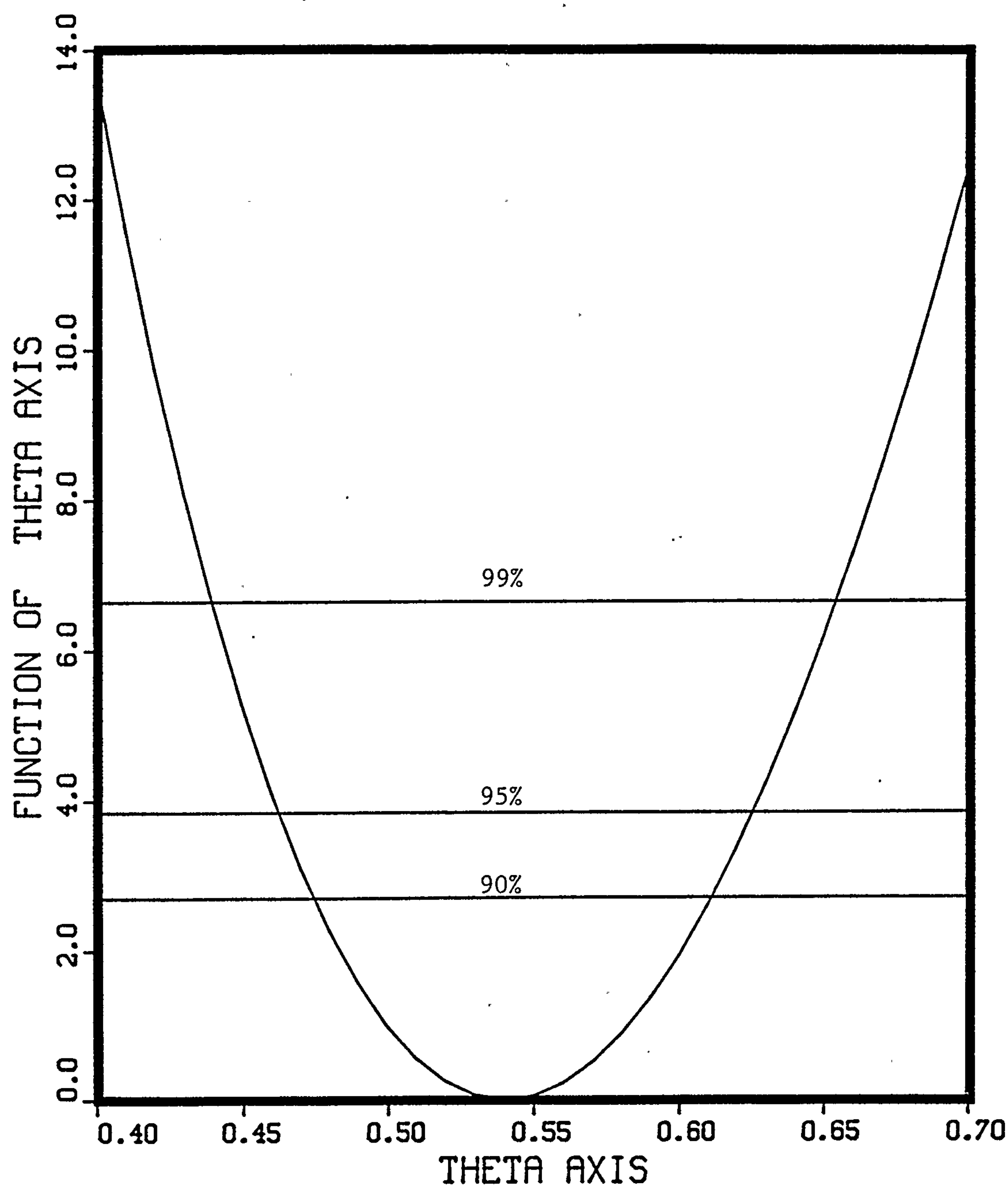
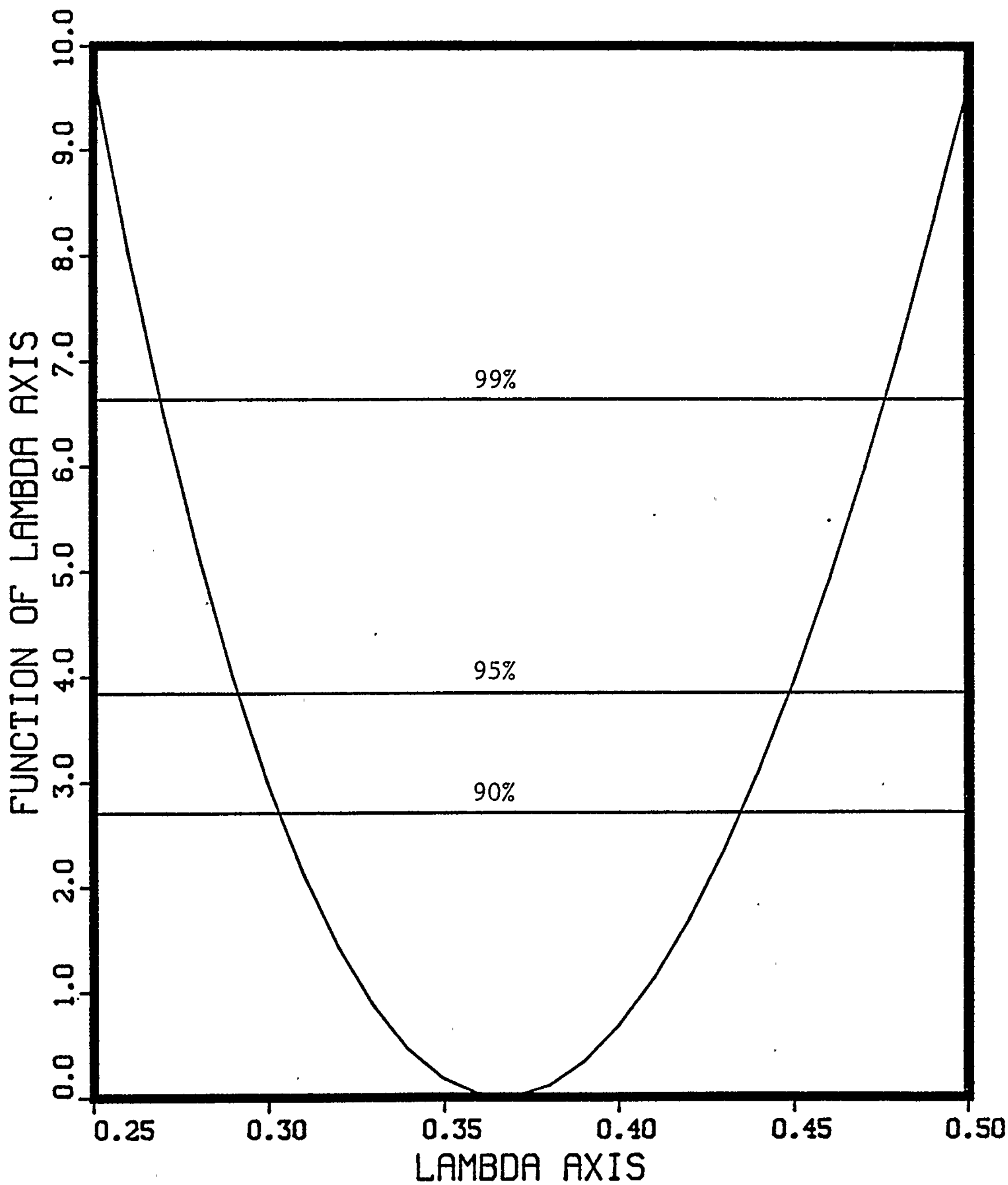


FIGURE 5.2

90, 95 and 99% confidence intervals for λ



The method of conditioning is also used in order to obtain C.I. for γ in the GPD model (1.3.2).

We substitute the data in equation (5.4.7) to get

$$0 = \frac{2(43)}{1+2\gamma} + \frac{6(18)}{1+3\gamma} + \frac{12(9)}{1+4\gamma} + \frac{20(6)}{1+5\gamma} + \frac{30(4)}{1+6\gamma} + \frac{42(3)}{1+7\gamma} + \frac{72}{1+9\gamma} - \frac{.8525(340)}{1+.8525\gamma}$$

$$= \frac{86}{1+2\gamma} + \frac{108}{1+3\gamma} + \frac{108}{1+4\gamma} + \frac{120}{1+5\gamma} + \frac{120}{1+6\gamma} + \frac{126}{1+7\gamma} + \frac{72}{1+9\gamma} - \frac{289.85}{1+.8525\gamma}. \tag{5.4.14}$$

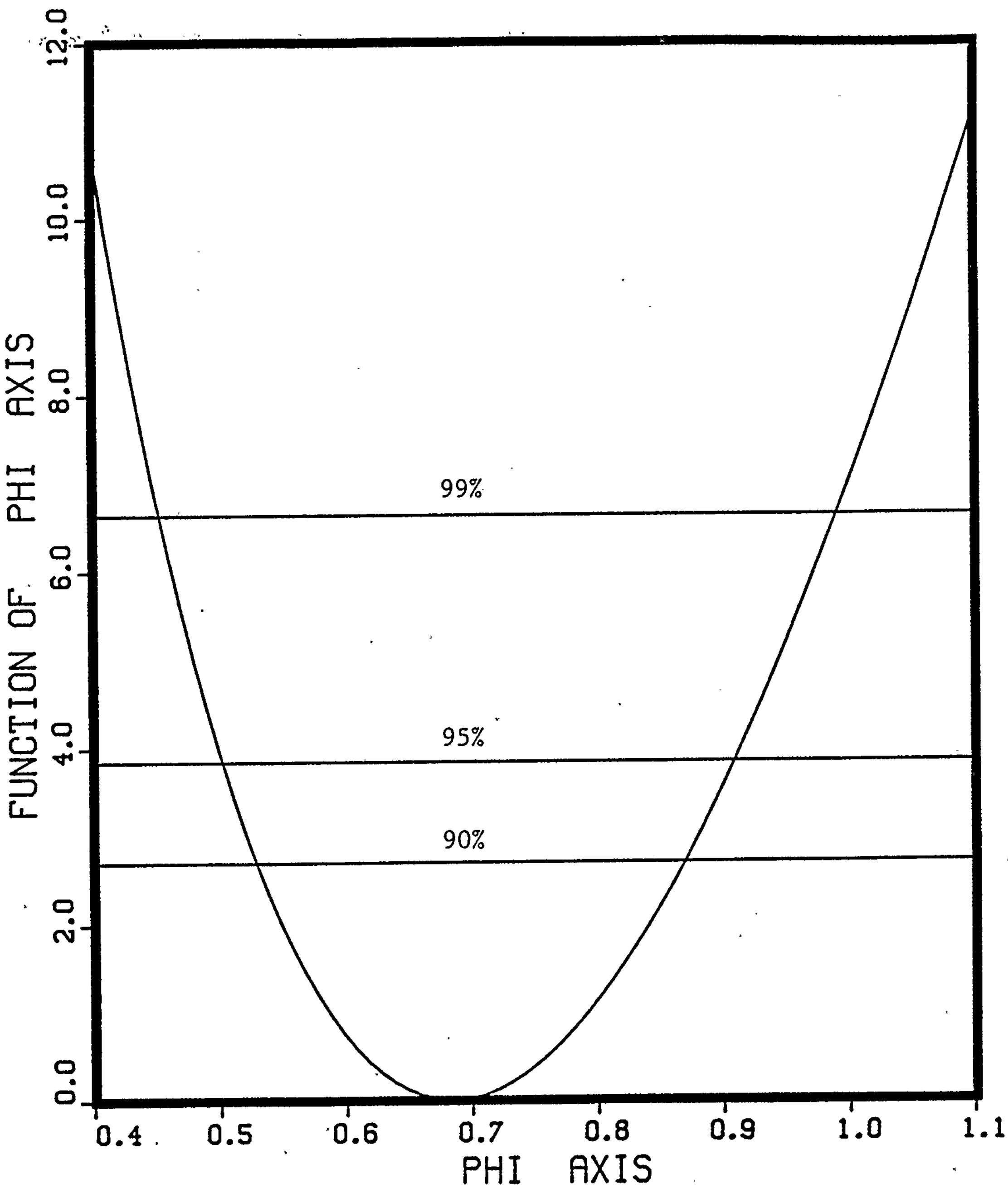
Also, when we substitute the data into (5.4.9), we obtain

$$T_c(\gamma) = -680 \, \ell n\left[\frac{1.5832}{1+.8525\gamma}\right] - 2\left\{43 \, \ell n\left[\frac{1+2\gamma}{2.3682}\right] + 36 \, \ell n\left[\frac{1+3\gamma}{3.0523}\right] + 27 \, \ell n\left[\frac{1+4\gamma}{3.7364}\right] + 24 \, \ell n\left[\frac{1+5\gamma}{4.4205}\right] + 20 \, \ell n\left[\frac{1+6\gamma}{5.1046}\right] + 18 \, \ell n\left[\frac{1+7\gamma}{5.7887}\right] + 8 \, \ell n\left[\frac{1+9\gamma}{7.1569}\right]\right\}. \tag{5.4.15}$$

By solving equation (5.4.14), we obtain the ML estimate of γ as 0.684148. This value is subsequently used to obtain (5.4.15). The expression in (5.4.15) is used to draw the U-shaped graph. All of these were done on the computer. The 95% C.I. for γ is found from the graph to be approximately (0.50, 0.91). This interval does not appear to be as short as those obtained for θ and λ. Both the 90% and 99% C.I. for φ are given as well in Figure 5.3.

FIGURE 5.3

90, 95 and 99% confidence intervals for φ



5.5 CONFIDENCE REGIONS

For uniparameter distributions, Bartlett (1953a) obtained approximate C.Is. for the parameter by assuming that the derivative of the log-likelihood function with respect to the parameter is approximately normally distributed with mean zero and known variance when the sample size n is large. Subsequently, Bartlett (1953b) extended the principle to multiparameter distributions and applied it to a distribution with two parameters θ_1 and θ_2. By using large sample approximation, he considered the partial derivatives of the log-likelihood function with respect to these parameters to be normally distributed with zero means and known covariance-matrix. These were used to obtain an approximate chi-square expression from which the confidence regions can be determined.

We shall apply Bartlett's results to the GPD model (1.3.1). By taking the partial logarithmic derivatives of the likelihood function in (4.1.1), we have

$$\frac{\partial \ell n L(\theta,\lambda)}{\partial\theta} = \frac{\partial}{\partial\theta}\left\{n\ell n\theta - \left[n\theta + \lambda \sum^{n} x_i\right]\right.$$

$$\left. + \sum_{i=1}^{n}\left[(x_i-1)\ \ell n(\theta+\lambda x_i) - \ell n(x_i!)\right]\right\}$$

$$= \frac{n}{\theta} - n + \sum_{i=1}^{n} \frac{x_i-1}{\theta+\lambda x_i},$$

$$\frac{\partial^2 \ell nL(\theta,\lambda)}{\partial\theta^2} = -\frac{n}{\theta^2} - \sum_{i=1}^{n} \frac{x_i - 1}{(\theta+\lambda x_i)^2}, \tag{5.5.1}$$

$$\frac{\partial^2 \ell nL(\theta,\lambda)}{\partial\lambda\partial\theta} = -\sum_{i=1}^{n} \frac{x_i(x_i-1)}{(\theta+\lambda x_i)^2}, \tag{5.5.2}$$

$$\frac{\partial \ell nL(\theta,\lambda)}{\partial\lambda} = -\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{x_i(x_i-1)}{\theta+\lambda x_i}$$

and

$$\frac{\partial^2 \ell nL(\theta,\lambda)}{\partial\lambda^2} = -\sum_{i=1}^{n} \frac{x_i^2(x_i-1)}{(\theta+\lambda x_i)^2}. \tag{5.5.3}$$

On taking the expectations of (5.5.1), (5.5.2) and (5.5.3) and using the principle that each r.v. X_i has a GPD model (1.3.1), we obtain,

$$I_{11} = E\left[-\frac{\partial^2 \ell nL(\theta,\lambda)}{\partial\theta^2}\right]$$

$$= E\sum_{i=1}^{n}\left[\frac{1}{\theta^2} + \frac{X_i-1}{(\theta+\lambda X_i)^2}\right]$$

$$= E\sum_{i=1}^{n}\frac{(\theta+\lambda)^2 X_i + \lambda^2 X_i(X_i-1)}{\theta^2(\theta+\lambda X_i)^2}$$

$$= \sum_{i=1}^{n}\left\{\left[\frac{\theta+\lambda}{\theta}\right]^2 E\left[\frac{X}{(\theta+\lambda X)^2}\right] + \left[\frac{\lambda}{\theta}\right]^2 E\left[\frac{X(X-1)}{(\theta+\lambda X)^2}\right]\right\}.$$

Now,

$$E\left[\frac{X}{(\theta+\lambda X)^2}\right] = \sum_{x=1}^{\infty} \frac{\theta(\theta+\lambda x)^{x-3}}{(x-1)!} e^{-\theta-\lambda x}$$

$$= \sum_{x=0}^{\infty} \frac{\theta(\theta+\lambda+\lambda x)^{x-2}}{x!} e^{-\theta-\lambda-\lambda x}$$

$$= \sum_{x=0}^{\infty} \frac{\theta[\theta+\lambda+\lambda x-\lambda x]}{\theta+\lambda} \cdot \frac{(\theta+\lambda+\lambda x)^{x-2}}{x!} e^{-\theta-\lambda-\lambda x}$$

$$= \frac{\theta}{\theta+\lambda} \sum_{x=0}^{\infty} \frac{(\theta+\lambda+\lambda x)^{x-1}}{x!} e^{-\theta-\lambda-\lambda x}$$

$$- \frac{\theta\lambda}{\theta+\lambda} \sum_{x=0}^{\infty} \frac{(\theta+2\lambda+\lambda x)^{x-1}}{x!} e^{-\theta-2\lambda-\lambda x}$$

$$= \frac{\theta}{(\theta+\lambda)^2} - \frac{\theta\lambda}{(\theta+\lambda)(\theta+2\lambda)}.$$

Also,

$$E\left[\frac{X(X-1)}{(\theta+\lambda X)^2}\right] = \sum_{x=2}^{\infty} \frac{\theta(\theta+\lambda x)^{x-3}}{(x-2)!} e^{-\theta-\lambda x}$$

$$= \sum_{x=0}^{\infty} \frac{\theta(\theta+2\lambda+\lambda x)^{x-1}}{x!} e^{-\theta-2\lambda-\lambda x}$$

$$= \frac{\theta}{\theta+2\lambda}. \tag{5.5.4}$$

Hence,

$$I_{11} = \sum_{i=1}^{n} \left\{ \left[\frac{\theta+\lambda}{\theta}\right]^2 \left[\frac{\theta}{(\theta+\lambda)^2} - \frac{\theta\lambda}{(\theta+\lambda)(\theta+2\lambda)}\right] + \left[\frac{\lambda}{\theta}\right]^2 \left[\frac{\theta}{\theta+2\lambda}\right] \right\}$$

$$= \sum_{i=1}^{n} \left\{ \frac{1}{\theta} - \frac{\lambda}{\theta+2\lambda} \right\}$$

$$= \frac{n[\theta(1-\lambda)+2\lambda]}{\theta(\theta+2\lambda)}. \tag{5.5.5}$$

$$I_{12} = E\left[- \frac{\partial^2 \ell n L(\theta,\lambda)}{\partial\lambda\partial\theta}\right]$$

$$= E \sum_{i=1}^{n} \left[\frac{X_i(X_i-1)}{(\theta+\lambda X_i)^2}\right]$$

$$= \sum_{i=1}^{n} E\left[\frac{X(X-1)}{(\theta+\lambda X)^2}\right]$$

$$= \frac{n\theta}{\theta+2\lambda} \tag{5.5.6}$$

by using the result in (5.5.4).

Also,

$$I_{22} = E\left[- \frac{\partial^2 \ell n L(\theta,\lambda)}{\partial\lambda^2}\right]$$

$$= E \sum_{i=1}^{n} \frac{X_i^2(X_i-1)}{(\theta+\lambda X_i)^2}$$

$$= \sum_{i=1}^{n} E\left[\frac{X^2(X-1)}{(\theta+\lambda X)^2}\right]$$

where

$$E\left[\frac{X^2(X-1)}{(\theta+\lambda X)^2}\right] = \sum_{x=2}^{\infty} x \frac{\theta(\theta+\lambda x)^{x-3}}{(x-2)!} e^{-\theta-\lambda x}$$

$$= \sum_{x=0}^{\infty} \theta(x+2) \frac{(\theta+2\lambda+\lambda x)^{x-1}}{x!} e^{-\theta-2\lambda-\lambda x}$$

$$= \frac{\theta}{\theta+2\lambda} \sum_{x=0}^{\infty} x(\theta+2\lambda) \frac{(\theta+2\lambda+\lambda x)^{x-1}}{x!} e^{-\theta-2\lambda-\lambda x}$$

$$+ \frac{2\theta}{\theta+2\lambda} \sum_{x=0}^{\infty} (\theta+2\lambda) \frac{(\theta+2\lambda+\lambda x)^{x-1}}{x!} e^{-\theta-2\lambda-\lambda x}$$

$$= \frac{\theta}{\theta+2\lambda}\cdot\frac{\theta+2\lambda}{1-\lambda} + \frac{2\theta}{\theta+2\lambda}$$

$$= \frac{\theta(\theta+2)}{(1-\lambda)(\theta+2\lambda)}.$$

Hence,

$$I_{22} = \sum_{i=1}^{n} \frac{\theta(\theta+2)}{(1-\lambda)(\theta+2\lambda)}$$

$$= \frac{n\theta(\theta+2)}{(1-\lambda)(\theta+2\lambda)}. \tag{5.5.7}$$

From (5.5.5), (5.5.6) and (5.5.7) we have

$$I_{22} - I_{12}^2/I_{11} = \frac{2n\theta}{(1-\lambda)[\theta(1-\lambda)+2\lambda]} = A. \tag{5.5.8}$$

Also,

$$I_{12}/I_{11} = \frac{\theta^2}{\theta(1-\lambda)+2\lambda}.$$

When the sample size n is large, the first partial derivatives

$$\frac{\partial \ln L(\theta,\lambda)}{\partial\theta} \quad \text{and} \quad \frac{\partial \ln L(\theta,\lambda)}{\partial\lambda}$$

of the log-likelihood function are normally distributed with zero means and known covariance-matrix.

Following the procedure suggested by Bartlett (1953b), the approximate chi-square expression for the confidence region (C.R.) is

$$T = \left[\frac{\partial \ln L(\theta,\lambda)}{\partial\theta}\right]^2 / I_{11} + \left[\frac{\partial \ln L(\theta,\lambda)}{\partial\lambda} - \frac{I_{12}}{I_{11}}\frac{\partial \ln L(\theta,\lambda)}{\partial\theta}\right]^2 / A$$

where A is defined in (5.5.8).

By substituting the values of the partial derivatives and of I_{11} and I_{12} in the above, we obtain

$$\frac{\partial \ln L(\theta,\lambda)}{\partial\lambda} - \frac{I_{12}}{I_{11}}\frac{\partial \ln L(\theta,\lambda)}{\partial\theta}$$

$$= -n\bar{x} + \sum_{i=1}^{n}\frac{x_i(x_i-1)}{\theta+\lambda x_i} - \frac{\theta^2}{\theta-\theta\lambda+2\lambda}\left[\frac{n(1-\theta)}{\theta} + \sum_{i=1}^{n}\frac{x_i-1}{\theta+\lambda x_i}\right]$$

$$= \sum_{i=1}^{n}\left[x_i - \frac{\theta^2}{\theta-\theta\lambda+2\lambda}\right]\frac{x_i-1}{\theta+\lambda x_i} - n\bar{x} - \frac{\theta n(1-\theta)}{\theta-\theta\lambda+2\lambda}$$

$$= \sum_{i=1}^{n}\frac{\left[x_i(\theta+2\lambda)-\theta(\theta+\lambda x_i)\right](x_i-1)}{(\theta+\lambda x_i)(\theta-\theta\lambda+2\lambda)} - n\bar{x} - \frac{\theta n(1-\theta)}{\theta-\theta\lambda+2\lambda}$$

$$= \frac{\theta+2\lambda}{\theta-\theta\lambda+2\lambda} \sum_{i=1}^{n} \frac{x_i(x_i-1)}{\theta+\lambda x_i} + \frac{n\theta(\theta-\bar{x})}{\theta-\theta\lambda+2\lambda} - n\bar{x}$$

and therefore,

$$T(\theta,\lambda) = T = \frac{\theta(\theta+2\lambda)}{n(\theta-\theta\lambda+2\lambda)} \left[\frac{n(1-\theta)}{\theta} + \sum_{i=1}^{n} \frac{x_i-1}{\theta+\lambda x_i}\right]^2$$

$$+ \frac{(1-\lambda)(\theta-\theta\lambda+2\lambda)}{2n\theta} \left[\frac{\theta+2\lambda}{\theta-\theta\lambda+2\lambda} \sum_{i=1}^{n} \frac{x_i(x_i-1)}{\theta+\lambda x_i} + \frac{n\theta(\theta-\bar{x})}{\theta-\theta\lambda+2\lambda} - n\bar{x}\right]^2$$

$$= \frac{\theta(\theta+2\lambda)}{n(\theta-\theta\lambda+2\lambda)} \left\{\left[\frac{n(1-\theta)}{\theta} + \sum_{i=1}^{n} \frac{x_i-1}{\theta+\lambda x_i}\right]^2\right.$$

$$\left. + \frac{1-\lambda}{2(\theta+2\lambda)} \left[\left[1 + \frac{2\lambda}{\theta}\right] \sum_{i=1}^{n} \frac{x_i(x_i-1)}{\theta+\lambda x_i} - n\bar{x}\left[2 - \lambda + \frac{2\lambda}{\theta}\right] + n\theta\right]^2\right\}. \tag{5.5.9}$$

The values of θ and λ for which the function $T(\theta,\lambda)$ in (5.5.9) remains less than $\chi^2_{\alpha,2}$, the critical value of chi-square distribution with 2 d.f., define an approximate C.R. contour. Thus, the set of values of θ and λ for which

$$T(\theta,\lambda) \leq \chi^2_{\alpha,2} \tag{5.5.10}$$

form an approximate 100(1−α)% C.R.. In actual practice, we find values of $T(\theta,\lambda)$ satisfying (5.5.10). These values are used to obtain a contour map of $T(\theta,\lambda)$ in the (θ,λ) plane. This will be called the

$100(1-\alpha)\%$ contour line. Points inside the obtained contour line are said to lie within the $100(1-\alpha)\%$ C.R.

EXAMPLE 5.1 CONTINUED

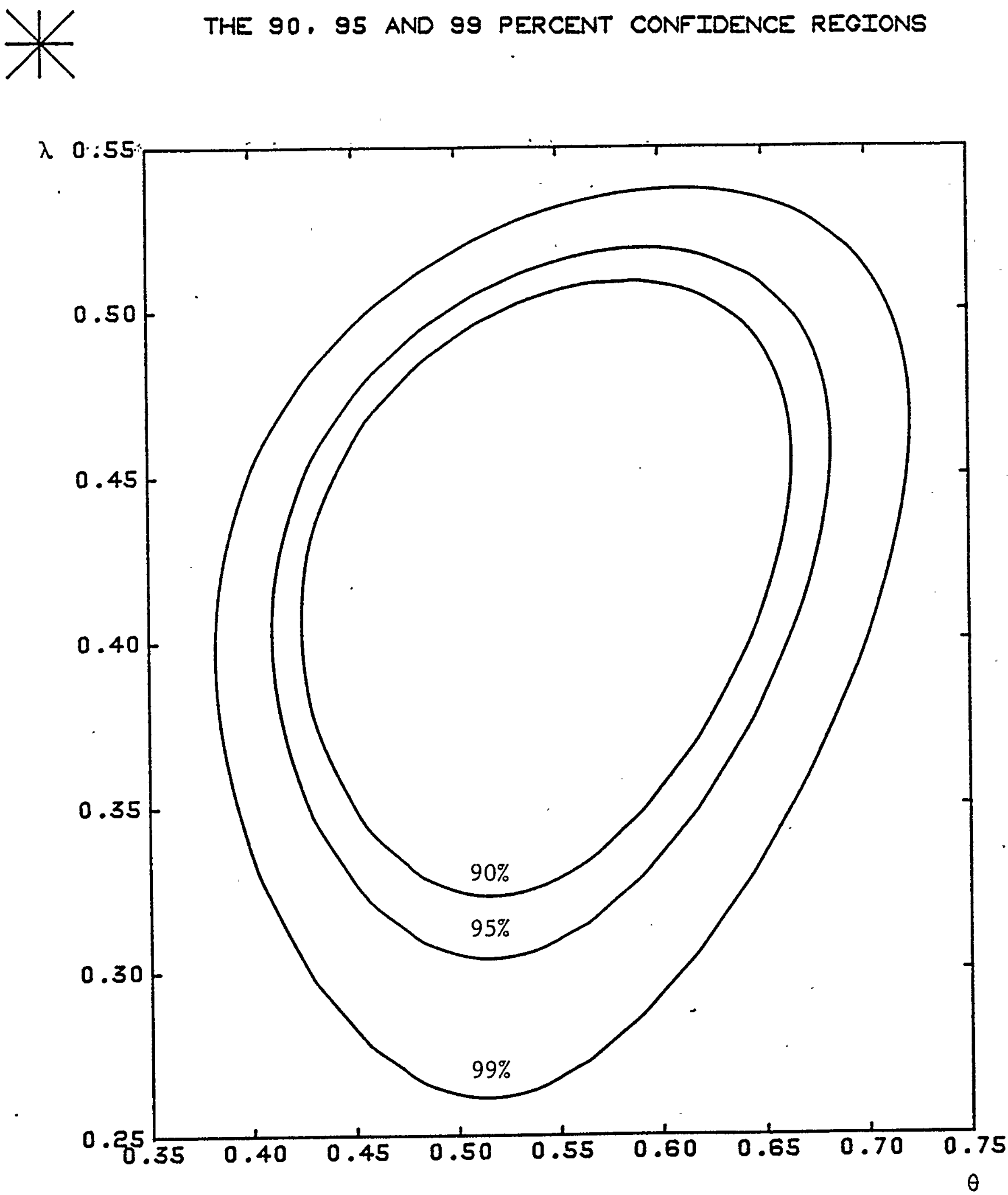
The data in Table 5.1 is used to set C.R. for the parameters θ and λ. By using the data in expression (5.5.9), we get

$$\begin{aligned} T(\theta,\lambda) = \frac{\theta(\theta+2\lambda)}{400(\theta-\theta\lambda+2\lambda)} \Bigg\{ & \left[\frac{165}{\theta} + \frac{43}{\theta+2\lambda} + \frac{36}{\theta+3\lambda} + \frac{27}{\theta+4\lambda}\right. \\ & \left. + \frac{24}{\theta+5\lambda} + \frac{20}{\theta+6\lambda} + \frac{18}{\theta+7\lambda} + \frac{8}{\theta+9\lambda} - 400\right]^2 \\ & + \frac{1-\lambda}{2(\theta+2\lambda)} \left[\frac{\theta+2\lambda}{\theta} \left[\frac{86}{\theta+2\lambda} + \frac{108}{\theta+3\lambda} + \frac{108}{\theta+4\lambda}\right.\right. \\ & \left. + \frac{120}{\theta+5\lambda} + \frac{120}{\theta+6\lambda} + \frac{126}{\theta+7\lambda} + \frac{72}{\theta+9\lambda}\right] \\ & \left. - 340\,\frac{(2\theta-\theta\lambda+2\lambda)}{\theta} + 400\theta\right]^2 \Bigg\}. \end{aligned} \tag{5.5.11}$$

A computer programme was written to find the values of the expression $T(\theta,\lambda)$ in (5.5.11) and these were used to obtain the contour lines for the 90%, 95% and 99% C.Rs. These are provided in Figure 5.4.

FIGURE 5.4

THE 90, 95 AND 99 PERCENT CONFIDENCE REGIONS



We notice that the 95% C.R. contour provides the 95% C.Is. for the parameters θ and λ as (0.44, 0.66) and (0.28, 0.48) respectively. The 95% C.Is. for each of the parameters θ and λ obtained in section 5.4 lie within the 95% C.R. for both parameters. The shapes of the C.Rs. are almost elliptical, which indicate the closeness of the anlaysed data to a normal distribution.

CHAPTER VI

SOME TESTS OF HYPOTHESES FOR GENERALIZED POISSON DISTRIBUTION

6.1 INTRODUCTION

Drawing inferences about the unknown parameters of a population by using the information contained in an observed sample data is an important problem in statistics. Usually, these inferences appear in either of two forms, as estimates of the respective parameters or as tests of hypotheses about their values. In this chapter, we are concerned with the latter.

The probability mass of the generalized Poisson distribution and its restricted form are given by (1.3.1) and (1.3.2) respectively.

The GPD model possesses the twin properties of over-dispersion and under-dispersion which make it to be a very good descriptive model in the fields of biology, ecology and many other areas. The under-dispersion is indicated by a negative value of the shape parameter λ in (1.3.1) while the over-dispersion is described by a positive value of λ. Tests of hypotheses can be applied to determine if there is over-dispersion or under-dispersion in a population. This will correspond to testing whether $\lambda > 0$ or $\lambda < 0$ in the GPD model (1.3.1).

Janardan and Schaeffer (1977) used the distribution (1.3.1) to model the number of aberrations awaiting restitution in human leukocytes. The parameter θ measures the rate of change while the parameter λ is related to the equilibrium constant. One may be interested in testing hypothesis about the magnitude of the rate of change, whether this is less than or greater than a specified quantity. This is an important area in which the tests of hypotheses are applicable.

Fazal (1977) considered the test $\lambda = 0$ against $\lambda \neq 0$ in the GPD model (1.3.1). This test was used to determine if a GPD model (1.3.1) should be used in place of a Poisson distribution model to fit a given data. Fazal based his test on the class of $c(\alpha)$ tests proposed by Neyman. We shall now discuss a number of tests for the GPD model.

Through out sections two to six of this chapter, we assume that a random sample $X_1, X_2, \ldots, X_n$ of size n is taken from a population with GP distribution.

6.2 UNIFORMLY MOST POWERFUL TEST

We assume that the random sample belongs to the restricted GPD model (1.3.2) where γ is known. A uniformly most powerful (UMP) test for the parameter θ can be constructed by using the following result by Lehmann (1959, page 70).

Lehmann's Result: Let Q be a real parameter, and let the random variable X has a probability density (with respect to some measure μ) given by

$$P_\theta(x) = C(\theta)\, e^{Q(\theta)T(x)}\, h(x), \tag{6.2.1}$$

where $Q(\theta)$ is a strictly monotone function of θ. If $Q(\theta)$ is an increasing function, then there exists a UMP test $\phi(x)$ for testing

$$H_0:\ \theta \le \theta_0 \qquad \text{against} \qquad H_a:\ \theta > \theta_0$$

which is given by

$$\phi(x) = \begin{cases} 1\ , & T(x) > C \\ \gamma\ , & T(x) = C \\ 0\ , & T(x) < C \end{cases} \tag{6.2.2}$$

where the boundary C of the critical region and the quantity γ are determined by

$$E\left[\phi(X)\ |\ \theta_0\right] = \alpha.$$

For the GPD model (1.3.2), the random variable X is discrete and μ is the counting measure, therefore the left hand side of (6.2.1) can be replaced by $P_x(\theta,\ \theta\gamma)$ to obtain

$$P_x(\theta,\ \theta\gamma) = e^{-\theta}\ e^{x[\ell n\theta - \theta\gamma]}\ (1+\gamma x)^{x-1}/x!,$$

where $C(\theta) = e^{-\theta}$, $Q(\theta) = \ell n\theta - \theta\gamma$, $T(x) = x$ and $h(x) = (1 + \gamma x)^{x-1}/x!$. When a random sample of size n is taken, we have

$$P_{\underline{x}}(\theta,\theta\gamma) \doteq e^{-n\theta}\ e^{Y[\ell n\theta - \theta\gamma]} \prod_{i=1}^{n} \left[\frac{(1+\gamma x_i)^{x_i-1}}{x_i!}\right] \tag{6.2.3}$$

where $Y = \sum^n X_i$ and $\underline{X} = (X_1, X_2, \ldots, X_n)$. It is well known by the convolution theorem that the random variable Y has a GPD with probability function (3.2.3)

On comparing (6.2.3) with (6.2.1), it is easy to note that

$$Q(\theta) = \ln\theta - \theta\Upsilon$$

is a strictly increasing function of θ as $Q'(\theta) > 0$. Hence the UMP test $\phi(x)$ for testing the null hypothesis H_0: $\theta \le \theta_0$ against the alternative hypothesis H_a: $\theta > \theta_0$ exists and is given by

$$\phi(x) = \begin{cases} 1, & Y > C \\ \gamma, & Y = C \\ 0, & Y < C \end{cases} \qquad (6.2.4)$$

where C and γ are determined from

$$\alpha = P(Y > C|H_0) + \gamma\, P(Y = C|H_0). \qquad (6.2.5)$$

The last term in (6.2.5) is only of interest if one is interested in randomization which will yield exact significance level α. Quite often, the statisticians ignore this term in order to avoid randomization and choose that value of C which gives a slightly lower probability value than α. In this way, the power of the test under the null hypothesis is being reduced.

If the last term in (6.2.5) is ignored, then C will be determined from

$$\sum_{y=C+1}^{\infty} n(n+\varphi y)^{y-1} \theta_0^y e^{-\theta_0(n+\varphi y)}/y! \le \alpha. \tag{6.2.6}$$

To determine the value of C numerically, a simple computer programme can be written to find C which will yield the largest probability $\le \alpha$. This method gives excellent results when the sample size n is small because the summation becomes easy. For large values of n the summation becomes pretty difficult even for the computer due to approximation errors.

Situations do arise in which n is large. It is well known by the central limit theorem that for large values of n the sample sum Y is approximately normally distributed with mean $n\mu$ and variance $n\sigma^2$ where

$$\mu = \theta(1-\theta\varphi)^{-1} \quad \text{and} \quad \sigma^2 = \theta(1-\theta\varphi)^{-3}.$$

Accordingly, the standardized variable

$$Z = \frac{Y - n\mu}{\sigma\sqrt{n}}$$

is approximately standard normal. With this assumption, and by ignoring the last term in (6.2.5), the quantity C can be determined as follows:

$$\begin{aligned}\alpha &= P\{Y > C|H_0\} \\ &= P\{Z > [C-n\theta_0(1-\varphi\theta_0)^{-1}]/\sqrt{[n\theta_0(1-\varphi\theta_0)^{-3}]}\} \\ &= P\{Z > z_\alpha\}\end{aligned}$$

where z_α is determined from the areas under the normal curve. Thus,

$$C = n\theta_0(1-\varphi\theta_0)^{-1} + z_\alpha\sqrt{[n\theta_0(1-\varphi\theta_0)^{-3}]}. \qquad (6.2.7)$$

The UMP size α test is to reject H_0 if $Y > C$.

For testing the hypothesis H_0: $\theta \geq \theta_0$ against H_a: $\theta < \theta_0$, a similar UMP test can be easily formulated. The only change will be that C will be replaced by C' and all the inequalities will be reversed.

6.3 AN APPROXIMATE TEST FOR φ OR λ

To carry out tests of hypotheses about the parameter λ when $\theta = \theta_0$ is known and large, we assume that $\lambda = \lambda_0$ is positive and that it is ≤ 0.5 in the GPD model (1.3.1). As indicated in section 4.2 and in the reference therein, the standardized variate

$$Z = \frac{X - \mu}{\sigma}$$

tends to standard normal form as θ increases without limit. For the GPD model (1.3.1), the mean μ and variance σ^2 of the random variable X are given in section 5.3 with θ replaced by θ_0. From one of the earlier papers on GPD, [Consul and Jain, 1973], the GPD model (1.3.1) is almost symmetrical in shape as the normal distribution for values of θ as high as 8.0 and $0 < \lambda \leq 0.5$. Therefore, a test for H_0: $\lambda = \lambda_0 \leq 0.5$ against H_a: $\lambda > \lambda_0$ can be based on normal approximation. It is interesting to note that this approximation does not depend on whether the sample size n is large or not.

Suppose a random sample of size n is taken. The test is based on the statistic

$$\bar{X} = n^{-1} \sum_{i=1}^{n} X_i. \tag{6.3.1}$$

The critical region for rejecting H_0 is $\bar{X} > C$ where C is determined from

$$\begin{aligned}
\alpha &= P\{\bar{X} > C \mid H_0\} \\
&= P\left\{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} > \frac{\sqrt{n}(C-\mu)}{\sigma} \mid H_0\right\} \\
&= P\left\{Z > \sqrt{n}\ [C-\theta_0(1-\lambda_0)^{-1}]/\sqrt{[\theta_0(1-\lambda_0)^{-3}]} \mid H_0\right\} \\
&= P\left\{Z > z_\alpha\right\}.
\end{aligned}$$

Thus, by finding the value of z_α from the standard normal table, the value of C is given by

$$C = \theta_0(1-\lambda_0)^{-1} + z_\alpha \sqrt{\{[\theta_0(1-\lambda_0)^{-3}]/n\}}. \tag{6.3.2}$$

A similar test can be carried out for the parameter φ in the GPD model (1.3.2) if it is known that $\varphi = \varphi_0$ is positive and that $\varphi\theta_0 \le 0.5$. In this case, the standardized variate

$$Z = \frac{X - \mu}{\sigma}$$

has an approximate normal distribution, where μ and σ^2 are given in section 5.3 with θ replaced by θ_0.

A test for H_0: $\varphi \le \varphi_0$ against H_a: $\varphi > \varphi_0$ is based on the statistic

$\bar{X}$. By following the earlier procedure, the value of the boundary C is given by

$$C = \theta_0(1-\varphi_0\theta_0)^{-1} + z_\alpha\sqrt{[\theta_0(1-\varphi_0\theta_0)^{-3}/n]}. \qquad (6.3.3)$$

One can easily formulate a two-sided test for testing a simple hypothesis. One will be testing H_0: $\varphi = \varphi_0$ against H_a: $\varphi \neq \varphi_0$. The critical region for rejecting H_0 is $|\bar{X}-\mu| > C-\mu$ where C is determined from

$$\alpha = P\{|\bar{X}-\mu| > C-\mu \mid H_0\}$$

which is equivalent to

$$\begin{aligned} \tfrac{1}{2}\alpha &= P\{\bar{X} - \mu > C - \mu \mid H_0\} \\ &= P\left\{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} > \frac{\sqrt{n}(C-\mu)}{\sigma} \mid H_0\right\} \\ &= P\{Z > z_{\alpha/2}\}. \end{aligned}$$

The value of $z_{\alpha/2}$ from the standard normal table will be used to obtain the value of C. On substituting the values of μ and σ, we get

$$C = \theta_0(1-\varphi_0\theta_0)^{-1} + z_{\alpha/2}\sqrt{[\theta_0(1-\varphi_0\theta_0)^{-3}/n]}. \qquad (6.3.4)$$

6.4 LIKELIHOOD RATIO TESTS FOR θ OR λ IN LARGE SAMPLES

We consider the composite hypotheses

$$H_0: \quad \theta = \theta_0, \ \lambda \quad \text{unknown}$$

against (6.4.1)

$$H_a: \ \theta \neq \theta_0, \ \lambda \quad \text{unknown.}$$

The likelihood function of the GPD model (1.3.1) is given by (4.1.1). Consul and Shoukri (1984) have shown that the ML estimates $\hat{\theta}$ and $\hat{\lambda}$ of θ and λ in the model (1.3.1) are unique and that they are obtained by solving equations

$$\sum_{i=1}^{n} \frac{x_i(x_i-1)}{\bar{x}+(x_i-\bar{x})\hat{\lambda}} - n\bar{x} = 0 \tag{6.4.2}$$

and

$$\hat{\theta} = \bar{x}(1-\hat{\lambda}). \tag{6.4.3}$$

Let the ML estimate of λ when θ is fixed at θ_0 be denoted by $\tilde{\lambda}(\theta_0)$. This estimate is used to obtain the likelihood ratio, when H_0 is true as

$$\ell = \frac{L\left[\theta_0, \ \tilde{\lambda}(\theta_0)\right]}{L(\hat{\theta},\hat{\lambda})} \tag{6.4.4}$$

which satisfies $0 \leq \ell \leq 1$. Large values of ℓ imply H_0 is reasonably acceptable since ℓ is the maximum likelihood under H_0 as a fraction of its largest possible value. The critical region will be of the form

$$\ell \leq c_\alpha$$

where c_α is determined from

$$\int_0^{c_\alpha} f(\ell) \, d\ell = \alpha$$

if $f(\ell)$ is the density of ℓ.

It is almost impossible to find the exact distribution of ℓ in some cases. As proved earlier in Chapter V, the statistic $-2 \log \ell$ is an approximate chi-square random variable with 1 d.f. when the sample size n is large. Thus, test of hypotheses in (6.4.1) can be carried out by using a chi-square random variable as the test statistic.

We now find the unconditional maximum $L(\hat{\theta},\hat{\lambda})$ of the likelihood function and the conditional maximum $L(\theta_0, \tilde{\lambda}(\theta_0))$ of the likelihood function where $\hat{\theta}$ and $\hat{\lambda}$ are the ML estimates obtained from equations (6.4.2) and (6.4.3). The ML estimate $\tilde{\lambda}(\theta_0)$ is obtained, through iterative procedure, from equation (4.3.6) with θ replaced by θ_0.

By substituting the values of the likelihood functions in (6.4.4), one obtains

$$\ell = \frac{\theta_0^n \, e^{-[n\theta_0+\tilde{\lambda}(\theta_0)\sum\limits^n x_i]} \prod\limits_{i=1}^{n} \left[\left[\theta_0+\tilde{\lambda}(\theta_0)x_i\right]^{x_i-1} / x_i!\right]}{\hat{\theta}^n \, e^{-[n\hat{\theta}+\hat{\lambda}\sum\limits^n x_i]} \prod\limits_{i=1}^{n} \left[\left[\hat{\theta}+\hat{\lambda}x_i\right]^{x_i-1} / x_i!\right]}$$

$$= \left[\frac{\theta_0}{\hat{\theta}}\right]^n e^{n[\hat{\theta}-\theta_0+\bar{x}(\hat{\lambda}-\tilde{\lambda}(\theta_0))]} \prod_{i=1}^{n} \left[\frac{\theta_0+\tilde{\lambda}(\theta_0)x_i}{\hat{\theta}+\hat{\lambda}x_i}\right]^{x_i-1} .$$

Finding the exact distribution of the statistic ℓ does not appear to be an easy task. Since the sample size n is large, the random variable $-2 \log \ell$ is approximately chi-square distributed with 1 d.f. Hence the likelihood ratio test (LRT) statistic T is given by

$$\begin{aligned} T &= -2 \log \ell \\ &= -2\left\{ n \log\left[\frac{\theta_0}{\hat{\theta}}\right] + n\left[\hat{\theta} - \theta_0 + \bar{x}(\hat{\lambda}-\tilde{\lambda}(\theta_0))\right] \right. \\ &\qquad \left. + \sum_{i=1}^{n} (x_i-1) \log\left[\frac{\theta_0+\tilde{\lambda}(\theta_0)x_i}{\hat{\theta}+\hat{\lambda}x_i}\right] \right\}. \end{aligned} \tag{6.4.5}$$

We reject the null hypothesis at significance level α if the calculated T in (6.4.5) from a random sample of size n is such that

$$T > \chi^2_{\alpha,1}$$

where $\chi^2_{\alpha,1}$ is the upper $100(1-\alpha)$ percent point of chi-square distribution with 1 d.f..

When the test of hypothesis is about the parameter λ, we modify (6.4.4) such that θ_0 is replaced by $\tilde{\theta}(\lambda_0)$ and $\tilde{\lambda}(\theta_0)$ is replaced by λ_0 where $\tilde{\theta}(\lambda_0)$ is the ML estimate of θ in the GPD model (1.3.1) when the parameter λ is fixed at λ_0.

Therefore, we have

$$\ell_1 = \frac{L(\tilde{\theta}(\lambda_0),\ \lambda_0)}{L(\hat{\theta},\hat{\lambda})}. \tag{6.4.6}$$

The ML estimate $\tilde{\theta}(\lambda_0)$ in the conditional maximum $L(\tilde{\theta}(\lambda_0), \lambda_0)$ of the likelihood function is obtained through iteration from equation (4.3.3) with λ replaced by λ_0. By using the values of the likelihood functions in (6.4.6), we obtain

$$\ell_1 = \left[\frac{\tilde{\theta}(\lambda_0)}{\hat{\theta}}\right]^n e^{n[\hat{\theta}-\tilde{\theta}(\lambda_0)+\bar{x}(\hat{\lambda}-\lambda_0)]} \prod_{i=1}^{n} \left[\frac{\tilde{\theta}(\lambda_0)+\lambda_0 x_i}{\hat{\theta}+\hat{\lambda} x_i}\right]^{x_i-1} .$$

As in the previous case, $-2 \log \ell_1$ is approximately chi-square distributed with 1 d.f.

6.5 CONDITIONAL LIKELIHOOD RATIO TEST

The conditional likelihood ratio test is applicable to tests about the parameter γ in the GPD model (1.3.2) when the sample size n is large. The likelihood function of the GPD model (1.3.2) is given by (4.1.2). We shall now formulate a procedure for testing the composite hypotheses

$$H_0: \gamma = \gamma_0 , \quad \theta \quad \text{unknown}$$

against

$$H_a: \gamma \neq \gamma_0, \quad \theta \quad \text{unknown.}$$

Since Y is a sufficient statistic for the nuisance parameter θ, one can easily eliminate the parameter θ by dividing the likelihood function by the probability function of Y as in subsection 5.4.2. From (5.4.6), the conditional likelihood function $C(\gamma)$ is given by

$$C(\gamma) = \frac{y!}{n(n+\gamma y)^{y-1}} \prod_{i=1}^{n} \left[\frac{(1+\gamma x_i)^{x_i-1}}{x_i!}\right]. \tag{6.5.1}$$

We define the conditional likelihood ratio function as

$$\ell_c = \frac{C(\gamma)}{C(\hat{\gamma})} \quad . \tag{6.5.2}$$

where $\hat{\gamma}$ is that value of γ which maximizes (6.5.1) and it is the root of equation (5.4.7). Although it appears as if equation (5.4.7) will have multiple roots, but from each of several examples we have worked on, the iterative procedures used to solve (5.4.7) have yielded a single solution in very few steps.

When H_0 is true, the conditional likelihood ratio (6.5.2) gives

$$\ell_c = \left[\frac{1 + \hat{\gamma}\bar{x}}{1 + \gamma_0\bar{x}}\right]^{n\bar{x}-1} \prod_{i=1}^{n} \left[\frac{1 + \gamma_0 x_i}{1 + \hat{\gamma} x_i}\right]^{x_i-1} .$$

It is interesting to note that this quantity behaves very much like the likelihood ratio ℓ. We have that $0 \leq \ell_c \leq 1$. By Corollary 5.2 of Chapter V, the conditional likelihood ratio test (CLRT) statistic $-2 \log \ell_c$ is approximately chi-square distributed with 1 d.f.. By applying this result, the test statistic

$$T_c = -2 \log \ell_c$$

$$= -2\left\{(n\bar{x}-1)\log\left[\frac{1+\hat{\gamma}\bar{x}}{1+\gamma_0\bar{x}}\right]\right.$$

$$\left. + \sum_{i=1}^{n}(x_i-1)\log\left[\frac{1+\gamma_0 x_i}{1+\hat{\gamma}x_i}\right]\right\} \qquad (6.5.3)$$

is distributed approximately as a chi-square random variable with 1 d.f. To test the composite hypothesis H_0 against the composite alternative H_a, the value of T_c is computed from (6.5.3) by using the observations from a random sample of size n and H_0 is rejected if

$$T_c > \chi^2_{\alpha,1}$$

where $\chi^2_{\alpha,1}$, as defined in section 6.4, will be obtained from the chi-square distribution table.

6.6 POWERS OF THE TESTS

We consider the procedures discussed in sections two through five and give the powers of the different tests for a specified value of the parameter under the alternative hypothesis.

For the UMP test in (6.2.4), the power of the test under H_a, when randomization is ignored is

$$\pi = 1-\beta = P\{Y > C \mid H_a\}$$

$$= \sum_{y=C+1}^{\infty} n(n+\varphi y)^{y-1} \theta_1^y e^{-\theta_1(n+\varphi y)}/y! \qquad (6.6.1)$$

where θ_1 is the specified value of θ under the alternative hypothesis and C is determined from (6.2.6).

For the approximate normal test in section 6.3, the power is given by

$$\pi = 1-\beta = P\{\bar{X} > C \mid H_a\}. \qquad (6.6.2)$$

In this case, C is determined from either (6.3.2) or (6.3.3) depending on whether the parameter we are testing for is λ or φ.

The test statistic T in (6.4.5) is, in general, a non-central chi-square random variable with 1 d.f. and non-centrality parameter

$$\nu_1 = (\theta-\theta_0)^2 \ \mathrm{Var}(\hat{\theta}) \qquad (6.6.3)$$

where

$$\mathrm{Var}(\hat{\theta}) = E\left[-\frac{\partial^2 \ell n L(\theta,\lambda)}{\partial\theta^2}\right],$$

[See Kendall and Stuart, 1977 page 247].

It has been shown in Chapter V for the GPD model (1.3.1) that

$$\mathrm{Var}(\hat{\theta}) = I_{11}(\theta) = \frac{n[\theta(1-\lambda)+2\lambda]}{\theta(\theta+2\lambda)} .$$

Under the null hypothesis, where $\theta = \theta_0$ and $\nu_1 = 0$ the test statistic T is a central chi-square random variable. Under H_a, the alternative hypothesis, ν_1 is as given in (6.6.3) with θ replaced by its specified value under H_a. Therefore, the power of the LRT is

$$\pi = 1-\beta = \int_{\chi^2_{\alpha,1}}^{\infty} d\,\chi^2(1,\nu_1). \tag{6.6.4}$$

But a non-central chi-square $\chi^2(1, \nu_1)$ can be approximated by an equivalent central chi-square distributed random variable. By following the procedure in Kendall and Stuart (1977, page 245), (6.6.4) can be approximated to give an approximate power for the LRT as

$$\pi \simeq \int_{b}^{\infty} d\,\chi^2 \left[1 + \frac{\nu_1^2}{1+2\nu_1}\right] \tag{6.6.5}$$

where $b = (1+\nu_1)(1+2\nu_1)^{-1}\,\chi^2_{\alpha,1}$ and $\chi^2(r)$ is a central chi-square variate with r d.f. and

$$\nu_1 = (\theta_1-\theta_0)^2\,\frac{n[\theta_1(1-\tilde{\lambda}(\theta_1))+2\tilde{\lambda}(\theta_1)]}{\theta_1(\theta_1+2\tilde{\lambda}(\theta_1))}. \tag{6.6.6}$$

In (6.6.6), θ_1 is the value of θ specified under the alternative hypothesis and $\tilde{\lambda}(\theta_1)$ is the ML estimate of λ in the GPD model (1.3.1) when the value of θ is taken to be θ_1. The ML estimate $\tilde{\lambda}(\theta_1)$ is the solution of equation (4.3.6) when θ is replaced by θ_1.

The procedure to obtain the power of the CLRT is similar to the above method and will not be given here.

6.7 SEQUENTIAL PROBABILITY RATIO TEST

The Neyman-Pearson theory considers the problem of constructing a most powerful test at a given sample size and significance level so as to keep both types of errors within a very reasonable limit. The choice of α and β (type I and II errors), remain somewhat arbitrary. A disadvantage of this approach is its failure to relate the choice of sample size and significance level (type I error) to the economic background.

A corrective approach to the above type of economic problem was introduced by Wald (1950). However, we shall not delve into the economic prospect of this elegant approach, we shall briefly outline how sequential probability ratio test (SPRT) can be applied to test hypotheses about the parameter of the restricted GP distribution.

The SPRT procedure achieves optimal economy in the required sample size in any problem. A rigorous mathematical proof has been given to this fact by Lehmann (1959). The SPRT, developed by Wald (1945, '47) will be applied to testing a null hypothesis against one-sided alternatives. Its application to two-sided alternatives is more complicated and will not be dealt with here.

Consider the hypothesis

$$H_0: \ \theta = \theta_0 \qquad \text{against}$$
$$H_a: \ \theta = \theta_1 \qquad (\theta_1 > \theta_0). \tag{6.7.1}$$

Let $X_1, X_2, X_3, \ldots$ be a sequence of independent random variables from the GPD given in (1.3.2) and suppose the value of the parameter φ is fixed. First, we obtain a sequential test for the hypotheses in (6.7.1). Subsequently, we shall obtain approximate expressions for the operating characteristic (OC) function and the average sample number (ASN) function of the SPRT.

Let A and B, $B < A$, be two given numbers. A SPRT for the test in (6.7.1) is defined as follows:

Observe $\{X_i\}$, $i = 1,2,3, \ldots$ successively, and at stage $N \geq 1$

(i) reject H_0 if $L(x) \geq A$

(ii) accept H_0 if $L(x) \leq B$

(iii) continue by observing X_{N+1} if $B < L(x) < A$ where

$$L(x) = \frac{P_{1N}}{P_{0N}} = \frac{P_{x_1}(\theta_1, \varphi\theta_1)\cdots\cdot P_{x_N}(\theta_1, \varphi\theta_1)}{P_{x_1}(\theta_0, \varphi\theta_0)\cdots\cdot P_{x_N}(\theta_0, \varphi\theta_0)}$$

$$= \frac{L(\theta_1, \varphi\theta_1)}{L(\theta_0, \varphi\theta_0)}$$

$$= \frac{\theta_1^{\overset{N}{\Sigma} x_i} e^{-\theta_1(N+\varphi\overset{N}{\Sigma} x_i)} \prod_{i=1}^{N} \left[(1+\varphi x_i)^{x_i-1}/(x_i!)\right]}{\theta_0^{\overset{N}{\Sigma} x_i} e^{-\theta_0(N+\varphi\overset{N}{\Sigma} x_i)} \prod_{i=1}^{N} \left[(1+\varphi x_i)^{x_i-1}/(x_i!)\right]}$$

$$= \left[\frac{\theta_1}{\theta_0}\right]^{\overset{N}{\Sigma} x_i} e^{(N+\gamma \overset{N}{\Sigma} x_i)(\theta_0-\theta_1)}. \tag{6.7.2}$$

Thus, we continue to sample as long as

$$\log B < \log L(x) < \log A \tag{6.7.3}$$

and accept H_0 or H_a according as

$$\log L(x) \leq \log B$$

or

$$\log L(x) \geq \log A.$$

By using (6.7.2) in (6.7.3), we get

$$\log B < \log\left\{\left[\frac{\theta_1}{\theta_0} e^{\gamma(\theta_0-\theta_1)}\right]^{\overset{N}{\Sigma} x_i} \cdot e^{N(\theta_0-\theta_1)}\right\} < \log A$$

i.e. $$\log\left[B\ e^{N(\theta_1-\theta_0)}\right] < \sum_{i=1}^{N} x_i \cdot \log\left[\frac{\theta_1}{\theta_0} e^{\gamma(\theta_0-\theta_1)}\right] < \log\left[A\ e^{N(\theta_1-\theta_0)}\right]. \tag{6.7.4}$$

But $0 < \theta_0 \gamma < \theta_1 \gamma < 1$ so that $\gamma < \theta_1^{-1}$. Also,

$$\frac{\theta_1}{\theta_0} e^{\gamma(\theta_0-\theta_1)} = \frac{\theta_1}{\theta_0} e^{\gamma\theta_1\left[\frac{\theta_0}{\theta_1} - 1\right]}$$

$$> \frac{\theta_1}{\theta_0} e^{\left[\frac{\theta_0}{\theta_1} - 1\right]}$$

$$= \frac{\theta_0}{\theta_1} \left\{1 + \left[\frac{\theta_0}{\theta_1} - 1\right] + \frac{1}{2}\left[\frac{\theta_0}{\theta_1} - 1\right]^2 + \frac{1}{3}\left[\frac{\theta_0}{\theta_1} - 1\right]^3 + \cdots\right\}$$

$$> \frac{\theta_1}{\theta_0} \left\{1 + \frac{\theta_0}{\theta_1} - 1\right\}$$

$$= 1.$$

Therefore,

$$\log\left[\frac{\theta_1}{\theta_0} e^{\Upsilon(\theta_0-\theta_1)}\right] > 0. \tag{6.7.5}$$

Because of (6.7.5), the inequalities in (6.7.4) can be written as

$$\frac{\log\left[B\ e^{N(\theta_1-\theta_0)}\right]}{\log\left[\frac{\theta_1}{\theta_0} e^{\Upsilon(\theta_0-\theta_1)}\right]} < \sum_{i=1}^{N} x_i < \frac{\log\left[A\ e^{N(\theta_1-\theta_0)}\right]}{\log\left[\frac{\theta_1}{\theta_0} e^{\Upsilon(\theta_0-\theta_1)}\right]}. \tag{6.7.6}$$

From Wald's analysis, the constants A and B can be approximated by

$$\begin{aligned} A &\simeq (1-\beta)/\alpha \\ B &\simeq \beta/(1-\alpha) \end{aligned} \tag{6.7.7}$$

where α and β are the probabilities of the first and second types of error, respectively. Also, the SPRT terminates with probability one. By using (6.7.6), we conclude that the SPRT in (6.7.1), with error probabilities α and β, is given by the statistic $\sum^{N} X_i$ and that the two boundaries are

$$\ell = \frac{\log B + N(\theta_1-\theta_0)}{\log\left[\frac{\theta_1}{\theta_0}\right] + \gamma(\theta_0-\theta_1)}$$

below and

$$u = \frac{\log A + N(\theta_1-\theta_0)}{\log\left[\frac{\theta_1}{\theta_0}\right] + \gamma(\theta_0-\theta_1)}$$

above, where A and B are given by (6.7.7).

6.7.1 OPERATING CHARACTERISTIC FUNCTION

Denote by $L(\theta)$ the probability that the sequential process will terminate by accepting the null hypothesis H_0 when θ is the true value of the parameter. Thus, $L(\theta)$ is the OC function. We are interested in finding an approximate value for $L(\theta)$ by neglecting the excess of $L(x)$ over the boundaries A and B when the process is terminated. By following Wald's argument, $L(\theta)$ can be approximated by

$$L(\theta) \simeq \frac{A^h - 1}{A^h - B^h}$$

$$= \frac{\left[\frac{1-\beta}{\alpha}\right]^h - 1}{\left[\frac{1-\beta}{\alpha}\right]^h - \left[\frac{\beta}{1-\alpha}\right]^h} \qquad (6.7.8)$$

where the function $h = h(\theta) \neq 0$ is obtained from

$$\sum_{x=0}^{\infty} \left[\frac{P_x(\theta_1, \varphi\theta_1)}{P_x(\theta_0, \varphi\theta_0)}\right]^h P_x(\theta, \varphi\theta) = 1.$$

Thus,

$$1 = \sum_{x=0}^{\infty} \left[\frac{\theta_1}{\theta_0}\right]^{hx} e^{-(\theta_1-\theta_0)(1+\varphi x)h} \frac{(1+\varphi x)^{x-1}}{x!} \theta^x e^{-\theta(1+\varphi x)}$$

$$= \sum_{x=0}^{\infty} \frac{(1+\varphi x)^{x-1}}{x!} \left[\theta\left[\frac{\theta_1}{\theta_0}\right]^h\right]^x e^{-[\theta+(\theta_1-\theta_0)h](1+\varphi x)} . \qquad (6.7.9)$$

By comparing the above summation with

$$\sum_{x=0}^{\infty} P_x(\theta, \varphi\theta) = 1,$$

it is easy to see that

$$\theta\left[\frac{\theta_1}{\theta_0}\right]^h = \theta + (\theta_1-\theta_0)h. \qquad (6.7.10)$$

Solving for $h = h(\theta)$ in (6.7.9) is equivalent to solving for h in (6.7.10). Thus, we get

$$\theta = \frac{(\theta_1-\theta_0)h}{\left[\theta_1/\theta_0\right]^h-1}. \qquad (6.7.11)$$

We now need to solve for h in (6.7.10) and these values of h will be subsequently used in (6.7.11) and (6.7.8) to obtain the points $(\theta, L(\theta))$ which are used to draw the OC function. However, the task of finding $h = h(\theta)$ in (6.7.10) does not appear to be an easy one. This problem is overcome by the fact that when any h is chosen arbitrarily the point $(\theta, L(\theta))$, computed from (6.7.11) and (6.7.8), will lie on the OC function. By considering many values of h which lead to large number of points $(\theta, L(\theta))$, one can draw the OC function.

From (6.7.11) and (6.7.8), the points in the following Table 6.1 are easy to obtain.

TABLE 6.1

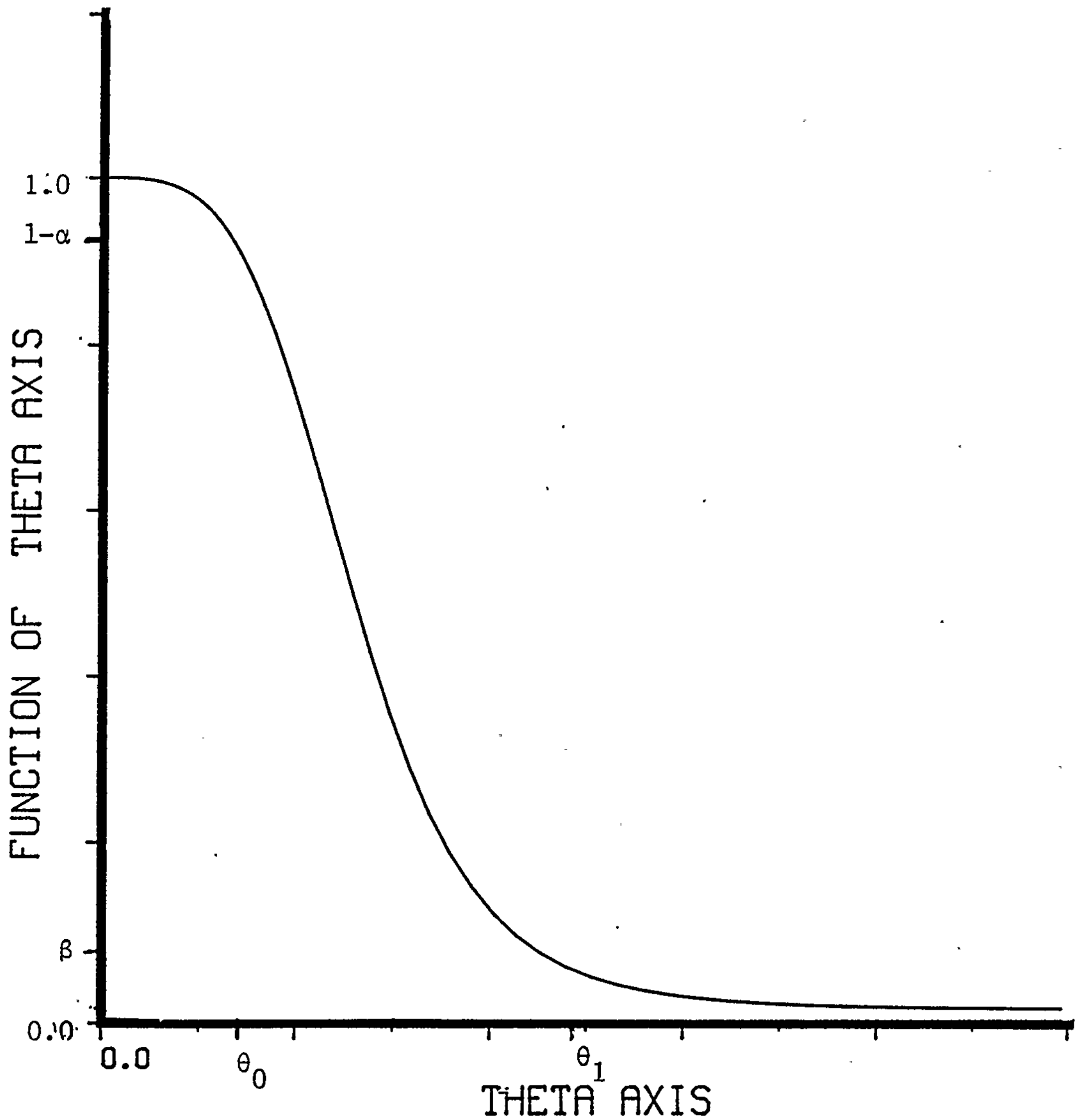
Some points on the OC function

h	θ	$L(\theta)$
1	θ_0	$1 - \alpha$
-1	θ_1	β
$+\infty$	0	1
$-\infty$	$+\infty$	0
0	$\dfrac{\theta_1-\theta_0}{\ln\theta_1-\ln\theta_0}$	$\dfrac{\ln[(1-\beta)/\alpha]}{\ln[(1-\beta)(1-\alpha)/\alpha\beta]}$
2	$\dfrac{2\theta_0^2}{\theta_1+\theta_0}$	$\dfrac{(1-\beta+\alpha)(1-\alpha)^2}{1-\alpha-\beta+2\alpha\beta}$
-2	$\dfrac{2\theta_1^2}{\theta_1+\theta_0}$	$\dfrac{(1-\beta+\alpha)\ \beta^2}{1-\alpha-\beta+2\alpha\beta}$

By using these seven points, a rough OC function which is given in Figure 6.1 can be obtained.

FIGURE 6.1

Operating Characteristic Function



6.7.2 AVERAGE SAMPLE NUMBER FUNCTION

From Wald's fundamental identity, it is known that the approximate formula for the average sample number function is

$$E_\theta(N) \simeq \frac{L(\theta) \log B + (1-L(\theta)) \log A}{E_\theta(Z)} \tag{6.7.12}$$

where
$$Z = \log \frac{P_x(\theta_1, \gamma\theta_1)}{P_x(\theta_0, \gamma\theta_0)}.$$

In subsection 6.7.1, the approximate formula for $L(\theta)$ has been given. We note here that $E_\theta(N)$, the expected value of N (the number of observations required by the SPRT) when θ is the true value of the parameter is a function of θ. For the GPD model (1.3.2), we have

$$E_\theta(Z) = E_\theta \log\left[\left[\frac{\theta_1}{\theta_0}\right]^X e^{-(\theta_1-\theta_0)(1+\gamma X)}\right]$$

$$= \log\left[\frac{\theta_1}{\theta_0}\right]\cdot E_\theta(X) - (\theta_1-\theta_0)[1 + \gamma E_\theta(X)]$$

$$= \frac{\theta}{1-\gamma\theta} \log\left[\frac{\theta_1}{\theta_0}\right] - \frac{\theta_1-\theta_0}{1-\gamma\theta}.$$

Hence,

$$E_\theta(N) \doteq \frac{L(\theta)\ \log B + (1-L(\theta))\log A}{\frac{\theta}{1-\gamma\theta}\log\left[\frac{\theta_1}{\theta_0}\right] - \frac{\theta_1-\theta_0}{1-\gamma\theta}}$$

$$= \frac{\frac{[(1-\beta)/\alpha]^h-1}{[(1-\beta)/\alpha]^h-[\beta/(1-\alpha)]^h}\log\left[\frac{\beta}{1-\alpha}\right] + \frac{1-[\beta/(1-\alpha)]^h}{[(1-\beta)/\alpha]^h-[\beta/(1-\alpha)]^h}\log\left[\frac{1-\beta}{\alpha}\right]}{\frac{\theta}{1-\gamma\theta}\log\left[\frac{\theta_1}{\theta_0}\right] - \frac{\theta_1-\theta_0}{1-\gamma\theta}}. \tag{6.7.13}$$

Since it is not easy to find the values of h as indicated in subsection 6.7.1, we consider a large number of values for h which are chosen arbitrarily. By substituting these values of h into (6.7.11) and (6.7.13) we obtain the points $(\theta,\ E_\theta(N))$ which are used to draw the ASN function.

The following three points are easy to obtain for the ASN function. When $\theta = \theta_0$, $L(\theta_0) = 1-\alpha$ and

$$E_{\theta_0}(N) = \frac{(1-\alpha)\ \log\left[\frac{\beta}{1-\alpha}\right] + \alpha\ \log\left[\frac{1-\beta}{\alpha}\right]}{\frac{\theta_0}{1-\gamma\theta_0}\log\left[\frac{\theta_1}{\theta_0}\right] - \frac{\theta_1-\theta_0}{1-\gamma\theta_0}}.$$

When $\theta = \theta_1$, $L(\theta_1) = \beta$ and

$$E_{\theta_1}(N) = \frac{\beta\ \log\left[\frac{\beta}{1-\alpha}\right] + (1-\beta)\ \log\left[\frac{1-\beta}{\alpha}\right]}{\frac{\theta_1}{1-\gamma\theta_1}\log\left[\frac{\theta_1}{\theta_0}\right] - \frac{\theta_1-\theta_0}{1-\gamma\theta_1}}.$$

When $\theta = 0$, $L(\theta) = 1$ and

$$E_0(N) = \frac{\log\left[\frac{\beta}{1-\alpha}\right]}{\theta_0-\theta_1}.$$

The work in this section can also be carried out for a composite hypothesis H_0: $\theta \leq \theta_0$ against an alternative hypothesis H_a: $\theta > \theta_0$. This is done by reducing the composite hypotheses to simple hypotheses of the form (6.7.1).

CHAPTER VII

ESTIMATION FOR GENERALIZED NEGATIVE BINOMIAL DISTRIBUTION

7.1 PROBLEMS INVOLVED

The GNB distribution in (1.2.1) is a three-parameter distribution. Jain and Consul (1971) applied the method of moments to estimate all the three parameters of the distribution. Charalambides (1974) considered a left truncated GNB distribution, truncated at point r, $r \geq 1$ and obtained the minimum variance unbiased estimators of functions of θ both for the cases when r is known and when r is unknown. The work of Kumar and Consul (1980) has been referenced in Chapter III.

The method of ML estimation was proposed as a general method of estimation by Fisher (1912). Later on, rigorous proofs of the asymptotic properties of the ML estimators were given in the works of Cramer (1946), Huzurbazar (1948) and Chanda (1954). Among the desirable properties of an estimator are

(i) consistency

(ii) efficiency and

(iii) unbiasedness.

One of the problems of an ML estimator is that it fails to satisfy some of the above desirable properties. Neyman and Scott (1948) and Kraft and LeCam (1956) have pointed out situations where the ML estimator is inefficient and not consistent. Their examples involve sampling from associated populations, that is, distinct but related populations. Thus, if we have a situation in which observations do not come from only the GNB population but also from a distinct but related population, we may end up with ML estimators which fail to satisfy some of the desirable properties.

When we have a uniparameter distribution, solving of likelihood equation may not be all that difficult. For two parameters, there may be problems if one (or all) of the likelihood equations is (or are) not well behaved. In general, as the number of parameters increases, the problems of solving the likelihood equations become more difficult. Therefore, obtaining the ML estimators in closed forms may prove to be a formidable task. This is in fact a problem in the ML estimation of the parameters in the GNBD. The expression in the likelihood equation looks horrible and cannot be easily evaluated. Another problem in the ML estimation of the parameters m, θ and β is that the parameters θ and β depend on one another on the boundary. This makes differentiation with respect to θ or β very difficult when both are unknown.

After getting the likelihood equations, some of which are ill-conditioned, one cannot talk about the existence of unique solutions. This has been a problem in the field of estimation of parameters in statistics. Consul and Shoukri (1984) proved a particular theorem for the existence of unique admissible ML estimators for the parameters of the GPD given in (1.3.1). Such a proof has not been given for the existence of unique ML estimators for the parameters of the GNB distribution.

Given a set of ML estimators, it will be of interest to know how well these estimators perform among other estimators from other methods. This can be measured by comparing their biases and relative efficiencies. The relative efficiency E is computed from

$$E = [(\text{generalized variance})\ (\text{information determinant})]^{-1}.$$

To compute the generalized variance, one needs the variances and covariances of the estimators. In this case of GNB distribution, it is very difficult to get the exact value of the generalized variance. Therefore, one has to take recourse to the asymptotic value.

Shoukri (1980) considered the simultaneous estimation of the parameters m and θ (with β known) by using the methods of moments and maximum likelihood estimation. He obtained the asymptotic biases, variances and covariance of the two sets of estimators by appealing to the method proposed by Shenton and Wallington (1962) for the negative binomial distribution. Shoukri's work involved very complicated expressions and the results obtained were more or less messy. If the expressions obtained for the estimation of two parameters are so messy, one begins to wonder as to how the expressions for the case of three parameters will look like.

Shoukri (1980) further obtained the relative asymptotic efficiencies of the ML and moment estimators of m and θ. The relative efficiencies were based on the first order biases, variances and covariance of these estimators. As a matter of fact, it is somehow difficult to give a reliable conclusion from any result based on first order approximations without any information on the behaviour of second order terms of biases, variances and covariance. This is the main reason why it is difficult to determine the subregion of the parameter space in which one estimator is superior to the other.

By examining the tables of relative efficiency of moment estimators with maximum likelihood estimators, it is noticed that the moment estimators are good when m is very small, say 1.0, θ is near 0.05 and $\beta < 10.0$. Thus, the moment estimators are good in a very small region of the parameter space if one is to rely on the results from first order approximations only.

The information determinant in the formula for finding the relative efficiency is not easy to obtain for the GNB distribution. An indication of this was given by Shenton (1949), who stated that the chief difficulty in computing the efficiencies in multiparameter distributions appeared to be the evaluation of the information determinant.

There is not much problem encountered in solving the equations that result from the moment method. The moment estimators are inefficient and biased. To evaluate the exact variances and covariances as well as the biases is a task that no one has attempted.

7.2 SOME UNSOLVED PROBLEMS

The following are some unsolved problems which require further research work in the estimation of the GNB distribution.

1. The existence of unique and admissible maximum likelihood estimators for m and θ, when β is assumed known, is still an open question.
2. The asymptotic biases, variances and covariance of the above ML estimators up to the second order are yet to be obtained. The same problem exists for the moment estimators. If these can be obtained, they will lead to a better assessment of the efficiencies of these estimators over the whole parameter space.
3. No work has been done on using the ML estimation method for estimating the three parameters simultaneously. This, however, may not be unconnected with the fact that the parameters θ and β are dependent on the boundary. Accordingly, the differentiation of the likelihood function with respect to θ or β becomes very difficult. The question of their uniqueness, biases, variances and covariances also remains open.
4. In view of the fact that the ML estimators have not been

proved to be unique and that the moment estimators seemed to be good in a small region of the parameter space, other methods of estimation are worth considering. These other methods which may be applied to the simultaneous estimation of the three parameters are:

(i) the method of the first two moments and the observed proportion of 'zeroes' and

(ii) the method of the first two moments and ratio of 'one' and 'zero' frequencies.

REFERENCES

1. Ali-Amidi (1978): "A note on the moments of the generalized negative binomial distribution and on certain properties of this distribution". SIAM J. Appl. Math. 34(2), 223-224.

2. Barndorff-Nielsen, I. (1976): "Plausibility inference". J. Roy. Statist. Soc. Ser. B 38, 103-131.

3. Bartlett, M.S. (1953a): "Approximate confidence intervals". Biometrika 40, 12-19.

4. Bartlett, M.S. (1953b): "Approximate confidence intervals II. More than one unknown parameter". Biometrika 40, 306-317.

5. Blischke, W.R. (1963): "Mixtures of discrete distributions." *Classical and contagious discrete distributions*, edited by G.P. Patil, Statistical Pub. Soc., Calcutta, pp.351-372.

6. Cassie, R.M. (1962): "Frequency distribution models in the ecology of plankton and other organisms". J. Animal Ecology 31, 65-92.

7. Chanda, K.C. (1954): "A note on the consistency and maxima of the roots of likelihood equations". Biometrika 41, 56-61.

8. Charalambides, C.A. (1974): "Minimum variance unbiased estimation for a class of left truncated discrete distributions". Sankhya A36, 397-418.

9. Consul, P.C. (1981): "Relation of modified power series distributions to Lagrangian probability distributions". Comm. Stat. - Theor. Meth. A 10(20), 2039-2046.

10. Consul, P.C. (1986): "On some models leading to the generalized Poisson distribution." (submitted).

11. Consul, P.C. and Gupta, R.C. (1980): "The generalized negative Binomial distribution and its characterization by zero regression". SIAM J. Appl. Math. 39(2), 231-237.

12. Consul, P.C. and Jain, G.C. (1973): "A generalization of the Poisson distribution". Technometrics 15, 791-799.

13. Consul, P.C. and Shenton, L.R. (1972): "Use of Lagrange expansion for generating discrete generalized probability distributions". SIAM J. Appl. Math. 23(2), 239-248.

14. Consul, P.C. and Shenton, L.R. (1973): "Some interesting properties of Lagrangian distributions". Comm. Stat. 2(3), 263-272.

15. Consul, P.C. and Shoukri, M.M. (1984): "Maximum likelihood estimation for the generalized Poisson distribution." Comm. Stat. - Theor. Meth. A13(12), 1533-1547.

16. Consul, P.C. and Shoukri, M.M. (1985): "The generalized Poisson distribution when the sample mean is larger than the sample variance". Comm. Stat. Simula. and Computa. 14(3), 667-681.

17. Cramer, H. (1946): *Mathematical methods of statistics.* Princeton University Press.

18. Cryer, J.D. and Robertson, T. (1975): "Isotonic estimation of the probability of extinction of a branching process." JASA 70, 905-912.

19. Dyczka, W. (1978): "A generalized negative binomial distribution as a distribution of power series distribution class." Zeszyty Nauk. Politech. Matematyka Z. 10, 5-14.

20. Fazal, S.S. (1977): "A test for a generalized Poisson distribution." Biometrical Journal 19(4), 245-251.

21. Fisher, R.A. (1912): "On an absolute criterion for fitting frequency curves." Messenger of Mathematics 41, 155-160.

22. Good, I.J. (1960): "Generalizations to several variables of Lagrange's expansion with applications to stochastic processes." Proc. Camb. Philos. Soc. 56(4), 367-380.

23. Good, I.J. (1965): "The generalization of Lagrange's expansion and the enumeration of trees." Proc. Camb. Philos. Soc. 61, 499-517.

24. Good, I.J. (1975): "The Lagrange distributions and branching processes." SIAM J. Appl. Math. 28, 270-275.

25. Gupta, R.C. (1974): "Modified power series distribution and some of its applications". Sankhya Ser. B 36, 288-298.

26. Gupta, R.C. (1975): "Maximum likelihood estimation of a modified power series distribution and some of its applications". Comm. Stat. 4(7), 689-697.

27. Gupta, R.C. (1976): "Distribution of the sum of independent generalized logarithmic series variables". Comm. Stat. - Theor. Meth. A 5(1), 45-48.

28. Gupta, R.C. (1977): "Minimum variance unbiased estimation in a modified power series distribution and some of its applications". Comm. Stat. - Theor. Meth. A6 (10), 977-991.

29. Gupta, R.C. and Singh, J. (1982): "Estimation of probabilities in the class of modified power series distributions". Math. Operationsforsch. Statist., Ser. Statistics 13(1), 71-77.

30. Huzurbazar, V.S. (1948): "The likelihood equation, consistency and maxima of the likelihood function". Annals of Eugen. Lond. 14, 185-200.

31. Jain, G.C. (1975): "On power series distributions associated with Lagrange expansion." Biom. Z. 17, 85-97.

32. Jain, G.C. and Consul, P.C. (1971): "A generalized negative binomial distribution". SIAM J. Appl. Math. 21(4), 501-513.

33. Jain, G.C. and Gupta, R.P. (1973): "A logarithmic type distribution". Trab. Estadist. 24, 99-105.

34. Jain, G.C. and Singh, N. (1975): "On bivariate power series distributions associated with Lagrange expansion." JASA 70, No.352, 951-954.

35. Janardan, K.G., Kerster, H.W. and Schaeffer, D.J. (1979): "Biological applications of the Lagrangian Poisson distribution". Bioscience 29, 599-602.

36. Janardan, K.G. and Schaeffer, D.J. (1977): "Models for the analysis of chromosomal aberrations in human leukocytes". Biometrical Journal 19(8), 599-612.

37. Johnson, N.L. and Kotz, S. (1969): *Discrete distributions*. John Wiley and Sons, New York.

38. Kalbfleisch, J.G. (1979): *Probability and statistical inference II*. Springer-Verlag, New York.

39. Kendall, M. and Stuart, A. (1977): *The advanced theory of statistics*, Vol.2, 4th ed. Macmillan Pub. Co. Inc., New York.

40. Khatri, C.G. (1959): "On certain properties of power series distributions". Biometrika 46, 486-490.

41. Keilson, J. and Gerber, H. (1971): "Some results for discrete unimodality". JASA 66, 386-389.

42. Kemp, A.W. and Kemp, C.D. (1968): "On a distribution associated with certain stochastic processes". J. Roy. Statist. Soc. B 30, 160-163.

43. Kosambi, D.D. (1949): "Characteristic properties of series distributions". Proc. Nat. Inst. Sci. India 15, 109-113.

44. Kraft, C. and LeCam, L. (1956): "A remark on the roots of the maximum likelihood equation". Ann. Math. Statist. 27, 1174-1176.

45. Kumar, A. (1981): "Some applications of Lagrangian distributions in queueing theory and epidemiology". Comm. Stat. - Theor. Meth. A10(14), 1429-1436.

46. Kumar, A. and Consul, P.C. (1979): "Negative moments of a modified power series distribution and bias of the maximum likelihood estimator". Comm. Stat. -Theor. Meth. A8(2), 151-166.

47. Kumar, A. and Consul, P.C. (1980): "Minimum variance unbiased estimation for modified power series distribution". Comm. Stat. - Theor. Meth. A9(12), 1261-1275.

48. Lehmann, E.L. (1959): *Testing statistical hypotheses*. John Wiley and Sons, New York.

49. Lüders, R. (1934): "Die statistik der seltenen ereignisse". Biometrika 26, 108-128.

50. Medgyessy, P. (1977): *Decomposition of superpositions of density functions and discrete distributions*. John Wiley and Sons, New York.

51. Mohanty, S.G. (1977): *Lattice path counting and applications*. Academic Press, Inc., New York.

52. Neyman, J. and Scott, E.L. (1948): "Consistent estimates based on partially consistent observations". Econometrica 16, 1-32.

53. Noack, A. (1950): "A class of random variables with discrete distributions". Ann. Math. Statist. 21, 127-132.

54. Otter, R. (1948): "The number of trees". Ann. of Math. 49, 583-599.

55. Pakes, A.G. and Speed, T.P. (1977): "Lagrange distributions and their limit theorems". SIAM J. Appl. Math. 32(4), 745-754.

56. Patel, J.K., Kapadia, C.H. and Owen, D.B. (1976): *Handbook of statistical distributions*. Marcel Dekker Inc., New York.

57. Patil, G.P. (1962): "Certain properties of the generalized power series distribution". Ann. Inst. Stat. Math., Tokyo 14, 179-182.

58. Patil, G.P. (1963): "On the multivariate generalized power series distribution and its application to the multinomial and negative multinomial". *Classical and contagious discrete distributions*, edited by G.P. Patil, Statistical Pub. Soc., Calcutta, pp. 183-194.

59. Prakasa Rao, B.L.S. (1969): "Estimation of a unimodal density". Sankhya Ser. A 31, 23-26.

60. Roy, J. and Mitra, S.K. (1957): "Unbiased minimum variance estimation in a class of discrete distributions". Sankhya 18, 371-378.

61. Robertson, T. (1967): "On estimating a density which is measurable with respect to a σ lattice". Ann. Math. Statist. 38, 482-493.

62. Shenton, L.R. (1949): "On the efficiency of the method of moments and Neyman's type A distribution". Biometrika 36, 450-454.

63. Shenton, L.R. and Wallington, P.A. (1962): "The bias of moment estimators with an application to the negative binomial distribution". Biometrika 49, 193-204.

64. Shoukri, M.M. (1980): *Estimation of generalized distributions*. Unpublished Ph.D. Thesis, University of Calgary, Calgary, Alberta, Canada.

65. Skellam, J.G. (1952): "Studies in statistical ecology I. Spatial pattern". Biometrika 39. 346-362.

66. Steutel, F.W. and van Harn, K. (1979): "Discrete analogues of self-decomposability and stability." Ann. of Prob. 7(5), 893-899.

67. Tweedie, M.C.K. and Veevers, A. (1968): "The inversion of cumulant operators for power-series distributions, and the approximate stabilization of variance by transformations". JASA 63, 321-328.

68. Wald, A. (1945): "Sequential tests of statistical hypotheses". Ann. Math. Statist. 16, 117-186.

69. Wald, A. (1947): *Sequential analysis*. John Wiley and Sons, New York.

70. Wald, A. (1950): *Statistical decision functions*. John Wiley and Sons, New York.

71. Wegman, E.J. (1970a): "Maximum likelihood estimation of a unimodal density function". Ann. Math. Statist. 41, 457-471.

72. Wegman, E.J. (1970b): "Maximum likelihood estimation of a unimodal density II". Ann. Math. Statist. 41, 2169-2174.

73. Wegman, E.J. (1972): "Nonparametric probability density estimation: I. A summary of available methods". Technometrics 14, 533-546.

74. Whittaker, E.T. and Watson, G.N. (1927): *A course of modern analysis*. Camb. Univ. Press.