UNLV Theses, Dissertations, Professional Papers, and Capstones

2009

Statistical inferences for functions of parameters of several pareto and exponential populations with application in data traffic

Sumith Gunasekera
University of Nevada Las Vegas

Follow this and additional works at: https://digitalscholarship.unlv.edu/thesesdissertations

Part of the Applied Mathematics Commons, Applied Statistics Commons, and the Statistical Methodology Commons

STATISTICAL INFERENCES FOR FUNCTIONS OF PARAMETERS OF

SEVERAL PARETO AND EXPONENTIAL POPULATIONS

WITH APPLICATIONS IN DATA TRAFFIC

by

Sumith Gunasekera

Bachelor of Science
University of Colombo, Sri Lanka
1995

A dissertation submitted in partial fulfillment
of the requirements for the

Doctor of Philosophy Degree in Mathematical Sciences
Department of Mathematical Sciences
College of Sciences

Graduate College
University of Nevada, Las Vegas
August 2009

.

.

ABSTRACT

**Statistical Inferences for Functions of Parameters of Several Pareto
and Exponential Populations with Applications in
Data Traffic**

by

Sumith Gunasekera

Dr. Malwane M. A. Ananda, Examination Committee Chair
Professor of Statistics
University of Nevada, Las Vegas

In this dissertation, we discuss the usability and applicability of three statistical

inferential frameworks – namely, the Classical Method, which is sometimes referred

to as the Conventional or the Frequentist Method, based on the approximate large

sample approach, the Generalized Variable Method based on the exact generalized

p-value approach, and the Bayesian Method based on prior densities – for solving ex-

isting problems in the area of parametric estimation. These inference procedures are

discussed through Pareto and exponential distributions that are widely used to model

positive random variables relevant to social, scientific, actuarial, insurance, finance,

investments, banking, and many other types of observable phenomena. Furthermore,

several Pareto and exponential populations, and the combination of several Pareto

and exponential distributions are widely used in the Computer Networking and Data

Transmission to model Self-Similar (SS) or Lthe ong-Range-Dependent (LRD) net-

work traffic that can be generated by multiplexing several Pareto and exponentially

distributed *ON/OFF* sources. One of the problems of interest in this dissertation is statistical inferences concerning common scale and common shape parameters of several Pareto distributions, and common location and common shape parameters of several exponential distributions based on the generalized p-value approach introduced by Tsui and Weerahandi where traditional frequentist or classical approaches do not provide useful solutions for the problems in the face of nuisance parameters. In this regard, we have developed exact tests and confidence intervals for common scale and common shape parameters of Pareto populations, and common location and common shape parameters of several exponential populations using ideas of generalized p-values and generalized confidence intervals. The resulting procedures are easy to compute and are applicable to small samples. We have also compared this test to a large sample test. Examples are given in order to illustrate results. In particular, using examples, it is pointed out that simply comparing classical and generalized p-values can produce a different conclusion that generalized pivotal quantities and generalized confidence intervals have proved to be very useful tools for making inference in practical problems. Furthermore, the Bayesian approach for the above problem is presented using the Gibbs sampling technique when shape parameters of several Pareto distributions and scale parameters of several exponential distributions are unknown. Their outcomes are compared with results based on classical and generalized approaches. The generalized inferential results derived for several Pareto and exponential populations are utilized extensively in finding exact solutions, as opposed to approximate solutions, for complicated functions of parameters of Pareto and exponential populations that are found in Computer Networking and Data Transmission. The Offered

Optical Network Unit Load (OOL), which is a direct result of the transmission of data files, generated at the Optical Network Units (ONUs) is discussed at length, through various aspects of inferential techniques, to find exact and non-misleading solutions to provide attractive, fast, reliable, and sophisticated online service to the customers. Network traffic flows generated by Hyper Text Transfer Protocol (HTTP), File Transfer Protocol (FTP), Variable-Bit-Rate (VBR), and Video Applications are injected into the system to simulate the system. Most of the simulations and real experiments described in this dissertation were performed with the self-similar traffic. The self-similar traffic is generated by aggregating the cumulative packet count at a certain time of multiple substreams, each consisting of alternating Pareto *ON*-and *OFF*-periods or exponentially distributed *ON*-and *OFF*-periods. These periods modeled by the fractional Brownian motion or the fractional Gaussian noise exhibit a time series whose process is characterized by the stochastic process. Detailed statistical inferences based on the classical framework, the generalized framework, and the Bayesian framework for the Offered Optical Network Unit Load (OOL) and the other related Computer Networking physical quantities are discussed. Examples are given through real data in order to illustrate the newly introduced the Generalized Variable Method procedure. A limited simulation study is given to demonstrate the performance of the proposed procedure.

.

TABLE OF CONTENTS

.

LIST OF TABLES

ix

.

LIST OF FIGURES

.

ACKNOWLEDGMENTS

This dissertation could not have been written without the support and friendship found at the University of Nevada-Las Vegas and elsewhere. The love of my family and friends provided my inspiration and was my driving force. It has been a long journey and completeing this work is definitely a high point in my academic career. I could not have come this far without the assistance of my family individuals and I want to express my deepest appreciation to them.

I was fortunate to gain a mentor in Professor Malwane M. A. Ananda. Ironically, it was he who started me on my path as a Statistician while I was getting ready for the graduate studies in Physics. His encouragement and advice led me to the University of Nevada-Las Vegas. I feel blessed to have worked with him. Without his belief in me I could not have developed confidence in my abilities as a Statistician and for that I am truly grateful. I have learned a great deal from him and I will never forget the valuable lessons he taught me.

The faculty and staff in the Department of Mathematical Sciences at the University of Nevada-Las Vegas are the most dedicated and generous people that I have ever met and I feel honored to have worked with them. Their guidance has served me well and I owe them my heartful appreciation. My committee members deserve a special note of praise, for they have watched over me since my first days as a graduate student. I wish to thank Professor Chih-Hsiang Ho, Professor Hokwon Anthony Cho,

.

CHAPTER I

INTRODUCTION

1.1 The background of exact statistical methods

Exact statistics has a history dated back to Fisher's era. Fisher's exact test – the first ever exact test in statistical history – has been playing a vital role in making inferences for parameters of interest. He discussed about the exact tests on the interpretation of χ^2 from contingency tables (Fisher 1922) claiming that they are based on the sampling distributions that are conditional on the marginals. As these exact test variables are functions only of random vector (random sample) and the parameter of interest, they are referred to as the classical exact test procedures. When statistical inferences are performed, these exact tests provide more reliable and accurate results outperforming procedures based on classical asymptotic and approximate statistical inference methods. Even though these exact tests frequently provide non-misleading solutions, sometimes they tend to provide misleading solutions too. In the late 80's this concept of classical exact test was broadened and widened by the introduction of the generalized test procedure by Tsui and Weerahandi as an extension to Fisher's exact tests. The most prominent and major characteristic of classical or generalized exact methods is that statistical inferences are mainly based on exact probability statements that are valid for any sample size. When the sample size is small, the asymptotic and other approximate results may lead to unreliable and

misleading conclusions. There are two branches in exact statistics as in approximate or asymptotic statistics:

1. exact parametric procedures where statistical inferences are performed under any parametric distributions – all assumptions of the distribution of the test statistic have to be met, and

2. exact nonparametric procedures where any distributional assumptions are not made.

Prompted by a conversation he had with Miss B. Muriel Bristol-Roach who was an alga biologist at the Rothamsted Experimental Station, Hertfordshire, England in 1919 about whether the tea or milk was added first to her cup, Sir Ronald Aylmer Fisher devised a comment from her to come up with the idea of "Exact Test". This exact test, introduced for the first time in statistical history, has been used in the analysis of contingency tables where sample sizes are small. When the cell counts are small – specifically, if more than twenty percent of the cells, when marginal totals are fixed, have an expected count that is less than five – the χ^2 distribution may not a suitable distributional candidate of the Pearson C^2 statistics or Likelihood Ratio G^2 statistics for testing independence of row and column variables. Such a situation is easily remedied by the Fisher's exact test.

Inspired by the Fisher's original treatment of hypothesis testing statistics (Fisher 1954), Weerahandi searched for an extreme region, an unbiased subset of sample space formed by minimal sufficient statistics having observed sample points on its boundary, to generalize the existing p-values to come up with exact solutions for different problems arise in hypothesis testing. For exact tests, readers are referred

to Fisher (1922), Weerahandi (1995, 2005), Metha (1995), Patel (1997), and many others.

Throughout this dissertation, the "Generalized Variable Method", an exact procedure introduced by Tsui and Weerahandi (1989) in making inferences of parameter(s) or functions of parameters, is used intensively to find exact solutions for the following problems:

1. Inferences of common scale and common shape parameters of several Pareto populations,

2. Inferences of common location and common scale parameters of several exponential populations,

3. Inferences of the Offered Optical Network Unit Load in Computer Networking and Data Transmission.

Keeping the original idea of the "Exact Test" introduced by Fisher, Tsui and Weerahandi (1989, 1993) generalized, based on exact probability statements, p-values as well as confidence intervals to remedy and overcome drawbacks of other conventional exact and approximate inference methods. Conventional methods alone do not always provide exact solutions to:

1. problems involving nuisance parameters such as that of comparing the means of two exponential distributions and making inferences of the second moments of a random variable whose underlying distribution is normal,

2. problems of making inferences of complicated functions of parameters of underlying distributions such as the Offered Optical Network Unit Load in

Data Transmission,

3. problems of making inferences in the face of small samples, especially that

   are found in biomedical researches.

Statisticians have always been using asymptotic procedures based on the large sample method to find solutions for above mentioned problems. But, this newly developed promising approach, the generalized variable method, provides exact solutions for such drastic, difficult, and intrigue problems.

Readers are referred to Weerahandi (1987, 1991, 1995, 2004), Weerahandi and Amaratunga (1999), Gamage and Weerahandi (1998), Weerahandi and Berger (1999), Ananda and Weerahandi (1997), Ananda (1995, 1998, and 1999), Gunasekera and Ananda (2009), Tian and Wu (2007), Krishnamoorthy and Lu (2003), Tian and Cappelleri (2004), Zhou and Mathew (1994), Thursby (1992), Griffiths and Judge (1992), Koschat and Weerahandi (1992) for application of the generalized p-value and confidence interval in various practical problems.

1.2 Motivation to perform generalized statistical inferences

1.2.1 Several Pareto distributions

The Power-law, fractal, right-skewed Pareto distribution that we present in the Chapter III of this dissertation is widely used to describe the distribution of positive random variables related to data found in different types of observable phenomena.

The classical (sometimes referred to as 'conventional, approximate, asymptotic, or frequentist') point and interval estimations of the scale parameter and the shape parameter (this is sometimes referred to as the "location parameter") of the Pareto

4

distribution have so far been extensively studied: *the Univariate Pareto* by Saksena and Johnson (1984), Rytgaard (1990), Quandt (1966), Chen (1996), Brazauskas and Serfling (2003), Moothathu (1990), and Malik (1970) ; *the Generalized Pareto* by Peng and Welsh (2001), Mahmoud, Sultan, and Moshref (2005) ; *the Mixtures of Pareto* by Moothathu (1993) , *Two Pareto Populations* by Rohatgi and Saleh (1987), *Two Multivariate Pareto* by Yeh (2000), and *Several Pareto* by De and Liu (1992), Baklizi (2002), Elfessi and Jin (1996), and Jin and Elfessi (2001).

Most of the above research are mainly based on classical approximate procedures; however, some of them are based on classical exact approaches with known nuisance parameters. Inspired by such work, the generalized variable procedure is intensively utilized to find exact solutions for common scale and common shape parameters of several Pareto populations. Practitioners in Computer Networking make use of several or combination of Pareto distributions intensively and heavily to, connecting theory with real world applications, model multiplexed (muxed) or demultiplexed (demuxed) traffic data and the data arise from other Computer Networking physical phenomena.

1.2.2 Several exponential distributions

The exponential distribution, another right-skewed distribution, discussed in Chapter IV of this dissertation is heavily used in describing the lengths of the inter-arrival times, waiting times, serving times, inter-leaving times, etc. that can be found in homogeneous Poisson process. This distribution considered as a continuous counterpart of the geometric distribution which describes the number of Bernoulli trials necessary for a discrete process to change state. Here, exponential distribution is used to describe the time for a continuous process to change state.

Extensive studies on the classical approximate and exact inferences of the common scale parameter and the common location parameter of several exponential distribution have so far been performed by many practitioners and researchers. Few of those works are listed below:

Pal and Sinha (1990), Costanza, Hamdy, and Son (1986), Lin and Shen (1997), Baklizi (2004), Madi and Tsui (1990), Chiou and Cohen (1984), Handa and Cambo (1999, 2005), Manna (2000), Akahira (1987), Moothathu (1993), Jin and Crouse (1998), Bai and Hong (1992), Cramer and Kamps (1997), Kanfer and Geertsema (2001), Carpenter and Pal (1995), Thomas and Kumar (2005), Sun and Sinha (1999), and Ghosh and Razmpour (1984).

As in the several Pareto populations case, the generalized p-value approach is used to remedy and overcome the difficulty of making exact inferences for common scale and common location parameters of several exponential populations. Prompted by the above mentioned work that are mostly based on approximate procedures except for few, the generalized variable method is utilized to find exact solutions in the face of nuisance parameters for common scale and common location parameters of several exponential populations. This setting of distribution is also a desirable candidate to model the traffic data and other related Computer Networking parameters.

1.2.3 WWW-and WAP-traffic data

When advanced internet technology is at an unprecedented peak, statistics plays a vital and dynamic role to provide an efficient, reliable, fast, and quality service to internet users. Thus, right-skewed Pareto and exponential distributions play very distinct and extraordinary roles in modeling data (sometimes referred to as 'files,

packets, or bytes') transmitted via Computer Networking systems that are designed to deliver multiple services and applications such as:

1. Voice Communications (Chatting over internet),

2. Standard and High-Definition Video (Movies, TV Shows, News Video Clips, etc.),

4. Video Games (Computer-Aided Designs),

4. Video Conferencing (discussions over internet),

5. WWW (World Wide Web)-data Traffic: web pages visited through Personal Computers (PCs) and Notebooks (laptops), and

6. WAP (Wireless Application Protocol)-data Traffic: web pages visited through hand-held electronic devices such as palmtops, i-phones, smart phones (blackberries), etc.

Pareto and exponential distributions are used mainly as file size ('bursts' or 'packet') distributions of internet traffic – which uses the Transmission Control Protocol (TCP) with the Internet Protocol (IP), the main protocol of the Internet. Many smaller-size files (e.g. files in kilobytes) with few larger ones (e.g. files in Megabytes, Gigabytes, or Terabytes) can be modeled by Pareto, exponential, and lognormal distributions, or by the combination of those distributions. Intensive studies on the suitable distributional candidate for modeling file sizes, transfer times, and burst lengths have been done by Downey (2005).

In Chapter V, the Offered Optical Network Unit Load (OOL), which is a direct result of the transmission of data files, generated at the Optical Network Units (ONUs) is discussed at length, through various aspects of inferential techniques, to find exact

and non-misleading solutions to provide attractive, fast, reliable, and sophisticated online service to the customers.

In addition to files sizes, the web-session interarrival time, the number of packet calls (or sometimes referred to as 'web-pages') per web-session, the reading time between web-pages, the number of items per web-page, the time interval between items in the same web-page, the web-item size on uplink–the substream from the computer to the central office, the web-item size on downlink–the substream from the central office to the computer, the time interval between two consecutive uplink packets inside an item, the time interval from uplink to downlink packet inside an item, the time interval between two consecutive downlink packets inside an item, and the time interval from downlink to uplink packet inside an item of web-traffic data are also modeled by several Pareto and exponential distributions.

Furthermore, the WAP-session interarrival time, the number of pages per web-session, the reading time between packet calls (sometimes referred to as 'WAP-items'), the WAP-item size on uplink, the WAP-item size on downlink, the transmission time of the uplink packets, the processing time of WAP-request, the web-transaction waiting time, the processing time of WAP-response, the transmission time of the downlink packets (sometimes referred to as 'WAP-response'), the acknowledgment time on uplink, and the acknowledgment time on downlink of WAP- traffic data are also extensively and exclusively modeled by several Pareto and exponential distributions.

The other well-known right-skewed parametric families of distributions such as the loglogistic (sometimes referred to as "Fisk distribution"), the lognormal, the Weibull, the inverse Weibull, the Burr (sometimes referred to as "the Burr Type XII distrib-

ution, the Singh-Maddala distribution, or the generalized log-logistic distribution"),
and the generalized Pareto are also considered as useful and suitable distributions
for modeling computer networking data. In addition, splicing, overlapping, superposition, merging, combining, and mixing of the existing parametric distributions can
also be used in modeling such data. Recently developed the lognormal-Pareto composite (LPC) distribution (Cooray and Ananda 2005) – which is a composite model
that takes the two-parameter lognormal density up to an unknown threshold value
and the two-parameter Pareto density for the rest of the model – with two unknown
parameters has been a more suitable candidate for modeling WWW-data traffic as
well as WAP-data traffic. This will be discussed thoroughly in future research in
order to find better solutions for the existing problems in Computer Networking and
Data Transmission.

.

CHAPTER II

THE GENERALIZED VARIABLE METHOD

2.1 Introduction

The *Generalized Variable Method* procedure, which is an attractive and robust *Exact Method*, was introduced by Tsui and Weerahandi (1989) as an extension to existing *Exact Methods*. The *Generalized p-Value* (Weerahandi 1989) and the *Generalized Confidence Interval* (Weerahandi 1993), direct derivations of the *Generalized Variable Method*, remedy and overcome difficulties of situations such as:

1. the limited availability of conventional *Fixed-Level Tests*,

2. the unavailability of *Exact Fixed-Level Tests* for comparing normal

   populations, and

3. the violation of assumption of equal variances in Analysis of Variance

   (ANOVA) and more complicated problems such as two-way ANOVA, the

   Multivariate Analysis of Variance (MANOVA), Mixed Models, and

   Repeated Measures Models including Crossover Experiments and Growth

   Curves.

In the application of comparing two regression models, Weerahandi (1987) gave the first introduction to the notion of the *Generalized p-Value* and showed that it is an exact probability of an unbiased extreme region, a well-defined subset of the

sample space formed by *Sufficient Statistics*. Having noticed the failure of the *Conventional Methods* that they do not always provide *Exact Solutions* to even simple problems involving nuisance parameters and to complicated functions of parameters, practitioners tend to work with asymptotic results to get *Approximate Solutions* for problems involved with nuisance parameters. This newly developed promising approach, the *Generalized Variable Method*, provides *Exact Solutions* for such drastic, difficult, intrigue problems. Ananda (1997), Ananda and Weerahandi (1997), Thursby (1992), Weerahandi and Johnson (1992), Gamage and Weerahandi (1998), and Park and Burdick (2004), through simulation studies, proved that the *Generalized p-Value* procedure outperforms, in terms of *Size* and *Power*, the existing *Approximate Tests* that have so far been used to find solutions in more complicated situations. For a detailed, complete, and clear coverage and applications of these *Generalized Tests* and *Generalized Confidence Intervals*, the interested parties are referred to Weerahandi (1995, 2004).

Even though many practitioners, researchers, or authors have developed different type of exact solutions for different situations, their solutions ended up with some failures, flaws, or faults: Kempthorne and Folks' (1971) attempt to show the *Classical Testing* approach's failure was overshadowed by vague and implicit explanations of the definitions of their new approach. Furthermore, Bernard (1984), and Rice and Gaines'(1989) contribution to one of the prominent problems in linear models – the *Behrens-Fisher Problem* – also became an incomplete procedure because of the lack of formal definitions, derivations, and formulae for computing *Exact p-Values*.

Some practitioners have so far compared the *Generalized Variable Method* with

the *Bayesian Method*, observing a clear relation between them:

Introducing a *Bayesian p-Value* with the aid of noninformative priors (sometimes referred to as 'vague, diffuse, or flat' priors), Meng (1994) showed that it is numerically equivalent to the *Generalized p-Value*. Weerahandi and Tsui (1996) also showed that the *Bayesian p-Values* numerically equivalent to the *Generalized p-Values*.

Furthermore, when existing exact procedures as well as the approximate procedures fail to find exact solutions, the *Generalized Variable Method* is heavily used in complicated models such as

1. discrete variable models,

2. categorical variable models,

3. nonlinear models, and

4. models based on non-normal distributions.

When the underlying family of distributions contains two or more unknown parameters, or parametric inferences are performed on complicated functions of parameters of underlying distributions, conventional tests are typically available only for special functions of the parameters. This is because it is not possible or easy to find test statistics having distributions free of nuisance parameters. *Exact Generalized Inferences* are involved with procedures of *Hypothesis Testings* and *Confidence Intervals* that are based on *Exact Probability Statements*. Throughout this dissertation problems of making *Inferences*, mainly based on the *Exact Parametric Procedures*, are discussed and their counterparts, *Exact Nonparametric Procedures*, are discussed in the future research.

In order to get a clear picture about the *Generalized Variable Method,* the defin-

itions of major components in *Statistical Inference Procedures* found in three major *Frameworks of Inferences* – the *Classical*, the *Generalized*, and the *Bayesian* – are given below:

1. the *Test Statistic*,

2. the *Pivotal Quantity*,

3. the *Test Variable*,

4. the *Extreme Region* or *Rejection Region*,

5. the *p-Value*

6. the *Fixed-Level Testing* or *Significance Testing*,

7. the *Size* of the *Test*: *Nominal (Intended) Size* and *True (Empirical/Actual) Size*,

8. the *Power Function*: without and after adjusting the *Size*,

9. the *Confidence Interval* or *Bound*, or *Credible Interval* or *Set*, and

10. the *Coverage Probability*: *Nominal (Intended) Coverage Probability* and *True (Empirical/Actual) Coverage Probability*.

2.2 The classical test statistic

Any real valued-valued function of an observed random sample of certain size taken from any distribution having nuisance parameters as well as parameter of interest is said to be a *Test Statistic*, if it satisfies the following properties:

1. the real valued-function must have a probability distribution that is free of nuisance parameters.

2. for given observed value of that function and the parameter of interest, the

distribution function of that real-valued function must be either stochastically increasing or decreasing function of parameter of interest.

The mathematical interpretation of such a *Test Statistic* is clearly presented in Weerahandi (*p. 27*, Weerahandi 1995; *p. 4*, Weerahandi 2004) as follows:

"A real-valued function of $\mathbf{x}^m$, denoted by $T^{(c)}(\mathbf{x}^m)$, where $\mathbf{x}^m = (x_1, x_2, ..., x_m)$ be the observed value of the random vector $\mathbf{X}^m = (X_1, X_2, ..., X_m)$ of size m taken from a distribution $F(X; \theta, \boldsymbol{\delta})$, θ being the parameter of interest and $\boldsymbol{\delta}$ a vector of nuisance parameters, is said to be a *Test Statistic* for θ, if it has the following properties:

1. $T^{(c)}(\mathbf{x}^m)$ has a probability distribution that is free of nuisance parameters.

2. given $t^{(c)}(\mathbf{x}^m), \theta$; $F_{T^{(c)}(\mathbf{x}^m)}(t^{(c)}(\mathbf{x}^m)) = Pr(T^{(c)}(\mathbf{x}^m) \leq t^{(c)}(\mathbf{x}^m))$, is a either stochastically increasing or decreasing function of θ."

Furthermore, a disadvantage (drawback) of this *Classical Test Statistic* is illustrated by citing an example:

Assume that X is normally distributed random variable with mean θ and variance σ^2. Therefore, from the *Central Limit Theorem*, the real-valued function of $\mathbf{x}$ is given by

$$T(\mathbf{x}) = \frac{\overline{x} - \theta_0}{\sigma/\sqrt{n}} \sim N(0, 1), \tag{2.1}$$

where θ_0 is a given value of θ, σ is a nuisance unknown parameter, and $N(a, b)$ stands for the normal distribution with mean a and variance b. Then, once the estimated variance S^2 replaces σ^2, this function drastically changes to

$$T(\mathbf{x}) = \frac{\overline{x} - \theta_0}{s/\sqrt{n}} \sim t_{df}; \tag{2.2}$$

14

$T(\mathbf{x})$ is free of nuisance parameters, and its distribution function is monotonically increasing function of θ; thus, it is considered as a *Test Statistic.* Here, t_{df} stands for the t-distribution with the degrees-of-freedom df and s is the observed value of S. The clear draw back here is that nuisance parameters have to be known, and if not to be estimated, something not necessary in the *Generalized Variable Method,* which we heavily and mainly, discuss and utilize throughout this dissertation.

Contrary to the *Classical Approximate Test* elaborated above, *Classical Exact Tests* that are based on *Probability Statements* such as:

1. the *Tippet's Method,*

2. the *Fisher's Method,*

3. the *Inverse Normal Method,* and

4. the *Logit Method*

have so far been discussed and have been active candidates for performing *Statistical Inferences* in statistical arena for a long period of time, providing the *Exact Solutions* for different statistical problems. However, major drawback of even on this method is related with the nuisance parameters: they have to be known. These draw backs are clearly illustrated in Chapter III and IV when the results based on this method are compared with that of the *Generalized Variable Method.*

The *Test Statistic* is not found in the *Generalized Variable Framework* of *Inference* because the *Generalized Tests Variable* is a function of random sample, its observed value, and all unknown parameters including nuisance parameters. *Bayesian* counter part of *Test Statistic* is also not available because *Bayesian Inferences* are performed mainly through the posterior distributions: the posterior distribution, the marginal

posterior distribution, or the conditional posterior distribution.

2.3 The classical pivotal quantity and test variable

A *Classical Pivotal Quantity* is a real-valued function of a random sample of certain size from any distribution as well as a parameter of interest whose distribution function is independent of parameter of interest.

This is clearly mathematically presented in (*p.*19 Weerahandi 1995) as follows:

" A *Classical Pivotal Quantity* $R^{(c)}(\mathbf{X}^m;\theta)$ is a real valued function of $\mathbf{X}^m$ and θ which satisfy the following property:

1. the distribution of the random variable $R^{(c)}(\mathbf{X}^m;\theta)$ does not depend

    upon θ."

Furthermore, Weerahandi (*pp.* 27-28, 1995) discussed that when there are no *Classical Pivotal Quantities* based on simple sufficient statistics are available, probability integral transform would be used in order to construct *Pivotal Quantities* as follows: Since

$$F_X(x;\theta) = U \sim UNIFORM(0,1), \qquad (2.3)$$

where $F_X(x;\theta)$ is *cdf* of a continuous random variable X and $UNIFORM(a,b)$ stands for the continuous uniform distribution with the minimum value parameter a and the maximum value parameter b,

$$-\log U \sim EXPONENTIAL(1) = GAMMA(1,1), \qquad (2.4)$$

where $EXPONENTIAL(\mu)$ stands for the exponential distribution with mean μ, and $GAMMA(a,b)$ for the gamma distribution with shape parameter a and scale parameter b.

Then the *Classical Pivotal Quantity* for testing hypotheses $\theta \leq \theta_0$ versus $\theta > \theta_0$, $\theta \geq \theta_0$ versus $\theta < \theta_0$, or $\theta = \theta_0$ versus $\theta \neq \theta_0$ as well as constructing confidence intervals for θ is given by

$$R^{(c)}(\mathbf{X}^m; \theta) = -\sum_{i=1}^{m} \log U_i = -\sum_{i=1}^{m} \log F_{X_i}(x_i; \theta) \sim GAMMA(m, 1),$$

(2.5)

where $X^m = (X_1, X_2, ..., X_m)$ is a random sample from $F_X(x; \theta)$. Moreover, a *Classical Pivotal Quantity* becomes a *Classical Test Variable* when the parameter of interest is substituted with a given value.

2.4 The generalized pivotal quantity and test variable

The *Generalized Pivotal Quantity* and *Test Variable* are considered to be the backbone of the *Generalized Variable Method*. They are functions of a random sample of any size taken from any parametric distribution, its observed value, and all the parameters – parameter of interest as well as nuisance parameters. Unlike in the *Classical Pivotal Quantity* or the *Test Variable*, the inclusion of nuisance parameters in the *Generalized Variable Method* paved the way for practitioners to come up with *Exact Solutions* for the problems when test variable based on *Sufficient Statistics* are unavailable.

Weerahandi (1995, 2004) discussed heavily and elaborately providing details about the *Generalized Pivotal Quantity*, which is a random variable that is a function of random sample of any size taken from any parametric distribution, its observed value as well as all the parameter satisfying the following two properties:

1. the observed value of the proposed random variable does not depend on

17

nuisance parameters.

2. the proposed random variable has a probability distribution that is free of unknown parameters.

A special attention has been given for this definition, and the mathematical aspect of this newly introduced technique, which use throughout this dissertation, is presented here as it appears in Weerahandi (*p.146, 1995; p.19, 2004*).

"A *Generalized Pivotal Quantity* is a random variable in the form $R^{(g)}(\mathbf{X}^m; \mathbf{x}^m, \theta, \boldsymbol{\delta})$, where

$\mathbf{X}^m = (X_1, X_2, ..., X_m)$; a random sample of size m from $F(X; \theta, \boldsymbol{\delta})$

$\mathbf{x}^m = (x_1, x_2, ..., x_m)$; observed value of $\mathbf{X}^m$,

$\theta =$ parameter of interest,

$\boldsymbol{\delta} =$ vector of nuisance parameters, and

having the following two properties:

1. The observed value of $R^{(g)}(\mathbf{X}^m; \mathbf{x}^m, \theta, \boldsymbol{\delta})$ is independent of of nuisance parameters.

2. When θ is specified, $R^{(g)}(\mathbf{X}^m; \mathbf{x}^m, \theta, \boldsymbol{\delta})$ has a probability distribution that is free of unknown parameters."

Then, a *Generalized Test Variable* directly derived from the *Generalized Pivotal Quantity* has also been discussed in Weerahandi (*p.115, 1995; p.8, 2004)* and is presented here for the mathematical tractability, consistency, and continuation of this discussion.

"A *Generalized Test Variable* is a random variable in the form of $T^{(g)}(\mathbf{X}^m; \mathbf{x}^m, \theta, \boldsymbol{\delta})$ that satisfy the following properties:

1. The observed value of $T^{(g)}(\mathbf{X}^m; \mathbf{x}^m, \theta, \boldsymbol{\delta})$, $t^{(g)}(\mathbf{x}^m; \mathbf{x}^m, \theta, \boldsymbol{\delta})$, does not depend on unknown parameters,

2. When θ is specified, $T^{(g)}(\mathbf{X}^m; \mathbf{x}^m, \theta, \boldsymbol{\delta})$ has a probability distribution that is free of nuisance parameters,

3. Given $t^{(g)}(\mathbf{x}^m; \mathbf{x}^m, \theta, \boldsymbol{\delta})$, $\mathbf{x}^m = (x_1, x_2, ..., x_m)$, and $\boldsymbol{\delta}$, the distribution function of $T^{(g)}(\mathbf{X}^m; \mathbf{x}^m, \theta, \boldsymbol{\delta})$ is a stochastically increasing or decreasing function of θ, ".

Furthermore, Weerahandi (1995, 2004) builds a bridge between the *Generalized Test Variable* and the *Generalized Pivotal Quantity* by connecting them through the parameter of interest, the parameter to be tested, or the parameter for which the confidence intervals are constructed as follows:

$$T^{(g)}(\mathbf{X}^m; \mathbf{x}^m, \theta, \boldsymbol{\delta}) = R^{(g)}(\mathbf{X}^m; \mathbf{x}^m, \theta, \boldsymbol{\delta}) - \theta \qquad (2.6)$$

2.5 The extreme region and critical region/rejection region

Even though an *Extreme Region* and a *Critical Region* seem to be the same, they are drastically different the way the testing of hypothesis for any parameter of interest is performed.

An *Extreme Region* is an unbiased subset of sample space having the observed value of the sample on the boundary of it – found in significance testing procedure practiced by giants in statistical arena such as Fisher, Gosset, and Pearson – presented as opposed to the *Rejection Region* that is an unbiased subset of sample space having the observed value of the sample outside of it – found in *Fixed-Level Testing* procedure.

For more details of the E*xtreme Region* as well as the *Critical Region/Rejection Region,* readers are referred to Weerahandi (1995, 2005).

2.6 The Substitution Method

Expressing all unknown parameters – nuisance parameters as well as the parameter of interest – in terms of *Sufficient Statistics* and the some random variables, Peterson, Berger, and Weerahandi (2003) defined a potential *Generalized Pivotal Quantity* by replacing sufficient statistics with their observed values. The method for which above mentioned procedures are adopted is called the *Substitution Method* in the *Generalized Variable Framework* of *Inference*. For more details, readers are referred to Weerahandi (2005).

2.7 The p-value

The probability of how well data supports or discredits the null hypothesis in *Statistical Testing Procedure* is a direct result of the combination of *Test Statistic*, *Test Variable* as well as the *Extreme Region* or the *Rejection Region*. Based on whether the *Classical Test Variable* or the *Generalized Test Variable* is used, it is referred to as either the *Classical* or *Generalized p-Value*, respectively. For more details readers are referred to Weerahandi (1995, 2005).

2.8 The size of the test

The *Size of the Test*, sometimes also referred to as

1. the *Level of Significance,*

2. the *Probability of Type-I Error Rate,*

3. the *Probability of α-error,* if α is the *Level of Significance*

20

4. the *Probability of False Positive,*

is the probability of rejecting a null hypothesis even when the it is true. The A*ctual (Empirical/True) Size of the Test* is sought for the classical as well as the the *Generalized Variable Framework* of *Inference* when the *Nominal (Intended) Size* is given.

2.9 The power of the test

Furthermore, *Power Function* – the probability of rejecting the null hypothesis when it is false – is also discussed along with the *Size* – without and after adjusting the *Size* – as it is connected with the *Size.* There are three types of *Power Functions:*

1. the *Power Function* of the *Classical Test Variable,*

2. the *Power Function of* the *Classical Test Statistic,*

3. the *Power Function of* the *Generalized Test Variable.*

For more details, readers are referred to Weerahandi (1995, 2005).

2.10 Confidence intervals and coverage probabilities

Once the *Test Variable* is developed, it is easily to construct *Confidence Intervals* based on that and to find *Empirical Coverage Probabilities* in *Classical* as well as *Generalized Variable Frameworks* of *Inferences.* For more details of:

1. *Classical Confidence Intervals,*

2. *Classical Coverage Probabilities,*

3. *Generalized Confidence Intervals,*

4. *Generalized Coverage Probabilities,*

which are discussed heavily throughout this dissertation, the interested parties are referred to Weerahandi (1995, 2005) or related literature found in the Reference.

.

CHAPTER III

THE GENERALIZED VARIABLE METHOD IN SEVERAL PARETO

POPULATIONS

3.1 Introduction

The positive-skewed, power-law, fractal, and heavy-tailed Pareto distribution (also referred to as the "Bradford distribution") has been defined in terms of the mode (which is also referred to as the 'scale' or the 'location' parameter) and a shape parameter. This distribution is used to model the data found in social, scientific, actuarial, insurance, finance, investments, banking, and many other fields. In the insurance field, the Pareto has been the most popular and the most used distribution to model payment data by setting the modal value as the minimum claim (deduction) and an infinitely large value as the maximum value (Klugman, Panjer, and Willmot 1998; Hogg and Klugman 1984). Meteorologists, climatologists, and weather experts also use this distribution to predict and forecast the climate and weather changes (Brabson and Palutikof 2000). The truncated shifted Pareto distribution is used to model size distributions of oil and gas fields for resource assessment (Houghton 1988). Furthermore, this distribution plays a vital role in Computer Networking and Data Transmission in modeling the sizes of files (Crovella and Bestavros 1996; Downey 2005). As files of larger-size are rarely and the smaller ones are quite often sent through the transmission lines, the file sizes are modeled with the Pareto distrib-

ution. The Pareto models with heavy tails have already been generalized and are discussed in the literature (Resnick 1997; Beirlant, Joossens, and Segers 2004). In addition, splicing, mixing, combining, superposition, composing of the existing Pareto distributions have also already been discussed (Klugman et al. 1998; Everitt and Hand 1981). Recently developed composite family of Pareto – the Lognomal-Pareto Composite (LPC) distribution by Cooray and Ananda (2005) – has drawn a special attention as it has become more suitable distributional candidate for modeling data found in Computer Networking and Data Transmission.

3.1.1 Types of the Pareto distribution

Four types of the Pareto distributions, based on the number of parameters involved in the distribution function, are available in the literature.

1. two-parameter: the European Pareto and the American Pareto

2. three-parameter: the shifted Pareto and the generalized Pareto

The European Pareto distribution:

The Pareto distribution of the random variable X_E with the scale parameter θ_E and the shape parameter α_E (i.e. $X_E \sim Pa_E(\theta_E, \alpha_E)$), and having the distribution function in the form of

$$F(x_E) = 1 - \left(\frac{\theta_E}{x_E}\right)^{\alpha_E}, \text{ where } \alpha_E > 0, \, 0 < \theta_E \leq x_E \qquad (3.1)$$

is referred to as the European Pareto distribution.

The American Pareto distribution:

The Pareto distribution of the random variable X_A with the scale parameter θ_A and the shape parameter α_A (i.e., $X_A \sim Pa_A(\theta_A, \alpha_A)$), and having the distribution

function in the form of

$$F(x_A) = 1 - \left(\frac{\theta_A}{\theta_A + x}\right)^{\alpha_A}, \quad \text{where } \alpha_A > 0, \ 0 \leq x_A \tag{3.2}$$

is referred to as the American Pareto distribution

The Shifted Pareto distribution:

The third version of the Pareto distribution of the random variable X_S with the scale parameter θ_S , the shift parameter β_S, and the shape parameter α_S (i.e. $X_S \sim Pa_S(\theta_S, \beta_S, \alpha_S)$), and having the distribution function in the form of

$$F(x_S) = 1 - \left(\frac{\beta_S + \theta_S}{\beta_S + x_S}\right)^{\alpha_S}, \quad \text{where } \theta_S, \beta_S, \alpha_S > 0, \ 0 < \theta_S \leq x_S \tag{3.3}$$

is referred to as the shifted Pareto distribution.

The generalized Pareto distribution:

The fourth type of the Pareto whose distribution function is in the form given below with location, scale, and shape parameters μ_G, θ_G, and α_G, respectively, is called the generalized Pareto distribution, where $X_G \sim Pa_G(\mu_G, \theta_G, \alpha_G)$:

$$F(x_G) = \begin{cases} 1 - \left(1 + \frac{x_G - \mu_G}{(\theta_G/\alpha_G)}\right)^{-\frac{1}{\alpha_G}} ; & \alpha_G, \mu_G \in \Re, \ \theta > 0, \ \mu_G \leq x_G \ (\alpha_G \geq 0), \\ & \theta_G \leq x_G \leq \theta_G/\alpha_G \ (\alpha_G < 0), \end{cases} \tag{3.4}$$

Note that, for our convenience, the European Pareto – which is henceforth referred to as 'Pareto' – is used throughout this dissertation.

In this Chapter, we consider $k \ (\geq 2)$ independent Pareto distributions with an unknown common scale parameter θ (sometimes referred to as the "location parameter" and also as the "truncation parameter") and unknown possibly unequal shape parameters α_i's $(i = 1, 2, ..., k)$.

Using the generalized variable approach (Tsui and Weerahandi 1989), we construct an exact test for testing θ. Furthermore, using the generalized confidence interval (Weerahandi 1993), we construct an exact confidence interval for θ as well. Simulation studies were carried out to compare the performance of these generalized procedures with that of the approximate procedures based on the large sample method. Moreover, performance of the classical exact test procedures based on the probability statements is compared with that of the newly introduced generalized variable approach.

The failure to draw exact statistical conclusions in the conventional framework of statistical inference - the classical approach – has laid down the foundation and paved the way for Tsui and Weerahandi (1989) to generalize the conventional definition of p-value so that the above mentioned problems can be easily resolved. As a common practice, even with small sample sizes, practitioners often resort to asymptotic methods that are overshadowed by the poor performance with these small samples. Once the generalized variable method is used, such poor performances are overcome by proving that the generalized p-value approach based on exact probability statements performs better than the approximate procedure based on the large sample approach performs.

Generalized inferential methods have now been successfully applied to obtain exact tests in various linear models: Weerahandi (1987), Thursby (1992), Griffiths and Judge (1992), and Koschat and Weerahandi (1992) for applications in *regressions*; Weerahandi (1991) and Zhou and Mathew (1994) for applications in *mixed models*; Weerahandi (1995) for applications in *one-way ANOVA*; Ananda and Weerahandi (1997), Gunasekera and Ananda (2009) for applications in *two-way ANOVA*; Ananda

25

(1998) for applications in *ANCOVA*; and Ananda (1995) for applications in *nested design*.

3.2 Classical inferences

The classical inferences of the scale and shape parameters of the Pareto distribution have so far been extensively studied by: Saksena and Johnson (1984), Rytgaard (1990), Quandt (1966), Chen (1996), Brazauskas and Serfling (2003), Moothathu (1990), and Malik (1970), Peng and Welsh (2001), Mahmoud, Sultan, and Moshref (2005), Moothathu (1993), Rohatgi and Saleh (1987), Yeh (2000), De and Liu (1992), Baklizi (2002), Elfessi and Jin (1996), and Jin and Elfessi (2001), and many others.

For a single Pareto distribution with the common scale parameter θ and the shape parameter α, Arnold (1983) described:

1. the confidence interval for θ, when α is known,

2. the confidence interval for α, when θ is known, and

3. the joint confidence region for θ and α, when both θ and α are unknown.

Using the certain classical independent tests based on the combination of probabilities: namely,

1. the Tippet method,

2. the Fisher method,

3. the inverse normal method, and

4. the logit method,

Baklizi (2002) constructed confidence intervals for θ.

3.3 Generalized inferences

The two-parameter Pareto distribution with the shape parameter α and the scale parameter θ has a cumulative distribution function given by

$$F(x) = \begin{cases} 1 - (\frac{\theta}{x})^\alpha & \text{if} \quad x \geq \theta \\ 0 & \text{if} \quad x < \theta \end{cases}, \tag{3.5}$$

where $\theta, \alpha > 0$ and $x \in [\theta, \infty)$.

The limiting case of the Pareto distribution is as follows:

$$\lim_{\alpha \to \infty} f(x) = \lim_{\alpha \to \infty} \frac{\alpha \theta^\alpha}{x^{\alpha+1}} = \delta(x - \theta),$$

where $\delta(x - \theta)$ is the Dirac delta function given by the heuristic definition

$$\delta(x - \theta) = \begin{cases} \infty, \ x = \theta \\ 0, \ x \neq \theta \end{cases},$$

constrained to satisfy the identity

$$\int_{-\infty}^{+\infty} \delta(x - \theta) dx = 1.$$

Furthermore, the rth moment about the origin of the Pareto distribution is given by

$$\mu'_r = \frac{\alpha \theta^r}{\alpha - r}, \ \text{where} \ \alpha > r. \tag{3.6}$$

In addition, the variance, mode, and median are, respectively, given by

$$Var(X) = \frac{\alpha \theta^2}{(\alpha - 1)^2 (\alpha - 2)}, \ \alpha > 2 \ \text{for finite variance}, \tag{3.7}$$

$$Mode(X) = \theta \ \text{and} \ Median(X) = \theta \sqrt[\alpha]{2}. \tag{3.8}$$

The Pareto distribution is related to the exponential distribution as follows:

$$f(x; \theta, \alpha) = Exp(\ln(x/\theta); \alpha), \text{ where } X \sim Pa(\theta, \alpha). \tag{3.9}$$

Let us suppose $\mathbf{X}^m = (X_1, X_2, ..., X_j, ..., X_m)$, $j = 1, 2, ..., m$ is a random sample of size m from (3.1). Quant (1966) showed that maximum likelihood estimators of θ and α – denoted by $\widehat{\theta}$ and $\widehat{\alpha}$, respectively – are

$$\widehat{\theta} = \min_{1 \leq j \leq m} X_j = X_{(1)} \text{ and } \widehat{\alpha} = mY^{-1}, \tag{3.10}$$

where $X_{(r)}$ denotes the rth order statistic and $Y = \sum_{j=1}^{m} \ln(X_j/X_{(1)})$. Furthermore, Malik (1970) derived distributions of $\widehat{\theta}$ and $\widehat{\alpha}$ that are given by

$$\widehat{\theta} \sim Pa\left(\theta, m\alpha\right) \text{ and } \widehat{\alpha} \sim \Gamma^{-1}(m - 1, \ m\alpha), \tag{3.11}$$

where $Pa\left(a, b\right)$ is the Pareto distribution with the scale parameter a and the shape parameter b, and $\Gamma^{-1}(c, d)$ is the inverse gamma distribution with the shape parameter c and the scale parameter d.

Now, let $\{X_{ij}\}, i = 1, 2, ..., k; \ j = 1, 2, ..., m_i$ be independently distributed as

$$X_{ij} \sim \frac{\alpha_i \theta^{\alpha_i}}{x_{ij}^{(\alpha_i + 1)}} I_{[\theta, \infty]}(x_{ij}), \quad \theta, \alpha_i > 0, \forall i, \tag{3.12}$$

where θ denotes the common unknown scale parameter and α_i's are unknown and possibly unequal shape parameters. I denotes the usual indicator function given by

$$I_{[\theta, \infty]}(y) = \begin{cases} 1 & \text{if } y \geq \theta \\ 0 & \text{otherwise} \end{cases}.$$

Elfessi and Jin (1996) showed the maximum likelihood estimators of the common θ, denoted by $\widehat{\theta}$, and the shape parameter, denoted by $\widehat{\alpha}_i$, of the ith Pareto population.

These results are expressed in the following:

$$\widehat{\theta} = T = \min_{1 \leq i \leq k} X_{i(1)} \text{ and } \widehat{\alpha}_i = A_i = m_i \left[\sum_{j=1}^{m_i} \ln(X_{ij}/\widehat{\theta}) \right]^{-1} \text{ for } \forall i,$$

(3.13)

where $X_{s(r)}$ denotes the rth order statistic of the sth population. Furthermore, Elfessi and Jin (1996) showed that

$$T \sim Pa\,(\theta, \alpha^*) \text{ and } A_i \sim \Gamma^{-1}(m_i - 1, m_i \alpha_i), \text{ for } \forall i,$$

(3.14)

where $\alpha^* = \sum_{i=1}^{k} m_i \alpha_i$ and $i = 1, 2, ..., k$.

Therefore, it can be shown that

$$2\alpha^* \ln(T/\theta) = V \sim \varkappa_2^2 \; ; \; 2m_i \alpha_i A_i^{-1} = W_i \sim \varkappa_{2m_i-2}^2, \text{ for } \forall i,$$

(3.15)

where $\varkappa_f^2$ is the chi-squared distribution with degrees-of-freedom f.

3.3.1 The statistical testing of hypothesis

Let us get started testing the hypothesis:

$$T^I \rightarrow H_0^I : \theta \leq \theta_0 \text{ vs. } H_a^I : \theta > \theta_0,$$

(3.16)

where θ_0 is a known quantity. Furthermore, we are also interested in testing:

$$T^{II} \rightarrow H_0^{II} : \theta \geq \theta_0 \text{ vs. } H_a^{II} : \theta < \theta_0, \text{ and}$$

(3.17)

$$T^{III} \rightarrow H_0^{III} : \theta = \theta_0 \text{ vs. } H_a^{III} : \theta \neq \theta_0.$$

(3.18)

Suppose $\mathbf{X}_i^{m_i} = (X_{i1}, X_{i2}, ..., X_{im_i})$ is a random sample of size m_i from an ith truncated Pareto population $Pa\,(\theta, \alpha_i)$, $i = 1, 2, ..., k$, where θ denotes the common

29

unknown scale parameter and α_i is an unknown and possibly unequal shape parameter of the ith Pareto population. Furthermore, suppose $\mathbf{x}_i^{m_i} = (x_{i1}, x_{i2}, ..., x_{im_i})$ is its observed value. Moreover, let us suppose $X_{DATA} = [X_{ij}]_{i=1,2,...,k;\ j=1,2,...,m_i}$ is the collection of all random samples from all k Pareto populations and $x_{DATA} = [x_{ij}]_{i=1,2,...,k;\ j=1,,...,m_i}$ is its observed value.

Now, from (3.15), generalized pivots for estimating θ, denoted by $R(X_{DATA};x_{DATA}, \theta, \boldsymbol{\alpha})$, and α_i, denoted by $R(\mathbf{X}_i^{m_i}; \mathbf{x}_i^{m_i}, \alpha_i)$, are given, respectively, by

$$R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha}) = te^{-V/(\sum_{i=1}^k W_i a_i)} \text{ and} \tag{3.19}$$

$$R(\mathbf{X}_i^{m_i}; \mathbf{x}_i^{m_i}, \alpha_i) = 0.5 W_i a_i / m_i; \text{ for } i = 1, 2, ..., k, \tag{3.20}$$

where $\boldsymbol{\alpha} = (a_1, a_2, ..., a_k)$, a_i is the observed value of A_i, or simply an estimate of α_i, and t is the observed value of T, or simply an estimate of θ.

Therefore, the potential generalized test variable for testing T^I, T^{II}, or T^{III} is defined by

$$T(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha}) = R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha}) - \theta,$$

$$T(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha}) = te^{-V/(\sum_{i=1}^k W_i a_i)} - \theta. \tag{3.21}$$

Since

1. the observed value of $T(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$, which is denoted by

   $t(x_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$, is 0,

2. $T(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$, when θ is specified, has a probability

   distribution that is free of nuisance parameters,

3. the cumulative distribution function of $T(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$, when x_{DATA}

   and nuisance parameters are fixed, is monotonically decreasing function of θ

   for any given $T(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$, denoted by $t(x_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$,

$T(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$ is a generalized test variable that can be used to test the given

hypotheses (Weerahandi 1989, 1995, and 2004).

Thus, the generalized p-value, sometimes referred to as the generalized observed level

of significance or the generalized significance level, for testing

1. T^I is given by $p_g^I = Pr(te^{-V/(\sum_{i=1}^{k} W_i a_i)} < \theta_0)$

2. T^{II} is given by $p_g^{II} = Pr(te^{-V/(\sum_{i=1}^{k} W_i a_i)} > \theta_0)$,

3. T^{III} is given by $p_g^{III} = 2\min[p_g^I, p_g^{II}]$.

These p-values can be evaluated through numerical procedures:

1. Numerical integration:

   Numerical integration is performed with respect to V and $W_{i(=1,2,...,k)}$, which

   are independent random variables with known density functions.

2. Monte Carlo Simulation method:

   Once the data are obtained, compute

   $$a_i = \frac{m_i}{\left[\sum_{j=1}^{m_i} \ln(x_{ij}/x_{i(1)})\right]} \text{ for } i = 1, 2, ..., k, \text{ and}$$

   $$t = \min\left[x_{1(1)}, x_{2(1)}, ..., x_{k(1)}\right].$$

   Then, generate a large number of random numbers $\left(V, W_{i(=1,2,...,k)}\right)$, where

   $$V \sim \varkappa_2^2 \text{ and } W_{i(=1,2,...,k)} \sim \varkappa_{2m_{i(=1,2,...,k)}-2}^2.$$

For each value of $\left(V, W_{i(=1,2,\ldots,k)}\right)$, compute

$$R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha}) = te^{-V/(\sum_{i=1}^{k} W_i a_i)},$$

and then compute

$$T(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha}) = R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha}) - \theta.$$

Now,

1. fraction of random numbers pairs for which $R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha}) < \theta_0$

   yields p_g^I,

2. fraction of random numbers pairs for which $R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha}) > \theta_0$

   yields p_g^{II}, and

3. fraction of random numbers pairs for which $\min[R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha}) < \theta_0,$

   $R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha}) > \theta_0]$ yields p_g^{III}.

3.3.2 Confidence intervals

Since

1. the value of $R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$ is θ, and

2. the distribution of $R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$ is independent of any unknown

   parameters,

$R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$ is a generalized pivotal quantity for constructing a $100(1 - \gamma)\%$ confidence interval for θ, where $1-\gamma$ is the confidence coefficient or the confidence level (Weerahandi 1993, 1995, and 2004).

Now, one-sided as well as two-sided confidence intervals are constructed as follows:

1. the lower bound, $R_{\gamma}(\theta; t, \mathbf{a})$, of a $100(1 - \gamma)\%$ one-sided confidence

interval for θ is sought such that

$$1 - \gamma = \Pr(te^{-V/(\sum_{i=1}^{k} W_i a_i)} \leq R_{1-\gamma}(\theta; t, \mathbf{a})), \qquad (3.22)$$

2. the upper bound, $R_{1-\gamma}(\theta; t, \mathbf{a})$, of a $100(1-\gamma)\%$ one-sided confidence

interval for θ is sought such that

$$\gamma = \Pr(te^{-V/(\sum_{i=1}^{k} W_i a_i)} \leq R_{\gamma}(\theta; t, \mathbf{a})), \qquad (3.23)$$

3. the lower limit, $R_{\gamma/2}^{l}(\theta; t, \mathbf{a})$, and the upper limit, $R_{1-\gamma/2}^{u}(\theta; t, \mathbf{a})$, of a

$100(1-\gamma)\%$ two-sided confidence interval for θ are sought such that

$$1 - \gamma = \Pr(R_{1-\gamma/2}^{l}(\theta; t, \mathbf{a}) \leq te^{-V/(\sum_{i=1}^{k} W_i a_i)} \leq R_{\gamma/2}^{u}(\theta; t, \mathbf{a})), \qquad (3.24)$$

where $\mathbf{a} = (a_1, a_2, ..., a_k)$ is the observed value of $\mathbf{A} = (A_1, A_2, ..., A_k)$, or simply an estimate of $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_k)$, and $R_{\gamma}(\theta; t, \mathbf{a})$ is the 100γ-th percentile of $R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$'s. Once the calculated $R(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\alpha})$'s are ordered, the desired percentiles can be obtained.

To obtain actual coverage probabilities (empirical confidence levels), it is necessary:

1. to repeat the above process for a large number of times (i.e. 100 000 –

1 000 000), and

2. to calculate the fraction of times θ falls within calculated (empirical)

generalized confidence intervals.

3.4 The Bayesian estimation

A Bayesian approach for statistical inferences of the common scale parameter θ, contrasting the conventional classical approach and the newly introduced generalized

variable approach, is introduced and discussed, and then the Monte Carlo method and commonly used Markov Chain Monte Carlo (MCMC) methods are introduced in this section.

The Bayesian statistics and Markov Chain Monte Carlo (MCMC) methods have been twins in statistical arena for more than 20 years as the former covers the philosophical aspect of the Bayesian approach and the latter is well suited for the calculations of probabilities and does not rely on conjugacy or asymptotic moment-based approximations.

When marginal posterior distributions are impossible to be summarized analytically, Bayesian statisticians tend to numerical approaches for the summarization of these marginal posterior distributions. The Monte Carlo method is the commonly used numerical approach in the Bayesian statistics. In order to use this method, it is necessary to have well-suited algorithms; there are two well-known algorithms:

1. the Gibbs sampling – uses a sequence of draws from conditional posterior distribution to characterize the joint posterior distribution:

special case of Metropolis-Hastings algorithm

2. the Metropolis-Hastings algorithm – used for all sorts of numerical integration and optimization.

For more details on this algorithm, interested parties are referred to Metropolis et al. (1953), Hastings (1979), and Chib and Greenberg (1995).

In the Gibbs sampling technique incorporated with the Meta-analysis – a statistical approach adopted to summarize and integrate a collection studies using many familiar techniques to draw general conclusions that was first performed by Karl

Pearson in 1904 – the information from several Pareto populations are combined to estimate the common scale parameter θ when shape parameters $\alpha_1, \alpha_2, ..., \alpha_k$ are unknown. The marginal posterior distribution of a parameter of interest is the target distribution in the Bayesian analysis for the estimation of the parameter of interest. But, there are few possible difficulties incorporated with handling those distributions:

1. when the marginal posterior distribution is a non-standard distribution,

2. when the marginal posterior distribution is a poly standard distribution,

3. when the marginal posterior distribution is a poly non-standard distribution,

4. when the dimensionality problem causes the numerical integration to be extremely difficult.

The Gibbs sampler provides considerable and fair robust solutions for such drastic and difficult situations. The methodology is illustrated by using an example providing the numerical values for Bayesian and MLE estimates along with the $100(1 - \gamma)\%$ credible region, where $1 - \gamma$ is the confidence coefficient.

Random samples from the Pareto distributed $Pa(\theta, \alpha_i)$ for $i = 1, 2, ..., k$ with the unknown shape parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ...\alpha_k)$ and an unknown common scale parameter θ are drawn. This is denoted by

$$x_{ij} \text{ for } i = 1, 2, ..., k; \ j = 1, 2, ..., m_i$$

Then the joint density of the sample variables is given by

$$f(\ X_{DATA}\,|\theta,\ \boldsymbol{\alpha}) = \prod_{i=1}^{k} \alpha_i^{m_i} \theta^{m_i \alpha_i} \exp\left(-(\alpha_i + 1) \sum_{j=1}^{m_i} \ln x_{ij}\right),$$

(3.25)

where

$$X_{DATA} = [x_{ij}]_{i=1,2,...,k;\ j=1,2,...,m_i} \quad \text{and} \quad f(x_{ij}) = \frac{\alpha_i \theta^{\alpha_i}}{x_{ij}^{(\alpha_i+1)}}, \quad \theta, \alpha_i > 0, \forall i.$$

3.4.1 The derivation of the Bayes estimate of θ when $\boldsymbol{\alpha}$ is known

If $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ...\alpha_k)$ is assumed to be known , then a conjugate family of prior distributions for θ is given by

$$\pi(\theta) = \prod_{i=1}^{k} \pi(\theta_i), 0 < \theta_i < \theta_0,$$

$$\pi(\theta) = \prod_{i=1}^{k} \delta_i \theta^{\delta_i-1} \theta_0^{-\delta_i}; 0 < \theta < \theta_0 \text{ and } \theta_0, \delta_i > 0 \qquad (3.26)$$

where θ_i is distributed as the power function distribution, i.e. $\theta_i \sim POWER(\theta_0, \delta_i)$, where θ_0 is one of the boundary parameters – or sometimes it is referred to as the scale parameter – and δ_i is the shape parameter. With the priors of this form, the joint posterior density of θ and $\boldsymbol{\alpha}$ may be seen to be of the form:

$$h(\theta, \boldsymbol{\alpha} | X_{DATA}) = \prod_{i=1}^{k} \alpha_i^{m_i} \delta_i \theta_0^{-\delta_i} \theta^{(m_i\alpha_i+\delta_i)-1} \exp\left(-(\alpha_i+1)\sum_{j=1}^{m_i} \ln x_{ij}\right).$$

$$(3.27)$$

Furthermore, it can be inferred that the the conditional posterior distribution of θ given the k parameters $\alpha_1, \alpha_2, ...\alpha_k$ is again in the form of power function as follows:

$$g(\theta | \alpha_1, \alpha_2, ...\alpha_k) = \tau(X_{DATA}, \boldsymbol{\alpha}) \theta^{\left(\sum_{i=1}^{k} m_i\alpha_i+\delta_i\right)-k}; \ 0 < \theta < \min(\theta_0, \min_{1 \le i \le k} X_{i(1)}),$$

$$(3.28)$$

where

$$\tau(X_{DATA}, \boldsymbol{\alpha}) = \prod_{i=1}^{k} \alpha_i^{m_i} \delta_i \theta_0^{-\delta_i} \exp\left(-(\alpha_i+1)\sum_{j=1}^{m_i} \ln x_{ij}\right). \qquad (3.29)$$

36

Equivalently,

$$\theta \mid \boldsymbol{\alpha} \sim POWER\left\{\min(\theta_0, \min_{1 \leq i \leq k} X_{i(1)}), \left[\left(\sum_{i=1}^{k} m_i\alpha_i + \delta_i\right) - k + 1\right]\right\}.$$

(3.30)

Note that if a random variable Y is distributed as a power function distribution with the the scale parameter a and the shape parameter b, i.e. $Y \sim POWER(a, b)$, then the mean and the median of the power function distribution are given by

$$E(Y) = \frac{ab}{(b+1)}, \quad Median(Y) = a2^{-b}.$$

(3.31)

Assuming the squared-error loss, the Bayes estimate of θ will be given by

$$\widehat{\theta}_B^{Mean} = \frac{\left[\sum_{i=1}^{k}(m_i\alpha_i + \delta_i) - k + 1\right]}{\left[\sum_{i=1}^{k}(m_i\alpha_i + \delta_i) - k + 2\right]}\left[\min(\theta_0, \min_{1 \leq i \leq k} X_{i(1)})\right],$$

(3.32)

where $\widehat{\theta}_B^{Mean}$ is the posterior mean (posterior mean estimate of θ). If instead, we assume the absolute-error loss, the Bayes estimate of θ will be given by

$$\widehat{\theta}_B^{Median} = 2^{-\left[\sum_{i=1}^{k}(m_i\alpha_i + \delta_i) - k + 1\right]^{-1}}\left[\min(\theta_0, \min_{1 \leq i \leq k} X_{i(1)})\right],$$

(3.33)

where $\widehat{\theta}_B^{Median}$ is the posterior median (posterior median estimate of θ).

3.4.2 The derivation of the Bayes estimate of θ when $\boldsymbol{\alpha}$ is unknown

As we know the joint posterior density of θ and $\boldsymbol{\alpha} = (a_1, a_2, ..., a_k)$ is in the form of

$$g(\theta, \boldsymbol{\alpha} \mid X_{DATA}) = \prod_{i=1}^{k} \alpha_i^{m_i}\delta_i\theta_0^{-\delta_i}\theta^{(m_i\alpha_i + \delta_i) - k}\exp\left(-(\alpha_i + 1)\sum_{j=1}^{m_i}\ln x_{ij}\right),$$

(3.34)

the marginal posterior distribution of θ is given by

$$m(\theta \,|X_{DATA}) \;\; = \;\; \prod_{i\,=1}^{k} \Gamma(m_i+1)\delta_i\theta_0^{-\delta_i}\theta^{(\delta_i-k)}\left[\left(\sum_{j\,=1}^{m_i}\ln x_{ij}\right)-m_i\ln\theta\right]^{-(m_i+1)}(3.35)$$
$$\left[\exp\left(\sum_{j\,=1}^{m_i}\ln x_{ij}\right)\right]$$

This distribution is a product of k non-standard distributions, called a poly non-standard distribution; it is difficult to work with and the numerical integration must be used to to determine the moments of the marginal distribution.

In general, given θ, the conditional posterior distribution of the parameter α_i is an independent gamma distribution shown as follows:

$$\alpha_i\,|\theta \sim GAMMA\left\{m_i+1,\left[\left(\sum_{j\,=1}^{m_i}\ln x_{ij}\right)-m_i\ln\theta\right]^{-1}\right\},\text{for}\;\; i=1,2,...,k.$$
$$(3.36)$$

The conditional posterior distribution of θ given $\boldsymbol{\alpha}$ is a power function distribution as follows:

$$\theta\,|\boldsymbol{\alpha} \sim POWER(\min(\theta_0,\min_{1\,\leq\,i\,\leq\,k}X_{i(1)}\;),\sum_{i\,=1}^{k}m_i\alpha_i+\delta_i).\qquad(3.37)$$

Under the squared-error loss, the Bayes estimate of θ is the mean value of the marginal posterior distribution. Because of the complexity of the function, prompted by Gregurich and Broemeling (1997) on the Bayesian analysis for the common mean of independent normal populations using the Gibbs sampler, the Gibbs sampler is used to evaluate the Bayes estimates of θ when $\boldsymbol{\alpha}$ is unknown.

The Gibbs sampling algorithm:

1. Find initial values for θ and $\alpha_i(i=1,2,...,k)-$ initial values are usually

    chosen close to the mode of the marginal posterior distribution or the

Maximum Likelihood Estimates (MLE). Here, we choose the MLE's of θ

and α_i's as the initial values: $\theta^{(0)}$ and $\alpha_i^{(0)}(i = 1, 2, ..., k)$, where $\theta^{(0)} =$

$\min_{1 \leq i \leq k} X_{i(1)}$, and $\alpha_i^{(0)} = m_i^{-1} \left[\sum_{j=1}^{m_i} \ln(X_{ij}/X_{i(1)}) \right]^{-1}$ for $i = 1, 2, ..., k$;

2. Draw $\theta^{(1)}$ from the conditional power distribution of θ given $\alpha_1 = \alpha_1^{(0)}, \alpha_2 =$

$\alpha_2^{(0)}, ..., \alpha_k = \alpha_k^{(0)}$

i.e., $\theta^{(1)} \sim \theta \left| \boldsymbol{\alpha} = (\alpha_1^{(0)}, \alpha_2^{(0)}, ..., \alpha_k^{(0)}) \right.$,

3. Draw $\alpha_i^{(1)'}$s from the conditional gamma distribution of α_i given $\theta = \theta^{(1)}$

for $i = 1, 2, ..., k$,

i.e., $\alpha_i^{(1)} \sim \alpha_i \left| \theta = \theta^{(1)} \right.$ for $i = 1, 2, ..., k$,

4. Draw $\theta^{(2)}$ from the conditional power distribution of θ given $\alpha_1 = \alpha_1^{(1)}, \alpha_2 =$

$\alpha_2^{(1)}, ..., \alpha_k = \alpha_k^{(1)}$

i.e. $\theta^{(2)} \sim \theta \left| \boldsymbol{\alpha} = (\alpha_1^{(1)}, \alpha_2^{(1)}, ..., \alpha_k^{(1)}) \right.$,

5. Draw $\alpha_i^{(2)'}$s from the conditional gamma distribution of α_i

given $\theta = \theta^{(2)}$ for $i = 1, 2, ..., k$;

i.e., $\alpha_i^{(2)} \sim \alpha_i \left| \theta = \theta^{(2)} \right.$ for $i = 1, 2, ..., k$,

Since we always condition on past draws, the resultant sequence yields

a Markov Chain Monte Carlo (MCMC),

6. This process is continued for $g = 1, 2, ...G$ iterations,

7. Then, the above Gibbs process is repeated $n = 1, 2, ..., N$ times.

This is schematically represented as in the Table 3.1. At this point, a random sample

of size N from the joint posterior distribution of θ and α_i's for $i = 1, 2, ..., k$, a random

sample of size N from the marginal posterior distribution of θ, and random samples

of size N from the marginal posterior distributions α_i's for $i = 1, 2, ..., k$, can easily be

obtained as follows:

Let's denote $\theta_n^{(G)}, \alpha_{1n}^{(G)}, ..., \alpha_{kn}^{(G)}$ by J_n for $n = 1, 2, ..., N$

$$\theta_1^{(G)}, \theta_2^{(G)}, ..., \theta_N^{(G)} \text{ by } M_\theta$$

$$\alpha_{i1}^{(G)}, \alpha_{i2}^{(G)}, ..., \alpha_{iN}^{(G)} \text{ by } M_{iN}^{\alpha_i} \text{ for } i = 1, 2, ..., k$$

Then $J_{n(=1,2,...,N)} \sim g(\theta, \boldsymbol{\alpha} \,|\, X_{DATA})$,

$$M_\theta \sim m(\theta \,|\, X_{DATA}), \text{and}$$

$$M_{iN(=1,2,...,k)}^{\alpha_i} \sim m(\alpha_{i(=1,2,...,k)} \,|\, X_{DATA}),$$

Table 3.1 Random sample of the joint posterior

distribution of $\theta, \alpha_1, ..., \alpha_k$

#	θ	α_1	α_2	.	.	.	α_i	.	.	.	α_k
1	$\theta_1^{(G)}$	$\alpha_{11}^{(G)}$	$\alpha_{21}^{(G)}$	.	.	.	$\alpha_{i1}^{(G)}$	.	.	.	$\alpha_{k1}^{(G)}$
2	$\theta_2^{(G)}$	$\alpha_{12}^{(G)}$	$\alpha_{22}^{(G)}$	.	.	.	$\alpha_{i2}^{(G)}$	.	.	.	$\alpha_{k2}^{(G)}$
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
n	$\theta_n^{(G)}$	$\alpha_{1n}^{(G)}$	$\alpha_{2n}^{(G)}$	.	.	.	$\alpha_{in}^{(G)}$	.	.	.	$\alpha_{kn}^{(G)}$
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
N	$\theta_N^{(G)}$	$\alpha_{1N}^{(G)}$	$\alpha_{2N}^{(G)}$	.	.	.	$\alpha_{iN}^{(G)}$	.	.	.	$\alpha_{kN}^{(G)}$

Each column in Table 3.1 shows that a sample of values of marginal distributions of $\theta, \alpha_1, \alpha_2, ...,$ and α_k that are generated by the Gibbs sampler from the conditional distributions. If N is large, mean and variance of the marginal posterior distribution of θ are given, respectively, by

$$E\left(\theta \,|X_{DATA}\right) = \sum_{n=1}^{N} \theta_n^{(G)} /N = \overline{\theta}, \text{ and} \qquad (3.38)$$

$$Var(\theta \,|X_{DATA}) = (N-1)^{-1} \sum_{n=1}^{N} \left(\theta_n^{(G)} - \overline{\theta}\right)^2 \qquad (3.39)$$

Therefore, the Bayes estimate of θ when $\boldsymbol{\alpha}$ is unknown using the Gibbs sampler is given by

$$\theta_B^G = \sum_{n=1}^{N} \theta_n^{(G)} /N \qquad (3.40)$$

Additional characteristics such as the median, mode, quartiles, deciles, percentile, octile, and the $100(1-\gamma)\%$ credible region, $1-\gamma$ being the confidence coefficient, of the posterior distribution of the parameter θ can also be calculated from the sample generated by the Gibbs technique. Furthermore, hypothesis testing can also be performed. The credible sets for θ are calculated approximately and exactly as follows:

1. the approximate $100(1-\gamma)\%$ credible region for θ :

$$E\left(\theta \,|X_{DATA}\right) \pm Z_{\gamma/2} \sqrt{\frac{Var(\theta \,|X_{DATA})}{N}}, \qquad (3.41)$$

2. the actual $100(1-\gamma)\%$ credible region for θ :

$$l = \prod_{i=1}^{k} \Gamma(m_i - 1)\delta_i \theta_0^{-\delta_i} \theta^{(\delta_i - 1)} \left[\left(\sum_{j=1}^{m_i} \ln x_{ij}\right) - m_i \ln \theta\right]^{(m_i - 1)} \left[\exp\left(\sum_{j=1}^{m_i} \ln x_{ij}\right)\right].$$

$$(3.42)$$

To obtain the actual credible region for θ, solve the equation (3.42) for θ with appropriate values of l until values of θ get as close as to values found in equation (3.41), while the area under the curve, given by the left hand side of equation (3.42), is $100(1-\gamma)\%$.

3.5 Illustrative examples for the common scale parameter

Example 1. Comparison of the proposed procedure with the classical approach

based on large sample method

This example deals with the Pareto distributions $X_{i(=1,2,3)} \sim Pa\left(\alpha_{i(=1,2,3)}, \theta\right)$ generated by the following population parameters: $\theta = 100$, $\boldsymbol{\alpha} = (0.5,\ 1.0,\ 1.5)$, and sample sizes $\mathbf{m} = (10, 10, 10)$. The data generated from these distributions are:

$X_1 \sim Pa\left(\theta, \alpha_1\right)$: $\quad$ $182.4447, 766.6342, 149.9515, 183.5521, 131.3459,$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad$ $184.8249, 403.8077, 314.5954, 1264.0143, 116.9585$

$X_2 \sim Pa\left(\theta, \alpha_2\right)$: $\quad$ $815.0133, 113.2192, 216.6859, 266.3277, 255.2327,$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad$ $354.8153, 640.5599, 417.5773, 109.8015, 167.6198$

$X_3 \sim Pa\left(\theta, \alpha_3\right)$: $\quad$ $102.8793, 142.2166, 101.4941, 104.4409, 247.1254,$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad$ $316.8746, 213.758, 227.4824, 164.4707, 335.9244$

I. Confidence intervals:

Assuming that all of the above parameters are unknown:

1. the lower bound of a 90% one-sided generalized empirical confidence interval for θ calculated from this data is 98.0647,

2. the upper bound of a 90% one-sided generalized empirical confidence

   interval for θ calculated from this data is 106.2982, and

3. the upper and lower limits of a 90% two-sided generalized empirical

   confidence interval for θ calculated from this data, respectively, are 96.8665,

   109.1483.

The comparison of classical and generalized confidence intervals are summarized in Table 3.2 as follows:

Table 3.2 Comparison of generalized and classical confidence intervals

Confidence interval	Generalized	Classical
90%	$(96.8665, 109.1483)$	$(95.6165, 110.7414)$

The results shows that the generalized variable method produces a shorter interval than the classical method produces. Therefore, the generalized variable method outperforms its counter part, the classical procedure, for this particular problem.

II. p-values:

Furthermore, the comparison of classical and generalized p-values of two tests are summarized in Table 3.3 as follows:

Table 3.3 Comparison of generalized and classical p-values

Test	Generalized	Classical
$\theta \leq 102$ vs. $\theta > 102$	0.9975	0.0065
$\theta \leq 98$ vs. $\theta > 98$	0.0019	0.2783

When θ is tested, assuming that it is unknown, to see whether it is greater than 102 (the claim or the alternative hypothesis being $\theta > 102$ versus $\theta \leq 102$), the generalized p-value, having a value greater than 0.05, provides evidence against the claim. However, the classical p-value, having a value less than 0.05, suggests that the claim, which is false since θ is 100, is accepted. This same argument can be used when θ, assuming that it is unknown, is tested to see whether it is less than 98. Both these arguments clearly show that the generalized variable method provides accurate, reliable, and non-misleading results, while the classical approach fails to do so for this particular case. Hence, the generalized variable method outperforms the classical approach for this particular case.

III. Sizes:

Table 3.2 shows classical and generalized empirical (actual) type I error rates (sizes of tests) for testing

1. $H_0^I : \theta \leq 100$ vs. $H_a^I : \theta > 100$,

2. $H_0^I : \theta \leq 500$ vs. $H_a^I : \theta > 500$,

when nominal (intended) type I error rate is at 0.1. All results are based on $100,000$ simulations.

Table 3.4 Comparison of generalized and classical actual type I
error rates

Parameters: k, θ, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$	Generalized	Classical
$k = 3$, $\theta = 100$, $\boldsymbol{\alpha} = (0.5, 1.0, 1.5)$	0.052	0.040
$k = 3$, $\theta = 100$, $\boldsymbol{\alpha} = (2.0, 2.5, 3.0)$	0.046	0.273
$k = 3$, $\theta = 100$, $\boldsymbol{\alpha} = (3.5, 4.0, 4.5)$	0.048	0.438
$k = 3$, $\theta = 500$, $\boldsymbol{\alpha} = (0.5, 1.0, 1.5)$	0.044	0.000
$k = 3$, $\theta = 500$, $\boldsymbol{\alpha} = (2.0, 2.5, 3.0)$	0.049	0.007
$k = 3$, $\theta = 500$, $\boldsymbol{\alpha} = (3.5, 4.0, 4.5)$	0.049	0.043

According to this simulation study, when compared with the classical procedure, actual type I error rates (actual sizes of tests) of the generalized procedure get much closer to the intended size. This paves the way for the generalized procedure to be considered as a better procedure than the classical procedure for this particualr problem.

IV. Powers:

Table 3.5 shows the power comparison for testing $H_0^I : \theta \leq 100$ vs. $H_a^I : \theta > 100$ without and after adjusting the size at $\gamma = 0.10$. All results are based on $100,000$ simulations.

Table 3.5 Comparison of powers for testing H_0^I without and after adjusting the size

Parameters: k, θ, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$	Without		After	
	GVM	CM	GVM	CM
$k = 3$, $\theta = 100$, $\boldsymbol{\alpha} = (0.5, 1.0, 1.5)$	0.052	0.040	0.010	0.010
$k = 3$, $\theta = 101$, $\boldsymbol{\alpha} = (2.0, 2.5, 3.0)$	0.115	0.054	0.110	0.108
$k = 3$, $\theta = 102$, $\boldsymbol{\alpha} = (3.5, 4.0, 4.5)$	0.412	0.238	0.405	0.403
$k = 3$, $\theta = 103$, $\boldsymbol{\alpha} = (0.5, 1.0, 1.5)$	0.716	0.580	0.698	0.695
$k = 3$, $\theta = 104$, $\boldsymbol{\alpha} = (2.0, 2.5, 3.0)$	0.823	0.704	0.812	0.809
$k = 3$, $\theta = 105$, $\boldsymbol{\alpha} = (3.5, 4.0, 4.5)$	0.951	0.912	0.932	0.925

When compared, values of generalized and classical powers, with and after adjusting the size, clearly suggest that generalized variable method outperforms the classical method in terms of power for this particular case.

V. Coverage probabilities:

The comparison of the generalized coverage probabilities with the classical coverage probabilities – with the intended confidence level, or the confidence coefficient, $1 - \gamma = 0.1$, or the intended significance level $\gamma = 0.9$ – is given in Table 3.6.

These coverage probabilities are based on 100 000 simulated random samples from the Pareto density given in equation (3.1). Random samples of size m are generated from the uniform distribution,

$$u \sim U(0, 1). \tag{3.43}$$

Then Pareto random samples are generated by substituting values of uniform random samples, different known Pareto parameters θ, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ into the Pareto quantile function, given by

$$Q(u) = \frac{\theta}{u^{1/\alpha}}. \tag{3.44}$$

Classical approximate $100(1-\gamma)\%$ confidence intervals for parameter θ are calculated by using $\left(\widehat{\theta} - Z_{\gamma/2}SE_{\widehat{\theta}}, \ \widehat{\theta} + Z_{\gamma/2}SE_{\widehat{\theta}}\right)$ while generalized approximate $100(1-\gamma)\%$ confidence intervals for parameter θ are calculated by using $(R_\theta(\gamma/2), R_\theta(1-\gamma/2))$.

Table 3.6 Comparison of generalized and classical probability coverages for 90% two-sided confidence intervals for θ

Parameters: k, θ, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$	Generalized	Classical
$k = 3$, $\theta = 100$, $\boldsymbol{\alpha} = (0.5, 1.0, 1.5)$	0.90	0.99
$k = 3$, $\theta = 100$, $\boldsymbol{\alpha} = (2.0, 2.5, 3.0)$	0.82	0.86
$k = 3$, $\theta = 100$, $\boldsymbol{\alpha} = (3.5, 4.0, 4.5)$	0.88	0.84
$k = 3$, $\theta = 500$, $\boldsymbol{\alpha} = (0.5, 1.0, 1.5)$	0.80	0.97
$k = 3$, $\theta = 500$, $\boldsymbol{\alpha} = (2.0, 2.5, 3.0)$	0.92	0.99
$k = 3$, $\theta = 500$, $\boldsymbol{\alpha} = (3.5, 4.0, 4.5)$	0.86	0.98

According to these simulation results, one can clearly see that actual (empirical) probability coverages for the generalized method get as much close as to intended (nominal) coverage probabilities. Thus, the generalized variable method outperforms the classical procedure for this particular case.

47

Example 2. Comparison of proposed procedure with the classical approach

based on the inverse normal method

Table 3.7 shows the comparison of expected lengths of $100(1 - \gamma)\%$ confidence intervals for θ, where $1 - \gamma$ being the confidence coefficient, based on the generalized variable method and the inverse normal method in Baklizi (2002). Confidence intervals based on the inverse normal method are constructrd using certain classical independent tests – namely, the Tippett's method, the Fisher's method, the inverse normal method, and the logit method using the Meta-analysis for combinations of p-values. This is done on the basis of simulation. Consider $k = 2$ and take $(m_1, m_2) = (10, 5), (10, 10), (10, 15), \theta = 100, \alpha_1 = 1, \alpha_2 = 0.5, 1$ and $\gamma = 0.05, 0.1$. Note that, since the inverse normal method outperforms other methods in Baklizi (2002), the comparison is done between the confidence lengths, based on the inverse normal method and the proposed generalized method.

Table 3.7 Comparison of expected lengths of $100(1 - \gamma)\%$
confidence intervals for θ based on the generalized
variable method and the inverse normal method

γ	α_1	α_2	m_1	m_2	Generalized	Inverse normal
0.10	1	0.5	10	5	0.0398	0.2386
				10	0.0249	0.1861
		1.0		5	0.0338	0.1942
				10	0.0212	0.1372
0.05		0.5		5	0.0400	0.3101
				10	0.0251	0.2380
		1.0		5	0.0341	0.2474
				10	0.0216	0.1762

The inverse normal method:

Let us consider that $\mathbf{x}_i^{n_i} = (x_{i1}, x_{i2}, ..., x_{im_i})$ is the observed random sample of $\mathbf{X}_i^{n_i} = (X_{i1}, X_{i2}, ..., X_{im_i})$ of size m_i from $X_i \sim Pa(\theta, \sigma_i), i = 1, 2, ..., k.$

Then, p-value for testing the hypothesis, based on the ith sample,

$$H_0^i : \theta \leq \theta_0 \ \text{ vs. } \ H_a^i : \theta > \theta_0 \tag{3.45}$$

is given by

$$P_i = \Pr(F_{2,2(m_i-1)} > f_i) = \left(1 + a_i \ln\left(\frac{t_i}{\theta}\right)\right)^{-m_i+1} \sim U(0,1), \tag{3.46}$$

49

where f_i, and t_i are, respectively, the observed value of F_i and T_i, with

$$F_i = \frac{V_i/2}{W_i/2m_i - 2} = (m_i - 1)A_i \ln(T_i/\theta) \sim F_{2,2(m_i-1)}. \qquad (3.47)$$

This F_i is derived from each random variables

$$W_i(i = 1, 2, ..., k) \text{ defined as } W_i = 2m_i\alpha_i A_i^{-1} \sim \varkappa^2_{2m_i-2}, \qquad (3.48)$$

$$V_i(i = 1, 2, ..., k) \text{ defined as } V_i = 2m_i\alpha_i \ln(T_i/\theta) \sim \varkappa^2_2. \qquad (3.49)$$

Then, the inverse normal method rejects hypotheses of θ, i.e., $H_0^{i=1,2,...,k} : \theta \leq \theta_0$,

if

$$\frac{\sum_{i=1}^{k} \Phi^{-1}(P_i)}{k} \geq -z_\delta, \qquad (3.50)$$

at the level δ where Φ is the standard normal *cdf* and $\Phi(z_\delta) = 1 - \delta$.

The results in Table 3.7 shows that the generalized variable method produces much shorter intervals than the classical method produces, for each parametric specification in Baklizi (2002). Therefore, the generalized variable method outperforms even the exact classical procedure for this particular problem.

Example 3. Bayesian approach

This example deals with the same Pareto distributions as of Example 1:

i.e., $X_i \sim Pa(\theta, \alpha_i)$ where $i = 1, 2, 3$ generated by the following population para-

meters: $\theta = 100$, $\alpha_1 = 0.5$, $\alpha_2 = 1.0$, $\alpha_3 = 1.5$, and sample sizes: $m_1 = m_2 = m_3 = 10$. Assuming that the shape parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and θ are unknown, the Bayesian estimate of the common scale parameter θ is evaluated from the Gibbs sampler using macros of MINITAB. The following are the Gibbs sampling parameter

specifications given in accordance with the Gelfand, Hills, Racine-Poon, and Smith's (1990) suggestion on G and N where they recommended that G be approximately 50 iterations and N be less than or equal to 1000:

$G = 50$, $\mathbf{N} = (250, 500, 750, 1000)$.

The estimated values of θ determined by the Gibbs sampling method are reported in Table 3.8.

Table 3.8 Bayes estimates and 95% credible regions using the Gibbs sampler for θ when $\boldsymbol{\alpha}$ is unknown

			95% Credible Region	
N	θ_B^{GS}	STD	Lower	Upper
250	100.31	0.94	95.11	108.21
500	100.42	0.90	94.11	109.24
750	100.36	0.93	95.23	110.23
1000	100.43	0.90	95.03	109.43

Results provide more accurate Bayesian estimate for the common scale parameter of several Pareto populations for different Gibbs sampling sizes.

3.6 Generalized inferences for the common shape parameter

Prompted by the parallel work by Tian and Wu (2007) based on the generalized variable method by Weerahandi (1995, 2004), we utilize the Meta-Analysis in making generalized inferences on the common shape parameter of several Pareto populations

Consider k independent Pareto populations with unknown shape parameter α_i and unknown scale parameters θ_i. Let $\mathbf{X}_i^{m_i} = (X_{i1}, X_{i2}, ..., X_{im_i})$ be a random sample from the ith Pareto population

$$X_{ij} \sim Pa\,(\theta_i\,,\alpha_i), \text{for } i = 1, 2, ..., k;\ \ j = 1, 2, ..., m_i$$

Thus we have

$$\alpha\ =\alpha_i,\ \text{ for } i = 1, 2, ..., k,$$

where α is the unknown common shape parameter of the Pareto populations.

3.6.1 The statistical testing of hypothesis for α

Let us suppose that $\mathbf{x}_i^{m_i} = (x_{i1}, x_{i2}, ..., x_{im_i})$ is the observed value of the above random sample. Then, as we know from Section 3.3, maximum likelihood estimators of θ_i and α_i are, respectively, given by

$$\widehat{\theta}_i = T_i = X_{i(1)} \text{ and } \widehat{\alpha}_i = A_i = m_i Y_i^{-1}, \tag{3.51}$$

where $X_{i(1)} = \min(X_{i1}, X_{i2}, ..., X_{m_i})$ and $\ Y_i = \sum_{j=1}^{m_i} \ln(X_{ij}/X_{i(1)})$,

and their distributions are given by

$$T_i \sim Pa\,(\theta_i, m_i\alpha_i\,) \text{ and } A_i \sim \Gamma^{-1}(m_i - 1, m_i\alpha_i). \tag{3.52}$$

Now, generalized pivotal quantities for estimating θ_i and α_i are given, respectively, by

$$R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \theta_i, a_i) = t_i e^{-Vi/(W_i a_i)} \text{ and } R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i) = 0.5 W_i a_i/m_i,$$

$$\tag{3.53}$$

where

$$2m_i\alpha_i \ln(T_i/\theta_i) = V_i \sim \varkappa_2^2 \text{ and } 2m_i\alpha_i A_i^{-1} = W_i \sim \varkappa_{2m_i-2}^2, \tag{3.54}$$

with a_i being the observed value of A_i, or simply an estimate of α_i, and t_i being the observed value of T_i, or simply an estimate of θ.

Obviously, both generalized pivotal quantities $R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \theta_i, a_i)$ and $R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)$ give the same result as the classical pivotal quantities do. From the ith sample, the maximum likelihood estimator of α_i is

$$\widehat{\alpha}_i = A_i = m_i \left(\sum_{j=1}^{m_i} \ln(X_{ij}/\widehat{\theta}_i) \right)^{-1}, \tag{3.55}$$

where $\widehat{\theta}_i = T_i = \min(X_{i1}, X_{i1}, ..., X_{im_i})$.

The classical population variance based on the large sample approach for $\widehat{\alpha}_i$ is

$$\sigma^2(\widehat{\alpha}_i) = \frac{(m_i\alpha_i)^2}{(m_i-2)^2(m_i-3)}, \quad m_i > 3. \tag{3.56}$$

as $\widehat{\alpha}_i \sim \Gamma^{-1}(m_i - 1, m_i\alpha_i)$.

Now, suppose $X_{DATA} = [X_{ij}]_{i=1,2,...,k;\ j=1,,...,m_i}$ is the collection of random samples of all k Pareto populations and $x_{DATA} = [x_{ij}]_{i=1,2,...,k;\ j=1,,...,m_i}$ is its observed value.

The generalized pivotal quantity for the Pareto common shape parameter α is a weighted average of the generalized pivot $R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)$ based on k individual samples. It is given as follows:

$$R(X_{DATA}; x_{DATA}, \alpha) = \frac{\sum_{i=1}^{k} R(w_i) R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)}{\sum_{i=1}^{k} R(w_i)}, \tag{3.57}$$

where

$$R(w_i) = \text{the generalized pivot of the precision of estimator } \widehat{\alpha}_i,$$

53

$$R(w_i) = \frac{(m_i - 2)^2(m_i - 3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]^2}. \tag{3.58}$$

Now, consider the potential generalized test variable for testing

$$H_0 : \alpha \leq \alpha_0 \quad \text{vs.} \quad H_a : \alpha > \alpha_0, \tag{3.59}$$

where α_0 is a known quantity, defined by

$$T(X_{DATA}; x_{DATA}, \alpha) = \frac{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]}}{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]^2}} - \alpha. \tag{3.60}$$

One can show that $T(X_{DATA}; x_{DATA}, \alpha)$ satisfy the three conditions of generalized test variable set forth by Weerahandi (1995, 2004). Therefore, $T(X_{DATA}; x_{DATA}, \alpha)$ is a generalized test variable for the given testing of hypothesis. Thus, the generalized p-value, for testing $H_0 : \alpha \leq \alpha_0$ vs. $H_a : \alpha > \alpha_0$ is given by

$$p_\alpha^g = Pr(T(X_{DATA}; x_{DATA}, \alpha) < t(x_{DATA}; x_{DATA}, \alpha \mid \alpha = \alpha_0),$$

$$\tag{3.61}$$

$$p_\alpha^g = Pr(\frac{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]}}{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]^2}} < \alpha_0). \tag{3.62}$$

As in the common scale parameter estimation case, this p-value can also be evaluated through numerical procedures:

1. Numerical integration:

   Numerical integration is performed with respect to $V_{i(=1,2,...,k)}$ and $W_{i(=1,2,...,k)}$, which are independent random variables with known density functions.

2. Monte Carlo Simulation method:

Once the data are obtained, compute

$$a_i = \frac{m_i}{\left[\sum_{j=1}^{m_i} \ln(x_{ij}/x_{i(1)})\right]} \text{ for } i = 1, 2, ..., k, \text{ and}$$

$$t = \min\left[x_{1(1)}, x_{2(1)}, ..., x_{k(1)}\right].$$

Then, generate a large number of random numbers $\left(V_{i(=1,2,...,k)}, W_{i(=1,2,...,k)}\right)$,

where

$$V_{i(=1,2,...,k)} \sim \varkappa_2^2 \text{ and } W_{i(=1,2,...,k)} \sim \varkappa_{2m_{i(=1,2,...,k)}-2}^2.$$

For each value of $\left(V_{i(=1,2,...,k)}, W_{i(=1,2,...,k)}\right)$, first compute

$$R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i) = 0.5 W_i a_i / m_i,$$

then, compute

$$R(X_{DATA}; x_{DATA}, \alpha) = \frac{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i};\mathbf{x}^{m_i},\alpha_i)]}}{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i};\mathbf{x}^{m_i},\alpha_i)]^2}},$$

and lastly compute

$$T(X_{DATA}; x_{DATA}, \alpha) = \frac{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i};\mathbf{x}^{m_i},\alpha_i)]}}{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i};\mathbf{x}^{m_i},\alpha_i)]^2}} - \alpha.$$

Now,

1. the fraction of random numbers pairs for which $R(X_{DATA}; x_{DATA}, \alpha) < \alpha_0$

   yields p_g^I,

2. the fraction of random numbers pairs for which $R(X_{DATA}; x_{DATA}, \alpha) > \alpha_0$

   yields p_g^{II}, and

3. the fraction of random numbers pairs for which

$$\min[R(X_{DATA}; x_{DATA}, \alpha) < \alpha_0, \ R(X_{DATA}; x_{DATA}, \alpha) > \alpha_0] \text{ yields } p_g^{III}.$$

3.6.2 Confidence intervals for α

Since

1. the value of $\dfrac{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]}}{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]^2}}$ is α, and

2. the distribution of $\dfrac{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]}}{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]^2}}$ is independent of any unknown

   parameters,

$R(X_{DATA}; x_{DATA}, \alpha)$ is a generalized pivotal quantity for constructing $100(1-\gamma)\%$

confidence interval for α, where $1-\gamma$ is the confidence coefficient (Weerahandi 1993).

Now, one-sided as well as two-sided confidence intervals are constructed as follows:

1. the lower bound, $R_\gamma(\alpha; a, \mathbf{t})$, of a $100(1-\gamma)\%$ one-sided confidence

   interval for α is sought such that

$$1 - \gamma = \Pr\left(\frac{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]}}{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]^2}} \leq R_{1-\gamma}(\alpha; a, \mathbf{t}) \right), \qquad (3.63)$$

2. the upper bound, $R_{1-\gamma}(\alpha; a, \mathbf{t})$, of a $100(1-\gamma)\%$ one-sided confidence

   interval for α is sought such that

$$\gamma = \Pr\left(\frac{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]}}{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]^2}} \leq R_\gamma(\alpha; a, \mathbf{t}) \right), \qquad (3.64)$$

3. the lower limit, $R_{1-\gamma/2}^{l}(\alpha; a, \mathbf{t})$, and the upper limit, $R_{\gamma/2}^{u}(\alpha; a, \mathbf{t})$, of a $100(1$

   $-\gamma)\%$ two-sided confidence interval for α, are sought such that

$$1 - \gamma = \Pr\left(R_{1-\gamma/2}^{l}(\alpha; a, \mathbf{t}) \leq \frac{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]}}{\sum_{i=1}^{k} \frac{(m_i-2)^2(m_i-3)}{[m_i R(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i)]^2}} \leq R_{\gamma/2}^{u}(\alpha; a, \mathbf{t}) \right),$$

$$(3.65)$$

where $\mathbf{t} = (t_1, t_2, ..., t_k)$ is the observed value of $\mathbf{T} = (T_1, T_2, ..., T_k)$,or simply an esti-

mate of $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)$, and $R_\gamma(\alpha; a, \mathbf{t})$ is the 100γ-th percentile of $R(X_{DATA}; x_{DATA}$

$, \alpha)$'s. Once the calculated $R(X_{DATA}; x_{DATA}, \alpha)$'s are ordered, the desired percentiles

can be obtained.

3.7 Illustrative examples for the common shape parameter

In order to illustrate the proposed procedure, the different number of samples

with sizes $m_{k(=2,6,12)} = 100\ 000$ are generated by keeping, for each case, the common

shape parameter and unequal scale parameters. Three cases are considered here for

the convenience and mathematical tractability:

1. 2-sample case with the same sizes m_2 for both samples –

   Parameter specifications for 4-repetitions are listed below:

   $\boldsymbol{\theta}_1^{(2)} = (100, 102)$, $\boldsymbol{\theta}_2^{(2)} = (104, 106)$,

   $\boldsymbol{\theta}_3^{(2)} = (108, 110)$, $\boldsymbol{\theta}_4^{(2)} = (112, 114)$, $\alpha^{(2)} = 1.2$

2. 6-sample case with the same sizes m_6 for all six samples –

   Parameter specifications for 4-repetitions are listed below:

   $\boldsymbol{\theta}_1^{(6)} = (\boldsymbol{\theta}_1^{(2)}, \boldsymbol{\theta}_2^{(2)}, \boldsymbol{\theta}_3^{(2)})$, $\boldsymbol{\theta}_2^{(6)} = (\boldsymbol{\theta}_1^{(2)}, \boldsymbol{\theta}_2^{(2)}, \boldsymbol{\theta}_4^{(2)})$,

   $\boldsymbol{\theta}_3^{(6)} = (\boldsymbol{\theta}_2^{(2)}, \boldsymbol{\theta}_3^{(2)}, \boldsymbol{\theta}_4^{(2)})$, $\boldsymbol{\theta}_4^{(6)} = (\boldsymbol{\theta}_3^{(2)}, \boldsymbol{\theta}_4^{(2)}, \boldsymbol{\theta}_1^{(2)})$, $\alpha^{(6)} = 1.3$

3. 12-sample case with the same sizes m_{12} for all twelve samples –

   Parameter specifications for 4-repetitions are listed below

   $\boldsymbol{\theta}_1^{(12)} = (\boldsymbol{\theta}_1^{(6)}, \boldsymbol{\theta}_2^{(6)})$, $\boldsymbol{\theta}_2^{(12)} = (\boldsymbol{\theta}_1^{(6)}, \boldsymbol{\theta}_3^{(6)})$,

   $\boldsymbol{\theta}_3^{(12)} = (\boldsymbol{\theta}_1^{(6)}, \boldsymbol{\theta}_4^{(6)})$, $\boldsymbol{\theta}_4^{(12)} = (\boldsymbol{\theta}_2^{(6)}, \boldsymbol{\theta}_3^{(6)})$, $\alpha^{(12)} = 1.4$

Probability coverages for all above cases are given in Tables 3.9 through 3.11 for

parameter specifications for $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_i, ..., \theta_k)_{k(=2,6,12)}$ for each case $i = 1, 2, ..., k$ and $\alpha^{(k)} (k = 2, 6, 12)$.

Table 3.9 Probability coverages for 95% two-sided confidence intervals for α :

2-sample case

Parameters:	Sample size					
$k = 2$, $\boldsymbol{\theta}^{(2)}$, $\alpha^{(2)}$	10		20		50	
	GV*	C**	GV	C	GV	C
$\boldsymbol{\theta}_1^{(2)} = (\theta_{11}^{(2)}, \theta_{12}^{(2)})$, $\alpha^{(2)}$	0.97	0.86	0.96	0.93	0.95	0.93
$\boldsymbol{\theta}_2^{(2)} = (\theta_{21}^{(2)}, \theta_{22}^{(2)})$, $\alpha^{(2)}$	0.97	0.85	0.96	0.93	0.95	0.94
$\boldsymbol{\theta}_3^{(2)} = (\theta_{31}^{(2)}, \theta_{32}^{(2)})$, $\alpha^{(2)}$	0.96	0.89	0.96	0.91	0.94	0.94
$\boldsymbol{\theta}_4^{(2)} = (\theta_{41}^{(2)}, \theta_{42}^{(2)})$, $\alpha^{(2)}$	0.97	0.86	0.95	0.92	0.95	0.95

$* = $ the Generalized Variable Method; $** = $ the Classical Method

Table 3.10 Probability coverages for 95% two-sided confidence intervals for α :

6-sample case

Parameters:	Sample size					
$k = 6,\ \boldsymbol{\theta}^{(6)},\ \alpha^{(6)}$	10		20		50	
	GV*	C**	GV	C	GV	C
$\boldsymbol{\theta}_1^{(6)} = (\theta_{11}^{(6)}, \theta_{12}^{(6)}, ..., \theta_{16}^{(6)}),\ \alpha^{(6)}$	0.98	0.85	0.96	0.90	0.95	0.93
$\boldsymbol{\theta}_2^{(6)} = (\theta_{21}^{(6)}, \theta_{22}^{(6)}, ..., \theta_{26}^{(6)}),\ \alpha^{(6)}$	0.98	0.72	0.96	0.82	0.95	0.94
$\boldsymbol{\theta}_3^{(6)} = (\theta_{31}^{(6)}, \theta_{32}^{(6)}, ..., \theta_{36}^{(6)}),\ \alpha^{(6)}$	0.97	0.75	0.97	0.79	0.96	0.92
$\boldsymbol{\theta}_4^{(6)} = (\theta_{41}^{(6)}, \theta_{42}^{(6)}, ..., \theta_{46}^{(6)}),\ \alpha^{(6)}$	0.97	0.80	0.96	0.88	0.96	0.91

$*$ = the Generalized Variable Method; $**$ = the Classical Method

Table 3.11 Probability coverages for 95% two-sided confidence intervals for α :

12-sample case

Parameters:	Sample size					
$k = 12,\ \boldsymbol{\theta}^{(12)},\ \alpha^{(12)}$	10		20		50	
	GV*	C**	GV	C	GV	C
$\boldsymbol{\theta}_1^{(12)} = (\theta_{11}^{(12)}, \theta_{12}^{(12)}, ..., \theta_{112}^{(12)}),\ \alpha^{(12)}$	0.97	0.76	0.96	0.93	0.95	0.94
$\boldsymbol{\theta}_2^{(12)} = (\theta_{21}^{(12)}, \theta_{22}^{(12)}, ..., \theta_{212}^{(12)}),\ \alpha^{(12)}$	0.98	0.69	0.96	0.94	0.95	0.93
$\boldsymbol{\theta}_3^{(12)} = (\theta_{31}^{(12)}, \theta_{32}^{(12)}, ..., \theta_{312}^{(12)}),\ \alpha^{(12)}$	0.97	0.70	0.95	0.94	0.96	0.94
$\boldsymbol{\theta}_4^{(12)} = (\theta_{41}^{(12)}, \theta_{42}^{(12)}, ..., \theta_{412}^{(12)}),\ \alpha^{(12)}$	0.97	0.75	0.95	0.92	0.96	0.93

$*$ = the Generalized Variable Method; $**$ = the Classical Method

Here, the classical confidence interval estimation of α is given by

$$\left(a - Z_{1-\gamma/2}\sqrt{s^2(A)}, a + Z_{1-\gamma/2}\sqrt{s^2(A)}\right) \tag{3.66}$$

where Z_β is βth quantile of $Z \sim N(0,1)$ and $(a,\, s^2(A))$ is the observed value of $(A, S^2(A))$, or simply an estimate of $(\alpha, \sigma^2(A))$. Equivalently,

$$\left(a - Z_{1-\gamma/2}\sqrt{\frac{1}{\sum_{i=1}^k \frac{1}{s^2(\widehat{\alpha}_i)}}}, a + Z_{1-\gamma/2}\frac{1}{\sum_{i=1}^k \frac{1}{s^2(\widehat{\alpha}_i)}}\right), \tag{3.67}$$

where $(a,\, s^2(\widehat{\alpha}_i))$ is the observed value of $(A, S^2(\widehat{\alpha}_i))$, or simply an estimate of $(\alpha, \sigma^2(\widehat{\alpha}_i))$. Since,

$$\alpha = \frac{\sum_{i=1}^k w_i \alpha_i}{\sum_{i=1}^k w_i}, \tag{3.68}$$

where

$$w_i = \frac{1}{\sigma^2(\widehat{\alpha}_i)}, \tag{3.69}$$

by equation (3.56),

$$\sigma^2(\widehat{\alpha}_i) = \frac{(m_i \alpha_i)^2}{(m_i - 2)^2(m_i - 3)}, m_i > 3. \tag{3.70}$$

as $\widehat{\alpha}_i \sim \Gamma^{-1}(m_i - 1, m_i \alpha_i)$.

Then, the classical estimate of the common shape parameter α of several Pareto populations is defined as

$$\widehat{\alpha} = A = \frac{\sum_{i=1}^k \widehat{w}_i \widehat{\alpha}_i}{\sum_{i=1}^k \widehat{w}_i}, \tag{3.71}$$

where

$$\widehat{w}_i = \frac{1}{S^2(\widehat{\alpha}_i)}. \tag{3.72}$$

This yields

$$A = \frac{\sum_{i=1}^{k} \frac{\widehat{\alpha}_i}{S^2(\widehat{\alpha}_i)}}{\sum_{i=1}^{k} \frac{1}{S^2(\widehat{\alpha}_i)}}, \tag{3.73}$$

where $\widehat{\alpha}_i = A_i = m_i \left(\sum_{j=1}^{m_i} \ln(X_{ij}/\widehat{\theta}_i) \right)^{-1}$ and $S^2(\widehat{\alpha}_i)$ is an estimator of $\sigma^2(\widehat{\alpha}_i)$.

Then,

$$S^2(A) = S^2 \left(\frac{\sum_{i=1}^{k} \frac{\widehat{\alpha}_i}{S^2(\widehat{\alpha}_i)}}{\sum_{i=1}^{k} \frac{1}{S^2(\widehat{\alpha}_i)}} \right) = \frac{1}{\sum_{i=1}^{k} \frac{1}{S^2(\widehat{\alpha}_i)}} \tag{3.74}$$

When the empirical confidence levels of the generalized variable method are carefully compared with that of the classical method for different sample sizes in different sample cases, it can be inferred that the generalized variable method outperforms the classical method for this particular case.

.

CHAPTER IV

GENERALIZED INFERENCES FOR SEVERAL EXPONENTIAL

POPULATIONS

4.1 Introduction

Generalized inferences for the common mean of several normal populations (Shin-Hui and Jack 2005), for the common mean of several log-normal populations (Tian and Wu 2007), and for the common scale parameter of several Pareto populations (Gunasekera and Ananda 2008) have been discussed in the literature. This Chapter discusses the exact inferences based on the generalized variable method (Tsui and Weerahandi 1989) of the location parameter of several exponential distributions with an unknown common location parameter and unknown possibly unequal scale parameters. These situations arise in life testing and reliability, where the common location parameter can be taken as the minimum guarantee-time of operation of several components, and scale parameters are interpreted as unknown, and possibly unequal, failure rates of those components.

The Maximum Likelihood Estimator (MLE), the Modified MLE (MMLE), and the Uniformly Minimum Variance Unbiased Estimator (UMVUE) of the common location parameter are proposed and compared asymptotically by Ghosh and Razmpour (1984) in terms of their biases and Mean Squared Errors (MSE's). Furthermore, Jin and Crouse (1998) under a class of convex loss functions obtained the MLE, MMLE

and UMVUE of the common location parameter. Using certain classical independent tests based on the combination of the probabilities: namely, the Tippet method, the Fisher method, the inverse normal method, and the logit method, Sun and Sinha (1999) constructed confidence intervals for the common location parameter.

Using the generalized variable approach, we perform an exact test (Weerahandi 1989) and construct an exact confidence interval (Weerahandi 1993) for the common location parameterof the several exponential populations. Simulation studies to compare the performance of these generalized procedures with those of the approximate procedures and other exact methods are carried out.

4.1.1 Types of the exponential distribution

The exponential distribution of the random variable X with the location parameter θ and the scale parameter σ, (i.e. $X \sim Exp(\theta, \sigma)$) has the distribution function in the form of

$$F(x) = 1 - \exp[(x - \theta)/\sigma], \text{ where } \sigma > 0, \ 0 < \theta \leq x. \tag{4.1}$$

This is the commonly used definition of the exponential, but when it is expressed in terms of the rate parameter λ, the reciprocal of the scale parameter, then the exponential distribution (i.e., $X \sim Exp(\theta, 1/\lambda)$) is given in the form of

$$F(x) = 1 - \exp[\lambda(x - \theta)], \text{ where } \lambda > 0, \ 0 \leq x. \tag{4.2}$$

Note that, for our convenience $X \sim Exp(\theta, \sigma)$, which is henceforth referred to as 'exponential', is used throughout this dissertation.

4.2 Generalized inferences

Suppose $\mathbf{X}_i^{n_i} = (X_{i1}, X_{i2}, ..., X_{in_i})$ is a random sample of size n_i from an ith truncated exponenetial population, $X \sim Exp(\theta, \sigma_i)$, $i = 1, 2, ..., k$, where θ denotes the common unknown location parameter and σ_i is an unknown and possibly unequal scale parameter of the ith exponential population. Furthermore, suppose $\mathbf{x}_i^{m_i} = (x_{i1}, x_{i2}, ..., x_{im_i})$ is its observed value. Therefore, $\{X_{ij}\}_{i=1,2,...,k;j=1,2,...,n_i}$ is independently distributed as

$$X_{ij} \sim \sigma_i^{-1} \exp(-(x_{ij} - \theta)/\sigma_i)I\left[x_{i(1)} \geq \theta\right], -\infty < \theta < \infty, \sigma_i > 0, \forall i,$$

$$(4.3)$$

where $x_{i(1)} = \min(x_{i1}, x_{i2}, ..., x_{in_i})$ and I denotes the usual indicator function given by

$$I\left[a \geq b\right] = \begin{cases} 1 & \text{if} \quad a \geq b \\ 0 & \text{if} \quad a < b \end{cases}.$$

The mean and variance of the truncated i^{th} exponential population are, respectively, given by

$$E(X_i) = \theta + \sigma_i, \ Var(X_i) = \sigma_i, \ \text{for} \ \forall i = 1, 2, ..., k \qquad (4.4)$$

In addition, the median and mode are also given, respectively, by

$$Median(X_i) = (\theta + \sigma_i)\ln(2), \text{ and } Mode(X_i) = \theta, \text{ for } \forall i = 1, 2, ..., k$$

$$(4.5)$$

Johnson, Kotz, and Balakrishnan (1994) have discussed the maximum likelihood estimators of the location and scale parameters of a single exponential distribution.

Sinha and Kale (1980) discussed about their distributions. Prompted by this work, Ghosh and Razmpour (1984) analysed the overall maximum likelihood estimators of the common location parameter θ and scale parameters σ_i's – denoted by $\widehat{\theta}$ and $\widehat{\sigma}_i$'s, respectively – of several exponential populations. Moreover, the distributions of $\widehat{\theta}$ and $\widehat{\sigma}_i$'s were also discussed by Ghosh and Razmpour (1984). These results are exprssed in the following:

$$\widehat{\theta} = T = \min_{1 \leq i \leq k} U_i(X_i) \text{ and } T - \theta \sim \Gamma(1, \sigma_*^{-1}), \tag{4.6}$$

where $U_i(X_i) = \min\limits_{1 \leq j \leq n_i} X_{ij}$ and $\sigma_* = \sum_{i=1}^{k} n_i \sigma_i^{-1}$, and

$$\widehat{\sigma}_i = S_i = n_i^{-1} \sum_{j=1}^{n_i} (X_{ij} - \widehat{\theta}) \text{ and } S_i \sim \Gamma(n_i - 1, n_i^{-1}\sigma_i), \text{ for } \forall i = 1, 2, ..., k$$

$$\tag{4.7}$$

4.2.1 The statistical testing of hypothesis

The problem of testing

$$H_0 : \theta \leq \theta_0 \text{ vs. } H_a : \theta > \theta_0, \tag{4.8}$$

where θ_0 is a known quantity, is considered.

Even though approximate tests and confidence intervals based on the large sample asymptotic approach are available, when σ_i's are unknown, there are no exact tests or confidence intervals for θ available in the literature. In other words, when σ_i's are unknown, it is difficult to find a pivotal quantity that is free of the parameter of interest, as well as of nuisance parameters. Tsui and Weerahandi (1989) introduced the exact generalized inference approach that can help address such problems.

65

Let us consider that $\mathbf{x}_i^{n_i} = (x_{i1}, x_{i2}, ..., x_{in_i})$ is the observed random sample of size n_i from $X_i \sim Exp(\theta, \sigma_i)$, $i = 1, 2, ..., k$. As discussed in Weerahandi (1995, 2004), θ can be expressed in terms of sufficient statistics and random variables as

$$\theta = T - V \left(\sum_{i=1}^{k} W_i/S_i \right)^{-1}, \tag{4.9}$$

where

$$V = 2\sigma_*(T - \theta) \sim \varkappa_2^2 \text{ and } W_i = 2n_i\sigma_i^{-1}S_i \sim \varkappa_{2(n_i-1)}^2, \tag{4.10}$$

with $\sigma_* = \sum_{i=1}^{k} n_i\sigma_i^{-1}$.

Then, the generalized pivot variable for θ is given by

$$R_g(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\sigma}) = t - V \left(\sum_{i=1}^{k} W_i/s_i \right)^{-1}, \tag{4.11}$$

where (t, s_i) is the observed value of (T, S_i), $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_n)$, x_{DATA} is the observed value of X_{DATA}; $X_{DATA} = [X_{ij}]_{i=1,...,k; j=1,...,m_i}$ and $x_{DATA} = [x_{ij}]_{i=1,2,...,k; j=1,,...,m_i}$.

Then, the generalized p-value for our procedure to test H_0 in (4.8) is given by

$$p_g = 1 - E_{W_{i(=1,2,...,k)}}[H_V(t - \theta_0 \sum_{i=1}^{k} W_i/s_i)], \tag{4.12}$$

where $E_{W_{i(=1,2,...,k)}}$ is the expectation with respect to independent chi-squared random variables $W_i(i = 1, 2, ..., k)$ defined in (4.10), and H_V is the *cdf* of V defined in (4.10).

This p-value can be either computed by numerical integration exact up to a desired level of accuracy or well approximated by a Monte Carlo method. When there are a large number of factor combinations, the latter method is more desirable and computationally more efficient.

The computations using the representation (4.12) are carried out as follows:

66

1. Generate a set of large number of random numbers from each chi-squared random variable $W_i \sim \varkappa^2_{2(n_i-1)}$ for $i = 1, 2, ..., k$.

2. Compute, for each set, the $cdf = H_V(t - \theta_0 \sum_{i=1}^{k} W_i/s_i)$, where $V \sim \varkappa^2_2$.

3. Compute their average with respect to $W_{i(=1,2,...,k)}$, say

$$E_{W_{i(=1,2,...,k)}}\left[H_V(t - \theta_0 \sum_{i=1}^{k} W_i/s_i) \right].$$

4. Estimate the generalized p-value by $1 - E_{W_{i(=1,2,...,k)}}\left[H_V(t - \theta_0 \sum_{i=1}^{k} W_i/s_i) \right].$

4.2.2 Confidence intervals

Since

1. the value of $R_g(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\sigma})$ is θ, and

2. the distribution of $R_g(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\sigma})$ is independent of any unknown parameters,

$R_g(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\sigma})$ is a generalized pivotal quantity for constructing $100(1-\gamma)\%$ confidence interval for θ, where $1 - \gamma$ is the confidence coefficient (Weerahandi 1993).

Now, one-sided as well as two-sided confidence intervals are constructed as follows:

1. the lower bound, $R_{1-\gamma}(\theta; t, \mathbf{s})$, of a $100(1 - \gamma)\%$ one-sided confidence interval for θ is sought such that

$$1 - \gamma = \Pr\left[t - V\left(\sum_{i=1}^{k} W_i/s_i \right)^{-1} \leq R_{1-\gamma}(\theta; t, \mathbf{s}) \right] \qquad (4.13)$$

2. the upper bound, $R_{\gamma}(\theta; t, \mathbf{s})$, of a $100(1 - \gamma)\%$ one-sided confidence interval for θ is sought such that

$$\gamma = \Pr\left[t - V\left(\sum_{i=1}^{k} W_i/s_i \right)^{-1} \leq R_{\gamma}(\theta; t, \mathbf{s}) \right], \qquad (4.14)$$

3. the lower limit, $R^{l}_{1-\gamma/2}(\theta; t, \mathbf{s})$, and the upper limit, $R^{u}_{\gamma/2}(\theta; t, \mathbf{s})$, of a $100(1-$

$\gamma)\%$ two-sided confidence interval for θ are sought such that

$$1 - \gamma = \Pr(R^l_{1-\gamma/2}(\theta; t, \mathbf{s})) \leq t - V \left(\sum_{i=1}^{k} W_i / s_i \right)^{-1} \leq R^u_{\gamma/2}(\theta; t, \mathbf{s})),$$

(4.15)

where $\mathbf{s} = (s_1, s_2, ..., s_n)$ is the observed value of $\mathbf{S} = (S_1, S_2, ..., S_n)$, or simply an estimate of $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, ..., \sigma_n)$, and $R_\gamma(\theta; t, \mathbf{s})$ is the 100γ-th percentile of $R_g(X_{DATA};$ $x_{DATA}, \theta, \boldsymbol{\sigma})$'s. Once the calculated $R_g(X_{DATA}; x_{DATA}, \theta, \boldsymbol{\sigma})$'s are ordered, the desired percentiles can be obtained.

To obtain actual coverage probabilities of θ (empirical confidence levels for θ), it is necessary:

1. to repeat the above process for larger number of times (i.e. 100 000−

    1 000 000), and

2. to calculate the fraction of times θ falls within calculated (empirical)

    generalized confidence intervals.

4.3 The Bayesian estimation

The Bayesian estimation of the common location parameter of several exponential distributions is discussed in this Section.

Consider k independent exponentially distributed populations – i.e. $Exp(\theta, \sigma_i)$ for $i = 1, 2, ..., k$, where parameters $\sigma_1, \sigma_2, ...\sigma_k$ are k unknown scale parameters and θ is the unknown common location parameter. Suppose the x_{ij} are k independent samples where $j = 1, 2, ..., m_i$, then the joint density of the sample variables is

$$f(X_{DATA} | \theta, \sigma_1, \sigma_2, ...\sigma_k) = \prod_{i=1}^{k} \sigma_i^{-m_i} \exp\left[-\sigma_i^{-1} \left(\sum_{j=1}^{m_i} x_{ij} - m_i\theta \right) \right],$$

(4.16)

where

$$X_{DATA} = [x_{ij}]_{i=1,2,...,k; \ j=1,2,...,m_i} \text{ and } f(x_{ij}) = \sigma_i^{-1} \exp(-(x_{ij}-\theta)/\sigma_i); \ \theta \text{ real}, \sigma_i > 0, \forall i.$$

4.3.1 The derivation of the Bayes estimate of θ when $\boldsymbol{\sigma}$ is known

A vague prior density is assumed for the parameters θ and $\sigma_1, \sigma_2, ...\sigma_k$, namely

$$\pi(\theta, \sigma_1, \sigma_2, ...\sigma_k) = \prod_{i=1}^{k} \sigma_i^{-1}. \tag{4.17}$$

With the priors of this form, the joint posterior density of θ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, ...\sigma_k)$ may be seen to be of the form

$$g(\theta, \sigma_1, \sigma_2, ...\sigma_k \,|X_{DATA}) = \prod_{i=1}^{k} \sigma_i^{-(m_i+1)} \exp\left[-\sigma_i^{-1}\left(\sum_{j=1}^{m_i} x_{ij} - m_i\theta\right)\right]. \tag{4.18}$$

The conditional posterior distribution of θ given k parameters $\sigma_1, \sigma_2, ...\sigma_k$ is in the from of the exponential distribution

$$g(\theta \,|\sigma_1, \sigma_2, ...\sigma_k) = \tau(X_{DATA}, \boldsymbol{\sigma}) \exp\left[\sigma_i^{-1} m_i\theta\right], \ -\infty < \theta < \infty, \tag{4.19}$$

where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, ...\sigma_k)$ and

$$\tau(X_{DATA}, \boldsymbol{\sigma}) = \prod_{i=1}^{k} m_i^{-1} \sigma_i^{-m_i} \exp\left(-\sigma_i^{-1} \sum_{j=1}^{m_i} x_{ij}\right). \tag{4.20}$$

Equivalently,

$$\theta \,|\sigma_1, \sigma_2, ...\sigma_k \sim EXP(\sum_{i=1}^{k} \sigma_i m_i^{-1}). \tag{4.21}$$

Assuming the squared-error loss, the Bayes estimate of θ will be given by

$$\widehat{\theta}_B^{Mean} = \sum_{i=1}^{k} \sigma_i m_i^{-1}, \tag{4.22}$$

where $\widehat{\theta}_B^{Mean}$ is the posterior mean (posterior mean estimate of θ). If instead, we assume the absolute-error loss, the Bayes estimate of θ will be given by

$$\widehat{\theta}_B^{Median} = \ln(2) \sum_{i=1}^{k} \sigma_i m_i^{-1} \tag{4.23}$$

where $\widehat{\theta}_B^{Median}$ is the the posterior median (posterior median estimate of θ).

4.3.2 The derivation of the Bayes estimate of θ when $\boldsymbol{\sigma}$ is unknown

The marginal posterior distribution of θ is obtained by integrating (4.18) with respect to σ_i's $(i = 1, 2, ..., k)$. But, using properties of the gamma distribution, the marginal posterior distribution of θ can be derived. It is given as follows:

$$m(\theta \, | X_{DATA}) = \prod_{i=1}^{k} \Gamma(m_i + 1) \left(\sum_{j=1}^{m_i} x_{ij} - m_i \theta \right)^{-(m_i+1)} \tag{4.24}$$

This distribution is a product of k non-standard distributions, called poly non-standard distributions; it is difficult to work with and numerical integration must be used to to determine moments and other characteristics of this marginal distribution.

In general, for a given θ, the conditional posterior distribution of the parameter σ_i $(i = 1, 2, ..., k)$ is the independent gamma density shown as follows:

$$\sigma_i \, |\theta \sim INVERSEGAMMA \left[(m_i + 1), \left(\sum_{j=1}^{m_i} x_{ij} \right)^{-1} \right], \text{ for } i = 1, 2, ..., k. \tag{4.25}$$

The conditional posterior distribution of θ given k parameters $\sigma_1, \sigma_2, ... \sigma_k$ is:

$$\theta \, |\sigma_1, \sigma_2, ... \sigma_k \sim EXP(\sum_{i=1}^{k} \sigma_i m_i^{-1}). \tag{4.26}$$

Using the Gibbs sampling technique, the Bayes estimate of θ, when $\boldsymbol{\sigma}$ is unknown, is given by

$$\theta_B^s = \sum_{n=1}^{N} \theta_n^{(G)} / N \, . \tag{4.27}$$

The steps to Gibbs sampling algorithms are as follows:

1. Find initial values for θ and $\alpha_{i(=1,2,...,k)}$, i.e., $\theta^{(0)}$ and $\sigma_{i(=1,2,...,k)}^{(0)}$, where

$$\theta^{(0)} = \min_{1 \leq i \leq k} X_{i(1)} \qquad \text{and} \qquad \sigma_i^{(0)} = n_i^{-1} \sum_{j=1}^{n_i} (X_{ij} - \theta^{(0)}),$$

2. Draw $\theta^{(1)}$ from the conditional power distribution of θ given

$$\sigma_1 = \sigma_1^{(0)}, \sigma_2 = \sigma_2^{(0)}, ..., \sigma_k = \sigma_k^{(0)}$$

i.e., $\theta^{(1)} \sim \theta \left| \boldsymbol{\sigma} = (\sigma_1^{(0)}, \sigma_2^{(0)}, ..., \sigma_k^{(0)}) \right.$,

3. Draw $\sigma_i^{(1)'}$s from the conditional gamma distribution of σ_i given $\theta = \theta^{(1)}$

for $i = 1, 2, ..., k$

i.e., $\sigma_i^{(1)} \sim \sigma_i \left| \theta = \theta^{(1)} \right.$ for $i = 1, 2, ..., k$,

4. Draw $\theta^{(2)}$ from the conditional power distribution of θ given $\sigma_1 = \sigma_1^{(1)}, \sigma_2 = $

$$\sigma_2^{(1)}, ..., \sigma_k = \sigma_k^{(1)}$$

i.e., $\theta^{(2)} \sim \theta \left| \boldsymbol{\sigma} = (\sigma_1^{(1)}, \sigma_2^{(1)}, ..., \sigma_k^{(1)}) \right.$,

5. Draw $\sigma_i^{(2)'}$s from the conditional gamma distribution of σ_i given $\theta = \theta^{(2)}$

for $i = 1, 2, ..., k$

i.e., $\sigma_i^{(2)} \sim \sigma_i \left| \theta = \theta^{(2)} \right.$ for $i = 1, 2, ..., k$,

6. This process is continued for $g = 1, 2, ...G$ iterations,

7. Then, the above Gibbs process is repeated $n = 1, 2, ..., N$ times.

This is schematically represented in Table 4.1. At this point, a random sample of size N from the joint posterior distribution of θ and σ_i's for $i = 1, 2, ..., k$, a random sample of size N from the marginal posterior distribution of θ, and a random samples of size N from the marginal posterior distributions σ_i's for $i = 1, 2, ..., k$, can easily be obtained as follows:

Let's denote $\theta_n^{(G)}, \alpha_{1n}^{(G)}, ..., \alpha_{kn}^{(G)}$ by J_n for $n = 1, 2, ..., N$

$$\theta_1^{(G)}, \theta_2^{(G)}, ..., \theta_N^{(G)} \text{ by } M_\theta$$

$$\sigma_{i1}^{(G)}, \sigma_{i2}^{(G)}, ..., \sigma_{iN}^{(G)} \text{ by } M_{iN}^{\sigma_i} \text{ for } i = 1, 2, ..., k$$

$$\text{Then } J_{n(=1,2,...,N)} \sim g(\theta, \boldsymbol{\alpha} | X_{DATA}),$$

$$M_\theta \sim m(\theta | X_{DATA}), \text{and}$$

$$M_{iN(=1,2,...,k)}^{\sigma_i} \sim m(\sigma_{i(=1,2,...,k)} | X_{DATA}).$$

Table 4.1 Random sample from the posterior distribution

of $(\theta, \sigma_1, ..., \sigma_k)$

#	θ	σ_1	σ_2	.	.	.	σ_i	.	.	.	σ_k
1	$\theta_1^{(t)}$	$\sigma_{11}^{(G)}$	$\sigma_{21}^{(G)}$	.	.	.	$\sigma_{i1}^{(G)}$	.	.	.	$\sigma_{k1}^{(G)}$
2	$\theta_2^{(G)}$	$\sigma_{12}^{(G)}$	$\sigma_{22}^{(G)}$	.	.	.	$\sigma_{i2}^{(G)}$	.	.	.	$\sigma_{k2}^{(G)}$
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
n	$\theta_n^{(G)}$	$\sigma_{1n}^{(G)}$	$\sigma_{2n}^{(G)}$	.	.	.	$\sigma_{in}^{(G)}$	.	.	.	$\sigma_{kn}^{(G)}$
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
N	$\theta_N^{(G)}$	$\sigma_{1N}^{(G)}$	$\sigma_{2N}^{(G)}$	.	.	.	$\sigma_{iN}^{(G)}$	.	.	.	$\sigma_{kN}^{(G)}$

Each column in Table 4.1 shows a sample of values of marginal distributions of

$\theta, \sigma_1, \sigma_2, ...,$ and σ_k that are generated by the Gibbs sampler from conditional distri-

butions. If N is large, the mean and variance of the marginal posterior distribution

of θ are given, respectively, by

$$E\left(\theta\,|X_{DATA}\right) = \sum_{n=1}^{N} \theta_n^{(G)} /N = \overline{\theta}, \text{ and} \tag{4.28}$$

$$Var(\theta\,|X_{DATA}) = (N-1)^{-1} \sum_{n=1}^{N} \left(\theta_n^{(G)} - \overline{\theta}\right)^2 \tag{4.29}$$

Therefore, the Bayes estimate of θ, when $\boldsymbol{\sigma}$ is unknown, using the Gibbs sampler is given by

$$\theta_B^G = \sum_{n=1}^{N} \theta_n^{(G)} /N \tag{4.30}$$

The sample variance is given by

$$\sigma_{(\theta|X_{DATA})}^2 = (N-1)^{-1} \sum_{n=1}^{N} \left(\theta_n^{(G)} - \overline{\theta}\right)^2. \tag{4.31}$$

Exact and approximate credible sets calculated for θ are given, respectively, as follows:

(i) the approximate $100(1-\gamma)\%$ credible region for θ

$$\sum_{n=1}^{N} \theta_n^{(G)} /N \pm Z_{\gamma/2} \sqrt{\frac{(N-1)^{-1} \sum_{n=1}^{N} \left(\theta_n^{(G)} - \overline{\theta}\right)^2}{N}}. \tag{4.32}$$

(ii) the actual $100(1-\gamma)\%$ credible region for θ

$$l = \prod_{i=1}^{k} \Gamma(m_i + 1) \left(\sum_{j=1}^{m_i} x_{ij} - m_i\theta\right)^{-(m_i+1)}, \tag{4.33}$$

To obtain the actual credible region for θ, we solve the above equation for θ with appropriate values of l, until the values of θ get as close as to the values found in (4.32); while the area under the curve, given by the left hand side of (4.33), is $100(1-\gamma)\%$.

4.4 Illustrative examples for the common location parameter

Example 1. Comparison of the proposed procedure with the classical approach

based on large sample method

This example deals with the exponential distributions $X_i \sim Exp\left(\theta, \sigma_i\right)$ where $i = 1, 2, 3$ generated by the following population parameters: $\theta = 100$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_2 = 3$ and sample sizes $n_1 = n_2 = n_3 = 10$. The data generated from these distributions are:

$X_1 \sim Exp\left(\theta, \sigma_1\right)$: $100.4, 100.7, 101.5, 100.1, 101.4, 101.4, 100.6, 101.2, 101.0, 103.0$

$X_2 \sim Exp\left(\theta, \sigma_2\right)$: $103.9, 101.7, 101.1, 101.0, 101.7, 104.2, 104.3, 100.4, 102.5, 101.6$

$X_3 \sim Exp\left(\theta, \sigma_3\right)$: $100.9, 103.5, 100.7, 105.4, 106.2, 100.9, 100.9, 101.7, 105.9, 100.0$

I. Confidence intervals:

Assuming that all of the above parameters are unknown:

1. the lower bound of a 95% one-sided generalized empirical confidence

   interval for θ calculated from this data is 98.0115,

2. the upper bound of a 95% one-sided generalized empirical confidence

   interval for θ calculated from this data is 100.0293,

3. the upper and lower limits of a 95% two-sided generalized empirical

   confidence interval for θ calculated from this data, respectively, are

   97.4630 and 100.0662.

The comparison of classical and generalized confidence intervals is summarized in Table 4.2 as follows:

Table 4.2 Comparison of generalized and classical confidence intervals

Confidence interval	Generalized	Classical
95%	$(97.4630, 100.0662)$	$(96.9688, 100.6983)$

When comparing these lengths for θ, it is clear that the generalized variable method, having a shorter length, outperforms its counter part, the classical procedure, for this particular problem.

II. p-values:

Furthermore, the comparison of classical and generalized p-values for various tests is summarized in Table 4.3 as follows:

Table 4.3 Comparison of generalized and classical p-values

Test	Generalized	Classical
$\theta \leq 102$ vs. $\theta > 102$	0.9998	0.4951
$\theta \leq 98$ vs. $\theta > 98$	0.0000	0.4770

When θ, assuming that it is unknown, is tested, to see whether it is greater than 102 (or the claim or the alternative hypothesis being $\theta > 102$ against $\theta \leq 102$), the generalized p-value, having a value greater than 0.05, is against the claim, which is true since θ is 100. However, the classical p-value, having a value greater than 0.05, also suggests that the claim is not accepted. But, when θ, assuming that it is unknown, is tested to see whether it is less than 98, the generalized p-value, having a value less than 0.05, is supportive of the claim which is true- since θ is100; however, the classical

p-value, having a value greater than 0.05, suggests that the claim is not accepted. Both these arguments clearly show that the generalized variable method provides accurate, reliable, and non-misleading results, while the classical approach fails to do so for this particular case. Hence, the generalized variable method outperforms the classical approach for this particular case.

III. Sizes:

Table 4.4 shows the classical and the generalized empirical (actual) type I error rates (sizes of tests) for testing

1. $H_0^I : \theta \leq 100$ vs. $H_a^I : \theta > 100$,

2. $H_0^I : \theta \leq 500$ vs. $H_a^I : \theta > 500$,

when nominal (intended) type I error rate is at 0.1. All results are based on $100,000$ simulations.

Table 4.4 Comparison of generalized and classical actual type I error rates

Parameters: k, θ, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$	Generalized	Classical
$k = 3$, $\theta = 100$, $\boldsymbol{\sigma} = (1, 2, 3)$	0.044	0.000
$k = 3$, $\theta = 100$, $\boldsymbol{\sigma} = (4, 5, 6)$	0.042	0.000
$k = 3$, $\theta = 100$, $\boldsymbol{\sigma} = (7, 8, 9)$	0.040	0.000
$k = 3$, $\theta = 500$, $\boldsymbol{\sigma} = (1, 2, 3)$	0.046	0.000
$k = 3$, $\theta = 500$, $\boldsymbol{\sigma} = (4, 5, 6)$	0.048	0.000
$k = 3$, $\theta = 500$, $\boldsymbol{\sigma} = (7, 8, 9)$	0.054	0.000

Here,

$$H_0 : \theta \leq 100 \quad \text{vs.} \quad H_a : \theta > 100, \tag{4.34}$$

is tested with respect to the classical framework based on the large sample approach. Suppose $\boldsymbol{v} = (\theta, \boldsymbol{\eta})$ is a vector of unknown parameters, θ is the parameter of interest, and $\boldsymbol{\eta}$ a vector of nuisance parameters. Then, the classical test variable, $T_\theta^c = T(\mathbf{X}; \boldsymbol{v})$, is derived. It is known that

$$(T - \mu_T)(\sigma_T^2)^{-1/2} = Z \sim N(0, 1), \tag{4.35}$$

where Z is the standard normal variate, $\mu_\theta = \theta$ is the mean, and $\sigma_T^2 = (\sum_{i=1}^{k} n_i \sigma_i^{-1})^{-2}$ is the asymptotic variance. Then, the classical pivotal quantity for estimating and testing θ is given by

$$R_\theta^c = (T - \theta)(\sum_{i=1}^{k} n_i \sigma_i^{-1}). \tag{4.36}$$

When θ is specified, T_θ^c has a probability distribution that is free of nuisance parameters. Furthermore, when $\mathbf{x}$ and nuisance parameters are fixed, the *cdf* of T_θ^c is a monotonically increasing function of θ for any given t_θ^c. Therefore, T_θ^c is a classical test variable that can be used to test the given hypothesis.

Thus, the classical p-value for testing $\theta \leq 100$ vs. $\theta > 100$, is given by

$$p_c = 1 - \Phi(r_\theta^c), \tag{4.37}$$

where $r_\theta^c = (t - 100)(\sum_{i=1}^{k} n_i \sigma_i^{-1})$ is the observed value of R_θ^c; (t, s_i) being the observed value of (T, S_i) and $\Phi(.)$ is the distribution function of Z.

According to this simulation study, when compared with the classical procedure, actual type I error rates (actual sizes of tests) of the generalized procedure get as close

77

as to the intended size. Hence, the generalized procedure outpeforms the classical procedure for this particualr problem.

IV. Powers:

Table 4.5 shows the power comparison for testing $H_0^I : \theta \leq 100$ vs. $H_a^I : \theta > 100$ without and after adjusting the size at $\gamma = 0.10$ based on 100 000 replications.

Table 4.5 Comparison of powers for testing H_0^I without and after adjusting the size

	Without		With	
Parameters: k, θ, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$	GVM	CM	GVM	CM
$k = 3$, $\theta = 100$, $\boldsymbol{\sigma} = (1, 2, 3)$	0.044	0.000	0.050	0.050
$k = 3$, $\theta = 101$, $\boldsymbol{\sigma} = (3, 4, 5)$	0.321	0.298	0.318	0.315
$k = 3$, $\theta = 102$, $\boldsymbol{\sigma} = (7, 8, 9)$	0.521	0.410	0.501	0.493
$k = 3$, $\theta = 103$, $\boldsymbol{\sigma} = (10, 11, 12)$	0.798	0.623	0.778	0.770
$k = 3$, $\theta = 104$, $\boldsymbol{\sigma} = (13, 14, 15)$	0.895	0.712	0.883	0.875
$k = 3$, $\theta = 105$, $\boldsymbol{\sigma} = (16, 17, 18)$	0.955	0.812	0.948	0.935

When compared, the values of generalized and classical powers, with and without adjusting the size, clearly suggest that the generalized variable method outperforms the classical method in terms of power for this particular case.

V. Coverage probabilities:

The comparison of generalized coverage probabilities with that of the classical counter part, with the intended confidence level $\gamma = 0.05$, is given in Table 4.6.

Table 4.6 Comparison of generalized and classical

probability coverages for 95% two-sided confidence

intervals for θ

Parameters: k, θ, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$	Generalized	Classical
$k = 3,\ \theta = 100,\ \boldsymbol{\sigma} = (1, 2, 3)$	0.958	0.998
$k = 3,\ \theta = 100,\ \boldsymbol{\sigma} = (4, 5, 6)$	0.938	0.997
$k = 3,\ \theta = 100,\ \boldsymbol{\sigma} = (7, 8, 9)$	0.942	0.999
$k = 3,\ \theta = 500,\ \boldsymbol{\sigma} = (1, 2, 3)$	0.940	0.996
$k = 3,\ \theta = 500,\ \boldsymbol{\sigma} = (4, 5, 6)$	0.944	1.000
$k = 3,\ \theta = 500,\ \boldsymbol{\sigma} = (7, 8, 9)$	0.940	0.999

Here, a $100(1 - \gamma)\%$ empirical large sample solution for confidence interval estimation of θ is given by

$$\left(t - Z_{\gamma/2} \sum\nolimits_{i=1}^{k} n_i s_i^{-1},\ \ t + Z_{\gamma/2} \sum\nolimits_{i=1}^{k} n_i s_i^{-1} \right), \qquad (4.38)$$

where $Z_{\gamma/2}$ is $\gamma/2$th quantile of $Z \sim N(0, 1)$.

According to these simulation results, one can clearly see that the actual (empirical) probability coverages for the generalized method get as close as to the intended (nominal) coverage probabilities. Thus, the generalized variable method outperforms the classical procedure for this particular case.

Example 2. The comparison of the proposed procedure with the classical

approach based on the inverse normal method

Table 4.7 shows the comparison of the expected lengths of $100(1-\gamma)\%$ confidence intervals for θ, $1-\gamma$ being the confidence coefficient or the confidence level, based on the generalized variable method and the inverse normal method (Sun and Sinha 1999). Confidence intervals based on the inverse normal method are constructed using certain classical independent tests – namely, the Tippett's method, the Fisher's method, the inverse normal method, and the logit method using the Meta-analysis for the combinations of p-values. This is done on the basis of simulation. Consider $k = 2$ and take $(n_1, n_2) = (10, 5)$, $(10, 10)$, $(10, 15)$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, 1, 2 and $\gamma = 0.05$. Note that, since the inverse normal method outperforms other methods in Sun and Sinha (1999), the comparison is done between the confidence lengths, based on the inverse normal method and the proposed generalized method.

Table 4.7 Comparison of expected lengths of $100(1-\gamma)\%$ confidence intervals for θ based on the generalized variable method and the inverse normal method

σ_1	σ_2	n_1	n_2	Generalized	Inverse normal
1	0.5	10	5	0.1864	0.1986
			10	0.1146	0.1274
	1		5	0.2484	0.2662
			10	0.1588	0.1878
	2		5	0.3055	0.3385
			10	0.2265	0.2552

The inverse normal method:

Let us consider that $\mathbf{x}_i^{n_i} = (x_{i1}, x_{i2}, ..., x_{in_i})$ is the observed random sample of $\mathbf{X}_i^{n_i} = (X_{i1}, X_{i2}, ..., X_{in_i})$ of size n_i from $X_i \sim Exp(\theta, \sigma_i)$, $i = 1, 2, ..., k$.

Then, p-value for testing the hypothesis, based on the ith sample,

$$H_0^i : \theta \leq \theta_0 \text{ vs. } H_a^i : \theta > \theta_0 \tag{4.39}$$

is given by

$$P_i = \Pr(F_{2,2(n_i-1)} > f_i) = \left(\frac{s_i}{\overline{x}_i - \theta}\right)^{n_i - 1} \sim U(0, 1), \tag{4.40}$$

where $\overline{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}$ and f_i, and s_i are, respectively, the observed value of F_i and S_i, with

$$F_i = \frac{V_i/2}{W_i/2n_i - 2} = \frac{(n_i - 1)(T_i - \theta)}{S_i} \sim F_{2,2(n_i-1)} \tag{4.41}$$

being derived from each random variables

$$W_i(i = 1, 2, ..., k) \text{ defined as } W_i = 2n_i\sigma_i^{-1}S_i \sim \varkappa_{2n_i-2}^2, \tag{4.42}$$

$$V_i(i = 1, 2, ..., k) \text{ defined as } V_i = 2n_i(T_i - \theta)\sigma_i^{-1} \sim \varkappa_2^2. \tag{4.43}$$

Then, the inverse normal method rejects hypotheses of θ, i.e., $H_0^{i=1,2,...,k} : \theta \leq \theta_0$, if

$$\frac{\sum_{i=1}^{k} \Phi^{-1}(P_i)}{k} \geq -z_\beta, \tag{4.44}$$

at the level β where Φ is the standard normal *cdf* and $\Phi(z_\beta) = 1 - \beta$.

The results in Table 4.7 show that the generalized variable method produces much shorter intervals than the classical method produces, for each parametric specification in Sun and Sinha (1999). Therefore, the generalized variable method outperforms even the exact classical procedure for this particular problem.

Example 3. The Bayesian approach

This example deals with the exponential distributions $X_i \sim Exp(\theta, \sigma_i)$ where $i = 1, 2, 3$ generated by the following population parameters: $\theta = 100$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_2 = 3$; and sample sizes $n_1 = n_2 = n_3 = 10$. Assuming that the shape parameters $\boldsymbol{\alpha} = (\sigma_1, \sigma_2, \sigma_3)$ and θ are unknown, the Bayesian estimate of the common scale parameter θ is evaluated from the Gibbs sampler, using macros of MINITAB. The following are the Gibbs sampling parameter specifications: $G = 50$, $\mathbf{N} = (250, 500, 750, 1000)$, given in accordance with the Gelfand, Hills, Racine-Poon, and Smith's (1990) suggestion on G and N where they recommended that G be approximately 50 iterations and N be less than or equal to 1000.

The estimated values of θ determined by Gibbs sampling method are reported in Table 4.8.

Table 4.8 Bayes estimates and 95% credible regions using the Gibbs sampler for θ when $\boldsymbol{\sigma}$ is unknown

m	θ	STD	95% Credible Region	
			Lower	Upper
250	100.01	0.84	96.12	101.21
500	100.02	0.85	95.23	102.25
750	100.05	0.89	95.25	101.23
1000	100.02	0.90	96.38	103.03

Results provide more accurate Bayesian estimates for the common scale parameter of several exponential populations for different Gibbs sampling sizes.

4.5 Generalized inferences for the common scale parameter.

Inspired by the parallel work by Tian and Wu (2007), as in several Pareto case, we utilize the Meta-Analysis in performing generalized inferences for the common scale parameter of several exponential populations.

Consider k independent exponential populations with an unknown scale parameter σ_i and unknown location parameter θ_i. Let $X_{i1}, X_{i1}, ..., X_{im_i}$ be a random sample from the ith exponential population $X_{ij} \sim Exp\left(\theta_i, \sigma_i\right)$.

4.5.1 The statistical testing of hypothesis for σ

Suppose $\mathbf{X}_i^{n_i} = (X_{i1}, X_{i2}, ..., X_{in_i})$ is a random sample of size n_i from the truncated exponential populations and suppose $\mathbf{x}_i^{m_i} = (x_{i1}, x_{i2}, ..., x_{in_i})$ is the observed sample.

Then, as we know from the Section 4.2, the maximum likelihood estimators of θ_i and σ_i – denoted $\widehat{\theta}_i$ and $\widehat{\sigma}_i$, respectively – and their respective distributions are given by

$$\widehat{\theta}_i = T_i = X_{i(1)} \text{ and } T_i - \theta_i \sim \Gamma\left(1, \sigma_i/n_i\right), \tag{4.45}$$

$$\widehat{\sigma}_i = S_i = n_i^{-1} \sum_{j=1}^{n_i} (X_{ij} - \widehat{\theta}_i) \text{ and } S_i \sim \Gamma(n_i - 1, \sigma_i/n_i), \tag{4.46}$$

where $\Gamma(a, b)$ is the inverse gamma distribution with the shape parameter a and the scale parameter b.

Now, the generalized pivotal quantities for estimating θ_i and σ_i are given, respectively, by

$$R_g(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \theta_i, a_i) = t_i - V_i\left(W_i/s_i\right)^{-1} \text{ and } R_g(\mathbf{X}^{m_i}; \mathbf{x}^{m_i}, \alpha_i) = 2n_i s_i/W_i,$$

$$\tag{4.47}$$

where

$$V_i = 2n_i \sigma_i^{-1}(T_i - \theta_i) \sim \varkappa_2^2 \text{ and } W_i = 2n_i \sigma_i^{-1} S_i \sim \varkappa_{2(n_i-1)}^2, \tag{4.48}$$

with (t_i, s_i) being the observed value of (T_i, S_i).

Obviously, both generalized pivotal quantities, $R_g(\mathbf{X}^{n_i}; \mathbf{x}^{n_i}, \theta_i, a_i)$ and $R_g(\mathbf{X}^{n_i}; \mathbf{x}^{n_i}, \sigma_i)$, give the same result as classical pivotal quantities do. From the ith sample, the maximum likelihood estimator of σ_i is

$$\widehat{\sigma}_i = S_i = n_i^{-1} \sum_{j=1}^{n_i} (X_{ij} - \widehat{\theta}_i), \tag{4.49}$$

where $\widehat{\theta}_i = T_i = \min(X_{i1}, X_{i1}, ..., X_{im_i})$. The large sample variance for $\widehat{\sigma}_i$ is

$$Var(\widehat{\sigma}_i) = (n_i - 1)(\sigma_i/n_i)^2. \tag{4.50}$$

The generalized pivotal quantity for the common exponential scale parameter σ is a weighted average of the generalized pivot $R_g(\mathbf{X}^{n_i}; \mathbf{x}^{n_i}, \sigma_i)$ based on k individual samples as in

$$R(X_{DATA}; x_{DATA}, \sigma) = \frac{\sum_{i=1}^{k} R_g(w_i) R_g(\mathbf{X}^{n_i}; \mathbf{x}^{n_i}, \sigma_i)}{\sum_{i=1}^{k} R_g(w_i)}, \qquad (4.51)$$

where

$$R_g(w_i) = \text{the generalized pivot of the precision of estimator } \widehat{\alpha}_i,$$
$$(4.52)$$

$$R_g(w_i) = (n_i/R_{\sigma_i})^2/(n_i - 1), \qquad (4.53)$$

$X_{DATA} = [X]_{i=1,2,\ldots,k;\ j=1,\ldots,m_i}$ and $x_{DATA} = [x]_{i=1,2,\ldots,k;\ j=1,\ldots,m_i}$; x_{DATA} is the observed value of X_{DATA}.

Now, consider the potential generalized test variable for testing

$$H_0: \sigma \leq \sigma_0 \text{ vs. } H_a: \sigma > \sigma_0, \qquad (4.54)$$

where σ_0 is a known quantity, defined by

$$T_g(X_{DATA}; x_{DATA}, \sigma) = \frac{\sum_{i=1}^{k} (n_i/R_g(\mathbf{X}^{n_i}; \mathbf{x}^{n_i}, \sigma_i))/(n_i - 1)}{\sum_{i=1}^{k} (n_i/R_g(\mathbf{X}^{n_i}; \mathbf{x}^{n_i}, \sigma_i))^2/(n_i - 1)} - \sigma.$$
$$(4.55)$$

One can show that $T(X_{DATA}; x_{DATA}, \sigma)$ is a test variable to make inferences about σ, satisfying the conditions set forth in Weerahandi (2004, 1995).

Thus, the generalized p-value, for testing $H_0: \sigma \leq \sigma_0$ vs. $H_a: \sigma > \sigma_0$ is given by

$$p_\sigma^g = Pr(T(X_{DATA}; x_{DATA}, \sigma) < T_g(x_{DATA}; x_{DATA}, \sigma) \mid \sigma = \sigma_0), \qquad (4.56)$$

$$p_\sigma^g = Pr(\frac{\sum_{i=1}^k (n_i/R_g(\mathbf{X}^{n_i}; \mathbf{x}^{n_i}, \sigma_i))/(n_i - 1)}{\sum_{i=1}^k (n_i/R_g(\mathbf{X}^{n_i}; \mathbf{x}^{n_i}, \sigma_i))^2/(n_i - 1)} < \sigma_0). \qquad (4.57)$$

p_σ^g is evaluated by the Monte Carlo method by generating a large number of random numbers from the chi-squared distributions with degree-of-freedom 2 and $(2m_i - 2)$ for $i = 1, 2, ..., k$, and

1. evaluating the fraction of random numbers pairs for which

    $R(X_{DATA}; x_{DATA}, \sigma) < \sigma_0$ for testing $\sigma \le \sigma_0$ vs. $\sigma > \sigma_0$,

2. evaluating the fraction of random numbers pairs for which $R(X_{DATA}; x_{DATA}, \sigma)$

    $> \sigma_0$ for testing $\sigma \ge \sigma_0$ vs. $\sigma < \sigma_0$, and

3. evaluating the fraction of random numbers pairs for which min

    $[R(X_{DATA}; x_{DATA}, \sigma) < \sigma_0, R(X_{DATA}; x_{DATA}, \sigma) > \sigma_0]$ for testing $\sigma = \sigma_0$ vs.

    $\sigma \ne \sigma_0$.

4.5.2 The confidence interval for σ

Since

1. the value of $R(X_{DATA}; x_{DATA}, \sigma)$ is σ, and

2. the distribution of $R(X_{DATA}; x_{DATA}, \sigma)$ is independent of any unknown

    parameters,

$R(X_{DATA}; x_{DATA}, \sigma)$ is a generalized pivotal quantity for constructing a $100(1-\gamma)\%$ confidence interval for σ, where $1 - \gamma$ is the confidence coefficient (Weerahandi 1993).

Now, one-sided as well as two-sided confidence intervals are constructed as follows:

1. the lower bound, $R_\gamma(\sigma; s, \mathbf{t})$, of a $100(1 - \gamma)\%$ one-sided confidence

    interval for σ is sought such that

$$1 - \gamma = \Pr(R(X_{DATA}; x_{DATA}, \sigma) \le R_{1-\gamma}(\sigma; s, \mathbf{t})), \qquad (4.58)$$

2. the upper bound, $R_{1-\gamma}(\sigma; s, \mathbf{t})$, of a $100(1-\gamma)\%$ one-sided confidence interval for σ is sought such that

$$\gamma = \Pr\left(R(X_{DATA}; x_{DATA}, \sigma) \leq R_\gamma(\sigma; s, \mathbf{t})\right), \qquad (4.59)$$

3. the lower limit, $R^l_{1-\gamma/2}(\sigma; s, \mathbf{t})$, and the upper limit, $R^u_{\gamma/2}(\sigma; s, \mathbf{t})$, of a $100(1-\gamma)\%$ two-sided confidence interval for σ are sought such that

$$1 - \gamma = \Pr\left(R^l_{1-\gamma/2}(\sigma; s, \mathbf{t}) \leq R(X_{DATA}; x_{DATA}, \sigma) \leq R^u_{\gamma/2}(\sigma; s, \mathbf{t})\right),$$

$$(4.60)$$

where $\mathbf{t} = (t_1, t_2, ..., t_k)$ is the observed value of $\mathbf{T} = (T_1, T_2, ..., T_k)$, or simply an estimate of $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)$, and $R_\gamma(\sigma; s, \mathbf{t})$ is the 100γ-th percentile of $R(X_{DATA}; x_{DATA}, \sigma)$'s. Once the calculated $R(X_{DATA}; x_{DATA}, \sigma)$'s are ordered, the desired percentiles can be obtained.

4.6 Illustrative examples for the common scale parameter

In order to illustrate the proposed procedure, a different number of samples with sizes $m_{k(=2,6,12)} = 100\,000$ is generated by keeping, for each case, the common shape parameter and unequal scale parameters. Three cases are considered here for the convenience and mathematical tractability:

1. 2-sample case with same sizes m_2 for both samples –

   the parameter specifications for 4-repetitions are listed below:

   $\boldsymbol{\theta}_1^{(2)} = (100, 102)$, $\boldsymbol{\theta}_2^{(2)} = (104, 106)$,

   $\boldsymbol{\theta}_3^{(2)} = (108, 110)$, $\boldsymbol{\theta}_4^{(2)} = (112, 114)$, $\sigma^{(2)} = 1$

2. 6-sample case with the same sizes m_6 for all six samples –

the parameter specifications for 4-repetitions are listed below:

$$\boldsymbol{\theta}_1^{(6)} = (\boldsymbol{\theta}_1^{(2)}, \boldsymbol{\theta}_2^{(2)}, \boldsymbol{\theta}_3^{(2)}), \boldsymbol{\theta}_2^{(6)} = (\boldsymbol{\theta}_1^{(2)}, \boldsymbol{\theta}_2^{(2)}, \boldsymbol{\theta}_4^{(2)}),$$

$$\boldsymbol{\theta}_3^{(6)} = (\boldsymbol{\theta}_2^{(2)}, \boldsymbol{\theta}_3^{(2)}, \boldsymbol{\theta}_4^{(2)}), \boldsymbol{\theta}_4^{(6)} = (\boldsymbol{\theta}_3^{(2)}, \boldsymbol{\theta}_4^{(2)}, \boldsymbol{\theta}_1^{(2)}), \sigma^{(6)} = 2$$

3. 12-sample case with the same sizes m_{12} for all twelve samples –

the parameter specifications for 4-repetitions are listed below

$$\boldsymbol{\theta}_1^{(12)} = (\boldsymbol{\theta}_1^{(6)}, \boldsymbol{\theta}_2^{(6)}), \boldsymbol{\theta}_2^{(12)} = (\boldsymbol{\theta}_1^{(6)}, \boldsymbol{\theta}_3^{(6)}),$$

$$\boldsymbol{\theta}_3^{(12)} = (\boldsymbol{\theta}_1^{(6)}, \boldsymbol{\theta}_4^{(6)}), \boldsymbol{\theta}_4^{(12)} = (\boldsymbol{\theta}_2^{(6)}, \boldsymbol{\theta}_3^{(6)}), \sigma^{(12)} = 3$$

Probability coverages for all above cases are given in Tables 4.9 through 4.11 for

parameter specifications for $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_i, ..., \theta_k)_{k(=2,6,12)}$ for each case $i = 1, 2, ..., k$

and $\alpha^{(k)}(k = 2, 6, 12)$.

Table 4.9 Probability coverages for 95% two-sided confidence intervals for σ :

2-sample case

Parameters:	Sample size					
$k = 2, \boldsymbol{\theta}^{(2)}, \sigma^{(2)}$	10		20		50	
	GVM	CM	GVM	CM	GVM	CM
$\boldsymbol{\theta}_1^{(2)} = (\theta_{11}^{(2)}, \theta_{12}^{(2)}), \sigma^{(2)}$	0.97	0.88	0.95	0.90	0.94	0.94
$\boldsymbol{\theta}_2^{(2)} = (\theta_{21}^{(2)}, \theta_{22}^{(2)}), \sigma^{(2)}$	0.96	0.89	0.95	0.91	0.95	0.95
$\boldsymbol{\theta}_3^{(2)} = (\theta_{31}^{(2)}, \theta_{32}^{(2)}), \sigma^{(2)}$	0.96	0.90	0.96	0.92	0.94	0.93
$\boldsymbol{\theta}_4^{(2)} = (\theta_{41}^{(2)}, \theta_{42}^{(2)}), \sigma^{(2)}$	0.96	0.85	0.95	0.92	0.95	0.95

GVM = the Generalized Variable Method; CM = the Classical Method

Table 4.10 Probability coverages for 95% two-sided confidence intervals for σ :

6-sample case

Parameters:	Sample size					
$k = 6,\ \boldsymbol{\theta}^{(6)},\ \sigma^{(6)}$	10		20		50	
	GVM	CM	GVM	CM	GVM	CM
$\boldsymbol{\theta}_1^{(6)} = (\theta_{11}^{(6)}, \theta_{12}^{(6)}, ..., \theta_{16}^{(6)}),\ \sigma^{(6)}$	0.97	0.86	0.96	0.91	0.95	0.94
$\boldsymbol{\theta}_2^{(6)} = (\theta_{21}^{(6)}, \theta_{22}^{(6)}, ..., \theta_{26}^{(6)}),\ \sigma^{(6)}$	0.97	0.80	0.97	0.88	0.95	0.94
$\boldsymbol{\theta}_3^{(6)} = (\theta_{31}^{(6)}, \theta_{32}^{(6)}, ..., \theta_{36}^{(6)}),\ \sigma^{(6)}$	0.98	0.88	0.97	0.79	0.95	0.93
$\boldsymbol{\theta}_4^{(6)} = (\theta_{41}^{(6)}, \theta_{42}^{(6)}, ..., \theta_{46}^{(6)}),\ \sigma^{(6)}$	0.97	0.89	0.96	0.89	0.96	0.92

GVM = the Generalized Variable Method; CM = the Classical Method

Table 4.11 Probability coverages for 95% two-sided confidence intervals for σ :

12-sample case

Parameters:	Sample size					
$k = 12,\ \boldsymbol{\theta}^{(12)},\ \sigma^{(12)}$	10		20		50	
	GM	CM	GM	CM	GM	CM
$\boldsymbol{\theta}_1^{(12)} = (\theta_{11}^{(12)}, \theta_{12}^{(12)}, ..., \theta_{112}^{(12)}),\ \sigma^{(12)}$	0.98	0.80	0.96	0.94	0.96	0.95
$\boldsymbol{\theta}_2^{(12)} = (\theta_{21}^{(12)}, \theta_{22}^{(12)}, ..., \theta_{212}^{(12)}),\ \sigma^{(12)}$	0.98	0.75	0.95	0.94	0.95	0.93
$\boldsymbol{\theta}_3^{(12)} = (\theta_{31}^{(12)}, \theta_{32}^{(12)}, ..., \theta_{312}^{(12)}),\ \sigma^{(12)}$	0.97	0.79	0.95	0.94	0.96	0.94
$\boldsymbol{\theta}_4^{(12)} = (\theta_{41}^{(12)}, \theta_{42}^{(12)}, ..., \theta_{412}^{(12)}),\ \sigma^{(12)}$	0.97	0.81	0.95	0.93	0.96	0.94

GM = the Generalized Variable Method; CM = the Classical Method

Here, the classical confidence interval estimation of α is given by

$$\left(s - Z_{1-\gamma/2}\sqrt{Var_{obs}(S)}, s + Z_{1-\gamma/2}\sqrt{Var_{obs}(S)} \right), \qquad (4.61)$$

where $Z_{\gamma/2}$ is $\gamma/2$th quantile of $Z \sim N(0,1)$ and $(s, \widehat{Var}_{obs}(S))$ is the observed value

of $(S, \widehat{Var}(S))$, or simply an estimate of $(\sigma, Var(S))$. Therefore,

$$\left(s - Z_{1-\gamma/2}\sqrt{\frac{1}{\sum_{i=1}^{k} \frac{1}{\widehat{Var}_{obs}(\widehat{\sigma}_i)}}}, s + Z_{1-\gamma/2}\frac{1}{\sum_{i=1}^{k} \frac{1}{\widehat{Var}_{obs}(\widehat{\sigma}_i)}} \right), \qquad (4.62)$$

where $(s, \widehat{Var}_{obs}(\widehat{\sigma}_i))$ is the observed value of $(S, \widehat{Var}(\widehat{\sigma}_i))$, or simplyan estimate of

$(\sigma, Var(\widehat{\sigma}_i))$.

As $\widehat{\sigma}_i \sim S_i \sim \Gamma(n_i - 1, \sigma_i/n_i)$,

$$Var(\widehat{\sigma}_i) = \frac{(\sigma_i/n_i)^2}{(n_i - 2)^2(n_i - 3)}, n_i > 3. \qquad (4.63)$$

Since,

$$\sigma = \frac{\sum_{i=1}^{k} w_i \sigma_i}{\sum_{i=1}^{k} w_i}, \qquad (4.64)$$

where

$$w_i = \frac{1}{Var(\widehat{\sigma}_i)}, \qquad (4.65)$$

Then, the classical estimate of the common shape parameter α of several exponential

populations is defined as

$$\widehat{\sigma} = S = \frac{\sum_{i=1}^{k} \widehat{w}_i \widehat{\sigma}_i}{\sum_{i=1}^{k} \widehat{w}_i}, \qquad (4.66)$$

where

$$\widehat{w}_i = \frac{1}{\widehat{Var}(\widehat{\sigma}_i)}. \qquad (4.67)$$

This yields

$$S = \frac{\sum_{i=1}^{k} \frac{\widehat{\alpha}_i}{\widehat{Var}(\widehat{\sigma}_i)}}{\sum_{i=1}^{k} \frac{1}{\widehat{Var}(\widehat{\sigma}_i)}}, \tag{4.68}$$

where $\widehat{\sigma}_i = S_i = n_i \left(\sum_{j=1}^{n_i} \ln(X_{ij}/\widehat{\theta}_i) \right)^{-1}$ and $S^2(\widehat{\sigma}_i)$ is an estimator of $Var(\widehat{\sigma}_i)$.

Then,

$$\widehat{Var}(S) = \widehat{Var}\left(\frac{\sum_{i=1}^{k} \frac{\widehat{\alpha}_i}{\widehat{Var}(\widehat{\sigma}_i)}}{\sum_{i=1}^{k} \frac{1}{\widehat{Var}(\widehat{\sigma}_i)}} \right) = \frac{1}{\sum_{i=1}^{k} \frac{1}{\widehat{Var}(\widehat{\sigma}_i)}} \tag{4.69}$$

As in the several Pareto populations case, a careful comparison of the empirical confidence levels of the generalized variable method with that of the classical method for different sample sizes in different sample cases, suggests that the generalized variable method outperforms the classical approach for this particular case.

.

CHAPTER V

THE GENERALIZED VARIABLE METHOD IN COMPUTER NETWORKING

AND DATA TRANSMISSION

5.1 Introduction

Today, computer networks have become the core and backbone of modern communication. Advancing and developing computer network systems have contributed in a large scale to improve the scope and the ability of communication significantly in the past decade. In order to increase the backbone capacity and the reliability of telecommunications, advanced optical technologies have been developed during the past decade such as

1. the Dense Wavelength Division Multiplexing (DWDM),

2. the Optical Amplification (OA),

3. the Optical Path Routing (OPR), and

4. the Wavelength Add-Drop Multiplexer Wide-Area Networks

(WADMWAN).

These are heavily used and are being experimented at the cost of knowledge, labor, sacrifice, determination, and dedication of scientists, practitioners, theorists, and experimentalists.

Computer networks are categorized based on their scale, scope, and purpose as follows:

1. the Personal Area Networks (PAN),

2. the Local Area Networks (LAN),

3. the Campus Area Networks (CAN),

4. the Metropolitan Area Networks (MAN),

5. the Wide Area Networks (WAN), and

6. the Global Area Networks (GAN).

In addition to the structural design of these computer networks, practitioners are interested in providing accurate, efficient, fast, dynamic services by modeling services like the World Wide Web, and in designing and implementing computer networks. This has been possible through statistical analyses, which provide accurate, less-error solutions for the problems encountered in Computer Networking and Data Transmission. This chapter is allocated for the critical statistical analysis of the Offered Optical Network Unit Load (OOL) – the most prominent physical quantity in Data Transmission to be tested – that is generated at the Optical Network Units (ONUs) or simply at the "switches." These critical analyses make sure that such a load would not exceed a certain threshold value, thus preventing any kind of crash in network systems as well as in computers. Newly introduced, the generalized variable method is compared with the other existing inference procedures in order to find exact solutions for such problems and to provide the best customer services.
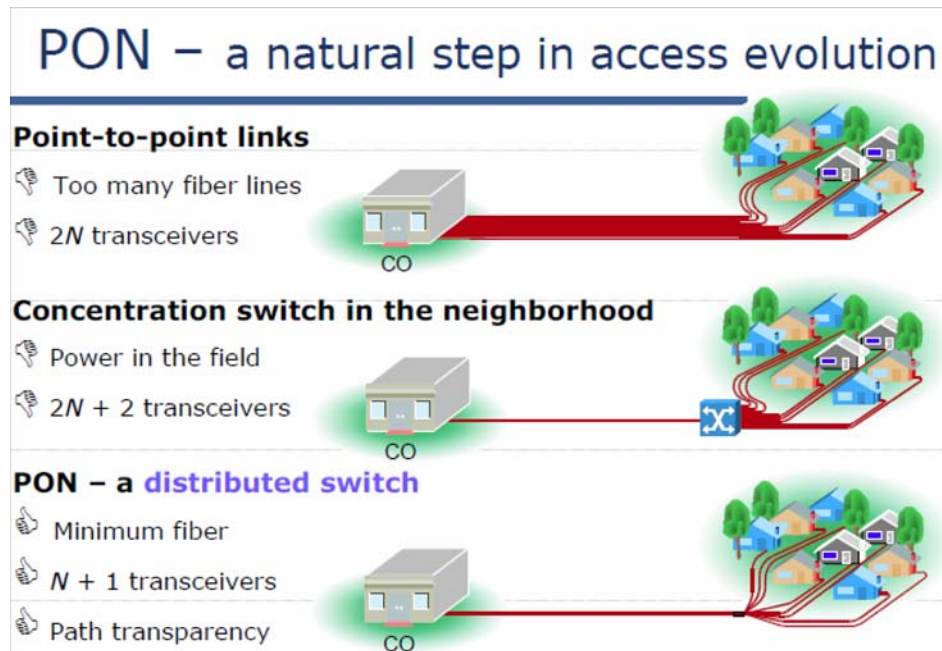
Figure 5.1 Passive Optical Network. Source: Kramer (2003).

5.2 Ethernet Passive Optical Networks (EPONs)

The well-known communication technologies, such as the Digital Subscriber Line (DSL) and the Cable Modem (CM) are used for the delivery of the services and applications such as:

1. Voice communications (chatting over internet),

2. Standard and High–Definition Video (STV and HDV),

3. Video Conferencing (interactive video),

4. WWW-data traffic, WAP-data traffic, and

5. Real-Time and Near- Real-Time Transactions.

The poor performance of the DSL as well as the Cable Modem, paved the way for Ethernet Passive Optical Network (EPON) to emerge as an alternative solution to those technologies of communications. As noted in Figure 5.1, EPONs are point to

94

point optical networks – which utilize optical fibers to transmit packets (sometimes referred to as "files, information, data, or bytes") – in the form of light, which has the highest velocity all objects or particles can attain. This technology provides a low-cost method of deploying optical access lines between an Optical Line Terminal (OLT) that resides in the service provider's Central Office (CO) (sometimes called either "Local Exchange (LE), Point of Presence (POP), Headend, or Hub") and Optical Network Units (ONUs) (sometimes referred to as "Optical Network Terminal" (ONT)). Furthermore, this technology connects optical access networks to MAN or WAN to communicate with IP-based networks such as:

1. the Session Initiation Protocol (SIP) servers,

2. the Voice-on-Demand (VoD) servers,

3. the Head End servers,

4. the Communication Satellites (CS) broadcasting stations, and

5. the Content providers.

The ONU is located, either at the end-user location, with Fiber-to-the-Home (FTTH) or with Fiber-to-the-Business (FTTB), or at the curb, with Fiber-to-the-Curb (FTTC) architecture.

5.3 Traffic data models

Traffic ("data, packets, files, or bandwidth") refers to the amount of data that is transferred from the Central Office of the internet provider to computers when individuals visit websites by browsing them using browsers such as Internet Explorer, Mozilla, Flock, Netscape Navigator, or AOL, etc. Technically, webpages are requested

and retrieved from the server. Webpage that are written in Hyper Text Markup Language (HTML) can be browsed through a Personal Computer (PC), a Note Book (Laptop), or a hand-held devices such as Smart phones, Blackberries, *i*-phones, or palmtops. The web pages retrieved through PC are called World Wide Web-traffic Data (or the WWW-traffic data) and those that are retrieved by the other devices mentioned above are called Wireless Application Protocol-traffic Data (or WAP-traffic data).

5.3.1 The WWW-traffic data model
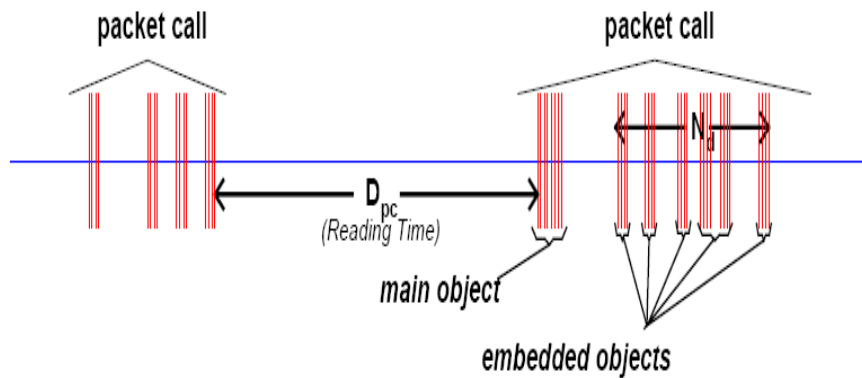
Figure 5.2 Content in a packet call. Source: Hauzner (2008).

The article Hauzner (2008), prepared to assist the IEEE P802.19 project, explains that the World Wide Web-Traffic Data Model (or simply the Web-Traffic Data Model) is a collection of eleven measurable parameters that can be modeled with any positive-skewed distribution.

The eleven measurable parameters are listed as follows:

1. The Web-session interarrival time,

2. The number of packets calls (pages) per Web-session,

3. The reading time between packet calls (Web-pages),

4. The number of items per Web-page,

5. The time intervals between items in the same Web-page,

6. The Web-item size on upstream,

7. The Web-item size on downstream,

8. The time interval between two consecutive upstream packets inside an item,

9. The time interval from upstream to downstream packet inside an item,

10. The time interval between two consecutive downstream packets inside an item,

11. The time interval from Downlink to upstream packet inside an item.

5.3.2 The WAP-traffic data model

Wireless Application Protocol-traffic Data Model (or simply Wireless-Traffic Data Model) is also a collection of twelve measurable distributional parameters.

1. The Wireless-session interarrival time

2. The number of packets calls (pages) per Web-session

3. The reading time between packet calls (Wireless-items)

4. The Wireless-item size on upstream,

5. The Wireless-item size on downstream,

6. The transmission time of the upstream packets,

7. The processing time of Wireless-request,

8. The Web-transaction waiting time,

9. The processing time of Wireless-response,

10. The transmission time of the downstream packets (Wireless-response),

11. The acknowledgment time on upstream,

12. The acknowledgment time on downstream.

Throughout this Chapter, the number of packets per Web-session is intensively discussed citing all the research work so far done to model this networking parameter.
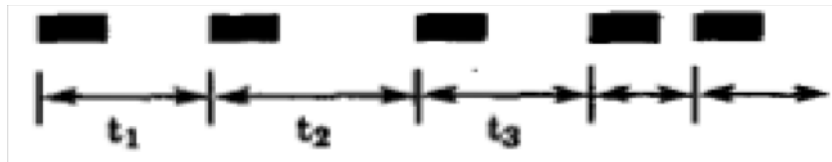
Figure 5.3 The Poisson model. Source: Jain and Routhier (1986).

5.4 Models of packet arrival

For a long period of time, for mathematical and analytical simplicity, network traffic was often modeled by the *Poisson arrival model* (Marathe and Hawe 1982, Ramakrishnan and Tripathi 1982, and Tobagi and Hunt 1979) and the *compound Poisson arrival model* (Heyman 1982, Meister 1980, and Mohanty 1978) even though packet interarrivals are not exponentially distributed. The failure of *Poisson modeling* is discussed at length by Paxson and Floyd (1995). The *Packet-train model,* in which packets were modeled by inter-cars of a train, was introduced by Jain and Routhier (1986) as an alternative to the *Poisson arrival model.* They pointed out that the *Poisson* as well as the *compound Poisson* treat packets as black boxes, but not distinguishing between the packets coming from sources or those going to different

destinations, caused the way for a loss of some information, which is easily available at the network layer. As this model also was criticized for the lack of a clear definition of "Train", of suggestions for choosing the crucial parameters, and of the physical interpretation, the *self-similar traffic model* was brought to the attention of probabilists by Taqqu and Levy (1986). However, this *self-similar traffic model* had already been introduced into the networking arena by Mendelbrot in 1969. As a straightforward extension of this *slef-similar traffic model*, Willinger, Taqqu, Sherman, and Wilson (1997) proposed the *ON/OFF source model,* which has strictly alternating *ON*-and *OFF*-periods.
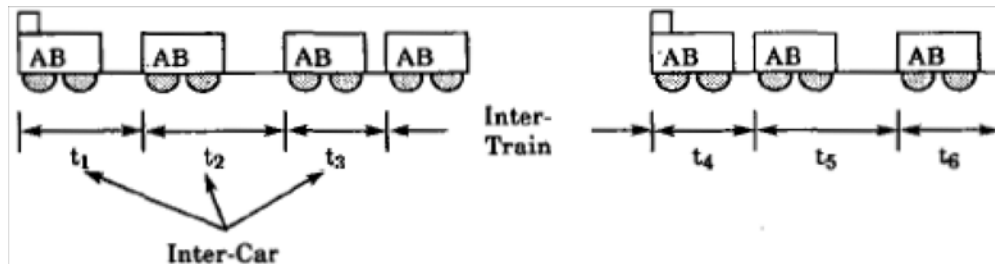
Figure 5.4 The packet train model. Source: Jain and Routhier (1986).

5.5 The self-similar traffic model

Taqqu, Willinger, and Sherman (1997) and Willinger *et al.* (1997) discussed the mathematical proof of fundamental results in the *self-similar traffic model*. In their proof, they suggested a stationary binary time series $\{P(t), t \geq 0\}$, $P(t) = 1$ for *ON* periods at time t and $P(t) = 0$ for *OFF* periods at time t for one source. When this is expanded for several n sources, and the aggregation of packet count is taken into

account, the aggregate cumulative packet counts in the interval $[0, \tau t]$ is given by

$$P(\tau t) = \int_0^{\tau t} \sum_{i=1}^{n} P_i(q) dq, \tag{5.1}$$

where $\tau = $ Rescaling time factor. Furthermore, Taqqu *et al.* (1997) and Willinger *et al.* (1997) showed – for homogeneous sources, large n and τ – the aggregate cumulative packet process $\{P(\tau t), t \geq 0\}$ behaves stochastically as

$$\tau n \frac{\mu_1}{\mu_1 + \mu_2} t + \tau^H \sqrt{N(t)n} c_{\lim} B_H(t), \tag{5.2}$$

where $B_H(t) = $ Fractional Brownian Process; $c_{\lim} = $ a finite constant; and $N(t) = $ normalization factor.

While the Joseph's effect exhibits the self-simlarity or long-range dependence, the Noah's effect exhibits the high variability. The Joseph's effect is created with the aid of distributions of *ON/OFF*-periods with finite mean values μ_1 and μ_2, and infinite variances σ_1 and σ_2, respectively. The quantity $n\mu_1/(\mu_1 + \mu_2)$ found in equation (5.2) is called the Offered Optical Network Unit Load (OOL).

For the heterogeneous sources, the aggregate cumulative packet process is given by

$$\tau \left(\sum_{s=1}^{S} n^{(s)} \frac{\mu_1^{(s)}}{\mu_1^{(s)} + \mu_2^{(s)}} \right) t + \sum_{s=1}^{S} \tau^{H^{(s)}} \sqrt{N^{(s)}(t)n^{(s)}} c_{\lim}^{(s)} B_{H^{(s)}}(t), \tag{5.3}$$

where all parameters are defined the same way as in the homogeneous case, but $()$ indicates the different sources in the same category. Also, the Offered Ooptical Network Unit Load of the heterogenous system is given by

$$\sum_{s=1}^{S} \sum_{i=1}^{n^{(s)}} \frac{\mu_{1i}^{(s)}}{\mu_{1i}^{(s)} + \mu_{2i}^{(s)}} = \sum_{s=1}^{S} \sum_{i=1}^{n^{(s)}} \frac{E(X_{1i}^{(s)})}{E(X_{1i}^{(s)}) + E(X_{2i}^{(s)})}. \tag{5.4}$$

100

5.6 The self-similarity and long-range dependence

Taqqu *et al.* (1997) explained that the Noah's effect (the high variability) produces aggregate network traffic exhibiting the Joseph's effect (the self-similarity or the long-range dependence). The measure of the intensity, given by the scale parameter α of a typical source, of the Noah's effect of the *ON*-and *OFF*-periods is related to the measure of the degree, given by the Hurst parameter H, of the Joseph's effect of the aggregate traffic stream as follows:

$$H = (3 - \alpha_{\min})/2, \tag{5.5}$$

where $\alpha_{\min} = \min(\alpha_1, ..., \alpha_n)$ for n *ON/OFF* sources.

The discovery of traffic similarity is credited to several Computer Networking researchers and practitioners :

1. the self similarity in Local Area Network (LAN)

    Leland *et al.* (1994),

2. the self similarity in pre-web Wide Area Network (pre-web WAN)

    Paxson and Floyd (1994),

3. the self similarity in modern Wide Area Network (modern WAN)

    Crovella and Bestavros (1996).

The mathematical aspect of the self-similarity was discussed by Kramer (2001) by defining the following specifications:

$X(t)$ = cumulative process (packets/bytes arrivals up to time $t = 1, 2, ..., T$)

X_t = increment process of $X(t)$,

$X_N^{(T)}$ = aggregated process of X_t, and

$\gamma(\tau) =$ auto covariance function

Kramer (2001) explained that the process is exactly self-similar if

$$\gamma^{(T)}(\tau) = \gamma(\tau), \tag{5.6}$$

and asymptotically self-similar if

$$lim_{T\to\infty}\ \gamma^{(T)}(\tau) = \gamma(\tau), \tag{5.7}$$

where

$$X_t = X(t+1) - X(t);\ t = 1, 2, ..., T$$

$$X_N^{(T)} = \tfrac{1}{T}\left[X_{NT-T+1} + X_{NT-T+2} + ... + X_{NT}\right];\ N = 1, 2, ..., \infty$$

$$\gamma(\tau) = E\left[(X_t - \mu)(X_{t+\tau} - \mu)\right];\ t, \tau = 1, 2, ..., T$$

$$\mu =$$ mean of the packet distribution

Kramer (2001), furthermore, statistically explained the self-similarity in different aspect, with parameter $H(0 < H < 1)$ for all $\tau > 0$ and $t \geq 0$, as follows:

Kramer (2001) considered X_t to be a stationary increment process, and $X^{(T)}$ to be a sample mean. Then,

$$X^{(T)} = \frac{1}{T}\sum_{t=1}^{T} X(t) = \frac{1}{T}\left[X(T) - X(0)\right],$$

$$X^{(T)} = \frac{T^H}{T}\sum_{t=1}^{T}[X(1) - X(0)] = T^{H-1}X. \tag{5.8}$$

the process is exactly self-similar if

$$X^{(T)} = T^{H-1}X, \tag{5.9}$$

and asymptotically self-similar if

$$lim_{T\to\infty}\ X^{(T)} = T^{H-1}X. \tag{5.10}$$

102

Moreover, Kramer (2001) explained that the system is long-range dependent for

$$0 < H < 1, H \neq \frac{1}{2} \quad r(\tau) \sim H(2H-1)\tau^{2H-2}, \text{as } \tau \to \infty, \tag{5.11}$$

and for

$$\frac{1}{2} < H < 1, \text{ then} \sum_{\tau = -\infty}^{\infty} r(\tau) = \infty, \tag{5.12}$$

where $r(\tau)$ = auto correlation function given by

$$r(\tau) = \frac{\gamma(\tau)}{\sigma^2} = \frac{E\left[(X_t - \mu)(X_{t+\tau} - \mu)\right]}{E\left[(X_t - \mu)^2\right]}, \tag{5.13}$$

σ^2 being the variance of the distribution. Note that the long-range dependence is different from the self-similarity, but for $1/2 < H < 1$, the process is both long-range dependent and self-similar.

5.7 The multiplexing (muxing)

Multiplexing (Muxing) – a process of analog or digital data combined into a single unit – and Demultiplexing (Demuxing) – the reverse process of Multiplexing – are common practices in Computer Networking and Data Transmission to make efficient, dynamic, high-speed, low-cost data (information), which is transmitted through a single channel/stream/ path/wire/medium. In electrical communications, there are two basic forms of multiplexing:

1. the Time-Division Multiplexing (TDM), and

2. the Frequency-Division Multiplexing (FDM)/Wavelength-Division

    Multiplexing (WDM).

When data are transmitted, each substream generates packets/packet trains/ bursts. The number of packets per burst (ON-periods) follows any heavy-tailed dis-

103

tribution. *OFF*-periods (intervals between the packet trains) also follow another heavy-tailed distribution or the same distribution that has been used for *ON*-periods. When packets (sometimes referred to as "files") are muxed, either at an ONU (optical Network Unit) or a switch, or at a combiner, the Offered Optical Unit Load (OOL), the Effective Optical Unit Load (EOL), the Offered Network Load (ONL), and the Effective Network Load (ENL) are generated. The generation of these different loads at the switch pave the way for statisticians to make sure that those loads do not exceed a certain threshold value, preventing any damage to any switch or computer, or eventually to the transmission line.

5.8 The Offered Optical Network Unit Load/Offered ONU Load (OOL)

When webpages are browsed through computers, downstream traffic – the data transmitted from the central office to the computer – passes through a passive optical splitter, which demuxes data making a point-to-multipoint network, and reaches each ONU. In upstream traffic – the data transmitted from the computer to the central office – the ONU aggregates (muxes) data from different sources making a multipoint-to-point network.
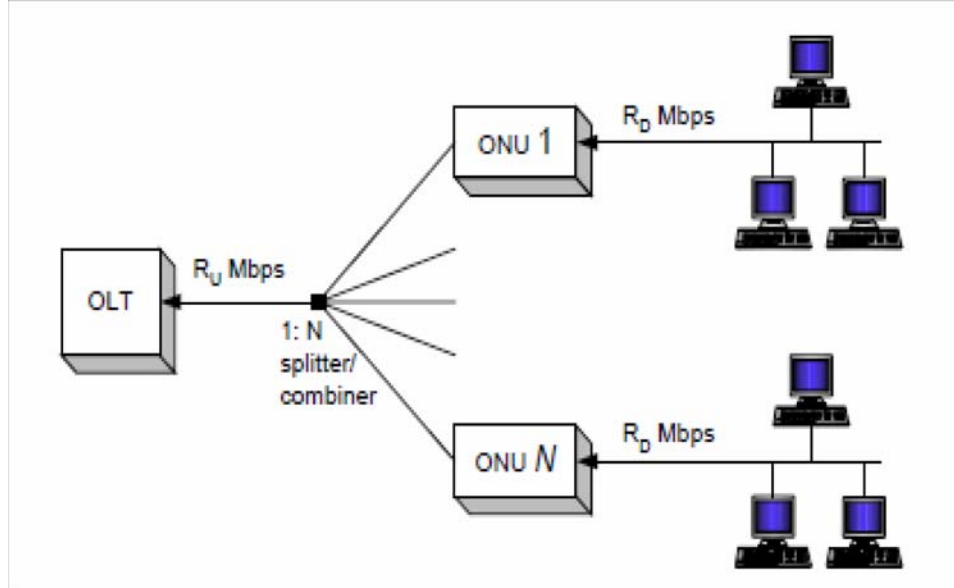
Figure 5.5 Access network based on PON. Source: Kramer *et al.* (2001).

In our implementation, each substream generates packets of constant size, although this size is different for different streams. Each substream generates packets in groups ("packet trains, bursts, train length, or *ON*-periods"). The number of packets per burst follows the Pareto distribution with a minimum of 1 (i.e., the smallest burst, a common location parameter, consists of only 1 packet) and with different shape parameters. *OFF*-periods (intervals between the packet trains or intertrain distance) also follow the Pareto distribution. We use heavier tail for the distribution of the *OFF*-periods because the *OFF*-periods represent a stable state in a network; i.e., a network can be in the *OFF* state (no packet transmission) for an unlimited time. On the other hand, the duration of the *ON*-periods are ultimately limited by network resources and are necessarily finite file sizes. The location parameter for the *OFF*-periods was chosen so as to obtain a desired load ϕ_i from the given substream

i (Kramer 2005):

$$\phi_i \;=\; \frac{E(X_{ON,i})}{E(X_{ON,i}) + E(X_{OFF,i})}, \tag{5.14}$$

where $E(X_{ON,i})$ and $E(X_{OFF,i})$ are expected lengths (durations) of an *ON*-and *OFF*-periods of source i, respectively, and are given in terms of parameters of Pareto distributions.
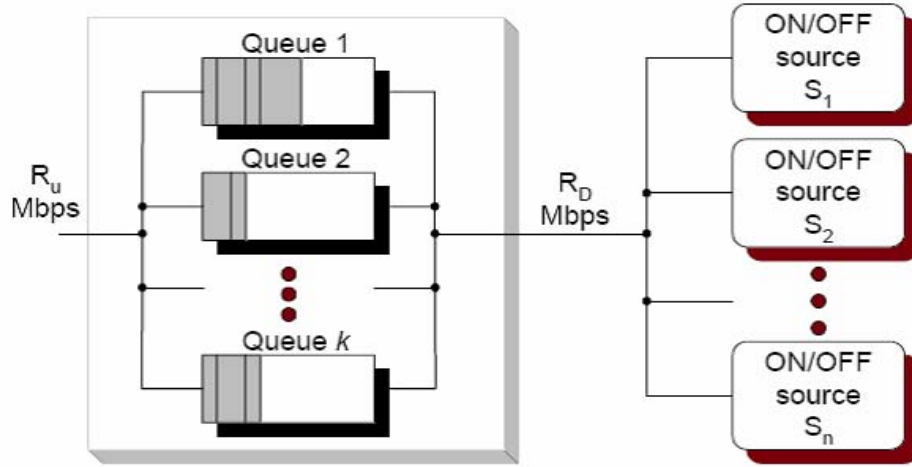
Figure 5.6 Traffic generation in the ONU. Source: Kramer *et al.* (2001).

When multiple substreams are aggregated (serialized or multiplexed), the net load from all n sources in an ONU is called the Offered Optical Network Unit Load (OOL), denoted by Ψ, and is given by (Kramer 2005):

$$\Psi \;=\; \sum_{i=1}^{n} \phi_i. \tag{5.15}$$

Consequently, the Offered Network Load (ONL) resulting from all the ONU ("N" ONUs) is given by $\Phi \;=\; R_D/R_U \sum_{j=1}^{N} \Psi_j,$ where R_D and R_U are data rate and the

upstream rate, respectively. Furthermore, the Effective Optical Network Unit Load (EOL) resulting from the data that are sent out by the ONU is given by ω. And, the Effective Network Load (ENL) resulting from the aggregation of all EOL, generated by each ONU and based on a scaling coefficient is given by $\Omega = R_D/R_U \sum_{j=1}^{N} \omega_j$.

We consider a networking system having $n \ (\geq 2)$ parallel independent substreams that are connected to an ONU with Pareto distributed ON-and OFF-packet sizes. They are Pareto distributed with the respective common scale parameters θ_1 and θ_2 (or θ_k where $k = 1(ON), 2(OFF)$) and with different shape parameters α_{1i} and α_{2i} (or α_{ki} ; $k = 1(ON), 2(OFF)$) where $i = 1, 2, ..., n$. Then, the inferences of the OOL of the system are performed; first, by the conventional classical method based on the large sample approach, and later, by the exact generalized method based on the generalized variable approach introduced by Tsui and Weerahandi (1989). We perform the approximate and exact tests, and construct the approximate and exact confidence intervals for the OOL. Real and simulated data sets are utilized to compare the performance of these generalized procedures with the other approximate procedures.

5.9 Heavy-tailed and power-law distributions in Network Traffic

Several Pareto populations as well as several exponential and lognormal populations have now been successfully applied in modeling the file sizes (sometimes referred to as 'packets, job sizes, or flow sizes') found in Computer Networking and Data Transmission. Transfer sizes, packet sizes, data sizes, number of bytes, number of packets per train, number of packets per burst, and ON-and OFF-packet sizes are

other similar terms used by the practitioners for file sizes. An important character-
istic of the traffic self-similarity is that the sizes of files being transferred are drawn
from a heavy-tailed distribution. A distribution is heavy-tailed if

$$[X > x] \sim \theta x^{-\alpha} \text{ as } x \to \infty, \text{ where } 0 < \alpha < 2, \tag{5.16}$$

where X is a random variable, θ is a location parameter, and α is a shape parameter.
When α is less than 2, the distribution has infinite variance, which is also required for
these models to produce self-similarity. The asymptotic shape of the distribution is
hyperbolic and follows a power law, i.e., the heavy-tailed distribution has a hyperbolic
tail or power law decay. The mathematical vehicle for modeling Noah's effect is a
heavy-tailed distribution with infinite variance, such as the Pareto distribution –
which is the simplest heavy-tailed distribution whose mean is finite, when $\alpha > 1$ and
variance is infinite, when $\alpha < 2$. The Pareto distribution is power-law over its entire
range and its probability density function is given by

$$f(x) = \begin{cases} 1 - (\frac{\theta}{x})^\alpha, \ x \geq \theta \\ \\ 0, \ x < 0 \end{cases}. \tag{5.17}$$

In order to have a self-similar traffic to be modeled, Pareto with $1 < \alpha < 2$ must
be used.

The degree of self-similarity as a measure of "burstiness" can be defined via the
Hurst parameter H, while the long -range dependence (heavy-tailed) is measured via
the scale parameter α of the Pareto distribution. The relationship between H and α
is given by $H = (3 - \alpha)/2$ (Wiilinger *et al.* 1995). In the presence of n *i.i.d.* sources,
$H = (3 - \alpha_{\min})/2$, where $\alpha_{\min} = \min(\alpha_1, ..., \alpha_n)$ for n *ON/OFF* sources. It should be

noted that to exhibit long-range dependence to which the heavy-tailed distribution is applied, α must be $1 < \alpha < 2$, which implies that $0.5 < H < 1$.

5.10 Statistical inferences for Ψ

Let us consider a passive optical network system having n parallel independent "last miles," which is the access portion of the network. The " first mile" is sometimes used for the "last mile" because the "last mile" of a network to the user is also the "first mile" from user to the world. Also, it is referred to as the "subscriber access network", or the "local loop," which connects the service provider's central office to businesses and residential subscriber. The last miles have Pareto distributed ON-and OFF-packet sizes.

For the ith $(i = 1, 2, ..., n)$ component in the system, suppose X_{1i} and X_{2i} are the independent ON-and OFF-packet sizes, respectively. They are Pareto distributed with the respective common scale parameters θ_1 and θ_2 and with different shape parameters α_{1i} and α_{2i}.
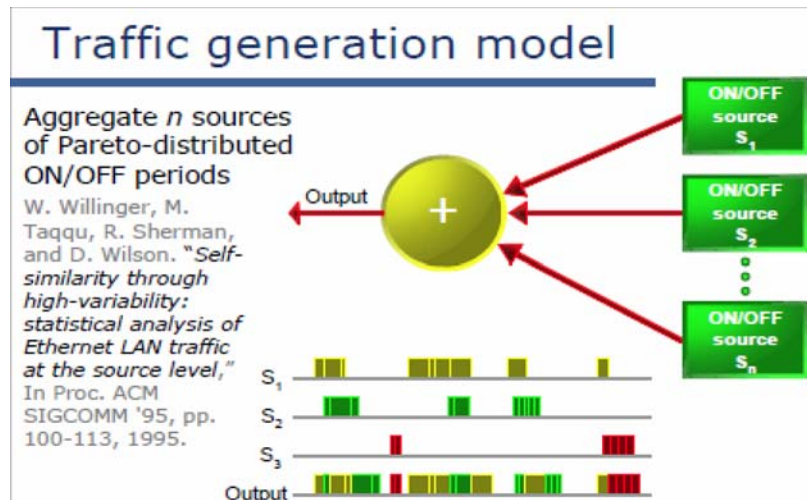
Figure 5.7 Traffic generation model. Source: Kramer (2001).

Then, the Offered Optical Network Unit Load (OOL) of the system is given by (Kramer 2005)

$$\Psi = \sum_{i=1}^{n} E(X_{1i})[E(X_{1i}) + E(X_{2i})]^{-1}, \tag{5.18}$$

where $E(X_{1i})$ and $E(X_{2i})$ are the expected lengths (durations) of an *ON*-and *OFF*-period of ith *ON/OFF* source, respectively. Furthermore, in terms of the parameters of two-parameter Pareto distributions with the common scale parameters θ_1 and θ_2 and the shape parameters α_{1i} and α_{2i} $i = 1, 2, ..., n$, the OOL of the system is given by

$$\Psi = \Psi(\boldsymbol{\theta}) = \sum_{i=1}^{n} (\alpha_{2i} - 1)\alpha_{1i}\theta_1[\alpha_{1i}\alpha_{2i}(\theta_1 + \theta_2) - (\alpha_{1i}\theta_1 + \alpha_{2i}\theta_2)]^{-1},$$

$$\tag{5.19}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \alpha_{11}, \alpha_{12}, ..., \alpha_{1n}, \alpha_{21}, \alpha_{22}, ..., \alpha_{2n})$.

Suppose $\mathbf{X}_i^{m_i} = (X_{i1}, X_{i2}, ..., X_{im_i})$ is a random sample of size m_i from an ith Pareto population, $X \sim Pa\ (\theta, \alpha_i),\ i = 1, 2, ..., k$, where θ denotes the common unknown scale parameter and α_i is an unknown and possibly unequal scale parameter of the ith exponential population. Furthermore, suppose $\mathbf{x}_i^{m_i} = (x_{i1}, x_{i2}, ..., x_{im_i})$ is its observed value. Therefore, $\{X_{ij}\}_{i=1,2,...,k;j=1,2,...,m_i}$ is independently distributed as

$$X_{ij} \sim \frac{\alpha_i \theta^{\alpha_i}}{x^{(\alpha_i+1)}} I_{[x_i \geq \theta]}, \quad \theta, \alpha_i > 0, \forall i, , \tag{5.20}$$

where I denotes the usual indicator function given by

$$I\left[a \geq b\right] = \begin{cases} 1 & \text{if}\ \ a \geq b \\ 0 & \text{if}\ \ a < b \end{cases}.$$

Inspired by Quant (1996) on maximum likelihood estimators of scale and shape parametrs of a single Pareto distribution, Elfessi and Jin (1996) showed, for the i-tuple $(i = 1, 2, ..., n)$ Pareto distributions, maximum likelihood estimators of θ and $\alpha_i (i = 1, 2, ..., n)$ – denoted by $\widehat{\theta}$ and $\widehat{\alpha}_i$, respectively – are given by

$$\widehat{\theta} = T = \min_{1 \leq i \leq k} \quad \min_{1 \leq j \leq m_i} X_{ij} \quad \text{and} \quad \widehat{\alpha}_i = A_i = m_i Y_i^{-1}, \tag{5.21}$$

where $Y_i = \sum_{j=1}^{m_i} \ln(X_{ij}/X_{i(1)})$ for $i = 1, 2, ..., n$.

Furthermore, inspired by Malik (1970) on the distributions of maximum likelihood estimators of scale and shape parametrs of a single Pareto distribution, Elfessi and Jin (1996) showed, for the i-tuple $(i = 1, 2, ..., n)$ Pareto distributions, distributions of the maximum likelihood estimators of θ and $\alpha_i (i = 1, 2, ..., n)$ are given by

$$T \sim Pa\,(\theta, \alpha^*) \quad \text{and} \quad A_i \sim \Gamma^{-1}(m_i - 1, \ m_i \alpha_i), \tag{5.22}$$

where $i = 1, 2, ..., n$ and $\alpha^* = \sum_{i=1}^{n} m_i \alpha_i$.

Now, suppose $\mathbf{X}_{ki}^{m_{ki}} = (X_{ki}^{(1)}, X_{ki}^{(2)}, ..., X_{ki}^{(m_{ki})})$ is a random sample of size m_{ki} from a k−period of the ith truncated Pareto population, say $X_{ki} \sim Pa\,(\theta_k, \alpha_{ki})$ for $i = 1, 2, ..., n$; $j_{ki} = 1, 2, ..., m_{ki}$, and $k = 1, 2$ for ON and OFF, respectively, where θ_k denotes the common unknown scale parameter of k−period and $\alpha_{k1}, \alpha_{k2}, ..., \alpha_{kn}$ are unknown and possibly unequal shape parameters of k−period. Then, maximum likelihood estimators of θ_k and α_{ki} – denoted by $\widehat{\theta}_k$ and $\widehat{\alpha}_{ki}$, respectively – are given by

$$\widehat{\theta}_k = T_k = \min_{1 \leq i \leq n} \ \min_{1 \leq j_{ki} \leq m_{ki}} X_{ki}^{((j_{ki}))} = X_{ki(1)} \text{ and } \widehat{\alpha}_{ki} = A_{ki} = m_{ki} Y_{ki}^{-1},$$

$$\tag{5.23}$$

where $Y_{ki} = \sum_{j_{ki}=1}^{m_{ki}} \ln(X_{ki}^{(j_{ki})}/X_{ki(1)})]^{-1}$ for $k = 1, 2$ for ON, OFF, respectively; $i = 1, 2, ..., n$. Their respective distributions are given by

$$T_k \sim Pa\left(\theta_k, \alpha_k^*\right) \text{ and } A_{ki} \sim \Gamma^{-1}\left(m_{ki} - 1, \, m_{ki}\alpha_{ki}\right), \tag{5.24}$$

where $\alpha_k^* = \sum_{i=1}^{n} m_{ki}\alpha_{ki}$ ($k = 1, 2$ for ON, OFF, respectively; $i = 1, 2, ..., n$).

Then, by using (5.23), the maximum likelihood estimator, denoted by $\widehat{\Psi}$, of the Offered Optical Network Unit Load in (5.19) is given by

$$\widehat{\Psi} = P = \sum_{i=1}^{n}(A_{2i} - 1)A_{1i}T_1[A_{1i}A_{2i}(T_1 + T_2) - (A_{1i}T_1 + A_{2i}T_2)]^{-1}. \tag{5.25}$$

5.10.1 Classical inferences for Ψ

We are interested in testing the hypothesis

$$H_0^I : \Psi \leq \Psi_0 \text{ vs. } H_a^I : \Psi > \Psi_0, \tag{5.26}$$

where Ψ_0 is a known quantity with respect to the classical framework, based on the large sample approach. Suppose $\boldsymbol{\theta} = (\theta_1, \theta_2, \alpha_{11}, \alpha_{12}, ..., \alpha_{1n}, \alpha_{21}, \alpha_{22}, ..., \alpha_{2n})$, $X_{DATA} = [X_{kij}]_{k=1,2;\ i=1,2,...,n;\ j=1,,...,m_{ki}}$, and $x_{DATA} = [x_{kij}]_{k=1,2;\ i=1,2,...,n;\ j=1,,...,m_{ki}}$.

Then, the classical test variable, $T_\Psi = T(X_{DATA}; \boldsymbol{\theta})$, based on the large sample approach, is derived. It is well known that

$$(P - \mu_P)(\sigma_P^2)^{-1/2} = Z \sim N(0, 1), \tag{5.27}$$

where $\mu_P = \Psi$ is the mean, $\sigma_P^2 = (\partial P)'\widehat{\Sigma}_{(A_{ki}, T_k)}(\partial P)$ is the asymptotical variance, $\widehat{\Sigma}_{(A_{ki}, T_k)}$ being the estimator of the covariance matrix of A_{ki}'s and T_k's, and Z is the standard normal variate. Then, the classical pivotal quantity for estimating and

testing Ψ is given by

$$T_\Psi = (P - \Psi)(\sigma_P^2)^{-1/2}. \tag{5.28}$$

When $\boldsymbol{\theta}$ is specified, T_Ψ has a probability distribution that is free of nuisance para-meters. Furthermore, when $\mathbf{x}$ and nuisance parameters are fixed, the *cdf* of T_Ψ is a monotonically increasing function of Ψ for any given t_Ψ. Therefore, T_Ψ is a classical test variable that can be used to test the given hypothesis.

Thus, the classical p-value for testing

1. $H_0^I : \Psi \leq \Psi_0$ vs. $H_a^I : \Psi > \Psi_0$ is given by $p_c^I = 1 - \Phi(t_\Psi)$,

2. $H_0^{II} : \Psi \geq \Psi_0$ vs. $H_a^{II} : \Psi < \Psi_0$ is given by $p_c^I = \Phi(t_\Psi)$, and

3. $H_0^{III} : \Psi = \Psi_0$ vs. $H_a^{III} : \Psi \neq \Psi_0$ is given by $p_c^{III} = 2\min[\Phi(t_\Psi), 1 - \Phi(t_\Psi)]$,

where $t_\Psi = (p - \Psi_0)\left(s_P^2\right)^{-\frac{1}{2}} - t_\Psi$, p, s_P^2 being the observed values of T_Ψ, P, and $S_P^2 = \widehat{\sigma}_P^2$, respectively – and $\Phi(.)$ is the distribution function of Z.

Furthermore, a $100(1 - \gamma)\%$ empirical large sample solution for the confidence interval estimation of Ψ is given by

$$\left(p - Z_{\gamma/2}\sqrt{s_P^2}, p + Z_{\gamma/2}\sqrt{s_P^2} \right), \tag{5.29}$$

where $Z_{\gamma/2}$ is $\gamma/2$th quantile of Z .

5.10.2 Generalized inferences for Ψ

Now, we are interested in performing inferences of Ψ with respect to the gener-alized variable method as opposed to the classical inferences of Ψ.

5.10.2.1 The testing of hypothesis for Ψ

Assume that for each component, m_{1i} and m_{2i} *ON*-and *OFF*-packet sizes are avail-able. Furthermore, assume that $\mathbf{x}_{1i}^{m_{1i}} = (x_{1i}^{(1)}, x_{1i}^{(2)}, ..., x_{1i}^{(m_{1i})})$ and $\mathbf{x}_{2i}^{m_{2i}} = (x_{2i}^{(1)}, x_{2i}^{(2)}, ...,$

$x_{2i}^{(m_{2i})}$), $k = 1, 2$ for ON, OFF, respectively; $i = 1, 2, ..., n$, are the observed values of

the random samples $\mathbf{X}_{1i}^{m_{1i}} = (X_{1i}^{(1)}, X_{1i}^{(2)}, ..., X_{1i}^{(m_{1i})})$ and $\mathbf{X}_{2i}^{m_{2i}} = (X_{2i}^{(1)}, X_{2i}^{(2)}, ..., X_{2i}^{(m_{2i})})$,

$k = 1, 2$ for ON, OFF, respectively; $i = 1, 2, ..., n$, respectively.

Now, from

$$2m_{ki}\alpha_{ki}A_{ki}^{-1} = W_{ki} \sim \varkappa_{2m_{ki}-2}^{2}, \tag{5.30}$$

where W_{ki} is $\varkappa^2$ variate with the degrees-of-freedom $(2m_{ki}-2)$, the generalized pivotal

quantity for estimating α_{ki} is given by

$$R_{\alpha_{ki}} = 0.5 W_{ki} a_{ki}/m_{ki}, \tag{5.31}$$

where a_{ki} is the observed value of A_{ki}, or simply anestimate of α_{ki} ($k = 1, 2$ for

ON, OFF, respectively; $i = 1, 2, ..., n$). Similarly, from

$$2\alpha_{k}^{*}log(T_{k}/\theta_{k}) = V_{k} \sim \varkappa_{2}^{2}, \tag{5.32}$$

where V_k is $\varkappa^2$ variate with degrees-of-freedom 2, the generalized pivotal quantity for

estimating θ_k is given by

$$R_{\theta_{k}} = t_{k}e^{-V_{k}/(\sum_{i=1}^{n} W_{ki}a_{ki})}, \tag{5.33}$$

where t_k is the observed value of T_k, or simply an estimate of θ_k ($k = 1, 2$ for

ON, OFF, respectively).

Now, consider the potential exact generalized test variable defined by

$$T(X_{DATA}; x_{DATA}, \boldsymbol{\theta}) = R(X_{DATA}; x_{DATA}, \boldsymbol{\theta}) - \Psi, \tag{5.34}$$

where $R(X_{DATA}; x_{DATA}, \boldsymbol{\theta})$ is the generalized pivotal quantity for Ψ, where α_{ki} and

θ_k have been replaced by $R_{\alpha_{ki}}$ and R_{θ_k} ($k = 1, 2$ for ON, OFF, respectively; $i =$

$1, 2, ..., n$), and is given by

$$R(X_{DATA}; x_{DATA}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{(R_{\alpha_{2i}} - 1) R_{\alpha_{1i}} R_{\theta_1}}{[R_{\alpha_{1i}} R_{\alpha_{2i}} (R_{\theta_1} + R_{\theta_2}) - (R_{\alpha_{1i}} R_{\theta_1} + R_{\alpha_{2i}} R_{\theta_2})]},$$

(5.35)

Since

1. the observed value of $T(X_{DATA}; x_{DATA}, \boldsymbol{\theta})$, which is denoted by

   $t(X_{DATA}; x_{DATA}, \boldsymbol{\theta})$, is 0,

2. it is clear that when Ψ is specified, $T(X_{DATA}; x_{DATA}, \boldsymbol{\theta})$ has probability

   distribution that is free of nuisance parameters,

3. when x_{DATA} and nuisance parameters are fixed, the

   cumulative density function of $T(X_{DATA}; x_{DATA}, \boldsymbol{\theta})$ is a monotonically

   decreasing function of Ψ for any given $t(x_{DATA}; x_{DATA}, \boldsymbol{\theta})$,

$T(X_{DATA}; x_{DATA}, \boldsymbol{\theta})$ is a generalized test variable (Weerahandi 1995 and 2004) that

can be used to test the given hypothesis. Thus, the generalized p-value for testing

1. $H_0^I : \Psi \leq \Psi_0$ vs. $H_a^I : \Psi > \Psi_0$ is given by $p_g^I = Pr(R(X_{DATA}; x_{DATA}, \boldsymbol{\theta}) >$

   $\Psi_0)$,

2. $H_0^{II} : \Psi \geq \Psi_0$ vs. $H_a^{II} : \Psi < \Psi_0$ is given by $p_g^{II} = Pr(R(X_{DATA}; x_{DATA}, \boldsymbol{\theta}) <$

   $\Psi_0)$, and

3. $H_0^{III} : \Psi = \Psi_0$ vs. $H_a^{III} : \Psi \neq \Psi_0$ is given by $2 \min[Pr(R(X_{DATA}; x_{DATA}, \boldsymbol{\theta}) >$

   $\Psi_0), R(X_{DATA}; x_{DATA}, \boldsymbol{\theta}) < \Psi_0)]$

These p-values can be evaluated through numerical procedures:

1. Numerical integration:

   is performed with respect to $V_k(k = ON, OFF)$ and $W_{ki}(i = 1, 2, ..., n; k = 1, 2)$,

which

are independent random variables with known density functions.

2. Monte Carlo Simulation method:

Once the data are obtained, compute

$$a_{ki} = \frac{m_i}{\left[\sum_{j=1}^{m_i} \ln(x_{ij}/x_{i(1)})\right]} \text{ for } i = 1, 2, ..., n; \ k = ON, OFF \text{ and}$$

$$t_k = \min_{1 \le i \le n} \ \min_{1 \le j_{ki} \le m_{ki}} x_{ki}^{(j_{ki})} \text{ for } k = ON, OFF.$$

Then, generate a large number of random numbers $(V_k, W_{ki})_{i=1,2,...,n; \ k=ON,OFF}$,

where

$$V_k \sim \varkappa_2^2(0) \text{ and } W_{ki} \sim \varkappa_{2m_{ki}-2}^2(0), \text{ for } \forall i, k.$$

For each value of $(V_k, W_{ki})_{i=1,2,...,n; \ k=ON,OFF}$, compute

$$R_{\theta_k} = t_k e^{-V_k/(\sum_{i=1}^{n} W_{ki} a_{ki})}, \text{ for } k = ON, OFF.$$

Then, compute

$$R(X_{DATA}; x_{DATA}, \boldsymbol{\theta}),$$

and finally compute

$$T(X_{DATA}; x_{DATA}, \boldsymbol{\theta}) = R(X_{DATA}; x_{DATA}, \boldsymbol{\theta}) - \Psi.$$

Now,

1. the fraction of random numbers pairs for which $R(X_{DATA}; x_{DATA}, \boldsymbol{\theta}) < \Psi_0$

   yields p_g^I,

2. the fraction of random numbers pairs for which $R(X_{DATA}; x_{DATA}, \boldsymbol{\theta}) > \Psi_0$

   yields p_g^{II}, and

3. the fraction of random numbers pairs for which $\min[R(X_{DATA};x_{DATA}, \boldsymbol{\theta}) < \theta_0,$

$R(X_{DATA};x_{DATA}, \boldsymbol{\theta}) > \Psi_0 \,]$ yields p_g^{III}.

5.10.2.2 The confidence interval for Ψ

Since

1. the value of $R(X_{DATA};x_{DATA}, \boldsymbol{\theta})$ is Ψ, and

2. the distribution of $R(X_{DATA};x_{DATA}, \boldsymbol{\theta})$ is independent of any unknown

   parameters,

$R(X_{DATA};x_{DATA}, \boldsymbol{\theta})$ is a generalized pivotal quantity for constructing a $100(1-\gamma)\%$

confidence interval for θ, where $1-\gamma$ is the confidence level or the confidence coefficient

(Weerahandi 1993).

Now, one-sided as well as two-sided confidence intervals are constructed as follows:

1. the lower bound, $R_{1-\gamma}(X_{DATA};x_{DATA}, \boldsymbol{\theta})$, of a $100(1-\gamma)\%$ one-sided

   confidence interval for Ψ is sought such that

$$1 - \gamma = \Pr(R(X_{DATA};x_{DATA}, \boldsymbol{\theta}) \leq R_{1-\gamma}(X_{DATA};x_{DATA}, \boldsymbol{\theta})), \qquad (5.36)$$

2. the upper bound, $R_{1-\gamma}(\theta; t, \mathbf{a})$, of a $100(1-\gamma)\%$ one-sided confidence

   interval for Ψ is sought such that

$$\gamma = \Pr(R(X_{DATA};x_{DATA}, \boldsymbol{\theta}) \leq R_{1-\gamma}(X_{DATA};x_{DATA}, \boldsymbol{\theta})),$$

$$(5.37)$$

3. the lower limit, $R_{\gamma/2}^{l}$, and the upper limit, $R_{1-\gamma/2}^{u}$, of a $100(1-\gamma)\%$ two-sided

   confidence interval for Ψ are sought such that

$$1 - \gamma = \Pr(R_{\gamma/2}^{l} \leq R(X_{DATA};x_{DATA}, \boldsymbol{\theta}) \leq R_{1-\gamma/2}^{u}), \qquad (5.38)$$

For the actual coverage probabilities (the empirical confidence level), repeat the above process for a larger number of times, and calculate the fraction of times θ falls within the calculated (empirical) generalized confidence intervals.

5.10.3 The Bayesian Estimation for Ψ

Prompted by Sinha and Sloan (1988) on the Bayes estimation of the parameters and reliability function of the 3-parameter Weibull distribution, the Bayes estimates of the Offered Optical Network Unit Load of the Networking system is discussed in this section.

Using an approximation due to Lindley (1980), the posterior s-expectation – which is the Bayes estimator of Ψ (or $\Psi(\boldsymbol{\theta})$) under a squared-error loss function – is asymptotically estimated by

$$E\left[\Psi(\boldsymbol{\theta})\,|\mathbf{x}\right] = \frac{\int_{\Theta} \Psi(\boldsymbol{\theta}) h(\boldsymbol{\theta}) \exp\left\{L(\boldsymbol{\theta})\right\} d\boldsymbol{\theta}}{\int_{\Theta} h(\boldsymbol{\theta}) \exp\left\{L(\boldsymbol{\theta})\right\} d\boldsymbol{\theta}}, \tag{5.39}$$

or equivalently by,

$$E\left[\Psi(\boldsymbol{\theta})\,|\mathbf{x}\right] \;=\; \left| \begin{array}{c} \psi + 1/2 \sum_i \sum_j (\psi_{ij} + 2\psi_i \rho_j)\sigma_{ij} + \\[2mm] 1/2 \sum_i \sum_j \sum_k \sum_l L_{ijk}\sigma_{ij}\sigma_{kl}\psi_l \end{array} \right|_{\widehat{\boldsymbol{\theta}}} \tag{5.40}$$
$$+\text{terms of order } n^{-2} \text{ or smaller},$$

where

$\Psi = \psi = \Psi(\boldsymbol{\theta}) = $ Offered Optical Network Unit Load which is a function of $\boldsymbol{\theta}$,

which is given by

$\Psi(\boldsymbol{\theta}) = \sum_{i=1}^{n} (\alpha_{2i} - 1)\alpha_{1i}\theta_1 [\alpha_{1i}\alpha_{2i}(\theta_1 + \theta_2) - (\alpha_{1i}\theta_1 + \alpha_{2i}\theta_2)]^{-1}$,

$\boldsymbol{\theta} = $ Pareto vector parameter: $(\theta_1, \theta_2, \alpha_{11}, \alpha_{12}, ..., \alpha_{1n}, \alpha_{21}, \alpha_{22}, ..., \alpha_{2n})$,

$\widehat{\boldsymbol{\theta}} = $ MLE vector estimator: $(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\alpha}_{11}, \widehat{\alpha}_{12}, ..., \widehat{\alpha}_{1n}, \widehat{\alpha}_{21}, \widehat{\alpha}_{22}, ..., \widehat{\alpha}_{2n})$,

Θ = range space of $\boldsymbol{\theta}$,

$\mathbf{x} = (x_1, x_2, ..., x_n)$,

$E\left[\Psi(\boldsymbol{\theta})\,|\mathbf{x}\right]$ = posterior s-expectation of $\Psi(\boldsymbol{\theta})$,

$h = h(\boldsymbol{\theta})$ = joint prior distribution,

$\psi_i, \psi_{ij} = \partial\psi/\partial\theta_i, \partial^2\psi/\partial\theta_i\partial\theta_j$; where $i, j = 1, 2, ..., 2(n+1)$,

$L = L(\boldsymbol{\theta})$ = logarithmic likelihood function,

$L_{ijk} = \partial^3\psi/\partial\theta_i\partial\theta_j\partial\theta_k$; where $i, j, k = 1, 2, ..., 2(n+1)$,

$\rho = \rho(\boldsymbol{\theta}) = \log[h(\boldsymbol{\theta})]$,

$\rho_j = \partial\rho/\partial\theta_j$; where $j = 1, 2, ..., 2(n+1)$,

$\sigma_{ij} = -\left\{L_{ij}\right\}^{-1}$; where $i, j = 1, 2, ..., 2(n+1)$, and

$i, j, k, l = 1, 2, ..., 2(n+1)$

Green (1980) discussed and recommended this linear Bayes estimator as a very good and operational approximation for the ratio of multi-dimensional integrals. Sinha (1986) has applied this method in many useful applications.

For mathematical simplicity and keeping with the view that this demonstration is only to show how this method works, our discussion will be restricted only to any 3 parameters that are involved in the Offered Optical Network Unit Load. Then, (5.40) reduces to

$$\widehat{\psi}_B = E\left[\psi\,|\mathbf{x}\right] = \psi + (\psi_1 a_1 + \psi_2 a_2 + \psi_3 a_3 + a_4 + a_5)$$

$$+1/2[A(\psi_1\sigma_{11} + \psi_2\sigma_{12})$$

$$+\quad B(\psi_1\sigma_{21} + \psi_2\sigma_{22} + \psi_3\sigma_{23})$$

$$+\quad C(\psi_1\sigma_{31} + \psi_2\sigma_{23} + \psi_3\sigma_{33}),$$

evaluated at $\widehat{\boldsymbol{\theta}} = (\theta_1, \theta_2, \theta_3)$,

where

$$a_1 = \rho_1\sigma_{11} + \rho_2\sigma_{12} + \rho_3\sigma_{13,}$$

$$a_2 = \rho_1\sigma_{21} + \rho_2\sigma_{22} + \rho_3\sigma_{23,}$$

$$a_3 = \rho_1\sigma_{31} + \rho_2\sigma_{32} + \rho_3\sigma_{33,}$$

$$a_4 = \psi_{12}\sigma_{12} + \psi_{13}\sigma_{13} + \psi_{23}\sigma_{23,}$$

$$a_5 = 1/2(\psi_{11}\sigma_{11} + \psi_{22}\sigma_{22} + \psi_{23}\sigma_{23}),$$

$$A = \sigma_{11}L_{111} + 2\sigma_{12}L_{121} + 2\sigma_{13}L_{131} + 2\sigma_{23}L_{231} + \sigma_{22}L_{221} + \sigma_{33}L_{331,}$$

$$B = \sigma_{11}L_{112} + 2\sigma_{12}L_{122} + 2\sigma_{13}L_{132} + 2\sigma_{23}L_{232} + \sigma_{22}L_{222} + \sigma_{33}L_{332,}$$

$$C = \sigma_{11}L_{113} + 2\sigma_{12}L_{123} + 2\sigma_{13}L_{133} + 2\sigma_{23}L_{233} + \sigma_{22}L_{223} + \sigma_{33}L_{333,}$$

subscripts $1, 2,$ and 3 refer to three parameters.

Further mathematical tractability and simplicity, two substreams with their Pareto distributed ON-and OFF-periods, say $X_{qi} \sim Pa\left(\alpha_{q,i}, \theta_q\right)$ for $i = 1, 2; q = ON, OFF,$ where θ_q denotes the common unknown scale parameter of $q-$ periods and $\alpha_{q,1}, \alpha_{q,2}$ are unknown and possibly unequal shape parameters of $q-$ periods, are considered. And, furthermore, for the mathematical simplicity

$$\boldsymbol{\theta} = (\theta_{ON}, \theta_{OFF}, \alpha_{ON,1} = \alpha_{ON,2} = \alpha_{OFF,1} = \alpha_{OFF,2} = \alpha), \tag{5.41}$$

is considered.

Then,

$$\psi = \Psi(\boldsymbol{\theta}) = \sum_{i=1}^{n} (\alpha_{2i} - 1)\alpha_{1i}\theta_1[\alpha_{1i}\alpha_{2i}(\theta_1 + \theta_2) - (\alpha_{1i}\theta_1 + \alpha_{2i}\theta_2)]^{-1}, \tag{5.42}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \alpha_{11}, \alpha_{12}, ..., \alpha_{1n}, \alpha_{21}, \alpha_{22}, ..., \alpha_{2n}),$ becomes

$$\psi = \Psi(\boldsymbol{\theta}) = \theta_{ON}/(\theta_{ON} + \theta_{OFF}). \tag{5.43}$$

Now, suppose that we are ignorant about the parameters $(\theta_{ON}, \theta_{OFF}, \alpha)$. Assume a vague prior for $(\theta_{ON}, \theta_{OFF}, \alpha)$ that is given by

$$h(\boldsymbol{\theta}) = v(\theta_{ON}, \theta_{OFF}, \alpha) \propto 1/(\theta_{ON}\theta_{OFF}\alpha). \tag{5.44}$$

Therefore,

$$\rho_1 = -1/\theta_{ON}, \rho_2 = -1/\alpha, \rho_3 = -1/\theta_{OFF} \tag{5.45}$$

where subscripts $1, 2,$ and 3 of ρ refer to $\theta_{ON}, \theta_{OFF},$ and α. Then, the Bayesian estimation of ψ under the squared-error loss is given by

$$\widehat{\psi}_B = E\left[\psi \,|\mathbf{x}\right] = \left[\psi + \sum_{i=1}^{3} \tau_i W_i + a_4 + a_5\right]_{\widehat{\theta}_{ON}, \widehat{\theta}_{OFF}, \widehat{\alpha}}, \tag{5.46}$$

$$\tau_1 = \left(\frac{A}{2} - \frac{1}{\theta_{ON}}\right), \tau_2 = \left(\frac{B}{2} - \frac{1}{\theta_{OFF}}\right), \tau_3 = \left(\frac{C}{2} - \frac{1}{\alpha}\right), \tag{5.47}$$

$$W_j = \sum_{i=1}^{3} \psi_i \sigma_{ji}; \ j = 1, 2, 3, \tag{5.48}$$

where subscripts $1, 2,$ and 3 of W and ψ refer to $\theta_{ON}, \theta_{OFF},$ and α.

In order to compare posterior standard-deviation estimate of ψ with the asymptotic standard-deviation estimate of its maximum likelihood counterpart $\widehat{\psi}$, define:

$$\widehat{\psi}_B^2 = E\left[\psi^2 \,|\mathbf{x}\right] = [\psi^2 + 2\psi\left(\sum_{i=1}^{3} \lambda_i W_i + a_4 + a_5\right) + \sum_{i=1i}^{3} \psi_2 \sigma_{ii} +$$

$$+ 2\psi_1\psi_2\sigma_{12} + \psi_1\psi_2\sigma_{13} + \psi_2\psi_3\sigma_{23}]_{\widehat{\theta}, \widehat{\alpha}_1, \widehat{\alpha}_2}. \tag{5.49}$$

Then,

$$Var\{\widehat{\psi}_B\} = \widehat{\psi}_B^2 - (\widehat{\psi}_B) < Var\{\widehat{\psi}\}. \tag{5.50}$$

121

5.11 Illustrative examples

Example 1. Simulated data

This example deals with the Pareto distributions, $X_{ki} \sim Pa\left(\alpha_{ki}, \theta_k\right)$ where $k = 1(ON), 2(OFF); i = 1, 2$, that are generated by the following population parameters: $\theta_1 = \theta_2 = 1$, $\alpha_{11} = 1.1$, $\alpha_{12} = 1.2$, $\alpha_{21} = 1.3$, $\alpha_{22} = 1.4$ and sample sizes $m_{11} = m_{12} = m_{21} = m_{22} = 10$. This parameter specification has helped to keep the Hurst parameter (H) between 0.5 and 1 so that a moderate load is generated in the system. The Offered Optical Network Unit Load of the system, with these parameter specification, is $\Psi = 1.2002$. *ON*-and *OFF*-periods of two substreams that were generated from the Pareto distributions with these parameter specifications are given below:

$X_{11} \sim Pa\left(\theta_1, \alpha_{11}\right) : 1.74, 1.39, 1.41, 1.04, 1.53, 5.37, 1.43, 12.68, 1.63, 3.96$

$X_{12} \sim Pa\left(\theta_1, \alpha_{12}\right) : 6.51, 5.11, 1.12, 1.19, 1.16, 4.63, 1.93, 1.06, 1.05, 06.25$

$X_{21} \sim Pa\left(\theta_2, \alpha_{21}\right) : 1.04, 6.36, 1.95, 1.05, 1.02, 1.01, 4.73, 1.24, 4.20, 11.54$

$X_{22} \sim Pa\left(\theta_2, \alpha_{22}\right) : 5.56, 1.01, 1.22, 1.24, 1.49, 1.54, 2.50, 10.17, 1.43, 1.18$

I. Confidence intervals:

Assuming that all of the above parameters are unknown:

1. the lower bound of a 95% one-sided generalized empirical confidence interval for Ψ calculated from this data is 0.0994,

2. the upper bound of a 95% one-sided generalized empirical confidence interval for Ψ calculated from this data is 5.6435,

3. the upper and lower limits of a 95% two-sided generalized empirical confidence interval for Ψ calculated from this data are 0.0579 and 11.3692.

The comparison of classical and generalized confidence intervals for Ψ are sum-

marized as follows:

Table 5.1 Comparison of confidence intervals for Ψ

Confidence interval	Generalized	Classical
95%	$(0.0579, 11.3692)$	$(0.0481, 11.7527)$

When comparing these lengths for Ψ, it is clear that the generalized variable method, having a shorter length, outperforms its counter part, the classical procedure, for this particular problem.

II. p-values:

Furthermore, the comparison of classical and generalized p-values for various tests are summarized as follows:

Table 5.2 Comparison of p-values

Test	Generalized	Classical
$\Psi \leq 12.0$ vs. $\Psi > 12.0$	0.6574	0.0095
$\Psi \leq 1.0$ vs. $\Psi > 1.0$	0.0473	0.4346

When Ψ,assuming that it is unknown, is tested, to see whether it is greater than 12.0 (or the claim or the alternative hypothesis being $\theta > 12.0$ against $\theta \leq 12.0$), the generalized p-value, having a value greater than 0.05, is against the claim which is true since Ψ is 1.2002. However, the classical p-value, having a value less than 0.05, suggests that the claim is acceptable which is false since Ψ is 1.2002. When θ, assuming that it is unknown, is tested to see whether it is less than 1.0, the generalized p-value,

123

having a value less than 0.05, is supportive of the claim which is true- since Ψ is 1.2002; however, the classical p-value, having a value greater than 0.05, suggests that the claim is not acceptable which is false – since Ψ is 1.2002. Both these arguments clearly show that the generalized variable method provides accurate, reliable, and non-misleading results, while the classical approach fails to do so for this particular case. Hence, the generalized variable method outperforms the classical approach for this particular case.

III. Size:

Table 5.3 shows the classical and the generalized empirical (actual) type I error rates (the rejection rate of the null hypothesis: the fraction of times the p-value is less than the nominal level) for the test $H_0 : \Psi \leq \Psi_0$ Vs. $H_a : \Psi > \Psi_0$ when nominal (intended) type I error rate is at 0.05. All results are based on $100,000$ replications.

Table 5.3 Comparison of actual type I error rates for testing H_0 when $\gamma = 0.05$

Parameters: $\boldsymbol{\theta} = (\theta_1, \theta_2)$ $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$, $\Psi = \Psi_0$	Generalized	Classical
$\boldsymbol{\theta} = (1,1)$, $\boldsymbol{\alpha} = (1.1, 1.2, 1.3, 1.4)$, $\Psi = \Psi_0 = 1.2002$	0.052	0.042
$\boldsymbol{\theta} = (1,1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.4)$, $\Psi = \Psi_0 = 1.4232$	0.056	0.264
$\boldsymbol{\theta} = (1,1)$, $\boldsymbol{\alpha} = (2.3, 1.2, 1.3, 0.5)$, $\Psi = \Psi_0 = 1.5277$	0.049	0.438
$\boldsymbol{\theta} = (1,1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.5)$, $\Psi = \Psi_0 = 1.5414$	0.054	0.124
$\boldsymbol{\theta} = (1,1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.6)$, $\Psi = \Psi_0 = 1.7708$	0.051	0.008
$\boldsymbol{\theta} = (1,1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.7)$, $\Psi = \Psi_0 = 2.4080$	0.050	0.048

According to this simulation study, when compared with the classical procedure, the actual type I error rates (actual sizes of tests) of the generalized procedure get as close as to the intended size. Therefore, the generalized procedure outperforms the classical procedure for this particular problem.

IV. Power:

Table 5.4 and 5.5, respectively, show the power comparison for testing $H_0^I : \theta \leq 100$ vs. $H_a^I : \theta > 100$ with and after adjusting the size at $\gamma = 0.10$ based on 100 000 replications.

Table 5.4 Comparison of powers for testing $\Psi \leq 1.2002$ vs. $\Psi > 1.2002$ without adjusting the size

Parameters: $\boldsymbol{\theta} = (\theta_1, \theta_2)$ $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$, Ψ	Generalized	Classical
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (1.1, 1.2, 1.3, 1.4)$, $\Psi = 1.2002$	0.052	0.042
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.4)$, $\Psi = 1.4232$	0.121	0.089
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.3, 1.2, 1.3, 0.5)$, $\Psi = 1.5277$	0.325	0.297
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.5)$, $\Psi = 1.5414$	0.514	0.501
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.6)$, $\Psi = 1.7708$	0.725	0.702
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.7)$, $\Psi = 2.4080$	0.879	0.822

Table 5.5 Comparison of powers for testing $\Psi \leq 1.202$ vs. $\Psi > 1.2002$

after adjusting the size

Parameters: $\boldsymbol{\theta} = (\theta_1, \theta_2)$ $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$, Ψ	Generalized	Classical
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (1.1, 1.2, 1.3, 1.4)$, $\Psi = 1.2002$	0.050	0.050
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.4)$, $\Psi = 1.4232$	0.118	0.088
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.3, 1.2, 1.3, 0.5)$, $\Psi = 1.5277$	0.322	0.296
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.5)$, $\Psi = 1.5414$	0.510	0.498
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.6)$, $\Psi = 1.7708$	0.722	0.693
$\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.7)$, $\Psi = 2.4080$	0.833	0.817

When compared, the values of generalized and classical powers, without and after adjusting the size, clearly suggest that generalized variable method outperforms the classical method in terms of power for this particular case.

Without adjusting the size, the values of powers of the generalized variable method for testing $\Psi \leq 1.202$ Vs. $\Psi > 1.2002$ clearly suggest that the generalized variable method outperforms the classical method. Even after adjusting the size, the generalized method still maintains a light advantage over the classical method. The size of the test has to be adjusted to get a meaningful comparison of power of tests. But, in reality practitioners, being less-concern about the size, are not interested in adjusting the nominal size in order to get the desired level γ.

Overall, according to these simulation studies, when compared with the classical procedure, not only is the actual type I error rates of the generalized procedure closer to the intended size, but the generalized procedure also outperforms the classical

procedure in terms of power.

V. Coverage probabilities:

The classical actual coverage probabilities for the maximum likelihood estimation method with intended confidence levels $\gamma = 0.1$ are given in Table 5.6. These coverage probabilities are based on 100 000 simulated random samples from the Pareto density given in equation (3.1). The random samples are generated for *ON*- and *OFF*- periods of two substreams of networking by plugging the known values of parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$, $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$ (say $\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (1.1, 1.2, 1.3, 1.4)$) to the quantile function $Q(u) = \theta/u^{1/\alpha}$ for each α value. In addition, an ordered random sample of size m (say $m = 10$) from the uniform distribution, $u \sim U(0, 1)$, is required to substitute for u. Thus, one random sample with size m (say $m = 10$) from Pareto distributions with parameters θ, $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$ (say $\theta = 1$, $\boldsymbol{\alpha} = (1.5, 1.2, 1.3, 1.4)$) can be generated. In this simulation study, seventy five thousand such samples are generated to get a single cell value in Table 5.6.

The classical approximate $100(1 - \gamma)\%$ confidence intervals for parameter Ψ are calculated by using $\left(\widehat{\Psi} - Z_{\gamma/2}SE_{\widehat{\Psi}}, \ \widehat{\Psi} + Z_{\gamma/2}SE_{\widehat{\Psi}}\right)$ while the generalized approximate $100(1 - \gamma)\%$ confidence intervals for parameter θ are calculated by using $(R_\Psi(\gamma/2), R_\Psi(1 - \gamma/2))$.

Table 5.6 Comparison of probability coverages for 90% two-sided
confidence intervals

Parameters: n, $\boldsymbol{\theta} = (\theta_1, \theta_2)$ $\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22})$	Generalized	Classical
$n = 2$, $\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (1.1, 1.2, 1.3, 1.4)$	0.91	0.89
$n = 2$, $\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.4)$	0.88	0.81
$n = 2$, $\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.3, 1.2, 1.3, 0.5)$	0.86	0.85
$n = 2$, $\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.5)$	0.83	0.88
$n = 2$, $\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.6)$	0.93	0.82
$n = 2$, $\boldsymbol{\theta} = (1, 1)$, $\boldsymbol{\alpha} = (2.1, 1.2, 1.3, 0.7)$	0.88	0.91

According to these simulation results, one can clearly see that the actual (empirical) probability coverage for the generalized method is getting closer to the intended (nominal) coverage probabilities, and furthermore see that the parameters are overly estimated under the maximum likelihood estimation method

Example 2. Cable TV and MCI backbone data

This example deals with the following data which appeared at http://www.csif.cs.ucdavis.edu/˜kramer, the homepage of Glen Kramer, University of California at Davis, California, USA.

MCI backbone: This packet distribution was measured on MCI backbone and reported in Claffy, Miller, and Thompson (1998).

CATV head-end (upstream) and CATV head-end (downstream): This packet distribution was measured at CATV head-end and reported in Sala and Gummalla (2001).

For our convenience and the mathematical tractability, small random samples were taken out of these real data sets, and, furthermore, *ON*-and *OFF*-periods are made representative from these Pareto distributed samples as follows :

$$X_{11} \sim Pa\left(\theta_1, \alpha_{11}\right) : 1, 95, 40, 2273562, 9612,$$

$$4710, 768, 906, 2763, 5304$$

$$X_{12} \sim Pa\left(\theta_1, \alpha_{12}\right) : 1433, 159, 12, 32, 1621,$$

$$187, 165, 191, 1903, 76$$

$$X_{21} \sim Pa\left(\theta_2, \alpha_{21}\right) : 3, 1676, 124, 38, 1733933,$$

$$1051, 3308, 1990, 2492, 90$$

$$X_{22} \sim Pa\left(\theta_2, \alpha_{21}\right) : 1433, 159, 12, 32, 162,$$

$$1187, 165, 191, 1903, 76$$

The Offered Optical Network Unit Load calculated by using the given data set is $\widehat{\Psi} = 0.0476$.

Confidence intervals:

Assuming that all of the above parameters are unknown:

1. the lower bound of one-sided 95% generalized empirical confidence interval for θ calculated from this data is 0.0016,

2. the upper bound of the one-sided 95% generalized empirical confidence interval for θ calculated from this data is 0.0617,and

3. the upper and lower limits of two-sided 95% generalized empirical confidence interval for θ calculated from this data are 0.0009 and 0.0811.

The comparison of classical and generalized confidence intervals are summarized as follows:

Table 5.7 Comparison of confidence intervals

Confidence interval	Generalized	Classical
95%	$(0.0009, 0.0811)$	$(0.0005, 0.0952)$

When comparing these lengths for Ψ, it is clear that the generalized variable method, having a shorter length, outperforms its counter part, the classical procedure, for this particular problem.

II. p-values:

Furthermore, the comparison of classical and generalized p-values for various tests are summarized as follows:

Table 5.8 Comparison of p-values

Test	Generalized	Classical
$\Psi \leq 1.00$ vs. $\Psi > 1.00$	0.8492	0.0000
$\Psi \leq 0.04$ vs. $\Psi > 0.04$	0.0000	0.5735

When Ψ,assuming that it is unknown, is tested, to see whether it is greater than 1.0 (or the claim or the alternative hypothesis being $\theta > 1.0$ against $\theta \leq 1.0$), the generalized p-value, having a value greater than 0.05, is against the claim which is true since the estimate of Ψ is 0.0476. However, the classical p-value, having a value less than 0.05, suggests that the claim is acceptable which is false since the estimate

of Ψ is 0.0476. Assuming that it is unknown, when θ is tested to see whether it is less than 0.04, the generalized p-value, having a value less than 0.05, is supportive of the claim which is true - since the estimate of Ψ is 0.0476; however, the classical p-value, having a value greater than 0.05, suggests that the claim is not acceptable which is false since the estimate of Ψ is 0.0476. Both these arguments clearly show that the generalized variable method provides accurate, reliable, and non-misleading results, while the classical approach fails to do so for this particular case. Hence, the generalized variable method outperforms the classical approach for this particular case.

Example 3. Bayesian approach

This example deals with two substreams having their *ON*-and *OFF*-periods distributed as Pareto population, say $X_{qi} \sim Pa\left(\theta_q, \alpha_{q,i}\right)$ for $i = 1, 2$; $q = ON, OFF$, where θ_q denotes the common unknown scale parameter of $q-$ periods and $\alpha_{q,1}, \alpha_{q,2}$ are unknown and possibly unequal shape parameters of $q-$ periods. Furthermore, For the mathematical simplicity $\boldsymbol{\theta} = (\theta_{ON}, \theta_{OFF}, \alpha_{ON,1} = \alpha_{ON,2} = \alpha_{OFF,1} = \alpha_{OFF,2} = \alpha)$, $m_{ON,1} = m_{ON,2} = m_{OFF,1} = m_{OFF,2} = 10$.

Table 5.9 shows the maximum likelihood and the Bayes estimates of the Offered Optical Network Unit Load for different common scale parameters of *ON*-and *OFF*-periods for two streams of file transmission lines.

Table 5.9 MLE and Bayes estimates of the Offered Optical

Network Unit Load Ψ based on varying (θ_1, θ_2)

θ_1	θ_2	True Ψ	$\widehat{\Psi}$	Ψ^*	Est. Var$\{\hat{\ }\}$	Posterior Var
1.0	1.1	0.4761	0.4562	0.4700	0.00001	0.00002
1.1	1.2	0.4782	0.4581	0.4710	0.00007	0.00008
1.2	1.3	0.4800	0.4703	0.4745	0.00009	0.00009
1.3	1.4	0.4814	0.4712	0.4985	0.00010	0.00009
1.4	1.5	0.4827	0.4789	0.4801	0.00012	0.00011

$\hat{\ }$ = Maximum Likelihood Estimate; * = Bayes Estimate

Different estimation techniques – the classical, the generalized, and the Bayesian – for performing the inferences of the Offered Optical Network Unit Load (OOL) Ψ, which is a function of several 2-parameter Pareto populations in this case, have been considered. When all these techniques are compared, the generalized variable method has become the more suitable candidate to perform inferences about Ψ in the sense of the numerical feasibility, the sample size, and the complexity of the function.

.

CHAPTER VI

OVERVIEW, SUMMARY, AND FUTURE RESEARCH

6.1 Overview

This dissertation has analyzed the intensive applications of the recently developed generalized inference method especially in the most sophisticated, fast developing Communication Technology and Information Technology, which have been made possible through Computer Networking and Data Transmission.

Computer Networks are inherently social networks, linking people, organizations, and knowledge. The Internet increases people's social capital, establishing contact with friends and relatives who live nearby and far away. New tools must be developed to help people navigate and find knowledge in complex, fragmented, and networked societies. With the improvements as well as the advancements of computer technology, statistics are heavily involved in providing a fast, reliable, efficient, competitive service to the society. In this regard, this study has developed exact inferential techniques that can be utilized to overcome difficulties of handling the other exact and asymptotic parametric estimation procedures in the face of nuisance parameters. This newly developed technique can be used to perform exact inferences on the complicated functions of parameters of underlying distributions, where classical inference methods cannot provide exact solutions. The generalized inference method was extensively utilized when inferences of the common scale and the common shape

parameter of several Pareto population, and the common location and the common shape parameter of exponential distribution were performed in Chapters III and IV of this dissertation.

6.2 Summary

In Chapter II, we reviewed and suggested the remedy for the problem of making inferences in the face of nuisance parameters from different populations by using the generalized p-value approach introduced by Tsui and Weerahandi (1987). This new development has a promising approach for data modeling in Computer Networking and Data Transmission, which have revolutionized modern society by their advanced technology. It also may be very useful for practitioners who have been performing inferences for small samples with the large sample approach for their research work. Network engineers encounter several streams of data generated by the *ON*-and *OFF*-periods and occasional larger file-size data with lower frequencies, which are exposed to a model that has a longer right tail. Inferences of functions of parameters of such heavy-tailed distributions are performed using this new model. In addition to Network engineering, this methodology is heavily used in fields of agriculture, mechanical engineering, insurance, banking, investment, and econometrics. This generalized p-value approach can easily be used to overcome the drawbacks of the F-test's failure to detect significant experimental results. Practitioners in biomedical research, where each sample point is vital and expensive, can comfortably use this generalized variable method to provide a significant test with the power of testing procedures.

In Chapter III, a combination of a two-parameter Pareto family of distribution

is presented as a mathematical vehicle for Data Transmission to perform inferences of the Offered Optical Network Unit Load (OOL). Generalized inferences for the common scale parameter as well as the common shape parameter of several Pareto populations are performed. The classical method based on the large sample approach is compared to the generalized variable method citing examples. Again, the other classical exact methods based on probability statements – such as the Fisher's, the Tippet's, the inverse normal, and logit methods – are compared to the generalized variable method citing previous research. The Bayesian estimation procedures for the common scale and the shape parameters are also discussed.

In Chapter IV, a combination of two-parameter exponential family of distribution is also presented as a mathematical vehicle for Data Transmission to perform inferences of the Offered Optical Network Unit Load (OOL). Generalized inferences for the common location parameter as well as the common shape parameter of several exponential populations are performed. Citing examples, the generalized variable method is compared to the classical method based on the large sample approach and to the other classical exact methods based on probability statements – such as the Fisher's, the Tippet's, the inverse normal, and the logit methods. Bayesian estimation procedures for the common scale and shape parameters are also discussed.

Chapter V investigates the statistical inferences of the Offered Optical Network Unit Load (OOL), which is one of the most important physical quantities found in Data Transmission. The newly introduced Passive Optical Networks (PON) have been considered as a solution for the subscriber access network. The OOL has been a direct result of the multiplexing of the long-range, self-similar data, or packets,

which have been model by several statistical distributions. In this Chapter V, several Pareto distribution and several exponential distributions are used to model such data. Statistical inferences of the OOL under the common scale as well the common shape parameter of several Pareto, and the common location and the shape parameter of several exponential populations are performed. These inferences are performed with respect the classical framework of inferences based on large sample approach, the generalized framework of inferences based on exact p-value approach, and the Bayesian framework of inferences based on prior densities. Complicated functions of parameters are not easily inferred exactly using the classical approach. Because of this reason, the emphasis has been placed on the importance of using the generalized variable method, which outperforms other available inferential methodologies in the face of nuisance parameters.

Finally, the following specific features give the importance of the inferential techniques presented in previous chapters:

1. inferences of common parameters of several Pareto populations,

2. inferences of common parameters of several exponential populations, and

3. applications of such densities and the performance of the inferences on complicated functions of parameters in Computer Networking and Data Transmission.

6.3 Future research

One of the major weaknesses or the drawbacks of the generalized variable method is its non-applicability when pivotal quantities are not distributed with standard

distributions. However, such situations are also resolved by using the intensive and tedious numerical approaches, which are to be explored as future research. Moreover, the power-guarantee has not been mathematically proved and is also a topic to be discussed. The advantages and drawbacks of the generalized variable method are, furthermore, summarized as follows.

The advantages of the proposed method are that it

1. can handle complicated functions of parameters,

2. involves with tests based on various distributions,

3. valid for smaller samples as well as for the larger samples,

4. can easily avoid the unnecessary large sample assumption, and

5. can find exact solutions in the face of nuisance parameters.

The drawbacks of the proposed procedure:

1. p-values are not uniformly distributed,

2. if the estimators are not distributed with distributions with closed forms, intensive numerical analysis has to be carried out, and

3. cannot be remedied all situations unless the test variable satisfy the properties of the generalized test variable.

Not only Offered Optical network Unit Load but also the other important physical quantities found in Computer Networking and Data Transmission such as the WWW-session interarrival time, the number of packets calls (pages) per WWW-session, the reading time between packet calls (WWW-pages), the number of items per WWW-page, the time intervals between items in the same WWW-page, the WWW-item size on uplink, the WWW-item size on Downlink, the time interval between two

137

consecutive Uplink packets inside an item, the time interval from Uplink to Downlink packet inside an item, the time interval between two consecutive Downlink packets inside an item, and the time interval from Downlink to Uplink packet inside an item of the World Wide Web traffic data; and the WAP-session interarrival time, the number of packets calls (pages) per WWW-session, the reading time between packet calls (WAP-items), the WAP-item sizes on uplink, the WAP-item sizes on Downlink, the transmission time of the Uplink packets (WAP-request, begin an item), the processing time of WAP-request, the WWW-transaction waiting time, the processing time of WAP-response, the transmission time of the Downlink packets (WAP-response), the acknowledgment time on Uplink, and the acknowledgment time on Downlink of the Wireless Application Protocol traffic data can be extensively and exclusively modeled by several Pareto and exponential distributions.

In addition to inferences on the Offered Optical Network Unit Load, the Offered Network Load, the Effective Optical Network Unit Load, and the Effective Network Load are performed. Furthermore, the Offered Optical Network Unit Load as well as the other important physical quantities found in Computer Networking and Data Transmission can be modeled using the other suitable distributions such as:

1. the log-normal distribution,

2. the log-logistic distribution,

3. the Weibull distribution,

4. the inverse Weibull distribution,

5. the Burr distribution, and

6. the lognormal-Pareto composite (LPC) distribution.

Their inferences are performed with this appealing generalized variable approach and can be compared with the other existing parametric estimation procedures.

Furthermore, rather than just analyzing the Offered Optical network Unit Load (OOL) at one time point, we can analyze it over time. Another analysis of OOL is to take randomized substreams rather than taking fixed number of them.

Applicability, accessibility, and usability of exact nonparametric procedures in Computer Networking and Data Transmission are also in consideration. New nonparametric approches for making inferences on the Offered Optical network Unit Load as well as other physical quantities found in Computer Networking and Data Transmission are also explored. To-be-new methods will be coupled with the old ones to develop a methodology to resolve the existing problems without taking the underlying distributions into account. Futhermore, applications of this generalized p-value methodology are sought not only in data traffic but also in other areas and fields such as reliability, survival analysis, econometrics, agriculture, actuary, insurance, banking, investment, etc.

.

APPENDIX A

INTERNET TRAFFIC TRACES

Traces available in the Internet Traffic Archive, a moderated repository to support widespread access to traces of Internet network traffic, sponsored by ACM SIG-COMM.

Here is a brief description of the traces currently in the archive. Following the links retrieves more information and a link for retrieving the trace.

* **BC** - 4 million-packet traces of LAN and WAN traffic seen on an Ethernet.

* **DEC-PKT** - 4 hour-long traces of all wide-area packets.

* **LBL-TCP-3** - 2 hours of wide-area TCP packets.

* **LBL-PKT** - 2 hour-long traces of all wide-area packets.

* **LBL-CONN**-7 - 30 days of wide-area TCP connections.

* **WorldCup98** - 1.3 billion Web requests recorded at servers for the 1998 World Cup.

* **EPA-HTTP** - a day of HTTP logs from a busy WWW server.

* **SDSC-HTTP** - a day of HTTP logs from a busy WWW server.

* **Calgary-HTTP** - a year of HTTP logs from a CS departmental WWW server.

* **ClarkNet-HTTP** - two weeks of HTTP logs from a busy Internet service

provider WWW server.

* **NASA-HTTP** - two months of HTTP logs from a busy WWW server.

* **Saskatchewan-HTTP** - seven months of HTTP logs from a University
WWW server.

* **BU-Web-Client** - Six months of Web client traces.

* **UC Berkeley Home IP Web Traces** - 18 days of HTTP traces.

* **Flattening Topology** - trace route measurements from 50 servers to the top
20 web servers circa 2007.

* **NPD-Routes** - Two datasets of repeated Internet route measurements.

following is one of the retrieved traces currently in the archive.

BC (BellCore -Bell Communications Research)

Description:

These four traces each contain a million packet arrivals seen on an Ethernet at the
Bellcore Morristown Research and Engineering facility. Two of the traces are LAN
traffic (with a small portion of transit WAN traffic), and two are WAN traffic.

Format:

The traces are in 2-column ASCII format, twenty bytes per line (including the
new line). The first column gives the time in seconds since the start of the trace. The
second column gives the Ethernet data length in bytes, not including the Ethernet
preamble, header, or CRC, though note that the Ethernet protocol forces all packets
to have at least a minimum size of 64 bytes and at most the maximum size of 1518
bytes.

Measurement:

The trace BC-pAug89 began at 11:25 on August 29, 1989, and ran for about 3142.82 seconds (until 1,000,000 packets had been captured). The trace BC-pOct89 began at 11:00 on October 5, 1989, and ran for about 1759.62 seconds. These two traces captured all Ethernet packets.

The trace BC-Oct89Ext began at 23:46 on October 3, 1989, and captured the first 1 million external arrivals (packets headed between Bellcore and the rest of the Internet), ending about 122797.83 seconds later. The trace BC-Oct89Ext4 comes from the 4th tape of a 307-hour trace begun at 14:37 on October 10, 1989. The tape started at timestamp 774018.987692, about 215 hours into the trace, and BC-Oct89Ext4 ends about 75943.08 seconds later.

All times are Eastern Daylight Time. The measurement hardware did not drop any packets, but corrupted packets (e.g., Ethernet collisions) are not included in the traces. 99.5% of the encapsulated packets were IP (Internet Protocol). The tracing was done at the Bellcore Morristown Research and Engineering facility, on the computing lab's Ethernet, which carried primarily local traffic, but also all traffic between Bellcore and the Internet. While timestamps are reported to 6 decimal places, they have 4-microsecond precision, and further analysis indicates that the actual accuracy is about 10 microseconds (primarily due to bus contention).

Here is a more detailed README (which refers to slightly different trace filenames).

Privacy:

The traces contain no packet contents and no host or protocol information.

Acknowledgements:

The traces were made by Will Leland (welbellcore.com) and Dan Wilson (dvwbellcore.com). In publications, please include appropriate citations to the papers mentioned below.

Publications:

The measurement techniques used in making the traces are described in High time-resolution measurement and analysis of LAN traffic: Implications for LAN interconnection, W. E. Leland and D. V. Wilson, Proc. IEEE INFOCOM '91, April 1991, pp. 1360-1366 (the Postscript for this paper has a missing figure).

These traces are a subset of those analyzed in Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management, H. J. Fowler and W. E. Leland, IEEE JSAC, 9(7), September 1991, pp. 1139-1149, and in On the Self-Similar Nature of Ethernet Traffic (Extended Version), W. Leland, M. Taqqu, W. Willinger, and D. Wilson, IEEE/ACM Transactions on Networking, 2(1), February 1994, pp. 1-15. The shorter, SIGCOMM '93 version of this last paper is available on-line.

Restrictions:

The traces may be freely redistributed.

Distribution:

Available from the Archive as BC-pAug89, compressed ASCII format (5 MB; 20 MB uncompressed); BC-pOct89, (5 MB; 20 MB uncompressed); BC-Oct89Ext, (6 MB; 20 MB uncompressed); and BC-Oct89Ext4 (6 MB; 20 MB uncompressed).

.

APPENDIX B

R CODES

Appendix B provides R codes that are related to the inferential techniques in real-data as well as simulated-data examples.

1. R codes related to inferences in several Pareto populations

Calculate p-values and confidence intervals

Number of generated Chi-squared values

k<-100 000

Number of Pareto populations

n<-3

Number of Simulations

N<-75 000

m1<-10

m2<-10

m3<-10

m<-c(m1,m2,m3)

pvaluegen<-0

pvaluecla<-0

SDthetahat<-0

```
Tobs<-0

P<-matrix(NA,m[1],n)

Rtheta<-matrix(NA,k,N)

constant<-rep(0.5,k)

alphahat<-matrix(NA,1,n)

W<-matrix(NA,k,n)

Rtheta<-0

theta_0<-100

theta<-100

gamma<-0.1

alpha1<-0.5

alpha2<-1

alpha3<-1.5

alpha<-c(alpha1,alpha2,alpha3)

countgen<-0

countcla<-0

Tobs1<-0

Tobs2<-0

P<-matrix(NA,m[1],n)

pvaluecla1<-0

pvaluegen1<-0

pvaluecla2<-0

pvaluegen2<-0
```

```r
# Number of generated Chi-squared values

for ( i in 1:n){

U<-runif(m[i],0,1)

P[,i]<-theta/(U)**(1/alpha[i])

alphahat[,i]<-m[i]*(sum(log(P[,i]/min(P[,i]))))**(-1)

W[,i]<-rchisq(k, 2*m[i]-2)

}

V<-rchisq(k, 2)

thetahat<-min(P)

for( t in 1:n)

{

Rtheta<-thetahat*exp(-constant*V/(sum(alphahat*W[t,])))

}

# Generalized p-value

numbergen<-sum(Rtheta < theta_0)

pvaluegen<-numbergen/k

pvaluegen

# Classical p-value

SDthetahat<-(sum(m*alphahat))*(thetahat)**2/((sum(m*alphahat)-1)**2*

(sum(m*alphahat)-2))

Tobs<-(thetahat- theta_0)/SDthetahat

pvaluecla<-1-pnorm( Tobs,mean =0, sd =1, lower.tail = TRUE,log.p=FALSE)

pvaluecla
```

```
# Exact (1-gamma)100% classical confidence interval for theta

x<-sort(Rtheta)

RgammaL<-x[k*10/100]

RgammaU<-x[k*90/100]

Rgammalower<-x[k*5/100]

Rgammaupper<-x[k*95/100]

RgammaL

RgammaU

Rgammalower

Rgammaupper

 # Calculate actual size of the test

for(j in 1:N)

{

for ( i in 1:n)

{

U<-runif(m[i],0,1)

P[,i]<-theta/(U)**(1/alpha[i])

alphahat[,i]<-m[i]*(sum(log(P[,i]/min(P[,i]))))**(-1)

W[,i]<-rchisq(k, 2*m[i]-2)

}

V<-rchisq(k, 2)

thetahat<-min(P)

for( t in 1:n){
```

```
Rtheta[,j]<-thetahat*exp(-constant*V/(sum(alphahat*W[t,])))

}

# Calculate generalized p-value

numbergen<-sum(Rtheta[,j] > theta)

pvaluegen[j]<-numbergen/k

# Calculate Classical p-value

SDthetahat[j]<-(sum(m*alphahat))*(thetahat)**2/((sum(m*alphahat)-1)**2

*(sum(m*alphahat)-2))

Tobs[j]<-(thetahat- theta)/SDthetahat[j]

pvaluecla[j]<-1-pnorm( Tobs[j],mean =0, sd =1, lower.tail = TRUE,

log.p=FALSE)

}

# Size of the test (classical)

sizenumbergen<-sum(pvaluegen < 0.1)

sizegen<-sizenumbergen/N

# Size of the test (generalized)

sizenumbercla<-sum(pvaluecla < 0.1)

sizecla<-sizenumbercla/N

sizegen

sizecla

 # Calculate coverage probabilities

for(j in 1:N)

{
```

```
for ( i in 1:n)

{

U<-runif(m[i],0,1)

P[,i]<-theta/(U)**(1/alpha[i])

alphahat[,i]<-m[i]*(sum(log(P[,i]/min(P[,i]))))**(-1)

W[,i]<-rchisq(k, 2*m[i]-2)

}

V<-rchisq(k, 2)

thetahat<-min(P)

for( t in 1:n)

{

Rtheta[,j]<-thetahat*exp(-V/(sum(alphahat*W[t,])))

}

# Approximate (1-gamma)100% classical confidence interval for theta

SDthetahat<-(sum(m*alphahat))*(thetahat)**2/((sum(m*alphahat)-1)**2

*(sum(m*alphahat)-2))

Zgammalower<-qnorm(gamma/2, mean = 0, sd = 1, lower.tail = TRUE, log.p

=FALSE)

ApConfidenceupper<-thetahat-Zgammalower*SDthetahat

ApConfidencelower<-thetahat+Zgammalower*SDthetahat

# Exact (1-gamma)100% generalized confidence interval for theta

x<-sort(Rtheta[,j])

Rgammalower<-x[k*5/100]
```

```
Rgammaupper<-x[k*95/100]

# Generalized coverage probabilities for theta

if (theta >= Rgammalower & theta <= Rgammaupper )

{

countgen<-countgen+1

}

# Generalized coverage probabilities for theta

if (theta >= ApConfidencelower & theta <= ApConfidenceupper )

{

countcla<-countcla+1

}

}

coveragegen<-countgen/N

coveragecla<-countcla/N

coveragegen

coveragecla

 # Calculate unadjusted power

for(j in 1:N)

{

for ( i in 1:n)

{

U<-runif(m[i],0,1)

P[,i]<-theta/(U)**(1/alpha[i])
```

```
alphahat[,i]<-m[i]*(sum(log(P[,i]/min(P[,i]))))**(-1)

W[,i]<-rchisq(k, 2*m[i]-2)

}

V<-rchisq(k, 2)

thetahat<-min(P)

for( t in 1:n)

{

Rtheta[,j]<-thetahat*exp(-constant*V/(sum(alphahat*W[t,])))

}

# Calculate generalized p-value

numbergen<-sum(Rtheta[,j] > theta_0)

pvaluegen[j]<-numbergen/k

# Calculate classical p-value

SDthetahat[j]<-(sum(m*alphahat))*(thetahat)**2/((sum(m*alphahat)-1)**2

*(sum(m*alphahat)-2))

Tobs[j]<-(thetahat- theta_0)/SDthetahat[j]

pvaluecla[j]<-1-pnorm( Tobs[j],mean =0, sd =1, lower.tail = TRUE, log.p

= FALSE)

}

# Size of the test (classical)

sizenumbergen<-sum(pvaluegen < 0.1)

powerunadjgen<-sizenumbergen/N

# Size of the test (generalized)
```

sizenumbercla<-sum(pvaluecla < 0.1)

powerunadjcla<-sizenumbercla/N

powerunadjgen

powerunadjcla

 # Calculated adjusted power

for(j in 1:N)

{

for (i in 1:n)

{

U<-runif(m[i],0,1)

P[,i]<-theta/(U)**(1/alpha[i])

alphahat[,i]<-m[i]*(sum(log(P[,i]/min(P[,i]))))**(-1)

W[,i]<-rchisq(k, 2*m[i]-2)

}

V<-rchisq(k, 2)

thetahat<-min(P)

for(t in 1:n)

{

Rtheta[,j]<-thetahat*exp(-constant*V/(sum(alphahat*W[t,])))

}

Calculate generalized p-value

numbergen1<-sum(Rtheta[,j] > theta)

pvaluegen1[j]<-numbergen1/k

```
# Calculate Classical p-value

SDthetahat[j]<-(sum(m*alphahat))*(thetahat)**2/((sum(m*alphahat)-1)**2*

(sum(m*alphahat)-2))

Tobs1[j]<-(thetahat- theta)/SDthetahat[j]

pvaluecla1[j]<-1-pnorm( Tobs1[j],mean =0, sd =1, lower.tail = TRUE, log.p

= FALSE)

}

xgen<-sort(pvaluegen1)

pgen<-xgen[k*10/100]

xcla<-sort(pvaluecla1)

pcla<-xcla[k*10/100]

for(j in 1:N)

{

# Calculate generalized p-value

numbergen2<-sum(Rtheta[,j] > theta_0)

pvaluegen2[j]<-numbergen2/k

# Calculate Classical p-value

SDthetahat[j]<-(sum(m*alphahat))*(thetahat)**2/((sum(m*alphahat)-1)**2*

(sum(m*alphahat)-2))

Tobs2[j]<-(thetahat- theta_0)/SDthetahat[j]

pvaluecla2[j]<-1-pnorm( Tobs2[j],mean =0, sd =1, lower.tail = TRUE, log.p

= FALSE)

}
```

Size of the test (classical)

sizenumbergen<-sum(pvaluegen2 < pgen)

poweradjgen<-sizenumbergen/N

Size of the test (generalized)

sizenumbercla<-sum(pvaluecla2 < pcla)

poweradjcla<-sizenumbercla/N

poweradjgen

poweradjcla

2. R codes related to inferences in several exponential populations

Number of Generated Chi-squared values

k<-1000

Number of pareto Populations

n<-2

Number of Simulations

N<-500

m1<-10

m2<-15

m<-c(m1,m2)

pvaluegen<-0

pvaluecla<-0

thetahat<-0

SDthetahat<-0

Tobs<-0

```
P<-matrix(NA,m[1],n)

W<-matrix(NA,k,n)

Rtheta<-matrix(NA,k,N)

alphahat<-matrix(NA,1,n)

Zgammalower<-0

ExConfidenceupper<-0

ExConfidencelower<-0

ApConfidenceupper<-0

ApConfidencelower<-0

gamma<-0.05

theta<-100

alpha1<-1

alpha2<-0.5

alpha<-c(alpha1,alpha2)

for(j in 1:N){

for ( i in 1:n){

U<-runif(m[i],0,1)

P[,i]<-theta + alpha[i]*log(1/U)

alphahat[,i]<-m[i]**(-1)*(sum(P[,i]-min(P[,i])))

W[,i]<-rchisq(k, 2*m[i]-2)

}

V<-rchisq(k, 2)

thetahat[j]<-min(P)
```

```r
for ( t in 1:k){

Rtheta[,j]<-thetahat[j]-V*(sum(alphahat/W[t,]))

}
# Calculate Generalized p-value

numbergen<-sum(Rtheta[,j] > theta)

pvaluegen[j]<-numbergen/k

# Calculate Classical p-value

SDthetahat[j]<-sum(m**(-1)*alphahat)

Tobs[j]<-(thetahat[j]- theta)/SDthetahat[j]

pvaluecla[j]<-1-pnorm( Tobs[j],mean =0, sd =1, lower.tail = TRUE, log.p =
FALSE)

# Approximate (1-gamma)100% Classical confidence Interval for theta

Zgammalower[j]<-qnorm(gamma, mean = 0, sd = 1, lower.tail = TRUE, log.p =
FALSE)

ApConfidenceupper[j]<-thetahat-Zgammalower[j]*SDthetahat[j]

ApConfidencelower[j]<-thetahat+Zgammalower[j]*SDthetahat[j]

}
pvaluecla[1]

pvaluegen[1]

# Size of the Test (Classical)

sizenumbergen<-sum(pvaluegen > 0.1)

sizegen<-sizenumbergen/N

# Size of the Test (Generalized)
```

156

sizenumbercla<-sum(pvaluecla > 0.1)

sizecla<-sizenumbercla/N

sizegen

sizecla

Confidence Interval

x<-sort(Rtheta)

x<-sort(Rtheta)

RgammaL<-x[k*N*10/100]

RgammaU<-x[k*N*90/100]

Rgammalower<-x[k*N*5/100]

Rgammaupper<-x[k*N*95/100]

length<-Rgammaupper - Rgammalower

length

3. R codes related to inferences in Computer networking and data

  Transmission

rON<-5000

rOFF<-5000

r<-5000

N<-1

n<-2

gamma<-0.05

psi_0<-1

#Read in Pareto data for the first substream from external files

```
Pdata11<-read.table("C:/Users/Sumith Gunasekera/Desktop/data11.txt",
header=FALSE)
Pdata1ON<-t(Pdata11)[c(1:10)]
Pdata12<-read.table("C:/Users/Sumith Gunasekera/Desktop/data12.txt",
header=FALSE)
Pdata1OFF<-t(Pdata12)[c(1:10)]
Pdata21<-read.table("C:/Users/Sumith Gunasekera/Desktop/data21.txt",
header=FALSE)
Pdata2ON<-t(Pdata21)[c(1:10)]
Pdata22<-read.table("C:/Users/Sumith Gunasekera/Desktop/data22.txt",
header=FALSE)
Pdata2OFF<-t(Pdata22)[c(1:10)]
m11<-length(t(Pdata1ON))
m12<-length(t(Pdata2ON))
m21<-length(t(Pdata1OFF))
m22<-length(t(Pdata2OFF))
mON<-c(m11,m21)
mOFF<-c(m12,m22)
PON<-matrix(NA,mON,n)
POFF<-matrix(NA,mOFF,n)
PON<-cbind(Pdata1ON,Pdata2ON)
POFF<-cbind(Pdata1OFF,Pdata2OFF)
pvaluegen<-0
```

```
pvaluecla<-0

RthetaON<-matrix(NA,r,n)

RthetaOFF<-matrix(NA,r,n)

psihat<-0

SDpsihat<-0

Tobs<-0

constant<-rep(0.5,r)

WON<-matrix(NA,rON,n)

WOFF<-matrix(NA,rOFF,n)

thetaONhat<-0

thetaOFFhat<-0

RthetaON<-matrix(NA,r,n)

RthetaOFF<-matrix(NA,r,n)

alphaONhat<-matrix(NA,1,n)

alphaOFFhat<-matrix(NA,1,n)

RalphaON<-matrix(NA,r,n)

RalphaOFF<-matrix(NA,r,n)

Rpsi<-matrix(NA,r,N)

for ( j in 1: N){

for ( i in 1:n){

alphaONhat[,i]<-mON[i]*(sum(log(PON[,i]/min(PON[,i]))))**(-1)

alphaOFFhat[,i]<-mOFF[i]*(sum(log(POFF[,i]/min(POFF[,i]))))**(-1)

WON[,i]<-rchisq(rON, 2*mON[i]-2)
```

WOFF[,i]<-rchisq(rOFF, 2*mOFF[i]-2)

RalphaON[,i]<-constant*WON[,i]*alphaONhat[,i]/mON[i]

RalphaOFF[,i]<-constant*WOFF[,i]*alphaOFFhat[,i]/mOFF[i]

}

VON<-rchisq(rON, 2)

VOFF<-rchisq(rOFF, 2)

thetaONhat<-min(PON)

thetaOFFhat<-min(POFF)

RthetaON1_2<-thetaONhat*exp(-constant*VON/(sum(RalphaON*mON)))

RthetaOFF1_2<-thetaOFFhat*exp(-constant*VOFF/(sum(RalphaOFF*m

OFF)))

RthetaON<-cbind(RthetaON1_2, RthetaON1_2)

RthetaOFF<-cbind(RthetaOFF1_2, RthetaOFF1_2)

alpha11hat<-(length(t(Pdata1ON)))*(sum(log(Pdata1ON))/min(Pdata1ON))**

(-1)

alpha12hat<-(length(t(Pdata1OFF)))*(sum(log(Pdata1OFF))/min(Pdata1OFF)

)**(-1)

alpha21hat<-(length(t(Pdata2ON)))*(sum(log(Pdata2ON))/min(Pdata2ON))

**(-1)

alpha22hat<-(length(t(Pdata2OFF)))*(sum(log(Pdata2OFF))/min(Pdata2OFF)

)**(-1)

psihat[j]<-abs(sum(((alphaOFFhat-1)*(alphaONhat)*(thetaONhat))/((alphaON

hat*alphaOFFhat)

```
*(thetaONhat+thetaOFFhat)-(alphaONhat*thetaONhat+alphaOFFhat*theta

OFFhat))))

for ( t in 1:r){

Rpsi[t,j]<-abs(sum((RalphaOFF[t,]-1)*RalphaON[t,]*RthetaON[t,]*(RalphaOFF

[t,]*RalphaON[t,]*(RthetaON[t,]+RthetaOFF[t,])-(RalphaON[t,]*RthetaON[t,]+

RalphaOFF[t,]*RthetaOFF[t,]))**(-1)))

}

numbergen<-sum(Rpsi[,j] > psi_0)

pvaluegen[j]<-numbergen/r

SDpsihat[j] <-sqrt(abs(((alpha21hat-1)**2*alpha21hat**2*alpha11hat**2)/((2*

alpha11hat*alpha21hat-alpha11hat-alpha21hat)**4*m11*(1-alpha11hat)) +

((alpha22hat-1)**2*alpha22hat**2*alpha12hat**2)/((2*alpha12hat*alpha22hat

-alpha12hat-alpha22hat)**4*m12*(1-alpha12hat)) +

((alpha11hat-1)**2*alpha11hat**2*alpha21hat**2)/((2*alpha11hat*alpha21hat

-alpha11hat-alpha21hat)**4*m21*(1-alpha21hat)) +

((alpha12hat-1)**2*alpha12hat**2*alpha22hat**2)/((2*alpha12hat*alpha22hat

-alpha12hat-alpha22hat)**4*m22*(1-alpha22hat))))

Tobs[j]<-(psihat[j] - psi_0)/SDpsihat[j]

pvaluecla[j]<-pnorm( Tobs[j],mean =0, sd =1, lower.tail = TRUE, log.p = FALSE)

}

pvaluegen

pvaluecla

x<-sort(Rpsi)
```

RgammaL<-x[r*N*gamma]

RgammaU<-x[r*N*(1-gamma)]

Rgammalower<-x[r*N*gamma/2]

Rgammaupper<-x[r*N*(1-gamma/2)]

.

REFERENCES

Akahira M. (1997). Second Order Asymptotic Comparison of Estimators of a Common Parameter in the Double Exponential Case. *Annals of the Institute of Statistical Mathematics* **39**, 25-36.

Ananda, M.M.A. (1995). Generalized F-tests for unbalanced nested designs under heteroscedasticity. *Annals of Institute of Statistical Mathematics* **47**(4) 731-742.

Ananda, M.M.A. and Weerahandi, S. (1997). Two-Way ANOVA with Unequal Cell Frequencies and Unequal variances. *Statistica Sinica* **7** 631-646.

Ananda, M. M. A. (1998). Bayesian and non-Bayesian solutions to analysis of covariance models under heteroscedasticity. *Journal of Econometrics* **86**, 177-192.

Ananda, M. M. A. (1999). Estimation and testing of availability of a parallel system with exponential failure and repair times. *Journal of Statistical Planning and Inference* **77,** 237-246.

Arnold, B. C. (1983). *Pareto Distributions.* International Co-operative Publishing House, Fairland, Maryland, USA.

Bai, D. S. and Hong, Y. W. (1992). Estimation of $Pr(X < Y)$ in the Exponential Case with Common Location Parameter. *Communication in Statistics–Theory and Methods* **21**(1), 269-282.

Baklizi, A. S. (2002). Confidence Lower Bounds for the Common Truncation Parameter of Several Pareto Distributions using Combined Independent Tests. *Far*

East Journal of Theoretical Statistics **8**, 115-121.

Baklizi, A. S. (2004). Shrinkage Estimation of the Common Location Parameter of Several Exponentials. *Communications in Statistics–Simulation and Computation* **33**(2), 321-339.

Beirlant, J., Joossens, E. and Segers, J. (2004). Generalized Pareto fit to the Society of Actuaries' large claims database. *North American Actuarial Journal* **8**, 108-111.

Bernard, G. A. (1984). Comparing the Means of two Independent samples. *Applied Statistics* **33**, 215-232.

Brabson, B. B. and Palutikof, J. P. (2000). Tests of the Generalized Pareto Distribution for Predicting Extreme Wind Speeds. *Journal of Applied Meteorology* **39**(9), 1627–1640.

Brazauskas, V. and Serfling, R. (2003). Favorable Estimators for Fitting Pareto Models: A study using Goodness-of-Fit Measures with Actual Data. *Asian Bulletin* **33**(2), 365-381.

Carpenter, M. and Pal, N. (1995). Estimation of the Smaller and Larger of the of Two Exponential Location Parameters with a Common Unknown Scale Parameter. (1995). *Calcutta Statistical Association Bulletin* **45**(177 & 178) 61-71.

Chi, E. M. and Weerahandi, S. (1998). Comparing treatments under growth curve models with compound-symmetric covariance structure: exact tests using generalized *p*-values. *Journal of Statistical Planning and Inference* **71**, 179-189.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician* **49**, 327–336.

Chiou, W. and Cohen, A. (1984). Estimating the Common Location Parameter of Exponential Distributions with Censored Samples. *Navel Research Logistics Quarterly* **31**, 475-482.

Cooray K. and Ananda M. M. A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal* **105**, 321-34.

Costanza, M. C., Hamdy, H. I., and Son, M. S. (1986). Two Stage Fixed Width Confidence Intervals for the Common Location Parameter of Several Exponential Distributions. *Communication In Statistics–Theory and Methods* **15**(8), 2305-2322.

Crovella, M. and Bestavros, A. (1996). Self-Similarity in the World Wide Web traffic: evidence and causes. *IEEE/ACM Transactions on Networking* **5**, 835-846.

Deaton, A. (1997). *Analysis of Household Surveys.* Johns Hopkins University Press, Baltimore, Maryland, USA.

Downey, A. B. (2005). Lognormal and Pareto Distributions in the Internet. *Computer Communications* **28**(7), 790-801.

Elfessi, A. and Jin, C. (1996). On Robust Estimation of the Common Scale Parameter of Several Pareto Distributions. *Statistics and Probability Letters* **29**, 345-352.

Everitt, B. S. and Hand, D. J. (1981). *Finite Mixture Distributions.* Chapman & Hall, Florida, USA.

Fisher, R. A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society* **85**(1), 87-94.

Fisher, R. A. (1954). *Statistical Methods for Research Workers.* Oliver and Boyd.

Gamage, J. and Weerahandi, S. (1998). Size performance of some tests in one-way

ANOVA. *Communications in Statistics: Simulation* **27**, 625-640.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990). Illustration of Bayesian Inference in Normal Data Models using Gibbs Sampling. *Journal of the American Statistical Association* **85**, 398-409.

Ghosh, M. and Razmpour, A. (1984). Estimation of the Common Location Parameter of Several Exponentials. *Sankhyā: The Indian Journal of Statistics* **46**, 383-394.

Green, J. (1980). Discussion on D. V. Lindley's (1980) paper on approximate Bayesian method. *Trabajos Estadistica* **31**, 241.

Gregurich, M. A. and Broemeling, L. D. (1997). A Bayesian Analysis for the Common Mean of Independent Normal Populations using Gibbs Sampler. *Communications in Statistics: Simulation* **26**(1), 35-51.

Gunasekera, S. and Ananda, M. M. A. (2008). Generalized Inferences for the Common Scale Parameter of Several Pareto Populations. Submitted for the publication.

Gunasekera, S. and Ananda, M. M. A. (2009). Unbalanced two-way ANOVA additive model under heteroscedasticity using generalized *p*-values. *Advances and Applications in Statistics* **11**(1), 1-13.

Handa, B. R. and Kambo, N. S. (1999). Estimation of Common Parameter of Several Exponential Distributions. *Calcutta Statistical Association Bulletin* **49**, 199-196.

Handa, B. R. and Kambo, N. S. (2005). Test of Equality of Parameters of Two Exponential Distributions with Common Known Coefficient of Variation. *Communi-*

cations in Statistics–Theory and Methods **34**, 2147-2155.

Hauzner, S. (2008). *Data Traffic Model for Wireless Coexistence in US 3.65GHz band Simulation*-Project*:* IEEE P802.19 Coexistence TAG, Tel-Aviv, Israel.

Heyman, D. P. (1982). An analysis of the carrier-sence multiple-access protocol. *Bell System Technical Journal* **61**(8), 2023-2051.

Hastings, W. K. (1979). Monte Carlo Sampling Method using Markov Chains and their Applications. *Biometrika* **57**, 97-109.

Hogg, R. V. and Klugman, S. A. (1984). *Loss Distributions.* John Wiley & Sons, New jersey, USA.

Houghton, J. C. (1988). Use of the truncated shifted Pareto distribution in assessing size distribution of oil and gas fields. *Mathematical Geology* **20**(8), 907-937.

Jain, R. and Routhier, S. A. (1986). Packet Trains – Measurements and a New Model for Computer Network Traffic. *IEEE Journal of Selected Areas in Telecommunications* **4**(6), 986-995.

Jin, C. and Crouse, R. H. (1998). An Identity for Exponential Distributions with the Common Location Parameter and its Applications. *Communications in Statistics: Theory and Methods* **27**(2), 409-422.

Jin, C. and Crouse, R. H. (1998). A Note on the Common Location Parameter of Several Exponential Populations. *Communications in Statistics–Theory and Methods* **27**, 2777-2789.

Jin, C. and Elfessi, A. (2001). On the Common Scale Parameter of Several Pareto Populations in Censored Samples. *Communications in Statistics–Theory and Methods* **30**(3), 451-462.

Johnson, N. L., Kotz, S., and Balakrishnan. (1994). *Continuous Univariate Distributions.* Vol 1. New York: Wiley.

Kanfer, F. H. J. and Geertsema, J. C. (2001). Robustness Properties of a Sequential Confidence Interval for the Common Location Parameter of a number of Exponential Populations. *Sequential Analysis* **20**(4), 293-323.

Kempthorne, O. and Folks, L. (1971). *Probability, Statistics, and Data Analysis.* Iowa State University Press, Ames.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (1998). *Loss Models From Data to Decisions.* John Wiley & Sons.

Koschat, M. A. and Weerahandi, S. (1992). Chow-type tests under heteroscedasticity. *Journal of Business and Economic Statistics* **10**, 221-228.

Kramer, G. (2001) *Self-similar Network Traffic.* University of California-Davis.

Kramer, G. (2003) EPON: *Challenges in Building Next Generation Access Network.* Workshop on Community Networks and FTTH/P/X. Teknovus, Inc.

Kramer, G. (2005) *Ethernet Passive Optical Networks.* McGraw-Hill, New York.

Kramer, G. (2008) The File size distribution, Glen Kramer Homepage at http://wwwcsif.cs.ucdavis.edu/~kramer/

Kramer, G., Mukherjee, B., and Pesavento, G. " Interleaved Polling with Adaptive Cycle Time (IPACT): A Dynamic bandwidth Distribution Scheme in an Optical Access Network." July 2001.

Krishnamoorthy, K. and Lu, Y. (2003). Inference on the common means of several normal populations based on the generalized variable method. *Biometrics* **59**, 237-247.

168

Leland, W., Taqqu, M., Willinger, W., and Wilson, D. (1994). On the Self-Similar Nature of Ethernet Traffic (Extended Version). *IEEE/ACM Transactions on Networking* **2**(1), 1-15.

Lin, Y. and Shen, W. (1997). Comparison of Power of Tests for a Common Location of k Truncated Exponentials. *Journal of Applied Statistical Science* **06**(4), 265-273.

Lindley, D. V. (1980). Bayesian approximation methods. *Trabajos Estadistica* **31**, 223-237.

Madi, M. and Tsui, K. (1990). Estimation of the Common Scale of Several Shifted Exponential Distributions with Unknown Locations. *Communications in Statistics–Theory and Methods* **19**(6), 2295-2313.

Malik, H. J. (1970). Estimation of the Parameters of the Pareto Distribution. *Metrika* **15**, 126-132.

Manna, D. K. (2000). Common Due-Date Assignment and Scheduling on Single Machine with Exponential Processing Times. *OPSEARCH* **37**, 221-236.

Marathe, M. and Hawe, W. (1982). Predicted capacity of Ethernet in university enviornment. in *Proceedings of the Southcon '82,* pp 1-10.

Mehta, C. R. (1995). SPSS 6.1 *Exact test for Windows.* Englewood Cliffs, Prentice Hall, New Jersey, USA.

Mehta, C. R. and Patel, N. R. (1997). Exact inference in categorical data. *Biometrics* **53**(1), 237-247.

Meister, B. (1980). Waiting time in a preemptive resume system with compound-Poisson input. *Computing* **25**(1), 17-28.

Mendelbrot, B. B. (1969). Long-run linearity, locally Gaussian processes, H-spectra and infinite variances. *International Economic Review* **10**, 82-113.

Meng, X. (1994). Posterior Predictive *p*-values. *Annals of Statistics* **22**, 1142-1160.

Metropolis, N., Rosenbluth, A. W., Roenbluth, M., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21**, 1087-1092.

Mohanty, N. C. (1978). Buffer behavior for batch arrivals and single output in satellite channel. In *Proceedings of the IEEE 1978 Nattional Telecommunication Confonference NTC' 78,* Birmingham, AL, pt III, pp 40.4/1-40.4/3.

Moothathu, T. S. K. (1990). A Characterization Property of Weibull, Exponential, and Pareto Distributions. *Journal of the Indian Statistical Association* **28**, 69-74.

Moothathu, T. S. K. (1993). A Characterization of the α-Mixture Pareto Distribution through a Property of Lorenz Curve. *Sankhyā: The Indian Journal of Statistics* **55**, Series B, Pt. 1, 130-134.

Pal, N. and Sinha, B. K. (1990). Estimation of a Common Location of Several Exponentials. *Statistics and Decisions* **8**, 27-36.

Pareto V. (1897). *Curs d'Economie Politique*, Paris: Rouge et Cie.

Park, D. J. and Burdick, R. K. (2003). Performance of Confidence Intervals in Regression Models with Unbalanced One-Fold Nested Error Structures. *Communications in Statistics - Simulation and Computation* **32**(3), 717-732.

Paxson, V. and Floyd, S. (1995). Wide-Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking* **3**(3), 226-244.

Quant, R. E. (1966). Old and new methods of estimation and the Pareto distribution. *Metrika* **10**, 55-82.

Ramakrishnan, K. K. and Tripathi, S. K. (1982). A common framework for studying the performance of channel access protocols. in *Proc. Comput. Performance Evaluation Users Group (CPEUG '82),* pp 365-373.

Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin* **27**, 139-51.

Rice, W. R. and Gaines, S. D. (1989). One-way Analysis of Variance with Unequal Variances. In *Proceedings of the National Academy of Sciences.* **86**, 8183-8184.

Rytgaard, M. (1990). Estimation in the Pareto Distribution. *Asian Bulletin* **20**(2), 201-216.

Sala, D. and Gummalla, A. (2001) *PON Functional Requirements: Services and Performance* presented at IEEE 802.3ah meeting in Portland, Oregon.

Shin-Hui, L. and Jack, L. C. (2005). Generalized inferences on the common mean of several normal populations. *Journal of Statistical Planning and Inference* **134**, 568-582.

Sinha, S. K. and Kale, B. K. (1980). *Life Testing and Reliability Estimation.* New York: Halsted Press, A Division of John Wiley and Sons, Inc.

Sinha, S. K. (1986). *Reliability and Life Testing.* New York: Wiley Eastern Ltd./Halstead Press.Sinha, S. K. (1986). *Reliability and Life Testing.* New York: Wiley Eastern Ltd./Halstead Press.

Sinha, S. K. and Sloan, J. A. (1988). Bayes Estimation of the Parameters and Reliability Function of the 3-parameter Weibull Distribution. *BIEE- Transactions on*

Reliability **37**(4), 364-369.

Sun, Y. and Sinha, K. (1999). Exact confidence intervals for a common location of several truncated exponentials. *Communications in Statistics- Theory and methods* **28**, 683-692.

Taqqu, M. S. and Levy, J. (1986). Using Renewal Processes to Generate Long-Range Dependent and High variability. In Eberlein, E. and Taqqu, M. S. editors, *Dependence in Probability and Statistics,* pp 73-89. Birkhäuser, Boston.

Taqqu, M. S., Willinger, W., and Sherman, R. (1997). Proof of a Fundamental Result in Self-Similar Traffic Modeling. *Computer Communication Review* **27**, 5-23.

Tian, L. and Wu, J. (2007). Inferences on the Common Mean of Several Log-normal Populations: The generalized Variable Approach. *Biometrical Journal* **49**, 944-951.

Thomas, Y. and Kumar, N. K. S. (2005). Estimating the Common Parameter of Several Exponential Distributions. *IAPQR Transactions* **30**(2), 125-134.

Tobagi, F. A. and Hunt, V. B. (1979). Performance analysis of carrier sence multiple access with collision detection. In *Proceedings of the LACN Symposium,* pp 217-244.

Tsui, K. and Weerahandi, S. (1989). Generalized *p*-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association* **84**, 602-607.

Weerahandi, S. (1987). Testing regression equality with unequal variances. *Econometrica* **55**, 1211-1215.

Weerahandi, S. (1991). Testing variance components in mixed models with gen-

eralized *p*-values. *Journal of the American Statistical Association* **86**, 151-153.

Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association* **88,** 899-905, correction in **89**, 726.

Weerahandi, S. (1995). ANOVA under unequal variances. *Biometrics* **51,** 589-599.

Weerahandi, S. (1995). *Exact Statistical Methods for Data Analysis.* Springer - Verlag, New York.

Weerahandi, S. and Tsui, K. (1996). Solving ANOVA problems by Bayesian Approach, Comment on "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies" by Gelman, Meng, and Stern. *Statistica Sinica* **6,** 792-796.

Weerahandi, S. and Amaratunga, D. (1999). A Performance Comparison of Methods in Mixed Models. Technical Report, Bell Communication Research (BellCore).

Weerahandi, S. and Berger, V. W. (1999). Exact Inference for Growth Curves with Intraclass Correlation Structure. *Biometrics* **55,** 921-924.

Weerahandi, S. and Peterson, J, J. (2003). Generalized Inference in Balaam Design. Technical Report.

Weerahandi, S. (2004). *Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models.* John Wiley, New Jersey.

Willinger, W., Taqqu, M. S., Sherman, R., and Wilson, D. V. (1997). Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Transactions on Networking* **5**(1), 71-86.

Zhou, L. and Mathew, T. (1994). Some Tests for Variance Components using generalized *p*-values. *Technometrics* **36,** 394-421.

.

VITA

Graduate College
University of Nevada, Las Vegas

Sumith Gunasekera

Local Address:
    4255 South Pecos Road #107
    Las Vegas, Nevada 89121- 5049, USA

Home Address:
    53/2 Kossinnawatte
    Katuwawala
    Boralesgamuwa
    Sri Lanka

Degrees:
    Bachelor of Science, Physics and Mathematics, 1995
    University of Colombo, Colombo 3, Sri Lanka

Special Achievements:
    The Doctoral Student with the highest overall GPA of 3.91 in the Department
    of Mathematical Sciences, College of Sciences, University of Nevada-Las Vegas,
    USA.

    Nominated as a biographical candidate to represent Las Vegas, NV, USA in the
    2008/2009 online Cambridge Who's Who Registry Among Executives and
    Professionals.

Publications:
    Cooray, K., Gunasekera, S., and Ananda, M. M. A. (2006). The Folded Logistic
    Distribution. *Communications in Statistics-Theory and Methods* **35**(3), 385-393.

    Gunasekera, S. and Ananda, M. M. A. (2009). Unbalanced two-way ANOVA
    additive model under heteroscedasticity using generalized p-values. *Advances*
    and Applications in Statistics **11**(1), 1-13.

Gunasekera, S. and Ananda, M. M. A. (2008). Generalized Inferences for the Common Scale Parameter of Several Pareto Populations. Submitted for publication in *International Journal of Statistics and Management System*.

Gunasekera, S. and Ananda M. M. A. (2008). Generalized Inferences for the Common Location Parameter of Several Exponential Populations. Submitted for publication in *Metrika*.

Gunasekera, S. and Ananda M. M. A. (2008). Statistical Inferences for the Offered Optical Network Unit Load of Data Transmitted via Ethernet Passive Optical Networks. Submitted for publication in *IEEE/ACM - Transactions on Networking*.

Cooray, K., Gunasekera, S. and Ananda, M. M. A. (2005). Modeling Actuarial Data with Pareto Composite Families. Submitted for publication in *American Journal of Mathematical and Management Sciences*.

Cooray, K., Gunasekera, S., and Ananda M. M. A. (2008). The Weibull-Inverse Weibull Composite Distribution for Modeling Reliability Data. Submitted for publication in *Model Assisted Statistics and Applications*.

Dissertation Title: Statistical Inferences for Functions of Parameters of Several Pareto and Exponential Populations with the Applications in Data Traffic

Dissertation Examination Committee:
 Chairperson, Dr. Malwane M. A. Ananda, Ph. D.
 Committee Member, Dr. Chih-Hsiang Ho, Ph. D.
 Committee Member, Dr. Hokwon Cho, Ph. D.
 Committee Member, Dr. Sandra Catlin, Ph. D.
 Graduate Faculty Representative, Dr. Chad Cross, Ph. D.