PROBABILISTIC PROPERTIES AND STATISTICAL INFERENCE FOR A FAMILY OF GENERALISED AND RELATED DISTRIBUTIONS

LIEW KIAN WAH

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

INSTITUTE OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2013

ACKNOWLEDGEMENT

First and foremost, I would like to express my highest gratitude to my thesis supervisor, Professor Dr. Ong Seng Huat, for his unstinting support throughout the process of completing my thesis as well as his immense patience. Besides taking out challenging problems from his treasure trove for me to solve, he also granted me the freedom to pursue my own projects and for this I am truly grateful. I would also like to thank him for providing the rare opportunities for me to interact with some of the distinguished scholars in the field, these were invaluable experiences.

Special thanks to the thesis examiners who have gone through the thesis meticulously and given constructive suggestions for improvements.

I would also like to thank my friends at Universiti Tunku Abdul Rahman and the University of Malaya for their continuous support and encouragements, especially Dr. Goh Yong Kheng and Dr. Ng Choung Min who are always there to lend me a helping hand. I am lucky to have met many students who have made my teaching life interesting and meaningful. Seeing the potential in the younger generation is always a big motivation for me to strive for better.

I will forever be thankful to my friends who shared the laughter and tears, fun and pain, food and drinks with me and showing me the silver lining during my cloudy days.

Last but not least, to my parents who were wondering what I did in my room all these years, thanks for their unconditional love and care.

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **LIEW KIAN WAH** (I.C/Passport No: **740822-14-5561**)

Registration/Matric No: **SHB080001**

Name of Degree: **DOCTOR OF PHILOSOPHY**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

PROBABILISTIC PROPERTIES AND STATISTICAL INFERENCE FOR A FAMILY OF GENERALISED AND RELATED DISTRIBUTIONS

Field of Study: **STATISTICS**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness's Signature Date

Name:
Designation:

ABSTRACT

Three generalised distributions are studied in this thesis from different aspects. The Hurwitz-Lerch zeta distribution (HLZD) that generalises the logarithmic distribution and a class of distributions that follows the power law is considered. To investigate the effects of parameters on the stochastic properties of the HLZD, stochastic orders between members in this large family are established. A relationship between the tail behaviours of the HLZD and that of a class of generalised logarithmic distribution is highlighted. The HLZD has shown good flexibilities in empirical modelling. A robust probability generating function based estimation method using Hellinger-type divergence is implemented in data-fitting and the results are compared with various other generalisations of logarithmic distribution. An augmented probability generating function is constructed to overcome the difficulties of this estimation procedure when some data are grouped. The Poisson-stopped sum of the Hurwitz-Lerch zeta distribution (Poisson-HLZD) is then proposed as a new generalisation of the negative binomial distribution. Several methods have been used in deriving the probability mass function for this new distribution to show the connections among different approaches from mathematics, statistics and actuarial science. Basic statistical measures and probabilistic properties of the Poisson-HLZD are examined and the usefulness of the model is demonstrated through examples of data-fitting on some real life datasets. Finally, the inverse trinomial distribution (ITD) is reviewed. Both Poisson-HLZD and ITD are proved to have mixed Poisson formulation, which extend the applications of the models for various phenomena. The associated mixing distribution for the ITD is obtained as an infinite Laguerre series and the result is compared to some numerical inversions of Laplace transform.

ABSTRAK

Tiga taburan teritlak telah dikaji dalam tesis ini dari pelbagai aspek. Keluarga Hurwitz-Lerch zeta (HLZ) yang mengandungi taburan logaritma dan suatu kelas taburan yang mematuhi hukum kuasa telah dipertimbangkan. Penertiban stokastik bagi pembolehubah-pembolehubah rawak dalam keluarga HLZ telah diterbit untuk mengkaji kesan-kesan parameter terhadap sifat-sifat taburan. Hubungan antara kebarangkalian ekor bagi taburan HLZ dan suatu taburan logaritma teritlak juga ditunjukkan. Demi menggambarkan kelenturan taburan HLZ, satu kaedah penganggaran teguh yang menggunakan divergen jenis Hellinger berasaskan fungsi penjana kebarangkalian telah dilaksanakan dan keputusannya dibanding dengan pelbagai pengitlakan taburan logaritma yang lain. Pembinaan fungsi penjana kebarangkalian imbuhan telah dicadang untuk menyelesaikan masalah pelaksanaan kaedah ini apabila terdapat data yang terkumpul. Seterusnya, taburan Poisson-Hurwitz-Lerch zeta (Poisson-HLZ) telah dibina sebagai pengitlakan baru bagi taburan binomial negatif. Beberapa pendekatan telah digunakan dalam penerbitan fungsi kebarangkalian taburan ini untuk menunjukkan perkaitan antara kaedah-kaedah daripada bidang-bidang seperti matematik, statistik dan sains aktuari. Ukuran statistik asas dan sifat-sifat berkebarangkalian taburan Poisson-HLZ telah diterbitkan dan kebergunaan model ini terus ditunjukkan melalui pemodelan beberapa set data yang benar. Akhirnya, taburan trinomial tersongsang (TTT) dikaji semula. Kedua-dua taburan Poisson-HLZ dan TTT telah dibukti sebagai taburan Poisson bercampur. Fakta ini boleh melanjutkan penggunaan taburan-taburan ini dalam pemodelan pelbagai fenomena. Taburan pencampuran bagi TTT juga diperoleh sebagai siri Laguerre tak terhingga dan hasilnya dibanding secara berangka dengan beberapa penyongsangan penjelmaan Laplace.

TABLE OF CONTENTS

NOTATIONS AND ABBREVIATIONS

$\setminus$	-	excluding
$\simeq$	-	is approximately equal to
$\triangleq$	-	is defined by
$=_{st}$	-	equality in law
$\xrightarrow{D}$	-	convergence in distribution
$n \mid m$	-	m is divisible by n
$\mathbb{N}$	-	$\{1, 2, \ldots\}$
$\mathbb{N}_0$	-	$\{0, 1, 2, \ldots\}$
$\mathbb{Z}_0^-$	-	$\{0, -1, -2, \ldots\}$
$\mathbb{R}$	-	the set of real numbers
$\mathbb{R}_0^+$	-	the set of nonnegative real numbers
$\mathbb{C}$	-	the set of complex numbers
$\mathbb{R}^d$	-	d-dimensional space of real numbers
$\lfloor x \rfloor$	-	the greatest integer smaller than or equal to x
$\lceil x \rceil$	-	the smallest integer greater than or equal to x
$[x]$	-	x rounded to the nearest positive integer
$\max(x, y)$	-	the larger of x and y
$\mathrm{Re}(x)$	-	real part of x
$n^{(k)}$	-	$n(n-1)\cdots(n-k+1)$
$(n)_k$	-	$n(n+1)\cdots(n+k-1)$
$\leq_{hr}$	-	is smaller in the hazard rate order than

$\leq_{lr}$	-	is smaller in the likelihood ratio order than
$\leq_{Lt}$	-	is smaller in the Laplace transform order than
$\leq_{rh}$	-	is smaller in the reverse hazard rate order than
$\leq_{st}$	-	is smaller in the usual stochastic order than
χ^2	-	chi-square statistic
$\chi^2_{\alpha,n}$	-	α-level critical value of χ^2 distribution with n degrees of freedom
$\Phi(\theta,s,a)$	-	Lerch transcendent
$\Gamma(s)$	-	Gamma function
$\Gamma(s,x)$	-	incomplete Gamma function
λ_f	-	umbral operator for f
μ	-	probability measure
μ'_r	-	r-th moment
$\mu'_{r:Y}$	-	r-th moment of Y
μ_r	-	r-th central moment
$\hat{\boldsymbol{\theta}}(\mathbf{X})$	-	maximum likelihood estimator for $\boldsymbol{\theta}$ based on the sample $\mathbf{X}$
${}_2F_1[a,b;c;x]$	-	Gauss hypergeometric function
D^k	-	k-th derivative operator
$D(\boldsymbol{\theta},\alpha,n)$	-	probability generating function based divergence statistic
$\overline{f}(t)$	-	compositional inverse of $f(t)$
$\hat{f}(s)$	-	Laplace transform of f
$F_\theta(x)$	-	distribution function with parameter θ

$G_X(t)$	-	probability generating function of random variable X
$I_\nu(z)$	-	modified Bessel function of order ν
log	-	natural logarithm
$L_n^a(x)$	-	Laguerre polynomials orthogonal with respect to $x^a\, e^{-x}$ over $(0,\infty)$
$L(\boldsymbol{\theta}\,\vert\,\mathbf{x})$	-	likelihood function based on sample $\mathbf{x}$
p_k	-	probability function
$s_k(x)$	-	k-th degree Sheffer polynomial in x
$T(\theta, s, a)$	-	$\theta\Phi(\theta, s+1, a+1)$
$T_F(x)$	-	tail probability of the distribution F
AIC	-	Akaikean information criterion
GLD	-	generalised logarithmic distribution
HLZ	-	Hurwitz-Lerch zeta
HLZD	-	Hurwitz-Lerch zeta distribution
ITD	-	inverse trinomial distribution
LD	-	logarithmic distribution
ML	-	maximum likelihood
MP	-	mixed Poisson
MPD	-	mixed Poisson distribution
NBD	-	negative binomial distribution
pdf	-	probability density function
pgf	-	probability generating function
pmf	-	probability mass function
Poisson-HLZD	-	Poisson-stopped sum of HLZD
PSSD	-	Poisson-stopped sum distribution

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

1.0 Probability Models for Count Data

Count data consists of the frequencies of random events. For example, the number of terrorists activities for 110 countries between 1971 and 2007 (Freytag *et al.*, 2011), number of single-vehicle fatal crashes at marked segments on multilane rural highways between 1997 and 2001 (Lord & Geedipally, 2011), counts of Leadbeater's possum on 151 three-hectare sites in the montane ash forests in Victoria, Australia (Dobbie & Welsh, 2001) and, not forgotten, the number of deaths by horse kicks in the Prussian army analysed by L. von Bortkiewicz in 1898 (Quine & Seneta, 1987). The list of examples could never be exhausted as new data are generated through experiments, observations, and surveys every day. Count data are usually fitted to lattice distributions with support on the set of nonnegative integers $\mathbb{N}_0$ or the set of positive integers $\mathbb{N}$ or their subsets. These distributions will hereafter be generally referred to as discrete distributions.

Empirical data fitting aims to find a parsimonious probability model that is adequate to accommodate the variations observed in real life data. Since the amazingly good fit obtained by Bortkiewicz in late nineteenth century (see Quine & Seneta, 1987), Poisson distribution has become a popular model for the counts of 'rare events'. However, due to its single parameter restriction and rigid variance to mean ratio, the Poisson distribution is not flexible enough to describe datasets that are over-dispersed (Famoye & Singh, 2006; Rigby *et al.*, 2008), under-dispersed (Ridout & Besbeas, 2004), with excess number of zero (Yip & Yau, 2005; Hu *et al.*, 2011), with multiple modes (Cortina-Borja, 2006) or heavy-tailed (Hougaard *et al.*, 1997; Gupta & Ong, 2005). The negative binomial distribution (NBD) that consists of two parameters shows more flexibility over the Poisson when the data is over-dispersed and has been applied extensively in biology, actuarial science, economics and other areas. The popularity of the NBD has also been partly attributed to its mixed Poisson (MP) formulation as derived by Greenwood & Yule (1920) and its representation as the Poisson-stopped sum of logarithmic distribution (LD) obtained in Quenouille (1949). These derivations render the NBD with natural interpretations to many real life phenomena and also provide effective ways to construct generalisations of the NBD. Nonetheless, the NBD also has limitations for many datasets as reported in Joe & Zhu (2005), Zhang *et al.* (2008), Sugita *et al.* (2011) and Geedipally *et al.* (2012).

In order to overcome the limitations of existing probability models, new distributions have been constructed in various ways and added into practitioners' inventory. Some commonly used construction techniques include relaxation of parameters, introducing new parameters into existing models, forming finite mixture of distributions, continuous mixing, convolution, stopped-sum formulation, truncation, transformation and so on. For a general survey of these methods, see Johnson *et al.* (2005). On top of flexibility, it is also desirable for the extended models to equip with some stochastic interpretations and probabilistic properties that are useful in modelling. In addition, existence of simple and robust statistical inference procedure is also important factor that decides the practical usefulness of a probability model.

1.1 Thesis Organisation

In the following chapters, some generalised distributions, including a new generalisation of the NBD, will be studied from different aspects. Literature review on the developments of relevant theories and a summary of research findings will be presented in the following two sections.

Chapter 2 consists of a collection of definitions, properties and theorems that will be used throughout the thesis. Stochastic orderings between distributions in the Hurwitz-Lerch zeta (HLZ) family will be established in Chapter 3 in which the tail behaviour of the HLZ distribution (HLZD) and a robust inference procedure will also be examined. Chapter 4 proposes the Poisson-Hurwitz-Lerch zeta distribution (Poisson-HLZD) as a new generalisation of the NBD and explores its various properties in depth. The inverse trinomial distribution proposed by Shimizu & Yanagimoto (1991) is reviewed in Chapter 5 and some new results have been obtained. The thesis is concluded with remarks and discussion on possible future research directions in Chapter 6.

1.2 Literature Review

Since the derivation of Pascal distribution by Montmort in 1713 as the probability distribution of the number of tosses of a coin necessary to achieve k heads (Johnson *et al.*, 2005), the NBD has gone through various generalisations in the passage of history. These include expansion of the parameter space, such as allow the k in a Pascal distribution to take non-integer positive value or even negative values as in Engen's extended NBD (Engen, 1974). To reflect the variations between the mean and variance, Jain & Consul (1971) proposed a generalised NBD with mean that is positively correlated to the variance by introducing new parameter through the Lagrangian transformation. Later, Ghitany & Al-Awadhi (2001) and Gupta & Ong (2004) mixed the Poisson mean with some generalised gamma distributions to construct larger families of distributions that include the NBD as special case. On the other hand, generalisations of the NBD can also be achieved by extending the LD in a Poisson-stopped sum formulation as shown in Kempton (1975), Ong (1995) and Khang & Ong (2007).

The Zipf's law or the power law that describes decay which is slower than the exponential rate has been applied to various fields where ranking of sizes or frequencies are of particular interest, such as quantitative linguistics (Calderon *et al.*, 2009), human demography (Gan *et al.*, 2006), economics (Zhang *et al.*, 2009), citations (Perc, 2010) and internet traffic modelling (Clegg *et al.*, 2010). Many well-known distributions such as the Zipf-Mandelbrot distribution, Lotka distribution and so on follow the Zipf's law. A unified representation of these distributions that generalised the LD was discussed under the setting of HLZD in Zörnig & Altmann (1995). A list of distributions in the HLZ family and their corresponding parameters can be found in Kemp (2010) where the HLZD was treated as a special case of the power series distribution. The reliability properties of the HLZD was examined in Gupta *et al.* (2008) in which the authors also demonstrated data-fitting by maximum likelihood (ML) procedure using global optimization technique. However, the parameters of the HLZD in their studies were restricted to a proper subset of the natural feasible parameter space. By relaxing the parameters, more flexible models can be obtained. The possibility of shifting the support of the HLZD to include zero has been addressed by Aksenov & Savageau (2005). Generalised HLZ functions have also been considered in constructing new continuous distributions as shown in Saxena *et al.* (2011) and Tomovski *et al.* (2012) recently.

The magnitudes of the random variables in the HLZ family subject to different parameters can be compared by using the idea of stochastic orders. Stochastic orders have found applications in many diverse areas such as reliability theory and survival analysis, life sciences, operations research, actuarial science, economics, and management science. These orders can be defined on random variables in different ways depending on the purpose of research. The usual stochastic ordering, which tells whether a random variable is more likely to take smaller values compared to another is useful, for example, in comparing risks. Some other orderings may be defined based on the hazard rate, reverse hazard rate or transforms of the random variables, see Shaked & Shanthikumar (2007). The relations between different types of orders are illustrated in Shaked & Wong (1997). Orderings of the Poisson and the MP random variables were established in Misra *et al.* (2003) whereas comparisons of the NBD and mixed NBD have been done by Pudprommarat & Bodhisuwan (2012). Similar results can be obtained for the HLZ family and the generalisation of NBD based on the HLZ family.

As mentioned earlier, the NBD is a special case of stopped sum distribution. An $\mathcal{F}_1$-stopped sum distribution models the sum of observations from a cluster size distribution $\mathcal{F}_2$, which is independent of $\mathcal{F}_1$, when the number of observations follows distribution $\mathcal{F}_1$. It has been called by different names in literatures, such as the contagious distribution (Neyman, 1939), generalised $\mathcal{F}_1$ distribution (Gurland, 1957), compound $\mathcal{F}_1$ distribution (Feller, 1967) or, more recently, the multiple Poisson distribution when $\mathcal{F}_1$ is a Poisson distribution (Wimmer & Altmann, 1996). Charalambides (2005) considered the stopped-sum distribution as the total number of balls in a random occupancy model where the number of urns follows $\mathcal{F}_1$ distribution and apply combinatorial techniques to derive some of its moment properties. Using these properties, Johnson *et al.* (2005) showed that a stopped-sum distribution always has index of dispersion that is greater than that of its cluster size distribution. The Poisson-stopped sum distribution (PSSD) is perhaps the most popular model in this family. It has simple probability generating function (pgf) that is convenient for manipulation and probability mass function (pmf) that can always be evaluated recursively. The moments, factorial moments, cumulants and factorial cumulants of a PSSD are closely related to their counterparts in the cluster size distribution as pointed out in Douglas (1980). Some of its subclasses that have attracted special attentions from researchers include the Hermite (Poisson-Binomial), Neyman Type A (Poisson-Poisson), Pólya-Aeppli (Poisson-Geometric), Delaporte (Willmot & Sundt, 1989), non-central negative binomial (Ong & Toh, 2001), inverse trinomial (Khang & Ong, 2007) and some of their generalisations.

The PSSD also attracts attentions from researchers due to its close relationship with infinitely divisible distribution that underlies every Lévy process at a fixed time (Sato, 1999). A comprehensive review on classical results of infinite divisibility can be found in Steutel & Van Harn (2004) and Bose *et al.* (2002). While all $\mathcal{F}_1$-stopped sum distributions with infinitely divisible $\mathcal{F}_1$ are infinitely divisible, de Finetti showed that all infinitely divisible distributions are actually weak limit of certain Poisson-stopped sum in 1931 (Johnson *et al.*, 2005; Mainardi & Rogosin, 2006). In fact, all infinitely divisible discrete distributions can be uniquely represented in the form of PSSD with a cluster size distribution on the set of positive integers (Steutel, 1973). Two other closely related concepts are self-decomposability and unimodality. Not all infinitely divisible distributions are unimodal and Masse & Theodorescu (2005) indicate that characterising the modality region of a PSSD could be a difficult problem even in the relatively simple case such as the Neyman Type A distribution. Part of this problem can be tackled through the concept of self-decomposable distributions that form a large subclass of infinitely divisible law that possesses the unimodal property (Steutel & Van Harn, 2004, Chapter V). Self-decomposable distribution is also of interest as it contains the subclass of stable law. Since unimodality is not always closed under convolution, it motivates the study on strongly unimodal property that preserves unimodality when a strongly unimodal distribution convolutes with any other unimodal distributions. Discrete strongly unimodality has been studied by Keilson & Gerber (1971), in which they proved that log-concavity of the probability function is a necessary and sufficient condition for a distribution to be strongly unimodal and show that the NBD is strongly unimodal under certain conditions. On the other hand, log-convexity of the probability function on $\mathbb{N}_0$ sufficiently leads to the infinite divisibility of a discrete distribution (Steutel & Van Harn, 2004, Chapter II).

Applications of finite mixture models show traces in various areas including engineering image analysis (Omar Mohd Rijal, Norliza Mohd Noor & Liew, 2005), medicine (Tohka *et al.*, 2007) and statistics (Titterington *et al.*, 1985). An important subclass of the finite mixture models is the zero-modified distribution in which the probability of getting zero is either suppressed or inflated to get a better fit to the data. In order to explain apparent contagious phenomena due to unobserved heterogeneity, more general mixed distribution has been introduced by treating the parameter of a distribution as the outcome of a random variable (Cameron, 1998). This feature is particularly attractive in actuarial risk theory when the policyholders are believed to carry different degrees of accident proneness (Klugman *et al.*, 2008). In the mixture family, the class of mixed Poisson distribution (MPD), in which many of the members are also PSSD's, has been the focus of interest. Maceda (1948) proved that the MPD resulting from an infinitely divisible mixing distribution on a nonnegative support is always infinitely divisible and is hence a PSSD. Conversely, a condition, under which a PSSD is MPD with discrete mixing distribution, was given in Gurland (1957). In order to cover the case when a non-infinitely divisible mixing distribution result in an infinitely divisible MPD, the idea of quasi-infinite divisibility has been introduced (Puri & Goldie, 1979). Holgate (1970) showed that an absolutely continuous unimodal mixing distribution always gives rise to a unimodal MPD and the result was extended by Bertin & Theodorescu (1995) to mixing distribution that is not necessarily absolutely continuous. Recently, Karlis & Xekalaki (2005) give a comprehensive survey on some important results related to the MPD while Gupta & Ong (2005) illustrated their applications in fitting very long-tailed data.

The MPD can be studied through its pgf which corresponds to the Laplace transform of the mixing distribution after a substitution. This allows characterisation of the MPD via proving the completely monotone of a function, see Bhattacharya & Waymire (2007). Only very few papers in the literature show the proofs of these types, among these, Joe & Zhu (2005) have proved that a generalised Poisson distribution is MPD. Miller & Samko (2001) provide some useful results in proving the complete monotonicity of a function. Due to the complexity of the transformed function, recovering the mixing distribution in a closed form expression through inverse Laplace transform is not always possible. One of the possible solutions is to express the inversion in infinite series based on the Fourier-series method proposed by Dubner & Abate (1968). Inversion that is based on the Laguerre series was studied in Abate *et al.* (1996) whereas Kabardov & Ryabov (2009) consider the accelerated convergence of these series. Approximation to the inversion by using the maximum entropy principle based on moments was proposed by Tagliani & Velasquez (2004). This method is particularly useful in approximating the inversion of the transform of probability function as it produces nonnegative result. Hassanzadeh & Pooladi-Darvish (2007) compared different numerical Laplace inversion methods with known analytical solution and found that the Fourier-series method is more accurate although computationally more expensive. Later, Masol & Teugels (2010) compared results of numerical inversions of the transformed Gamma distribution and suggested the Post-Widder algorithm or Gaver-Stehfest algorithm for approximations of the tail probabilities. Some of these methods require high-ordered differentiation, which could be a luxury demand from the computational point of view for more complicated function. Convergence of infinite series, stability of round-off errors and selection of truncation criteria should also be considered in the execution of these algorithms. Agreement of results from two or more methods is recommended for getting a reliable conclusion.

Parameter estimation is an important step in modelling. The method of moments, ML estimation and Bayesian estimation are among the popular classical approaches. The famed ML estimator possesses many desirable features such as consistency, invariance, and asymptotically unbiased, efficient and normal under some regularity conditions. However, it is also known to be sensitive to the presence of outliers (Pawitan, 2001; Millar, 2011). More robust estimation methods that based on minimizing statistical distances between the model and the empirical distribution have been proposed and are getting increasingly popular following the advancement of computing technology. On the other hand, using transforms of distribution such as the pgf in estimation and hypothesis testing have been proposed by some researchers when the pgf has a simpler form than the probability functions; see, for example, Kocherlakota & Kocherlakota (1986) and Kemp & Kemp (1988). A novel approach through minimizing the generalized Hellinger-type divergence between the pgf and the empirical pgf has been proposed in Sim & Ong (2010) to eliminate the dependency of estimates on the dummy variable in the pgf. They have also demonstrated the robustness of the estimators in contaminated data through a series of simulation study and suggested the forms of divergence that are computationally more efficient. This approach, however, leads to certain arbitrariness in constructing the empirical pgf for grouped data. Reducing the effect of this arbitrariness and enhancing the estimation outcome are possible through some graduation schemes designed exclusively for some families of distributions.

1.3 Contributions of the Thesis

The main contributions of this thesis are listed as follows. Results obtained in Chapter 3 have been accepted for publication and two other submissions will be based on the findings in Chapter 4 and Chapter 5.

- Further properties of the HLZD are derived and the feasibility of a pgf based estimation procedure for this family of distributions is examined.
    - Likelihood ratio orderings between the members of the HLZ class of distributions are established and will be applied to derive corresponding results for the Poisson-HLZD.
    - Comparison between the tail probabilities of the HLZD and that of the generalisation of LD in Khang & Ong (2007) is illustrated.
    - The pgf based estimation method suggested in Sim & Ong (2010) is modified to handle the problem arising from grouped data for the HLZ class of distributions.
- A new class of distributions is constructed as the Poisson stopped-sum of the HLZD. This new class generalises the NBD to a 4-parameter model and is shown to be more flexible and useful in data-fitting.
    - Basic properties of the Poisson-HLZD, including expressions for probability function, generating function, moments and other measures, are derived.
    - The Poisson-HLZD is shown to be self-decomposable and strongly unimodal under certain conditions.
    - A moment-ratio diagram is used to depict the versatility of the Poisson-HLZD as compare to other well-known frequency models.

    - The Poisson-HLZD is proved to have a MP formulation under certain conditions and the mixing distribution is obtained numerically.
- The inverse trinomial distribution is formulated as a MPD under certain condition.
    - A necessary and sufficient condition for the inverse trinomial distribution to have MP formulation is obtained.
    - The mixing distribution of the inverse trinomial distribution is derived in terms of the Laguerre polynomials and is compared to the result obtained from numerical inversion.

CHAPTER 2 PRELIMINARIES

2.0 Introduction

Concepts and theorems that are useful for the derivations of results in subsequent chapters are presented in this chapter. The interrelations among some properties and their importance in statistics are also highlighted.

2.1 Absolutely Monotone and Completely Monotone

A nonnegative function $f(x)$ on $(0,\infty)$ is said to be absolutely monotone if its *n*-th derivative $f^{(n)}(x)$ exists for all $n \in \mathbb{N}$ and $f^{(n)}(x) \geq 0$ on $(0,\infty)$. If, on the other hand, $(-1)^n f^{(n)}(x) \geq 0$ on $(0,\infty)$ then $f(x)$ is said to be completely monotone.

The following results will be used in proving the complete monotonicity of functions in later chapters.

Theorem 2.1 (Miller & Samko, 2001; Theorem 2): If $f(x)$ is completely monotonic and $h(s)$ is nonnegative with a completely monotonic derivative, then $f\left(h(s)\right)$ is completely monotone with respect to *s*.

Theorem 2.2 (Miller & Samko, 2001; Theorem 3): Given a completely monotone function $y = h(s)$, if the power series $\varphi(y) = \sum_{k=0}^{\infty} a_k y^k$, where $a_k \geq 0$ for all *k*, converges for all *y* in the range of $h(s)$, then $\varphi(h(s))$ is completely monotone in *s*.

Theorem 2.3 (Miller & Samko, 2001; Theorem 4): Let $K(x,t)$ be completely monotone in x for all $t \in (0,\infty)$ and $f(t)$ be a nonnegative locally integrable function such that all the integrals

$$\int_a^b \frac{\partial^n}{\partial x^n} K(x,t) f(t) dt, \quad n \in \mathbb{N}_0$$

converge uniformly in a neighbourhood of any point $x \in (0,\infty)$. Then

$$F(x) = \int_a^b \frac{\partial^n}{\partial x^n} K(x,t) f(t) dt, \quad 0 \le a < b \le \infty;$$

is completely monotone.

2.2 Probability Generating Function

Let X be a discrete random variable with pmf $\{p_k\}$, the probability generating function (pgf) of X, $G_X(t)$ is defined by

$$G_X(t) \triangleq E[t^X] = \sum_k p_k t^k \ . \tag{2.1}$$

$G_X(t)$ always exists for $-1 \le t \le 1$ with $G_X(0) = p_0$ and $G_X(1) = 1$. The pgf uniquely determines a distribution and has the following properties:

(a) Let $G_X(t)$ and $G_Y(t)$ be the pgf's of independent random variables X and Y respectively, then the pgf of the convolution of X and Y is $G_X(t)G_Y(t)$.

(b) $$E[X] = \frac{d}{dt} G_X(t)\bigg|_{t=1} \ . \tag{2.2}$$

(c) The set of absolutely monotone functions $M(t)$ on $[0,1)$ such that $\lim_{t \to 1^-} M(t) = 1$, is equal to the set of all pgf's (Steutel & Van Harn, 2004, Theorem A.4.3).

2.3 Stopped Sum Distribution

Consider a discrete random variable N with pgf $G_N(t)$, and a sequence of independent and identically distribution random variables Y_i having a common distribution as the random variable Y with pgf $G_Y(t)$. The sum $S_N = \sum_{i=1}^{N} Y_i$, which is stopped by the outcome of N, is said to have a stopped sum distribution. It is easy to show that the pgf of S_N,

$$G_{S_N}(t) = G_N(G_Y(t)). \tag{2.3}$$

The distribution of Y is also called the cluster size distribution. When N is a Poisson random variable, we have the widely studied class of PSSD with pgf in the form

$$\exp[\lambda(G_Y(t)-1)], \tag{2.4}$$

where $G_Y(t)$ is the pgf of the cluster size distribution and λ is the Poisson mean of N.

2.4 Mixture Distribution

Given $k \in \mathbb{N}$ and $\omega_1,...,\omega_k \in \mathbb{R}_0^+$ such that $\sum_{i=1}^{k} \omega_i = 1$, a k-component finite-mixture distribution $F(x)$ is defined as the weighted average of k distributions $F_1,...,F_k$, that is $F(x) = \sum_{i=1}^{k} \omega_i F_i(x)$. This formulation, which has natural interpretation when the sample is drawn from a population that consists of heterogeneous subpopulations, has made finite mixture models appealing in modelling.

On the other hand, consider a parameter space Θ and a σ-algebra $\mathcal{A}$ in Θ. Let μ be a probability measure on $\mathcal{A}$ and $F_\theta(x)$ be a distribution on $\mathbb{R}$ for every $\theta \in \Theta$ such that $\theta \mapsto F_\theta(x)$ is $\mathcal{A}$-measurable. A mixed distribution $F(x)$ can be constructed by defining $F(x)=\int_{\Theta} F_{\theta}(x)\mu(d\theta)$ (see Steutel & Van Harn, 2004, Chapter VI). In this case, μ is the probability measure induced by the mixing distribution that plays the role of a weight function. Furthermore, if $F_{\theta}(x)$ is absolutely continuous with density $f_{\theta}(x)$ for every θ, then $F(x)$ is also absolutely continuous with density $f(x)=\int_{\Theta} f_{\theta}(x)\mu(d\theta)$.

When $F_{\theta}(x)$ is the distribution function of a discrete random variable, the pmf of the mixed distribution is given by $p_k=\int_{\Theta} p_{\theta}(k)\mu(d\theta)$. The MPD, obtained by ascribing $p_{\theta}(k)=\dfrac{e^{-\theta}\theta^k}{k!}$, forms an important subclass that includes many well-known distributions such as the NBD, Delaporte, Poisson-inverse Gaussian (Sichel), Poisson-Tweedie and so on (Johnson *et al.*, 2005). The pmf and pgf of a MPD with mixing distribution $f(\theta)$ on $[0,\infty)$ are, respectively, given by

$$p_k=\int_0^{\infty}\frac{e^{-\theta}\theta^k}{k!}f(\theta)d\theta \tag{2.5}$$

and $$G(t)=\int_0^{\infty}e^{\theta(t-1)}f(\theta)d\theta. \tag{2.6}$$

Theorem 2.4 (Bernstein's Theorem; Bhattacharya & Waymire, 2007; Theorem 8.6): A function $\phi(s)$ is completely monotonic on $(0,\infty)$ if and only if there is a measure μ on $[0,\infty)$ such that

$$\phi(s)=\int_0^{\infty}e^{-\lambda s}\mu(d\lambda), \quad s>0. \tag{2.7}$$

In particular, μ is a probability measure if and only if $\lim_{s\to 0^+}\phi(s)=1$.

Expressions (2.6) and (2.7) also suggest that a substitution of $t=1-s$ in the pgf $G(t)$ gives the Laplace transform of the mixing distribution $f(\lambda)$, see also Section 2.10. Furthermore, if $G(1-s)$ is completely monotone with respect to s on $(0,\infty)$ such that

$\lim_{s \to 0^+} G(1-s) = 1$ then, by the Bernstein's theorem, the random variable with pgf $G(t)$ must have a MPD.

2.5 Moments, Skewness and Kurtosis for PSSD

Moments for the random variable X with pgf (2.4) will be considered in this section. Using (2.2), it is easy to show that the random variable X with pgf (2.4) has mean and variance

$$E[X] = \lambda E[Y] \quad \text{and} \quad Var(X) = \lambda E[Y^2]. \tag{2.8}$$

If denote the r-th moment of X, $E[X^r]$ by μ'_r and the r-th moment of Y by $\mu'_{r:Y}$, Ross (1996) gives the recursive formula

$$\mu'_r = \lambda \sum_{i=0}^{r-1} \binom{r-1}{i} \mu'_i \mu'_{r-i:Y} ,$$

which can be used repeatedly to obtain higher moments of X. For example,

$$\mu'_3 = \lambda E[Y^3] + 3\lambda E[X]E[Y^2] + \lambda^3 (E[Y])^3 \tag{2.9}$$

and

$$\mu'_4 = \lambda \left\{ E[Y^4] + 3E[X]E[Y^3] + 3E[X^2]E[Y^2] + E[X^3]E[Y] \right\}. \tag{2.10}$$

Define the r-th central moments of X, $\mu_r = E[(X-\mu)^r]$. Using (2.9) and (2.10) and some simple algebra, the third and fourth central moments of X can be expressed in the moments of the cluster size distribution,

$$\mu_3 = \lambda E[Y^3] \quad \text{and} \quad \mu_4 = \lambda E[Y^4] + 3\lambda^2 (E[Y^2])^2 . \tag{2.11}$$

The skewness and kurtosis of a distribution can be defined through its central moments. For PSSD, they are closely linked to the Poisson mean and the moments of the cluster size distribution after substitution using (2.8) and (2.11).

$$\text{Skewness } = \frac{E[Y^3]}{\sqrt{\lambda E[Y^2]^3}} \text{ and Kurtosis } = 3 + \frac{E[Y^4]}{\lambda E[Y^2]^2}$$

2.6 Sheffer Sequences and Generalised Laguerre Polynomials

Let $q_k(x)$ be a *k*-degree polynomial in *x* over a field of characteristic zero. The formal power series $\sum_{k=0}^{\infty} q_k(x)t^k$ is called the generating function of the polynomial sequence. Given $g(t) = \sum_{n=0}^{\infty} a_n t^n$ and $f(t) = \sum_{n=1}^{\infty} b_n t^n$ with $a_0 b_1 \neq 0$, the compositional inverse of $f(t)$ always exists and is denoted by $\bar{f}(t)$. Consider generating function of the form

$$\frac{1}{g\left(\bar{f}(t)\right)} e^{x\bar{f}(t)} = \sum_{k=0}^{\infty} \frac{s_k(x)}{k!} t^k \tag{2.12}$$

where $s_k(x)$ in (2.12) is a *k*-degree polynomial in *x*. $s_k(x)$ is called the Sheffer polynomial for $(g(t), f(t))$. When $g(t) = 1$, $s_k(x)$ is named the associated sequence for $f(t)$. The Sheffer system of polynomials includes many well-known classical polynomials such as Hermite, Laguerre, Charlier, Abel, Bell, Poisson-Charlier, actuarial polynomials and so on, see Roman (2005).

To facilitate further discussion, let D^k denote the *k*-th derivative operator on the algebra of Sheffer polynomials, that is

$$D^k x^n = \begin{cases} n^{(k)} x^{n-k}, & k \leq n, \\ 0, & k > n, \end{cases}$$

where $n^{(k)} = n(n-1)...(n-k+1)$. When $s_k(x)$ is associated for $f(t)$, defined the umbral operator λ_f for $f(t)$ on the algebra of Sheffer polynomials by $\lambda_f x^k = s_k(x)$. With these notations, the recurrence formula for associated sequences for $f(t)$ can be stated as follows.

Theorem 2.5 (Roman, 2005; Corollary 3.6.6): If $s_k(x)$ is associated to $f(t)$, then $s_{k+1}(x) = x\lambda_f \frac{d}{dt}\left[\bar{f}(t)\right]\Big|_{t=D} x^k$.

Given $a > -1$ and $n \in \mathbb{N}_0$, the generalised Laguerre polynomials, denoted by $L_n^a(x)$, can be defined from the recurrence relation

$$(n+1)L_{n+1}^a(x) = (2n+a+1-x)L_n^a(x) - (n+a)L_{n-1}^a(x)$$

with $L_0^a(x) = 1$ and $L_1^a(x) = a+1-x$ (see Erdélyi *et al.*, 1953; Vol II, pg 188). The generalised Laguerre polynomials are orthogonal with respect to $x^a e^{-x}$ over $(0,\infty)$ and has generating function $\sum_{k=0}^{\infty} L_n^a(x)t^n = \frac{1}{(1-t)^{a+1}} \exp\left[\frac{xt}{t-1}\right]$.

If $\frac{1}{g\left(\bar{f}(t)\right)}$ and $e^{x\bar{f}(t)}$ are, respectively, pgf of independent random variables *X* and *Y*, then (2.12) is the pgf of $X+Y$ and $P(X+Y=k) = \frac{s_k(x)}{k!}$. The non-central NBD and the discrete Charlier distribution are examples with pmf that are expressible in terms of orthogonal polynomials, see Ong (1987) and Ong (1988).

2.7 Probabilistic Properties of a Distribution

In this section, a few probabilistic properties that have attracted attentions from researchers since the last century will be introduced.

2.7.1 Infinite Divisibility

A random variable X is said to be infinitely divisible if for every $n \in \mathbb{N}$ it can be written as $X \stackrel{d}{=} X_{n,1} + X_{n,2} + ... + X_{n,n}$, where $X_{n,1}, X_{n,2}, ..., X_{n,n}$ are independent with $X_{n,j} \stackrel{d}{=} X_n$ for all j and some random variable X_n, the n-th order factor of X. The practical interest of infinite divisibility is mainly in modelling the sum of several independent quantities with the same distribution.

All non-degenerate discrete infinitely divisible distributions must be unbounded, with positive mass at 0 and closed under convolution. Other useful properties are given in the following theorems.

Theorem 2.6 (Steutel & Van Harn, 1979; Lemma 1.2.): A random variable with $0 < p_0 < 1$ and pgf $G(t)$ is infinitely divisible if, and only if,

$$G(t) = \exp\left\{\lambda\left[H(t) - 1\right]\right\}, \tag{2.13}$$

for some $\lambda > 0$ and $H(t)$ is a unique pgf with $H(0) = 0$.

For an infinitely divisible distribution with pgf $G(t)$, the canonical measure r_k can be defined via its generating function

$$R(t) = \frac{d}{dt}\log G(t) = \sum_{k=0}^{\infty} r_k t^k \ . \tag{2.14}$$

Theorem 2.7 (Steutel & Van Harn, 1979; Lemma 1.2.): A distribution is infinitely divisible if, and only if, the probabilities $\{p_n\}$ satisfy

$$(n+1)p_{n+1} = \sum_{j=0}^{n} r_{n-j}\, p_j \tag{2.15}$$

with canonical measure $r_k \geq 0$ for all *k*.

In other words, the non-negativity of the canonical measure r_k characterizes the infinite divisibility of the distribution. Consider an infinitely divisible pgf $G(t)$ written in the form of (2.13) with $H(t) = \sum_{k=1}^{\infty} q_k t^k$. From (2.14),

$$R(t) = \frac{d}{dt}\{\lambda H(t)\} = \frac{d}{dt}\left\{\sum_{k=1}^{\infty} \lambda q_k t^k\right\} = \sum_{k=1}^{\infty} \lambda k q_k t^{k-1} = \sum_{k=0}^{\infty} r_k t^k \ .$$

Comparing the coefficient of t^k ,

$$r_k = \lambda(k+1)q_{k+1}; \qquad (2.16)$$

which shows the connection between the canonical measure of a Poisson-stopped sum distribution and its cluster size distribution.

2.7.2 Log-convexity and Log-concavity

Log-convexity and log-concavity are useful in deriving other properties of a distribution. A discrete distribution $\{p_n\}$ is said to be log-convex if $p_n^2 \le p_{n-1} p_{n+1}$ for all *n*. Log-concave is defined with the inequality sign reverses.

Theorem 2.8 (Steutel & Van Harn, 2004; Chapter II, Theorem 10.3): If the canonical measures r_k of an infinitely divisible distribution $\{p_n\}$ is log-concave, then $\{p_n\}$ is log-concave if, and only if $r_1 \le r_0^2$.

It is easy to see that the product of two log-concave (log-convex) sequences is log-concave (log-convex). The sequence of natural numbers is log-concave. From (2.16), it is trivial that the canonical measure of a Poisson-stopped sum distribution is log-concave if the cluster size distribution is log-concave.

2.7.3 Discrete Self-Decomposability

Following Steutel & Van Harn (1979), a discrete distribution on $\mathbb{N}_0$ with pgf $G(t)$ is said to be discrete self-decomposable if

$$G(t) = G\left(1-\alpha+\alpha t\right)G_{\alpha}(t), \quad |t|\leq 1, 0\leq\alpha\leq 1;$$

where $G_{\alpha}(t)$ is a pgf. From the definition, it is clear that a discrete self-decomposable distribution can be expressed as the convolution of two discrete distributions in which one is the thinning version of the original distribution. Discrete self-decomposability can be determined through its canonical measure as stated in the next theorem.

Theorem 2.9 (Steutel & Van Harn, 1979; Theorem 2.2): A discrete distribution $\{p_n\}$ is discrete self-decomposable if, and only if, it is infinitely divisible and has a non-increasing canonical measure r_n.

2.7.4 Unimodality and Strong Unimodality

A discrete distribution $\{p_n\}$ is said to be unimodal if there exists an M such that $p_n \geq p_{n-1}$ for all $n \leq M$ and $p_n \leq p_{n-1}$ for all $n \geq M$. A discrete distribution $\{p_n\}$ is strongly unimodal if the convolution of $\{p_n\}$ with any unimodal distribution is unimodal. The class of strongly unimodal discrete distribution is closed under reversal ($\bar{p}_n = p_{-n}$), convolution and passage to limit. Strongly unimodal distributions also have all moments.

Theorem 2.10 (Keilson & Gerber, 1971; Theorem 2): A necessary and sufficient condition for strongly unimodal is log-concavity, that is $p_n^2 \geq p_{n-1}p_{n+1}$ for all *n*.

2.8 Stochastic Orders

Consider two random variables X and Y with probability functions $f(x)$ and $g(x)$ respectively. Denote by $F(x)$ and $G(x)$ the respective cumulative distribution function of X and Y, and define their reliability functions to be $\bar{F}(x)=1-F(x)$ and $\bar{G}(x)=1-G(x)$. We say that X is smaller than Y in the usual stochastic order (written $X \le_{st} Y$) if $\bar{F}(x) \le \bar{G}(x)$ for all *x*. X is said to be smaller than Y in the hazard rate order (written $X \le_{hr} Y$) if $\bar{F}(x)\bar{G}(y) \ge \bar{F}(y)\bar{G}(x)$ for all $x \le y$. Similarly, X is said to be smaller than Y in the reverse hazard rate order (written $X \le_{rh} Y$) if $F(x)G(y) \ge F(y)G(x)$ for all $x \le y$. On the other hand, if $f(x)g(y) \ge f(y)g(x)$ for all $x \le y$ or equivalently $\dfrac{f(x)}{g(x)}$ is monotone decreasing in *x*, then we say X is smaller than Y in the likelihood ratio order (written $X \le_{lr} Y$). The likelihood ratio order is important in the sense that it implies the previous three orderings and is easy to verify. The relations between the stochastic orders are summarized in Shaked & Wong (1997) in a diagram partially shown in Figure 2.1.

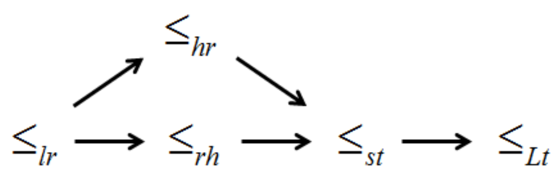

Figure 2.1 Relations between stochastic orders

If the relevant distributions are not in closed form, ordering that are defined based on other characteristics of the distributions could be used, for example, we say that X is smaller than Y in the Laplace transform order, denoted by $X \le_{Lt} Y$ if two nonnegative random variables X and Y are such that $E[\exp\{-sX\}] \geq E[\exp\{-sY\}]$ for all $s > 0$.

The following theorems are useful in deriving the orderings of a PSSD and MPD.

Theorem 2.11 (Shaked & Shanthikumar, 2007; Theorem 1.A.4): Given two sequences of nonnegative independent random variables $\{X_j, j = 1, 2, \ldots\}$ and $\{Y_j, j = 1, 2, \ldots\}$. Let M be a nonnegative integer-valued random variable which is independent of the X_i's. and N be a nonnegative integer-valued random variable which is independent of the Y_i's. If $X_i \leq_{st} Y_i$, $i = 1, 2, \ldots$, and if $M \leq_{st} N$, then $\sum_{j=1}^{M} X_j \leq_{st} \sum_{j=1}^{N} Y_j$.

Using the notation $=_{st}$ to denote equality in law, an analogous result for MPD can be stated as follows.

Theorem 2.12 (Shaked & Shanthikumar, 2007; Theorem 1.A.6): Consider a family of distribution function $\{G_\theta, \theta \in \Omega\}$. Let Θ_1 and Θ_2 be two random variables having supports in Ω with distribution functions F_1 and F_2 respectively and $Y_i =_{st} X(\Theta_i)$ be the random variables with distribution function given by $H_i(y) = \int_\Omega G_\theta(y) dF_i(y)$, $y \in \mathbb{R}$, $i = 1, 2$. If $X(\theta) \leq_{st} X(\theta')$ whenever $\theta \leq \theta'$, and if $\Theta_1 \leq_{st} \Theta_2$, then $Y_1 \leq_{st} Y_2$.

2.9 Parameter Estimation and Model Selection

Let $\mathcal{A}$ be a σ-algebra on the sample space Ω. A parametric family is a set of probability measures $P_{\boldsymbol{\theta}}$ on the measurable space $(\Omega, \mathcal{A})$ such that $P_{\boldsymbol{\theta}}$ is determined when $\boldsymbol{\theta}$, which is an element in the parameter space $\Theta \subseteq \mathbb{R}^d$ for some fixed positive integer d, is known. Parameter estimation is a crucial step in statistical inference to find an appropriate $\boldsymbol{\theta}$ such that the probability measure $P_{\boldsymbol{\theta}}$ of a parametric family fits well to a set of given sample, see, for example, Shao (2003) .

The popular quote “Everything should be made as simple as possible, but not simpler” points out the essence of the principle of parsimony that, as described in Box *et al.* (1994), “employs the smallest number of parameters for adequate representations”. Generalisation of probability distribution introduces more parameters into the models and hence usually produces a better fit to the data. To avoid over-fitting, several information theoretic based criteria have been developed for model selection. These include the Akaikean information criterion (AIC), Bayesian information criterion (BIC) that make use of different functional forms of the maximum log-likelihood statistic to measure the divergences between the proposed models and the real model (Van Der Hoeven, 2005). Other hypothesis testing approaches such as the likelihood ratio test, Wald’s test and the score test are also available, see Hogg *et al.* (2005). In this thesis, the likelihood ratio test will be used to select the appropriate model from a nested family of distributions.

2.9.1 ML Estimation

Given $\boldsymbol{\theta} = (\theta_1, ..., \theta_d) \in \Theta$, let $X_1, X_2, ..., X_n$ be a random sample from a population with probability function $f(k \mid \boldsymbol{\theta})$. For each fixed sample point $\mathbf{x} = (x_1, x_2, ..., x_n)$, the likelihood function is defined by $L(\boldsymbol{\theta} \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i \mid \boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}(\mathbf{x}) \in \Theta$ be a parameter value at which $L(\boldsymbol{\theta} \mid \mathbf{x})$ attains its maximum. The ML estimator of the parameter $\boldsymbol{\theta}$ based on a sample $\mathbf{X}$ is $\hat{\boldsymbol{\theta}}(\mathbf{X})$ (Casella & Berger, 2002).

Finding the ML estimates could be complicated step for certain parametric families as the system of ML equations obtained may not have a closed form solution and hence numerical methods such as the Newton-Raphson method or the EM algorithm will be implemented to seek for numerical solution. These numerical methods usually return a local maximum that depends on the choice of initial point. With the advancement in computing technology, stochastic optimization algorithm that returns the global maxima is preferred when the equations has no analytic solution. A simulated annealing type of random search algorithm will be used in this thesis to obtain the ML estimates for various models, see Robert & Casella (2004).

Consider $\Theta = \prod_{k=1}^{d} I_k \subset \mathbb{R}^d$, where I_k are intervals such that its supremum $\sup I_k = u_k < \infty$ and infimum $\inf I_k = l_k > -\infty$. Define $dia\ \Theta = \max\{u_1 - l_1, ..., u_d - l_d\}$. The algorithm to find the parameter $\hat{\boldsymbol{\theta}}$ in Θ that returns the maximum of an objective function is outlined as follows.

<u>**Algorithm 2.1** Global optimization algorithm for estimation</u>

1. Initialise the search space Θ based on the naturally bounded parameter space or defined a bounded space based on intuition or experience.
2. Choose a stopping criterion ε such that the programme terminates when $dia\ \Theta < \varepsilon$.
3. Set the counter k = 1 and choose a transition criterion T such that the search is refined whenever $k > T$.
4. Starting with a random point $\boldsymbol{\theta}_k$ in Θ, set $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_k$ and calculate the value of the objective function $E_{\hat{\boldsymbol{\theta}}}$ based on $\hat{\boldsymbol{\theta}}$.

5. **while** $dia\ \Theta > \varepsilon$ **do**

a. **while** $k \le T$, **do**

i. Randomly move to another point $\boldsymbol{\theta}_{k+1}$ in Θ and calculate the value of the objective function $E_{\boldsymbol{\theta}_{k+1}}$ based on $\boldsymbol{\theta}_{k+1}$.

1. **if** $E_{\boldsymbol{\theta}_{k+1}} > E_{\hat{\boldsymbol{\theta}}}$, **then** set $k = 1$, $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_{k+1}$ and $E_{\hat{\boldsymbol{\theta}}} = E_{\boldsymbol{\theta}_{k+1}}$. **repeat** i.

2. **if** $E_{\boldsymbol{\theta}_{k+1}} \le E_{\hat{\boldsymbol{\theta}}}$, **then** $k = k+1$.

end do

b. Define the refined search space Θ_s such that $\hat{\boldsymbol{\theta}} \in \Theta_s \subset \Theta$ and $dia\ \Theta_s < dia\ \Theta$. Set $\Theta = \Theta_s$ and $k = 1$.

end do

6. Return the estimate $\hat{\boldsymbol{\theta}}$ and $E_{\hat{\boldsymbol{\theta}}}$.

2.9.2 Likelihood Ratio Test

Using the same notations in 2.9.1 and consider a simpler nested model with reduced parameter space Θ_0 such that $\Theta_0 \subset \mathbb{R}^{d_0} \subset \Theta$. The likelihood ratio test statistic for testing $H_0 : \boldsymbol{\theta} \in \Theta_0$ versus $H_1 : \boldsymbol{\theta} \in \Theta_0^c$ is $\Lambda = \dfrac{\sup_{\Theta_0} L(\boldsymbol{\theta} \mid \mathbf{x})}{\sup_{\Theta} L(\boldsymbol{\theta} \mid \mathbf{x})}$. Let $\hat{\boldsymbol{\theta}}_0(\mathbf{X})$ be the ML estimator for $\boldsymbol{\theta}$ in Θ_0, the test statistic can be written as $\Lambda = \dfrac{L(\hat{\boldsymbol{\theta}}_0(\mathbf{X}))}{L(\hat{\boldsymbol{\theta}}(\mathbf{X}))}$. When $\hat{\boldsymbol{\theta}}_0(\mathbf{X})$ is a consistent estimator, under certain regularity conditions,

$$-2\log\Lambda \xrightarrow{D} \chi^2_{d-d_0}$$

under H_0, see Casella & Berger (2002) and Hogg *et al.* (2005).

2.9.3 AIC

The AIC proposed in Akaike (1974) is a statistic which is penalized to reflect the trade-off between the fit to a model as quantified by $L(\hat{\boldsymbol{\theta}}(\mathbf{X}))$ and model complexity (Millar, 2011) and is given by

$$\mathrm{AIC} = 2[k - \log(L(\hat{\boldsymbol{\theta}}(\mathbf{X}))],$$

where k is the number of unknown parameters to be estimated in the model. The model with smallest fitted AIC is preferred in model selection.

2.10 Numerical Inversion of Laplace Transform

The Laplace transform of a function $f:[0,\infty)\to\mathbb{R}$ is the function $\hat{f}$ defined by $\hat{f}(s)=\int_0^\infty e^{-st} f(t)dt$ whenever the integral converges. Some useful properties of the Laplace transform are listed in the following, see Bartle (1976).

(a) If there exist a real number c such that $|f(t)|\le e^{ct}$ for sufficiently large t, then the Laplace transform exists for $s>c$ and the convergence of the integral is uniform for $s \ge c+\delta$ if $\delta > 0$.

(b) Under the boundedness condition in (a), $\hat{f}$ has derivatives of all orders for $s>c$ and that the n-th derivative

$$\hat{f}^{(n)}(s)=\int_0^\infty (-t)^n e^{-st} f(t)dt \,. \tag{2.17}$$

If $f(t)$ is a probability density function (pdf), it is trivial that the boundedness condition is satisfied for any $c \ge 0$ since $f(t)\to 0$.

If $\hat{f}(s)$ is analytic throughout the finite complex plane except for a finite number of isolated singularities and L_R is a vertical line segment in the complex plane joining $\sigma + Ri$ and $\sigma - Ri$ such that all singularities of $\hat{f}$ lie on the left of L_R. The inverse transform of $\hat{f}$ can be obtained by evaluating the Bromwich integral

$$f(t) = \frac{1}{2\pi i} \lim_{R\to\infty} \int_{L_R} e^{st} \hat{f}(s) ds, \quad t > 0.$$

However, the integral is usually not easy to be evaluated for complicated $\hat{f}$. Hence, some numerical methods have been developed to recover $f(t)$ based on the approximation to the integral through trapezoidal rules or others.

2.10.1 Post-Widder Formula (Widder, 1941)

This celebrated formula gives the inverse Laplace transform in analytic form and was first proved for special case by Post in 1930 and generalised by Widder in 1934. The implementation requires derivatives of all orders for $\hat{f}$ on the entire positive half-line and the inversion is given by

$$f(t) = \lim_{k\to\infty} \frac{(-1)^k}{k!} \hat{f}^{(k)}\left(\frac{k}{t}\right)\left(\frac{k}{t}\right)^{k+1}.$$

The convergence of the limit is slow and is often not efficient to be used although it has simpler form compares to other methods.

2.10.2 Gaver-Stehfest Formula (Abate & Whitt, 2006; Masol & Teugels, 2010)

This algorithm is a combination of Gaver's formula with acceleration scheme for convergence proposed by Stehfest. The formula depends on a positive integer M to control the precision of results as follows

$$f(t) \simeq f_g(t, M) = \frac{\log 2}{t} \sum_{k=1}^{2M} \zeta_k \hat{f}\left(\frac{k \log 2}{t}\right),$$

where

$$\zeta_k = (-1)^{M+k} \sum_{j=\lfloor (k+1)/2 \rfloor}^{\min(k,M)} \frac{j^{M+1}}{M!} \binom{M}{j} \binom{2j}{j} \binom{j}{k-j}.$$

The selection of $M = \lceil 1.1n \rceil$ will return a result with n significant digits when the computation is run on a computer with precision $\lceil 2.2M \rceil$. Here, $\lfloor k \rfloor$ represents the greatest integer less than or equal to k whereas $\lceil k \rceil$ represents the smallest integer bigger than or equal to k. For transforms that have singularities only on the negative real axis and function f that are infinitely differentiable for $t > 0$, the inversion is reported to have small relative error which approximately $10^{-0.9M}$.

2.10.3 The Fourier Series Method (Abate & Whitt, 1999, 2006)

This inversion formula is derived as the trapezoidal approximation to the real part of the Bromwich integral. The approximant is given by

$$f_e(t,M) = \frac{10^{M/3}}{t} \sum_{k=0}^{2M} \eta_k \operatorname{Re}\left(\hat{f}\left(\frac{\beta_k}{t} \right) \right),$$

where

$$\beta_k = \frac{M \log 10}{3} + \pi i k \,, \quad \eta_k = (-1)^k \xi_k$$

with $i = \sqrt{-1}$ and

$$\xi_0 = \frac{1}{2}, \;\; \xi_k = 1 \text{ for } 1 \le k \le M \,, \;\; \xi_{2M-k} = \xi_{2M-k+1} + 2^{-M} \binom{M}{k} \text{ for } 0 < k < M \,, \;\; \xi_{2M} = 2^{-M} \,.$$

Similar to the Gaver-Stehfest algorithm, the selection of $M = \lceil 1.7n \rceil$ returns a result with n significant digits. The accuracy of this algorithm is reported to be satisfactory in most trial experiment although it takes a longer computer time to obtain the results.

CHAPTER 3 FURTHER PROPERTIES AND PARAMETER ESTIMATION FOR THE HURWITZ-LERCH ZETA DISTRIBUTION

3.0 Introduction

In this chapter, further properties of the HLZD, including stochastic orderings and tail behaviour, are considered. A pgf based estimation method suggested in Sim & Ong (2010) will be modified to illustrate its feasibility in handling grouped data. Section 3.1 introduces the HLZD while Section 3.2 discusses the stochastic orderings among the members in this family. The tail behaviour of the HLZD is examined in Section 3.3 and the last section considers parameter estimation with the pgf based method illustrated by a number of datasets.

3.1 The HLZD

The HLZD has pmf

$$p_k = \frac{1}{T(\theta,s,a)}\frac{\theta^k}{(k+a)^{s+1}}, \quad k=1,2,...;$$

where $0<\theta\le 1$; $s\in\mathbb{R}$; $a>-1$ and $s>0$ when $\theta=1$. Here, we use the standard notation

$$T(\theta,s,a)=\sum_{k=1}^{\infty}\frac{\theta^k}{(k+a)^{s+1}}=\theta\Phi(\theta,s+1,a+1)$$

and

$$\Phi(z,s,a)=\sum_{k=0}^{\infty}\frac{z^k}{(k+a)^s},$$

is the Lerch transcendent defined for $a \notin Z_0^-$ and $s \in \mathbb{C}$ when $|z|<1$ or $\operatorname{Re} s>1$ when $|z|=1$. If $\operatorname{Re}(a)>0$ and either $\operatorname{Re}(s)>0$ when $|z|\le 1$, $(z \ne 1)$ or $\operatorname{Re}(s)>1$ when $z=1$, the integral representation

$$\Phi(z,s,a)=\frac{1}{\Gamma(s)}\int_0^\infty \frac{e^{-(a-1)t}}{e^t-z}t^{s-1}dt \tag{3.1}$$

holds (Erdélyi *et al.*, 1953; Vol I, p. 27). Furthermore, by using (3.1), the series representation of the Lerch transcendent can be extended to $\Phi(z,s,a)$ which is an analytic function of *z* for $z \in \mathbb{C}\setminus[1,\infty)$ provided $\operatorname{Re}(s)>0$ and $\operatorname{Re}(a)>0$.

Let *Y* be an HLZ random variable. Its pgf can be written as

$$G_Y(t)=E\left[t^Y\right]=\frac{t\Phi\left(\theta t,s+1,a+1\right)}{\Phi\left(\theta,s+1,a+1\right)}, \text{ for } 0<\theta t\le 1. \tag{3.2}$$

The mean of the distribution is $E[Y]=\dfrac{T(\theta,s-1,a)}{T(\theta,s,a)}-a$. Some of the common distributions in the HLZ family are shown in Table 3.1.

Table 3.1

Some common distributions in the HLZ family

Distribution	θ	s	a
Lotka	1	1	0
Riemann zeta	1	s	0
Zipf-Mandelbrot	1	s	a
Good	θ	s	0
Logarithmic	θ	0	0

Under this definition, the HLZD has a larger parameter space compares to the definition in Gupta *et al.* (2008), where the authors allowed $s \ge 0$, $0 \le a \le 1$ and $0<\theta \le 1$. Although the relaxation of parameters has the advantage of increasing the flexibility of the model in empirical modelling, some of the properties that are derived in their paper no longer hold in general as indicated in the following.

- The HLZD hence defined may not be log-convex.

$$\frac{\theta^{k+1}}{(k+a+1)^{s+1}}\frac{\theta^{k-1}}{(k+a-1)^{s+1}}$$
$$=\frac{\theta^{2k}}{[(k+a)^2-1]^{s+1}}\begin{cases}\geq\left[\dfrac{\theta^k}{(k+a)^{s+1}}\right]^2, s\geq -1\\ <\left[\dfrac{\theta^k}{(k+a)^{s+1}}\right]^2, s< -1\end{cases}$$

- The mode of the HLZD may not be 1. For $s \geq -1$,

$$\frac{p_{k+1}}{p_k}=\theta(1-\frac{1}{k+1+a})^{s+1}<1.$$

If $s < -1$, then

$$\frac{p_{k+1}}{p_k}=1 \Rightarrow (1+\frac{1}{k+a})^{-(s+1)}=\frac{1}{\theta}\Rightarrow \frac{\sqrt[-(s+1)]{\theta}}{1-\sqrt[-(s+1)]{\theta}}-a=k.$$

When $\frac{\sqrt[-(s+1)]{\theta}}{1-\sqrt[-(s+1)]{\theta}}-a$ is not positive, it indicates that the mode is at 1 or else the mode is at $\max\left[1,\left[\frac{\sqrt[-(s+1)]{\theta}}{1-\sqrt[-(s+1)]{\theta}}-a\right]\right]$ as shown in Figure 3.1.

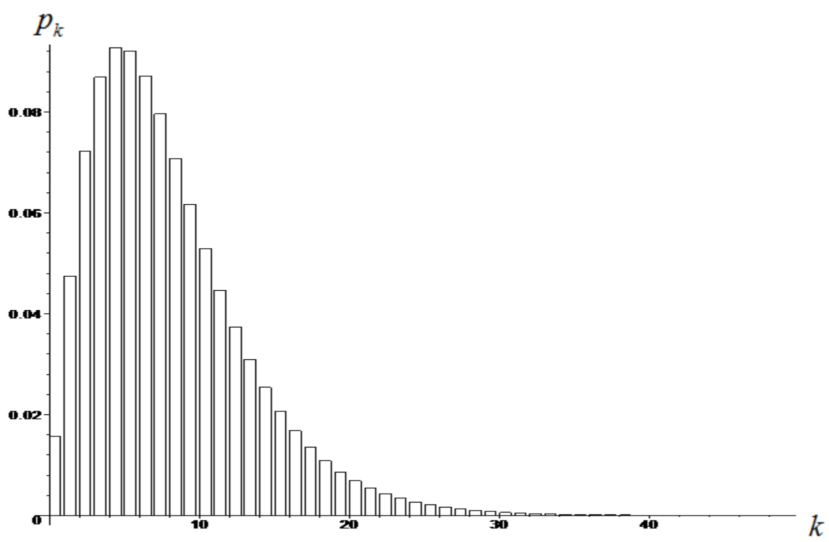

Figure 3.1

Pmf of the HLZD with $\theta = 0.73660$, $a = -0.34031$ and $s = -2.5370$

3.2 Likelihood Ratio Order in the HLZ Family

Since the HLZ family consists of many distributions, the stochastic orderings shed light on their probabilistic behaviour relative to each other. Establishing the orderings in the HLZ family is also useful to infer orderings in the Poisson-HLZD in the next chapter. Figure 2.1 has clearly shown the hierarchical relationships among the stochastic orderings with the likelihood ratio order to be the strongest in the sense that it implies all other orders shown in that figure.

The likelihood ratio order of the members in the HLZ family will be considered in three different cases depending on the number of different parameters between the models.

Case 1: When only one parameter is different

(a) If $\theta_1 < \theta_2$, then $\dfrac{\dfrac{1}{T(\theta_1,s,a)}\dfrac{\theta_1^k}{(k+a)^{s+1}}}{\dfrac{1}{T(\theta_2,s,a)}\dfrac{\theta_2^k}{(k+a)^{s+1}}} = \dfrac{T(\theta_2,s,a)}{T(\theta_1,s,a)}\left(\dfrac{\theta_1}{\theta_2}\right)^k$ is a decreasing function in *k*. Hence, $HLZ(\theta_1,s,a) <_{lr} HLZ(\theta_2,s,a)$.

(b) When $a_1 < a_2$, the ratio $\dfrac{\dfrac{1}{T(\theta,s,a_1)}\dfrac{\theta^k}{(k+a_1)^{s+1}}}{\dfrac{1}{T(\theta,s,a_2)}\dfrac{\theta^k}{(k+a_2)^{s+1}}} = \dfrac{T(\theta,s,a_2)}{T(\theta,s,a_1)}\left(1+\dfrac{a_2-a_1}{k+a_1}\right)^{s+1}$ is a decreasing function in *k* when $s \geq -1$. Hence, $HLZ(\theta,s,a_1) <_{lr} HLZ(\theta,s,a_2)$ for $s \geq -1$.

(c) If $s_2 < s_1$, then $\dfrac{\dfrac{1}{T(\theta,s_1,a)}\dfrac{\theta^k}{(k+a)^{s_1+1}}}{\dfrac{1}{T(\theta,s_2,a)}\dfrac{\theta^k}{(k+a)^{s_2+1}}} = \dfrac{T(\theta,s_2,a)}{T(\theta,s_1,a)}(k+a)^{s_2-s_1}$ is a decreasing function in *k*, implying $HLZ(\theta,s_1,a) <_{lr} HLZ(\theta,s_2,a)$.

Case 2: When two parameters are different

(a) Consider θ is common, then $\dfrac{\dfrac{1}{T(\theta,s_1,a_1)}\dfrac{\theta^k}{(k+a_1)^{s_1+1}}}{\dfrac{1}{T(\theta,s_2,a_2)}\dfrac{\theta^k}{(k+a_2)^{s_2+1}}}=\dfrac{T(\theta,s_2,a_2)}{T(\theta,s_1,a_1)}\dfrac{(k+a_2)^{s_2+1}}{(k+a_1)^{s_1+1}}$.

Clearly the ratio is a decreasing function of *k* if

$$\frac{\left(k+a_2\right)^{s_2+1}}{\left(k+a_1\right)^{s_1+1}}>\frac{\left(k+1+a_2\right)^{s_2+1}}{\left(k+1+a_1\right)^{s_1+1}},$$

that is,

$$\left(\frac{k+a_2}{k+1+a_2}\right)^{s_2+1}>\left(\frac{k+a_1}{k+1+a_1}\right)^{s_1+1}.$$

Hence, $HLZ(\theta,s_1,a_1)<_{lr} HLZ(\theta,s_2,a_2)$ if, and only if, $s_2\le s_1$ and $a_2\ge a_1$.

(b) If a is common, then $\dfrac{\dfrac{1}{T(\theta_1,s_1,a)}\dfrac{\theta_1^k}{(k+a)^{s_1+1}}}{\dfrac{1}{T(\theta_2,s_2,a)}\dfrac{\theta_2^k}{(k+a)^{s_2+1}}}=\dfrac{T(\theta_2,s_2,a)}{T(\theta_1,s_1,a)}\left(\dfrac{\theta_1}{\theta_2}\right)^k(k+a)^{s_2-s_1}$. The ratio is a decreasing function of *k* if $\theta_1\le\theta_2$ and $s_2\le s_1$. It is a strictly decreasing function of *k* if $\theta_1<\theta_2$ and $s_2<s_1$. Hence, $HLZ(\theta_1,s_1,a)<_{lr} HLZ(\theta_2,s_2,a)$.

(c) When s is common, then $\dfrac{\dfrac{1}{T(\theta_1,s,a_1)}\dfrac{\theta_1^k}{(k+a_1)^{s_1+1}}}{\dfrac{1}{T(\theta_2,s,a_2)}\dfrac{\theta_2^k}{(k+a_2)^{s_2+1}}}=\dfrac{T(\theta_2,s,a_2)}{T(\theta_1,s,a_1)}\left(\dfrac{\theta_1}{\theta_2}\right)^k\left(\dfrac{k+a_2}{k+a_1}\right)^{s+1}$

and the ratio is a decreasing function of k if $\theta_1\le\theta_2$ and $a_1\ge a_2$. Hence $HLZ(\theta_1,s,a_1)<_{lr} HLZ(\theta_2,s,a_2)$.

Case 3: When all three parameters are different

Consider the ratio $\dfrac{\dfrac{1}{T(\theta_1,s_1,a_1)}\dfrac{\theta_1^k}{(k+a_1)^{s_1+1}}}{\dfrac{1}{T(\theta_2,s_2,a_2)}\dfrac{\theta_2^k}{(k+a_2)^{s_2+1}}}=\dfrac{T(\theta_2,s_2,a_2)}{T(\theta_1,s_1,a_1)}\dfrac{(k+a_2)^{s_2+1}}{(k+a_1)^{s_1+1}}\left(\dfrac{\theta_1}{\theta_2}\right)^k$. By applying the results in Case 2, this ratio is a decreasing function of k if $\theta_1 \le \theta_2$, $a_1 \ge a_2$ and $s_2 \le s_1$.Thus $HLZ(\theta_1,s_1,a_1) <_{lr} HLZ(\theta_2,s_2,a_2)$.

Based on the discussion above, some of the distributions in Table 3.1 can be ranked. For example, by assuming $0 < s < 1$, $a > 0$ and $\theta < 1$ to have same values in different models, we have

Good $\le_{lr}$ LD $\le_{lr}$ Lotka (if $2\theta \le 1$) $\le_{lr}$ zeta $\le_{lr}$ Zipf-Mandelbrot.

3.3 Tail Behaviour of HLZD

The rate of decay of a probability distribution determines whether the distribution has a short or long tail. Usually this can be assessed by finding the limit of the ratio of consecutive probabilities. For the HLZD,

$$\lim_{k\to\infty}\frac{p_{k+1}}{p_k}=\lim_{k\to\infty}\left(\frac{a+k}{a+k+1}\right)^{s+1}\theta=\theta.$$

When $\theta=1$, $p_k \propto (a+k)^{-(s+1)}$ giving the distribution the long-tail property. It shows that the HLZD exhibits either exponential decay or power law decay.

The upper tail probability of a random variable is often of interest. For the HLZD, define $T_{HLZ}(x)=\sum_{i=x}^{\infty}p_i$. The tail probability can be expressed in terms of the Lerch transcendent as follows.

$$
\begin{aligned}
T_{HLZ}(x) &= \frac{1}{T(\theta,s,a)}\sum_{i=x}^{\infty}\frac{\theta^i}{(a+i)^{s+1}} \\
&= \frac{\theta^x}{T(\theta,s,a)}\sum_{i=0}^{\infty}\frac{\theta^i}{(a+x+i)^{s+1}} \\
&= \frac{\theta^{x-1}\Phi(\theta,s+1,a+x)}{\Phi(\theta,s+1,a+1)}
\end{aligned}
$$

Compare this to the asymptotic upper tail probability of the generalised logarithmic distribution (GLD) in Khang & Ong (2007), which takes the form $T_{GLD}(x) \simeq -\frac{(q+2\sqrt{pr})^{x+1/2}}{2(\pi\sqrt{pr})^{1/2}\log p}\Phi(q+2\sqrt{pr},\frac{3}{2},x)$. By taking $\theta = q+2\sqrt{pr}$, $s = 0.5$ and $a = 0$, it is interesting to see that the asymptotic ratio of the tail probabilities of both distributions can be made independent of *x*, that is

$$
\frac{T_{GLD}(x)}{T_{HLZ}(x)} \simeq -\frac{\Phi(q+2\sqrt{pr},1.5,1)(q+2\sqrt{pr})^{3/2}}{2(\pi\sqrt{pr})^{1/2}\log p}.
$$

Figure 3.2 shows the graph of the above ratio over all possible combinations of *p*, *q* and *r*, we see that the ratio is within the range from 1 to less than 4.5.

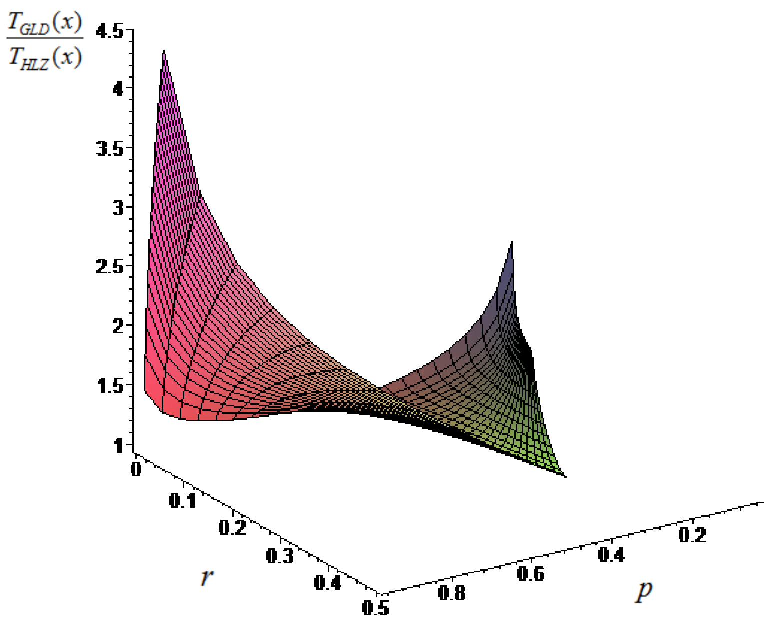

Figure 3.2

Ratio of tail probabilities of GLD to HLZD for all possible combinations of *p* and *r*

3.4 Estimation of Parameters in the HLZD

In Gupta *et al.* (2008), the ML equations for the parameters in the HLZD are derived and the authors pointed out that the equations result in method of moment equations with different functional forms. However, closed form solution of these equations is unavailable and a global optimization technique is needed to find the ML estimates. In Sim & Ong (2010), an estimation procedure by minimizing the pgf-based divergence statistic

$$D(\boldsymbol{\theta},\alpha,n)=\int_0^1\left(f_n(t)^{\alpha}-g(t;\boldsymbol{\theta})^{\alpha}\right)^2\beta(t)dt,\quad 0<\alpha\le 1;$$

is proposed, where $f_n(t)=\frac{1}{n}\sum_{i=1}^{n}t^{x_i}$ is the empirical pgf based on n observations $x_1,x_2,...,x_n$, $g(t;\boldsymbol{\theta})$ is the pgf with parameter vector $\boldsymbol{\theta}$ and $\beta(t)$ is a weight function. For simplicity and shorter execution time, the statistics $D_1(\boldsymbol{\theta},n)=\int_0^1\left(f_n(t)-g(t;\boldsymbol{\theta})\right)^2dt$ and $D_2(\boldsymbol{\theta},n)=\int_0^1\left(\sqrt{f_n(t)}-\sqrt{g(t;\boldsymbol{\theta})}\right)^2dt$ suggested in the paper will be used in this section to estimate the parameters of the HLZD.

A number of datasets will be used to demonstrate the parameter estimation and goodness-of-fit: (1) the distribution of number of moth species represented by n individuals in a sample from the lightly logged rainforest (Khang & Ong, 2007); (2) distribution of 1534 biologists according to the number of research papers to their credit (Jain & Gupta, 1973); (3) number of boards that contains at least one sowbug (Doray & Luong, 1997); (4) Corbet's Malayan butterflies data (Gupta *et al.*, 2008). These datasets have been fitted to various generalisations of the LD and will be used to illustrate the competency of the HLZD.

Algorithm 2.1 has been used to minimize either $D_1(\boldsymbol{\theta},n)$ or $D_2(\boldsymbol{\theta},n)$ in the following estimations with $g(t;\boldsymbol{\theta})$ equals to the HLZ pgf in (3.2). To accelerate computations, the integrals involved are evaluated by a 6-point Gaussian quadrature. The performances of these estimates can be assessed based on the *p*-value.

The moth data from Chey (2002) will be used as the first illustrative example. In Khang & Ong (2007), this distribution of number of moth species represented by n individuals in a sample from the lightly logged rainforest in Sabah, Malaysia was fitted to a GLD arises as the cluster size distribution in the Poisson-stopped sum representation of the inverse trinomial distribution. This GLD has interesting feature such as oscillatory behaviour. Using the ML estimates obtained from simulated annealing algorithm, their χ^2 statistic shows a good fit with *p*-value of 0.77. Following the discussion in Section 3.3, the HLZD is expected to perform equally well for this set of data. In Table 3.2, using the same grouping, the results obtained by minimizing $D_1(\boldsymbol{\theta}, n)$ and $D_2(\boldsymbol{\theta}, n)$ show adequate fit with a slightly smaller *p*-value of 0.73 and 0.76 respectively.

Table 3.2

Number of moth species represented by n individuals in a sample from the lightly logged rainforest fitted to the HLZD by minimizing $D_1(\boldsymbol{\theta}, n)$ and $D_2(\boldsymbol{\theta}, n)$

N	Number of species	$D_1(\boldsymbol{\theta}, n)$	$D_2(\boldsymbol{\theta}, n)$
1	140	139.30	139.68
2	36	38.01	37.21
3	17	17.61	17.58
4	13	9.81	9.93
5	6	6.00	6.13
6	2	3.89	3.99
7	4	2.62	2.69
8	2	1.82	1.86
9	2	1.29	1.31
10	0	0.93	0.94
11	1	0.67	0.68
12	0	0.50	0.50
13	0	0.37	0.37
14	0	0.28	0.27
15	1	0.21	0.20
>15	0	0.70	0.64
Total	224	224	224
df		3	3
χ^2		1.29	1.17
p-value		0.7305	0.7604

$D_1(\boldsymbol{\theta}, n): \theta = 0.8237;\ s = 0.2416;\ a = -0.3029$

$D_2(\boldsymbol{\theta}, n): \theta = 0.8061;\ s = 0.0763;\ a = -0.4438$

The second dataset shows the distribution of 1534 biologists according to the number of research papers to their credit in The Review of Applied Entomology, Vol 24. This data has been fitted by Jain & Gupta (1973) to their GLD with a *p*-value of 0.238. Later, Tripathi & Gupta (1988) fitted the same set of data to another generalisation of the LD developed from taking limit in the truncated mixed generalised Poisson distribution proposed in Tripathi & Gupta (1984). They managed to improve the fit by using estimates from the method of moments, obtaining a *p*-value of 0.323. Our trials, on minimizing $D_1(\boldsymbol{\theta}, n)$ and $D_2(\boldsymbol{\theta}, n)$ are marginal, with p-values of 0.1359 and 0.1117 respectively, see Table 3.3.

Table 3.3

Distribution of 1534 biologists according to the number of research papers to their credit fitted to the HLZD by minimizing $D_1(\boldsymbol{\theta}, n)$ and $D_2(\boldsymbol{\theta}, n)$

Number of papers per author	Observed Frequencies	$D_1(\boldsymbol{\theta}, n)$	$D_2(\boldsymbol{\theta}, n)$
1	1062	1056.52	1057.44
2	263	283.80	282.75
3	120	104.47	103.84
4	50	44.90	44.76
5	22	21.17	21.25
6	7	10.61	10.77
7	6	5.56	5.71
8	2	3.01	3.14
9	0	1.67	1.77
10	1	0.95	1.02
11	1	0.55	0.60
>11	0	0.80	0.93
Total	1534	1534	1534
df		4	4
χ^2		7.00	7.50
p-value		0.1359	0.1117
θ		0.6814	0.7050
s		0.8443	1.0184
a		0.5232	0.6217

As shown in Table 3.1, the Good distribution is a special case in the HLZ family. Doray & Luong (1997) compared the efficiency of the ML estimator and a quadratic distance estimator that is constructed based on least square method for the Good family. The authors also pointed out that the usage of the quadratic distance estimator instead of ML estimator can be regarded as trading off efficiency versus robustness. By fitting of the data of number of boards that contains at least one sowbug to the Good distribution, their reported p-values for the two methods of estimations are 0.5351 and 0.117

respectively. Minimizing $D_1(\boldsymbol{\theta}, n)$ and $D_2(\boldsymbol{\theta}, n)$ gives p-values of 0.4819 and 0.4543 respectively, which are also comparable to the ML estimates, see Table 3.4.

Table 3.4

Number of boards that contains at least one sowbug fitted to Good's distribution by minimizing $D_1(\boldsymbol{\theta}, n)$ and $D_2(\boldsymbol{\theta}, n)$

Number of sowbug	Observed	$D_1(\boldsymbol{\theta}, n)$	$D_2(\boldsymbol{\theta}, n)$
1	28	26.91	27.24
2	14	16.29	16.12
3	11	11.32	11.12
4	8	8.33	8.16
5	11	6.32	6.20
6	2	4.89	4.81
7	3	3.84	3.79
8	3	3.04	3.02
9	3	2.43	2.43
10	3	1.95	1.96
11	2	1.58	1.59
12	0	1.28	1.30
13	1	1.04	1.07
14	2	0.85	0.88
15	1	0.69	0.72
16	0	0.57	0.60
17	2	0.47	0.49
>18	0	2.21	2.49
Total	94	94.00	94.00
df		7	7
χ^2		6.51	6.76
p-value		0.4819	0.4543
θ		0.8450	0.8551
s		-0.5186	-0.4691

The LD has been proposed to describe data on insect and species counts that exhibit long-tailed pattern (Fisher *et al.*, 1943). However, the LD seems to be too restrictive for extra-long tail datasets as shown in Khang & Ong (2007). Another long tail example is given in Table 3.5 that consists of the Corbet's Malayan butterflies data fitted to the HLZD Gupta *et al.* (2008) with ML estimates. Since the data is not primary, an immediate problem in implementing the pgf based estimation method to this dataset is that classes larger than 24 has been grouped into one class. If we ignore the last class completely and minimizing the divergence between the ‘empirical pgf constructed only from the first 24 classes’ and the ‘pgf that only consists of 24 terms’, the deviance is very large. On the other hand, if we assume the 119 grouped observations to be uniformly distributed with one observation in each class after 24, the fit is still not satisfactory. The arbitrariness in handling the grouped data is a disadvantage in the pgf based estimation. This drawback can be removed by using a more objective data graduation scheme as shown below.

When s is small, the ratios between the expected frequencies from consecutive classes in the HLZD are approximately θ. In the global optimisation search for optimal solution, for each new random search of θ, we make use of the observation from class 24 and construct the sequence $\{[3\theta],[3\theta^2],...\}$, where $[x]$ refer to x rounded to the nearest positive integer. The process continues until the sum of the sequence is 119. This sequence will be used to represent the observations in classes larger than 24 when constructing the empirical pgf. Using this augmented empirical pgf, the estimates obtained have greatly improved the fit as shown in the last column in Table 3.5. However, the smaller chi-square value compared to that in Gupta *et al.* (2008) could also be due to the extended parameter space that includes negative values for s.

Table 3.5

Corbet's Malayan butterflies data fitted to the HLZD based on three different ways to construct the empirical pgf: a) by truncating the tail, b) by assuming uniformly distributed observations at the tail and c) by redistribute the grouped observations according to the expected ratios in the HLZD

Individuals/Species	Number of Species	Truncating the tail at 24	Assuming one observation in each class after 24	Graduation of data using HLZ ratio
1	118	119.14	117.75	118.93
2	74	67.25	68.89	65.64
3	44	45.87	47.33	45.67
4	24	34.36	35.42	35.03
5	29	27.22	27.96	28.37
6	22	22.39	22.88	23.78
7	20	18.91	19.24	20.42
8	19	16.30	16.50	17.84
9	20	14.27	14.38	15.80
10	15	12.65	12.69	14.14
11	12	11.33	11.31	12.77
12	14	10.23	10.18	11.61
13	6	9.31	9.23	10.62
14	12	8.52	8.42	9.76
15	6	7.85	7.72	9.02
16	9	7.26	7.12	8.36
17	9	6.74	6.59	7.78
18	6	6.28	6.13	7.25
19	10	5.87	5.72	6.79
20	10	5.51	5.35	6.36
21	11	5.18	5.02	5.98
22	5	4.88	4.72	5.63
23	3	4.61	4.45	5.31
24	3	4.37	4.21	5.02
>24	119	143.70	140.79	122.13
Total	620	620	620	620
df		21	21	21
χ^2		31.26	31.88	20.32
p-value		0.0695	0.0602	0.5011
θ		0.9924	0.9942	0.9803
s		0.1340	0.2339	-0.1311
a		0.5508	0.8630	0.0674

CHAPTER 4 THE POISSON-HURWITZ-LERCH ZETA DISTRIBUTION

4.0 Introduction

In this chapter, the Poisson-stopped sum of HLZD is proposed as a new generalisation of the NBD. Section 4.1 introduces the construction of the models and some basic results are derived. Probabilistic properties of the model such as infinite divisibility, discrete self-decomposability, and so on are studied in Section 4.2 followed by Section 4.3, in which the Poisson-HLZD is proved to have a MP formulation and some numerical examples of inversions to recover the mixing distributions are given. Section 4.4 establishes the stochastic orderings between the members in the Poisson-HLZ family. Finally, in Section 4.5, application of the model in data-fitting are illustrated.

4.1 The Poisson-HLZD

Consider a pgf of the form

$$G_X(t) = \exp\{\lambda[G_Y(t) - 1]\}, \tag{4.1}$$

where $G_Y(t)$ is the pgf of the HLZD in (3.2). A random variable X with pgf $G_X(t)$ is said to have a Poisson-HLZD and it can be expressed as

$$X = \sum_{i=0}^{N} Y_i \tag{4.2}$$

where N is a Poisson random variable with mean λ and Y_i are independent HLZ random variables having common distribution as Y. When s = 0, a = 0, $G_Y(t)$ reduces to the pgf of a LD and $G_X(t)$ becomes a pgf for the NBD.

4.1.1 Probability Mass Function of the Poisson-HLZD

Let p_k be the probability function of the Poisson-HLZD. By the definition of pgf,

$$e^{-\lambda} \exp\left[\frac{\lambda t \Phi(\theta t, s+1, a+1)}{\Phi(\theta, s+1, a+1)}\right] = \sum_{j=0}^{\infty} p_j t^j . \qquad (4.3)$$

Differentiating both sides of (4.3) with respect to t, we get

$$\sum_{j=0}^{\infty} j p_j t^{j-1}$$

$$= \frac{\lambda e^{-\lambda}}{\Phi(\theta, s+1, a+1)} \exp\left[\frac{\lambda t \Phi(\theta t, s+1, a+1)}{\Phi(\theta, s+1, a+1)}\right]\left(\Phi(\theta t, s+1, a+1) + t\frac{d}{dt}\left(\Phi(\theta t, s+1, a+1)\right)\right)$$

$$= \frac{\lambda}{\Phi(\theta, s+1, a+1)}\left(\Phi(\theta t, s+1, a+1) + t\frac{d}{dt}\left(\Phi(\theta t, s+1, a+1)\right)\right)\left(\sum_{j=0}^{\infty} p_j t^j\right)$$

$$= \frac{\lambda}{\Phi(\theta, s+1, a+1)}\left(\sum_{j=1}^{\infty} \frac{j(\theta t)^{j-1}}{(a+j)^{s+1}}\right)\sum_{j=0}^{\infty} p_j t^j$$

By comparing the coefficient of t^n from both sides, we obtain the recurrence formula

$$(n+1) p_{n+1} = \frac{\lambda}{\Phi(\theta, s+1, a+1)} \sum_{j=0}^{n} \frac{(j+1)\theta^j}{(a+1+j)^{s+1}} p_{n-j} . \qquad (4.4)$$

Since the HLZD has support on the positive integers, we can use the probability of zero occurrence $p_0 = e^{-\lambda}$ to initiate the calculation in (4.4).

In a more general setting, pgf (4.1) can also be regarded as a special case of (2.12) and the recurrence formula (4.4) followed by Theorem 2.5. The details of this approach are given in Appendix A. Another approach to derive the pmf in the context of actuarial science is shown in Appendix B.

The Poisson-HLZD can take various shapes as shown in the following figures. Figure 4.1 also suggests the possibility of using Poisson-HLZD in modelling zero-inflated data when there is no trivial reason to assume the data coming from different subpopulations.

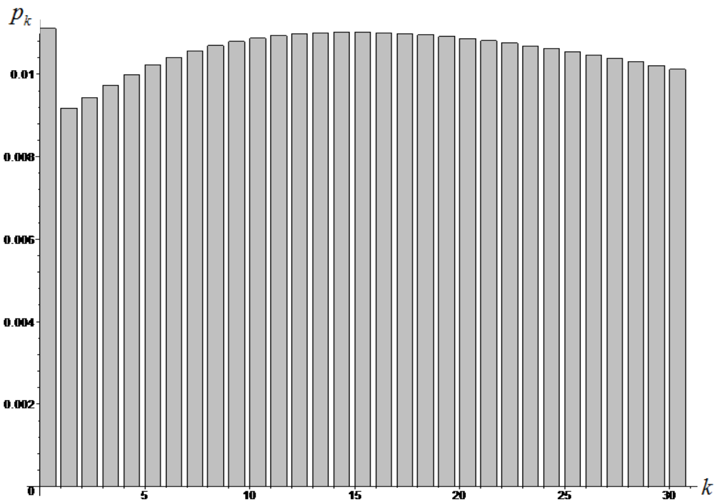

Figure 4.1 Pmf of Poisson-HLZD with $\lambda = 4.5; \theta = 0.98; a = 0.7, s = 0.01$

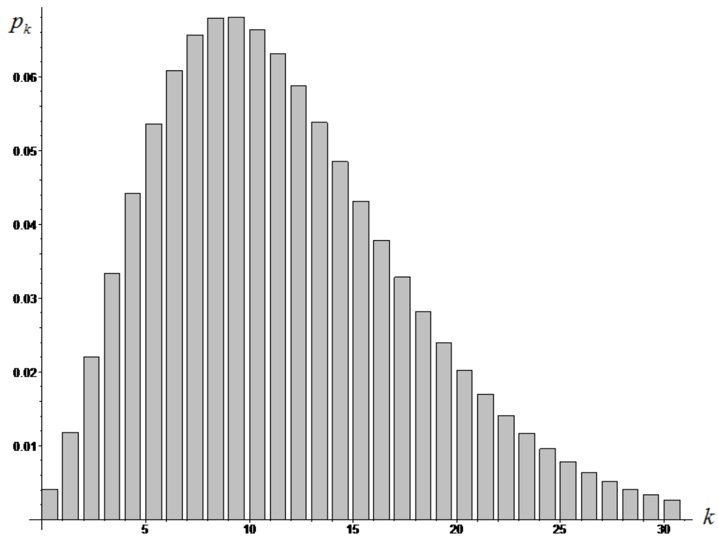

Figure 4.2 Pmf of Poisson-HLZD with $\lambda = 5.5; \theta = 0.7; a = 0.7, s = 0.1$

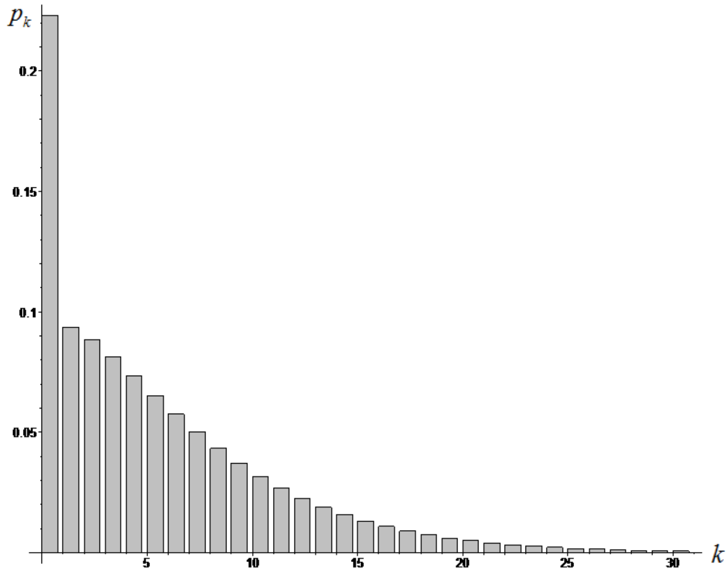

Figure 4.3 Pmf of Poisson-HLZD with $\lambda = 5.5; \theta = 0.7; a = 0.7, s = -1.1$

4.1.2 Moment-Ratio Diagram

The mean and variance of the Poisson-HLZD can be obtained using (2.8) and the first two moments of the HLZD from Gupta *et al.* (2008).

$$E[X] = \lambda\left(\frac{T(\theta, s-1, a)}{T(\theta, s, a)} - a\right)$$

$$Var(X) = \lambda[\frac{T(\theta, s-2, a)}{T(\theta, s, a)} - 2a\frac{T(\theta, s-1, a)}{T(\theta, s, a)} + a^2]$$

The index of dispersion (ID), defined as the ratio of variance to the mean, is a useful indicator in model selection. A distribution which has ID > 1 is said to be over-dispersed with respect to the Poisson distribution.

Property 1 The Poisson-HLZD is over-dispersed.

Proof:

Let Y be a HLZ random variable. Since the support of the HLZD is the set of positive integers, we must have $E[Y^2] > E[Y]$. The ID of Poisson-HLZD is given by $\frac{\lambda E[Y^2]}{\lambda E[Y]} > 1$. ■

As seen from the proof, Property 1 is also a general property of the PSSD with cluster size distribution on $\mathbb{N}$.

Comparisons of probability models can be done in various ways depending on the objectives of studies. When flexibility of the models is concerned, the moment-ratio diagram or skewness-kurtosis plot, gives an insightful picture to the relative versatility of the models. Vargo *et al.* (2010) provide the moment-ratio diagrams for some common distributions and illustrate their applications in model selection. Skewness-kurtosis plot of well-known MPD including the Sichel, Delaporte, Poisson-Tweedie and some zero-inflated models are shown in Rigby *et al.* (2008). For easy comparison with these models, the skewness-kurtosis plot for the Poisson-HLZD is superimposed to Fig. 5 in the paper as shown in Figure 4.5.

From the discussion in Section 2.5, the skewness and kurtosis can be calculated based on formula (2.2) in Gupta (1974) by treating the HLZD as a special case of the modified power series distribution. The system of simultaneous equations

$$\begin{cases} \lambda\left(\dfrac{T(\theta,s-1,a)}{T(\theta,s,a)}-a\right)=1 \\ \lambda[\dfrac{T(\theta,s-2,a)}{T(\theta,s,a)}-2a\dfrac{T(\theta,s-1,a)}{T(\theta,s,a)}+a^2]=2 \end{cases} \tag{4.5}$$

is solved for λ and s for some given pairs (θ,a). To be specific, in order to obtained the lower boundary in Figure 4.4, the value of a is fixed at -0.99999 with θ changes from 0.05 to 0.85. The points on the upper boundary are obtained by fixing $\theta = 0.99$ and a changing from 0.07 to 9.9 with a step size of 0.1. The values of a near –1 are not considered because the corresponding combinations produce very large kurtoses that distort the graphical presentation. The values of θ are chosen such that the computational time is acceptable. Construction of one boundary could take a few days due to the difficulties in solving (4.5). To avoid the case of not plotting the true boundaries, points of random combinations of parameters are generated as shown in Figure 4.4. All the points fall into the region between the two boundaries. In Figure 4.5, the boundaries shown in Figure 4.4 are superimposed on Fig 5. in Rigby *et al.* (2008) for easy comparison. It is observed that the region corresponding to the Poisson-HLZD includes both the Sichel and Poisson-Tweedie distributions and also cover part of the two regions representing zero-inflated Sichel and Poisson-shifted generalised inverse Gaussian. Part of the Maple programme is shown in Appendix C.

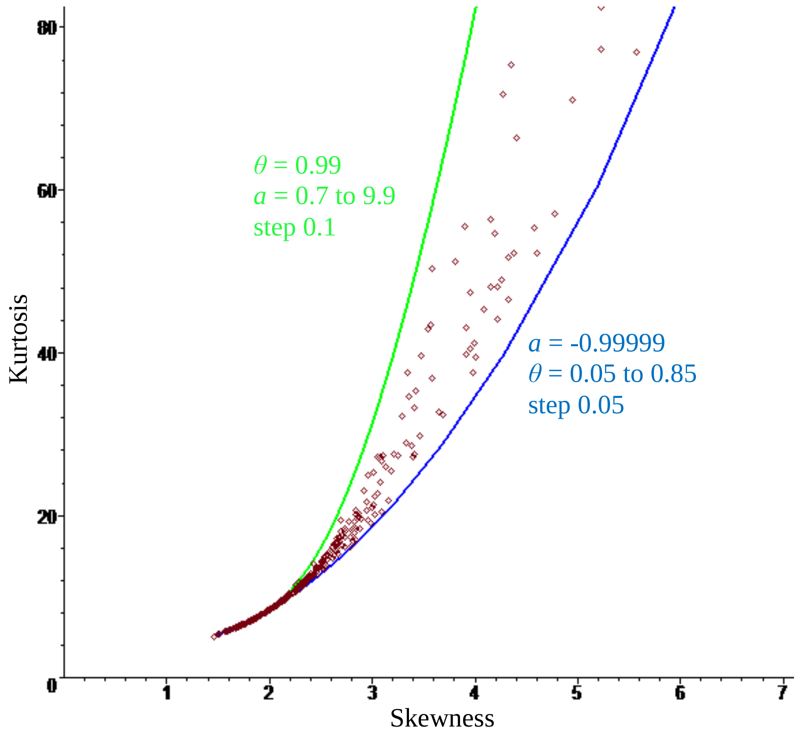

Figure 4.4

Moment-ratio plot for Poisson-HLZD with mean =1 and variance = 2

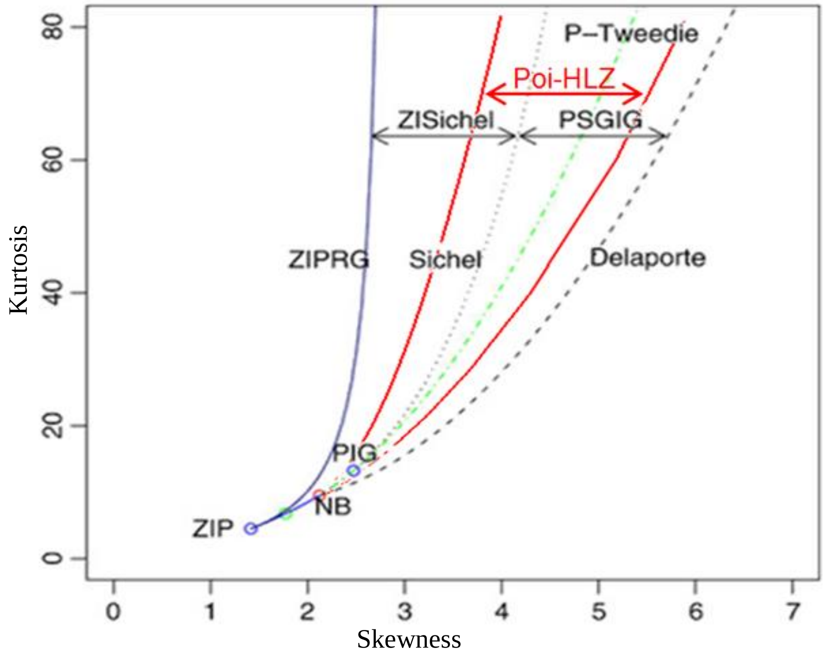

Figure 4.5

Comparison of moment-ratio plot for various MPDs

4.2 Probabilistic Properties of the Poisson-HLZD

Some modern probabilistic structures of the Poisson-HLZD which may be used to prove some other properties of the distribution are examined. Some of these properties may only be true under some restrictions in parameters.

4.2.1 Infinite Divisibility

By Theorem 2.6, the pgf in (4.1) clearly shows that the Poisson-HLZD is infinitely divisible. Hence, the recursive formula for the pmf of the Poisson-HLZD can also be obtained easily as a consequence of infinite divisibility.

From either (2.14) or (2.16), the canonical measure r_k of the Poisson-HLZD is given in terms of the pmf of the HLZD as

$$r_k = \frac{\lambda(k+1)\theta^{k+1}}{T(\theta,s,a)(a+k+1)^{s+1}} . \tag{4.6}$$

Using Theorem 2.9, an alternative derivation for (4.4) can be obtained.

$$\begin{aligned}(n+1)p_{n+1} &= \sum_{k=0}^{n}\left[p_k \cdot \lambda(n-k+1)q_{n-k+1}\right] \\ &= \frac{\lambda}{\Phi(\theta,s+1,a+1)}\sum_{j=0}^{n}\frac{(j+1)\theta^k}{(j+1+a)^{s+1}}p_{n-j}\end{aligned}$$

4.2.2 Discrete Self-Decomposable

As shown in Figure 4.1, the Poisson-HLZD may not be unimodal. However, we can show that it is unimodal when a = 0.

Property 2 When $s \ge 0$, the Poisson-HLZD is discrete self-decomposable when $a = 0$.

Proof:

From (4.6) and Theorem 2.9, this is equivalent to finding conditions such that the canonical measure r_n is non-increasing.

$$\begin{aligned}\frac{r_n}{r_{n-1}} &= \frac{\dfrac{\lambda(n+1)\theta^{n+1}}{T(\theta,s,a)(a+n+1)^{s+1}}}{\dfrac{\lambda n\theta^n}{T(\theta,s,a)(a+n)^{s+1}}} \\ &= (1+\frac{1}{n})(1-\frac{1}{a+n+1})^{s+1}\theta\end{aligned}$$

Under the stated condition, $(1+\frac{1}{n})(1-\frac{1}{n+1})^{s+1}\theta \le (1+\frac{1}{n})(1-\frac{1}{n+1})\theta = \theta \le 1$. ■

The condition in Property 2 is satisfied by distributions such as the LD and Lotka distribution in the HLZ family.

4.2.3 Poisson-HLZD as a Convolution of Two Random Variables

From the definition of self-decomposable and Property 2, we know that under certain conditions the Poisson-HLZD may be expressed as convolution of two random variables with pgf's

$$e^{-\lambda}\exp\left[\frac{\lambda(1-\alpha+\alpha t)\Phi(\theta(1-\alpha+\alpha t),s+1,a+1)}{\Phi(\theta,s+1,a+1)}\right]$$

and

$$\exp\left[\frac{\lambda}{\Phi(\theta,s+1,a+1)}\left(\sum_{k=0}^{\infty}\frac{\theta^k(t^{k+1}-(1-\alpha+\alpha t)^{k+1})}{(k+a+1)^{s+1}}\right)\right],$$

where $0\le\alpha\le1$. The first pgf represents a thinning version of the Poisson-HLZD whereas the second distribution could be a new distribution that needs further exploration.

4.2.4 Unimodality and Strong Unimodality

Proving unimodality of a PSSD in general is not always direct. On the other hand, strong unimodality which implies unimodality could be easier to verify by using Theorem 2.8 and Theorem 2.10.

Property 3 The canonical measures r_n of Poisson-HLZD is log-concave if $s\le0$ and $\left(\frac{a+1}{a+2}\right)^{s+1}\theta\le\frac{1}{2}$. The same conditions lead to strong unimodality of the Poisson-HLZD.

Proof:

From (4.6),

$$\frac{r_n^2}{r_{n+1}r_{n-1}}=\left[\frac{(a+n+2)(a+n)}{(a+n+1)^2}\right]^{s+1}\frac{(n+1)^2}{n(n+2)}$$

$$\ge\left[\frac{(a+n+2)(a+n)}{(a+n+1)^2}\right]^{s}=\left[1-\frac{1}{(a+n+1)^2}\right]^{s}.$$

This is because the function $f(x)=\frac{(x+1)^2}{x(x+2)}>0$ and is decreasing in x for $x>0$. Since $\left[1-\frac{1}{(a+n+1)^2}\right]^s \ge 1 \Leftrightarrow s\le 0$, from Theorem 2.8, the probability distribution is log-concave if, and only if,

$$\frac{r_1}{r_0^2}=\frac{\dfrac{2\lambda\theta^2}{T(\theta,s,a)(a+2)^{s+1}}}{\dfrac{\lambda\theta}{T(\theta,s,a)(a+1)^{s+1}}}=2\left(\frac{a+1}{a+2}\right)^{s+1}\theta\le 1\cdot$$

The conclusion then follows by Theorem 2.10. ■

When s = 0 and a = 0, we have the NBD.

4.3 MP Formulation of Poisson-HLZD

The Poisson-HLZD will be shown to have a MP formulation, which is equivalent to showing that after substitution of $t=1-u$ in the pgf of the Poisson-HLZD, the resulting function

$$F(u)=e^{-\lambda}\exp\left[\frac{\lambda(1-u)\Phi(\theta(1-u),s+1,a+1)}{\Phi(\theta,s+1,a+1)}\right] \tag{4.7}$$

is completely monotone in u on $(0,\infty)$. Since the pgf is a power series with positive coefficients and always exists on the interval $[-1,1]$, according to Theorem 2.2, a substitution of $t=1-u$ will make $F(u)$ completely monotone on $(0,1]$. Hence, only the complete monotonicity of $F(u)$ for $u>1$ will be proved in the following and the Weierstrass M-test will be used to justify the uniform convergent requirement in Theorem 2.3.

(*Weierstrass M-test*) Suppose that $f(x,t)$ is Riemann integrable over $[a,c]$ for all $c\ge a$ and all $t\in J$. Suppose there exists a positive function M defined for $x\ge a$ such that $|f(x,t)| \le M(x)$ for all $x \ge a$ and $t \in J$, and such that the infinite integral $\int_a^\infty M(x)dx$ exists. Then for each $t \in J$, the integral $\int_a^\infty f(x,t)dx$ converges uniformly on J (Bartle, 1976).

Theorem 4.1 The Poisson-HLZD has MP formulation when $s > -1$.

Proof:

Applying Theorem 2.1 by taking $f(y) = e^{-\lambda} \exp\left[-\frac{\lambda y}{\Phi(\theta, s+1, a+1)}\right]$, it is easy to see that *f* is completely monotone in *y*. It remains to show that

$$h(u) = (u-1)\Phi(\theta(1-u), s+1, a+1)$$

is completely monotone for $u > 1$. Using the integral representation (3.1),

$$h(u) = (u-1)\frac{1}{\Gamma(s+1)}\int_0^\infty \frac{1}{e^t + \theta(u-1)} e^{-at} t^s dt$$

Hence, $h(u)$ is nonnegative for $u > 1$.

The derivative of $h(u)$,

$$h'(u) = (u-1)\frac{d}{du}\Phi(\theta(1-u), s+1, a+1) + \Phi(\theta(1-u), s+1, a+1)$$

$$= \Phi(\theta(1-u), s, a+1) - a\Phi(\theta(1-u), s+1, a+1)$$

$$= \int_0^\infty \frac{1}{e^t + \theta(u-1)}\frac{1}{\Gamma(s)} e^{-at} t^{s-1} dt - \int_0^\infty \frac{1}{e^t + \theta(u-1)}\frac{a}{\Gamma(s+1)} e^{-at} t^s dt$$

Let $I_1 = \int_0^\infty \frac{1}{e^t + \theta(u-1)}\frac{1}{\Gamma(s)} e^{-at} t^{s-1} dt$ and $I_2 = \int_0^\infty \frac{1}{e^t + \theta(u-1)}\frac{a}{\Gamma(s+1)} e^{-at} t^s dt$.

By taking $f(t) = \frac{1}{\Gamma(s)} e^{-at} t^{s-1}$, $K(u,t) = \frac{1}{e^t + \theta(u-1)}$ in Theorem 2.3, it is easy to see that

$$K(u,t) = \frac{1}{e^t + \theta(u-1)} \ge 0 ,$$

and

$$(-1)^k \frac{\partial^k}{\partial u^k} K(u,t) = k!\left(e^t + \theta(u-1)\right)^{-(k+1)} \theta^k \geq 0.$$

Observed that $\frac{\partial^k}{\partial u^k} K(u,t) f(t)$ is a product of two bounded continuous functions and is hence Riemann integrable on $[0,c]$ for any $c > 0$. For any given $k \in \mathbb{N}_0$,

$$\left| \frac{\partial^k}{\partial u^k} K(u,t) f(t) \right| = \frac{\theta^k k!}{\Gamma(s)} \frac{1}{[e^t + \theta(u-1)]^{(k+1)}} e^{-at} t^{s-1} \leq \frac{\theta^k k!}{\Gamma(s)} e^{-(a+k+1)t} t^{s-1},$$

for all $u > 1$ and $\int_0^\infty \frac{\theta^k k!}{\Gamma(s)} e^{-(a+k+1)t} t^{s-1} dt = \theta^k k! (a+k+1)^s$. By the Weierstrass *M*-test, the uniform convergence of the integral $\int_0^\infty \frac{\partial^n}{\partial u^n} K(u,t) f(t) dt$ can be justified and Theorem 2.3 can be applied to conclude that I_1 and I_2 are both completely monotone on $(1,\infty)$.

Although the sum of two completely monotone functions is completely monotone, there is no analogous argument about the difference of two completely monotone functions. To proceed, apply change of variable $x = at$,

$$\begin{aligned}
&I_1 - I_2 \\
&= \frac{1}{\Gamma(s)} \int_0^\infty \frac{e^{-at}}{e^t + \theta(u-1)} t^{s-1} dt - \frac{a}{\Gamma(s+1)} \int_0^\infty \frac{e^{-at}}{e^t + \theta(u-1)} t^s dt \\
&= \frac{1}{a^s} \left[\int_0^\infty \frac{1}{e^{x/a} + \theta(u-1)} \frac{1}{\Gamma(s)} e^{-x} x^{s-1} dx - \int_0^\infty \frac{1}{e^{x/a} + \theta(u-1)} \frac{1}{\Gamma(s+1)} e^{-x} x^s dx \right]
\end{aligned}$$

Let $f_X(x) = \frac{1}{\Gamma(s)} e^{-x} x^{s-1}$ and $f_Y(y) = \frac{1}{\Gamma(s+1)} e^{-y} y^s$ be the respective pdf of Gamma random variables *X* and *Y*. Let $\Gamma(s,x) = \int_x^\infty e^{-t} t^{s-1} dt$ be the upper incomplete Gamma function, $P(X > x) = \frac{\Gamma(s,x)}{\Gamma(s)}$ and $P(Y > x) = \frac{\Gamma(s+1,x)}{\Gamma(s+1)}$.

Using the recurrence relation of the incomplete Gamma function $\Gamma(s+1,x)=s\Gamma(s,x)+x^s e^{-x}$ (Erdélyi *et al.*, 1953; Vol II, p.134),

$$P(Y>x)=\frac{s\Gamma(s,x)+x^s e^{-x}}{s\Gamma(s)}$$

$$=\frac{\Gamma(s,x)}{\Gamma(s)}+\frac{x^s e^{-x}}{s\Gamma(s)}>P(X>x),\ \text{for all } x>0\,.$$

This shows that X is stochastically smaller than Y. The following shows that if X is stochastically smaller than Y then $E[\frac{1}{e^X+\theta(u-1)}]\geq E[\frac{1}{e^Y+\theta(u-1)}]$.

$$\begin{aligned}
&E[\frac{1}{e^X+\theta(u-1)}]\\
&=\int_0^\infty P(\frac{1}{e^X+\theta(u-1)}>t)dt\\
&=\int_0^{\frac{1}{\theta(u-1)}} P(X<\ln\left[\frac{1}{t}-\theta(u-1)\right])dt\\
&\geq\int_0^{\frac{1}{\theta(u-1)}} P(Y<\ln\left[\frac{1}{t}-\theta(u-1)\right])dt\\
&=E[\frac{1}{e^Y+\theta(u-1)}]
\end{aligned}$$

Similarly,

$$E\left[\frac{k!\theta^k}{\left[e^X+\theta(u-1)\right]^{k+1}}\right]\geq E\left[\frac{k!\theta^k}{\left[e^Y+\theta(u-1)\right]^{k+1}}\right].$$

These results show that

$$\left|\int_0^\infty \frac{\partial^k}{\partial u^k}\left(\frac{1}{e^{x/a}+\theta(u-1)}\right)\frac{1}{\Gamma(s)}e^{-x}x^{s-1}dx\right|$$
$$\geq\left|\int_0^\infty \frac{\partial^k}{\partial u^k}\left(\frac{1}{e^{x/a}+\theta(u-1)}\right)\frac{1}{\Gamma(s+1)}e^{-x}x^{s}dx\right|$$

for all $k \in \mathbb{N}_0$ and $u > 1$. Hence, the sign of $h^{(k)}(u)$ is dominated by the sign of $\frac{\partial^k}{\partial u^k} I_1$, which is completely monotone. Therefore, $h'(u)$ is completely monotone. ■

Using the Laplace transform of the mixing distribution, the moments of the mixing distribution can be obtained easily and the moment generating function $M(t) = e^{-\lambda} \exp\left[\frac{\lambda(1+t)\Phi(\theta(1+t), s+1, a+1)}{\Phi(\theta, s+1, a+1)}\right]$. Various approaches have been developed to approximate the pdf of a distribution from its moments. These include the applications of the Pearson curves, orthogonal polynomials (Provost, 2005) or the maximum entropy principle (Tagliani & Velasquez, 2004). With the advancements of modern computing technology, numerical inversion as described in Section 2.10 also gives satisfactory results as shown in the following examples.

Figure 4.6 shows the performance of the Gaver-Stehfest formula and the Fourier series method. The inversions highly coincide for most of the domain of *x*, except for small values of *x* where the Fourier series method fails to produce any value. In Figure 4.7, the inversions are fine for all values of *x*, except at a point near 0.5. The diagram shows a spike from the Fourier series method which could represent a singularity or point mass in the mixing distribution. In general, the Bernstein's theorem does not guarantee the mixing distribution to be absolutely continuous.

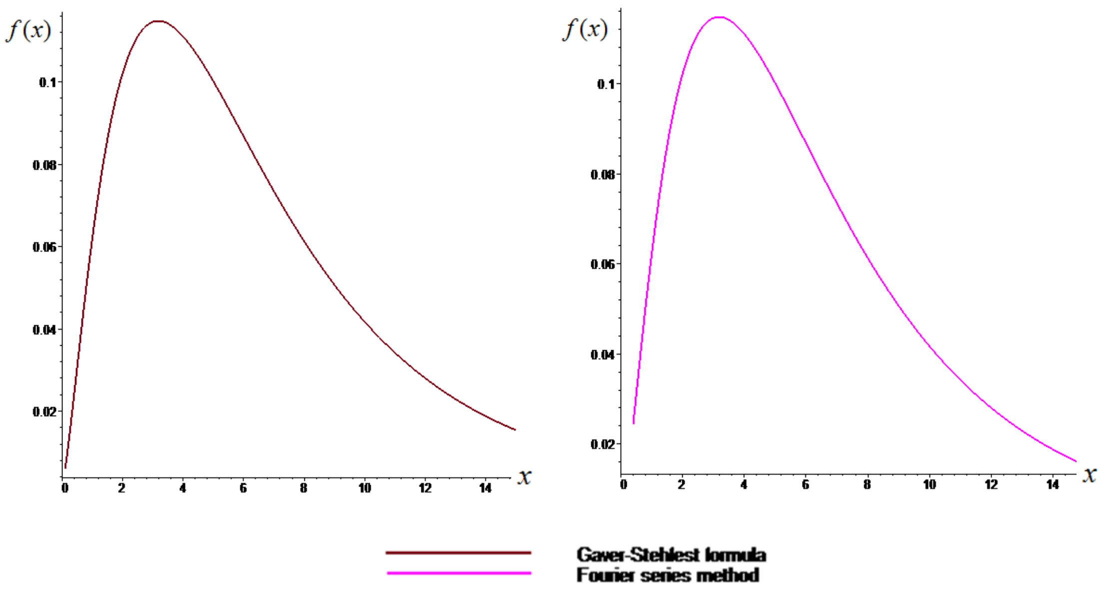

Figure 4.6

Pdf of mixing distribution of Poisson-HLZD with

$\lambda = 3,\ \theta = 0.8949,\ s = 0.9509,\ a = 0.6905$

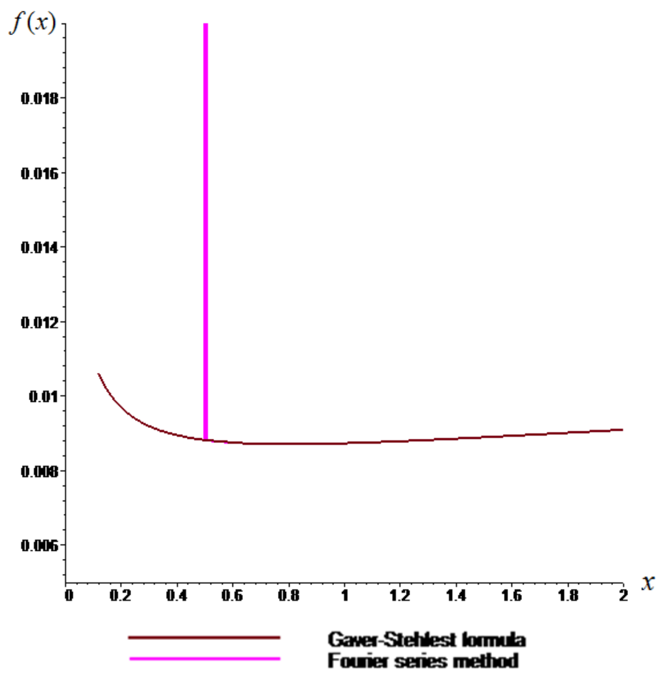

Figure 4.7

Pdf of mixing distribution of Poisson-HLZD with

$\lambda = 4.5,\ \theta = 0.98,\ s = 0.01,\ a = 0.7$

4.4 Stochastic Orders Related to the Poisson-HLZ Family

The results established in Section 3.2 will be applied to deduce similar properties for the Poisson-HLZD based on Theorem 2.11 and Theorem 2.12. Before that, the following property for the Poisson distribution is needed.

Lemma 4.1 The likelihood ratio orders among the Poisson random variables follow the orders between their means.

Proof: Let λ_1 and λ_2 with $\lambda_1 < \lambda_2$, be the means of two Poisson random variables X and Y respectively. The ratio of corresponding pmf's,

$$\frac{\dfrac{e^{-\lambda_1}\lambda_1^x}{x!}}{\dfrac{e^{-\lambda_2}\lambda_2^x}{x!}} = e^{-(\lambda_2-\lambda_1)}\left(\frac{\lambda_1}{\lambda_2}\right)^x$$

is a monotone decreasing function in *x*. Hence, $X \leq_{lr} Y$. ■

Since the likelihood ratio order implies the usual stochastic order, hazard rate order and reverse hazard rate order, the conclusion above also holds for these orders.

Using Theorem 2.11 and Lemma 4.1, some orderings among the Poisson-HLZDs can be determined. However, when the conditions in the theorems are not satisfied, the orders are not trivial.

4.4.1 Comparison of Poisson and Poisson-HLZD

Let N be a Poisson random variable with mean $\lambda*$ and M be a MP random variable with mean Θ that has pdf $f(\theta)$. The orders, $\leq_{lr}$ and $\leq_{st}$, between N and M can be determined by using the following theorem from Misra *et al.* (2003).

Lemma 4.2 (Misra *et al.*, 2003; Lemma 3.1 & Theorem 3.1): Define $a_1(k)=E[e^{-\Theta}\Theta^{k+1}]/E[e^{-\Theta}\Theta^{k}]$, $\lambda_0=a_1(0)$, $\lambda_1=-\ln E[e^{-\Theta}]$, $\lambda_2=E[\Theta]$ and $h_k(x)=x\sum_{j=0}^{k}\left[\frac{(-\ln x)^j}{j!}\right]$, for $0<x<1$, $k=0,\ 1,\ \ldots$. Then,

- $0<\lambda_0\le\lambda_1\le\lambda_2$.
- $M\ge_{lr} N$ if, and only if $0<\lambda^*\le\lambda_0$.
- $M\ge_{st} N$ if, and only if $0<\lambda^*\le\lambda_1$.

To apply Lemma 4.2, λ_0, λ_1, and λ_2 can be found from the Laplace transform of the mixing distribution. From $\hat{f}(u)=\int_0^\infty e^{-u\theta}f(\theta)d\theta$, we have $E[e^{-\Theta}]=\hat{f}(1)$, $E[e^{-\Theta}\Theta]=-\hat{f}\,'(1)$, and $E[\Theta]=-\hat{f}\,'(0)$. For Poisson-HLZD,

$$\hat{f}(u)=e^{-\lambda}\exp\left[\frac{\lambda(1-u)\Phi(\theta(1-u),s+1,a+1)}{\Phi(\theta,s+1,a+1)}\right]$$

and

$$\hat{f}\,'(u)=\frac{\lambda e^{-\lambda}}{\Phi(\theta,s+1,a+1)}\exp\left[\frac{\lambda(1-u)\Phi(\theta(1-u),s+1,a+1)}{\Phi(\theta,s+1,a+1)}\right]\times\left[a\Phi\left(\theta(1-u),s+1,a+1\right)-\Phi\left(\theta(1-u),s,a+1\right)\right].$$

Therefore,

$$\lambda_0=\frac{E[e^{-\Theta}\Theta]}{E[e^{-\Theta}]}=\frac{\lambda}{\Phi(\theta,s+1,a+1)}\left(\frac{1}{(a+1)^s}-\frac{a}{(a+1)^{s+1}}\right)$$

$$\lambda_1=-\ln E[e^{-\Theta}]=\lambda$$

$$\lambda_2=E[\Theta]=\frac{\lambda}{\Phi(\theta,s+1,a+1)}\left(\Phi\left(\theta,s,a+1\right)-a\Phi\left(\theta,s+1,a+1\right)\right)$$

These three numbers are sufficient to determine the order of Poisson-HLZD as compare to a Poisson random variable. The following property gives a more general result on the PSSD.

Property 4 If *M* is a Poisson-stopped sum random variable that has MPD, then *M* is stochastically larger than a Poisson random variable with mean λ^* if, and only if $\lambda^* \le \exp[\lambda\left(G(0)-1\right)]\lambda G'(0)$, where $G(t)$ is the pgf of the cluster size distribution of *M*.

Proof:

Let $G_M(t)=\exp[\lambda\left(G(t)-1\right)]$, the Laplace transform of the mixing distribution is given by $\hat{f}(u)=\exp[\lambda\left(G(1-u)-1\right)]$. Hence,

$$\hat{f}'(u)=-\exp[\lambda\left(G(1-u)-1\right)]\lambda G'(1-u)$$

and

$$\lambda_0=\frac{-\hat{f}'(1)}{\hat{f}(1)}=\frac{\exp[\lambda\left(G(0)-1\right)]\lambda G'(0)}{\exp[\lambda\left(G(0)-1\right)]}=\lambda G'(0)$$

$$\lambda_1=-\ln\hat{f}(1)=\lambda\left(1-G(0)\right)$$

$$\lambda_2=-\hat{f}'(1)=\exp[\lambda\left(G(0)-1\right)]\lambda G'(0)$$

■

4.5 Examples of Data-Fitting

The Poisson-HLZD is fitted to some well-known datasets to the model to illustrate its usefulness. Although there is no closed form expression for the pmf of the Poisson-HLZD, by (4.4), the ML estimates of the parameters can be obtained using Algorithm 2.1.

Table 4.1 shows the well-known accident claim data from Bühlmann (1970), which has been fitted to different models by many authors. It is shown in Gathy & Lefevre (2010) that NBD and a generalised NBD that is derived from the Lagrangian Katz family (3 parameters) do not provide good fit whereas generalised Poisson (2 parameters) gives only marginal *p*-value of 0.0632. Our results show that the full Poisson-HLZD give very good fit to the data with AIC = 109227.52. From the values of the estimates for s and a, a simpler model with $a = 0$ and $s = 2$ is selected based on smaller AIC of 109225.32. Nevertheless, in Klugman *et al.* (2008), the authors were able to get an AIC of 109223.6 by using the Sichel distribution. Although the difference in AIC between the reduced model and Sichel is small, from the χ^2 statistic, the Sichel is still preferred.

Table 4.1

Fitting Bühlmann's accident claim data to Poisson-HLZD

Claim numbers	Observed frequencies	$a = 0, s = 2$	Full model	Poisson-inverse Gaussian*
0	103704	103654.28	103700.25	103710.0
1	14075	14143.94	14072.69	14054.7
2	1766	1745.29	1775.36	1784.9
3	255	252.41	252.91	254.5
4	45	44.62	41.19	40.4
5	6	9.38	7.88	6.9
6	2	2.25	1.65	1.3
>6	0	0.84	0.5	0.3
Df		4	2	4
χ^2		2.2378	0.8786	0.9719
p-value		0.6921	0.6445	0.9143
log-likelihood		-54610.66	-54609.76	-54609.8
λ		0.14486	0.14471	
θ		0.44173	0.35946	
s		2	1.84329	
a		0	0.11579	

* The result in this column follows Klugman (2009) except that the author combines the last three groups while calculating the χ^2 statistic.

In Table 4.2, the Poisson-HLZD is fitted to data of automobile accident claims on 9461 contracts as illustrated in Thyrion (1960). This data was shown to have bad fit to Poisson and NBD. In Ruohonen (1988), the author also fitted a three parameter model, which represents two processes with 'good' and 'bad' risk, to this set of data and get marginal fit at 10% significance level. Table 4.2 shows that the χ^2 value can be reduced significantly when fitted to the Poisson-HLZD. From the result of the full model, a simpler model that will produce slightly heavy-tail distribution requires $a = 0$. Testing $H_0 : a = 0$ versus $H_1 : a \neq 0$ at 5% significance level, the likelihood ratio test statistic $\Lambda \simeq 0.004 < \chi^2_{0.05,1} = 3.841$. The AIC of reduced model is also smaller than that of the full model. Hence, the Poisson-stopped sum of Good distribution gives an adequate parsimonious model for the data.

Table 4.2

Fitting Thyrion's accident claim data to Poisson-HLZD

Claim numbers	Observed frequencies	NBD	a = 0	Full model	Weighted Poisson**
0	7840	7848.25	7838.94	7838.04	7837.4
1	1317	1285.1	1323.09	1323.71	1326.16
2	239	257.5	227.63	228.45	222.76
3	42	54.74	49.78	49.44	52.68
4	14	11.97	13.82	13.57	15.08
5	4	2.66	4.61	4.53	4.66
6	4	0.6	1.76	1.75	1.5
7	1	0.14	0.73	0.76	0.5
>7	0	0.04	0.65	0.74	0.26
Total	9461	9461	9461.01	9460.99	9461
df		3	2	1	2
χ^2		12.69	2.02	1.85	4.12
p-value		0.0053	0.3649	0.1742	0.1277
log-likelihood		-5348.056	-5342.26	-5342.258	
λ		0.1868875	0.1880749	0.188	
θ		0.2370001	0.68918	0.7817	
s		0	1.975	2.665	
a		0	0	0.2287	

**The data in this column follows Ruohonen (1988).

CHAPTER 5 THE INVERSE TRINOMIAL DISTRIBUTION AS A MIXED POISSON DISTRIBUTION

5.0 Introduction

The univariate inverse trinomial distribution (ITD) can be derived from various mechanisms as mentioned in Shimizu *et al.* (1997). In this chapter, the ITD will be proved to be a MPD under certain condition and its mixing distribution is obtained. The evolution of the ITD as a random walk model and some related results in the literature will be introduced in Section 5.1. In Section 5.2, the Bernstein's theorem is applied again to show that ITD can be regarded as a MPD and its mixing distribution is represented as an infinite Laguerre series in Section 5.3.

5.1 The ITD as a Random Walk Model

In Vol I, Section XIV.4, Feller (1967) describes the fortune of a gambler, who plays against an infinitely rich opponent, by using a simple random walk model starting at $z \in \mathbb{N}$ with absorbing barrier at 0. Let the probabilities of steps of –1 and +1 be p and r respectively, with $p \geq r > 0$ and $p + r = 1$. The random variable M, which represents the number of steps until the process ends at 0, has pmf

$$p_m = \begin{cases} \dfrac{z}{m}\dbinom{m}{(m-z)/2} r^{(m-z)/2} p^{(m+z)/2}; & 2 \mid (m-z), \\ 0 & ; \quad \text{otherwise.} \end{cases}$$

Applying the transformation $M = 2Y + z$, Yanagimoto's (1989) inverse binomial random variable Y which has the lost-games distribution is obtained, see Kemp & Kemp, (1968). The random variable Y has pmf

$$p_y = \frac{\Gamma(2y+z)z}{\Gamma(y+1)\Gamma(y+z+1)} p^{y+z} r^y,\ y \in \mathbb{N}_0. \tag{5.1}$$

The inverse binomial distribution is further generalised in Shimizu & Yanagimoto (1991) to obtain the ITD by introducing a nonzero probability of stagnant in the random walk model.

Consider the random walk model illustrated in Figure 5.1 that has a starting point at $\lambda > 0$ and an absorbing barrier at 0. The probabilities of a –1, 0, or +1 step are given by *p*, *q* and *r* respectively, such that $p \geq r$ and $p+q+r=1$. An inverse trinomial random variable *X* is defined by $X = N - \lambda$, where N is the number of steps until the process is absorbed by 0. The pmf of *X* is

$$p_k = \frac{\lambda p^\lambda q^k}{k+\lambda} \sum_{j=0}^{\lfloor k/2 \rfloor} \binom{k+\lambda}{j,\, j+\lambda, k-2j} \left(\frac{pr}{q^2}\right)^j,\ k \in \mathbb{N}_0; \tag{5.2}$$

where $\binom{k+\lambda}{j,\, j+\lambda, k-2j} = \frac{(k+\lambda)!}{j!(j+\lambda)!(k-2j)!}$. When λ is not an integer, the factorial will be replaced by Gamma function.

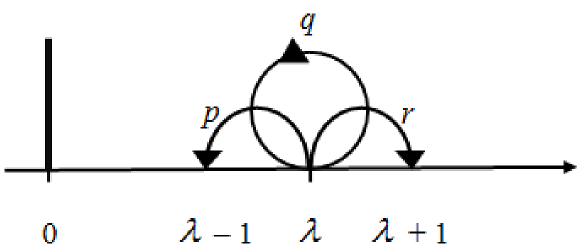

Figure 5.1

Random walk model in the ITD

The ITD is named to reflect its relationship with the trinomial random variable *Z* which takes values $\frac{1}{\lambda}$, 0, and $-\frac{1}{\lambda}$ with probabilities *p*, *q* and *r* respectively. The inverse of $\log E[e^{-tZ}]$ is equal to $\log E[e^{-tN}]$ for $t \leq \left(\frac{\lambda}{2}\right)\log\left(\frac{p}{r}\right)$, where $N = X + \lambda$; see Shimizu *et al.* (1997). In the same paper, the authors also derived two expressions for the pgf of the multivariate ITD, from which we obtain two expressions for the pgf of univariate ITD,

$$G_X(t)=\left(\frac{2p}{\left[(1-qt)+\sqrt{(1-qt)^2-4prt^2}\right]}\right)^{\lambda} \tag{5.3}$$

and

$$G_X(t)=\left(\frac{1-q}{1-qt}\right)^{\lambda}\left(\frac{p}{1-q}\right)^{\lambda}\,{}_2F_1\left[\frac{\lambda}{2},\,\frac{\lambda+1}{2};\lambda+1;\frac{4prt^2}{(1-qt)^2}\right], \tag{5.4}$$

where ${}_2F_1\left[a,\,b;c;x\right]=\sum_{j=0}^{\infty}\frac{(a)_j(b)_j}{(c)_j}\frac{x^j}{j!}$ is the Gauss hypergeometric function. If r = 0, (5.3) reduces to the pgf of NBD.

Khang & Ong (2007) studied the ITD as a PSSD and derived a new generalisation of LD from the cluster size distribution. Aoyama *et al.* (2008) further generalised the shifted ITD to a distribution of random walk on a half plane with 5 transition probabilities. More recently, a subclass of this generalisation, the $\mathrm{GIT}_{3,1}$ which exhibits under-, equi- and over-dispersed properties was extended as a convolution of binomial and negative binomial random variables in Imoto (2012). Since the ITD is both a PSSD and a generalisation of the NBD, it is therefore interesting to investigate whether it is also a MPD.

5.2 ITD as a MPD

Using expression (5.3), let

$$\phi(s) = G_X(1-s)$$
$$= \left(\frac{2p}{\left[(1-q+qs) + \sqrt{(1-q+qs)^2 - 4pr(1-s)^2} \right]} \right)^{\lambda}$$

It is easy to verify that $\phi(0)=1$. $\phi(s)$ is nonnegative on [0,1] as this part is the reflection image of a pgf about the line $s = 0.5$. Since

$$(1-q+qs)^2 - 4pr(1-s)^2 = (q^2 - 4pr)(1-s)^2 - 2q(1-s) + 1, \tag{5.5}$$

a nonnegative leading coefficient for the quadratic in $(1-s)$ in (5.5) is required for $\phi(s)$ to remain nonnegative on $(1,\infty)$, that is

$$q^2 - 4pr \geq 0 \Leftrightarrow q \geq 2\sqrt{pr} \Leftrightarrow 1-p-r \leq 2\sqrt{pr} \Leftrightarrow 1 \geq (\sqrt{p}+\sqrt{r})^2.$$

Hence, a necessary condition for ITD to be a MPD is $\sqrt{p}+\sqrt{r} \leq 1$. Argument for sufficiency is more complicated and can be done by treating $f(x) = x^{-\lambda}$ and $h(s) = (1-q+qs) + \sqrt{(1-q+qs)^2 - 4pr(1-s)^2}$ in Theorem 2.1 and apply the following lemmas.

Lemma 5.1 $f(x) = x^{-\lambda}$ is completely monotone on $(0,\infty)$ for $\lambda \geq 0$.

Proof:

Let $f(x) = x^{-\lambda}$, $(-1)^n f^{(n)}(x) = (-1)^{2n}(\lambda)_n x^{-(\lambda+n)} \geq 0$. ■

Lemma 5.2 When $\sqrt{p}+\sqrt{r} = 1$, $h(s) = (1-q+qs) + \sqrt{(1-q+qs)^2 - 4pr(1-s)^2}$ is nonnegative and has completely monotone derivative on $(0,\infty)$.

Proof:

$$\sqrt{p}+\sqrt{r}=1 \Leftrightarrow 4pr=q^2,$$

$$(1-q+qs)+\sqrt{(1-q+qs)^2-4pr(1-s)^2}$$

$$=(1-q+qs)+\sqrt{(1-q+qs)^2-q^2(1-s)^2}$$

$$=(1-q+qs)+\sqrt{1-2q+2qs}$$

$=x+\sqrt{2x-1}=k(x)$, with substitution $x=1-q+qs$.

Under the condition $\sqrt{p}+\sqrt{r}=1$, $q=1-(1-\sqrt{r})^2-r=2(\sqrt{r}-r)$ has a global maximum at (0.25, 0.5), hence we conclude that $1-q+qs\geq 1-q\geq \frac{1}{2}$ for $s>0$ and $h(s)$ is nonnegative.

To show that $h(s)$ has completely monotone derivative $\frac{d}{ds}h(s)$. By the chain rule,

$$\frac{d}{ds}h(s)=\frac{d}{dx}k(x)\cdot\frac{dx}{ds}=q\frac{d}{dx}k(x).$$

Since multiplying a positive constant does not change the completely monotone property, we only need to look at the complete monotonicity of $k(x)$. It is easy to deduce that

$$\frac{d^n}{dx^n}k(x)=(-1)^{n-1}\cdot 1\cdot 3\cdot\ldots\cdot(2n-3)\cdot(2x-1)^{-\frac{2n-1}{2}}\begin{cases}\geq 0, \text{ when } n=3,5,\ldots\\ \leq 0, \text{ when } n=2,4,\ldots\end{cases}$$

Hence, $h(s)$ has completely monotone derivative on $(0,\infty)$. ■

Lemma 5.3 When $\sqrt{p}+\sqrt{r}<1$, $h(s)=(1-q+qs)+\sqrt{(1-q+qs)^2-4pr(1-s)^2}$ is nonnegative and has completely monotone derivative on $(0,\infty)$.

Proof:

$$\sqrt{p}+\sqrt{r}<1 \Leftrightarrow q^2>4pr \text{ or } \frac{4pr}{q^2}<1.$$

Denote $a=\frac{4pr}{q^2}$ and use the substitution $x=1-q+qs$ or $(1-s)=\frac{1-x}{q}$.

$$h(s)=\left(1-q+qs\right)+\sqrt{\left(1-q+qs\right)^2-4pr(1-s)^2}$$

$$=x+\sqrt{x^2-a(1-x)^2}$$

$$=x+\sqrt{\left(1-a\right)x^2+2ax-a}=k(x)$$

Since $s\in(0,\infty)\Rightarrow x\in(1-q,\infty)\subset(0,\infty)$, to show that $h(s)$ is nonnegative, we only need to show that the function under the square root is positive for $s\in(0,\infty)$. The quadratic equation $\left(1-a\right)x^2+2ax-a=0$ has roots $\frac{-2a\pm\sqrt{4a^2+4a(1-a)}}{2(1-a)}=\frac{\pm\sqrt{a}-a}{1-a}$ and is positive whenever

$$x>\frac{\sqrt{a}-a}{1-a}=\frac{\sqrt{a}}{1+\sqrt{a}}=\frac{2\sqrt{pr}/q}{1+2\sqrt{pr}/q}=\frac{2\sqrt{pr}}{1-(\sqrt{p}-\sqrt{r})^2}\geq 2\sqrt{pr}.$$

Since $x\in(1-q,\infty)$ and $2\sqrt{pr}<1-q$, $h(s)$ is nonnegative.

The next step is to show that $\frac{d}{ds}h(s)$ is completely monotone. Similar to the proof in Lemma 5.2, we only need to look at the complete monotonicity of $k(x)$. For $x\in\left(1-q,\infty\right)$, we have

$$\frac{d}{dx}k(x)=1+\frac{2x+2a(1-x)}{2\sqrt{x^2-a(1-x)^2}}=1+\frac{\left(1-a\right)x+a}{\sqrt{x^2-a(1-x)^2}}\geq 0$$

and

$$\frac{d^2}{dx^2}k(x)=\frac{-a}{(\sqrt{x^2-a(1-x)^2})^3}=-a[(1-a)x^2+2ax-a]^{-\frac{3}{2}}\leq 0.$$

Higher derivatives are very complicated and it is not easy to determine their positivities. Instead of taking higher derivatives, consider the function

$$k_1(x) = [(1-a)x^2 + 2ax - a]^{-\frac{3}{2}}$$

$$= (1-a)^{-\frac{3}{2}}\left[\left(x+\frac{a}{1-a}\right)^2 - \frac{a}{(1-a)^2}\right]^{-\frac{3}{2}}. \tag{5.6}$$

For $x+\dfrac{a}{1-a} > \dfrac{\sqrt{a}}{1-a}$ or $x > \dfrac{\sqrt{a}-a}{1-a}$, using formula 2.1.5.5 from (Prudnikov *et al.*, 1986; Vol 5, pg 27), expression (5.6) has inverse Laplace transform,

$$l(t) = \frac{(1-a)^{-\frac{3}{2}}}{\sqrt{\frac{a}{(1-a)^2}}} e^{-\frac{a}{1-a}v} t I_1\left(\sqrt{\frac{a}{(1-a)^2}}t\right)$$

where $I_\nu(z)$ is the modified Bessel function of order ν defined by

$$I_\nu(z) = \left(\frac{1}{2}z\right)^\nu \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}z^2\right)^k}{k!\Gamma(\nu+k+1)}.$$

It is easy to see from the definition that the modified Bessel function is nonnegative when the argument is real. Since $k_1(x)$ is the Laplace transform of a nonnegative function $l(t)$, $k_1(x)$ must be alternate in sign after every differentiation according to (2.17). Hence, completes the proof that $h(s)$ has completely monotone derivative. ■

Theorem 5.1 When $\sqrt{p}+\sqrt{r} \le 1$, ITD has a MP formulation.

Proof:

Take $f(x) = x^{-\lambda}$, $h(s) = (1-q+qs) + \sqrt{(1-q+qs)^2 - 4pr(1-s)^2}$ and apply Theorem 2.1, together with Lemma 5.1, Lemma 5.2, and Lemma 5.3. ■

The admissible values for p and q in the original parameter space form a triangular region bounded by the lines, $p=\frac{1}{2}(1-q)$ and $p=1-q$ for $0\le q\le 1$. The additional condition

$$\begin{aligned}
&\sqrt{p}+\sqrt{r}\le 1\\
&\Leftrightarrow p+2\sqrt{p(1-p-q)}+1-p-q\le 1\\
&\Leftrightarrow 2\sqrt{p(1-p-q)}\le q\\
&\Leftrightarrow 4p^2+(4q-4)p+q^2\ge 0
\end{aligned}$$

The solutions for p in the quadratic equation $4p^2+(4q-4)p+q^2=0$ are $p=\frac{1-q\pm\sqrt{1-2q}}{2}$. Hence, for $q\ge 0.5$, the ITD is always a MPD. However, for $q<0.5$, the ITD has MP formulation if and only if $\frac{1-q+\sqrt{1-2q}}{2}\le p\le 1-q$. These pairs of (q,p) are shown in the shaded region in Figure 5.2.

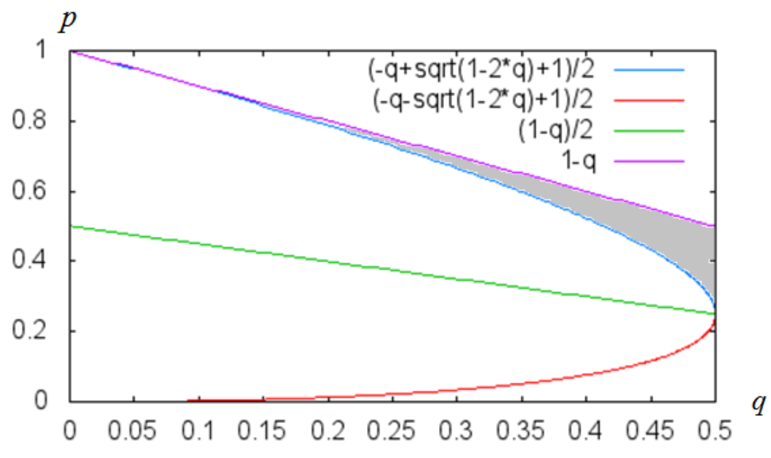

Figure 5.2

Region in the ITD parameter space that is corresponding to MPD for $q<0.5$

5.3 Mixing Distribution of the ITD

Theorem 5.2 When $\sqrt{p}+\sqrt{r}\le 1$, the ITD with pmf given by (5.2) has a mixed Poisson formulation and the mixing pdf $f(x)$ is given by

$$f(x)=\left(\frac{p}{\phi}\right)^{\lambda}\sum_{i=0}^{\infty}\frac{(\lambda)_i\left(\lambda+1/2\right)_i}{\left(2\lambda+1\right)_i}\left(\frac{4\sqrt{pr}}{\phi}\right)^{i}\frac{x^{\lambda-1}}{\Gamma(\lambda+i)}e^{-\frac{1-\phi}{\phi}x}L_i^{\lambda-1}\left(\frac{x}{\phi}\right), \quad (5.7)$$

when the infinite series convergent. Here, $L_n^a\left(x\right)$ is the Laguerre polynomial orthogonal over $(0,\infty)$ with respect to $x^{a-1}e^{-x}$, and $\phi=q+2\sqrt{pr}$.

Proof:

In the Euler transformation

$${}_2F_1\left[\lambda,\,\delta;2\delta;2z\right]=\left(1-z\right)^{-\lambda}\,{}_2F_1\left[\frac{\lambda}{2},\frac{\lambda+1}{2};\delta+\frac{1}{2};\left(\frac{z}{1-z}\right)^2\right],$$

set $z=\dfrac{-2\sqrt{pr}\;u}{1-qu-2\sqrt{pr}\,u}$, and substitute into (5.4),

$$G_X(t)=\left(\frac{p}{1-(q+2\sqrt{pr}\,)t}\right)^{\lambda}\,{}_2F_1\left[\lambda,\,\lambda+1/2;2\lambda+1;\,\frac{-4\sqrt{pr}\,t}{1-(q+2\sqrt{pr}\,)t}\right].$$

Let $\phi=q+2\sqrt{pr}$ and substitute $t=1-s$ in the expression above gives

$$F(s)=\left(\frac{p}{1-\phi\left(1-s\right)}\right)^{\lambda}\,{}_2F_1\left[\lambda,\,\lambda+1/2;2\lambda+1;\,\frac{-4\sqrt{pr}\left(1-s\right)}{1-\phi\left(1-s\right)}\right],$$

which can be rewritten as

$$F(s)=p^{\lambda}\sum_{i=0}^{\infty}\frac{(\lambda)_i(\lambda+1/2)_i}{(2\lambda+1)_i\,i!}\frac{(4\sqrt{pr}\,)^i(s-1)^i}{\phi^{\lambda+i}\left(s+(1-\phi)/\phi\right)^{\lambda+i}} \quad (5.8)$$

To find the inverse Laplace transform of (5.8), we apply the inverse Laplace transform (Roberts & Kaufman, 1966; pg 223, Formula 20),

$$L^{-1}\left\{\frac{(s-a)^n}{(s-b)^{n+\nu}}\right\}=\frac{n!}{\Gamma(n+\nu)}t^{\nu-1}e^{bt}L_n^{\nu-1}[(a-b)t];\ \mathrm{Re}\,\nu>0;\,n=0,1,....;\,\mathrm{Re}\,s>\mathrm{Re}\,b;$$

with $a=1$, $b=-\dfrac{1-\phi}{\phi}$, $n=i$ and $\nu=\lambda$. Since $1-\phi=(\sqrt{p}-\sqrt{r})^2\geq 0$, the conditions of the formula are satisfied. Hence,

$$f(x)=\left(\frac{p}{\phi}\right)^{\lambda}\sum_{i=0}^{\infty}\frac{(\lambda)_i\left(\lambda+1/2\right)_i}{\left(2\lambda+1\right)_i i!}\left(\frac{4\sqrt{pr}}{\phi}\right)^{i}\frac{i!x^{\lambda-1}}{\Gamma(\lambda+i)}e^{\frac{\phi-1}{\phi}x}L_i^{\lambda-1}(x/\phi)$$

$$=\left(\frac{p}{\phi}\right)^{\lambda}\sum_{i=0}^{\infty}\frac{(\lambda)_i\left(\lambda+1/2\right)_i}{\left(2\lambda+1\right)_i}\left(\frac{4\sqrt{pr}}{\phi}\right)^{i}\frac{x^{\lambda-1}}{\Gamma(\lambda+i)}e^{-\frac{1-\phi}{\phi}x}L_i^{\lambda-1}(x/\phi) \qquad \blacksquare$$

Note that if $r=0$, (5.7) reduces to the gamma pdf $f(x)=\dfrac{(p/q)^{\lambda}}{\Gamma(\lambda)}x^{\lambda-1}e^{-\frac{p}{q}x}$.

As pointed out in Abate *et al.* (1996), one of the problems with an infinite Laguerre series is slow convergence for large argument. To evaluate the performance of (5.7), comparisons with the numerical methods introduced in Section 2.10 are performed by using some random combinations of parameters. The graphs of the mixing distributions obtained by using (1) Post-Widder formula; (2) Gaver-Stehfest algorithm; (3) Laguerre series are shown in Figure 5.3 and Figure 5.4.

In Figure 5.3, the blue line that represents the Laguerre series completely conincides with the red line from the Gaver-Stehfest algorithm and is not visible in the diagram. The Post-Widder method with n = 15 took more than 100 seconds to obtain a not so accurate approximation where as the other two methods took just a few seconds.

Figure 5.4, on the other hand, shows some undesirable results when the computation becomes unstable. In (a), the Post-Widder formula produces a spike near $x=1$, whereas the results of two other methods coincide. In (b), all three methods shows high degree of agreement for small *x*, for larger *x*, the Laguerre series become unstable and exhibits oscillation behaviour and also produces negative values.

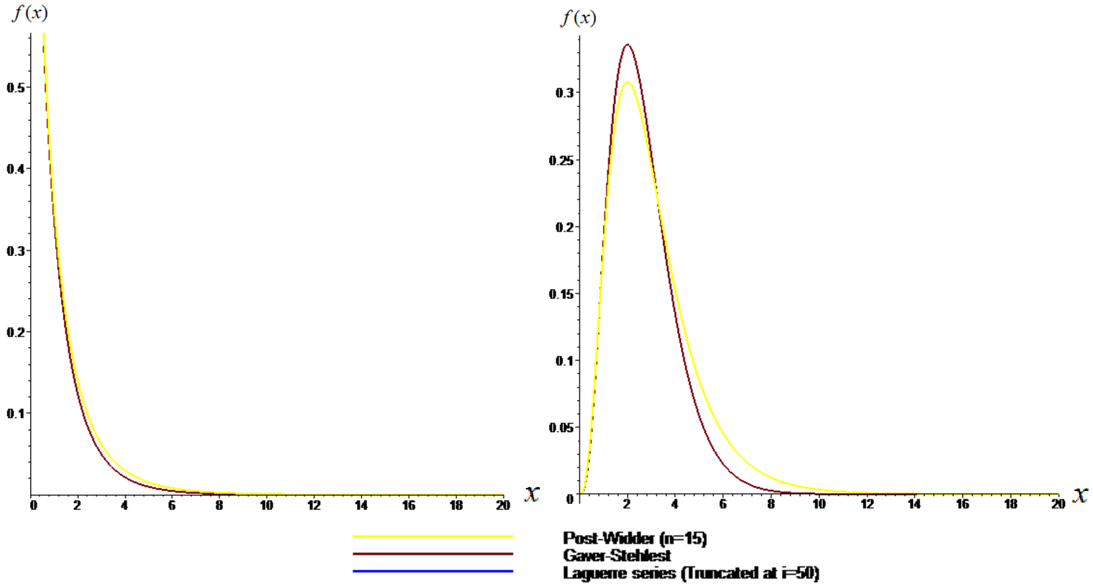

Figure 5.3

Inversions of Laplace transform of mixing distribution of ITD

(a) $\lambda = 0.966, p = 0.5, q = 0.487, r = 0.013$ (b) $\lambda = 4, p = 0.6, q = 0.39997, r = 0.00003$

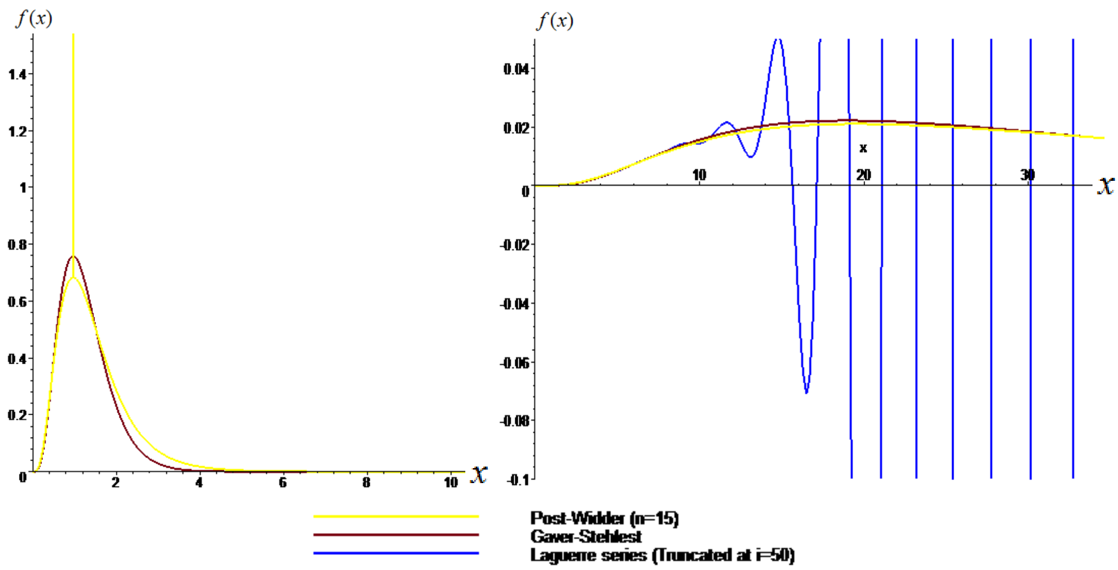

Figure 5.4

Inversions of Laplace transform of mixing distribution of ITD

(a) $\lambda = 4.982, p = 0.8, q = 0.1981, r = 0.0019$ (b) $\lambda = 7.5, p = 0.33, q = 0.5, r = 0.17$

CHAPTER 6 CONCLUSION AND FUTURE WORK

A number of generalised distributions, both old and new, have been considered in this research and some new results have been obtained. These models generalise some classical distributions and have shown advantages in describing the variations observed in some real life datasets. Judging from the goodness-of-fit statistics to the tested datasets, the HLZD has shown good competency among other generalisations of the LD. The proposed pgf based estimation method performed well with the HLZD on many datasets reflected in goodness-of-fit that are comparable to that given by the ML estimation. A simple data graduation scheme, which can be easily applied to other families of distribution with similar property, is proposed to avoid the arbitrariness in handling grouped data when constructing the empirical pgf and the result obtained is encouraging. Although the method works well in the given examples, large scale simulation experiment and sensitivity analysis of the method to the construction of the augmented pgf deserves further investigations.

The Poisson-HLZD has been constructed and its properties are explored from different aspects. Although some of the probabilistic properties are true only under certain conditions, they are not too restrictive as they are satisfied by most of the well-known distributions in the family.

In deriving the recursive formula for the pmf, a number of methods that are scattered in various disciplines have been applied, hence building up some connections among the different approaches. However, as mentioned in Chapter 4, the pmf of the Poisson-HLZD has a k-term recurrence formula where k is varying. This varying k makes the ML estimation procedure time consuming and require more memory especially when the data has long tail. Implementation of the pgf based estimation method as described in Chapter 3 can be considered in the future to reduce the computing power. On the other hand, the De Pril's formula in Appendix B is an application of the Sundt and Jewell's recursion for the Poisson case which has a 2-term recurrence formula. When a stopped sum distribution is stopped by a random variable with distribution that has a k-term recurrence formula in specific form for a constant k, the stopped sum distribution can then be evaluated recursively. The advantage of this type of recursion is evident. With a varying k, high-ordered convolution may involve whereas for a fixed k, only a maximum of k-convolution is needed. This drawback hence restricted the willingness of using such distribution to model quantity such as the claim number. It is therefore of practical interest to derive a k-term recurrence formula for the pmf of the Poisson-HLZD for some fixed k although the existence of such formula is still unclear. There also exists in some literatures, recursive formula for aggregate sum that is stopped by a MPD with specific form. Since the Poisson-HLZD has been proved to have such formulation, it provides an alternative way to tackle the problem. The difficulty in this approach is that, the mixing distribution for Poisson-HLZD is expressed in Laplace transform, justifying the condition of existing result is not a trivial problem.

The Poisson-HLZD and ITD share some common characteristics. Both distributions are PSSD with cluster size distributions that are related as described in Section 3.3. It is therefore natural to study the possible relation between the tail probabilities of the Poisson-HLZD and that of the ITD.

Following the MP property of the Poisson-HLZD and ITD, a few new distributions has occurred as by-products in this research. Two mixing distributions arise from the Poisson-HLZD and the ITD are supposed to have generalised the gamma distribution in some ways. Another possibly new distribution appears in the decomposition of the Poisson-HLZD following its self-decomposability property.

Investigating the properties of these ‘new’ distributions could lead to new research directions. From computational point of view, the inversion of the mixing distribution is not always a smooth process. Problems like singularities or convergence are definitely some important considerations that the numerical analysts would like to investigate.

REFERENCES

Abate, J., Choudhury, G. L. & Whitt, W. (1996). On the Laguerre Method for Numerically Inverting Laplace Transforms. *INFORMS Journal on Computing, 8*(4), 413-427.

Abate, J. & Whitt, W. (1999). Infinite-Series Representations of Laplace Transforms of Probability Density Functions for Numerical Inversion. *Journal of the Operations Research Society of Japan*, *42*(3), 268-285.

Abate, J. & Whitt, W. (2006). A Unified Framework for Numerically Inverting Laplace Transforms. *INFORMS Journal on Computing, 18*(4), 408-421.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723.

Aksenov, S.V., Savageau, M.A., (2005). Some properties of the Lerch family of discrete distributions. Preprint downloaded from arXiv:math.PR/054485v1.

Aoyama, K., Shimizu, K. & Ong, S. H. (2008). A First-Passage Time Random Walk Distribution with Five Transition Probabilities: A Generalization of the Shifted Inverse Trinomial. *Annals of the Institute of Statistical Mathematics, 60*(1), 1-20.

Bartle, R. G. (1976). *The Elements of Real Analysis*. New York: Wiley.

Bertin, E. & Theodorescu, R. (1995). Preserving Unimodality by Mixing. *Statistics & Probability Letters*, *25*(3), 281-288.

Bhattacharya, R. N. & Waymire, E. C. (2007). *A Basic Course in Probability Theory*. New York: Springer.

Bose, A., Dasgupta, A. & Rubin, H. (2002). A Contemporary Review and Bibliography of Infinitely Divisible Distributions and Processes. *Sankhya-the Indian Journal of Statistics Series A, 64*(3), 763-819.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control* (3rd ed.). Englewood Cliffs: Prentice Hall.

Bühlmann, H. (1970). *Mathematical Methods in Risk Theory*. New York: Springer-Verlag.

Calderon, F., Curilef, S. & Ladron de Guevara, M. L. (2009). Probability Distribution in a Quantitative Linguistic Problem. *Brazilian Journal of Physics, 39*(2A), 500-502.

Cameron, A. C. (1998). *Regression Analysis of Count Data*. Cambridge and New York: Cambridge University Press.

Casella, G. & Berger, R. L. (2002). *Statistical Inference* (2nd ed.). Pacific Grove: Thomson Learning.

Chadjiconstantinidis, S. & Pitselis, G. (2009). Further Improved Recursions for a Class of Compound Poisson Distributions. *Insurance: Mathematics and Economics, 44*(2), 278-286.

Charalambides, C. A. (2005). *Combinatorial Methods in Discrete Distributions*. Hoboken: Wiley-Interscience.

Chey, V. K. (2002). Comparison of Moth Diversity between Lightly and Heavily Logged Sites in a Tropical Rain Forest. *Malayan Nature Journal*, *56*(1), 23-41.

Clegg, R. G., Di Cairano-Gilfedder, C. & Zhou, S. (2010). A Critical Look at Power Law Modelling of the Internet. *Computer Communications, 33*(3), 259-268.

Cortina-Borja, M. (2006). Some Remarks on the Generalized Hermite and Generalized Gegenbauer Probability Distributions and Their Applications. In R. Köhler & P. Grzybek (Eds.), *Exact Methods in the Study of Language: Festchrift for the 75th Birthday of Gabriel Altmann* (pp. 49-60). Berlin and New York: de Gruyter.

Doray, L. G. & Luong, A. (1997). Efficient Estimators for the Good Family. *Communications in Statistics-Simulation and Computation, 26*(3), 1075-1088.

Douglas, J. B. (1980). *Analysis with Standard Contagious Distributions*. Fairland: International Co-operative Pub. House.

Dubner, H. & Abate, J. (1968). Numerical Inversion of Laplace Transforms by Relating Them to the Finite Fourier Cosine Transform. *J. ACM*, *15*(1), 115-123.

Engen, S. (1974). On Species Frequency Models. *Biometrika, 61*(2), 263-270.

Erdélyi, A., Magnus, W., Oberhettinger, F. & Tricomi, F. G. (1953). *Higher Transcendental Functions*. New York: McGraw-Hill.

Famoye, F. & Singh, K. P. (2006). Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data. *Journal of Data Science, 4*(1), 117-130.

Feller, W. (1967). *An Introduction to Probability Theory and Its Applications* (3d ed.). New York: Wiley.

Fisher, R. A., Corbet, A. S. & Williams, C. B. (1943). The Relation between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. *Journal of Animal Ecology, 12*, 42-58.

Freytag, A., Krüger, J. J., Meierrieks, D. & Schneider, F. (2011). The Origins of Terrorism: Cross-Country Estimates of Socio-Economic Determinants of Terrorism. *European Journal of Political Economy, 27, Supplement 1*(0), S5-S16.

Gan, L., Li, D. & Song, S. (2006). Is the Zipf Law Spurious in Explaining City-Size Distributions? *Economics Letters, 92*(2), 256-262.

Gathy, M. & Lefevre, C. (2010). On the Lagrangian Katz Family of Distributions as a Claim Frequency Model. *Insurance：Mathematics & Economics, 47*(1), 76-83.

Geedipally, S. R., Lord, D. & Dhavala, S. S. (2012). The Negative Binomial-Lindley Generalized Linear Model: Characteristics and Application Using Crash Data. *Accident Analysis and Prevention, 45*, 258-265.

Ghitany, M. E. & Al-Awadhi, S. A. (2001). A Unified Approach to Some Mixed Poisson Distributions. *Tamsui Oxford Journal of Mathematical Sciences, 17*(2), 147-161.

Greenwood, M. & Yule, G. U. (1920). An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents. *Journal of the Royal Statistical Society, Series A, 83*, 255-279.

Gupta, P. L., Gupta, R. C., Ong, S.-H. & Srivastava, H. M. (2008). A Class of Hurwitz-Lerch Zeta Distributions and Their Applications in Reliability. *Applied Mathematics and Computation, 196*(2), 521-531.

Gupta, R. C. (1974). Modified Power Series Distribution and Some of Its Applications. *Sankhya-the Indian Journal of Statistics Series B, 36*(3), 288-298.

Gupta, R. C. & Ong, S. H. (2004). A New Generalization of the Negative Binomial Distribution. *Computational Statistics & Data Analysis, 45*(2), 287-300.

Gupta, R. C. & Ong, S. H. (2005). Analysis of Long-Tailed Count Data by Poisson Mixtures. *Communications in Statistics-Theory and Methods*, *34*(3), 557-573.

Gurland, J. (1957). Some Interrelations among Compound and Generalized Distributions. *Biometrika*, *44*(1/2), 265-268.

Hassanzadeh, H. & Pooladi-Darvish, M. (2007). Comparison of Different Numerical Laplace Inversion Methods for Engineering Applications. *Applied Mathematics and Computation, 189*(2), 1966-1981.

Hogg, R. V., McKean, J. W. & Craig, A. T. (2005). *Introduction to Mathematical Statistics* (6th ed.). Upper Saddle River.: Pearson Education.

Holgate, P. (1970). The Modality of Some Compound Poisson Distributions. *Biometrika*, *57*(3), 666-667.

Hougaard, P., Lee, M. L. T. & Whitmore, G. A. (1997). Analysis of Overdispersed Count Data by Mixtures of Poisson Variables and Poisson Processes. *Biometrics*, *53*(4), 1225-1238.

Hu, M.-C., Pavlicova, M. & Nunes, E. V. (2011). Zero-Inflated and Hurdle Models of Count Data with Extra Zeros: Examples from an Hiv-Risk Reduction Intervention Trial. *American Journal of Drug and Alcohol Abuse, 37*(5), 367-375.

Imoto, T. (2012). An Extension of $GIT_{3,1}$: A Convolution of Binomial Variable and Negative Binomial Variable. *Proceedings of the 1st ISM International Statistical Conference*, 13-20.

Jain, G. C. & Consul, P. C. (1971). A Generalized Negative Binomial Distribution. *SIAM Journal on Applied Mathematics*, *21*(4), 501-513.

Jain, G. C. & Gupta, R. P. (1973). A Logarithmic Series Type Distribution. *Trabajos de Estadistica, 24*, 99-105.

Joe, H. & Zhu, R. (2005). Generalized Poisson Distribution: The Property of Mixture of Poisson and Comparison with Negative Binomial Distribution. *Biometrical Journal, 47*(2), 219-229.

Johnson, N. L., Kemp, A. W. & Kotz, S. (2005). *Univariate Discrete Distributions* (3rd ed.). Hoboken, N.J.: Wiley.

Kabardov, M. M. & Ryabov, V. M. (2009). Acceleration of the Convergence of the Laguerre Series in the Problem of Inverting the Laplace Transform. *Computational Mathematics and Mathematical Physics, 49*(4), 579-588.

Karlis, D. & Xekalaki, E. (2005). Mixed Poisson Distributions. *International Statistical Review, 73*(1), 35-58.

Keilson, J. & Gerber, H. (1971). Some Results for Discrete Unimodality. *Journal of the American Statistical Association, 66*(334), 386-389.

Kemp, A. W. (2010). Families of Power Series Distributions, with Particular Reference to the Lerch Family. *Journal of Statistical Planning and Inference, 140*(8), 2255-2259.

Kemp, A. W. & Kemp, C. D. (1968). On a Distribution Associated with Certain Stochastic Processes. *Journal of the Royal Statistical Society. Series B, 30*(1), 160-163.

Kemp, C. D. & Kemp, A. W. (1988). Rapid Estimation for Discrete Distributions. *Journal of the Royal Statistical Society. Series D (The Statistician), 37*(3), 243-255.

Kempton, R. A. (1975). A Generalized Form of Fisher's Logarithmic Series. *Biometrika, 62*(1), 29-38.

Khang, T. F. & Ong, S. H. (2007). A New Generalization of the Logarithmic Distribution Arising from the Inverse Trinomial Distribution. *Communications in Statistics-Theory and Methods, 36*(1), 3-21.

Klugman, S. A., Panjer, H. H. & Willmot, G. E. (2008). *Loss Models : From Data to Decisions* (3rd ed.). Hoboken, N.J.: John Wiley & Sons.

Kocherlakota, S. & Kocherlakota, K. (1986). Goodness of Fit Tests for Discrete Distributions. *Communications in Statistics-Theory and Methods*, *15*(3), 815-829.

Lord, D. & Geedipally, S. R. (2011). The Negative Binomial–Lindley Distribution as a Tool for Analyzing Crash Data Characterized by a Large Amount of Zeros. *Accident Analysis & Prevention, 43*(5), 1738-1742.

Maceda, E. C. (1948). On the Compound and Generalized Poisson Distributions. *The Annals of Mathematical Statistics, 19*(3), 414-416.

Mainardi, F. & Rogosin, S. (2006). The Origin of Infinitely Divisible Distributions: From De Finetti's Problem to Levy-Khintchine Formula. *Mathematical Methods in Economics and Finance, 1*(1), 37-55.

Masol, V. & Teugels, J. L. (2010). Numerical Accuracy of Real Inversion Formulas for the Laplace Transform. *Journal of Computational and Applied Mathematics, 233*(10), 2521-2533.

Masse, J. C. & Theodorescu, R. (2005). Neyman Type A Distribution Revisited. *Statistica Neerlandica, 59*(2), 206-213.

Millar, R. B. (2011). *Maximum Likelihood Estimation and Inference: With Examples in R, SAS, and ADMB.* United Kingdom: John Wiley & Sons, Ltd.

Miller, K. S. & Samko, S. G. (2001). Completely Monotonic Functions. *Integral Transforms and Special Functions*, *12*(4), 389-402.

Misra, N., Singh, H. & Harner, E. J. (2003). Stochastic Comparisons of Poisson and Binomial Random Variables with Their Mixtures. *Statistics & Probability Letters, 65*(4), 279-290.

Neyman, J. (1939). On a New Class of "Contagious" Distributions, Applicable in Entomology and Bacteriology. *Annals of Mathematical Statistics, 10*(1), 35-57.

Omar Mohd Rijal, Norliza Mohd Noor, & Liew, K. W. (2005). Two-component mixture distributions for measuring area and compactness of intermetallic formation from the Au-Al wire bonding image. *WSEAS Transactions on Circuits and Systems*, *4*(9), 1085-1093.

Ong, S. H. (1987). Some Notes on the Non-Central Negative Binomial Distribution. *Metrika, 34*, 225-236.

Ong, S. H. (1988). A Discrete Charlier Series Distribution. *Biometrical Journal, 30*(8), 1003-1009.

Ong, S. H. (1995). Computation of Probabilities of a Generalized Log-Series and Related Distributions. *Communications in Statistics-Theory and Methods, 24*(1), 253-271.

Ong, S. H. & Toh, K. K. (2001). Cluster Size Distribution of the Non-Central Negative Binomial Distribution. *Communications in Statistics-Theory and Methods, 30*(11), 2415-2426.

Pawitan, Y. (2001). *In All Likelihood : Statistical Modelling and Inference Using Likelihood*. New York: Oxford University Press.

Perc, M. (2010). Zipf's Law and Log-Normal Distributions in Measures of Scientific Output across Fields and Institutions: 40 Years of Slovenia's Research as an Example. *Journal of Informetrics, 4*(3), 358-364.

Provost, S. B. (2005). Moment-Based Density Approximants. *The Mathematica Journal, 9*(4), 727-756.

Prudnikov, A. P., Brychkov, I. U. A. & Marichev, O. I. (1986). *Integrals and Series*. New York: Gordon and Breach Science Publishers.

Pudprommarat, C. & Bodhisuwan, W. (2012). Stochastic Orders Comparisons of Negative Binomial Distribution with Negative Binomial-Lindley Distribution. *Open Journal of Statistics, 2*(2), 208-212.

Puri, P. S. & Goldie, C. M. (1979). Poisson Mixtures and Quasi-Infinite Divisibility of Distributions. *Journal of Applied Probability, 16*(1), 138-153.

Quenouille, M. H. (1949). A Relation between the Logarithmic, Poisson, and Negative Binomial Series. *Biometrics, 5*, 162-164.

Quine, M. P. & Seneta, E. (1987). Bortkiewicz's Data and the Law of Small Numbers. *International Statistical Review, 55*(2), 173-181.

Ridout, M. S. & Besbeas, P. (2004). An Empirical Model for Underdispersed Count Data. *Statistical Modelling, 4*(1), 77-89.

Rigby, R. A., Stasinopoulos, D. M. & Akantziliotou, C. (2008). A Framework for Modelling Overdispersed Count Data, Including the Poisson-Shifted Generalized Inverse Gaussian Distribution. *Computational Statistics & Data Analysis, 53*(2), 381-393.

Robert, C. P. & Casella, G. (2004). *Monte Carlo Statistical Methods* (2nd ed.). New York: Springer.

Roberts, G. E. & Kaufman, H. (1966). *Table of Laplace Transforms*. Philadelphia: Saunders.

Roman, S. (2005). *The Umbral Calculus*. Mineola, New York: Dover Publications.

Ross, S. M. (1996). *Stochastic Processes* (2nd ed.). New York: Wiley.

Ruohonen, M. (1988). A Model for the Claim Number Process. *Astin Bulletin, 18*(1), 57-68.

Sato, K.-i. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge and New York: Cambridge University Press.

Saxena, R. K., Pogány, T. K., Saxena, R. & Jankov, D. (2011). On Generalized Hurwitz-Lerch Zeta Distributions Occuring in Statistical Inference. *Acta Univ. Sapientiae, Mathematica, 3*(1), 43-59.

Shaked, M. & Shanthikumar, J. G. (2007). *Stochastic Orders*. New York: Springer.

Shaked, M. & Wong, T. (1997). Stochastic Orders Based on Ratios of Laplace Transforms. *Journal of Applied Probability, 34*(2), 404-419.

Shao, J. (2003). *Mathematical Statistics* (2nd ed.). New York: Springer.

Shimizu, K., Nishii, N. & Minami, M. (1997). The Multivariate Inverse Trinomial Distribution as a Lagrangian Probability Model. *Communications in Statistics-Theory and Methods, 26*(7), 1585-1598.

Shimizu, K. & Yanagimoto, T. (1991). The Inverse Trinomial Distribution. *Japanese Journal of Applied Statistics, 20*(2), 89-96.

Sim, S. Z. & Ong, S. H. (2010). Parameter Estimation for Discrete Distributions by Generalized Hellinger-Type Divergence Based on Probability Generating Function. *Communications in Statistics-Simulation and Computation, 39*(2), 305-314.

Steutel, F. W. (1973). Some Recent Results in Infinite Divisibility. *Stochastic Processes and their Applications, 1*(2), 125-143.

Steutel, F. W. & van Harn, K. (1979). Discrete Analogues of Self-Decomposability and Stability. *The Annals of Probability, 7*(5), 893-899.

Steutel, F. W. & van Harn, K. (2004). *Infinite Divisibility of Probability Distributions on the Real Line*. New York: Marcel Dekker.

Sugita, T., Shimizu, K., Ong, S. H. & Ng, C. M. (2011). A Family Which Integrates the Generalized Charlier Series and Extended Noncentral Negative Binomial Distributions. *Communications in Statistics-Theory and Methods, 40*(11), 1977-1998.

Tagliani, A. & Velasquez, Y. (2004). Inverse Laplace Transform for Heavy-Tailed Distributions. *Applied Mathematics and Computation, 150*, 337-345.

Thyrion, P. (1960). Contribution Á L'étude Du Bonus Pour Non Sinistre En Assurance Automobile. *Astin Bulletin, 1*(3), 142-162.

Titterington, D. M., Smith, A. F. M. & Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester; New York: Wiley.

Tohka, J., Krestyannikov, E., Dinov, I. D., Graham, A. M., Shattuck, D. W., Ruotsalainen, U. & Toga, A. W. (2007). Genetic Algorithms for Finite Mixture Model Based Voxel Classification in Neuroimaging. *IEEE Transactions on Medical Imaging, 26*(5), 696-711.

Tomovski, Ž., Saxena, R. K. & Pogány, T. K. (2012). Probability Distributions Associated with Mathieu Type Series. *ProbStat Forum, 5*, 86-96.

Tripathi, R. C. & Gupta, R. C. (1984). Statistical Inference Regarding the Generalized Poisson Distribution. *Sankhya-the Indian Journal of Statistics Series B*, *46*, 166-173.

Tripathi, R. C. & Gupta, R. C. (1988). Another Generalization of the Logarithmic Series and the Geometric Distributions. *Communications in Statistics-Theory and Methods, 17*(5), 1541-1547.

van der Hoeven, N. (2005). The Probability to Select the Correct Model Using Likelihood-Ratio Based Criteria in Choosing between Two Nested Models of Which the More Extended One Is True. *Journal of Statistical Planning and Inference, 135*(2), 477-486.

Vargo, E., Pasupathy, R. & Leemis, L. (2010). Moment-Ratio Diagrams for Univariate Distributions. *Journal of Quality Technology, 42*(3), 276-286.

Widder, D. V. (1941). *The Laplace Transform*. Princeton: Princeton University Press.

Willmot, G. E. & Sundt, B. (1989). On Evaluation of the Delaporte Distribution and Related Distributions. *Scandinavian Actuarial Journal*, *1989*(2), 101-113.

Wimmer, G. & Altmann, G. (1996). The Multiple Poisson Distribution, Its Characteristics and a Variety of Forms. *Biometrical Journal, 38*(8), 995-1011.

Yanagimoto, T. (1989). The Inverse Binomial Distribution as a Statistical Model. *Communications in Statistics-Theory and Methods, 18*(10), 3625-3633.

Yip, K. C. H. & Yau, K. K. W. (2005). On Modeling Claim Frequency Data in General Insurance with Extra Zeros. *Insurance: Mathematics & Economics, 36*(2), 153-163.

Zhang, J., Chen, Q. & Wang, Y. (2009). Zipf Distribution in Top Chinese Firms and an Economic Explanation. *Physica A: Statistical Mechanics and its Applications, 388*(10), 2020-2024.

Zhang, Z. J., Ong, S. H., Lynn, H. S., Peng, W. X., Zhou, Y. B., Zhao, G. M. & Jiang, Q. W. (2008). Generalized Negative Binomial Distribution: A Promising Statistical Distribution for Oncomelania Hupensis in the Lake- and Marsh-Land Regions of China. *Annals of Tropical Medicine and Parasitology, 102*(6), 541-552.

Zörnig, P. & Altmann, G. (1995). Unified Representation of Zipf Distributions. *Computational Statistics & Data Analysis*, *19*(4), 461-473.

APPENDIX A: DERIVATION OF THE PROBABILITY MASS FUNCTION FOR POISSON-HURWITZ-LERCH ZETA DISTRIBUTION USING THEOREM 2.5

Consider the pgf of the Poisson-HLZD in (4.1) expressed in the form (2.12)

$$\exp\left[\lambda\left(\frac{t\Phi(\theta t, s+1, a+1)}{\Phi(\theta, s+1, a+1)} - 1\right)\right] = \sum_{k=0}^{\infty}\frac{k!\,p_k t^k}{k!} = \sum_{k=0}^{\infty}\frac{p_k(x)t^k}{k!}.$$

Take $x = \dfrac{\lambda}{\Phi(\theta, s+1, a+1)}$, $\overline{f}(t) = t\Phi(\theta t, s+1, a+1)$ and $p_n(x) = e^{\lambda} n!\, p_n$. From Theorem 2.5,

$$\begin{aligned}
p_{n+1}(x) &= x\lambda_f[t\Phi(\theta t, s+1, a+1)]'\, x^n \\
&= x\lambda_f[t\Phi'(\theta t, s+1, a+1) + \Phi(\theta t, s+1, a+1)]x^n \\
&= x\lambda_f[\sum_{k=1}^{\infty}\frac{k(\theta t)^{k-1}}{(a+k)^{s+1}}]x^n \\
&= x\lambda_f[\sum_{k=1}^{n+1}\frac{k\theta^{k-1}}{(a+k)^{s+1}}n^{(k-1)}x^{n-k+1}] \\
&= x\sum_{k=1}^{n+1}\frac{k\theta^{k-1}}{(a+k)^{s+1}}n^{(k-1)}p_{n-k+1}(x) \\
&= x\sum_{j=0}^{n}\frac{(j+1)\theta^{j}}{(a+1+j)^{s+1}}n^{(j)}p_{n-j}(x) \\
&= x\sum_{j=0}^{n}\frac{(j+1)\theta^{j}}{(a+1+j)^{s+1}}\frac{n!}{(n-j)!}p_{n-j}(x).
\end{aligned}$$

Substitute $x = \dfrac{\lambda}{\Phi(\theta, s+1, a+1)}$ and $p_n(x) = e^{\lambda} n!\, p_n$, gives (4.4).

APPENDIX B: DERIVATION OF THE PROBABILITY MASS FUNCTION FOR POISSON-HURWITZ-LERCH ZETA DISTRIBUTION USING DE PRIL'S FORMULA

The following theorem of De Pril is given in Chadjiconstantinidis & Pitselis (2009).

Theorem: Let $G_Y(t)=\sum_{k=1}^{\infty} q_k t^k$ be the pgf of the claim amounts. If its derivative is of the form $G_Y'(t)=\frac{\sum_{k=0}^{\infty} a(k)t^k}{\sum_{k=0}^{\infty} b(k)t^k}$, with $b(0)=1$, then the aggregate claim stopped by a Poisson random variable with mean λ has pmf p_k which can be recursively calculated by $p_n=\frac{1}{n}\sum_{k=1}^{n}\left[\lambda a(k-1)-(n-k)b(k)\right]p_{n-k}$ with $p_0=e^{-\lambda}$.

In the Poisson-HLZD, $G_Y(t)=\sum_{k=1}^{\infty}\frac{(\theta t)^k}{T(\theta,s,a)(a+k)^{s+1}}$ and $G_Y'(t)=\sum_{k=1}^{\infty}\frac{k\theta(\theta t)^{k-1}}{T(\theta,s,a)(a+k)^{s+1}}$ satisfy the condition in the theorem with $a(k)=\frac{(k+1)\theta^{k+1}}{T(\theta,s,a)(a+k+1)^{s+1}}$. Hence,

$$p_n=\frac{1}{n}\sum_{k=1}^{n}\left[\lambda\frac{k\theta^k}{T(\theta,s,a)(a+k)^{s+1}}\right]p_{n-k}$$

$$=\frac{1}{n}\sum_{k=1}^{n}\left[\lambda\frac{k\theta^{k-1}}{\Phi(\theta,s+1,a+1)(a+k)^{s+1}}\right]p_{n-k}$$

$$=\frac{1}{n}\frac{\lambda}{\Phi(\theta,s+1,a+1)}\sum_{k=0}^{n-1}\left[\frac{(k+1)\theta^k}{(a+k+1)^{s+1}}\right]p_{n-k-1}$$

APPENDIX C: MAPLE PROGRAMME FOR THE CONSTRUCTION OF MOMENT-RATIO DIAGRAM

```
with(Statistics):
with(stats[statplots]):
with(stats[random]):
K:=[]:U:=[]:

k:=(t*diff(LerchPhi(t,s-1,a+1)/LerchPhi(t,s+1,a+1)-
2*a*LerchPhi(t,s,a+1)/LerchPhi(t,s+1,a+1)+a^2,t)+(LerchPhi(t,s-
1,a+1)/LerchPhi(t,s+1,a+1)-
2*a*LerchPhi(t,s,a+1)/LerchPhi(t,s+1,a+1)+a^2)*(LerchPhi(t,s,a+1)/Lerc
hPhi(t,s+1,a+1)-a))/(l^0.5*((LerchPhi(t,s-1,a+1)/LerchPhi(t,s+1,a+1)-
2*a*LerchPhi(t,s,a+1)/LerchPhi(t,s+1,a+1)+a^2))^1.5):

u:=3+(t*diff(t*diff(LerchPhi(t,s-1,a+1)/LerchPhi(t,s+1,a+1)-
2*a*LerchPhi(t,s,a+1)/LerchPhi(t,s+1,a+1)+a^2,t)+(LerchPhi(t,s-
1,a+1)/LerchPhi(t,s+1,a+1)-
2*a*LerchPhi(t,s,a+1)/LerchPhi(t,s+1,a+1)+a^2)*(LerchPhi(t,s,a+1)/Lerc
hPhi(t,s+1,a+1)-a),t)+(t*diff(LerchPhi(t,s-1,a+1)/LerchPhi(t,s+1,a+1)-
2*a*LerchPhi(t,s,a+1)/LerchPhi(t,s+1,a+1)+a^2,t)+(LerchPhi(t,s-
1,a+1)/LerchPhi(t,s+1,a+1)-
2*a*LerchPhi(t,s,a+1)/LerchPhi(t,s+1,a+1)+a^2)*(LerchPhi(t,s,a+1)/Lerc
hPhi(t,s+1,a+1)-a))*(LerchPhi(t,s,a+1)/LerchPhi(t,s+1,a+1)-
a))/(l*( LerchPhi(t,s-1,a+1)/LerchPhi(t,s+1,a+1)-
2*a*LerchPhi(t,s,a+1)/LerchPhi(t,s+1,a+1)+a^2)^2):

for i from 1 to 110 do
      a:=-1+i*0.1:
      t:=0.99:ans:=fsolve({l1*(LerchPhi(t,s1,a+1)/LerchPhi(t,s1+1,a+1)
      -a)=1,l1*(LerchPhi(t,s1-1,a+1)-
      a*LerchPhi(t,s1,a+1))/LerchPhi(t,s1+1,a+1)=2+a},{l1,s1}):
      l:=subs(ans,l1):
      s:=subs(ans,s1):
      p1:=evalf(k):
      p2:=evalf(u):
      K:=[op(K),p1]:
      U:=[op(U),p2]:
od;
xylist:=zip((x,y)->[x,y],K,U):
plot(xylist,style=line);
```