New Classes of Multivariate Gamma Survival and Reliability Models

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

Norou Dini Diawara

Certificate of Approval:

Asheber Abebe
Assistant Professor
Mathematics and Statistics

Mark Carpenter, Chair
Associate Professor
Mathematics and Statistics

Nedret Billor
Associate Professor
Mathematics and Statistics

Olav Kallenberg
Professor
Mathematics and Statistics

Jerzy Szulga
Professor
Mathematics and Statistics

Stephen L. McFarland
Dean
Graduate School

New Classes of Multivariate Gamma Survival and Reliability Models

Norou Dini Diawara

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama
August 7, 2006

New Classes of Multivariate Gamma

Survival and Reliability Models

Norou Diawara

Signature of Author

August 7, 2006

Date of Graduation

Vita

Norou Dini Diawara, son of Moustapha Bassirou Diawara and Khady Diatou Dieng, was born in Ziguinchor, Sénégal on November 15, 1971. He attented public school in Dakar, Senegal. After graduating from Lamine Gueye's High School, he entered Cheick Anta Diop University (U.C.A.D.) in September 1991, from which he received his B.S. in Mathematics and Physics in 1993. He started his graduate studies in the Department of Mathematics at U.C.A.D. With the help of his family and uncle Alioune Diop, he went on to complete his Licence and Maîtrise at the University of Le Havre in Le Havre, France in 1996. He continued his studies at the University of South Alabama, in Mobile, AL, USA where he earned his M.S. in Mathematics and Statistics in 1999. After two years at the University of Georgia, in Athens, GA. as graduate teaching assistant and working for the Canadian Consulate General in Atlanta, GA., he came to Auburn University in August 2002 where he completed his Ph.D. in Mathematics and Statistics in 2006.

Dissertation Abstract

New Classes of Multivariate Gamma

Survival and Reliability Models

Norou Dini Diawara

Doctor of Philosophy, August 7, 2006
(M.S., University of South Alabama, Mobile,AL., 1999)
(M.S., Le Havre University, Le Havre, France, 1996)

121 Typed Pages

Directed by Mark Carpenter

Multivariate modeling and analysis based on the multivariate normal distribution is well established and widely used. However, when the marginal distributions have only a positive support, such as time-to-event models, that are positively skewed, often the multivariate normal theory and resulting approximations fail. Accordingly, over the last fifty years, thousands of papers have been published suggesting many ways of generating families of positive support multivariate distributions, such as gamma, Weibull and exponential. As evidenced by recent literature, this quest is still rigorously pursued even today. In this dissertation, we provide a large and flexible class of multivariate gamma distributions that contains both absolutely continuous and discontinuous distributions on the positive hypercube support. All of these models are applicable to the area of reliability and survival modeling.

Acknowledgments

I would like to thank my advisor, Dr. Carpenter, for his academic guidance, generous advice, his sharp comments and support, during our many discussions. I am very appreciative of his encouragement and patience when things seemed fuzzy. Without his support, this thesis would not have been completed. It has been a privilege working with him.

I also wish to express my sincere gratitude to the members of my committee: Dr. Ash Abebe, Dr. Nedret Billor, Dr. Olav Kallenberg and Dr. Jerzy Szulga for their time, patience and suggestions that led to me improving this work. I have truly enjoyed working with each one of them, and I really appreciate their full support. Special thanks to Dr. Olav Kallenberg for the knowledge I have gained through the seminars and years of research we had started. I thank him for his continuous support since I came to Auburn University, his kindness, time, and collaboration. Thanks to the faculty and staff of our Department, and to my fellow students for the great experience at Auburn.

I am deeply indebted to my family, who has always supported me, and specially to my parents and my siblings Papa Amadou, Amadou Belal, Ndeye Maguette, and Ndeye Fatou, for their love, guidance and vision. Special thanks to my uncle Alioune Diop and my aunt Ndiaya Mbengue for their kindness, support and encouragement. I am grateful to my wife, Oumy Ba, for her endless love, friendship, support and encouragement. I thank her for her patience during all these stages when we had to spend time apart, and her alone taking care of our baby Sidy Manoumbé. She is an inspiration to me.

I dedicate this dissertation to my beloved brother Sidy Manoumbé, and all those who truly believed in me, encouraged and supported me during this long path.

Style manual or journal used Transactions of the American Mathematical Society (together with the style known as "auphd"). Bibliography follows van Leunen's *A Handbook for Scholars.*

Computer software used The document preparation package TEX (specifically LATEX) together with the departmental style-file `auphd.sty`.

Table of Contents

List of Figures

Abbreviations

Abbreviation	Description
iid	Independent and identically distributed
EM	Expectation Maximization
LST	Laplace-Stieltjes transform
MCMC	Markov chain Monte Carlo
MLE	Maximum Likelihood Estimator
MSE	Mean Square Error
cdf	Cumulative density function
pdf	Probability density function (or mass function in discrete case)
pp.	Pages
prob.	Probability
et al.	and others
a.s.	almost surely
i.e.	that means
r.s.	Random sample
r.v.	Random variable
$a \wedge b$	Minimum between a and b
$a \vee b$	Maximum between a and b

Notations

Notation	Description
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^+$	Set of positive real numbers
$\mathbb{Z} = \{0, \pm 1, ..\}$	Set of integer numbers
$\mathbb{N} = \{1, 2, \ldots\}$	Set of positive integer numbers
X	Random variable
p	Dimension of multivariate distribution
$[a]$	The entire part of a number a
$\mathcal{P}$	The power set
$\lvert A \rvert$	Cardinal of a set A
$A_{(k)}$	k^{th} element of the set A
$A^c_{(k)}$	k^{th} element of the set A^c, complement set of A
$E(X)$, $Var(X)$	Expectation and variance of the r.v. X
$Ga(\mu, \lambda, \alpha)$	Gamma distribution with location μ, scale λ, and shape α
$\Gamma(\alpha)$	Gamma function with parameter α
θ	Parameter set of full model
$\hat{\theta}$	Maximum likelihood estimate for parameter set of full model
$F_{X\|Y}$ and $f_{X\|Y}$	conditional cdf and pdf, respectively
$X \sim f(x)$	The random variable X follows the density $f(x)$
L_X	The Laplace transform of the random variable X
$\propto$	Proportional to

Chapter 1

Introduction

1.1 Motivation

Multivariate survival or reliability analysis refers to the multivariate modeling of the different times to multiple events that are recorded on the same experimental unit. In such cases, when more than one time-to-event is observed for each individual under study, it is reasonable to assume that the event times are dependent. However, it is common practice to assume "working independence" (see Lawless (2003) [56] for example) where each event is analyzed separately, while ignoring this dependence (effectively treating them as independent events). This seemingly pragmatic approach can lead to confusing interpretations and incorrect results. Carpenter et al. (2006) [12] showed that ignoring dependence between variables comes at a great cost in terms of bias, mean square error (MSE), and other objective criteria in the bivariate exponential case.

This idea is quite intuitive, because ignoring dependence requires maximizing the wrong or misspecified likelihood function. That is, the maximum likelihood estimators (MLE's) of the parameters based on the marginal likelihoods are not necessarily the same as the MLE's derived from the joint likelihood.

Another problem with ignoring dependence is that such an analysis would require one to adjust for the multiple testing problem. In clinical trials, the Food and Drug Administration, for example, routinely requires conservative adjustments for multiple endpoints such as a primary endpoint of time-to-death due to a brain tumor, and a secondary endpoint of time-to-recurrence of the tumor after surgery.

Because the marginal distributions are of positive support and often very positively skewed, treating the observations as multivariate normal can lead to incorrect results and/or convergence issues. For example, the multivariate mixture model is used in many applications for unsupervised classification. Hougaard (2001) [37] advises cautions using the multivariate normal when the underlying distribution does not apply to it.

Also, these methods can be computationally intensive (see Borovkov and Utev (1984) [7], Chen (1982) [14], Mclachlan and Peel (2000) [67], Johnson and Wichern (1998) [48]) even if the underlying latent populations are multivariate normal. In fact, Mclachlan and Peel (2000) [67] provide a whole chapter on mixture models for failure time data.

Therefore, over the last fifty years there has been great interest in deriving and characterizing multivariate distributions with positive support and skewed marginals. A few of the many widely known references are from Arnold (1967) [2], Barlow and Proschan (1965) [5], Fang (1990) [24], Furman (2005) [28], Gosh and Gelfand (1998) [30], Hanagal (1996) [32], Hougaard (2000) [36], Hougaard (2001) [37], Joe (1997) [44], Kotz et al. (2000) [53], Lee (1979) [57], Lu and Bhattacharyya (1990) [59], Marshall and Olkin (1967) [61], Mathai and Moschopoulos (1991) [63], Mathai and Moschopoulos (1992) [64], Moran (1969) [69], Sivazlian (1981) [76], and Walker and Stephens (1999) [79].

In this dissertation, we provide a large and flexible class of multivariate gamma distributions that contains both absolutely continuous and discontinuous distributions on the positive hypercube support. All of these models are applicable to the area of reliability and survival modeling. Also, since this theoretical research resulted from a real world application we provide methods for parameter estimation for a variety of applications.

1.2 Research Question

Our focus is on generating multivariate survival and reliability models with specified marginal distributions of the same family, i.e., exponential, Weibull and gamma. One common use of these models, in survival analysis, is to examine the survival times of patients in medical or biological settings, based on factors such as treatment, age, education, to predict the time to death, or time to certain events. In the industrial engineering setting, the study of time to failure of devices, machines or components in a system, are referred to as reliability analysis. Although the applications are very different, these models can be used in both contexts of survival and reliability analysis.

Our approach to creating a dependence structure between multiple events, is to indirectly associate the random variables (r.v.'s) linearly with latent, unobservable r.v.'s. Specially, let p be some fixed integer, and $X_0, X_1, ..., X_p$ be r.v.'s with specified distributions of the same family, say $f(x|\theta)$, that are linked through the mutually independent r.v.'s $Z_1, Z_2, ..., Z_p$ in the following way:

$$X_i = a_i X_0 + Z_i \tag{1.1}$$

where a_i's are some nonnegative constants, and Z_i independent of X_0 for $i = 1, 2, \cdots, p$. Note that $X_0, Z_1, Z_2, \cdots, Z_p$ are considered latent, unobservable r.v.'s , that generate the observable multivariate vector $\mathbf{X} = (X_1, X_2, \cdots, X_p)'$.

Much of the research in this dissertation is on the derivation and characterization of the classes of distributions of $Z_1, Z_2, \cdots, Z_p$, that produce the specified marginal distributions of $X_1, X_2, \cdots, X_p$, respectively. We derive and study estimators of the parameters under the resulting dependence structure.

Interestingly, if we treat the latent variables, $X_0, Z_1, \cdots, Z_p$, as missing observations, then methods based on conditional expectation and/or expectation-maximization (EM) algorithms have great potential in estimating the parameters associated with the latent population. We develop such approaches as well as a missing value substitution techniques in improving our estimates of the parameters of the marginal distributions and parameters associated with the latent terms.

Since in this dissertation we focus on a new multivariate gamma distribution, in the sections that follow, we review the univariate gamma and other multivariate gammas in their various forms found in the literature and in practice.

1.3 The Univariate Gamma Distribution

Positively skewed distributions on a positive support occur quite often in practical applications such as in reliability and survival analysis. Lawless (2003) [56] gives some examples of the most common models, and their applications. The gamma family of distributions is right tailed, and seems a natural choice for positive valued distributions with heavy tails. It is widely used in reliability and survival. For example, the gamma distribution is used in statistical studies of times between earthquakes. So estimation of its parameters is very important.

The gamma distribution belongs to the class of exponential distributions on the real line with respect to Lebesgue measure. See McCullagh and Nelder (1989) [66] for example.

This family plays an important role in many areas of probability and statistics through the normal, exponential, gamma, Poisson, binomial distributions, and many other ones.

The exponential family distribution is a large class that allow us to handle a wide category of distributions, and many of the properties are similar to the normal distribution. Lehmann(1986) [58] gives some results for that class of distributions.

McCullagh and Nelder (1989) [66] give some examples of models where the gamma distribution is used in the regression structure under generalized linear model (GLM). GLM allows a unified theory for many models in practical statistics, including models for gamma responses and survival data.

We consider the univariate three parameter gamma distribution X, denoted here as $Ga(\mu, \lambda, \alpha)$. If $\mu = 0$, we denote the two parameter gamma distribution X as $Ga(\lambda, \alpha)$.

Definition 1.1 *The three parameter gamma distribution, see Cohen and Whitten (1986) [16] and Billingsley (1986) [6] for examples, is defined by the density function* $f_X(.;\mu,\lambda,\alpha) : t \in (\mu,\infty) \mapsto f_X(t;\mu,\lambda,\alpha) \in \mathbb{R}^+$ *of the three parameters* $\mu \in \mathbb{R}, \lambda > 0$ *and* $\alpha > 0$ *given as*

$$f_X(t;\mu,\lambda,\alpha) = \frac{\lambda^\alpha}{\Gamma(\alpha)}(t-\mu)^{\alpha-1}e^{-\lambda(t-\mu)}I_{[\mu,\infty)}(t) \tag{1.2}$$

where

$\mu \in \mathbb{R}, \quad \lambda > 0,$ *and* $\alpha > 0$ *are the location, scale and shape parameters, respectively.*
$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ *is called the gamma function.*

The three parameter gamma distribution is also known as Type III of distributions in the Pearson's system of distributions. Rather than considering the distribution at the origin, it is sometimes more appropriate to shift it on the real axis, allowing the domain to be $[\mu, \infty)$, as it is done in (1.2). The scale parameter λ is also called the rate or inverse scale under inverse parametrization. The smaller the value of λ, the wider the spread, and vice versa. Whereas for the normal distribution, the scale parameter is the standard deviation. When $0 < \alpha < 1$, the distribution has a vertical asymptote. When $\alpha > 1$, the distribution has a mode at $\dfrac{\alpha - 1}{\lambda} + \mu$ and heavy tails. See Figure 1.1 and Figure 1.2 below for the cdf and pdf, respectively.

The gamma function $\Gamma(\alpha) = \displaystyle\int_0^\infty x^{\alpha-1}e^{-x}dx$ with $\Gamma\left(\dfrac{1}{2}\right) = \sqrt{\pi}$ is the normalizing constant of the distribution, and satisfies the relation $\alpha\Gamma(\alpha) = \Gamma(\alpha + 1)$. The incomplete gamma is defined as $I(\alpha, t) = \dfrac{1}{\Gamma(\alpha)}\displaystyle\int_0^t x^{\alpha-1}e^{-x}dx$, as in Lawless (2003) [56], and $\Gamma(\alpha, t) = \dfrac{1}{\Gamma(\alpha)}\displaystyle\int_t^\infty x^{\alpha-1}e^{-x}dx$ is called the complementary incomplete gamma function.

Several special cases of the gamma , such as the Erlang or the exponential, exist, and are widely used. The gamma distribution of (1.2) with location at the origin and scale 1 is known as the standard exponential distribution. Most statistical packages, such as SAS® and S-Plus generate standard exponential distributions. Its pdf is:

$$f(t) \quad = \quad \lambda e^{-\lambda t}, \quad \text{where } \lambda > 0, \text{ and } t \geq 0. \tag{1.3}$$

The displaced exponential is the exponential shifted by some value μ, denoted $\exp(\mu, \lambda)$ and defined as

$$f(t) \quad = \quad \lambda e^{-\lambda(t-\mu)}, \quad \text{where } \lambda > 0, \text{ and } t \geq \mu. \tag{1.4}$$

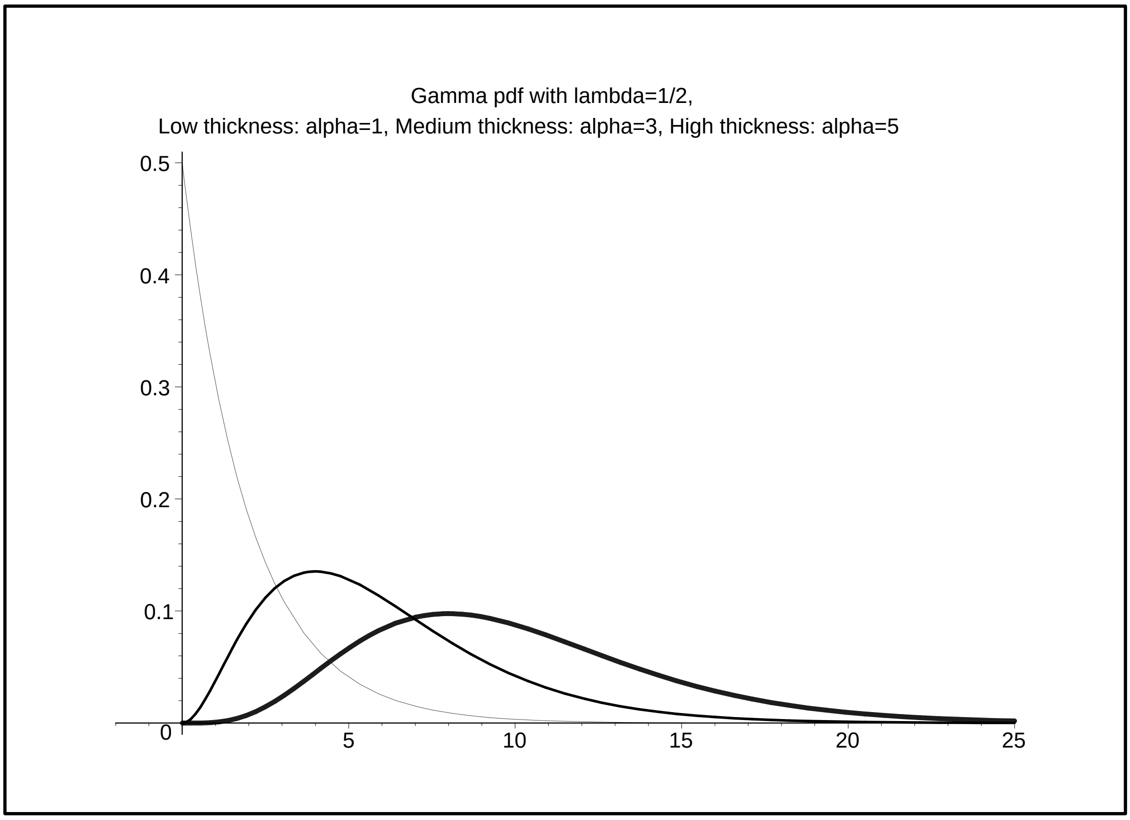

Figure 1.1: Graph of gamma pdf

The Erlang distribution is also a particular case of the gamma distribution where the shape parameter α in (1.2), is a positive integer value, denoted here as n. Its pdf is:

$$f(t) \quad = \quad \frac{\lambda^n}{\Gamma(n)}(t-\mu)^{n-1}e^{-\lambda(t-\mu)} \in \mathbb{R}^+, \quad \text{where } \lambda > 0, \text{ and } n \in \mathbb{N}. \tag{1.5}$$

The Erlang distribution is often used to model the waiting times in queuing systems, particularly in the case of telephone traffic engineering. The Erlang distribution is the probability distribution of the waiting time until the n^{th} arrival in a one-dimensional Poisson process with intensity λ. The case $n = 1$ reduces to the exponential distribution. The

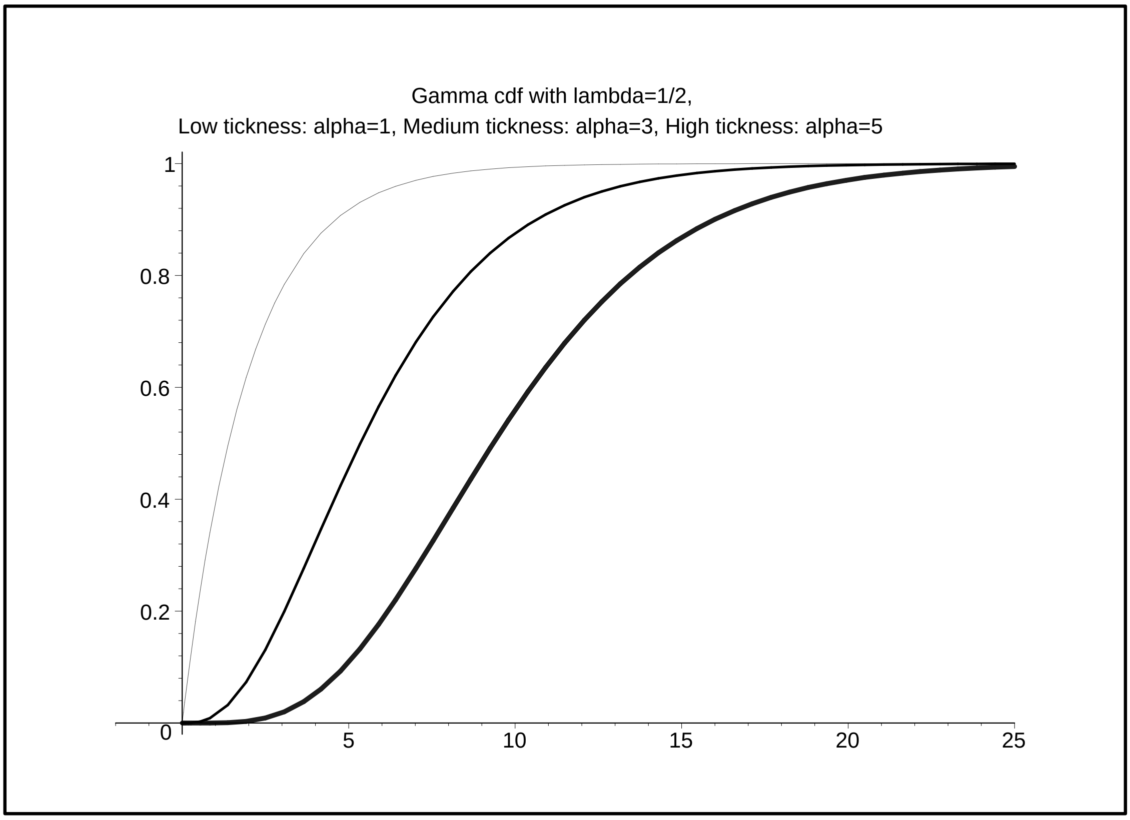

Figure 1.2: Graph of gamma cdf

exponential distribution is the difference between successive occurrences of events given by a Poisson process.

Another interpretation due to Barlow and Proschan (1965) [5], explains the Erlang distribution as the waiting time until failure by a random shock in a device. The device fails when there are exactly α shocks that occur randomly over time, $\alpha \in \mathbb{N}$, at a Poisson rate with parameter λ. That is the convolution of α exponential distributions.

1.3.1 Properties

The class of gamma distributions which includes the exponential, the Erlang and the Chi-Square distributions, is widely used in applied and theoretical statistics. Unlike the normal distribution, the gamma distribution is asymmetric, and cannot be distinguished by its mean and variance alone, known as first and second moments, respectively. In addition to the mean and variance, this distribution is characterized by its third and fourth order moments known as skewness and kurtosis, and by higher moments. See Grove and Coddington (2005) [31] for example.

The expected value of the gamma distribution in (1.2) and its variance are

$$EX = \frac{\alpha}{\lambda} + \mu \quad \text{and} \quad Var(X) = \frac{\alpha}{\lambda^2}, \tag{1.6}$$

respectively.

The probability that an event of interest has not occurred by time x, the survival or reliability function, is given by

$$S_X(x) = 1 - F_X(x) = 1 - I\Big(\alpha; \lambda(x-\mu)\Big), \tag{1.7}$$

where $I(\alpha, t) = \frac{1}{\Gamma(\alpha)} \int_0^t x^{\alpha-1} e^{-x} dx$.

The instantaneous rate of occurrence of an event, also called hazard function, or failure rate function, is obtained from (1.7) and is defined as:

$$h(x) = \frac{f(x)}{S(x)} = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{(x-\mu)^{\alpha-1}}{(1 - I(\alpha; \lambda(x-\mu)))} e^{-\lambda(x-\mu)}, \tag{1.8}$$

and the cumulative hazard rate is

$$H(x) = \int_{\mu}^{x} h(u)du = -\log S(x). \tag{1.9}$$

The hazard function (1.8), for the gamma density in (1.2), is increasing when $\alpha > 1$ from 0 to λ, decreasing when $\alpha < 1$ from ∞ to λ, and constant when $\alpha = 1$ (exponential case).

Many authors have recognized the importance of modeling and studying the hazard function (1.8) or the cumulative hazard rate (1.9). For example, Cox (1972) [20] used the hazard function to get estimators of the effects of different covariates on the times to failures of a system, by his famously proposed Cox proportional hazard model.

Balakrishnan and Wang (2000) [4] showed that the estimation problem for the three parameter gamma has always been a challenging and interesting one, in particular when the shape parameter α is small (less than 2.5). As the shape $\alpha \to \infty$, the gamma distribution in (1.2) approaches normality.

Another property related with the gamma pdf is that it is closed under scalar product. More precisely, if $X \sim Ga(\mu, \lambda, \alpha)$, then

$$Y = cX \sim Ga(c\mu, \frac{\lambda}{c}, \alpha) \tag{1.10}$$

for all $c > 0$,

and $$f_Y(y) = f_X(\frac{y}{c}) * \frac{1}{c} = \frac{(\lambda/c)^{\alpha}}{\Gamma(\alpha)}(y - c\mu)^{\alpha-1}e^{-\frac{\lambda}{c}(y-c\mu)}.$$

The Laplace transform of (1.2) is given by

$$L_X(t) = e^{-\mu t}\left(\frac{\lambda}{\lambda + t}\right)^{\alpha}. \tag{1.11}$$

Some advantages of the gamma distribution are in the simple forms of its density (1.2) and its Laplace transform (1.11) further discussed in Example 2.2.

Another important property is that in the Erlang type distribution, $G(\mu, \lambda, m)$ can be thought of as the sum of m independent distributions that are each displaced exponential as in (1.4) with parameters μ_i and λ such that $\sum_{i=1}^{m} \mu_i = \mu$.

More precisely, if $X_i \sim \exp(\mu_i, \lambda)$ $i = 1, 2, \cdots, m$, then $X = \sum_{i=1}^{m} X_i \sim G(\sum_{i=1}^{m} \mu_i, \lambda, m)$. Dudewicz and Mishra (1988) [23], and Billingsley (1986) [6] showed that a sum of independent gamma r.v.'s with same location and scale parameters μ and λ, is again a gamma r.v. with the added shape parameters. The class of distributions $\{Ga(\mu, \lambda, m);\ m \in \mathbb{N}\}$ is then additive as its sum operation is well defined. This is analogous to the informally defined multiplicative distribution for distributions that have a well defined multiplicative operation.

1.3.2 Estimation of Parameters

Since the gamma distribution has many applications and is widely used, obtaining reliable estimates of its parameters is a key issue, especially since the gamma can model a variety of hazard functions. There are several methods of estimation such as graphical methods, methods of moments, the least squares method, and the maximum likelihood method.

The estimations of the shape and scale are affected if the location has to be estimated using method of moments. Refer to Bowman and Shenton (2002) [9] for an example. The location parameter μ, in most cases non-negative, is typically called the threshold parameter. Here, it is interpreted as the minimum possible lifetime and can be estimated by the minimum order statistics, as the condition $x \geq \mu$ must be always satisfied, although it is only an MLE in the exponential case. Note that Balakrishnan and Wang (2000) [4] also considered a general linear combination of the order statistics,

$$\hat{\mu} = \textstyle\sum_{i=1}^{n} c_i x_{(i)},$$

where $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ are the order statistics from a given sample of size n.

However, as mentioned for example in Bowman and Shenton (1988) [8], there is a high degree of deviation of the estimates from the parent distribution if one uses the method of moments. In exploratory data analysis, the sample mean and the sample standard deviation are two quantities often computed as measures of center and spread, respectively. In the normal distribution case, the population mean and variance are mathematically independent, and the sample mean and variance are known to be statistically independent. However, in the non normal case, the population mean and variance are often linked mathematically (See McCullagh and Nelder (1989) [66]), i.e. the higher the mean, the higher the standard deviation, and in fact the sample mean and sample variance are not independent.

The coefficient of variation establishes a similar independence property for the gamma case. The population coefficient of variation, CV, for a distribution X with mean μ_X and standard deviation σ_X is defined as $CV = \dfrac{\sigma_X}{\mu_X}$. Estimates are obtained for example by Hwang and Huang (2002) [41], among other authors. They derive the results, establishing that the sample mean $\bar{X}$ and the sample variance S_n^2 are unbiased estimators of (1.6), respectively. (See Hwang and Hu (1999)[39] and Hwang and Hu (2000)[40]). We recall from Dudewicz and Mishra (1986) [23] and Casella and Berger (1990) [13], that the independence of the sample mean and sample standard deviation characterizes the normal distribution. Theorem 1.2 establishes a similar characterization for the gamma distribution through the sample coefficient of variation $CV_n = \dfrac{S_n}{\bar{X}_n}$.

Theorem 1.2 *Suppose that $n \geq 3$. $X_1, X_2, \ldots, X_n$ are independent $Ga(\mu, \lambda, \alpha)$ r.v.'s iff $\bar{X}_n$ and CV_n are independent.*

Proof: Proved in Hwang and Hu (1999) [39]. □

Hwang and Hu (1999) [39] use this characterization to obtain estimates of shape and scale parameters for the gamma distribution. They showed that these new estimators perform better that the MLE, and the method of moments estimators for a sample of size $n \leq 25$.

Using (1.2), the likelihood function based on an independent identically gamma distributed random sample $X_1, X_2, \cdots, X_n$ of size n is called , is given by:

$$L(\mu, \lambda, \alpha) = \frac{\lambda^{n\alpha}}{\Gamma^n(\alpha)} \prod_{i=1}^{n} (x_i - \mu)^{\alpha-1} e^{-\lambda \sum_{i=1}^{n} (x_i - \mu)} I_{(\mu,\infty)}(x_i).$$

The likelihood is viewed as a function of the parameters μ, λ, and α conditioned on the data. To ease the computations, one often deals with the logarithmic of the likelihood, and the log likelihood is given as

$$l(\mu, \lambda, \alpha) = n\alpha \log \lambda - n \log \mathit{\Gamma}(\alpha) + (\alpha - 1) \sum_{i=1}^{n} \log(x_i - \mu) - \lambda \sum_{i=1}^{n} (x_i - \mu).$$

Based on the strong law of large numbers or ergodic theorems, the maximum likelihood estimators have useful properties such as consistency and sufficiency. They achieve the Cramér-Rao minimum variance asymptotically. They have the smallest variance among all possible estimators asymptotically as the sample size increases. In other words, MLE's are statistically efficient and given better and often more reliable estimates than the method of moments.

1.4 Literature Review

Several approaches and models have been suggested and developed for constructing the multivariate gamma distribution by many authors such as Johnson and Kotz (1970) [45], Hougaard (2001) [37], Marshall and Olkin (1967) [61], Tanner (1996) [77], Moran (1969) [69], Mathai and Moschopoulos (1992) [64], and Walker and Stephens (1999) [79].

Defining the multivariate gamma distribution can be quite intuitive, but properties are difficult to illustrate. That could explain why numerous versions of multivariate gamma distributions are proposed in the literature. In particular, various forms of bivariate models have been described from combinations of independent gamma distributions. Johnson et al. (1997) [53] gives a coverage of a useful variety. One approach is to start with certain desirable statistical properties in the univariate case, and build the multivariate extension through various mechanisms. We take one such approach in this dissertation. See for example Mathai and Moschopoulos (1991) [63], Gaver and Lewis (1980) [29], and Hanagal (1996) [32]. Another approach, which defines the multivariate gamma from the bivariate, is not as straightforward, since many bivariate gamma do not have a natural extension to multivariate dependence. For example, see Marshall and Olkin (1967) [61], Marshall and Olkin (1983) [62], and Sivazlian (1981) [76]. In fact, Marshall and Olkin (1967) [61], and Marshall and Olkin (1983) [62] proposed an approach that does not extend to the multivariate case easily. Sivazlian (1981) [76] gives a form of multivariate gamma with independent structure.

Walker and Stephens (1999) [79] define a multivariate family of distributions on $\mathbb{R}_+^p$ in the context of survival analysis. It includes the Weibull, with interesting properties for which the marginal distribution is of the same family, with a flexible relationship between the components, straightforward analysis, and is applicable for censored data. As they stated, the extension to the multivariate setting has proved quite problematic. Our goal is to solve that problem and maintain these properties as well in our construction of the multivariate gamma distribution.

Another classical multivariate gamma is known as Cheriyan and Ramabhadran's gamma distribution and is given in Kotz et al. (2000) [53]. It is a multivariate additive distribution which is a simple case of the one we define in (1.1) where the nonnegative constant is taken to be unity. Results for this distribution are also noted in Mathai and Moschopoulos (1991) [63], which they further extend.

Henderson and Shimakura (2003) [34] presented a version of multivariate gamma distribution with gamma marginal distributions and correlation matrix without a closed form for the density. Moreover, the estimation of parameters and asymptotic properties are not well studied in those cases.

One needs to develop a multivariate gamma distribution that can be used in modeling processes, such as those defined by Marshall and Olkin (1983) [62] for the bivariate logistic model, or the bivariate exponential model by Mardia (1970) [60]. These methods and others are given in Coles and Tawn (1994) [19], where it is shown how statistical methods in the multivariate context may be applied to problems of data analysis in the extremes in particular.

As we said earlier, for the multivariate normal $\mathbf{X} = (X_1, X_2, \cdots, X_p)'$, with $p \times p$ positive definite covariance matrix Σ having $Var(X_i) = \sigma_i^2$ $(1 \le i \le p)$ as diagonal elements of Σ, Chen (1987) [15], among others, obtained a characterization of $\mathbf{X}$ as

$$U(\mathbf{X}, \Sigma) = 1, \quad \text{where}$$

$$U(\mathbf{X}, \Sigma) = sup_{g \in C} \frac{Var[g(X)]}{E[\nabla^t g(X) \Sigma \nabla g(X)]}.$$

Here C is the class of measurable functions $g : \mathbb{R}^p \mapsto \mathbb{R}$ such that $E[g(\mathbf{X})]^2 < \infty$ and $Var[g(\mathbf{X})] > 0$ and $\nabla g(\mathbf{X})$ represents the gradient operator.

Then for the multivariate gamma, $U(\mathbf{X}, \Sigma) > 1$. While in the Gaussian case, where the correlation matrix and the marginal distributions completely specify the joint distribution, those two components (correlation matrix and marginal distributions) do not induce a unique joint distribution for correlated gamma r.v.'s.

We cannot arbitrarily set $\mathbf{X}$ such that the marginal distributions are gamma, i.e $X_i \sim Ga(0, \lambda_i, \alpha_i)$ for $i = 1, \cdots, n$, and $Corr(X_i, X_j) = \rho_{ij} \in [-1, 1] \quad \text{for} \quad i, j \in \{1, 2, \cdots, n\}$. The difficulty arises from the fact that the mean and variance/covariance cannot be modeled arbitrary as in the normal case, because in the gamma case, the $\{\rho_{ij}\}_{1 \le i,j \le n}$ are functions of the shape parameters. This is true for all multivariate gamma motivated in the literature to date.

Using the spirit of the multivariate normal density, several versions of the multivariate gamma can be found. Fang and Zhang (1990) [24] describe a family of multivariate distributions called p-dimensional r.v. $\mathbf{X} = (X_1, X_2, \cdots, X_p)'$ with symmetric Kotz type distribution (KTD) denoted as $\mathbf{X} \sim KTD_p(\mu, \Sigma_{p\times p}, \alpha, m, \beta)$ where

- μ represents the location vector in $\mathbb{R}^p$,
- $\Sigma_{p\times p}$ is the positive definite covariance matrix,
- α is such that $\alpha > (2-p)/2$,
- $m > 0$ and $\beta > 0$.

It is defined as having a joint density of the form

$$f_{\mathbf{X}}(\mathbf{x}) = C_p|\Sigma|^{-1}[(\mathbf{x}-\mu)^{\mathbf{t}}\boldsymbol{\Sigma}^{\mathbf{-1}}(\mathbf{x}-\mu)]^{\alpha-\mathbf{1}} \exp\{-\mathbf{m}[(\mathbf{x}-\mu)^{\mathbf{t}}\boldsymbol{\Sigma}^{\mathbf{-1}}(\mathbf{x}-\mu)]^{\beta}\},$$

where C_p is the normalizing constant given by,

$$C_p = \frac{\beta\Gamma(p/2)m^{p/2\beta+(\alpha-1)/\beta}}{\pi^{p/2}\Gamma(p/2\beta+(\alpha-1)/p)}.$$

When $\beta = 1, \alpha = 1$, and $m = \frac{1}{2}$, we obtain the multivariate normal distribution.

Using Fang and Zhang (1990) [24] approach, and modifying the multivariate normal, we hope to have properties similar to the multivariate normal. We define the square root of a vector as the square root of its components. We introduce the type of multivariate gamma as follows:

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= C_p|\Sigma|^{-1}[(m[\mathbf{x}-\mu])^{(\mathbf{1/2})}]^{\mathbf{t}}\boldsymbol{\Sigma}^{\mathbf{-1}}(\mathbf{m}[\mathbf{x}-\mu])^{\mathbf{1/2}}]^{\alpha-\mathbf{1}} \\ &\quad \times \exp\{-[(m[\mathbf{x}-\mu])^{(\mathbf{1/2})\mathbf{t}}\boldsymbol{\Sigma}^{\mathbf{-1}}(\mathbf{m}[\mathbf{x}-\mu])^{\mathbf{1/2}}]^{\beta}\}. \end{aligned}$$

Taking $\beta = 1, m = (\sqrt{\alpha_1}, \cdots, \sqrt{\alpha_p})^t$, we have the form of the multivariate gamma distribution. When the X_i's are independent, we have the desired form. The result is not obtained when there is dependence. This multivariate gamma is an example of the multivariate dispersion model introduced by Jorgensen (1987) [49].

1.5 Marginal, Conditional and Joint Distributions

Here, we review the relevant concepts of marginal, joint and conditional distributions that are referred to frequently in this dissertation. These concepts can be found in textbooks such as Feller (1971) [25], Dudewicz and Mishra (1988) [23], Folland (1999) [26], and Kallenberg (2002) [51]. We also define the concept of conditional independence.

Definition 1.3 *Let $(\Omega, \mathcal{A}, \mathcal{P})$ be a probability space, i.e. Ω is a set, $\mathcal{A}$ is a σ-algebra, a family of subsets of Ω and P is a mapping, $P : \mathcal{A} \to [\mathbf{0}, \mathbf{1}]$ such that (i) $P(\Omega) = 1$ and (ii) if $\{A_n, n \in \mathbb{N}\}$ disjoints sets in $\mathcal{A}$, then $P(\bigcup_n A_n) = \sum_{n=1}^{\infty} P(A_n)$.*

Definition 1.4 *Let $\mathcal{B}(\mathbb{R})$ be the Borel $\sigma-$field generated on $\mathbb{R}$. A random variable (r.v.) X is a measurable function on a probability space, i.e. X is a mapping $X : \Omega \to \mathbb{R}$ such that for every $B \in \mathcal{B}(\mathbb{R})$, $X^{-1}B \in \mathcal{A}$.*

Definition 1.5 *The cumulative distribution function (cdf) of X is a function F_X, also denoted as F if there is no confusion, $F : \mathbb{R} \to [0, 1]$ defined as*

$$F(x) = (P \circ X^{-1})((-\infty, x]) = P(X \leq x), \quad x \in \mathbb{R}.$$

The cdf is a nondecreasing, right continuous, and has the following limits:
$\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

For a fixed integer $p \in \mathbb{N}$, the joint cumulative distribution of $p-$r.v.'s, or $p-$variate denoted $\mathbf{X} = (X_1, \cdots, X_p)'$ is an extension of the above concept, and is defined on a probability space $(\Omega, \mathcal{A}, P)$. When $\Omega = \mathbb{R}^p$, the joint cdf of $\mathbf{X}$ is defined as $F : \mathbb{R}^p \to [0, 1]$ by:

$$F(x_1, \cdots x_p) = P(X_1 \leq x_1, \cdots, X_p \leq x_p). \tag{1.12}$$

The joint cdf is right continuous, and satisfies the following:

$\lim_{x_i \to \infty} F(x_1, \cdots, x_i, \cdots, x_p) = F(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_p)$,

which is the joint distribution of the remaining $p-1$-variates and

$\lim_{x_1, \cdots, x_p \to -\infty} F(x_1, \cdots, x_p) = 0$ and $\lim_{x_1, \cdots, x_p \to \infty} F(x_1, \cdots, x_p) = 1$.

The marginal cdf of X_k, with $1 \leq k \leq p$, is defined as

$F(x_k) = \lim_{x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_p \to \infty} F(x_1, \cdots, x_p)$.

The $p-$r.v's. are said to be independent if (1.12) can be written as

$F(x_1, \cdots, x_p) = F_{X_1}(x_1) \cdots F_{X_p}(x_p)$,

and then $E(X_1 \cdots X_p) = E(X_1) \cdots E(X_p)$ provided that all the expectations exist. In addition, the r.v.'s $\{g_i(X_i), 1 \leq i \leq p\}$ are independent. It is also known that uncorrelated r.v.'s can be deduced from independence, but uncorrelated does not always imply independence.

Another concept that is useful is that of conditional distributions.

Definition 1.6 *Suppose $(X_1, X_2, \cdots, X_p)'$ is a random vector. The conditional distribution of $(X_1, X_2, \cdots, X_k)'$, with $1 \leq k \leq p$, given $(X_{k+1}, \cdots, X_p)'$, is given by the ratio of the joint distribution of all the r.v.'s over the joint marginal distribution of the conditioned* *r.v.'s. It is given by:*

$$f_{X_1,X_2,\cdots,X_k|X_{k+1},\cdots,X_p}(x_1,x_2,\cdots,x_k|x_{k+1},\cdots,x_p) = \frac{f_{X_1,X_2,\cdots,X_p}(x_1,x_2,\cdots,x_p)}{f_{X_{k+1},\cdots,X_p}(x_{k+1},\cdots,x_p)},$$

provided that $f_{X_{k+1},\cdots,X_p}(x_{k+1},\cdots,x_p) > 0$, *where* f *could be a pdf or a pmf.*

Definition 1.7 *The* $p-$*variate vector* $\mathbf{X} = (X_1, X_2, \cdots, X_p)'$ *is said to be conditionally independent, conditioned on a variable, say* X_0, *if*

$$f(x_1, x_2, \cdots, x_p|x_0) = f_1(x_1|x_0)f_2(x_2|x_0)\cdots f_p(x_p|x_0). \tag{1.13}$$

where $f_i(\cdot|x_0)$ *represents the conditional pdf of* X_i, $i = 1, \cdots, p$ *given* X_0.

(1.13) is equivalent to $F(x_1, x_2, \cdots, x_p|x_0) = F_1(x_1|x_0)F_2(x_2|x_0)\cdots F_p(x_p|x_0)$ where $F_i(\cdot|x_0)$ represents the conditional cdf of X_i, for $i = 1, \cdots, p$ given X_0.

Theorem 1.8 *The joint density of* $(X_0, X_1, \ldots, X_p)$ *satisfying (1.1) is expressed as:*

$$f(x_0, x_1, x_2, \cdots, x_p) = f(x_0)\prod_{i=1}^{p} f_{Z_i}(x_i - a_i x_0).$$

and $(X_1, X_2, \cdots, X_p)'$ *are conditionally independent given the latent variable* X_0.

Proof: Using the independence of Z_i's, $i = 1, .., p$, of each other and of X_0, we have that the joint density of $X_0, Z_1, Z_2, \cdots, Z_p$, given as:

$$f(x_0, z_1, z_2, \cdots, z_p) = f(x_0)f(z_1)f(z_2)\cdots f(z_p).$$

Hence based on the linear transformation $X_i = a_i X_0 + Z_i \Longleftrightarrow Z_i = X_i - a_i X_0$, for $i = 1, 2, .., p$, we have that:

$$f(x_0, x_1, x_2, \cdots, x_p) \quad = \quad f(x_0)f(x_1 - a_1 x_0)\cdots f(x_p - a_p x_0)$$

$$= f(x_0)\prod_{i=1}^{p} f_{Z_i}(x_i - a_i x_0).$$

From Definition 1.7, $(X_1, X_2, \cdots, X_p)'$ given X_0 are independent as the conditional distribution of $(X_1, X_2, \cdots, X_p)'$ given X_0 is:

$$\begin{aligned}
f(x_1, x_2, \cdots, x_p|x_0) &= \frac{f(x_0, x_1, x_2, \cdots, x_p)}{f(x_0)} \\
&= f_{Z_1}(x_1 - a_1x_0)f_{Z_2}(x_2 - a_2x_0)\cdots f_{Z_p}(x_p - a_px_0) \\
&= \prod_{i=1}^{p} f_{Z_i}(x_i - a_ix_0) \\
&= \prod_{i=1}^{p} f_{X_i}(x_i|x_0).
\end{aligned}$$

$\square$

Note that, given x_0, a_ix_0 plays the role of a location parameter. So $x_i \geq a_ix_0$ for all $i = 1, 2, \cdots, p$. In Chapter 3, we see that when x_0 is unknown, a possible estimate for it is the minimum of the x_i/a_i's, $\quad i = 1, \cdots, p$.

Definition 1.9 : *The random vector* $\mathbf{X} = (X_1, X_2, \cdots, X_p)'$ *is multivariate gamma if its components* X_i, $\quad i = 1, \cdots, p$, *marginally are of gamma family following the structure described in (1.1), and each subset of* $\mathbf{X}$ *has a multivariate gamma form.*

1.6 Thesis Outline

In this dissertation, we are concerned with the three parameter gamma distribution. The characterization of the joint gamma distribution of $\mathbf{X} = (X_1, X_2, \cdots, X_p)$ based on (1.1):

- has Mathai and Moschopoulos (1992) [64] as a special case and a more flexible linear structure,
- has Iyer et al. (2004) [43] as a special case both in terms of structure and dimension,
- has Nadarajah and Kotz (2005) [72] as a special case,
- and shares similar properties to the Marshall and Olkin (1967) [61] model, that includes the continuous and discontinuous portions. Marshall and Olkin's expression for the bivariate exponential model has received a lot of attention due to its allowance for the simultaneous failure of several system components induced from a fatal shock.

We provide different inference procedures for the parameters, and characterize location-scale families of multivariate gamma distributions. The marginal distributions are traditional location-scale gammas, and their joint distribution contains absolutely continuous classes, as well as, the Marshall-Olkin (1967) [61] type of distributions with a positive probability mass on a set of measure zero. MLE's are developed in the bivariate case.

The dissertation is organized as follows. In Chapter 2, we start with preliminary materials and fundamental concepts used in these new classes of survival and reliability models. In Chapter 3, we derive and characterize the generalized multivariate gamma distribution. The continuous case of the multivariate gamma is studied in Chapter 4. In Chapter 5, a discontinuous case of the multivariate gamma, in the form of multivariate exponential, is given.

Chapter 2

Preliminaries

The necessary mathematical concepts for this dissertation are concentrated in this chapter. We review the notions of Laplace transforms (See Feller (1971) [25], Abramowitz and Stegun (1972) [1], and Billingsley (1986) [6]) and Dirac delta (See Cohen-Tannoudji et al. (1977), and Khuri (2004) [52]), infinite and stable distributions (See Feller (1971) [25], and Hougaard (2000) [36]).

2.1 The Laplace Transform

The Laplace transform (the equivalent concept of moment generating function) is used in many areas of statistics, probability theory, and risk analysis. We are interested in the Laplace transform associated with real arguments. The characterization by the means of the real Laplace transform does not presume the existence of moments

Definition 2.1 *If X is a r.v. defined on $\mathbb{R}_+$ with cdf F_X, satisfying $P(X = 0) < 1$, then its Laplace-Stieltjes transform (LST) is the function valued in $\mathbb{R}$ defined in Abramowitz and Stegun (1972) [1] as:*

$$L_X(s) = Ee^{-sX} = \int_0^\infty e^{-sx}\, dF_X(x). \tag{2.1}$$

Here are some properties associated with the LST:

- existence: the integral in (2.1) is with respect to Lebesgue-Stieltjes integration. In our cases of positive support distributions, (2.1) always exists. In fact, $0 < L_X(s) \leq 1$.
- $L_X(s)$ is infinitely differentiable, and $\dfrac{d^n L_X}{ds^n}(s)$ exists for all $n \in \mathbf{N}$.

- For $m \in \mathbb{N}$, the m^{th} moment of X is given by $EX^m = (-1)^m L_X^{(m)}(0)$.

- additivity: the LST of the sum of independent r.v.'s is obtained by taking the product of the LST of the individual r.v. For $X_1, \ldots, X_n$ independent r.v.'s, then $X = \sum_{i=1}^n X_i$ has LST:
$$L_X(s) = Ee^{-sX} = E\prod_{i=1}^{n} e^{sX_i} = \prod_{i=1}^{n} Ee^{sX_i} = \prod_{i=1}^{n} L_{X_i}(s).$$

- the LST can be defined for a $p-$variate distribution, say $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$ as:

  $L_{\mathbf{X}}(\mathbf{s}) = Ee^{-(s_1 X_1 + \cdots + s_p X_p)}$, for $\mathbf{s} = (s_1, s_2, \cdots, s_p)' \in \mathbb{R}_+^p$.

- uniqueness: if X_1 and X_2 are two r.v.'s such that $L_{X_1}(s) = L_{X_2}(s)$ then $f_{X_1}(x) = f_{X_2}(x)$, for all x except on a set of measure 0. Therefore, the LST completely characterizes the distribution.

The LST helps in the computations and in the linear combinations of r.v.'s associated with some distributions.

Example 2.2 *If $X \sim Ga(\mu, \lambda, \alpha)$ with pdf (1.2), then its LST is given by (1.11), as:*

$$\begin{aligned}
L_X(s) &= \int_\mu^\infty e^{-sx}\frac{\lambda^\alpha}{\Gamma(\alpha)}(x-\mu)^{\alpha-1}e^{-\lambda(x-\mu)}dx \\
&= e^{-s\mu}\int_\mu^\infty e^{-s(x-\mu)}\frac{\lambda^\alpha}{\Gamma(\alpha)}(x-\mu)^{\alpha-1}e^{-\lambda(x-\mu)}dx \\
&= e^{-s\mu}\int_\mu^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)}(x-\mu)^{\alpha-1}e^{-(\lambda+s)(x-\mu)}dx \\
&= e^{-s\mu}\int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)}t^{\alpha-1}e^{-(\lambda+s)t}dt \\
&= e^{-s\mu}\frac{\lambda^\alpha}{(\lambda+s)^\alpha}\int_0^\infty \frac{(\lambda+s)^\alpha}{\Gamma(\alpha)}t^{\alpha-1}e^{-(\lambda+s)t}dt \\
&= e^{-s\mu}\left(\frac{\lambda}{\lambda+s}\right)^\alpha.
\end{aligned}$$

The gamma distribution, shifted at the origin, with unit mean has LST

$$L_X(t) = e^{-\mu t}\left(\frac{1}{1+t/\alpha}\right)^{\alpha} \longrightarrow e^{-\mu t}e^{-t} = e^{-(\mu+1)t} \quad \text{as } \alpha \to \infty.$$

Example 2.3 *: Let $X_1, X_2, \cdots, X_n$, $n > 1$, be independent r.v.'s distributed as $Ga(\mu_i, \lambda, \alpha_i)$ for $1 \le i \le n$. It is well known that $X = X_1 + X_2 + \cdots X_n$ has density $Ga(\sum_{i=1}^{n} \mu_i, \lambda, \sum_{i=1}^{n} \alpha_i)$. This is easily shown using LST. The Laplace transform of the gamma distribution X_i, $i = 1, \cdots, n$, given in Example 2.2 becomes:*

$$L_{X_i}(s) = e^{-s\mu_i}\left(\frac{\lambda}{\lambda+s}\right)^{\alpha_i}.$$

Using the properties of the LST, we have

$$L_X(s) = \prod_{i=1}^{n} L_{X_i}(s) = \prod_{i=1}^{n} e^{-s\mu_i}\left(\frac{\lambda}{\lambda+s}\right)^{\alpha_i} = e^{-s\sum_{i=1}^{n}\mu_i}\left(\frac{\lambda}{\lambda+s}\right)^{\sum_{i=1}^{n}\alpha_i}$$

$$\Rightarrow X \sim Ga(\sum_{i=1}^{n} \mu_i, \lambda, \sum_{i=1}^{n} \alpha_i).$$

In the remaining cases, we consider situations where the λ's are different. When two or more of the gamma distributions have same parameter λ, we can add the r.v.'s to obtain another gamma distribution with the same parameter λ, but a different shape parameter.

Example 2.4 *Suppose X and Y are independent discrete and positive support continuous distributions with pmf and pdf $p(x)$ and $f(y)$, respectively. Then*

$$\begin{aligned} L_{XY}(s) &= Ee^{-sXY} = \int_0^\infty \sum_x e^{-sxy} f(y)p(x)dy \\ &= \sum_x \left(\int_0^\infty e^{-sxy} f(y)dy\right) p(x) \\ &= \sum_x L_Y(sx)p(x). \end{aligned}$$

We later use Example 2.4 in the case when X is a Bernoulli random variable with probability p, and Y is a positive support distribution. In that case, we have that:

$$L_{XY}(s) = p + (1 - p)L_Y(s).$$

We are also interested in the sum of r.v.'s. For two continuous r.v.'s $X_1 \sim f_1$ and $X_2 \sim f_2$, their sum, $X = X_1 + X_2$, has density obtained from the joint pdf as:

$$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{x-x_1} f(x_1, x_2)dx_1dx_2.$$

The convolution of two functions $f, g : \mathbb{R} \to \mathbb{R}$ is the function

$$(f \star g)(x) = \int_{-\infty}^{\infty} f(t)g(x - t)dt.$$

If the r.v.'s are independent, the density of their sum is the convolution of their densities, and can be represented as below.

Theorem 2.5 *Assume that X_1 and X_2 are independent r.v.'s defined on $\mathbb{R}_+$ with pmf/pdf f_1 and f_2, respectively. Then $X = X_1 + X_2$ has density*

$$f_X(x) = (f_1 \star f_2)(x) = \int_0^x f_1(t)f_2(x - t)dt.$$

Proof: Proved in Hunter and Nachtergaele (2001) [38] □

However, a lot of work can be alleviated as the LST of X in Theorem 2.5 is given by: $L_X(t) = L_{X_1}(t)L_{X_2}(t)$, and is recognized in some distributional form.

Example 2.6 *Consider X to be the sum of two iid of exponential type r.v.'s with parameter λ. Then $f_X(x) = \int_0^x f_1(t)f_2(x - t)dt = \int_0^x \lambda e^{-\lambda t}\lambda e^{-\lambda(x-t)}dt = \lambda^2 xe^{-\lambda x}$, which is a $Ga(0, \lambda, 2)$.*

Theorem 2.7 *For $X_1 \sim Ga(\mu_1, \lambda_1, \alpha_1)$ and $X_2 \sim Ga(\mu_2, \lambda_2, \alpha_2)$, the r.v. $X = X_1 + X_2$ has a gamma distribution $Ga(\mu_1 + \mu_2, \lambda, \alpha_1 + \alpha_2)$ iff $\lambda_1 = \lambda_2 = \lambda$.*

Proof: The LST of X is given by $L_X(s) = e^{-(\mu_1+\mu_2)s} \dfrac{\lambda_1^{\alpha_1} \lambda_2^{\alpha_2}}{(\lambda_1 + s)^{\alpha_1} (\lambda_2 + s)^{\alpha_2}}$

If $\lambda_1 \neq \lambda_2$, then the LST is not representative of a gamma distribution.

Considering the case where $\lambda_1 = \lambda_2 = \lambda$, then the LST and the density of the sum $X = X_1 + X_2$ become:

$$
\begin{aligned}
L_X(s) &= e^{-s\mu} \left(\frac{\lambda}{\lambda + s} \right)^{\alpha} \\
f_X(x) &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} (x - \mu)^{\alpha - 1} e^{-\lambda(x - \mu)}
\end{aligned}
$$

where $\alpha = \alpha_1 + \alpha_2$ and $\mu = \mu_1 + \mu_2$. $\square$

We generalize this above result later in Theorem 3.3. The sum of independent gamma distributions with same scale parameter gives a gamma distribution with same scale parameter with the location and scale updated. In that sense, the gamma distribution follows an additive property. Mathai and Moschopoulos (1992) [64] use this property to generate their class of multivariate gamma distributions.

2.2 The Dirac Delta Function

Another important collection of tools that we use are the properties of the Dirac distribution, sometimes referred to as the Dirac delta function. See Cohen-Tannoudji et al. (1977) [18], and Khuri (2004) [52].

Definition 2.8 *Dirac delta function at the point $c \in \mathbb{R}$, is a point mass distribution denoted δ_c, and we say that a r.v. X has point mass δ_c distribution at c if its pmf is given by*

$$
f(x|c) = \delta_c(x) = \delta(x - c) = 0 \quad \text{if} \quad x \neq c, \quad \text{and} \quad \int_{-\infty}^{\infty} f(x|c)dx = 1. \tag{2.2}
$$

The integral notation in (2.2) is not mathematically justified. δ is not rigourously defined as a function, but as a distribution, and there have been several references on this function. Cohen-Tannoudji et al. (1977) [18] gives a complete discussion.

Despite its name, the Dirac delta function is not a function in the classical sense. One reason for this is that because the function $f(x) = \delta(x)$, and $g(x) = 0 \quad a.e.$, are equal almost everywhere, yet their (Lebesgue) integrals are different. See Folland (1999) [26] and Rudin (1976) [74]. Another reason is that it is singular. Instead, it is said to be a distribution. It is a generalized idea of functions, and can be used inside integrals. The well known mathematician Laurent Schwartz gave it in 1947 a rigorous mathematical definition as a linear functional on the space of test functions $\mathbf{D}$, the set of all real valued infinitely differentiable functions with compact support on $\mathbb{R}$ such that for a given $f(x) \in \mathbf{D}$, the value of the functional is given by Kallenberg (1986) [50], and Hunter and Nachtergaele (2001) [38]. Such linear functionals are called generalized functions or distributions. For this reason, the delta function is more appropriately called Dirac's delta distribution. Thus, the value of the Dirac delta function δ_x is defined by its action of a function $f(x) \in \mathbf{D}$ when used in integral as in formula (2) in Khuri (2004) [52].

The theory of distributions in mathematics has been highly developed, and as a result, the Dirac delta function is well established and accepted in mathematics as a generalized function or distribution. Note also that it has been modified from the original version defined by Dirac in 1920.

It is well known that the Heaviside step function is an antiderivative of the Dirac distribution. The Heaviside step function, also called unit step function, see for example

Abramowitz and Stegun (1972) [1], is a discontinuous function defined as

$$H(x) = \int_{-\infty}^{x} \delta(t)dt = \begin{cases} 0, & \text{if } x \leq 0; \\ 1, & \text{if } x > 0. \end{cases} \tag{2.3}$$

The value of the Heaviside function at 0 is sometimes taken to be 0, or $\frac{1}{2}$ (most popular for symmetry purposes) or 1. Here, we take it to be 0.

Both Dirac and Heaviside functions have been used in a variety of fields of science and engineering. Their use in statistics is relatively new. For another reference see the paper by Pazman and Pronzato (1996) [73].

The Dirac delta function is a very useful tool in approximating tall narrow spike functions (also called impulse functions), and the following integral

$$\int_{-\infty}^{-\infty} f(x)\delta(x)dx = \delta[f] = f(0)$$

for any (test) function $f(x)$, is more a notation for convenience, and not a true integral. It can be regarded as an "operator" or a linear functional on the space of (test) functions, which gives the value of the function at 0. It is important to see that the integral is simply a notational convenience, and not a true integral.

More details are given in Kallenberg (1986) [50], Williamson (1962) [80], Au and Tam (1999) [3] and Shilov and Gurevich (1977) [75] for examples. So as a distribution, the Dirac delta function $\delta(x-s)$ is a pdf with mean median and mode s, cdf $H(x-s)$, variance and skewness 0 satisfying the following:

- $\displaystyle\int_{-\infty}^{\infty} \delta(\alpha x)dx = \int_{-\infty}^{\infty} \delta(u)\frac{du}{|\alpha|} = \frac{1}{|\alpha|}$
- $\delta(\alpha x) = \dfrac{\delta(x)}{|\alpha|}$

- $\delta(x) = \lim_{a \to 0} \delta_a(x)$ where $\delta_a(x) = \frac{1}{a\sqrt{\pi}} e^{-x^2/a^2}$ as limit of a normal distribution.

To end this review, we note the following results:

$$c > 0, \quad H(cx) = H(x),$$

$$H(x-a) = 1 - H(-x+a) = 1 - H(a-x) \quad \text{and}$$

$$\int H(x-a)dx = (x-a)H(x-a).$$

The Dirac delta distribution can be thought as the limit case of a distribution whose density must be concentrated at the origin point. So for a r.v. X with Dirac density $\delta(x-c), c \geq 0$, the LST is given by $L_X(t) = e^{-ct}$.

The moments for the Dirac delta function δ_c are given by: $EX^k = c^k, \quad Var(X) = 0$, and its characteristic function is given by $\phi(t) = e^{itc}$.

The Dirac function provides a very helpful tool in mathematical statistics as it provides a unifying approach in the treatment of discrete and continuous distributions and their transformations. We give two examples in each case below.

Example 2.9 *Khuri (2004) [52]*

Suppose X is a discrete r.v. that takes values $a_1, a_2, \cdots, a_n$ with corresponding probabilities $p_1, p_2, \cdots, p_n$, respectively. Assume that $\sum_{i=1}^n p_i = 1$. Then the pdf $p(x)$ of X can be represented as $p(x) = \sum_{i=1}^n p_i \delta(x - a_i)$.

For example, if $X \sim B(n,p)$ where $B(n,p)$ is the binomial distribution, then $p(x) = \sum_{i=0}^n p^i (1-p)^{n-i} \delta(x-i)$ and the k^{th} noncentral moment of X is given by

$$\int_{-\infty}^{\infty} x^k p(x)dx = \sum_{i=1}^n p_i \int_{-\infty}^{\infty} x^k \delta(x-a_i)dx = \sum_{i=1}^n a_i^k p_i.$$

Example 2.10 *Khuri (2004) [52]*

Let X_1 and X_2 be independent r.v. distributed as $X_1 \sim Ga(\lambda = \dfrac{1}{2}, \alpha)$ and $X_2 \sim Ga(\lambda = \dfrac{1}{2}, \beta)$. The joint distribution of $Y = \frac{X_1}{X_1+X_2}$ and $Z = X_1 + X_2$ can expressed as:

$$\begin{aligned}
f(y,z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2)\delta[\frac{x_1}{x_1+x_2} - y]\delta(x_1+x_2-z)dx_1dx_2 \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)2^{\alpha+\beta}}\int_0^{\infty} x_2^{\beta-1}dx_2 \int_0^{\infty} x_1^{\alpha-1}e^{-\frac{1}{2}(x_1+x_2)} \\
&\quad \times\delta(\frac{x_1}{x_1+x-2} - y)\delta[x_1 - (z-x_2)]dx_1 \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)2^{\alpha+\beta}}\int_0^{\infty} x_2^{\beta-1}(z-x_2)^{\alpha-1}e^{-\frac{1}{2}z}\delta[\frac{z-x_2}{z} - y]dx_2 \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)2^{\alpha+\beta}} y^{\alpha-1}(1-y)^{\beta-1}z^{\alpha+\beta-1}e^{-\frac{1}{2}z}.
\end{aligned}$$

The cumulants are alternative ways of summarizing the properties of the r.v.'s specially when the moments are not easily obtained. The moments of X are not directly related to the moments of $aX+b$, for a and b positive constants. The moments of $X = X_1 + X_2$ do not have simple relation with the moments of X_1 and X_2. The idea is then to use the cumulants.

Definition 2.11 *The function $\kappa(s) = \log L_X(-s)$ is called the cumulant generating function of X.*

Expanding $\kappa(s)$ in its power series, $\kappa(s) = \sum_{i=1}^{\infty} \frac{\kappa_i}{i!}s^i$, gives coefficients κ_i called i^{th} cumulant of the distribution X. The cumulant generating function allows one to relate the cumulants to the moments. It allows us to characterize infinitely divisible LST's which is given in Subsection 2.3. See Feller (1971) [25].

2.3 Infinite Divisibility

There are important results obtained from the concept of infinite divisibility which have many applications in the theory of limit distributions for the sum of independent r.v.'s. In general it is difficult to determine whether a given distribution is infinitely divisible or not. We would like to consider what conditions are required for the pdf of the gamma distributions and its mixture to be infinitely divisible. We also consider the exponential distribution.

We first give notations and a definition. Let the symbol $\stackrel{d}{=}$ denote equality in distribution, and $\stackrel{d}{\longrightarrow}$ denote convergence in distribution.

Definition 2.12 *Consider a random vector X. Its distribution is said to be infinitely divisible if for every $n \in \mathbb{N}$ there exist iid random vectors $X_{n1}, X_{n2}, \cdots, X_{nn}$ with $\sum_k X_{nk} \stackrel{d}{=} X$. In other words, an infinitely divisible r.v. X has pdf $f(x)$ that can be represented as the sum of an arbitrary number of iid r.v.'s $X_1, X_2, \cdots, X_n$, with cdf F_n , that is:*

$$X \stackrel{d}{=} X_1 + X_2 + \cdots + X_n$$

hence the term infinitely divisible. Borrowing from Billingsley (1985) [6] pp. 383-384, the distribution F of X is the n-fold convolution $F_n * F_n * \cdots * F_n$ where F_n is the distribution function of X_i, $1 \leq i \leq n$.

We use that property of infinitely divisible to characterize the distribution of the unknown r.v. from a linear relationship as given in (1.1).

Two simple examples of infinitely divisible distributions are the Poisson distribution and the negative Binomial distribution.

Definition 2.13 *: A function ϕ on the interval $I = [0, \infty)$ is completely monotone if it possesses derivatives $\phi^{(n)}$ at all orders which alternate in sign, i.e. if*

$(-1)^n \phi^{(n)}(s) \geq 0$, *for all s in the interior of I, and $n = 0, 1, 2, \ldots$*

Theorem 2.14 *ϕ is completely monotone iff ϕ is the Laplace transform of some measure.*

Proof: See Feller (1971) [25]. $\square$

For two real valued functions ϕ_1 and ϕ_2 that are completely monotone, so is their product and their compositions, when appropriately chosen.

It is important to note that any completely monotone probability density function is infinitely divisible. See Feller (1971) [25]. Moreover, if ϕ is completely monotone on $[0, \infty)$ and $\phi(c) = 0$ for some $c > 0$, then ϕ must be identically zero on $[0, \infty)$.

All Erlang type distributions are infinitely divisible. The Erlang distribution defined in (1.5), is infinitely divisible in the sense that it can be represented as the sum of n independent r.v.'s with a common distribution. (See Feller (1971) [25] pages 176-177). In fact, n need not be an integer as the gamma can be expressed in a canonical form of an infinitely divisible distribution.

Notice that the distribution of the sum of mutually independent gamma r.v.'s with different scale parameters, is not gamma, even if they are mutually independent. Rather, it is described as a mixed gamma with mixing shape parameter. Using ideas from Mathai and Moschopoulos (1991) [63] and Mathai and Saxena (1978) [65], we characterize the distribution of the sums in Chapter 4.

2.4 Stable Distributions

There are many results about the limit theorems. See Kallenberg (2002) [51] and Billingsley (1986) [6]. We study the class of positive stable distributions. The results could also apply to distributions other than gamma. Here we introduce the necessary definitions and recall limit laws for the joint distribution of the sum of independent distributions from Feller (1971) [25].

Definition 2.15 *(Stable distribution, Feller (1971) [25]) Let $X_0, X_1, X_2, \cdots$ be an iid sequence of r.v.'s with common distribution F. The r.v. X_0 is stable or the distribution F is stable (in the broad sense) if for each n there exist constants $c_n > 0$ and γ_n such that $S_n \stackrel{d}{=} c_n X_0 + \gamma_n$, where $S_n = X_1 + X_2 + \cdots + X_n$ and F is not concentrated at one point. We say F is stable in the strict sense if the above holds with $\gamma_n = 0$.*

Stable distributions are a rich class of probability distributions that allow one to generalize the Central Limit Theorem (CLT), replacing the assumption of finite variance with a less restrictive condition on the infinite variance. For data with heavy tails, stable distributions should fit better than a normal distribution.

From *Definition 2.12* above, the two r.v.'s X_0 and S_n are of the same type as

$$f_{S_n}(x) = f_X\left(\frac{x - \gamma_n}{c_n}\right).$$

Stable distributions can be characterized with the concept of domain of attraction.

Definition 2.16 *(Domain of Attraction, Feller (1971) [25]) The distribution of the independent sequence of r.v.'s $X_1, X_2, \cdots$ belongs to the domain of attraction of a r.v. X_0 with distribution F, if there exist norming or scaling constants $a_n > 0$ and b_n such that*

$$\frac{S_n}{a_n} - b_n = \frac{X_1 + X_2 + \cdots + X_n}{a_n} - b_n \xrightarrow{d} X_0 \ \ as \ \ n \to \infty.$$

The above definition is different from Definition 2.12 in the sense that there is not equality in distribution, but convergence in distribution. In fact, Billingsley (1985) [6] page 389, gives a more restrictive version of Definition 2.13, where for each n, there exit constants $a_n > 0$ and b_n such that $\frac{S_n}{a_n} - b_n$ has same distribution as the iid r.v.'s $X_1, X_2, \cdots, X_n$, and calls it stable. Also, the distribution of X_0 is not necessarily the same as the distribution of the sequence. Finally, this definition recaptures the CLT when we take a sequence of r.v.'s with mean μ and variance σ^2, where $a_n = \frac{\sqrt{n}}{\sigma}$ and $b_n = \frac{n^{3/2}\mu}{\sigma}$ with X_0 being the standard normal distribution. The definition also states that similar limit theorems are possible for distributions without variance, that is stable distributions.

The next result states the equivalence of Definition 2.12 and Definition 2.13.

Theorem 2.17 *A r.v. X_0 with distribution F possesses a domain of attraction iff it is stable.*

Proof: Given in Feller (1971) [25]. $\square$

Hence the non-degenerate limiting distribution of a sum of gamma distributions must be a member of a stable class.

However, gamma distributions are not stable, although they are infinitely divisible. Stable distributions are infinitely divisible, but not the opposite. For another example, consider the case of the Poisson distribution. The Poisson distribution, although infinitely divisible, is not stable (See Billingsley (1986) [6] Section 28 page 389). The class of infinitely divisible distribution is larger than the class of stable distributions. That leads us to explore other possibilities.

Consider now two independent gamma distributions $X_1 \sim Ga(\mu_1, \lambda_1, \alpha_1)$ and $X_2 \sim Ga(\mu_2, \lambda_2, \alpha_2)$. The graph of the joint distribution is given below in Figure 2.1.

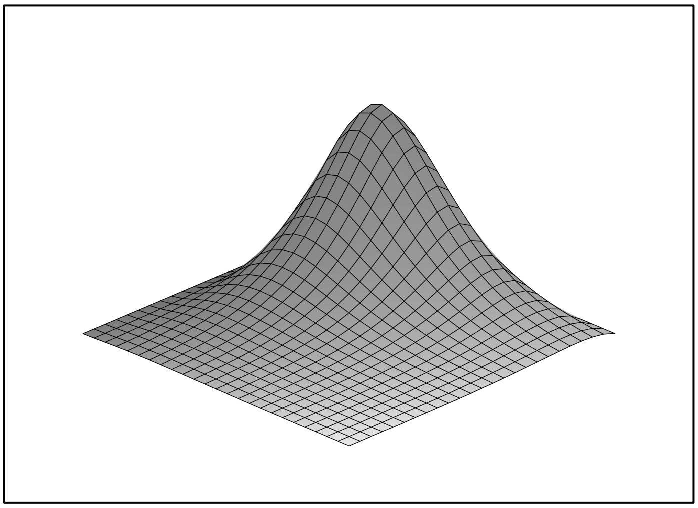

Figure 2.1: Graph of bivariate gamma distribution

We characterize the multivariate gamma distribution in the next chapter.

Chapter 3

The Multivariate Gamma

In this chapter, we propose and characterize a quite general multivariate gamma distribution, with specified three parameter gamma marginals of Type III in the Pearson's system. We introduce some notations and basic concepts of multivariate gamma distributions. Because a multivariate distribution function is completely characterized by its mean, variance and covariance structure, we give these properties. We also propose a series of parameter estimation techniques that can be adapted to special classes of distributions such as those discussed in Chapter 4 (Continuous class) and Chapter 5 (Fatal shock).

Multivariate gamma distributions belong in the class of multivariate lifetime distributions, since all the components of the random vector have positive support and their marginals follow a typical lifetime distribution. Hougaard (2000) [36] among other authors give many applications of the multivariate lifetime distributions. As he mentioned, the field has gaps and is still growing both in theory and application.

The bivariate exponential, one simple case of bivariate gamma, plays a fundamental role in survival and reliability analysis because of its adaptation to many practical situations. The complexity of the model, in its notation and estimation technique, increases with the number of components. Within this framework, Moran (1969) [69], among other authors, had also described a form of a bivariate gamma process and given estimation techniques of one parameter using the other variate as a control. It has been difficult to work with until the development of proportional hazard models or frailty models. See Cox (1972) [20], Ghosh and Gelfand (1998) [30] and Hougaard (2000) [36] to mention a few. However, we do not deal with the proportional hazard model here, since it is primarily used for the univariate survival case. In fact, Cox's proportional hazard model, sometimes considered as nonparametric, is typically treated as multivariate in only the recurring events situations, such as monitoring times until repeated hospitalizations of the same individual, or repeated failures of a device in a system. Our model is truly multivariate in the sense that the associated parameters are mathematically independent and we can model components with different means and variances.

3.1 Properties and Characterization

In the next definition, we formally define our multivariate gamma that was briefly described in (1.1).

Definition 3.1 *: Let* $X_0, X_1, \ldots, X_p$ *gamma r.v.'s with shape parameters* α_i*, scale parameters* λ_i *and location parameters* μ_i*, i.e* $X_i \sim Ga(\mu_i, \lambda_i, \alpha_i)$ *for* $i = 0, 1, \ldots, p$ *as given in (1.2). Let* $Z_1, Z_2, \ldots, Z_p$ *be mutually independent r.v.'s satisfying* $X_i = a_i X_0 + Z_i$ *for some* $a_i > 0$ *and* Z_i *independent of* X_0 *for* $i = 1, 2, \ldots, p$*. We define the joint distribution of the random vector* $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)'$ *as the multivariate gamma distribution for a fixed integer* p*.*

Although the density function is not explicitly expressed in Definition 3.1, we can study various properties of the multivariate gamma distribution by using the linear relationship given in (1.1), and the fact that $Z_i,\ i = 1, .., p$, are independent of each other and of X_0. Note that the distribution of $Z_i,\ i = 1, .., p$, has not been specified. As we will explain later, depending on the choice of the constants a_i, the distributions of Z_i take on many forms. In fact, in some cases, Z_i is a gamma, and in other cases, Z_i is a mixture of distributions.

The linear correlation between X_i and the latent term X_0 is expressed as:

$$\rho(X_i, X_0) \quad = \quad \frac{Cov(X_i, X_0)}{\sqrt{Var(X_i)}\sqrt{Var(X_0)}},$$

where $Cov(X_i, X_0) = E(X_0 X_i) - E(X_0)E(X_i)$ is the covariance between X_0 and X_i, and $Var(X_0)$ and $Var(X_i)$ denote the variances of X_0 and X_i, respectively.

Note that if the Z_i were degenerate, say $Z_i \equiv b_i$, or $P(Z_i = b_i) = 1$, $b_i \in \mathbb{R}^+$, $i = 1, 2, .., p$, then the linear correlation, a measure of linear dependence, would be perfect. That is $Z_i \equiv b_i \Rightarrow X_i = a_i X_0 + b_i$, which gives $\rho(X_0, X_i) = \pm 1$, or more precisely $\rho(X_i, X_j) = sgn(a_i a_j)$. The degenerate case is the only case where $|\rho| = 1$.

This linear relation (1.1) seems to be natural as pointed out and motivated in Carpenter et al. (2006) [12], where properties for the exponential model are given. Similarly, from this structural relationship, we derive the following values of the means, variances, and covariances:

$$\begin{aligned}
E(X_i) &= \frac{\alpha_i}{\lambda_i} + \mu_i, && (3.1)\\
E(Z_i) &= E(X_i) - a_i E(X_0) = (\frac{\alpha_i}{\lambda_i} - a_i \frac{\alpha_0}{\lambda_0}) + (\mu_i - a_i \mu_0), && (3.2)\\
Var(X_i) &= \frac{\alpha_i}{\lambda_i^2}, && (3.3)\\
Var(Z_i) &= Var(X_i) - a_i^2 Var(X_0) = \frac{\alpha_i}{\lambda_i^2} - a_i^2 \frac{\alpha_0}{\lambda_0^2} > 0 \Longleftrightarrow \frac{\alpha_i}{\alpha_0} > \left(\frac{a_i \lambda_i}{\lambda_0}\right)^2, && (3.4)\\
Cov(X_i, X_j) &= a_i a_j Var(X_0) = a_i a_j \frac{\alpha_0}{\lambda_0^2} \quad \Longleftrightarrow \rho(X_i, X_j) = a_i a_j \frac{\lambda_i \lambda_j}{\lambda_0^2} \frac{\alpha_0}{\sqrt{\alpha_i \alpha_j}}, && (3.5)\\
Cov(Z_i, Z_j) &= 0, && (3.6)
\end{aligned}$$

for $i \neq j, \quad i, j = 1, .., p$.

It is clear from Definition 3.1 and Equation (3.5) that X_i and X_j, $\quad i \neq j,\ i,j = 1,..,p$, are positively correlated. This is a restriction in our model, but not unusual in this field. See He and Lawless (2005) [33] for more examples of positive correlations in lifetime data. Also, note that (3.1), (3.3), (3.4) and (3.5) can be more concisely expressed in the mean vector of the mean of $\mathbf{X}$:

$$E(\mathbf{X}) = \left(\frac{\alpha_1}{\lambda_1} + \mu_1, \frac{\alpha_2}{\lambda_2} + \mu_2, \cdots, \frac{\alpha_p}{\lambda_p} + \mu_p\right)', \tag{3.7}$$

the variance/covariance matrix:

$$\Sigma = \begin{pmatrix} \frac{\alpha_1}{\lambda_1^2} & a_1a_2\frac{\alpha_0}{\lambda_0^2} & \cdots & a_1a_p\frac{\alpha_0}{\lambda_0^2} \\ a_2a_1\frac{\alpha_0}{\lambda_0^2} & \frac{\alpha_2}{\lambda_2^2} & \cdots & a_2a_p\frac{\alpha_0}{\lambda_0^2} \\ \vdots & \vdots & \vdots & \vdots \\ a_pa_1\frac{\alpha_0}{\lambda_0^2} & \cdots & a_pa_{p-1}\frac{\alpha_0}{\lambda_0^2} & \frac{\alpha_p}{\lambda_p^2} \end{pmatrix} = \frac{\alpha_0}{\lambda_0^2}\begin{pmatrix} \frac{\lambda_0^2}{\lambda_1^2}\frac{\alpha_1}{\alpha_0} & a_1a_2 & \cdots & a_1a_p \\ a_2a_1 & \frac{\lambda_0^2}{\lambda_2^2}\frac{\alpha_2}{\alpha_0} & \cdots & a_2a_p \\ \vdots & \vdots & \vdots & \vdots \\ a_pa_1 & \cdots & a_pa_{p-1} & \frac{\lambda_0^2}{\lambda_p^2}\frac{\alpha_p}{\alpha_0} \end{pmatrix},$$

and the correlation matrix:

$$\rho = \frac{\alpha_0}{\lambda_0^2}\begin{pmatrix} \frac{\lambda_0^2}{\alpha_0} & \frac{a_1a_2\lambda_1\lambda_2}{\sqrt{\alpha_1\alpha_2}} & \cdots & \frac{a_1a_p\lambda_1\lambda_p}{\sqrt{\alpha_1\alpha_p}} \\ \frac{a_2a_1\lambda_2\lambda_1}{\sqrt{\alpha_1\alpha_2}} & \frac{\lambda_0^2}{\alpha_0} & \cdots & \frac{a_2a_p\lambda_2\lambda_p}{\sqrt{\alpha_2\alpha_p}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{a_pa_1\lambda_p\lambda_1}{\sqrt{\alpha_p\alpha_1}} & \cdots & \frac{a_pa_{p-1}\lambda_p\lambda_{p-1}}{\sqrt{\alpha_p\alpha_{p-1}}} & \frac{\lambda_0^2}{\alpha_0} \end{pmatrix}.$$

For $X_i \sim Ga(\mu_i, \lambda_i, \alpha_i)$ for $i = 0, 1, \cdots, p$, based on the LST given in (2.1) and using the binomial theorem, we have the following:

$$\frac{d^m L_{X_i}(s)}{ds^m} = \lambda^\alpha \sum_{k=0}^{m} \binom{m}{k} \left\{\frac{d^k(\lambda_i + s)^{-\alpha}}{ds^k}\right\}\left\{\frac{d^{m-k}e^{-\mu_i s}}{ds^{m-k}}\right\}, \quad \text{for } m \in \mathbb{N}.$$

Setting $s = 0$, we have that

$$E(X_i^m) = \sum_{k=0}^{m} \binom{m}{k} \lambda_i^{\alpha+k} \mu_i^{m-k} (\alpha_i)_k,$$

where $(\alpha)_k = \alpha(\alpha+1)\cdots(\alpha+k-1)$.

The moments and cross-moments are given for m and $l \in \mathbb{N}$, by:

$$\begin{aligned}
E(X_i^m) &= E(a_i X_0 + Z_i)^m = \sum_{r=0}^{m} \binom{m}{r} a_i^r E X_0^r E Z_i^{m-r} \\
&= \sum_{r=0}^{m} \binom{m}{r} E Z_i^{m-r} a_i^r \sum_{k=0}^{r} \binom{r}{k} \lambda_i^{\alpha+k} \mu_i^{r-k} (\alpha_i)_k
\end{aligned}$$

$$\begin{aligned}
\text{and } E(X_i^m X_j^l) &= E(a_i X_0 + Z_i)^m (a_j X_0 + Z_j)^l \\
&= E \sum_{r=0}^{m} \binom{m}{r} a_i^r X_0^r Z_i^{m-r} \sum_{s=0}^{l} \binom{l}{s} a_j^s X_0^s Z_j^{l-s} \\
&= \sum_{r=0}^{m} \sum_{s=0}^{l} \binom{m}{r} \binom{l}{s} a_i^r a_j^s E X_0^{r+s} E Z_i^{m-r} E Z_j^{l-s}, \ i \neq j, \ i,j = 1,2,\ldots,p.
\end{aligned}$$

The LST has many helpful properties as discussed in Section 2.1. In particular, in this dissertation, we use it for many of the derivations of the distributions of the latent variables. The multivariate LST for any p-variate random vector $\mathbf{X}$ defined in (1.1) is:

$$\begin{aligned}
L_{\boldsymbol{X}}(\boldsymbol{s}) &= E e^{-\mathbf{s'X}} = E e^{-\sum_{i=1}^{p} s_i X_i} \\
&= E e^{-\sum_{i=1}^{p} s_i (a_i X_0 + Z_i)} = E \prod_{i=1}^{p} e^{-a_i s_i X_0} e^{-s_i Z_i} \\
&= E e^{-X_0 \sum_{i=1}^{p} a_i s_i} \prod_{i=1}^{p} E e^{-s_i Z_i} = L_{X_0}\left(\sum_{i=1}^{p} a_i s_i \right) \prod_{i=1}^{p} L_{Z_i}(s_i) \\
&= L_{X_0}\left(\sum_{i=1}^{p} a_i s_i \right) \prod_{i=1}^{p} \frac{L_{X_i}(s_i)}{L_{X_0}(a_i s_i)} \quad ,
\end{aligned}$$

for $\mathbf{s} = (s_1, s_2, \ldots, s_p)'$, regardless of the underlying distributions as long as the LST exists. If $X_0 \sim Ga(\mu_0, \lambda_0, \alpha_0)$, then the above becomes

$$L_{\boldsymbol{X}}(\boldsymbol{s}) \quad = \quad e^{-\mu_0 \sum_{i=1}^{p} a_i s_i} \left(\frac{\lambda_0}{\lambda_0 + \sum_{i=1}^{p} a_i s_i} \right)^{\alpha_0} \prod_{i=1}^{p} \frac{L_{X_i}(s_i)}{L_{X_0}(a_i s_i)}. \tag{3.8}$$

Note that from the above, the LST of Z_i is:

$$L_{Z_i}(s) = \frac{L_{X_i}(s)}{L_{X_0}(a_i s)}, \quad \mathit{for} \quad i = 1, 2, .., p. \tag{3.9}$$

As long as L_{Z_i} in Equation (3.9) is completely monotone as in Definition 2.13, the conditions of Definition 3.1 are satisfied, and the marginals are gamma. Equation (3.9) will be used extensively throughout this dissertation, but for now we present an important theorem that gives some properties on the scale and shift transformations of the multivariate gamma distribution given in Definition 3.1.

Theorem 3.2 *The class of multivariate gamma distributions based on Definition 3.1 is closed under scale transformation in the sense of products with diagonal matrices with positive entries, shift transformations, and under finite independent convolutions.*

Proof: Consider a multivariate gamma $\mathbf{X} = (X_1, \cdots, X_p)'$, and let $C = (c_i)_{i=1,..,p}$ be a diagonal matrix with positive entries. Set $\mathbf{Y} = C\mathbf{X}$. Then we have that

$$\begin{aligned} L_{\boldsymbol{Y}}(\boldsymbol{s}) &= Ee^{-\mathbf{s'Y}} = Ee^{-\sum_{i=1}^{p} s_i Y_i} \\ &= Ee^{-\sum_{i=1}^{p} c_i s_i (a_i X_0 + Z_i)} = E\prod_{i=1}^{p} e^{-a_i c_i s_i X_0} e^{-c_i s_i Z_i} \\ &= Ee^{-X_0 \sum_{i=1}^{p} a_i c_i s_i} \prod_{i=1}^{p} Ee^{-c_i s_i Z_i} = L_{X_0}(\sum_{i=1}^{p} a_i c_i s_i) \prod_{i=1}^{p} L_{Z_i}(c_i s_i) \end{aligned}$$

$$= \quad e^{-\mu_0 \sum_{i=1}^{p} a_i c_i s_i} \left(\frac{\lambda_0}{\lambda_0 + \sum_{i=1}^{p} a_i c_i s_i} \right)^{\alpha_0} \prod_{i=1}^{p} \frac{L_{X_i}(c_i s_i)}{L_{X_0}(a_i c_i s_i)},$$

for $\mathbf{s} = (s_1, s_2, \ldots, s_p)$, which is of the same form as in (3.8). The mean $E\mathbf{Y} = CE\mathbf{X}$ and the variance is expressed as: $Var(\mathbf{Y}) = C'Var(\mathbf{X})C$.

Let $\mathbf{d} = (d_1, d_2, \cdots, d_p)$ and $\mathbf{W}$ defined as: $\mathbf{W} = (W_1, W_2, \cdots, W_p)' := \mathbf{X} + \mathbf{d} = (X_1 + d_1, X_2 + d_2, \cdots, X_p + d_p)$. Then, we have the following:

$$\begin{aligned}
L_{\boldsymbol{W}}(\boldsymbol{s}) &= Ee^{-\mathbf{s'W}} = Ee^{-\sum_{i=1}^{p} s_i W_i} = Ee^{-\sum_{i=1}^{p} s_i(X_i + d_i)} \\
&= Ee^{-\sum_{i=1}^{p} s_i d_i} e^{-\sum_{i=1}^{p} s_i(a_i X_0 + Z_i)} = Ee^{-\sum_{i=1}^{p} s_i d_i} \prod_{i=1}^{p} e^{-a_i s_i X_0} e^{-s_i Z_i} \\
&= e^{-\sum_{i=1}^{p} s_i d_i} Ee^{-X_0 \sum_{i=1}^{p} a_i s_i} \prod_{i=1}^{p} Ee^{-s_i Z_i} \\
&= e^{-\sum_{i=1}^{p} s_i d_i} L_{X_0}(\sum_{i=1}^{p} a_i s_i) \prod_{i=1}^{p} L_{Z_i}(s_i) \\
&= e^{-\sum_{i=1}^{p} s_i d_i} e^{-\mu_0 \sum_{i=1}^{p} a_i s_i} \left(\frac{\lambda_0}{\lambda_0 + \sum_{i=1}^{p} a_i s_i} \right)^{\alpha_0} \prod_{i=1}^{p} \frac{L_{X_i}(s_i)}{L_{X_0}(a_i s_i)} \\
&= e^{-\sum_{i=1}^{p} (d_i + \mu_0 a_i) s_i} \left(\frac{\lambda_0}{\lambda_0 + \sum_{i=1}^{p} a_i s_i} \right)^{\alpha_0} \prod_{i=1}^{p} \frac{L_{X_i}(s_i)}{L_{X_0}(a_i s_i)},
\end{aligned}$$

which is of the same form as in (3.8). So the multivariate gamma is closed under shift transformation of the form $\mathbf{W} = \mathbf{X} + \mathbf{d}$.

Consider now n independent p-variate gamma distributions $\mathbf{X_i}$=$(X_{i1}, X_{i2}, \cdots, X_{ip})'$ for $i = 1, \cdots, n$, and set

$\mathbf{X} = \mathbf{X_1} + \cdots + \mathbf{X_n}$= $(X_{11} + \cdots + X_{n1}, X_{12} + \cdots + X_{n2}, \cdots, X_{1p} + \cdots + X_{np})'$.

$$L_{\mathbf{X}}(\mathbf{s}) \quad = \quad Ee^{-\mathbf{s}'(\mathbf{X_1} + \cdots + \mathbf{X_n})} \quad = \quad Ee^{-\sum_{i=1}^{p} s_i(X_{1i} + \cdots + X_{ni})}$$

$$= \prod_{i=1}^{p} Ee^{-s_i(X_{1i}+\cdots+X_{ni})} \quad = \quad L_{\mathbf{X_1}}(s)\cdots L_{\mathbf{X_n}}(s),$$

which has the form of n product of (3.8). ∎

Recall that the distribution of Z_i has not been specified, but its LST is given in (3.9). In the next theorem, we state a very interesting property, as it gives a necessary and sufficient condition for all the latent variables Z_i to be of the gamma type distributions (1.2).

Theorem 3.3 *Suppose that for some* $a_i > 0$, *i=1,...,p*, $X_i \sim Ga(\lambda_i, \alpha_i)$ *i=0,1,...,p, are related as in Definition 3.1 with* Z_i, $i = 1, \ldots, p$; *then*

$$Z_i \sim Ga(\lambda_i, \alpha_i - \alpha_0) \Longleftrightarrow a_i = \frac{\lambda_0}{\lambda_i} \quad \textit{and} \quad p_i = P(Z_i = 0) = 0.$$

Proof: From the expression of $a_i = \dfrac{\lambda_0}{\lambda_i}$ and the result in (1.10), we have that $a_i X_0 \sim Ga(\lambda_i, \alpha_0)$. Since $X_i \sim Ga(\lambda_i, \alpha_i)$, it follows from Example 2.3 (the additive property of the gamma distribution) that Z must be distributed as $Z_i \sim Ga(\lambda_i, \alpha_i - \alpha_0)$.
Conversely, assume that $Z_i \sim Ga(\lambda_i, \alpha_i - \alpha_0)$. Then from Example 2.2 and the independence between X_0 and Z_i, we have that:

$$L_{X_i}(t) = L_{X_0}(a_i t) L_{Z_i}(t) \Longrightarrow L_{X_0}(a_i t) = \left(\frac{\lambda_0}{\lambda_0 + a_i t}\right)^{\alpha_0} = \left(\frac{\lambda_i}{\lambda_i + t}\right)^{\alpha_i} \left(\frac{\lambda_i}{\lambda_i + t}\right)^{\alpha_0 - \alpha_i}.$$

Hence $\qquad \dfrac{\lambda_0}{\lambda_0 + a_i t} = \dfrac{\lambda_i}{\lambda_i + t},$ $\quad$ and solving for a_i gives

$$\frac{\lambda_0 + a_i t}{\lambda_0} = \frac{\lambda_i + t}{\lambda_i} \quad \Longrightarrow \quad 1 + \frac{a_i t}{\lambda_0} = 1 + \frac{t}{\lambda_i} \quad \text{or} \quad a_i = \frac{\lambda_0}{\lambda_i}.$$ ∎

Mathai and Moschopoulos (1991) [63] among other authors made use of this particular coefficient of linear relationship between two gamma distributions, and developed properties, approximations, and inequalities based on that particular model. In this dissertation, we allow for more general representations.

In Chapter 4, we examine further this specific form, and call it the continuous class of distributions. In all other cases where the conditions of Theorem 3.3 are not satisfied, we call that form the "fatal shock" class, because there is a positive probability of simultaneous or proportional occurrences. In the Section 3.2, we focus on the characterization of the distribution of $Z_i,\ i = 1, 2, .., p$ in the fatal shock class. This last type of distributions is further studied in Chapter 5.

All the Z_i's take on similar structures. To simplify the notations, without lost of generality, we drop the indices, and we rewrite (1.1) as:

$$Y = aX + Z, \qquad \text{for some } a > 0, \tag{3.10}$$

where Y, X, a and Z in (3.10) play the role of X_i, X_0, a_i and Z_i in (1.1) respectively.

A simple computation shows that the cdf of Z is:

$$\begin{aligned}
P(Z \leq z) &= P(Y - aX \leq z) \\
&= P(Y = aX) + (1-p)P(X \geq \frac{Y-z}{a}) \\
&= P(Y = aX) + (1-p)\Big[1 - P(X \leq \frac{Y-z}{a})\Big] \\
&= P(Y = aX) + (1-p)\Big[1 - \int_z^\infty F_X(\frac{y-z}{a}) f_Y(y)dy\Big].
\end{aligned}$$

Since

$$\int_z^\infty F_X(\frac{y-z}{a}) f_Y(y)dy = F_X(\frac{y-z}{a})F_Y(y)\Big]_z^\infty - \int_z^\infty \frac{1}{a} f_X(\frac{y-z}{a})F_Y(y)dy$$

$$= \quad 1 - \int_z^{\infty} \frac{1}{a} f_X(\frac{y-z}{a}) F_Y(y) dy,$$

the cdf can be further simplified as

$$P(Z \leq z) \quad = \quad P(Y = aX) + \frac{1-p}{a} \int_z^{\infty} f_X(\frac{y-z}{a}) F_Y(y) dy,$$

where $P(Y = aX) = P(Z = 0)$ is the probability of proportional occurrence.

Note that if the conditions of Theorem 3.3 are not satisfied, then $P(Y = aX) = P(Z = 0) > 0$. In Section 3.2, we further characterize the distribution of Z.

3.2 The Fatal Shock Model

In this section, we study the form of the distribution of Z obtained in (3.10) for a general $a > 0$ where $a \neq \dfrac{\lambda_0}{\lambda_i}$.

Suppose that X and Y are distributed as $X \sim Ga(\mu, \lambda, \alpha)$ and $Ga(\mu', \alpha', \lambda')$, respectively. Assuming a linear relationship as in (3.10), and from Equation (3.9), where a is some positive constant, and Z is independent of X, the LST of Z becomes:

$$L_Z(s) = \frac{L_Y(s)}{L_X(as)} = e^{(\mu-\mu')s} \left(\frac{\lambda'}{\lambda' + s} \right)^{\alpha'} \left(\frac{\lambda + as}{\lambda} \right)^{\alpha}.$$

All distributions Z satisfying this relation (3.10) for all $a \in (0, 1)$, are called members of the self decomposable r.v.'s. See also Gaver and Lewis (1980) [29].

Previously, we have studied the properties of the r.v. Z, derive its LST in (3.9), mean in (3.2), variance and covariance in (3.4) and (3.6), respectively. However, the form of the distribution of Z has not been given. In the next theorems, we give the required forms that the distribution of Z must take on to give gamma marginals. We state the main theorems, and start with the Erlang case defined in (3.10).

Theorem 3.4 *Erlang type distributions.*

Let $X \sim Ga(\mu, \lambda, \alpha)$ and $Y \sim Ga(\mu', \lambda', \alpha')$ such that $\alpha, \alpha' \in \mathbb{N}, \alpha \leq \alpha'$ and as in (3.10), Z is independent of X. Then Z is the shifted sum of independent:

- *α r.v.'s, each being product of Bernoulli and exponential distribution with parameter λ',*
- *and of $\alpha' - \alpha$ gamma r.v.'s with scale λ' and shape $k = 1, .., \alpha' - \alpha$.*

Proof: Using the Laplace transform technique, we have:

$$\begin{aligned} L_Z(s) &= \frac{L_Y(s)}{L_X(as)} = e^{-(\mu-\mu')s}\frac{\lambda'^{\alpha'}}{\lambda^{\alpha}}\frac{(\lambda+as)^{\alpha}}{(\lambda'+s)^{\alpha'}} \\ &= e^{-(\mu-\mu)s}\frac{\lambda'^{\alpha'}}{\lambda^{\alpha}}\frac{(\lambda+as)^{\alpha}}{(\lambda'+s)^{\alpha}}\left[\frac{1}{(\lambda'+s)^{\alpha'-\alpha}}\right] \end{aligned}$$

By the theorem of expansion in partial fractions, with α instead of $\alpha' - \alpha$,

$$\frac{1}{(\lambda'+s)^{\alpha}} = \sum_{k=1}^{\alpha}\frac{C_k}{(\lambda'+s)^k} \qquad \text{where } C_k \text{ are real numbers.}$$

Hence:

$$L_Z(s) = e^{-(\mu-\mu')s}\frac{\lambda'^{\alpha'}}{\lambda^{\alpha}}\Big[p + (1-p)L_Y(s)\Big]^{\alpha}\sum_{k=1}^{\alpha'-\alpha}\frac{C_k}{\lambda'^k}\left(\frac{\lambda'}{\lambda'+s}\right)^k.$$

where $C_k,\ k = 1, 2, \ldots, \alpha' - \alpha$, are constant and $p = a\dfrac{\lambda}{\lambda'}$. Based on the properties of the LST from Section 2.1, we have that $\Big[p + (1-p)L_Y(s)\Big]$ is the product of Bernoulli and exponential r.v.'s as discussed in Example 2.4 and in Carpenter et al. (2006) [12]. So $\Big[p + (1-p)L_Y(s)\Big]^{\alpha}$ is the product of α independent of those terms.

Also, since the LST of a sum of functions is the sum of their LST, then $\sum_{k=1}^{\alpha'-\alpha} \frac{C_k}{\lambda'^k}\left(\frac{\lambda'}{\lambda'+s}\right)^k$ is the LST of the sum of $\alpha' - \alpha$ gamma distributions.

$\square$

The special case of the exponential where $\alpha = \alpha' = 1$ has been studied in Carpenter et al. (2006) [12], and by Iyer et al. (2002) [42]. A similar result in the gamma context has been given in Carpenter and Diawara (2006) [11].

Theorem 3.5 *Suppose $X \sim Ga(\mu, \lambda, \alpha)$, with $\alpha \in \mathbb{N}$, and $Y \sim Ga(\mu', \lambda', \alpha')$ as in (3.10), and $\alpha \leq \alpha'$. Then Z is the sum of two r.v.'s:*

- *a $Ga(\mu - \mu', \lambda', \alpha' - \alpha)$ and*
- *another distribution which is the sum of α independent r.v.'s each of which is a product of a Bernoulli and an exponential with parameter λ'.*

Proof: From the expression of L_Z,

$$L_Z(s) = \left[\frac{\lambda'}{\lambda}\frac{(\lambda + as)}{(\lambda' + s)}\right]^{\alpha} e^{(\mu-\mu')s}\left[\frac{\lambda'}{\lambda' + s}\right]^{\alpha'-\alpha},$$

the result follows. $\square$

Also, assuming $Z = X_2 - aX_1$, and using the Taylor approximation of

$$\log(1+s) = \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{(n+1)!} s^{n+1},$$

the cumulants of Z are obtained as

$$\begin{aligned}
\log L_Z(s) &= \log L_{X_2}(s) - \log L_{X_1}(as) \\
&= \alpha_1 \log(1 - \frac{as}{\lambda_1}) - \alpha_2 \log(1 - \frac{s}{\lambda_2})
\end{aligned}$$

$$
\begin{aligned}
&= \alpha_1 \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{(n+1)!} (\frac{as}{\lambda_1})^{n+1} - \alpha_2 \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{(n+1)!} (\frac{s}{\lambda_2})^{n+1} \\
&= \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{(n+1)!} s^{n+1} \left[\alpha_1 (\frac{a}{\lambda_1})^{n+1} - \alpha_2 (\frac{1}{\lambda_2})^{n+1} \right]
\end{aligned}
$$

Another characterization of Z is obtained by considering several cases for the $Ga(\mu, \lambda, \alpha)$ and $Ga(\mu', \lambda', \alpha')$.

Corollary 3.6 *Let $\alpha = \alpha' \in \mathbb{N}$, then Z is a shifted sum of α independent r.v.'s, each being a product of a Bernoulli with mean $(1-p)$ and an exponential with parameter λ', where the value of the shift is $\mu - \mu'$.*

Proof: $\alpha \in \mathbb{N}$. Then the LST of Z can be expressed as:

$$
\begin{aligned}
L_Z(s) &= e^{(\mu-\mu')s} \left[\frac{\lambda'}{\lambda} \frac{(\lambda + as)}{(\lambda' + s)} \right]^{\alpha} \\
&= e^{(\mu-\mu')s} \left[(1-p)(\frac{\lambda'}{\lambda' + s}) + p \right]^{\alpha}.
\end{aligned}
$$

□

Theorem 3.7 *Suppose $X \sim Ga(\mu, \lambda, \alpha)$, and $Y \sim Ga(\mu', \lambda', \alpha')$ as in (3.10), with $\alpha < \alpha'$, and assume that $[\alpha + 1] \leq \alpha'$. Then Z is is the sum of*

- *$[\alpha + 1]$ independent r.v.'s, each being a product of a Bernoulli with mean $(1-p)$ and an exponential with parameter λ'*
- *$Ga(\mu - \mu', \lambda', \alpha' - [\alpha + 1])$*
- *$Ga(0, \lambda/a, [\alpha + 1] - \alpha)$*

Proof:
$$L_Z(s) = \left[\frac{\lambda'}{\lambda}\frac{\lambda+as}{\lambda'+s}\right]^{[\alpha+1]} e^{(\mu-\mu')s}\left[\frac{\lambda'}{\lambda'+s}\right]^{\alpha'-[\alpha+1]}\left[\frac{\lambda+as}{\lambda}\right]^{\alpha-[\alpha+1]}$$
and notice that:
$$\left[\frac{\lambda+as}{\lambda}\right]^{\alpha-[\alpha+1]} = \left[\frac{\lambda}{\lambda+as}\right]^{[\alpha+1]-\alpha}.$$

∎

In fact, instead of assuming $[\alpha+1] \leq \alpha'$, it suffices to assume that $[\alpha+\epsilon] \leq \alpha'$ for some $\epsilon \in [0,1]$.

Theorem 3.8 *Suppose $X \sim Ga(\mu,\lambda,\alpha)$, and $Y \sim Ga(\mu',\lambda',\alpha')$ as in (3.10), with $\alpha = \alpha' \notin \mathbb{N}$, then Z does not have the LST of some measure.*

Proof: $\alpha \notin \mathbb{N}$ The explicit formula in terms of the LST is

$$L_Z(s) \quad = \quad \left[\frac{\lambda'}{\lambda}\frac{(\lambda+as)}{(\lambda'+s)}\right]^{[\alpha]} * \left[\frac{\lambda'}{\lambda'+s}\right]^{\alpha-[\alpha]} * e^{(\mu-\mu')s}\left[\frac{\lambda+as}{\lambda}\right]^{\alpha-[\alpha]}$$

The question we have is how to interpret

$$\left[\frac{\lambda+as}{\lambda}\right]^{\alpha-[\alpha]}.$$

However, set $\phi(s) = \left[\frac{\lambda+as}{\lambda}\right]^{\beta}$, where $\beta = \alpha - [\alpha] \in [0,1]$.

$$\begin{aligned}
\phi^{(1)}(s) &= \beta\frac{a}{\lambda}\left[\frac{\lambda+as}{\lambda}\right]^{\beta-1}\\
\phi^{(2)}(s) &= \beta(\beta-1)\left(\frac{a}{\lambda}\right)^2\left[\frac{\lambda+as}{\lambda}\right]^{\beta-2}\\
&\vdots\\
\phi^{(n)}(s) &= \beta(\beta-1)\cdots(\beta-n+1)\left(\frac{a}{\lambda}\right)^n\left[\frac{\lambda+as}{\lambda}\right]^{\beta-n},
\end{aligned}$$

So $\phi(0) = 1$, and has derivatives that alternate in sign. However, by Theorem 2.14, from Feller (1971) [25], $(-1)^n \phi(s) \leq 0$, for all $n = 0, 1, 2, \ldots$ ϕ is not completely monotone. We have that ϕ is not the Laplace transform of some probability distribution. $\square$

Theorem 3.9 *Let $X \sim Ga(\mu, \lambda, \alpha)$ and $Y \sim Ga(\mu', \lambda', \alpha')$ such that $\alpha \in \mathbb{N}$, $\alpha' \in \mathbb{R}_+$, $\alpha' \leq \alpha$ and as in (3.10). Then Z has LST of some measure iff $\alpha = \alpha'$.*

Proof:

$$L_Z(s) = \frac{L_Y(s)}{L_X(as)} = e^{(\mu-\mu')s}\left(\frac{\lambda'}{\lambda'+s}\right)^{\alpha'}\left(\frac{\lambda+as}{\lambda}\right)^{\alpha}$$
$$= e^{(\mu-\mu')s}\left[\frac{\lambda'}{\lambda}\frac{(\lambda+as)}{(\lambda'+s)}\right]^{\alpha}\left[\frac{\lambda+as}{\lambda}\right]^{\alpha-\alpha'}.$$

The expression $\left[\frac{\lambda+as}{\lambda}\right]^{\alpha-\alpha'}$ is completely monotone iff $\alpha \leq \alpha'$. Since $\alpha' \leq \alpha$ must be satisfied also, then $\alpha = \alpha'$. $\square$

We have seen that the distribution of Z takes on many different forms depending on the parameters that are given. In Theorem 3.3, we found the only case where the latent variable is of continuous gamma type as the marginals. This case is further studied in Chapter 4. The results from Theorem 3.4 up to Theorem 3.9 give the various forms that the distribution of Z can take. Theorem 3.4 and Theorem 3.5 are similar, as they express the distribution of Z as a sum of product of Bernoulli and exponential, and that of gamma r.v.'s. The special case, given in the Corollary 3.6, states that the Erlang case where $\alpha = \alpha' \in \mathbf{N}$, leads to just the product of α Bernoulli and exponential. These results give alternative ways of characterizing the distribution of the latent variables, an opportunity to replace the unconditional expectation, density by the unconditional expectation, density. That allow to increase the efficiency in the variance covariance matrix of parameters. In all those cases,

$\alpha \leq \alpha'$ is a required condition for the model validation. The Theorem 3.7 shows that there must be an integer value between α and α' to justify the gamma marginals for Y and X in the Equation (3.10). If there is no such integer value between α and α', then Z does not have the LST of some measure. It is not be represented by a distributional form. Theorem 3.9 confirms that finding in the sense that it gives back the result obtained in Corollary 3.6.

3.3 The Multivariate Gamma

The previous section overviews the possible forms of distribution that Z, based on Equation (3.10), can take. Our primary goal is to obtain the multivariate gamma form of $\mathbf{X}$ as in Definition 3.1. That means we need to study the distribution of $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_p)$ based on Equation (1.1) and Definition 3.1.

In Chapter 4, we study the case where the latent variables Z_i are all of gamma types as in the result obtained from Theorem 3.3. In that case the probability of proportional occurrence is $p_i = 0$, for $i = 1, 2, .., p$. Many authors have suggested that formulation. Mathai and Moschopoulos (1991) [63] use that form to give a version of multivariate gamma. They give properties of such model in terms of means, variance and covariance of parameters, approximations, and inequalities. In this dissertation, we have proposed a general model as in Definition 3.1 of a multivariate gamma, that simplifies to the one proposed by Mathai and Moschopoulos (1991) [63], and Mathai and Moschopoulos (1992) [64]. Nadarajah and Kotz (2005) [72] derive a linear combination of exponential and gamma r.v.'s. They assume that there is no probability of simultaneous occurrence between X_i and X_0. However, as given in the famous paper by Marshall and Olkin (1969) [61], that is not meaningful in "shock model" conditions. Our model captures the shock model that Marshall and Olkin

(1967) [61] studied in the simple case of the multivariate exponential distribution. That lead us to elaborate on the multivariate exponential distribution, and the results are given in Chapter 5.

When $a_i = \dfrac{\lambda_0}{\lambda_i}, \ , i = 1, 2, .., p$ in Equation (1.1), then $\mathbf{X}$ reduces to multivariate gamma proposed by Mathai and Moschopoulos (1991) [63]. We give the form of the joint distribution of $\mathbf{Z}$ in the next chapter, its interprtation and estimation procedures of its parameters.

Iyer et al. (2002) [42] and Iyer et al. (2004) [43] dealt with the case of positive fixed a_i's in the one dimensional case assuming that X_0 and X_i are exponential with parameters λ_0 and $\lambda_i, \ , i = 1, 2, .., p$, respectively. Carpenter et al. (2006) [12] showed that the density of Z_i is then expressed as:

$$f_{Z_i}(z) = p_i \delta(z) + (1 - p_i) f_{X_i}(z) I_{(z>0)}, \tag{3.11}$$

where $p_i = P(Z_i = 0) = a_i \frac{\lambda_i}{\lambda_0}$, $i = 1, 2, .., p$ and δ is the Dirac delta function given in Equation (2.2). The result from Carpenter et al. (2006) [12] is an adaptation of Corollary 3.6, and the multivariate form of the distribution of $\mathbf{Z}$ is given in Chapter 5.

Based on the multivariate form of $\mathbf{Z}$, we derive the form of the multivariate distribution of $\mathbf{X}$. We then avoid problems that many authors, say Fang and Zhang (1990) [24] and Walker and Stephens (1999) [79] for examples, run into when trying to characterize the multivariate gamma distribution. We have a close form expression, a feature that was difficult to derive from Henderson and Shimakura (2003) [34]. The expression is not simple, but it is tractable in contrast to what Mathai and Saxena (1978) [65] have suggested in terms of infinite series.

3.4 The Autoregressive Model

In this section, we focus on another approach to obtain a model similar to the direct linkage. This is an important extension to the multivariate distribution with correlation structure following the first order autoregressive scheme as stated for further research in the conclusion of Minhajuddin et al. (2003) [68]. In some applications, the parameter of interest depends on time. Time series could then be applied to the phenomena which is described linearly as

$$X_t = a_t X_{t-1} + Z_t \quad \text{for} \quad t \in N.$$

Brockwell and Davis (1996) [10] describe a similar model with some fixed $a_t = a$, $|a_t| < 1$ and the Z_t are iid normal with $EZ_t^2 = \sigma^2$. That gives that $E(X_t|\mathcal{F}_{t-1}) = aX_{t-1}$ where $\mathcal{F}_{t-1}$ is the set of measurable information up to time $t-1$.

Let X_t denote the observation at time t. The sequence of r.v.'s $(X_1, X_2, \cdots)$ is known as a time series. A graph of X_t against $t = 1, 2, \cdots$ is a time plot, and can reveal trends (upward, downward, variant) or seasonal variation. A model for a time series is described by the joint distributions of the r.v.'s $(X_1, X_2, \cdots, X_n)$, which is $P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n)$ for every integer $n = 1, 2, \ldots$

However, in practice, only partial information (such as mean, variance) may be available. Also, we cannot write the joint density as the product of the marginal densities since independence between the X_i's is not a workable assumption in our case. The time series model appears in the class of self decomposable distributions.

Definition 3.10 *A r.v.* X *is called self decomposable if for every* $0 < a < 1$, *there exists a r.v.* $Z = Z(a)$ *independent of* X *such that* $X \stackrel{d}{=} aX + Z$.

Some desirable properties follow:

- Self decomposable r.v.'s are contained in the class of infinitely divisible r.v's. See Feller (1971) [25].
- The time series version is $X_t = aX_{t-1} + Z_t$ where Z_t are independent of each other of of $X_s, \quad s < t$.
- If $a \to 0$, then X_t are independent. If $a \to 1$, then X_t are perfectly dependent.

Joe (1997) [44] chapter 8, discusses some special cases.

- If X is exponential with mean 1, then Z is a mixture of a point mass with weight a and the exponential with mean 1.
- If X is an Erlang distribution $Ga(\theta, 1)$, $\theta \in \mathbb{N}$, then Z is again a mixture $Z = \begin{cases} 0, & \text{with prob. } a^\theta; \\ Erlang \quad Ga(\text{j}, 1), & \text{with prob. } p_j \end{cases}$
  where p_j, $1 \leq j \leq \theta$, are prob. from binomial $(\theta, 1-a)$.
- If X is $Ga(\theta, 1)$, $\theta \in \mathbb{R} - \mathbb{N}$, then $\epsilon_t = \begin{cases} 0, & \text{with prob. } a^\theta; \\ Ga(\text{j,a}), & \text{with prob. } p_j, \end{cases}$
  where p_j, $j \geq 1$, are prob. from negative binomial (θ, a), $\quad p_j = \frac{\Gamma(\theta+j)}{j!\Gamma(\theta)} a^\theta (1-a)^j$.

The class of linear relations is quite large, and may contain unobservable factors. Among the alternatives, we consider the compound autoregressive model given by $X = (X_t)_{t \geq 0}$.

Since the beginning of time series analysis, one popular model assume that the variable X_t of the process is related to its past development and on a random disturbance sequence

Z_n which is not related to the past. The process is called a Markov process if the densities of X_t are normally distributed.

Definition 3.11 *The process X is called a compound autoregressive process of order p, denoted CAR(p) if and only if the conditional distribution of X_t given X_{t-1} admits the conditional Laplace form:*

$$E[e^{-sX_t}|X_{t-1}] = e^{-a_1X_{t-1}-a_2X_{t-2}-\cdots-a_pX_{t-p}+b}$$

where $a_i = a_i(s)$ $i = 1,\ldots,p$ and $b = b(s)$ are functions of the admissible values s in $\mathbb{R}$.

Our goal is to represent the data based on our probability model (1.2) and on this family of model (1.1). Then it becomes possible to estimate parameters, check for goodness of fit, and use our understanding of the mechanism to generate adjustments.

We assume that $X_t = aX_{t-1} + Z_t, t = 1, 2, \ldots$ where Z_t are independent, and also each Z_t is independent of $X_{t'}$ for each $t' < t$.

If $Z = \{Z_t\}_t$ were normally distributed with mean 0 and variance σ^2, then the model would be an autoregressive model, denoted AR(1), and the Z_t are commonly referred to as white noise. The process would then be stationary linear if $|a| < 1$.

Here, we have that:

$$\begin{aligned}
X_t &= aX_{t-1} + Z_t = a(aX_{t-2} + Z_{t-1}) + Z_t = a^2X_{t-2} + aZ_{t-1} + Z_t \\
&= \vdots \\
&= a^tX_0 + a^{t-1}Z_1 + a^{t-2}Z_2 + \cdots + aZ_{t-1} + Z_t \\
&= a^tX_0 + \sum_{k=0}^{t-1} a^kZ_{t-k}
\end{aligned}$$

Therefore $\mu_t = EX_t = a^tEX_0 + \sum_{k=0}^{t-1} a^kEZ_{t-k}$

and $Var(X_t) = a^{2t}Var(X_0) + \sum_{k=0}^{t-1} a^{2k}Var(Z_{t-k})$.

Also, it follows that:

$$\begin{aligned} Cov(X_{t+k}, X_t) &= Cov(aX_{t+k-1} + Z_t, X_t) = aCov(X_{t+k-1}, X_t) \\ &= a^2Cov(X_{t+k-2}, X_t) = \cdots = a^kCov(X_t, X_t) \\ &= a^kVar(X_t) \end{aligned}$$

The dependence decreases with the lag. That is, as k increases, (X_{t+k}, X_t) has less dependence. We would like the model to be stationary. The process $X = (X_n)_{(n=0,1,2,\cdots)}$ and its k shift, $\theta_k X = (X_{n+k})_{n=0,1,2,\cdots}$, share similar similar properties, and we write: $X \stackrel{d}{=} \theta_k X$.

We say that X is weakly stationary if it satisfies the following two properties:

1. the mean of X_t does not depend on t
2. for each $k = 1, 2, \cdots$, $Cov(X_{t+k}, X_t)$ does not depend on t.

This weakly stationarity property is achieved only if $a = 1$, and $EZ_t = 0$.

Rather than considering only nonnegative values of t, one could consider a shifted sequence on $\mathbb{Z}$.

Assume that $X_0 = aX_{-1} + Z_0$ and that $X_{-1} = aX_{-2} + Z_{-1}$ and so on... It follows that:

$$\begin{aligned} X_t &= a^tX_0 + \sum_{k=0}^{t-1} a^kZ_{t-k} = a^{t+1}X_{-1} + \sum_{k=0}^{t} a^kZ_{t-k} \\ &= \vdots \\ &= \sum_{k=0}^{\infty} a^kZ_{t-k}. \end{aligned}$$

Then $X_t = \sum_{k=0}^{\infty} a^k Z_{t-k}$.

For the series X to be stationary, i.e its mean and covariance independent of t, the required condition is that of absolute summability of the coefficients $\{a^k\}_{k=0,1,2,...}$ i.e. $\sum_{j=0}^{\infty} |a^k| = \sum_{j=0}^{\infty} a^k < \infty$. This is achieved if and only if $|a| = a < 1$.

Let us denote the coefficients a^k as ϕ_k. Then $X_t = \sum_{j=0}^{\infty} \phi_k Z_{t-k}$.

Moreover, if we can express Z_t in terms of X_t's, then we say that the process is invertible. This is achieved, and we can express $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$, where $\pi_0 = 1,\ \pi_1 = -a,\ and\ \pi_j = 0\ for\ j > 1$.

The compound autoregressive process defined earlier in terms of the Laplace transform with autoregressive path dependence can be expressed as:

$$E[e^{-sX_t}|X_{t-1}] = e^{-asX_{t-1}-sZ_t}$$

and the log of the conditional Laplace transform is a linear function of X_{t-1}.

3.5 Summary

In this chapter, we have generated a general class of multivariate gamma distributions. We call it generalized gamma distribution, not because the marginals are the generalized gamma distribution as defined in Kotz et al. (2000) [53], but because it generates two very distinct and broad subclasses of distributions:

- The continuous class model, where the latent variable is of the same gamma family. We study that class in Chapter 4. This class has no discontinuities in the joint densities.

- The fatal shock model, where all marginal distributions are gamma, but all the latent variables are gamma mixtures, and there are discontinuities due to the probability of simultaneous occurrence. In Chapter 5, we study that model under the exponential case, because of the wide use and applications of the exponential distributions.

Chapter 4

The Continuous Case Model

This chapter studies the continuous class of the model given in Chapter 3. The latent variables are all continuous, and of specific same family gamma type distributions. Our goal is to find the joint multivariate density of the latent variables Z_i's defined in (1.1), assuming that the conditions of Theorem 3.3 are satisfied, that is the Z_i's are independent gamma distributed for $i = 1, 2, .., p$. Nadarajah and Kotz (2005) [72] describe the distribution of the linear combinations of exponential and gamma type r.v.'s. Our approach is different as the mariginal distributions are specified, and the coefficients a_i in (1.1) are chosen such that the Z_i's are of the same family of gamma distributions, for $i = 1, 2, .., p$. Methods of estimations of parameters associated with the latent variables are also proposed.

4.1 Definition of Model

Mathai and Moschopoulos (1991) [63] considered a multivariate gamma model where the dependence structure is obtained by adding a factored common r.v. to every univariate marginal, exploiting the additive property. In their following paper [64], they consider an additive form of gamma distributions with common scale parameter. In Kotz et al. (2000) [53, page 468], a model in reliability applications is described as a system where a failing component is replaced by an identical one so that the process never stops until the total items in the system fail.

The Definition 3.1 is in the spirit of the Mathai and Moschopoulos (1991) [63] multivariate gamma with a flexible dependence structure for non-symmetric and nonnegative support. It provides a model fit for many situations. A slightly similar idea has been suggested by Hougaard (1986) [35], where the class of multivariate lifetime distributions had a dependence created by an unobserved quantity. We next recall the definitions and most important properties of the multivariate gamma distribution in the sense of Mathai and Moschopoulos (1991) [63]. The distribution of the sum with any shape and scale parameters is also analyzed.

Note that $\mathbf{X}$ in Definition 3.1 does not necessarily follow the multivariate gamma unless we pick appropriate values for the $a_i, \quad i = 1, 2, \cdots, p$, as it is proved in Mathai and Moschopoulos (1991) [63] and (1992) [64].

If $\alpha_0 \to 0$, then $X_0 \to 0$ $a.s.$, and then the multivariate gamma consists of n independent gamma distributions.

Carpenter et al. (2006) [12] marginally defined the X_i's as exponential, and showed that Z_i is then a product of a point mass and an exponential distribution. Despite the theoretical universality, the difference is in the concept. They start from a bivariate distribution, say X_1 and X_0, and derive the conditional distribution of the associated Z_1, whereas in most general situations, one describes the joint distribution of X_0 and X_1 with Z_1 known.

Mathai and Moschoupolos (1992) [64] showed that for $X_0 = Z_0, Z_1, Z_2, \cdots, Z_p$, mutually independent gamma r.v.'s with parameters such that $X_0 = Z_0 \sim Ga(\mu_0, \lambda_0, \alpha_0)$ and $Z_i \sim Ga\left(\mu_i - \frac{\alpha_0}{\lambda_0}\lambda_i, \lambda_i, \alpha_i - \alpha_0\right)$ for $i = 1, \cdots, p$, then X_i satisfying $X_i = a_i X_0 + Z_i$ for $a_i = \frac{\lambda_i}{\lambda_0}$, are also gamma distributed as $X_i \sim Ga(\mu_i, \lambda_i, \alpha_i)$. This result is now in this dissertation an application of Theorem 3.3. We next give some properties and the form of the multivariate density function.

4.2 Properties

Some important features, as in Section 4.1, follow from Equation (3.8) and Equation (3.9) of the Laplace transform $L_{\mathbf{X}}$.

$$L_{\boldsymbol{X}}(\boldsymbol{s}) \quad = \quad e^{-\mu_0 \sum_{i=1}^{p} a_i s_i} \left(\frac{\lambda_0}{\lambda_0 + \sum_{i=1}^{p} a_i s_i} \right)^{\alpha_0} \prod_{i=1}^{p} \frac{L_{X_i}(s_i)}{L_{X_0}(a_i s_i)}.$$

And then, it is sufficient to explain the nature of $\dfrac{L_{X_i}(s_i)}{L_{X_0}(a_i s_i)}$. We next present two important theorems of this chapter.

Theorem 4.1 *The joint pdf of the multivariate gamma* $\boldsymbol{X} = (X_1, X_2, \cdots, X_p)$ *based on Definition 3.1, when* $X_i \sim Ga(\mu_i, \lambda_i, \alpha_i)$, $i = 0, 1, \cdots, p$ *is explicitly given in terms of hypergeometric series (the Lauricella function of type B) in the special case where* $Z_i, 1 \leq i \leq p$, *are gamma type.*

Proof:

We follow the ideas in Mathai and Saxena (1978) [65].

Assume that $Z_i \sim Ga(\mu_i', \lambda_i', \alpha_i')$ for $i = 1, 2, \cdots, p$. Using the idea from Mathai and Moschopoulos (1991) [63], the joint density of $X_0, Z_1, \cdots, Z_p$ is given as:

$$\begin{aligned} f(x_0, z_1, \cdots, z_p) \quad &= \quad C(x_0 - \mu_0)^{\alpha_0 - 1} exp\{-\lambda_0(x_0 - \mu_0)\} \\ &\quad \times \prod_{i=1}^{p} (z_i - \mu_i')^{\alpha_i' - 1} exp\{-\lambda_i(z_i - \mu_i')\} \end{aligned}$$

where $C = \frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)} \prod_{i=1}^{p} \frac{\lambda_i'^{\alpha_i'}}{\Gamma(\alpha_i')}$

Even in the simplified case, integrating out x_0 is possible only in very special cases as it is noted in Krzanowski and Marriot (1994) [54].

Now putting $z_i = x_i - a_i x_0$, we have that the joint density of $(X_0, X_1, X_2, \cdots, X_p)$ is given by:

$$\begin{aligned} g(x_0, x_1, \cdots, x_p) &= C(x_0 - \mu_0)^{\alpha_0 - 1} exp\{-\lambda_0(x_0 - \mu_0)\} \\ &\quad \times \prod_{i=1}^{p} (x_i - a_i x_0 - \mu_i')^{\alpha_i' - 1} exp\{-\lambda_i'(x_i - a_i x_0 - \mu_i')\}. \end{aligned}$$

Set $u_0 = x_0 - \mu_0$ and $u_i + \mu_0 = \dfrac{1}{a_i}(x_i - \mu_i')$ for $1 \le i \le p$.

Then the density of can be written as:

$$\begin{aligned} g(u_0, u_1, \cdots, u_p) &= C \prod_{i=1}^{p} a_i^{\alpha_i' - 1} u_0^{\alpha_0 - 1} exp\{-\lambda_0 u_0)\} \\ &\quad \times \prod_{i=1}^{p} (u_i - u_0)^{\alpha_i' - 1} exp\{-\lambda_i' a_i (u_i - u_0), \} \end{aligned}$$

where $0 < u_0 < min\{u_1, \cdots, u_p\}$

The joint density of $(u_1, \cdots, u_p)$ is available by integrating out u_0. Hence

$$\begin{aligned} g(u_1, \cdots, u_p) &= C_1 \int_0^{min\{u_1, \cdots, u_p\}} u_0^{\alpha_0 - 1} \exp\{-\lambda_0 u_0)\}(u_1 - u_0)^{\alpha_1' - 1} \cdots (u_p - u_0)^{\alpha_p' - 1} \\ &\quad \times \exp\{-[\lambda_1' a_1 (u_1 - u_0) + \cdots + \lambda_p' a_p (u_p - u_0)]\} du_0 \end{aligned}$$

where $C_1 = C \prod_{i=1}^{p} a_i^{\alpha_i' - 1}$.

It follows that the density depends on the different forms of the $k!$ possible orderings of $u_1, u_2, \cdots, u_p$. For example $u_1 < u_2 < \cdots < u_p$. We have

$$\begin{aligned} g(u_1, \cdots, u_p) &= C_1\{\prod_{i=1}^{p} u_i^{\alpha_i'-1}\} \int_0^{u_1} u_0^{\alpha_0-1} \exp\{-\lambda_0 u_0)\}(1-\frac{u_0}{u_1})^{\alpha_1-1} \cdots (1-\frac{u_0}{u_p})^{\alpha_p-1} \\ &\quad \times \exp\{-[\lambda_1' a_1 u_1(1-\frac{u_0}{u_1}) + \cdots + \lambda_p' a_p u_p(1-\frac{u_0}{u_p})]\}du_0. \end{aligned}$$

Set $y = \dfrac{u_0}{u_1}$. Then the above is expressed as:

$$\begin{aligned} g(u_1, \cdots, u_p) &= C_1\{\prod_{i=1}^{p} u_i^{\alpha_i'-1}\} u_1^{\alpha_0} \int_0^1 y^{\alpha_0-1} exp\{-\lambda_0 u_1 y)\} \\ &= \times(1-y)^{\alpha_1'-1} \times (1-\frac{u_1}{u_2}y)^{\alpha_2'-1} \cdots (1-\frac{u_1}{u_p}y)^{\alpha_p'-1} \\ &= \times \exp\{-[\lambda_1' a_1 u_1(1-y) + \lambda_2' a_2 u_2(1-\frac{u_1}{u_2}y) \cdots + \lambda_p' a_p u_p(1-\frac{u_1}{u_p}y)]\}dy. \end{aligned}$$

Expanding the exponentials in series form, we obtain:

$$\begin{aligned} g(u_1, \cdots, u_p) &= C_1\{\prod_{i=1}^{p} u_i^{\alpha_i'-1}\} u_1^{\alpha_0} \\ &\quad \times \sum_{r_0}^{\infty}\sum_{r_1}^{\infty}\cdots\sum_{r_p}^{\infty} \frac{(-\lambda_0 u_1)^{r_0}}{r_0!}\frac{(-\lambda_1' a_1 u_1)^{r_1}}{r_1!}\cdots\frac{(-\lambda_p' a_p u_p)^{r_p}}{r_p!} \\ &\quad \times \int_0^1 y^{\alpha_0+r_0-1}(1-y)^{\alpha_1'+r_1-1} \times (1-\frac{u_1}{u_2}y)^{\alpha_2'+r_2-1} \cdots (1-\frac{u_1}{u_p}y)^{\alpha_p'+r_p-1}dy \\ &= C_1\{\prod_{i=1}^{p} u_i^{\alpha_i'-1}\} u_1^{\alpha_0} \times \frac{\Gamma(\alpha_0+r_0)\Gamma(\alpha_1'+r_1)}{\Gamma(\alpha_0+r_0+\alpha_1'+r_1)} \\ &\quad \times \sum_{r_0}^{\infty}\sum_{r_1}^{\infty}\cdots\sum_{r_p}^{\infty} \frac{(-\lambda_0 u_1)^{r_0}}{r_0!}\frac{(-\lambda_1' a_1 u_1)^{r_1}}{r_1!}\cdots\frac{(-\lambda_p' a_p u_p)^{r_p}}{r_p!} \\ &\quad \times F_D(\alpha_0+r_0; \alpha_2'+r_2, \cdots, \alpha_p'+r_p; \alpha_0+r_0+\alpha_1'+r_1; \frac{u_1}{u_2}, \cdots, \frac{u_1}{u_p}), \end{aligned}$$

where F_D is sometimes referred to as the Lauricella function of n variables.

The theorem is proved when substituting back x_i in terms of u_i.

Recall that Lauricella F_D, or more precisely $F_D^{(n)}$, is a function of n variables and n+2 parameters defined by the power series as:

$$F_D^{(n)}(a; b_1, \cdots, b_n; c; x_1, \cdots, x_n) = \sum_{m_1=0}^{\infty} \cdots \sum_{m_n=0}^{\infty} \frac{(a)_{m_1+\cdots+m_2}(b_1)_{m_1} \cdots (b_n)_{m_n}}{(c)_{m_1+\cdots+m_n} m_1! \cdots m_n!} x_1^{m_1!} \cdots x_n^{m_n}.$$

$\square$

We have a characterization in terms of an infinite sum of distributions. The question of how to generate variates from this distribution is then raised. Also relevant is how to implement a fit for this distribution. To answer these questions, a further study is then needed. We want to obtain another representation although we can tell by now that there is no unique way of expressing the joint density of a set of correlated gamma marginals.

Note from the above theorem, relations between means and variances of X_i and Z_i can be obtained. In particular,

$$\mu_i' = \mu_i - a_i \mu_0 \quad \text{and} \quad \frac{\alpha_i'}{\lambda_i'} = \frac{\alpha_i}{\lambda_i} - a_i \frac{\alpha_0}{\lambda_0},$$

$$\frac{\alpha_i'}{\lambda_i'^2} = \frac{\alpha_i}{\lambda_i^2} - a_i^2 \frac{\alpha_0}{\lambda_0^2},$$

Hence

$$\lambda_i' = \frac{\frac{\alpha_i}{\lambda_i} - a_i \frac{\alpha_0}{\lambda_0}}{\frac{\alpha_i}{\lambda_i^2} - a_i^2 \frac{\alpha_0}{\lambda_0^2}}$$

and

$$\alpha_i' = \frac{(\frac{\alpha_i}{\lambda_i} - a_i \frac{\alpha_0}{\lambda_0})^2}{\frac{\alpha_i}{\lambda_i^2} - a_i^2 \frac{\alpha_0}{\lambda_0^2}}.$$

However, in order to obtain a better parameter estimates, a more tractable version of the joint density is desired.

The notion of *frailty* is used as a measure of general random effect. Vaupel et al. (1979) [78] gave a description of frailty as a measure of susceptibility to all causes of death in heterogeneous populations, with applications assuming gamma distributions, based on the advantages of the gamma in the simple forms of its density in Equation (1.2) and its LST in Example 2.2.

In the particular case where $X_1 \sim Ga(\mu_1 = 0, \lambda_1, \alpha_1 = 1)$ and $X_2 \sim Ga(\mu_2 = 0, \lambda_2, \alpha_2 = \frac{1}{2})$, the density of the sum $X = X_1 + X_2$ can be expressed as:

$$f_X(x) = \sqrt{\lambda_1 \lambda_2} e^{-\frac{(\lambda_1+\lambda_2)}{2}x} I_0[\frac{(\lambda_2 - \lambda_1)}{2}x],$$

where I_0 is the Bessel function of first kind and zero order

$$I_0(t) = 1 + \frac{t^2}{2^2} + \frac{t^4}{2^2 \cdot 4^2} + \frac{t^6}{2^2 \cdot 4^2 \cdot 6^2} + \cdots .$$

The result of the convolution of n independent gammas, is given by Moschopoulos (1985) [70], and is as follows:

$$f_X(x) = \frac{\prod_{i=1}^n \lambda_i^{\alpha_i}}{\Gamma(\sum_{i=1}^n \alpha_i)} x^{\sum_{i=1}^{n-1} \alpha_i - 1} e^{-\lambda_n x} \sum_{k=0}^{\infty} \frac{b_n(k)(\sum_{i=1}^{n-1} \alpha_i)}{k!(\sum_{i=1}^n \alpha_i)_k} [(\lambda_n - \lambda_{n-1})x]^k$$

where the coefficients b_i are obtained recursively by:

$$b_i = \begin{cases} 1, & i = 2; \\ \sum_{j=0}^k \frac{b_{i-1}(\sum_{p=1}^{n-2} \alpha_p)_j (-k)_j}{j!(\sum_{p=1}^{i-1} \alpha_p)_j} C_i^j, & i = 3, 4, \cdots, n, \end{cases} \quad \text{with} \quad C_i = \frac{\lambda_{i-2} - \lambda_{i-1}}{\lambda_i - \lambda_{i-1}}.$$

Moschopoulos (1985) [70], and Furman and Landsman (2005) [28], give an interpretation of the distribution of X as a sum of gamma distributions the location being at the origin.

Theorem 4.2 *The distribution of the sum X is a mixed gamma with mixing shape parameter given by $\alpha+K$ and scale parameter $\lambda_{max} = \max\{\lambda_1, \cdots, \lambda_p\}$ where $\alpha = \sum_{i=1}^{p} \alpha_i$ and K is a non negative integer r.v. with distribution*

$$p_k = f(k) = Cd_k, \quad k = 0, 1, 2, ...$$

$$C = \prod_{i=1}^{p} (\frac{\lambda_i}{\lambda_{max}})^{\alpha_i} \quad \textit{and} \quad d_0 = 1, \quad d_k = \frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{p} \alpha_j (1 - \frac{\lambda_j}{\lambda_{max}})^i d_{k-i}$$

$$i.e \quad f(x) = \sum_{k=0}^{\infty} p_k f_{Y_k}(x), \quad Y_k \sim Ga(\lambda_{max}, \alpha + k)$$

Proof: See Furman (2005) [28]. □

In the case of the sum of two gamma $X_1 \sim Ga(\lambda_1, \alpha_1)$ and $X_2 \sim Ga(\lambda_2, \alpha_2)$, the distribution of $X = X_1 + X_2$ is given as:

$$f(x) = \sum_{k=0}^{\infty} p_k f_{Y_k}(x), \quad p_k = Cd_k, \quad C = \frac{\lambda_1^{\alpha_1} \lambda_2^{\alpha_2}}{\lambda_{max}^{\alpha_1+\alpha_2}}$$

$$d_0 = 1, \quad d_1 = \alpha_1(1 - \frac{\lambda_1}{\lambda_{max}}) + \alpha_2(1 - \frac{\lambda_2}{\lambda_{max}})$$

$$d_2 = \frac{d_1}{2}[\alpha_1(1 - \frac{\lambda_1}{\lambda_{max}}) + \alpha_2(1 - \frac{\lambda_2}{\lambda_{max}})] + \frac{d_0}{2}[\alpha_1(1 - \frac{\lambda_1}{\lambda_{max}})^2 + \alpha_2(1 - \frac{\lambda_2}{\lambda_{max}})^2]$$

$$d_k = \frac{1}{k} \sum_{i=1}^{k} \sum_{j=1}^{p} \alpha_j (1 - \frac{\lambda_j}{\lambda_{max}})^i d_{k-i}, \quad \text{for} \quad k \geq 3.$$

$$Y_k \sim Ga(\lambda_{max}, \alpha_1 + \alpha_2 + k).$$

The form of the exact distribution Theorem 4.1, and in Theorem 4.2, in an infinite series, is quite complicated in practical situations. Because the sum involves an infinite number of distributions, there is a need in developing a mathematically tractable theory and result. We consider some results in the gamma cases, and explain the case of the exponential distribution.

4.3 Estimation Procedures

Because of the complexity of the likelihood, no MLE can be done in the usual way in estimating model parameters. Alternatives for the MLE have been suggested by many authors such as Cohen and Whitten (1986) [16] and Cohen and Whitten (1988) [17]. They used method called modified moments estimators or modified MLE. Our approach builds the probability criteria from the EM algorithm. It can be used to compute the MLE for incomplete data. Disregarding the missing data is usually not safe. After separating the data into missing and non missing, we start by taking a guess of parameters within a corresponding domain. The idea is to replace the missing value by its expectation or predicted score given the starting guessed parameter values. We use an expected value of the likelihood function, computed using sufficient statistics, in turn computed from the data, and maximize the likelihood function to obtain new parameter estimates. From the famous paper by Dempster et al. (1977) [22], the basic EM theorem states that improving the expected log likelihood leads to an increase in the likelihood itself $l(\theta|\boldsymbol{X}) = \log L(\theta|\boldsymbol{X})$. The data $\mathbf{X} = (\mathbf{X}_{obs}, \mathbf{X}_{mis})$ has 2 parts: an observed part denoted as $\mathbf{X}_{obs}$ and a missing part we denote as $\mathbf{X}_{mis}$ and is assumed to follow a gamma distribution with unknown parameter θ, i.e. its pdf is expressed as in (1.2).

The idea is to fill in the missing values and iterate. Each iteration is done in two steps: the expectation (or E step) and the maximization (or M step).

At the expectation or E step, we find the conditional expectation of the log likelihood function for the complete data denoted here as $\mathbf{X}$ given the observed data say $\mathbf{X}_{obs}$, i.e.

$$Q(\theta|\theta^{(k)}) = E\bigg[\log f(\mathbf{X}|\theta)|\theta^{(k)}, \mathbf{X}_{obs}\bigg].$$

The next step of the EM algorithm, the M-step, is to maximize the Q function; i.e. we find the parameter $\theta = \theta^{k+1}$ maximizing $Q(\theta|\theta^{(k)})$. The EM algorithm guarantees that the likelihood function is increased by increase of the function Q.

The EM algorithm has a generalization called generalized EM (GEM), where in the M step, one finds only a value θ such that $Q(\theta|\theta^{(k)})$ is strictly increasing in k.

The density in our example is of multivariate type. Using ideas from Mathai and Moschopoulos (1991) [63], the likelihood function can be thought as a sum of p terms. One idea is to maximize each term separately.

The data vectors $\mathbf{X}_j = (x_{1j}, x_{2j}, \cdots, x_{pj}) \in \mathbb{R}^p$ for $j = 1, \cdots, n$ are observed and what are missing (unobserved) is the group of units $\mathbf{X}_0 = (x_{01}, x_{02}, \cdots, x_{0n})$. We wish to calculate the ML estimates of θ.

There are $pn + n = n(p+1)$ observations, and among them pn are observed and n are missing.

Then setting $\theta = (\mu_0, \lambda_0, \alpha_0, \mu_1, \lambda_1, \alpha_1, \cdots, \mu_p, \lambda_p, \alpha_p)'$, a vector of $3(p+1)$ parameters, the log likelihood is

$$
\begin{aligned}
l(\theta|\boldsymbol{X}) &= \log \prod_{j=1}^{n} g(x_{0j}, x_{1j}, \cdots, x_{pj}|\theta) = \sum_{j=1}^{n} \log\left[g(x_{0j}, x_{1j}, \cdots, x_{pj})|\theta\right] \\
&= n \log C + \sum_{j=1}^{n} \log\left[(x_{0j} - \mu_0)^{\alpha_0 - 1} exp\{-\lambda_0(x_{0j} - \mu_0)\}\right. \\
&\quad \left. \times \prod_{i=1}^{p} (x_{ij} - a_i x_{0j} - \mu_i')^{\alpha_i' - 1} exp\{-\lambda_i'(x_{ij} - a_i x_{0j} - \mu_i')\}\right] \\
&= n \log C + \sum_{j=1}^{n} \left[(\alpha_0 - 1)\log(x_{0j} - \mu_0) - \lambda_0(x_{0j} - \mu_0)\right. \\
&\quad \left. + \log\{\prod_{i=1}^{p} (x_{ij} - a_i x_{0j} - \mu_i')\} + \log\{\prod_{i=1}^{p} e^{-\lambda_i'(x_{ij} - a_i x_{0j} - \mu_i')}\}\right] \\
&= n \log C + (\alpha_0 - 1) \sum_{j=1}^{n} \log(x_{0j} - \mu_0) - \lambda_0 \sum_{j=1}^{n} (x_{0j} - \mu_0) \\
&\quad + \sum_{j=1}^{n} \sum_{i=1}^{p} (\alpha_i' - 1)\log(x_{ij} - a_i x_{0j} - \mu_i') - \sum_{j=1}^{n} \sum_{i=1}^{p} \lambda_i'(x_{ij} - a_i x_{0j} - \mu_i'),
\end{aligned}
$$

where $\quad C = \prod_{i=0}^{p} \frac{\lambda_i^{\alpha_i}}{\Gamma(\alpha_i)} \quad$ and $\quad \log C = \sum_{i=0}^{p} \Big\{ \alpha_i \log(\lambda_i) - \log(\Gamma(\alpha_i)) \Big\}.$

The log likelihood is not linear in the data. The sample and arithmetic means are jointly sufficient and complete statistics for the parameters of the data. See Casella and Berger (1990) [13]. They are functions of:

$$\sum_{j=1}^{n} x_{ij}, \quad \text{and} \quad \sum_{j=1}^{n} x_{ij} x_{i'j}, \quad i, i' = 1, \cdots, p.$$

We wish to calculate

$$E[x_{0j}|x_{1j}, x_{2j}, \cdots, x_{pj}], \quad j = 1, \cdots, n$$

$$E_{X_{0j}}[\log(x_{0j} - \mu_0)], \quad j = 1, 2, \cdots, n$$

$$E_{X_0}[x_{ij} - a_i x_{0j} - \mu_i], \quad i = 1, 2, \cdots, p, \quad j = 1, 2, \cdots, n$$

$$E_{X_0}[\log\{x_{ij} - a_i x_{0j} - \mu_i\}], \quad i = 1, 2, \cdots, p, \quad j = 1, 2, \cdots, n$$

$$Cov(x_{0j}|x_{1j}, x_{2j}, \cdots, x_{pj}; \hat{\mu}, \hat{\Sigma}), \quad j = 1, \cdots, n$$

Using the same idea as the Bayes theorem, which states that

$$f(x_0|\mathbf{x}) = \frac{f(x_0, \mathbf{x})}{f(\mathbf{x})} = \frac{f(x_0) f(\mathbf{x}|x_0)}{f(\mathbf{x})},$$

we have that the conditional density of x_{0j} given $\mathbf{x} = (x_{1j}, x_{2j}, \cdots, x_{pj})'$ for $j = 1, 2, \cdots, n$, is given by:

$$f(x_{0j}|x_{1j}, x_{2j}, \cdots, x_{pj}) = \frac{(x_{0j} - \mu_0)^{\alpha_0 - 1} e^{-\lambda_0 (x_{0j} - \mu_0)} G(x_{0j}, \mathbf{x})}{\int_{\mu_0}^{x_j^{min}} (x_{0j} - \mu_0)^{\alpha_0 - 1} e^{-\lambda_0 (x_{0j} - \mu_0)} G(x_{0j}, \mathbf{x}) dx_{0j}}$$

where

- the denominator represents the joint distribution of the j^{th} response vector,
- $x_j^{min} = \min_{i=1,2,\cdots,p}\{\frac{x_{ij}}{a_i}\}$,

- $G(x_{0j}, \mathbf{x}) = \prod_{i=1}^{p} (x_{ij} - a_i x_{0j} - \mu_i')^{\alpha_i' - 1} e^{-\lambda_i'(x_{ij} - a_i x_{0j} - \mu_i')}$.

It is then not possible to just have the conditional density of the observed data. We cannot separate the component likelihood of the known parameter from the variance component parameter. This complexity of the conditional density of x_{0j} given $x_{1j}, x_{2j}, \cdots, x_{pj}$ for $j = 1, 2, \cdots, n$, (not available in closed or simple form), prevents us from using the EM algorithm. The problem is in the integration. To go around this problem, we could approximate the integral or we could replace x_{0j} by a reasonable estimate. Based on the following relations,

$$x_{ij} = a_i x_{0j} + z_{ij} \Longleftrightarrow x_{0j} = \frac{x_{ij} - z_{ij}}{a_i}, \quad \text{for} \quad i = 1, 2, \cdots, p, \quad \text{and} \quad j = 1, 2, \cdots, n.$$

Considering the minimum of the x_{0j} is a natural idea one could use, i.e.

$$x_{0j} = \min_{i=1,..,p} \left\{ \frac{x_{ij} - z_{ij}}{a_i} \right\}.$$

In other words, the x_{0j}'s are estimable based on the nature of the structure. And hence the parameters $\mu_0, \lambda_0, \alpha_0$ are then estimated.

Now replace z_{ij} by $E(z_{ij}) = \dfrac{\alpha_i}{\lambda_i} - a_i \dfrac{\alpha_0}{\lambda_0} + \mu_i - a_i \mu_0$. Then

$$x_{0j} = \min\nolimits_{i=1,..,p} \left\{ \frac{x_{ij} - \frac{\alpha_i}{\lambda_i} - \mu_i}{a_i} \right\} + \left(\frac{\alpha_0}{\lambda_0} + \mu_0 \right).$$

So

$$E(x_{0j}) = \min_{i=1,..,p} \left\{ \frac{x_{ij} - \frac{\alpha_i}{\lambda_i} - \mu_i}{a_i} \right\} + \left(\frac{\alpha_0}{\lambda_0} + \mu_0 \right).$$

$$E \log(x_{0j} - \mu_0) = \log \left[\min_{i=1,..,p} \left\{ \frac{x_{ij} - \frac{\alpha_i}{\lambda_i} - \mu_i}{a_i} \right\} + \left(\frac{\alpha_0}{\lambda_0} \right) \right].$$

$$E \log(x_{ij} - a_i x_{0j} - \mu_i) = \log \left[x_{ij} - a_i \min_{i=1,..,p} \left\{ \frac{x_{ij} - \frac{\alpha_i}{\lambda_i} - \mu_i}{a_i} \right\} + \left(\frac{\alpha_0}{\lambda_0} + \mu_0 \right) - \mu_i \right].$$

Hence,

$$
\begin{aligned}
Q(\theta|\theta^{(k-1)}) &= E\Big\{ l(\theta|\theta^{(k-1)}, \boldsymbol{x}) \Big\} \\
&= n\log C + (\alpha_0 - 1)\sum_{j=1}^{n} E\log(x_{0j} - \mu_0) - \lambda_0 \sum_{j=1}^{n} E(x_{0j} - \mu_0) \\
&\quad + \sum_{i=1}^{p}\sum_{j=1}^{n} (\alpha_i - 1) E\log(x_{ij} - a_i x_{0j} - \mu_i) - \sum_{i=1}^{p}\sum_{j=1}^{n} \lambda_i E(x_{ij} - a_i x_{0j} - \mu_i) \\
&= n\log C + (\alpha_0 - 1)\sum_{j=1}^{n} \log\left[\min_{i=1,..,p}\left\{\frac{x_{ij} - \frac{\alpha_i}{\lambda_i} - \mu_i}{a_i}\right\} + \frac{\alpha_0}{\lambda_0}\right] \\
&\quad - \lambda_0 \sum_{j=1}^{n}\left[\min_{i=1,..,p}\left\{\frac{x_{ij} - \frac{\alpha_i}{\lambda_i} - \mu_i}{a_i}\right\} + \frac{\alpha_0}{\lambda_0}\right] \\
&\quad + \sum_{i=1}^{p}\sum_{j=1}^{n} (\alpha_i - 1)\log\left[x_{ij} - a_i(\min_{i=1,..,p}\left\{\frac{x_{ij} - \frac{\alpha_i}{\lambda_i} - \mu_i}{a_i}\right\} + \left(\frac{\alpha_0}{\lambda_0} + \mu_0\right)\Big) - \mu_i\right] \\
&\quad - \sum_{i=1}^{p}\sum_{j=1}^{n} \lambda_i \left[x_{ij} - a_i\left(\min_{i=1,..,p}\left\{\frac{x_{ij} - \frac{\alpha_i}{\lambda_i} - \mu_i}{a_i}\right\} + \left(\frac{\alpha_0}{\lambda_0} + \mu_0\right)\right) - \mu_i\right],
\end{aligned}
$$

where $\theta^{(k-1)} = (\alpha_0^{(k-1)}, \lambda_0^{(k-1)}, \mu_0^{(k-1)}, \alpha_1^{(k-1)}, \lambda_1^{(k-1)}, \mu_1^{(k-1)}, \cdots, \alpha_p^{(k-1)}, \lambda_p^{(k-1)}, \mu_p^{(k-1)})'$.

Then using maximum likelihood, find

$\theta^{(k)} = (\mu_0^{(k)}, \lambda_0^{(k)}, \alpha_0^{(k)}, \mu_1^{(k)}, \lambda_1^{(k)}, \alpha_1^{(k)}, \cdots, \mu_p^{(k)}, \lambda_p^{(k)}, \alpha_p^{(k)})'$

that maximizes $Q(\theta|\theta^{(k-1)})$.

So in here, the parameters are computed iteratively. The following steps are suggested to get the MLE's of the parameters:

- step 1: update the unknown values x_{0j} for each iteration along with the parameters of interest.
- step 2: find estimates for μ_0, λ_0 and α_0.
- step 3: find estimates for μ_i, λ_i and α_i for $1 \leq i \leq p$.

Continue the process at to convergence, when there is no much difference gained in the next iteration. The vector of parameters $\theta = (\mu_0, \lambda_0, \alpha_0, \mu_1, \lambda_1, \alpha_1, \cdots, \mu_p, \lambda_p, \alpha_p)'$ has two components: $\theta = (\theta_1, \theta_2)'$ where $\theta_1 = (\mu_0, \lambda_0, \alpha_0)'$ and $\theta_2 = (\mu_1, \lambda_1, \alpha_1, \cdots, \mu_p, \lambda_p, \alpha_p)'$ is estimated by MLE for θ_1 fixed.

We continue the iterative process until some stopping criteria is reached. Although the process seems very similar to the EM algorithm, it is not based on the conditional distribution. It seems nevertheless very related to the EM algorithm. The estimation is performed by a generalization of the EM algorithm which Tanner (1996) [77] called the Expectation-Solution (ES) algorithm. For the ES algorithm, the M-step of the EM algorithm is replaced by a step requiring the solution of an estimating equation, which does not necessarily correspond to a maximization problem. The ES algorithm is a general algorithm for solving equations with incomplete data. These authors used this algorithm to fit a mixture model for independent and correlated data. The use of mixture of densities gives more flexibility to the model the distribution. With this approach, we have a method to calculate the variance-covariance matrix of the parameters.

The Markov Chain Monte Carlo (MCMC) allows the simulation of many complex and multivariate types of random data. It also deals with calculating expectations, finding conditional expectations for maximizing likelihood in missing data patterns, predicting the missing data based on a normal distribution. Although the assumption of normality is not always a negative aspect (specially when the missing information is not too large), this new estimation technique can provide better answers as the values x_0 are not missing at random.

To understand a statistical model, one idea would be to simulate many realizations from that model, and study it. For example, consider a r.v. X with pdf $f(x)$ and a function g such that $E(g(X)) = \int_X g(x)f(x)dx$ is difficult to integrate. The integral can be approximated after n realizations, $x_1, x_2, \cdots, x_n$, of X by the following sum:

$$\frac{1}{n}\sum_{i=1}^{n} g(x_i).$$

In our example, the joint density has the form $f(\mathbf{x}) = f(x_0, x_1, x_2, \cdots, x_p)$, and we are interested in some features of the marginal density. Assume we can sample the $p+1$ many univariate conditional densities:

$$\begin{gathered} f(x_0|x_1, x_2, \cdots, x_p) \\ f(x_1|x_0, x_2, \cdots, x_p) \\ f(x_2|x_0, x_1, x_3, \cdots, x_p) \\ \vdots \\ f(x_p|x_0, x_1, x_2, \cdots, x_{p-1}). \end{gathered}$$

Indeed

$$\begin{aligned} f(x_k|x_0, x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_p) &= \frac{f(x_0, x_1, \cdots, x_p)}{f(x_0, x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_p)} \\ &= \frac{\lambda_k^{\alpha_k}}{\Gamma(\alpha_k)}(x_k - a_k x_0)^{\alpha_k - 1} e^{-\lambda_k(x_k - a_i x_0)} \end{aligned}$$

which is a gamma distribution with location $a_k x_0$ and parameters α'_k and λ'_k denoted here as α_k and λ_k, respectively.

So choose arbitrarily p initial values: $x_1 = x_1^{(0)}, x_2 = x_2^{(0)}, \cdots, x_p = x_p^{(0)}$.

Then create

- $x_0^{(1)}$ by drawing from $f(x_0|x_1^{(0)}, \cdots, x_p^{(0)})$
- $x_1^{(1)}$ by drawing from $f(x_1|x_0^{(1)}, x_2^{(0)}, \cdots, x_p^{(0)})$

- $x_2^{(1)}$ by drawing from $f(x_2|x_0^{(1)}, x_1^{(1)}, x_3^{(0)}, \cdots, x_p^{(0)})$
- $\vdots$
- $x_p^{(1)}$ by drawing from $f(x_p|x_0^{(1)}, x_1^{(1)}, x_2^{(1)}, \cdots, x_{p-1}^{(1)})$

This is one Gibbs pass through the $p+1$ conditional densities that gives values $(x_0^{(1)}, x_1^{(1)}, x_2^{(1)}, \cdots, x_p^{(1)})$.

Thus, we can take the last n of x_0 values after many Gibbs passes and we set that:

$$E(X_0) = \frac{1}{n} \sum_{i=m}^{m+n} x_0^i.$$

The Monte Carlo EM algorithm is an important sampling technique to generate random variates to construct Monte Carlo approximations.

As said earlier, the identification of parameters is an important topic. We have proposed a novel procedure to identify the parameters which uses ideas of the EM algorithm, with a computation cost not so large.

Chapter 5

The Multivariate Exponential

The multivariate exponential plays an important role in survival and reliability analysis. Marshall and Olkin (1967) [61], Joe (1997), Ghosh and Gelfand (1998) [30], and Hougaard (2000) [36], to mention a few authors, gave several examples to motivate this research problem. In this chapter, we focus on the multivariate exponential distribution. More specifically, we consider the $p-$variate $X_1, X_2, \cdots,$ and X_p be fixed marginally as exponential r.v.'s with hazard rates $\lambda_1, \lambda_2, \cdots,$ and λ_p, respectively. Then by introducing two types of latent non-negative variables, X_0 and $Z_1, Z_2, \cdots,$ and Z_p, statistically independent between themselves and of X_0, a linear relationship is formed between X_0 and $X_1, X_2, \cdots, X_p$ as in (1.1). Note that this structure was introduced in Chapter 3, and is a special family within the generalized gamma distribution.

We focus on the exponential class of distributions because of its importance in the literature. This class is large, and includes the continuous (examined in Chapter 4) and discontinuous cases. We also study the bivariate case approach where X_0 is unobserved, and missing, and lay out the joint density function along with the joint survival function. We also developed estimators for the parameters associated with the model. Carpenter et al. (2006) [12] defined a similar approach at the univariate level. They characterize through Laplace transforms, the distribution of the latent variable in the exponential case as mixture of a point mass at zero and an exponential with hazard rate λ_i as shown in Corollary 3.6. Note that when $Z_i = 0,$ there is a positive probability that X_i is proportional to X_0 with proportionality constant a_i, i.e. $P(X_i = a_i X_0) > 0, \quad i = 1, \ldots, p.$

We show that our model possesses the property of conditional independence in Definition 1.7 and Equation (1.13), given a random latent effect. Hougaard (2000) [36] calls this random latent effect frailty. Frailty models are common in describing the dependence. See Vaupel et al. (1979) [78], Hougaard (2000) [36], and Henderson and Shimakura (2003) [34]. Hougaard (1986) [35] and Hougaard (2000) [36] consider the random latent effect as parameter, and we are treating it as latent variable. As described in Carpenter et al. (2005) [12], the joint distribution is not absolutely continuous. The multivariate survival data of the experiment gives multiple events and involves several members or components in a system.

The multivariate lifetime distributions by Hougaard (2000) [36], has a dependence created by an unobservable quantity. This is a typical scenario we adopt here also. Hougaard (1986) [35] proposed a continuous multivariate lifetime distribution where the marginal distributions are Weibull (continuous) and does not allow the property of simultaneous or proportional failures of individuals or components.

5.1 The Multivariate Exponential

In this section, we give our version of the $p-$variate exponential distribution through Definition 5.1. We study its various properties, and characterize its density function.

Definition 5.1 *Let $X_0, X_1, \cdots, X_p$ be exponential r.v.'s as in (1.3) with scale parameters $\lambda_i,\ \ 0 \le i \le p$. Let $Z_i,\ \ i = 1, .., p$, be independent r.v.'s satisfying (1.1). We define the joint distribution of $\boldsymbol{X} = (X_1, X_2, \cdots, X_p)$ as the $p-$variate exponential distribution.*

From Equation (3.7), the mean of $\mathbf{X}$ is given as

$$E(\mathbf{X}) = \left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots, \frac{1}{\lambda_p}\right)', \tag{5.1}$$

and the variance/covariance matrix is given as

$$\Sigma = \frac{1}{\lambda_0^2}\begin{pmatrix} \frac{\lambda_0^2}{\lambda_1^2} & a_1a_2 & \cdots & a_1a_p \\ a_2a_1 & \frac{\lambda_0^2}{\lambda_2^2} & \cdots & a_2a_p \\ \vdots & \vdots & \vdots & \vdots \\ a_pa_1 & \cdots & a_pa_{p-1} & \frac{\lambda_0^2}{\lambda_p^2} \end{pmatrix}. \tag{5.2}$$

From (5.2), the correlation matrix is given as

$$\rho = \frac{1}{\lambda_0^2}\begin{pmatrix} \lambda_0^2 & a_1a_2\lambda_1\lambda_2 & \cdots & a_1a_p\lambda_1\lambda_p \\ a_2a_1\lambda_2\lambda_1 & \lambda_0^2 & \cdots & a_2a_p\lambda_2\lambda_p \\ \vdots & \vdots & \vdots & \vdots \\ a_pa_1\lambda_p\lambda_1 & \cdots & a_pa_{p-1}\lambda_p\lambda_{p-1} & \lambda_0^2 \end{pmatrix}.$$

Note that since we consider the exponential case, the shape parameters $\alpha_i = 1$ in (1.2) for $i = 1, 2, .., p$. From Theorem 3.4 and Equation (1.1), the LST of Z_i is

$$L_{Z_i}(s) = p_i + (1 - p_i)L_{X_i}(s), \quad \textit{with} \quad p_i = a_i\frac{\lambda_i}{\lambda_0}, \quad i = 1, 2, .., p.$$

That is Z_i is a mixture of Bernoulli r.v. with probability p_i and exponential r.v. with parameter λ_i as in Example 2.4.

From Equation (3.11), the conditional survival function of X_i given X_0, is obtained as

$$\begin{aligned}
S(t|x_0) &= P(X_i > t|x_0) = P(a_i x_0 + Z_i > t|x_0) = P(Z_i > t - a_i x_0|x_0) \\
&= \int_{t-a_i x_0}^{\infty} f_{Z_i}(z)dz = \int_{t-a_i x_0}^{\infty} \left[p_i \delta(z) + (1-p_i)\lambda_i e^{-\lambda_i z}\right] dz \\
&= p_i \int_{t-a_i x_0}^{\infty} \delta(z)dz + (1-p_i) \int_{t-a_i x_0}^{\infty} e^{-\lambda_i z} dz \\
&= p_i(1 - H(t - a_i x_0)) + (1-p_i)e^{-\lambda_i(t-a_i x_0)}, \quad i = 1, .., p,
\end{aligned}$$

where H is the Heaviside function defined in Equation (2.3).

From the conditional independence property and (1.13), the joint conditional survival function is

$$S(x_1, \cdots, x_p|x_0) = \prod_{i=1}^{p} \left[p_i(1 - H(x_i - a_i x_0)) + (1-p_i)e^{-\lambda_i(x_i - a_i x_0)}\right]. \tag{5.3}$$

The distribution of the minimum lifetime distribution $X_{(1)} = min\left\{\frac{X_1}{a_1}, \frac{X_2}{a_2}, \cdots, \frac{X_p}{a_p}\right\}$ can be derived directly from (5.3) and from properties of the Heaviside function in Equation (2.3). It is given as

$$\begin{aligned}
P(X_{(1)} > t|x_0) &= \prod_{i=1}^{p} P\left(\frac{X_i}{a_i} > t|x_0\right) = \prod_{i=1}^{p} P(X_i > a_i t|x_0) \\
&= \prod_{i=1}^{p} \left[p_i(1 - H(x_i - x_0)) + (1-p_i)e^{-\lambda_i a_i(x_i - x_0)}\right].
\end{aligned}$$

Note that, as we should expect,

$$S(0, 0, \cdots, 0) = 1 \quad \text{and} \quad S(\infty, \infty, \cdots, \infty) = 0.$$

The Equation (5.3) above is quite interesting and could be deduced intuitively from the nature of the conditional distribution. It also captures the method of copula that many authors have suggested. According to that method, the joint survival function of $X_1, \cdots, X_p$ is represented as:

$$S(x_1, x_2, \cdots, x_p) = C(P(X_1 > x_1), P(X_2 > x_2), \cdots, P(X_p > x_p))$$

$$= C(S_{X_1}(x_1), S_{X_2}(x_2), \cdots, S_{X_p}(x_p)),$$

where $C(u_1, u_2, \cdots, u_p)$ is a copula: a function mixing the univariate survival function u_1, u_2 and u_p. For example, Hougaard (2001) [37] proposed the following copula function C at the bivariate level:

$$C(u_1, u_2) = \exp\left[-\{(-\ln u_1)^\beta + (-\ln u_2)^\beta\}^{1/\beta}\right]$$

with $\beta \geq 1$ coefficient of association. The case $\beta = 1$ corresponds to the independence between u_1 and u_2.

There are some problems associated with the form of the copula, as it can lead to misrepresentation of the association. The false estimation of β could then lead to false estimates for the parameters of interest.

Taking the derivative of the i^{th} survival function in Equation (5.3), and using Equation (2.3), the conditional density is given by

$$f_{X_i|X_0}(t) \quad = \quad p_i\delta(t - a_i x_0) + (1 - p_i)\lambda_i e^{-\lambda_i(t - a_i x_0)}, \quad i = 1, .., p. \tag{5.4}$$

and from (5.4), the conditional expectation is

$$\begin{aligned}
E_{X_i|X_0}(X_i) &= \int_0^\infty t f_{X_i|X_0}(t)dt \\
&= p_i \int_0^\infty t\delta(t - a_i x_0)dt + (1-p_i)\int_{a_i x_0}^\infty \lambda_i t e^{-\lambda_i(t-a_i x_0)}dt \\
&= p_i a_i x_0 + (1-p_i)(\frac{1}{\lambda_i} + a_i x_0) = a_i x_0 + (1-p_i)\frac{1}{\lambda_i}, \quad i = 1, .., p.
\end{aligned}$$

Note that taking the expectation of the above with respect to X_0 gives

$$\begin{aligned}
EX_i &= E_{X_0}E_{X_i|X_0}(X_i) = E_{X_0}[p_i a_i x_0 + (1-p_i)(\frac{1}{\lambda_i} + a_i x_0)] \\
&= p_i \frac{a_i}{\lambda_0} + (1-p_i)(\frac{1}{\lambda_i} + \frac{a_i}{\lambda_0}) \\
&= \frac{a_i}{\lambda_0} + (1-p_i)\frac{1}{\lambda_i} \\
&= \frac{1}{\lambda_i}, \quad i = 1, .., p,
\end{aligned}$$

since $p_i = a_i \frac{\lambda_i}{\lambda_0}$, confirming earlier results in Equation (5.1) and in Carpenter et al. (2006) [12].

Although the joint density of $(X_0, X_1, X_2, \cdots, X_p)'$ is easy to find (see Equation in proof of Theorem 1.8), the density of $(X_1, X_2, \cdots, X_p)'$ is not. However, we can study the density of $(X_0, X_1, X_2, \cdots, X_p)'$ through the latent variables $Z_1, Z_2, \cdots, Z_p$, with relative ease. Using the independence of the Z_i's, $i = 1, 2, .., p$, between each other and of X_0, and Theorem 1.8, we have that:

$$f(x_0, x_1, \cdots, x_p) = \lambda_0 e^{-\lambda_0 x_0} \prod_{i=1}^{p} \left[p_i \delta(x_i - a_i x_0) - (1-p_i)\lambda_i e^{-\lambda_i(x_i - a_i x_0)} I_{(x_i > a_i x_0)} \right], \quad (5.5)$$

To demonstrate the complicated form of the above density, we give the explicit form in the case of three r.v.'s $X_1 = a_1X_0 + Z_1$, $X_2 = a_2X_0 + Z_2$, and $X_3 = a_3X_0 + Z_3$. Hence, Equation (5.5) becomes:

$$
\begin{aligned}
f(x_0, x_1, x_2, x_3) &= f(x_0)f(x_1 - a_1x_0)f(x_2 - a_2x_0)f(x_3 - a_3x_0) \\
&= f(x_0)[p_1\delta(x_1 - a_1x_0) + (1 - p_1)f_{X_1}(x_1 - a_1x_0)I_{(x_1 > a_1x_0)}] \\
&\quad [p_2\delta(x_2 - a_2x_0) + (1 - p_2)f_{X_2}(x_2 - a_2x_0)I_{(x_2 > a_2x_0)}] \\
&\quad [p_3\delta(x_3 - a_3x_0) + (1 - p_3)f_{X_3}(x_3 - a_3x_0)I_{(x_3 > a_3x_0)}] \\
&= f(x_0)[p_1p_2p_3\delta(x_1 - a_1x_0)\delta(x_2 - a_2x_0)\delta(x_3 - a_3x_0) \\
&\quad +p_1(1 - p_2)p_3\delta(x_1 - a_1x_0)\delta(x_3 - a_3x_0)f_{X_2}(x_2 - a_2x_0) \\
&\quad +p_1p_2(1 - p_3)\delta(x_1 - a_1x_0)\delta(x_2 - a_2x_0)f_{X_3}(x_3 - a_3x_0) \\
&\quad +p_1(1 - p_2)(1 - p_3)\delta(x_1 - a_1x_0)f_{X_2}(x_2 - a_2x_0)f_{X_3}(x_3 - a_3x_0) \\
&\quad +(1 - p_1)p_2p_3\delta(x_2 - a_2x_0)\delta(x_3 - a_3x_0)f_{X_1}(x_1 - a_1x_0) \\
&\quad +(1 - p_1)(1 - p_2)p_3\delta(x_3 - a_3x_0)f_{X_1}(x_1 - a_1x_0)f_{X_2}(x_2 - a_2x_0) \\
&\quad +(1 - p_1)p_2(1 - p_3)\delta(x_2 - a_2x_0)f_{X_1}(x_1 - a_1x_0)f_{X_3}(x_3 - a_3x_0) \\
&\quad +(1 - p_1)(1 - p_2)(1 - p_3)f_{X_1}(x_1 - a_1x_0)f_{X_2}(x_2 - a_2x_0)f_{X_3}(x_3 - a_3x_0)].
\end{aligned}
$$

Coming back to the multivariate form obtained from Equation (5.5), there are 2^p terms in the sum expressed from the power set of $\mathcal{P}$, the set of subsets of $\{1, 2, \cdots, p\}$. Hence,

$$
\begin{aligned}
f(x_0, x_1, \cdots, x_p) = \sum_{|A|=k, A\in\mathcal{P}}^{p} \quad & p_{A_{(1)}} \cdots p_{A_{(k)}}\delta(x_{A_{(1)}} - a_{A_{(1)}}x_0) \cdots \delta(x_{A_{(k)}} - a_{A_{(k)}}x_0) \\
& \times f(x_0)(1 - p_{A^c_{(1)}}) \cdots (1 - p_{A^c_{(p-k)}})
\end{aligned}
$$

$$\times f_{A^c_{(1)}}(x_{A^c_{(1)}} - a_{A^c_{(1)}} x_0) \cdots f_{A^c_{(p-k)}}(x_{A^c_{(p-k)}} - a_{A^c_{(p-k)}} x_0),$$

where A is a subset of $\mathcal{P}$, with k elements, $0 \leq k \leq p$, denoted as $A_{(1)}, A_{(2)}, \cdots, A_{(k)}$, and we then can deduce the subset $A^c = \{1, 2, \ldots, p\} \setminus A$, whose elements are denoted as $A^c_{(1)}, A^c_{(2)}, \cdots, A^c_{(p-k)}$. Hence,

$$\begin{aligned} f(x_0, x_1, \cdots, x_p) = \sum_{|A|=k, A\in\mathcal{P}}^{p} \quad & p_{A_{(1)}} \cdots p_{A_{(k)}} \delta(x_{A_{(1)}} - a_{A_{(1)}} x_0) \cdots \delta(x_{A_{(k)}} - a_{A_{(k)}} x_0) \\ & \times (1 - p_{A^c_{(1)}}) \cdots (1 - p_{A^c_{(p-k)}}) \\ & \times \lambda_0 \lambda_{A^c_{(1)}} \cdots \lambda_{A^c_{(p-k)}} e^{-\lambda_{A^c_{(1)}} x_{A^c_{(1)}}} \cdots e^{-\lambda_{A^c_{(p-k)}} x_{A^c_{(p-k)}}} \\ & \times e^{-(\lambda_0 - a_{A^c_{(1)}} \lambda_{A^c_{(1)}} - \cdots - a_{A^c_{(p-k)}} \lambda_{A^c_{(p-k)}}) x_0}. \end{aligned}$$

Therefore the joint density of $(X_1, X_2, \cdots, X_p)'$ the multivariate exponential is obtained by integrating the above expression with respect to x_0, giving

$$\begin{aligned} f(x_1, \cdots, x_p) = \sum_{|A|=k, A\in\mathcal{P}}^{p-1} \quad & \frac{p_{A_{(1)}} \cdots p_{A_{(k)}}}{a_{A_{(1)}} \cdots a_{A_{(k)}}} \delta\Big(\frac{x_{A_{(1)}}}{a_{A_{(1)}}} - \frac{x_{A_{(k)}}}{a_{A_{(k)}}}\Big) \delta\Big(\frac{x_{A_{(k-1)}}}{a_{A_{(k-1)}}} - \frac{x_{A_{(k)}}}{a_{A_{(k)}}}\Big) \\ & \times (1 - p_{A^c_{(1)}}) \cdots (1 - p_{A^c_{(p-k)}}) \\ & \times \lambda_0 \lambda_{A^c_{(1)}} \cdots \lambda_{A^c_{(p-k)}} e^{-\lambda_{A^c_{(1)}} x_{A^c_{(1)}}} \cdots e^{-\lambda_{A^c_{(p-k)}} x_{A^c_{(p-k)}}} \\ & \times e^{-(\lambda_0 - a_{A^c_{(1)}} \lambda_{A^c_{(1)}} - \cdots - a_{A^c_{(p-k)}} \lambda_{A^c_{(p-k)}}) \frac{x_{A_{(1)}}}{a_{A_{(1)}}}} \\ + \quad & (1 - p_1) \cdots (1 - p_p) \frac{\lambda_0 \lambda_1 \cdots \lambda_p}{\lambda_0 - a_1 \lambda_1 - \cdots - a_p \lambda_p} \\ & \times e^{-\lambda_1 x_1} \cdots e^{-\lambda_p x_p} (1 - e^{-(\lambda_0 - a_1 \lambda_1 - \cdots - a_p \lambda_p) \varphi}), \end{aligned}$$

where $\varphi = min_i \left(\dfrac{x_i}{a_i} \right)$.

Similarly, to get the unconditional survival function, one would derive it from (5.3), and therefore compute

$$S(x_1, x_2, \cdots, x_p) = \int_0^\infty \prod_{i=1}^{p} S(x_i|x_0) f_{X_0}(x_0) dx_0$$

$$= \lambda_0 \int_0^\infty \prod_{i=1}^{p} \left[p_i(1 - H(x_i - a_i x_0)) + (1 - p_i) e^{-\lambda_i (x_i - a_i x_0)} \right] e^{-\lambda_0 x_0} dx_0.$$

As we can see, we cannot interchange integration and product in this above expression, and then there are substantial number of cases, 2^p precisely, to consider in order to find the survival function. We discuss the case of $p = 2$ in the next section and explain each of the components of the survival in that bivariate case. We also obtain estimates of the unknown parameters in that case.

5.2 The Bivariate Exponential Model and MLE

In this section, we study the bivariate case of the multivariate exponential distribution of the previous section. The bivariate exponential distribution from Definition 5.1 can be expressed as:

$$\begin{cases} X_1 = a_1 X_0 + Z_1 & ; \\ X_2 = a_2 X_0 + Z_2 & . \end{cases} \tag{5.6}$$

Then, from Carpenter et al. (2006) [12], the joint density of (X_0, X_1) is given by

$$f(x_0, x_1) \quad = \quad \begin{cases} p_1 f_{X_0}(x_0) \delta(x_1 = a_1 x_0) \\ (1 - p_1) f_{X_0}(x_0) f_{X_1}(x_1 - a_1 x_0) I_{(x_1 > a_1 x_0)} & , \end{cases}$$

$$= \begin{cases} p_1\lambda_0 e^{-\lambda_0 x_0}, & \text{if } x_0 = \frac{x_1}{a_1} \\ (1-p_1)\lambda_0\lambda_1 e^{-\lambda_0 x_0} e^{-\lambda_1(x_1 - a_1 x_0)}, & \text{if } x_0 < \frac{x_1}{a_1} \end{cases},$$

where $p_1 = a_1 \dfrac{\lambda_1}{\lambda_0}$. Similarly based on the expression $X_2 = a_2 X_0 + Z_2$, we have for $p_2 = a_2 \dfrac{\lambda_2}{\lambda_0}$,

$$\begin{aligned} f(x_0, x_2) &= \begin{cases} p_2 f_{X_0}(x_0)\delta(x_2 = a_2 x_0) \\ (1-p_2) f_{X_0}(x_0) f_{X_2}(x_2 - a_2 x_0) I_{(x_2 > a_2 x_0)} \end{cases}, \\ &= \begin{cases} p_2\lambda_0 e^{-\lambda_0 x_0}, & \text{if } x_0 = \frac{x_2}{a_2} \\ (1-p_2)\lambda_0\lambda_2 e^{-\lambda_0 x_0} e^{-\lambda_2(x_2 - a_2 x_0)}, & \text{if } x_0 < \frac{x_2}{a_2}. \end{cases} \end{aligned}$$

Hence, using the independence between X_0, Z_1, Z_2 and Theorem 1.8, the joint density of (X_0, X_1, X_2) related as in (5.6), is given from (5.5) by:

$$\begin{aligned} f(x_0, x_1, x_2) &= \lambda_0 e^{-\lambda_0 x_0} \Big[p_1 \delta(x_1 - a_1 x_0) + (1-p_1) f_{X_1}(x_1 - a_1 x_0) I_{(x_1 > a_1 x_0)} \Big] \\ &\quad \times \Big[p_2 \delta(x_2 - a_2 x_0) + (1-p_2) f_{X_2}(x_2 - a_2 x_0) I_{(x_2 > a_2 x_0)} \Big] \\ &= p_1 p_2 \lambda_0 e^{-\lambda_0 x_0} \delta_{(x_1 - a_1 x_0)} \delta_{(x_2 - a_2 x_0)} \\ &\quad + p_1(1-p_2)\lambda_0 e^{-\lambda_0 x_0} f_{X_2}(x_2 - a_2 x_0) \delta_{(x_1 - a_1 x_0)} \\ &\quad + (1-p_1) p_2 \lambda_0 e^{-\lambda_0 x_0} f_{X_1}(x_1 - a_1 x_0) \delta_{(x_2 - a_2 x_0)} \\ &\quad + (1-p_1)(1-p_2)\lambda_0 e^{-\lambda_0 x_0} f_{X_1}(x_1 - a_1 x_0) f_{X_2}(x_2 - a_2 x_0) I_{(x_1 > a_1 x_0, x_2 > a_2 x_0)}. \end{aligned}$$

The expression $f(x_0, x_1, x_2)$ is one way to obtain an estimate for x_0 or the parameter associated with it, λ_0. Let's assume that $\varphi = min\left(\dfrac{x_1}{a_1}, \dfrac{x_2}{a_2}\right) = \dfrac{x_1}{a_1} \le \dfrac{x_2}{a_2}$.

Also, set $r_i^{(1)} = I_{(z_{1i}=0)} = I_{(x_1=a_1x_0)} = \begin{cases} 1 & , \text{if } z_{1i} = 0 \\ 0 & , \text{if } z_{1i} \neq 0, \end{cases}$

and $r_i^{(2)} = I_{(z_{2i}=0)} = I_{(x_2=a_2x_0)} = \begin{cases} 1 & , \text{if } z_{2i} = 0 \\ 0 & , \text{if } z_{2i} \neq 0. \end{cases}$

Then the full likelihood function based on a random sample of size n is the product of n contributed likelihoods and is given as:

$$
\begin{aligned}
L(\lambda_0, \lambda_1, \lambda_2) &= \prod_{i=1}^{n} \left[p_1 p_2 \lambda_0 e^{-\lambda_0 x_{0i}} \right]^{r_i^{(1)} r_i^{(2)}} \\
&\quad \times \left[p_1(1-p_2)\lambda_0\lambda_2 e^{-(\lambda_0 - a_2\lambda_2)x_{0i}} e^{-\lambda_2 x_{2i}} \right]^{(1-r_i^{(1)})r_i^{(2)}} \\
&\quad \times \left[(1-p_1)p_2\lambda_0\lambda_1 e^{-(\lambda_0 - a_1\lambda_1)x_{0i}} e^{-\lambda_1 x_{1i}} \right]^{r_i^{(1)}(1-r_i^{(2)})} \\
&\quad \times \left[(1-p_1)(1-p_2)\lambda_0\lambda_1\lambda_2 e^{-(\lambda_0 - a_1\lambda_1 - a_2\lambda_2)x_{0i}} e^{-\lambda_1 x_{1i}} e^{-\lambda_2 x_{2i}} \right]^{(1-r_i^{(1)})(1-r_i^{(2)})} \\
&= \prod_{i=1}^{n} \left[a_1 a_2 \frac{\lambda_1\lambda_2}{\lambda_0} e^{-\lambda_0 x_{0i}} \right]^{r_i^{(1)} r_i^{(2)}} \\
&\quad \times \left[a_1 \frac{\lambda_1\lambda_2}{\lambda_0} (\lambda_0 - a_2\lambda_2) e^{-(\lambda_0 - a_2\lambda_2)x_{0i}} e^{-\lambda_2 x_{2i}} \right]^{(1-r_i^{(1)})r_i^{(2)}} \\
&\quad \times \left[a_2 \frac{\lambda_1\lambda_2}{\lambda_0} (\lambda_0 - a_1\lambda_1) e^{-(\lambda_0 - a_1\lambda_1)x_{0i}} e^{-\lambda_1 x_{1i}} \right]^{r_i^{(1)}(1-r_i^{(2)})} \\
&\quad \times \left[\frac{\lambda_1\lambda_2}{\lambda_0} (\lambda_0 - a_1\lambda_1)(\lambda_2 - a_2\lambda_2) \right. \\
&\quad \left. \times e^{-(\lambda_0 - a_1\lambda_1 - a_2\lambda_2)x_{0i}} e^{-\lambda_1 x_{1i}} e^{-\lambda_2 x_{2i}} \right]^{(1-r_i^{(1)})(1-r_i^{(2)})} \\
&= \left(a_1 a_2 \frac{\lambda_1\lambda_2}{\lambda_0} \right)^{\sum_i r_i^{(1)} r_i^{(2)}} e^{-\lambda_0 \sum_i x_{0i} r_i^{(1)} r_i^{(2)}} \\
&\quad \times \left(a_1 a_2 \frac{\lambda_1\lambda_2}{\lambda_0} \right)^{\sum_i (1-r_i^{(1)}) r_i^{(2)}} (\lambda_0 - a_2\lambda_2)^{\sum_i (1-r_i^{(1)})} e^{-(\lambda_0 - a_2\lambda_2)\sum_i x_{0i}(1-r_i^{(1)}) r_i^{(2)}}
\end{aligned}
$$

$$\times\left(a_1a_2\frac{\lambda_1\lambda_2}{\lambda_0}\right)^{\sum_i r_i^{(1)}(1-r_i^{(2)})}(\lambda_0-a_1\lambda_1)^{\sum_i(1-r_i^{(2)})}e^{-(\lambda_0-a_1\lambda_1)\sum_i x_{0i}r_i^{(1)}(1-r_i^{(2)})}$$
$$\times\left(\frac{\lambda_1\lambda_2}{\lambda_0}\right)^{\sum_i(1-r_i^{(1)})(1-r_i^{(2)})}e^{-\lambda_1\sum_i x_{1i}(1-r_i^{(2)})}e^{-\lambda_2\sum_i x_{2i}(1-r_i^{(1)})}$$
$$\times e^{-(\lambda_0-a_1\lambda_1-a_2\lambda_2)\sum_i x_{0i}(1-r_i^{(1)})(1-r_i^{(2)})}.$$

Hence,

$$\begin{aligned}
L(\lambda_0,\lambda_1,\lambda_2) &= a_1^{\sum_i r_i^{(2)}}a_2^{\sum_i r_i^{(1)}}(\frac{\lambda_1\lambda_2}{\lambda_0})^n e^{-\lambda_0\sum_i x_{0i}r_i^{(1)}r_i^{(2)}}\\
&\quad (\lambda_0-a_2\lambda_2)^{\sum_i(1-r_i^{(1)})}e^{-(\lambda_0-a_2\lambda_2)\sum_i x_{0i}(1-r_i^{(1)})r_i^{(2)}}\\
&\quad (\lambda_0-a_1\lambda_1)^{\sum_i(1-r_i^{(2)})}e^{-(\lambda_0-a_1\lambda_1)\sum_i x_{0i}r_i^{(1)}(1-r_i^{(2)})}\\
&\quad e^{-\lambda_1\sum_i x_{1i}(1-r_i^{(2)})}e^{-\lambda_2\sum_i x_{2i}(1-r_i^{(1)})}\\
&\quad e^{-(\lambda_0-a_1\lambda_1-a_2\lambda_2)\sum_i x_{0i}(1-r_i^{(1)})(1-r_i^{(2)})}.
\end{aligned}$$

Hence the log likelihood is

$$\begin{aligned}
LL(\lambda_0,\lambda_1,\lambda_2) &= \log(a_1)\sum_i r_i^{(2)}+\log(a_2)\sum_i r_i^{(1)}+n\log(\frac{\lambda_1\lambda_2}{\lambda_0})\\
&\quad +\log(\lambda_0-a_1\lambda_1)\sum_i(1-r_i^{(2)})+\log(\lambda_0-a_2\lambda_2)\sum_i(1-r_i^{(1)})\\
&\quad -(\lambda_0-a_1\lambda_1)\sum x_{0i}r_i^{(1)}(1-r_i^{(2)})-(\lambda_0-a_1\lambda_1)\sum x_{0i}(1-r_i^{(1)})r_i^{(2)}\\
&\quad -\lambda_1\sum_i x_{1i}(1-r_i^{(2)})-\lambda_2\sum_i x_{2i}(1-r_i^{(1)})\\
&\quad -(\lambda_0-a_1\lambda_1-a_2\lambda_2)\sum_i x_{0i}(1-r_i^{(1)})(1-r_i^{(2)}),
\end{aligned}$$

$$\text{and}\quad \frac{\partial LL}{\partial\lambda_0} = -\frac{n}{\lambda_0}+\frac{\sum_i(1-r_i^{(2)})}{\lambda_0-a_1\lambda_1}+\frac{\sum_i(1-r_i^{(1)})}{\lambda_0-a_2\lambda_2}$$

$$
\begin{aligned}
& -\sum_i x_{0i} r_i^{(1)}(1-r_i^{(2)}) - \sum_i x_{0i}(1-r_i^{(1)}) r_i^{(2)} \\
& -\sum_i x_{0i}(1-r_i^{(1)})(1-r_i^{(2)}) \\
= \quad & -\frac{n}{\lambda_0} + \frac{\sum_i (1-r_i^{(2)})}{\lambda_0 - a_1\lambda_1} + \frac{\sum_i (1-r_i^{(1)})}{\lambda_0 - a_2\lambda_2} - \sum_i x_{0i}(1-r_i^{(1)} r_i^{(2)}).
\end{aligned}
$$

Similarly

$$
\frac{\partial LL}{\partial \lambda_1} = \frac{n}{\lambda_1} - a_1 \frac{\sum_i (1-r_i^{(2)})}{\lambda_0 - a_1\lambda_1} + a_1 \sum_i x_{0i}(1-r_i^{(2)}) - \sum_i x_{1i}(1-r_i^{(2)})
$$

So setting $\dfrac{\partial LL}{\partial \lambda_1} = 0$ gives

$$
\frac{\sum_i (1-r_i^{(2)})}{\lambda_0 - a_1\lambda_1} = \frac{n}{a_1\lambda_1} + \sum_i x_{0i}(1-r_i^{(2)}) - \frac{\sum_i x_{1i}(1-r_i^{(2)})}{a_1},
$$

and $$\frac{\partial LL}{\partial \lambda_2} = \frac{n}{\lambda_2} - a_2 \frac{\sum_i (1-r_i^{(1)})}{\lambda_0 - a_2\lambda_2} + a_2 \sum_i x_{0i}(1-r_i^{(1)}) - \sum_i x_{2i}(1-r_i^{(1)}).$$

So setting $\dfrac{\partial LL}{\partial \lambda_2} = 0$ gives

$$
\frac{\sum_i (1-r_i^{(1)})}{\lambda_0 - a_2\lambda_2} = \frac{n}{a_2\lambda_2} + \sum_i x_{0i}(1-r_i^{(1)}) - \frac{\sum_i x_{2i}(1-r_i^{(1)})}{a_2}.
$$

Now setting $\dfrac{\partial LL}{\partial \lambda_0} = 0$, and substituting values for $\dfrac{\sum_i (1-r_i^{(2)})}{\lambda_0 - a_1\lambda_1}$ and $\dfrac{\sum_i (1-r_i^{(1)})}{\lambda_0 - a_2\lambda_2}$ gives

$$
\frac{1}{\lambda_0} = \frac{1}{a_1\lambda_1} + \frac{1}{a_2\lambda_2} - \frac{\sum_i \{x_{2i}(1-r_i^{(1)}) - a_2 x_{0i}(1-r_i^{(1)})\}}{a_2 n}
$$

$$
\begin{aligned}
& -\frac{\sum_i \{x_{1i}(1-r_i^{(2)}) - a_1 x_{0i}(1-r_i^{(2)})\}}{a_1 n} - \frac{\sum_i x_{0i}(1-r_i^{(1)} r_i^{(2)})}{n} \\
= & \frac{1}{a_1\lambda_1} + \frac{1}{a_2\lambda_2} - \frac{\sum_i z_{1i}(1-r_i^{(2)})}{a_1 n} - \frac{\sum_i z_{2i}(1-r_i^{(1)})}{a_2 n} - \frac{\sum_i x_{0i}(1-r_i^{(1)} r_i^{(2)})}{n} \\
= & \frac{1}{a_1}(\frac{1}{\lambda_1} - \frac{\sum_i z_{1i}(1-r_i^{(2)})}{n}) + \frac{1}{a_2}(\frac{1}{\lambda_2} - \frac{\sum_i z_{2i}(1-r_i^{(1)})}{n}) - \frac{\sum_i x_{0i}(1-r_i^{(1)} r_i^{(2)})}{n}.
\end{aligned}
$$

The above likelihood equations can be used to estimate λ_0, λ_1 and λ_2 if the x_{0i}'s, $1 \leq i \leq n$, were known. We address this question, by developing estimators of these latent terms, and use results that were obtained previously in Section 4.3. It is worth noting that no approximations has been used here in contrast with the technique used in Chapter 3 and Chapter 4.

To develop the unconditional MLE, we integrate out x_0 from the joint density $f(x_0, x_1, x_2)$, and we have that:

$$
\begin{aligned}
f(x_1, x_2) &= \int_{x_0} f(x_0, x_1, x_2) dx_0 \\
&= p_1 p_2 \int \lambda_0 e^{-\lambda_0 x_0} \delta_{(x_1 - a_1 x_0)} \delta_{(x_2 - a_2 x_0)} dx_0 \\
&\quad + p_1(1-p_2) \int \lambda_0 e^{-\lambda_0 x_0} f_{X_2}(x_2 - a_2 x_0) I_{(x_2 > a_2 x_0)} \delta_{(x_1 - a_1 x_0)} dx_0 \\
&\quad + (1-p_1) p_2 \int \lambda_0 e^{-\lambda_0 x_0} f_{X_1}(x_1 - a_1 x_0) I_{(x_1 > a_1 x_0)} \delta_{(x_2 - a_2 x_0)} dx_0 \\
&\quad + (1-p_1)(1-p_2) \int \lambda_0 e^{-\lambda_0 x_0} f_{X_1}(x_1 - a_1 x_0) f_{X_2}(x_2 - a_2 x_0) I_{(\frac{x_1}{a_1} > x_0, \frac{x_2}{a_2} > x_0)} dx_0 \\
&= p_1 p_2 PartA_1 + p_1(1-p_2) PartA_2 \\
&\quad + (1-p_1) p_2 PartA_3 + (1-p_1)(1-p_2) PartA_4,
\end{aligned}
$$

$$
\text{where} \quad PartA_1 = \int \lambda_0 e^{-\lambda_0 x_0} \delta_{(x_1 - a_1 x_0)} \delta_{(x_2 - a_2 x_0)} dx_0
$$

$$
\begin{aligned}
&= \frac{1}{a_1 a_2}\lambda_0 \int e^{-\lambda_0 x_0}\delta_{(\frac{x_1}{a_1}-x_0)}\delta_{(\frac{x_2}{a_2}-x_0)}dx_0 \\
&= \begin{cases} \frac{1}{a_1 a_2}\lambda_0 e^{-\lambda_0 \frac{x_2}{a_2}}\delta_{(\frac{x_1}{a_1}-\frac{x_2}{a_2})} & ; \\ or & \\ \frac{1}{a_1 a_2}\lambda_0 e^{-\lambda_0 \frac{x_1}{a_1}}\delta_{(\frac{x_1}{a_1}-\frac{x_2}{a_2})} & . \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
PartA_2 &= \int \lambda_0 e^{-\lambda_0 x_0} f_{X_2}(x_2 - a_2 x_0) I_{(x_2 > a_2 x_0)} \delta_{(x_1 - a_1 x_0)} dx_0 \\
&= \frac{1}{a_1}\int \lambda_0 e^{-\lambda_0 x_0} f_{X_2}(x_2 - a_2 x_0) I_{(x_2 > a_2 x_0)} \delta_{(\frac{x_1}{a_1} - x_0)} dx_0 \\
&= \frac{1}{a_1}\lambda_0 e^{-\lambda_0 \frac{x_1}{a_1}} f_{X_2}(x_2 - a_2 \frac{x_1}{a_1}) I_{(x_2 > a_2 \frac{x_1}{a_1})} \\
&= \frac{1}{a_1}\lambda_0 \lambda_2 e^{-\lambda_0 \frac{x_1}{a_1}} e^{-\lambda_2 (x_2 - a_2 \frac{x_1}{a_1})} I_{(x_2 > a_2 \frac{x_1}{a_1})} \\
&= \frac{1}{a_1}\lambda_0 \lambda_2 e^{-\frac{x_1}{a_1}(\lambda_0 - a_2\lambda_2)} e^{-\lambda_2 x_2} I_{(x_2 > a_2 \frac{x_1}{a_1})} \\
&= \frac{1}{a_1}\lambda_0 \lambda_2 e^{-\frac{x_1}{a_1}(\lambda_0 - a_2\lambda_2)} e^{-\lambda_2 x_2} I_{(\frac{x_2}{a_2} > \frac{x_1}{a_1})} \\
&= \frac{1}{a_1}\lambda_0 \lambda_2 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} e^{-\frac{x_1}{a_1}(\lambda_0 - a_1\lambda_1 - a_2\lambda_2)} I_{(\frac{x_2}{a_2} > \frac{x_1}{a_1})} \\
&= \frac{1}{a_1}\lambda_0 \lambda_2 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} e^{-\lambda^* \varphi} I_{(\frac{x_2}{a_2} > \frac{x_1}{a_1})},
\end{aligned}
$$

where $\qquad \varphi = min\left(\dfrac{x_1}{a_1}, \dfrac{x_2}{a_2}\right)$ and $\lambda^* = \lambda_0 - a_1\lambda_1 - a_2\lambda_2.$

Similarly,
$$
\begin{aligned}
PartA_3 &= \int \lambda_0 e^{-\lambda_0 x_0} f_{X_1}(x_1 - a_1 x_0) I_{(x_1 > a_1 x_0)} \delta_{(x_2 - a_2 x_0)} dx_0 \\
&= \frac{1}{a_2}\int \lambda_0 e^{-\lambda_0 x_0} f_{X_1}(x_1 - a_1 x_0) I_{(x_1 > a_1 x_0)} \delta_{(\frac{x_2}{a_2} - x_0)} dx_0 \\
&= \frac{1}{a_2}\lambda_0 e^{-\lambda_0 \frac{x_2}{a_2}} f_{X_1}(x_1 - a_1 \frac{x_2}{a_2}) I_{(x_1 > a_1 \frac{x_2}{a_2})}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{a_2}\lambda_0\lambda_1 e^{-\lambda_0\frac{x_2}{a_2}} e^{-\lambda_1(x_1-a_1\frac{x_2}{a_2})} I_{(x_1>a_1\frac{x_2}{a_2})}\\
&= \frac{1}{a_2}\lambda_0\lambda_1 e^{-\frac{x_2}{a_2}(\lambda_0-a_1\lambda_1)} e^{-\lambda_1 x_1} I_{(x_1>a_1\frac{x_2}{a_2})}\\
&= \frac{1}{a_2}\lambda_0\lambda_1 e^{-\frac{x_2}{a_2}(\lambda_0-a_1\lambda_1)} e^{-\lambda_1 x_1} I_{(\frac{x_1}{a_1}>\frac{x_2}{a_2})}\\
&= \frac{1}{a_2}\lambda_0\lambda_1 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} e^{-\frac{x_2}{a_2}(\lambda_0-a_1\lambda_1-a_2\lambda_2)} I_{(\frac{x_1}{a_1}>\frac{x_2}{a_2})}\\
&= \frac{1}{a_2}\lambda_0\lambda_1 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} e^{-\lambda^*\varphi} I_{(\frac{x_1}{a_1}>\frac{x_2}{a_2})},
\end{aligned}
$$

$$
\begin{aligned}
\text{and} \quad PartA_4 &= \int \lambda_0 e^{-\lambda_0 x_0} f_{X_1}(x_1-a_1x_0) f_{X_2}(x_2-a_2x_0) I_{(x_1>a_1x_0, x_2>a_2x_0)} dx_0\\
&= \int_0^{\varphi} \lambda_0\lambda_1\lambda_2 e^{-\lambda_0 x_0} e^{-\lambda_1(x_1-a_1x_0)} e^{-\lambda_2(x_2-a_2x_0)} dx_0\\
&= \int_0^{\varphi} \lambda_0\lambda_1\lambda_2 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} e^{-(\lambda_0-a_1\lambda_1-a_2\lambda_2)x_0} dx_0\\
&= \lambda_0\lambda_1\lambda_2 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} \int_0^{\varphi} e^{-\lambda^* x_0} dx_0\\
&= \frac{\lambda_0\lambda_1\lambda_2}{\lambda^*} e^{-\lambda_1 x_1} e^{-\lambda_2 x_2}(1-e^{-\lambda^*\varphi}).
\end{aligned}
$$

Hence the expression for the joint density becomes:

$$
\begin{aligned}
f(x_1,x_2) &= p_1p_2\frac{\lambda_0}{a_1a_2} e^{-\lambda_0\frac{x_2}{a_2}} \delta_{(\frac{x_1}{a_1}-\frac{x_2}{a_2})}\\
&\quad + p_1(1-p_2)\frac{1}{a_1}\lambda_0\lambda_2 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} e^{-\lambda^*\varphi} I_{(\frac{x_2}{a_2}>\frac{x_1}{a_1})}\\
&\quad + (1-p_1)p_2\frac{1}{a_2}\lambda_0\lambda_1 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} e^{-\lambda^*\varphi} I_{(\frac{x_1}{a_1}>\frac{x_2}{a_2})}\\
&\quad + (1-p_1)(1-p_2)\frac{\lambda_0\lambda_1\lambda_2}{\lambda^*} e^{-\lambda_1 x_1} e^{-\lambda_2 x_2}(1-e^{-\lambda^*\varphi}),
\end{aligned}
$$

$$
\text{where} \quad \varphi = min\left(\frac{x_1}{a_1}, \frac{x_2}{a_2}\right) \quad \text{and} \quad \lambda^* = \lambda_0 - a_1\lambda_1 - a_2\lambda_2,
$$

with a similar form as in previous section.

Based on a random sample of size n denoted $(x_{11}, x_{21}), (x_{12}, x_{22}), \cdots, (x_{1n}, x_{2n})$, let's define $r_j^{(1)} = \begin{cases} 1, & \text{if } x_{0j} = \frac{x_{1j}}{a_1} \le \frac{x_{2j}}{a_2}; \\ 0, & \text{if } \frac{x_{1j}}{a_1} > \frac{x_{2j}}{a_2} \end{cases}$ and $r_j^{(2)} = \begin{cases} 1, & \text{if } x_{0j} = \frac{x_{2j}}{a_2} \le \frac{x_{1j}}{a_1}; \\ 0, & \text{if } \frac{x_{2j}}{a_2} > \frac{x_{1j}}{a_1}. \end{cases}$ Then

$$\begin{aligned}
L(\lambda_0, \lambda_1, \lambda_2) &= \prod_{j=1}^{n} f(x_{1j}, x_{2j}) \\
= \prod_{j=1}^{n} & \left[p_1 p_2 \frac{\lambda_0}{a_1 a_2} e^{-\lambda_0 \frac{x_2}{a_2}} \right]^{r_j^{(1)} r_j^{(2)}} \\
& \left[p_1 (1-p_2) \frac{1}{a_1} \lambda_0 \lambda_2 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} e^{-\lambda^* \varphi} \right]^{r_j^{(1)}(1-r_j^{(2)})} \\
& \left[(1-p_1) p_2 \frac{1}{a_2} \lambda_0 \lambda_1 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} e^{-\lambda^* \varphi} \right]^{(1-r_j^{(1)}) r_j^{(2)}} \\
& \left[(1-p_1)(1-p_2) \frac{\lambda_0 \lambda_1 \lambda_2}{\lambda^*} e^{-\lambda_1 x_1} e^{-\lambda_2 x_2} (1 - e^{-\lambda^* \varphi}) \right]^{(1-r_j^{(1)})(1-r_j^{(2)})}.
\end{aligned}$$

In order to obtain estimators, the log likelihood is:

$$\begin{aligned}
l(\lambda_0, \lambda_1, \lambda_2) \quad & = \log L(\lambda_0, \lambda_1, \lambda_2) \\
= \sum_i \quad & r_i^{(1)} r_i^{(2)} \Big[\log(p_1 p_2) + \log(\lambda_0) - \log(a_1 a_2) - \lambda_0 \varphi_i \Big] \\
+ \quad & r_i^{(1)} (1 - r_i^{(2)}) \Big[\log p_1 (1-p_2) + \log \lambda_0 + \log \lambda_2 - \log a_1 - \lambda_1 x_{1i} - \lambda_2 x_{2i} - \lambda^* \varphi_i \Big] \\
+ \quad & (1 - r_i^{(1)}) r_i^{(2)} \Big[\log(1-p_1) p_2 + \log \lambda_0 + \log \lambda_1 - \log a_2 - \lambda_1 x_{1i} - \lambda_2 x_{2i} - \lambda^* \varphi_i \Big] \\
+ \quad & (1 - r_i^{(1)})(1 - r_i^{(2)}) \Big[\log(1-p_1)(1-p_2) + \log \lambda_0 + \log \lambda_1 \lambda_2 - \log \lambda^* \\
& - \lambda_1 x_{1i} - \lambda_2 x_{2i} + \log(1 - e^{-\lambda^* \varphi_i}) \Big].
\end{aligned}$$

Assume that p_1 and p_2 are fixed known constant. Then

$$
\begin{aligned}
\frac{\partial l}{\partial \lambda_0} &= \sum_i r_i^{(1)} r_i^{(2)} \left[\frac{1}{\lambda_0} - \varphi_i\right] + r_i^{(1)}(1-r_i^{(2)})\left[\frac{1}{\lambda_0} - \varphi_i\right] \\
&\quad + (1-r_i^{(1)})r_i^{(2)}\left[\frac{1}{\lambda_0} - \varphi_i\right] + (1-r_i^{(1)})(1-r_i^{(2)})\left[\frac{1}{\lambda_0} - \frac{1}{\lambda^*} - \frac{\varphi e^{-\lambda^*\varphi_i}}{1-e^{-\lambda^*\varphi_i}}\right] \\
&= \sum_i \left[\frac{1}{\lambda_0} - \varphi_i\left(r_i^{(1)} + r_i^{(2)} - r_i^{(1)} r_i^{(2)}\right)\right. \\
&\quad \left. -(1-r_i^{(1)})(1-r_i^{(2)})\left(\frac{1}{\lambda^*} + \frac{\varphi_i e^{-\lambda^*\varphi_i}}{1-e^{-\lambda^*\varphi_i}}\right)\right] \\
&= \frac{n}{\lambda_0} - \sum_i \varphi_i\left(r_i^{(1)} + r_i^{(2)} - r_i^{(1)} r_i^{(2)}\right) \\
&\quad - \sum_i (1-r_i^{(1)})(1-r_i^{(2)})\left(\frac{1}{\lambda^*} + \frac{\varphi_i e^{-\lambda^*\varphi_i}}{1-e^{-\lambda^*\varphi_i}}\right).
\end{aligned}
$$

$$
\begin{aligned}
\text{Similarly,}\quad \frac{\partial l}{\partial \lambda_1} &= \sum_i r_i^{(1)}(1-r_i^2)[-x_{1i} + a_1\varphi_i] + (1-r_i^{(1)})r_i^{(2)}[\frac{1}{\lambda_1} - x_{1i} + a_1\varphi_i] \\
&\quad + (1-r_i^{(1)})(1-r_i^{(2)})[\frac{1}{\lambda_1} + \frac{a_1}{\lambda^*} - x_{1i} + \frac{a_i\varphi_i e^{-\lambda^*\varphi_i}}{1-e^{-\lambda^*\varphi_i}}] \\
&= \sum_i -x_{1i}(1-r_i^{(1)} r_i^{(2)}) + a_1\varphi_i[r_i^{(1)}(1-r_i^{(2)}) + (1-r_i^{(1)})r_i^{(2)}] \\
&\quad + \frac{1}{\lambda_1}(1-r_i^{(1)}) + a_1(1-r_i^{(1)})(1-r_i^{(2)})(\frac{1}{\lambda^*} + \frac{\varphi_i e^{-\lambda^*\varphi_i}}{1-e^{-\lambda^*\varphi_i}}).
\end{aligned}
$$

So setting $\frac{\partial l}{\partial \lambda_1} = 0$ and $\frac{\partial l}{\partial \lambda_0} = 0$ gives

$$
a_1 \sum_i (1-r_i^{(1)})(1-r_i^{(2)})\left(\frac{1}{\lambda^*} + \frac{\varphi_i e^{-\lambda^*\varphi_i}}{1-e^{-\lambda^*\varphi_i}}\right) =
$$

$$
\sum_i x_{1i}(1-r_i^{(1)} r_i^{(2)}) - a_1\varphi_i\left[r_i^{(1)}(1-r_i^{(2)}) + (1-r_i^{(1)})r_i^{(2)}\right] - \frac{1}{\lambda_1}(1-r_i^{(1)}),
$$

and hence $\quad \dfrac{n}{\lambda_0} = \dfrac{1}{a_1}\sum_i x_{1i}(1 - r_i^{(1)} r_i^{(2)}) + \sum_i \varphi_i r_i^{(1)} r_i^{(2)} - \dfrac{1}{a_1\lambda_1}\sum_i (1 - r_i^{(1)}),$

or $\quad \dfrac{1}{\lambda_0} = \dfrac{1}{a_1}\sum_i \dfrac{x_{1i}(1 - r_i^{(1)} r_i^{(2)})}{n} + \sum_i \dfrac{\varphi_i r_i^{(1)} r_i^{(2)}}{n} - \dfrac{1}{a_1\lambda_1}\sum_i \dfrac{(1 - r_i^{(1)})}{n}.$

A similar formula can be obtained by taking $\dfrac{\partial l}{\partial \lambda_2}$ and setting it equal to zero.

$$\begin{aligned}
\frac{\partial l}{\partial \lambda_2} &= \sum_i r_i^{(1)}(1 - r_i^2)\left[\frac{1}{\lambda_2} - x_{2i} + a_2\varphi_i\right] + (1 - r_i^{(1)}) r_i^{(2)}[-x_{2i} + a_2\varphi_i] \\
&\quad + (1 - r_i^{(1)})(1 - r_i^{(2)})\left[\frac{1}{\lambda_2} + \frac{a_2}{\lambda^*} - x_{2i} + \frac{a_2\varphi_i e^{-\lambda^*\varphi_i}}{1 - e^{-\lambda^*\varphi_i}}\right] \\
&= \sum_i -x_{1i}(1 - r_i^{(1)} r_i^{(2)}) + a_2\varphi_i\left[r_i^{(1)}(1 - r_i^{(2)}) + (1 - r_i^{(1)}) r_i^{(2)}\right] \\
&\quad + \frac{1}{\lambda_2}(1 - r_i^{(2)}) + a_2(1 - r_i^{(1)})(1 - r_i^{(2)})\left(\frac{1}{\lambda^*} + \frac{\varphi_i e^{-\lambda^*\varphi_i}}{1 - e^{-\lambda^*\varphi_i}}\right).
\end{aligned}$$

So, $\quad a_2\sum_i (1 - r_i^{(1)})(1 - r_i^{(2)})\left(\dfrac{1}{\lambda^*} + \dfrac{\varphi_i e^{-\lambda^*\varphi_i}}{1 - e^{-\lambda^*\varphi_i}}\right) =$

$$\sum_i x_{2i}(1 - r_i^{(1)} r_i^{(2)}) - a_2\varphi_i\left[r_i^{(1)}(1 - r_i^{(2)}) + (1 - r_i^{(1)}) r_i^{(2)}\right] - \frac{1}{\lambda_2}(1 - r_i^{(2)})$$

and hence, $\quad \dfrac{n}{\lambda_0} = \dfrac{1}{a_2}\sum_i x_{2i}(1 - r_i^{(1)} r_i^{(2)}) + \sum_i \varphi_i r_i^{(1)} r_i^{(2)} - \dfrac{1}{a_2\lambda_2}\sum_i (1 - r_i^{(2)}),$

or $\quad \dfrac{1}{\lambda_0} = \dfrac{1}{a_2}\sum_i \dfrac{x_{2i}(1 - r_i^{(1)} r_i^{(2)})}{n} + \sum_i \dfrac{\varphi_i r_i^{(1)} r_i^{(2)}}{n} - \dfrac{1}{a_2\lambda_2}\sum_i \dfrac{(1 - r_i^{(2)})}{n}.$

This above equation gives estimate of the parameter associated with the unknown latent variable x_0.

The distribution of the minimum is also exponential, so the need to characterize its distribution. For $X_{(1)}$ denoting the minimum between $\dfrac{X_1}{a_1}$ and $\dfrac{X_2}{a_2}$, we have, based on the independence of $X_1|x_0$ and $X_2|x_0$ that:

$$P(X_{(1)} > t|x_0) = P(X_1 > a_1 t|x_0)P(X_2 > a_2 t|x_0). \tag{5.7}$$

Equation (5.7) is obtained from Theorem 1.8 and the fact that $X_1|x_0$ and $X_2|x_0$ are independent. But $X_1|x_0$ and $X_2|x_0$ are not identically distributed. So we cannot use the results in say Dudewicz and Mishra (1988) [23] of the minimal order statistic. Hence, we need to find $P(X_i > a_i t|x_0)$ for $i = 1, 2$.

$$\begin{aligned}
P(X_i > a_i t|x_0) &= P(a_i x_0 + Z_i > a_i t|x_0) \\
&= P(Z_i > a_i(t - x_0)|x_0) \\
&= \int_{a_i(t-x_0)}^{\infty} f_{Z_i}(z_i)dz_i \\
&= \int_{a_i(t-x_0)}^{\infty} \Big[p_i\delta(z_i) + (1-p_i)\lambda_i e^{-\lambda_i z_i}\Big]dz_i \\
&= p\int_{a_i(t-x_0)}^{\infty} \delta(z_i)dz_i + (1-p_i)\lambda_i \int_{a_i(t-x_0)}^{\infty} e^{-\lambda_i z_i}dz_i \\
&= p_i\Big[1 - H(a_i(t-x_0))\Big] + (1-p_i)e^{-a_i\lambda_i(t-x_0)} \\
&= p_i\Big[1 - H(t-x_0)\Big] + (1-p_i)e^{-a_i\lambda_i(t-x_0)}
\end{aligned}$$

From the above, we can deduce the conditional density of $X_i|x_0$ which is:

$$f_{X_i|x_0}(t) = -\frac{d}{dt}P(X_i > a_i t|x_0) = p_i\delta(t-x_0) + (1-p_i)\lambda_i e^{-a_i\lambda_i(t-x_0)}.$$

Our goal is to find $\quad f_{X_{(1)}|X_0}(t) = -\dfrac{d}{dt}P(X_{(1)} > t|x_0).$

So from Theorem 1.8, the hazard function of the minimum lifetimes is given by:

$$\begin{aligned}
P(X_{(1)} > t|x_0) &= P(X_1 > a_1 t|x_0)P(X_2 > a_2 t|x_0) \\
&= \left[p_1(1 - H(t - x_0)) + (1 - p_1)e^{-a_1\lambda_1(t-x_0)}\right] \\
&\quad \left[p_2(1 - H(t - x_0)) + (1 - p_2)e^{-a_2\lambda_2(t-x_0)}\right] \\
&= p_1 p_2\Big(1 - H(t - x_0)\Big) \\
&\quad + p_1(1 - p_2)\Big(1 - H(t - x_0)\Big)e^{-a_2\lambda_2(t-x_0)} \\
&\quad + (1 - p_1)p_2\Big(1 - H(t - x_0)\Big)e^{-a_1\lambda_1(t-x_0)} \\
&\quad + (1 - p_1)(1 - p_2)e^{-(a_1\lambda_1 + a_2\lambda_2)(t-x_0)}.
\end{aligned}$$

Hence the conditional density of $X_{(1)}$ given X_0 is:

$$\begin{aligned}
f_{(X_{(1)}|x_0}(t) &= -\frac{d}{dt}P(X_{(1)} > t|x_0) \\
&= p_1 p_2 \delta(t - x_0) \\
&\quad + p_1(1 - p_2)\left[\delta(t - x_0)e^{-a_2\lambda_2(t-x_0)} + \underbrace{(1 - H(t - x_0))a_2\lambda_2 e^{-a_2\lambda_2(t-x_0)}}\right] \\
&\quad + (1 - p_1)p_2\left[\delta(t - x_0)e^{-a_1\lambda_1(t-x_0)} + \underbrace{(1 - H(t - x_0))a_1\lambda_1 e^{-a_1\lambda_1(t-x_0)}}\right] \\
&\quad + (1 - p_1)(1 - p_2)(a_1\lambda_1 + a_2\lambda_2)e^{-(a_1\lambda_1 + a_2\lambda_2)(t-x_0)} \\
&= p_1 p_2 \delta(t - x_0) \\
&\quad + p_1(1 - p_2)\delta(t - x_0)e^{-a_2\lambda_2(t-x_0)} \\
&\quad + (1 - p_1)p_2\delta(t - x_0)e^{-a_1\lambda_1(t-x_0)}
\end{aligned}$$

$$+(1-p_1)(1-p_2)(a_1\lambda_1+a_2\lambda_2)e^{-(a_1\lambda_1+a_2\lambda_2)(t-x_0)}$$

where the underlined expression is zero by definition of the Heaviside function in Equation (2.3).

Note that $\int_{-\infty}^{\infty} f_{(X_{(1)}|x_0}(t)dt = 1.$ Also,

$$\begin{aligned}
E(X_{(1)}|x_0) &= \int_{-\infty}^{\infty} t f_{X_{(1)}|x_0}(t)dt \\
&= p_1p_2\int_{-\infty}^{\infty} t\delta(t-x_0)dt \\
&\quad +p_1(1-p_2)\int_{-\infty}^{\infty} t\delta(t-x_0)e^{-a_2\lambda_2(t-x_0)}dt \\
&\quad +(1-p_1)p_2\int_{-\infty}^{\infty} t\delta(t-x_0)e^{-a_1\lambda_1(t-x_0)}dt \\
&\quad +(1-p_1)(1-p_2)\int_{-\infty}^{\infty}(a_1\lambda_1+a_2\lambda_2)e^{-(a_1\lambda_1+a_2\lambda_2)(t-x_0)}dt \\
&= p_1p_2x_0+p_1(1-p_2)x_0+(1-p_1)p_2x_0+(1-p_1)(1-p_2)\left(\frac{1}{a_1\lambda_1+a_2\lambda_2}+x_0\right) \\
&= x_0+\frac{(1-p_1)(1-p_2)}{a_1\lambda_1+a_2\lambda_2}.
\end{aligned}$$

$$\text{So,}\qquad E(X_{(1)}|x_0) = x_0+\frac{(1-p_1)(1-p_2)}{a_1\lambda_1+a_2\lambda_2}.$$

$$\text{Hence,}\quad EX_{(1)} = E_{X_0}E(X_{(1)}|x_0) = E(X_0)+E\left[\frac{(1-p_1)(1-p_2)}{a_1\lambda_1+a_2\lambda_2}\right]$$

$$EX_{(1)} = \frac{1}{\lambda_0}+E\left[\frac{(1-p_1)(1-p_2)}{a_1\lambda_1+a_2\lambda_2}\right].$$

Another way to obtain $EX_{(1)}$ using the densities is as follows:

$$f_{X_{(1)},X_0}(t,x_0) = f_{X_{(1)}|X_0}(t)f_{X_0}(x_0)$$

$$
\begin{aligned}
&= p_1p_2\lambda_0 e^{-\lambda_0 x_0}\delta(t-x_0)\\
&\quad +p_1(1-p_2)\lambda_0 e^{-\lambda_0 x_0}e^{-a_2\lambda_2(t-x_0)}\delta(t-x_0)\\
&\quad +(1-p_1)p_2\lambda_0 e^{-\lambda_0 x_0}e^{-a_1\lambda_1(t-x_0)}\delta(t-x_0)\\
&\quad +(1-p_1)(1-p_2)\lambda_0(a_1\lambda_1+a_2\lambda_2)e^{-\lambda_0 x_0}e^{-(a_1\lambda_1+a_2\lambda_2)(t-x_0)}\delta(t-x_0).
\end{aligned}
$$

Hence the form of the density of the minimum order statistic is:

$$
\begin{aligned}
f_{X_{(1)}}(t) &= \int_{-\infty}^{\infty} f_{X_{(1)},X_0}(t,x_0)dx_0\\
&= p_1p_2\lambda_0\int_{-\infty}^{\infty} e^{-\lambda_0 x_0}\delta(t-x_0)dx_0\\
&\quad +p_1(1-p_2)\lambda_0\int_{-\infty}^{\infty} e^{-\lambda_0 x_0}e^{-a_2\lambda_2(t-x_0)}\delta(t-x_0)dx_0\\
&\quad +(1-p_1)p_2\lambda_0\int_{-\infty}^{\infty} e^{-\lambda_0 x_0}e^{-a_1\lambda_1(t-x_0)}\delta(t-x_0)dx_0\\
&\quad +(1-p_1)(1-p_2)\lambda_0(a_1\lambda_1+a_2\lambda_2)\int_0^t e^{-\lambda_0 x_0}e^{-(a_1\lambda_1+a_2\lambda_2)(t-x_0)}\delta(t-x_0)dx_0\\
&= p_1p_2\lambda_0 e^{-\lambda_0 t}\\
&\quad +p_1(1-p_2)\lambda_0 e^{-\lambda_0 t}\\
&\quad +(1-p_1)p_2\lambda_0 e^{-\lambda_0 t}\\
&\quad +(1-p_1)(1-p_2)\frac{\lambda_0(a_1\lambda_1+a_2\lambda_2)}{\lambda^*}(1-e^{-\lambda^* t})e^{-(a_1\lambda_1+a_2\lambda_2)t}\\
&= \lambda_0 e^{-\lambda_0 t}+(1-p_1)(1-p_2)[\frac{\lambda_0(a_1\lambda_1+a_2\lambda_2)}{\lambda^*}(1-e^{-\lambda^* t})e^{-(a_1\lambda_1+a_2\lambda_2)t}-\lambda_0 e^{-\lambda_0 t}]\\
&= \lambda_0 e^{-\lambda_0 t}+(1-p_1)(1-p_2)\frac{\lambda_0}{\lambda^*}\left[(a_1\lambda_1+a_2\lambda_2)e^{-(a_1\lambda_1+a_2\lambda_2)t}-\lambda_0 e^{-\lambda_0 t}\right].
\end{aligned}
$$

Hence

$$
EX_{(1)} = \frac{1}{\lambda_0}+(1-p_1)(1-p_2)\frac{\lambda_0}{\lambda^*}\left[\frac{1}{a_1\lambda_1+a_2\lambda_2}-\frac{1}{\lambda_0}\right].
$$

Recall that:

$$
\begin{aligned}
P(X_i > x_i|x_0) &= P(a_i x_0 + Z_i > x_i|x_0) \\
&= P(Z_i > x_i - a_i x_0|x_0) \\
&= \int_{x_i - a_i x_0}^{\infty} f_{Z_i}(z)dz \\
&= \int_{x_i - a_i x_0}^{\infty} \Big[p_i \delta(z) + (1-p_i)\lambda_i e^{-\lambda_i z}\Big] dz \\
&= p \int_{x_i - a_i x_0}^{\infty} \delta(z)dz + (1-p_i)\lambda_i \int_{x_i - a_i x_0}^{\infty} e^{-\lambda_i z} dz \\
&= p_i \Big[1 - H(x_i - a_i x_0)\Big] + (1-p_i) e^{-\lambda_i (x_i - a_i x_0)}.
\end{aligned}
$$

Hence $$
\begin{aligned}
F_{X_i|x_0}(x_i) = P(X_i \leq x_i|x_0) &= 1 - P(X_i > x_i|x_0) \\
&= p_i H(x_i - a_i x_0) + (1-p_i)\Big[1 - e^{-\lambda_i (x_i - a_i x_0)}\Big].
\end{aligned}
$$

And, $$
\begin{aligned}
F(x_1, x_2|x_0) &= \Big\{p_1 H(x_1 - a_1 x_0) + (1-p_1)\Big[1 - e^{-\lambda_1 (x_1 - a_1 x_0)}\Big]\Big\} \\
&\quad \times \Big\{p_2 H(x_2 - a_2 x_0) + (1-p_2)\Big[1 - e^{-\lambda_2 (x_2 - a_2 x_0)}\Big]\Big\} \\
&= p_1 p_2 H(x_1 - a_1 x_0) H(x_2 - a_2 x_0) \\
&\quad + p_1(1-p_2) H(x_1 - a_1 x_0)\Big[1 - e^{-\lambda_2 (x_2 - a_2 x_0)}\Big] \\
&\quad + (1-p_1) p_2 H(x_2 - a_2 x_0)\Big[1 - e^{-\lambda_1 (x_1 - a_1 x_0)}\Big] \\
&\quad + (1-p_1)(1-p_2)\Big[1 - e^{-\lambda_1 (x_1 - a_1 x_0)}\Big]\Big[1 - e^{-\lambda_2 (x_2 - a_2 x_0)}\Big].
\end{aligned}
$$

$$\begin{aligned}
\text{Also,} \quad F(x_1, \infty|x_0) &= \lim_{x_2 \to \infty} F(x_1, x_2|x_0) \\
&= p_1 p_2 H(x_1 - a_1 x_0) + p_1(1-p_2) H(x_1 - a_1 x_0) \\
&\quad + (1-p_1)p_2 \left[1 - e^{-\lambda_2(x_2 - a_2 x_0)}\right] + (1-p_1)(1-p_2)\left[1 - e^{-\lambda_1(x_1 - a_1 x_0)}\right] \\
&= p_1 H(x_1 - a_1 x_0) + (1-p_1)\left[1 - e^{-\lambda_1(x_1 - a_1 x_0)}\right].
\end{aligned}$$

$$\text{Similarly,} \qquad F(\infty, x_2|x_0) = p_2 H(x_2 - a_2 x_0) + (1-p_2)\left[1 - e^{-\lambda_2(x_2 - a_2 x_0)}\right].$$

$$\begin{aligned}
\text{Hence} \quad S(x_1, x_2|x_0) &= 1 + p_1 p_2 H(x_1 - a_1 x_0) H(x_2 - a_2 x_0) \\
&\quad - p_1 p_2 H(x_1 - a_1 x_0) - p_1(1-p_2) H(x_1 - a_1 x_0) e^{-\lambda_2(x_2 - a_2 x_0)} \\
&\quad - p_1 p_2 H(x_2 - a_2 x_0) - (1-p_1) p_2 H(x_2 - a_2 x_0) e^{-\lambda_1(x_1 - a_1 x_0)} \\
&\quad + (1-p_1)(1-p_2)\left[1 - e^{-\lambda_1(x_1 - a_1 x_0)}\right]\left[1 - e^{-\lambda_2(x_2 - a_2 x_0)}\right] \\
&\quad - (1-p_1)\left[1 - e^{-\lambda_1(x_1 - a_1 x_0)}\right] - (1-p_2)\left[1 - e^{-\lambda_2(x_2 - a_2 x_0)}\right].
\end{aligned}$$

We have proposed a very general bivariate exponential class distributions, which includes all the work considered in Chapter 3 and in Chapter 4. We have described the form of the joint distribution and survival functions. The distribution of the minimum has been characterized. Estimations of the parameters are given based on the likelihood equation. We have done all that retaining the form of the marginal exponential distribution, and the fatal shock idea as in Marshall and Olkin (1967) [61].

These results are not easily generalized to the p-variate gamma distribution case. One reason is that there is no need to approximate the likelihood function in the multivariate exponential in contrast with the results obtained from Chapter 4. In the next section, we examine a simulated example to illustrate our proposed model.

5.3 Simulation Example

In this section, we perform a simulation study of the multivariate exponential to examine the properties of various estimators of the parameter from the latent distribution, λ_0. We focus on λ_0 because it is an important portion of the correlation structure, and all of the parameters associated with X_1 through X_p can be easily estimated marginally, which is well documented in the literature. We examine the multivariate exponential given in (1.1) and (1.3) for various dimensions, from $p = 2$ to $p = 5$.

Based on 10,000 replications of sample size 50 each, of $(X_1, X_2, \ldots, X_p)'$, we choose all a_i's to be 1, λ_0 to be 1, and solving for λ_i w.r.t. ρ, we have that $\lambda_i = \frac{\rho}{a_i}\lambda_0 = \rho$. Separate simulations are done for $\rho = 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90$ and 0.95.

Results for the bias and MSE are presented in Table 5.1 and Table 5.2, respectively. $Bias0_2$ and $Mse0_2$ represent the bias and MSE for $\widehat{\lambda}_0 = 1/\bar{x}_0$, where $\bar{x}_0 = \sum_{i=1}^{n} x_{0i}$, if the latent unobservable values, $x_{01}, \ldots, x_{0n}$, were actually known. It is important to point out that $\widehat{\lambda}_0$ is not observable. However, if these values were observable, then $\widehat{\lambda}_0$ would be MLE and the best unbiased estimator for λ_0. Therefore, the performance of $\widehat{\lambda}_0$ serves as a good benchmark to compare with all other estimators described in this chapter, which are based on approximations. More precisely, if we denote x_{min_p} to be the minimum of x_1/a_1 up to x_p/a_p for $p = 2, .., 5$, then

$$Bias0_2 = \frac{1}{\overline{x}_0} - \lambda_0 \quad \text{and} \quad Bias1_p = \frac{1}{\overline{x}_{min_p}} - \lambda_0, \qquad \text{respectively,}$$

and the MSE are

$$Mse0_2 = Bias0_2^2 \quad \text{and} \quad Mse1_p = Bias1_p^2.$$

x_{diff} represents the difference between the estimates and the true value of x_0.

x_{diffsq} represents the MSE of the difference.

ρ	$Bias0_2$	$Bias1_2$	$Bias1_3$	$Bias1_4$	$Bias1_5$	x_0	x_{min}	x_{diff}
0.05	-0.011197	-0.574	-0.39937	-0.28855	-0.17101	0.98249	1.23238	0.24989
0.10	0.009905	-0.421	-0.23330	-0.12258	-0.06381	0.99339	1.08727	0.09389
0.20	0.010300	-0.24054	-0.08718	-0.05341	-0.01757	1.01166	1.03754	0.02588
0.30	0.038332	-0.12648	-0.01558	-0.01427	-0.00735	1.02107	1.02898	0.00791
0.40	0.032919	-0.06926	-0.01205	0.01867	0.00780	1.00860	1.01026	0.00166
0.50	0.014231	-0.04644	0.02828	0.02922	0.01344	1.00713	1.00778	0.00065
0.60	0.002170	-0.03026	-0.01516	0.00538	0.00991	1.01039	1.01066	0.00027
0.70	0.017750	0.00292	0.02846	0.03072	0.03005	0.99122	0.99122	0.00001
0.80	0.003247	-0.00332	-0.01022	0.03713	0.00960	1.01068	1.01068	0.00000
0.90	0.0223014	0.02106	0.01377	0.00629	0.02209	0.99810	0.99810	0.00000
0.95	-0.009249	-0.00944	0.03343	0.00253	0.00857	1.00892	1.00892	0.00000

Table 5.1: Table of Bias and Estimation of x_0 for different correlations

From Table 5.1, when the number of variates increasing, the bias reduces in magnitude. That is something we could expect as the prediction of x_0 becomes more accurate with the higher number of variates. In fact, each one gives partial information about x_0.

Also, as ρ increases, the estimate of x_0 becomes more efficient.

It is also observable that the bias becomes satisfactory with higher correlation for the 2, 3, 4 and 5 variates.

We also present the MSE table of the estimates in Table 5.2.

ρ	$Mse0_2$	$Mse1_2$	$Mse1_3$	$Mse1_4$	$Mse1_5$	x_{diffsq}
0.05	0.018739	0.33305	0.16586	0.090732	0.044181	0.45044
0.10	0.019985	0.18484	0.06545	0.030232	0.020013	0.12077
0.20	0.018835	0.07029	0.02061	0.016686	0.019382	0.02934
0.30	0.020357	0.02646	0.01864	0.021683	0.022462	0.00687
0.40	0.020278	0.02136	0.02126	0.020382	0.017916	0.00080
0.50	0.019973	0.01946	0.02365	0.027534	0.022398	0.00022
0.60	0.019373	0.01839	0.01619	0.016020	0.021058	0.00026
0.70	0.031892	0.02974	0.01811	0.029901	0.023984	0.00000
0.80	0.017505	0.01730	0.02042	0.029010	0.020515	0.00000
0.90	0.018066	0.01788	0.01857	0.019959	0.022949	0.00000
0.95	0.015202	0.01517	0.02429	0.027867	0.018026	0.00000

Table 5.2: Table of MSE of x_0 for different correlations

These estimated values show the effectiveness of the proposed estimation techniques developed. As we see from the Table 5.2 of MSE, the difference does appear to be consistently small, although the high values of correlations do appear to give lower MSE's.

The algorithm to the proposed estimation is not difficult to implement, maybe time consuming. We implemented it using SAS® program.

Chapter 6

Conclusion

In this dissertation, we defined and characterized a new multivariate generalized location-scale family of gamma distributions with potential applications in survival and reliability modeling. This family possesses three-parameter gamma marginals (in most cases) and it contains absolutely continuous classes, as well as, the Marshall Olkin type of distributions with a positive probability mass on a set of measure zero. Interestingly, the variables making up the multivariate vector were made linearly related indirectly through a collection of latent random variables and the multivariate distribution is not necessarily restricted to those with gamma marginal distributions. Maximum likelihood estimators and estimators based on the EM algorithm were proposed for the unknown parameters, and, in addition, methods were given to estimate the latent terms in the model.

We have shown that this distribution is shift invariant, closed under finite independent convolutions, and closed under scale transformations. We have also shown that our model has as special cases those models proposed by Mathai and Moschopoulos (1991) [63], Iyer et al. (2002) [42], and Iyer et al. (2004) [43], and corrected some of their omissions.

The possible implication of this work is enormous. It takes into account the non iid properties of real data. Assuming that a_i's are unknown is their structures will add a lot more applications to the model. Further investigations of the shapes parameters will allow a characterizations of more distributions. The estimation techniques can be refined, and that will improve a lot in statistical decision approach. Extending the model to include censoring is an attractive option for several applications.

Bibliography

[1] Abramowitz, M and Stegun, I.A. (1972), *Handbook of Mathematical Functions*, Selected Government Publications, Chapter 6, New York, Dover.

[2] Arnold, B.C. (1967), *A note on multivariate distributions with specified marginals*, Journal of the American Statistical Association, Vol. 62, pp. 1460-1461.

[3] Au, Chi and Tam, Judy (1999), *Transforming Variables using the Dirac Generalized Function*, The American Statistician, Vol. 53, 3, pp. 270-273.

[4] Balakrishnan, N. and Wang, J. (2000), *Simple efficient estimation of the three-parameter gamma distribution,* Journal of Statistical Planning and Inference, 85, pp. 115-126.

[5] Barlow, R.E. and Proschan, F. (1965), *Mathematical Theory of Reliability*, John Wiley & Sons, Library of Congress.

[6] Billingsley, Patrick (1986), *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, 2^{nd} edition.

[7] Borovkov, A.A. and Utev, S.A. (1984), *On an inequality and a related characterization of the normal distribution*, Theory of Prob. Appl. Vol 28, pp. 219-228.

[8] Bowman, K.O. and Shenton, L.R. (1988), *Properties of estimators of the gamma distribution*, New York: Marcel Dekker, Taylor Series.

[9] Bowman, K.O. and Shenton, L.R. (2002), *Problems with Maximum Likelihood Estimation and the 3-Parameter Gamma Distribution*, J. Statist. Comput. Simul. Vol.72(5), pp. 391-401.

[10] Brockwell, P.J. and Davis, R.A. (1996), *Introduction to Time Series and Forecasting,* Springer Texts in Statistics.

[11] Carpenter, M. and Diawara, N. (2006), *A Multivariate Gamma Distribution, and Its Characterizations,* Accepted pending revision J. Mathematical and Management Sciences.

[12] Carpenter, M., Diawara, N. and Han, Yi, (2006), *A New Class of Bivariate Weibull Survival Distributions*, J. Mathematical and Management Sciences, Vol. 26 (1 & 2), pp. 164-184.

[13] Casella, G. and Berger, R.L. (1990), *Statistical Inference*, Duxbury Press 2^{nd} edition.

[14] Chen, L.H.Y. (1982), *An inequality for the multivariate normal distribution*, J. Multivariate Anal., Vol 12, pp 306-315.

[15] Chen, L.H.Y. (1987), *Characterization of probability distributions by Poincare-type inequalities*, Ann. Inst. H. Poincare Sect. B (N.S.) Vol 23, pp. 91-110.

[16] Cohen, A.C. and Whitten, B.J. (1986), *Modified moment estimation for the three-parameter gamma distribution,* J. Quality Technol. 18, pp. 53-62.

[17] Cohen, A.C. and Whitten, B.J. (1988), *Parameter Estimation in Reliability and Life Span Models,*Dekker, New York.

[18] Cohen-Tannoudji, C., Diu, B., and Laloë, F. (1977), *Quantum Mechanics*, Vol. 2, John Wiley & Sons.

[19] Coles, S.G. and Tawn, J.A. (1994), *Statistical Methods for Multivariate Extremes: An Application to structural Design*, Appl. Statist., Vol. 43, pp. 1-48.

[20] Cox, D.R. (1972), *Regresssion Models and Life Tables*, J. Royal Statist. Society, Series B, Vol. 34, pp 187-220.

[21] Csorgo, S. and Welsh, A.H. (1989), *Testing for Exponential and Marshall-Olkin Distributions*, Journal of Statistical Planning and Inferences, 23, 287-300.

[22] Dempster, A., Laird, N. and Rubin, D. (1977), *Maximum Likelihood from Incomplete Data via the EM algorithm,*, Journal of the Royal Statistical Society, Series B, Vol 39 (1), pp 1-38.

[23] Dudewicz, E.J. and Mishra, S. (1988), *Modern Mathematical Statistics*, Wiley Series in Probability and Mathematical Statistics.

[24] Fang, K. and Zhang, Y. (1990), *Generalized Multivariate Analysis,* Springer.

[25] Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, Second Edition, Wiley Series in Probability and Mathematical Statistics.

[26] Folland, G.B. (1999), *Real Analysis: Modern Techniques, and Their Applications*, Wiley-Interscience Series.

[27] Friedl, H. and Kauermann, G. (2000), *Standard errors for EM estimates in generalized linear models with random effects,* Biometrics, Vol. 56, pp. 761-767.

[28] Furman, E. and Landsman, Z. (2005), *Risk capital decomposition for a multivariate dependent gamma portfolio,* Insurance: Mathematics and Economics, 2005, Vol. 37, issue 3, pages 635-649.

[29] Gaver, D.P., and Lewis, P.A.W., (1980), *First-order autoregressive gamma sequences and point processes*, Adv. appl. Prob, 12, 727-745.

[30] Ghosh, S.K. and Gelfand, A.E. (1998), *Latent Waiting Time Models for Bivariate Event Times with Censoring*, Sankya, Vol. 60, Series A, pp. 31-47.

[31] Grove, D.A. and Coddington, P.D. (2005), *Analytic Models of Probability Distributions for MPI Point-to-Point Communication Times on Distributed Memory Parallel Computers,* Proc. 6^{th} International Conference on Algorithms and Architectures for Parallel Processing, Melbourne, Australia.

[32] Hanagal, D.D. (1996), *A multivariate Weibull distribution*, Economic Quality Control, 11, pp. 193-200.

[33] He, W. and Lawless, J.F. (2005), *Bivariate location-scale models for regression analysis, with applications to lifetime data,* J. R. Statist. Soc. B, Vol. 67, part 1, pp. 63-78.

[34] Henderson, R. and Shimakura, S. (2003), *A Serial Correlated Gamma Frailty Model for Longitudinal Count Data*, Biometrika, 90, 2, pp. 355-366.

[35] Hougaard, Philip (1986), *A class of multivariate failure time distributions*, Biometrika, 73, 3, pp. 671-678.

[36] Hougaard, Philip (2000), *Analysis of Multivariate Survival Data,* Springer.

[37] Hougaard, Philip (2001), *The normality assumption: Pro and Contra*, Biometrika, 73, 3, pp. 671-678.

[38] Hunter, J.K., and Nachtergaele, B. (2001), *Applied Analysis*, River Edge, NJ, World Scientific.

[39] Hwang, T.Y. and Hu, C.Y. (1999), *On a characterization of the gamma distribution: the independence of the sample mean and the sample coefficient of variation*, Ann. Inst. Statist. Math., 51, pp. 749-753.

[40] Hwang, T.Y. and Hu, C.Y. (2000), *On some characterizations of population distributions,* Taiwanese Journal of Mathematics, Vol 4, No 3, pp. 427-437.

[41] Hwang, T.Y. and Huang, P.H. (2002), *On new moment estimation of parameters of gamma distribution using its characterization*, Ann. Inst. Stat. Math., 54, pp. 840-847.

[42] Iyer, Srilanth K., Manjunath, D. and Manivasakan, R. (2002), *Bivariate Exponential Distributions Using Linear Structures*, Sankhya, V.64, Series A, pp. 156-166.

[43] Iyer, Srilanth K. and Manjunath, D. (2004), *Correlated Bivariate Sequence for queueing and reliability applications*, Communications in Statistics, Vol. 33, pp. 331-350.

[44] Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall.

[45] Johnson, N. and Kotz, S. (1970), *Continuous Univariate Distributions-1* Wiley, New York.

[46] Johnson, N., Kotz, S. and Kemp, A. (1993), *Univariate Discrete Distributions*, 2^{nd} edition Wiley Series in Probability and Mathematical Statistics.

[47] Johnson, N., Kotz, S., Balakrishnan, N. (1997), *Discrete Multivariate Distributions*, Wiley Series in Probability and Statistics.

[48] Johnson, R.A. and Wichern, D.W. (1998), *Applied Multivariate Statistical Analysis*, Prentice Hall, 4^{th} edition.

[49] Jorgensen, B. (1987), *Exponential Dispersion Models,* J. R. Statist. Soc. B, Vol. 49, No. 2, pp. 127-162.

[50] Kallenberg, O. (1986), *Random Measures,* 4^{th} Edition, Akademie-Verlag Berlin.

[51] Kallenberg, O. (2002), *Foundations of Modern Probability,* 2^{nd} Edition, Springer.

[52] Khuri, Andre (2004), *Applications of Dirac's Delta Function in Statistics*, Int. J. Math. Educ. Sci. Technol., 35, 2, pp. 185-195.

[53] Kotz, S., Balakrishnan, N. and Johnson, N. (2000), *Continuous Multivariate Distributions*, Volume 1 Wiley Series in Probability and Statistics.

[54] Krzanowski, W.J. and Marriot, F.H.C. (1994), *Multivariate Analysis: Part 1: Distributions, Ordination and Inference,* Edward Arnold, Libary of Congress.

[55] Lawrence, A.J. (1978), *Some Autoregressive Models for Point Processes*, Colloquia Mathematica Societatis János Boylai.

[56] Lawless, J. F. (2003), *Statistical Models and Methods for Lifetime Data*, 2^{nd} edition Wiley Series in Probability and Statistics.

[57] Lee, Larry (1979), *Multivariate Distributions having Weibull Properties*, Journal of Multivariate Analysis, 9, pp. 267-277.

[58] Lehmann, E.L. (1986), *Testing Statistical Hypotheses,* 2^{nd} edition, New York, Chapman & Hall.

[59] Lu, Jye-Chyi and Bhattacharyya, G.K. (1990), *Some New Constructions of Bivariate Weibull Models*, Ann. Inst. Statist. Math., 43(3): pp. 543-559.

[60] Mardia, K.V. (1970), *Measures of multivariate skewness and kurtosis with applications,* Biometrika, Vol. 57., pp. 519-530.

[61] Marshall, A.W. and Olkin, I. (1967), *A Multivariate Exponential Distribution*, J. Amer. Stat. Assoc., 63: pp. 30-44.

[62] Marshall, A.W. and Olkin, I. (1983), *Domains of Attraction of Multivariate Extreme Value Distributions*, The Annals of Probability, 111, pp. 168-177.

[63] Mathai, A.M. and Moschopoulos, P.G. (1991), *On a Multivariate Gamma*, Journal of Multivariate Analysis, 39, pp. 135-153.

[64] Mathai, A.M. and Moschopoulos, P.G. (1992), *A Form of Multivariate Gamma Distribution*, Annals of the Institute of Statistical Mathematics, 44, pp. 97-106.

[65] Mathai, A.M. and Saxena, R.K. (1978), *The H-function with Applications in Statistics and Other Disciplines*, Wiley New York.

[66] McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models,* Chapman and Hall: London.

[67] McLachlan, G.L. and Peel, D. (2000), *Finite Mixture Models,* New York, Wiley.

[68] Minhajuddin, A.T.M., Harris, I.R., and Schucany, W.R. (2003), *Simulating Multivariate Distributions with Specific Correlations,* J. Statist. Comp. Simul., Vol. 74, pp. 599-607.

[69] Moran, P.A.P. (1969), *Statistical Inference with Bivariate Gamma Distributions,* Biometrika, Vol. 56, No.3, pp. 627-634.

[70] Moschopoulos, P.G. (1985), *The Distribution of the Sum of Independent Gamma Random Variables,* Annals of the Institute of Statistical Mathematics, Vol. 37, pp. 541-544.

[71] Moschopoulos, P.G. and Staniswalis, J. G. (1994), *Estimation Given Conditionals from an Exponential Family,* The American Statistician, Vol. 48, No. 4, pp. 271-275.

[72] Nadarajah, S. and Kotz, S. (2005), *On the Linear Combinations of Exponential and Gamma Random Variables,* Entropy, Vol 7[2], pp. 161-171.

[73] Pazman, A. and Pronzato, L. (1996), *A Dirac-function method for densities of nonlinear statistics and for marginals densities in nonlinear regression,* Statistics and Probability Letters, Vol 26, pp 159-167.

[74] Rudin, W. (1976), *Principles of Mathematical Analysis*, McGraw-Hill.

[75] Shilov, G.E. and Gurevich, B.L. (1977), *Integral, Measure, and Derivative: a Unified Approach,* New York, Dover Publications.

[76] Sivazlian, B.D. (1981), *On a Multivariate Extension of the Gamma and Beta Distributions*, SIAM Journal on Applied Mathematics, Vol. 41, No. 2, pp. 205-209.

[77] Tanner, M.A. (1996), *Tools for Statistical Inference: Methods for the Expectation of Posterior Distributions and Likelihood Functions,* 3^{rd} edition, New York: Springer.

[78] Vaupel, J.W., Manton, K.G., and Stallard, E. (1979), *The impact of heterogeneity in individual frailty on the dynamics of mortality*, Demogarphy, 16(3), pp. 439-454.

[79] Walker, S.G. and Stephens, D.A. (1999), *A Multivariate Family of Distributions on* $(0, \infty)^p$, Biometrika, Vol. 86, No. 3, pp. 703-709.

[80] Williamson, J.H. (1962), *Lebesgue Integration*, New York, Holt, Rinehart and Winston, Athena Series.