Mathematics & Statistics Theses & Dissertations

Mathematics & Statistics

Summer 2010

# Semi-Parametric Likelihood Functions for Bivariate Survival Data

S. H. Sathish Indika
*Old Dominion University*

Follow this and additional works at: https://digitalcommons.odu.edu/mathstat_etds

Part of the Mathematics Commons, and the Statistics and Probability Commons

# SEMI-PARAMETRIC LIKELIHOOD FUNCTIONS FOR BIVARIATE SURVIVAL DATA

by

S. H. Sathish Indika
BS Mathematics, 1997, University of Colombo
MS Mathematics-Statistics, 2002, New Mexico Institute of Mining and Technology
MS Mathematical Sciences-Operations Research, 2004, Clemson University
MS Computer Science, 2006, College of William and Mary

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS AND STATISTICS

OLD DOMINION UNIVERSITY
August 2010

Approved by:

Norou Diawara (Director)

Dayanand N. Naik

Larry D. Lee

Cynthia M. Jones

# ABSTRACT

# SEMI-PARAMETRIC LIKELIHOOD FUNCTIONS FOR BIVARIATE SURVIVAL DATA

S. H. Sathish Indika
Old Dominion University, 2010
Director: Dr. Norou Diawara

Because of the numerous applications, characterization of multivariate survival distributions is still a growing area of research. The aim of this thesis is to investigate a joint probability distribution that can be derived for modeling nonnegative related random variables. We restrict the marginals to a specified lifetime distribution, while proposing a linear relationship between them with an unknown (error) random variable that we completely characterize. The distributions are all of positive supports, but one class has a positive probability of simultaneous occurrence. In that sense, we capture the absolutely continuous case, and the Marshall-Olkin type with a positive probability of simultaneous event on a set of measure zero. In particular, the form of the joint distribution when the marginals are of gamma distributions are provided, combining in a simple parametric form the dependence between the two random variables and a nonparametric likelihood function for the unknown random variable. Associated properties are studied and investigated and applications with simulated and real data are given.

# ACKNOWLEDGMENTS

I would like to give special acknowledgement to my advisor Dr. Norou Diawara for his valuable guidance, support and constant attention throughout my graduate studies. I also would like to extend my sincere appreciation to my Ph.D. committee members, Dr. Dayanand N. Naik, Dr. Larry D. Lee, and Dr. Cynthia M. Jones, for their help and support. I would like to thank my father and the rest of the family members for all their support and encouragement. Finally, I would like to thank all my friends, both in Sri Lanka and in the United States for all their help, support and encouragement.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER I

# INTRODUCTION

## I.1  PROBLEM DESCRIPTION

It is in the nature of many processes that are linked that a number of predictors must be considered simultaneously in modeling. Assumptions made in the structural analysis also effect the quality of the procedure and outcome result. The exponential family of distributions is very useful and is often used for modeling phenomena in life testing, reliability, and other types of engineering applications. Bivariate exponential distributions, with exponential marginal densities have been known in the literature for some time. However, the majority of these models have been theoretically motivated rather than emphasizing applicability. A notable exception was the "fatal shock model" of Marshall and Olkin [32], where the model links two non negative random variables (rv's) with an unobservable latent random variable (rv) via the proportional hazard model. It is also important to note that the modeling and analysis based on the classical bivariate normal distribution is well established and widely used. However, the bivariate normal theory is inadequate for modeling positive support distributions. Hougaard [19] discusses the theoretical advantages of normal distribution theory as well as the situations where positive support distributions are of relevance. Therefore, many bivariate distributions such as exponential, Weibull and gamma, have been proposed in the literature to study many diseases or behavioral problems, with associated non negative rv's. Traditionally, the two variables in a system or model have been assumed to be independent therefore ignoring the dependence between these variables. This is termed as the "working independence" assumption in Lawless [29]. We will show, using Mean Square Errors, Bias, and other objective criteria, that the independence assumption comes at a greater cost than the proposed procedure in estimation of the parameters and in the model validity as the estimates lose asymptotic properties.

The problem that describes the association in disease occurrences has been difficult to explain. Many diseases and events are linked, and interrelated in their action/reaction. If one considers pairs of organs, there are analytic difficulties in describing the joint distribution of the two disease events. One reason is that the

distribution has a discrete and continuous part. The discrete part is motivated by the fact that the two diseases can occur exactly at the same time, (i.e. simultaneous) or proportional to each other. The continuous part described the association between the two rv's with positive continuous common dependence on an unobserved rv. The analysis of bivariate data then presents difficulties when one has to incorporate the discrete and continuous parts, and find estimates. Various models of associations between occurrence of events are motivated (Iyer and Manjunath [22]). In the medical field, such phenomena is known as co-morbidity condition, where many diseases can occur together. For example, diabetes, that affects millions of people in the world, can lead to blindness, kidney failure, stroke, amputations and an increased risk of cardiovascular disease. It occurs when the body does not produce enough insulin or cannot properly utilize it. This makes it difficult for blood sugar to enter the body cells, and it is the seventh leading cause of death in the United States. Other famous example is that of asthma and HIV/AIDS which can occur together with other diseases. Recent studies have confirmed that asthma-related mortality is increasing. Patients with asthma represent a considerable challenge for physicians because of high risks of asthma related morbidity and mortality as reported in Holgate et al. [17], and in Crane et al. [9] for examples. It is then of interest to develop a model distribution that can be used to study the growth and interrelationships between many diseases and phenomena. In that sense, such analysis has also potential applications in HIV/AIDS and other diseases as well. Attention Deficit Hyperactivity Disorder (ADHD) is another example of disease that does not travel alone. Usually depression, anxiety, unconducive behavior, are other conditions found during or after diagnosis of ADHD. Our methods will have applications in this field also. In the past, analysis of co-morbidity has been done using the multivariate normal distribution, and assuming independence between the simultaneous occurring events. Each of those events can be described in terms of lifetime distribution. However, because in such problems, the response variables and the error models have only positive support and non-normal behavior (such as asymmetry), the multivariate normal theory fails. Hougaard [19] discusses the cases where the multivariate normal model is not appropriate. The approximation to normal theory is also not reliable since there is no primarily large data. Optimal parameters need to be developed, and the diabetes example is one case that shows the importance of detailed description, characterization of the inter-relationship between events. Data transformation to better approximate

with normal distribution is not appropriate if a more suitable theoretical model can be found, and/or the sample sizes are not large enough. Accordingly, our goal is to establish a large and flexible class of bivariate gamma distribution models that contain both absolutely continuous and discontinuous distributions on the positive hypercube. We consider the class of bivariate distribution functions for two observed dependent events, linked through an unobserved rv. A linear relationship is considered between the two events first with a nonzero probability of simultaneous or proportional occurrence. The discontinuous part provides a great advantage as there are situations where two events occur at the same time or where one event is proportionally related to the other event. All these models will have applications in survival and reliability analysis, and they will allow us to study the growth and the relationship between diseases. The study of interrelationship between various diseases will involve developing a linear model among positive rv's with a random effect. The estimation of the parameters associated with the rv's will be given. The estimation is computationally intensive, and we propose to develop computational methods using algorithm to handle this problem. Thus, developed bivariate probability models and innovative computing tools will enable us to understand the growth and interrelationships between events that can occur simultaneously or linearly. Through these models and their analysis, we will be able to compute more accurately the probability of survival of an individual when suffering from related disease events.

## I.2  LITERATURE REVIEW

Many authors have considered the univariate exponential distribution in modeling phenomena in life testing, reliability, and other types of engineering applications. The exponential rv is nonnegative and has desirable properties such as memorylessness and constant hazard function. The need for bivariate/multivariate models arise, whenever there is the need to model two or more variables in a system. But in many analysis, these variables were assumed to be independent. For example in queueing theory applications the inter arrival times and the service times were traditionally assumed to be independent. This is not a realistic assumption particularly in packet communication networks. Also in reliability analysis, the failure of one component in a multi component system influences the lifetime of the other component(s).

Marshall and Olkin [32] introduced their classical bivariate exponential model,

with exponential marginal densities, in which they considered the case where one component of a two component system fails after receiving a fatal shock. They termed this as the "fatal shock model". It is assumed that the occurrence of shocks to these two components are governed by three independent Poisson processes. The Poisson processes $Z_1(t; \lambda_1)$, $Z_2(t; \lambda_2)$ and $Z_{12}(t; \lambda_{12})$ governs the shocks to the first component, shocks to the second component and the simultaneous shocks to both components. Subsequently, many authors have developed bivariate/multivariate models to account the dependencies in a system model. Steel and Roux [46] have considered the lifetimes of the two components $C_1$ and $C_2$ operating in a system. These two components $C_1$ and $C_2$ are subjected to shocks and it is assumed that $C_1$ fails after receiving $h$ shocks and $C_2$ fails after receiving $l$ shocks. The occurrence of the shocks in the two components, $C_1$ and $C_2$, are governed by a Poisson process with parameters $1/\alpha_1$ and $1/\alpha_2$, respectively. If $C_1$ fails in the first instance, then it is assumed that the occurrence of the subsequent shocks in $C_2$ is governed by a Poisson process with parameters $\lambda_2/\alpha_2$, whereas, if $C_2$ fails in the first instance, then it is assumed that the occurrence of the shocks in $C_1$ is governed by a Poisson process with parameters $\lambda_1/\alpha_1$. Mathai and Moschopoulos [35] discussed two cases of applications of their multivariate gamma. The first case is where a $k$-variate system $(Z_1, \ldots, Z_k)$ is subjected to disturbances such that the new system is $(Z_1 + \varepsilon_1, \ldots, Z_k + \varepsilon_k)$, where $Z_j$'s and $\varepsilon_j$'s are mutually independent gamma rvs. The second case is where the rv's $(Z_1, \ldots, Z_k)$ represent the runoffs to a dam from $k$ different streams and the new random components, $(Z_1 + \delta_1 X, \ldots, Z_k + \delta_k X)$, where $\delta_j$'s are constants and $X$ is a new gamma rv which is independent of $(Z_1, \ldots, Z_k)$ and represents the variation in the dam from rainfall from the catchment areas. Mathai and Moschopoulos [36] also introduced a new form of multivariate gamma that can be applied in reliability and stochastic processes where partial sums of independent positive rv's are of importance. Csorgo and Welsh [10] reiterated the need of bivariate distributions that can be used to model the failure of paired components such as aircraft engines, paired organs such as eyes, kidneys while developing a test for the Marshal-Olkin distribution.

## I.3 SOLUTION METHODOLOGY

Using likelihood methods, our aim is to develop multivariate probability models and find tools to quantify the interrelationships between various events that can

occur simultaneously in human beings during development or across the span of a related diseases and events process. Through these models and the related statistical analysis, more accurate computation of the probability of survival of an individual will be given when suffering from many interrelated diseases or developing significant events in individuals. The model parameters will be estimated and their improvement from the working independence models will be assessed. Old approach based on the working independence leads to unreliable estimates, which may have particularly grave consequences in the case of assessing co-morbidity. The alternative is to adopt a general class of distributions which is flexible and analytically tractable in describing the related types of events. The new approach we propose offers considerable insight in understanding the relationship between these related events. The interaction between related events is taken into account, hence including the stochastic processes highly non-Gaussian and computationally challenging. Another focus is on getting efficient estimation in model parameters with an emphasis on non- and semi-parametric regression models, on bivariate models with partial knowledge about the marginals. In that sense, fully imputed estimators will be studied. We show that our methodology suggests that the joint estimatior procedure is typically better than the available estimator which only assumes independent events. We also show that fully implemented model related estimators will be more efficient than the proposed ones from independence assumption. The work on bivariate models will deal with fine-tuning results for known and equal marginals obtained so far, and on extending such results to more general models including those in which the marginal distributions are linked through a parameter.

## I.4  ORGANIZATION OF THE DISSERTATION

In this dissertation we consider two rv's $X_1$ and $X_2$ which have specified marginal distributions. A linear relationship is defined between them with the aide of an unknown latent rv, $Z$. We next investigate the bivariate density with a non-zero probability of simultaneous occurrence. More precisely, we set

$$X_2 = aX_1 + Z, \text{ where } a > 0, \text{ is a fixed constant.} \tag{1}$$

In that sense, there is a non zero probability of simultaneous or proportional occurrence on a set of measure zero. Carpenter and Diawara [6] described the forms of the parameters associated with $X_1$ and $X_2$ when these latter are of the exponential

type distributions. In this dissertation, we extend the result for the gamma type distributions.

In Chapter III, the mean of $Z$ is obtained by taking the difference of $X_2$ and $aX_1$ The sum of independent distribution is used to describe the bivariate model. To that end, Chapter IV describes such sums including the case where distributions are discrete and continuous. Properties of the random sum are presented and will be relevant in subsequent chapters. Chapter V describes the bivariate model when the marginal distributions are gamma, Weibull and exponential. Having a direct known density allows one to develop simulated results where parameters can be verified. Applications with simulated and real data are then presented.

# CHAPTER II

# BACKGROUND AND PRELIMINARIES

In this chapter, we review the mathematical concepts and the notations that will be used in the rest of the thesis. We review the univariate gamma and Weibull distributions (See Johnson and Kotz [23] and Casella and Berger [7] , the multivariate extensions of those distributions, the notion of Laplace transforms (See Feller [14]). Also the application of the Dirac delta function in Statistics (See Au and Tam [3], and Khuri [26]) is reviewed.

## II.1   UNIVARIATE GAMMA AND WEIBULL DISTRIBUTIONS

Positive skewed distributions with nonnegative support occur quite often in practical applications such as in reliability and survival analysis. The gamma family of distributions is one such with a heavy right tail. It is flexible and widely used in reliability and survival analysis. The univariate three parameter gamma distribution denoted here as $Ga(\mu, \lambda, \alpha)$, also known as Type III of distributions in the Pearson's system of distributions. It's probability density function (pdf) is defined as:

$$f_X(x; \mu, \lambda, \alpha) = \frac{\lambda^\alpha}{\Gamma(\alpha)} (x - \mu)^{\alpha-1} e^{-\lambda(x-\mu)} I_{[\mu, \infty)}(x), \tag{2}$$

where

- $\mu \in \mathbb{R}$, $\quad \lambda > 0$, and $\alpha > 0$ are the location, scale and shape parameters, respectively.

- $\Gamma(\alpha) = \displaystyle\int_0^\infty t^{\alpha-1} e^{-t} dt$ is called the gamma function.

- when $\alpha = 1$ $\quad$ and $\quad \mu = 0$, the gamma reduces to the simple exponential distribution with parameter $\lambda$.

- when $\alpha$ is an integer value, the gamma distribution is also called the Erlang distribution. Barlow and Proschan [4] present interesting applications of the Erlang distributions.

Sometimes (2) is parameterized as a function of $1/\lambda$ instead of $\lambda$. In this dissertation, we will use the parametrization $Ga(\mu, \lambda, \alpha)$. In the next section, the properties

of this distribution will be described. Johnson and Kotz [23] gives a comprehensive account of this distribution, including its characterization and estimation of its parameters. The two-parameter gamma distribution with location parameter, $\mu = 0$, scale parameter $\lambda$ and shape parameter $\alpha$, is a particular form of (2). Here also we denote this as $Ga(\lambda, \alpha)$. It's pdf is defined as:

$$f_X(x; \lambda, \alpha) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} I_{[0,\infty)}(x), \quad \lambda > 0, \text{ and } \alpha > 0. \tag{3}$$
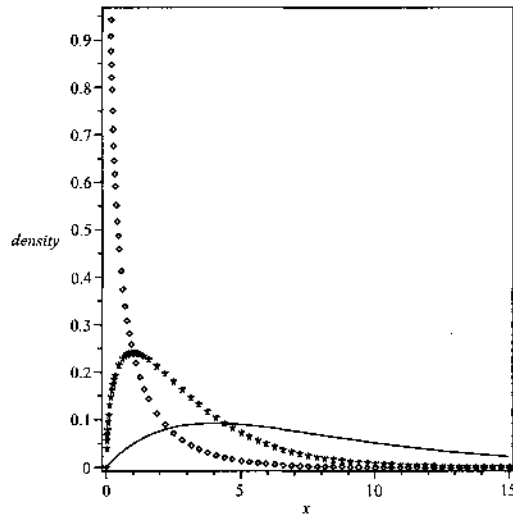


FIG. 1: The graph of the gamma pdf with $\mu = 0$: diamond $\lambda = 0.5$, $\alpha = 0.5$; asterisk: $\lambda = 0.5$, $\alpha = 1.5$; line: $\lambda = 0.25$, $\alpha = 2.0$.

The shape parameter $\alpha$ explains when the hazard function is increasing ($\alpha - 1 > 0$), decreasing ($\alpha = 1$), or constant ($\alpha = 1$). The $k^{th}$ moment of (2) is:

$$\begin{aligned}
E(X^k) &= \int_\mu^\infty x^k f_X(x; \mu, \lambda, \alpha)\, dx, \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_\mu^\infty x^k (x - \mu)^{\alpha-1} e^{-\lambda(x-\mu)}\, dx.
\end{aligned}$$

By change of variables, we get

$$
\begin{aligned}
E(X^k) &= \frac{1}{\Gamma(\alpha)} \int_0^\infty \left(\mu + \frac{t}{\lambda}\right)^k t^{\alpha-1} e^{-t}\, dt, \\
&= \frac{1}{\lambda^k \Gamma(\alpha)} \sum_{i=0}^k \binom{k}{i} (\lambda\mu)^i \int_0^\infty t^{k+\alpha-i-1} e^{-t}\, dt, \\
&= \frac{1}{\lambda^k \Gamma(\alpha)} \sum_{i=0}^k \binom{k}{i} (\lambda\mu)^i \Gamma(k+\alpha-i).
\end{aligned}
$$

In particular, the first and the second moments are:

$$
\begin{aligned}
E(X) &= \frac{\alpha}{\lambda} + \mu \text{ and} \\
E(X^2) &= \frac{\alpha(\alpha+1)}{\lambda^2} + \frac{2\mu\alpha}{\lambda} + \mu^2.
\end{aligned}
$$

The $k^{th}$ moment of (2) can also be written as:

$$
E(X^k) = \frac{\Gamma(k+\alpha)}{\lambda^k \Gamma(\alpha)} + \frac{1}{\lambda^k \Gamma(\alpha)} \sum_{i=1}^k \binom{k}{i} (\lambda\mu)^i \Gamma(k+\alpha-i).
$$

Therefore, when $\mu = 0$,

$$
E(X^k) = \frac{\Gamma(k+\alpha)}{\lambda^k \Gamma(\alpha)}.
$$

Therefore, the expected value and the variance of the three parameter gamma distribution in (2) are:

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

respectively.

The cumulative distribution function (cdf) and the survival function are as follows:

$$
\begin{aligned}
F_X(x) &= \int_\mu^x f(t)\, dt = \frac{\gamma(\alpha, \lambda(x-\mu))}{\Gamma(\alpha)}, \\
S_X(x) &= 1 - F_X(x),
\end{aligned}
$$

where $\gamma(\alpha, x) = \int_0^x t^{\alpha-1} e^{-t}\, dt$ is the incomplete gamma function as in Casella and Berger [7].

With respect to the model defined in (1), equating the moment of $Z$ as the

difference of the moments of $X_2$ and $aX_1$ will illuminate the unknown parameters of $Z$.

The gamma pdf is closed under convolution provided the scale parameter $\lambda$ is common to each of the independent gamma rvs. It is also closed under scalar multiplication. Specifically, if $X \sim Ga(\mu, \lambda, \alpha)$, then $Y = cX \sim Ga(c\mu, \frac{\lambda}{c}, \alpha)$, for all $c > 0$. In particular, consider the rv $X$ with $\mu = 0$ and $\alpha = 1$. It's pdf is:

$$f_X(x; \lambda) = \lambda x e^{-\lambda x} I_{[0,\infty)}(x). \tag{4}$$

The pdf in (4) is called the exponential pdf. Now consider the transformation $Y = X^{\frac{1}{\beta}}$, $\beta > 0$. $Y$ is called Weibull rv and its pdf is:

$$f_Y(y; \lambda, \beta) = \lambda e^{-\lambda y^\beta} \beta y^{\beta-1} = \lambda \beta\, y^{\beta-1}\, e^{-\lambda y^\beta},\ y > 0. \tag{5}$$

The Weibull distribution is also another example of a commonly used lifetime distribution. A rv is called stable (Feller [14]) if it can be written as a sum of independent copies of that same family distribution. Figure 2 describes the Weibull pdf for different parameter values.
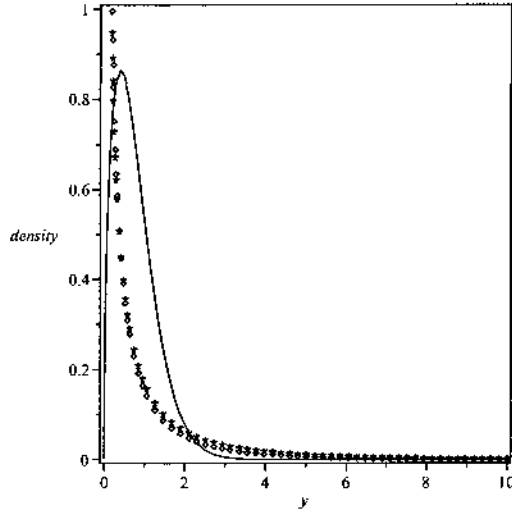


FIG. 2: The graph of the Weibull pdf diamond $\lambda_1 = 1.25$, $\beta_1 = 1.5$; asterisk: $\lambda_2 = 1.5$, $\beta_2 = 0.5$; line: $\lambda_3 = 1.75$, $\beta_3 = 0.5$.

The $k^{th}$ moment of (4) is:

$$
\begin{aligned}
E(Y^k) &= \int_0^\infty y^k f_Y(y; \lambda, \beta)\, dy, \\
&= \lambda\beta \int_0^\infty y^{\beta+k-1} e^{-\lambda y^\beta}\, dy.
\end{aligned}
$$

By change of variables we get

$$
\begin{aligned}
E(Y^k) &= \frac{1}{\lambda^{k/\beta}} \int_0^\infty t^{k/\beta+1-1} e^{-t}\, dt, \\
&= \frac{1}{\lambda^{k/\beta}} \Gamma\left(1 + \frac{k}{\beta}\right).
\end{aligned}
$$

In particular, the first and the second moments are:

$$
\begin{aligned}
E(Y) &= \frac{1}{\lambda^{1/\beta}} \Gamma\left(1 + \frac{1}{\beta}\right), \text{ and} \\
E(Y^2) &= \frac{1}{\lambda^{2/\beta}} \Gamma\left(1 + \frac{2}{\beta}\right).
\end{aligned}
$$

Therefore expected value and the variance of the two parameter Weibull distribution in (5) are:

$$
E(Y) = \frac{1}{\lambda^{1/\beta}} \Gamma\left(1 + \frac{1}{\beta}\right) \quad \text{and}
$$
$$
Var(Y) = \frac{1}{\lambda^{2/\beta}} \left[ \Gamma\left(1 + \frac{2}{\beta}\right) - \left\{ \Gamma\left(1 + \frac{1}{\beta}\right) \right\}^2 \right],
$$

respectively.

## II.2   THE MULTIVARIATE GAMMA

The random vector components corresponding to a multivariate lifetime distribution have positive support distributions. There is a difficulty in generalizing the univariate gamma distributions to the multivariate case. There should be a mechanism to derive a distribution that can be categorized as multivariate gamma. In other words, the dependence structure of the rv's have to be specifically defined. Different approaches and models have been suggested for developing bivariate/multivariate

gamma distributions. Kotz et al. [27] have extended the standard gamma distribution to the multivariate case. In their construction, they have considered $(m + 1)$ mutually independent standard gamma rv's, $X_j, j = 0, \ldots, m$ and considered the linear relationship $Y_j = X_0 + X_j, j = 1, \ldots, m$. The joint distribution of $Y_1, \ldots, Y_m$ is defined to be multivariate gamma. Its joint pdf for $m = 2$ is:

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{e^{-(y_1 + y_2)}}{\Gamma(\theta_0)\Gamma(\theta_1)\Gamma(\theta_2)} \int_0^{\tilde{y}} x_0^{\theta_0 - 1}(y_1 - x_0)^{\theta_1 - 1}(y_2 - x_0)^{\theta_2 - 1} e^{x_0} \, dx_0,$$

where $\tilde{y} = min(y_1, y_2)$.

The multivariate gamma distribution of Krishnaiah and Rao [28] is defined as the joint distribution of the diagonal elements of a Wishart matrix. They have defined it as the multivariate chi-square distribution. They first considered a $p$-variate normally distributed random vector, $X_u = (X_{1u}, \ldots, X_{pu}), u = 1, \ldots, n$, with mean vector zero and variance-covariance matrix $M$, where $M = (\sigma_{ij})$. Defining $Z_i = \sum_{u=1}^{n} X_{iu}^2, i = 1, \ldots, p$, the marginal distribution of $Z_i$ is gamma with scale parameter $2\sigma_{ii}$ and shape parameter $n/2$. Moran [40] utilized the three parameter gamma distribution in discussing the need for multivariate models in rainmaking experiments. Steel and Roux [46] have introduced bivariate gamma models in reliability analysis, where the components of operating systems are subjected to shocks. They considered $X_1$ and $X_2$ to be the lifetimes of the two components $C_1$ and $C_2$ operating in a system. These two components $C_1$ and $C_2$ are subjected to shocks and it is assumed that $C_1$ fails after receiving $h$ shocks and $C_2$ fails after receiving $l$ shocks. The occurrence of the shocks in the two components, $C_1$ and $C_2$, are governed by a Poisson process with parameters $1/\alpha_1$ and $1/\alpha_2$ respectively. If $C_1$ fails in the first instance, then it is assumed that the occurrence of the subsequent shocks in $C_2$ is governed by a Poisson process with parameters $\lambda_2/\alpha_2$, whereas, if $C_2$ fails in the first instance, then it is assumed that the occurrence of the shocks in $C_1$ is governed by a Poisson process with parameters $\lambda_1/\alpha_1$. The joint pdf of $X_1$ and $X_2$ is:

$$f_{X_1, X_2}(x_1, x_2) = \begin{cases} \frac{\lambda_2 x_1^{h-1}(\lambda_2(x_2 - x_1) + x_1)^{l-1}}{\Gamma(h)\Gamma(l)\alpha_1^h \alpha_2^l} e^{[-(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} - \frac{\lambda_2}{\alpha_2})x_1 - \frac{\lambda_2}{\alpha_2}x_2]}, & \text{if } 0 < x_1 < x_2; \\ \frac{\lambda_1 x_2^{l-1}(\lambda_1(x_1 - x_2) + x_2)^{h-1}}{\Gamma(h)\Gamma(l)\alpha_1^h \alpha_2^l} e^{[-\frac{\lambda_1}{\alpha_1}x_1 - (\frac{1}{\alpha_1} + \frac{1}{\alpha_2} - \frac{\lambda_1}{\alpha_1})x_2]}, & \text{if } 0 < x_2 < x_1. \end{cases}$$

In their multivariate gamma model Mathai and Moschopoulos [35] considered a linear combination of independent gamma rvs. In their model, they defined $V_i \sim Ga(\gamma_i, \beta_i, \alpha_i), i = 0, \ldots, k$, to be mutually independent and considered the linear relationship $Z_i = \frac{\beta_i}{\beta_0} V_0 + V_i, i = 1, \ldots, k$. The random vector $(Z_1, \ldots, Z_k)$ is then defined to be the multivariate gamma distribution. They have discussed two cases that can be used to model their multivariate gamma. The first case is where a $k$-variate system $(Z_1, \ldots, Z_k)$ is subjected to disturbances such that the new system is $(Z_1 + \varepsilon_1, \ldots, Z_k + \varepsilon_k)$, where $Z_j$'s and $\varepsilon_j$'s are mutually independent gamma rvs. The second case is where the rv's $(Z_1, \ldots, Z_k)$ represent the runoffs to a dam from $k$ different streams and the new random components, $(Z_1 + \delta_1 X, \ldots, Z_k + \delta_k X)$, where $\delta_j$'s are constants and $X$ is a new gamma rv which is independent of $(Z_1, \ldots, Z_k)$ and represents the variation in the dam from rainfall from the catchment areas. Mathai and Moschopoulos [36] also introduced a new form of multivariate gamma that can be applied in reliability and stochastic processes where partial sums of independent positive rv's are of importance. Here also they considered $V_i \sim Ga(\gamma_i, \beta, \alpha_i), i = 1, \ldots, k$, to be mutually independent and considered their partial sums, $Z_i = Z_{i-1} + V_i, i = 1, \ldots, k$ with $Z_0 = 0$, to construct the multivariate gamma distribution. Their joint pdf for $k = 2$ is:

$$f_{Z_1, Z_2}(z_1, z_2) = \frac{(z_1 - \gamma_1)^{\alpha_1 - 1} (z_2 - z_1 - \gamma_2)^{\alpha_2 - 1}}{\beta^{\alpha_1 + \alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)} e^{-\frac{(z_2 - (\gamma_1 + \gamma_2))}{\beta}}.$$

By considering a Negative Binomial randomizing procedure, Gaver [16] constructively generated a bivariate gamma with gamma marginal distributions. He considered the rv's $X_1$ and $X_2$, with the following relationship: $X_1 = X_{1n} + X_{1k}$, $X_2 = X_{2n} + X_{2k}$, where $X_{1n}, X_{1k}, X_{2n}, X_{2k}$ are mutually independent gamma rv's with shape parameters $n, k, n, k$ and unit scale parameters. The joint Laplace-Stieltjes transform (JLST) of this bivariate gamma density is:

$$\psi_k(s_1, s_2) = \left[ \frac{\alpha}{(1 + \alpha)(1 + s_1)(1 + s_2 - 2) - 1} \right]^k, \ \forall k > 0,$$

where

$$G_k(z) = \sum_{n=0}^{\infty} b_n(k) z^n = \left[ \frac{\alpha}{1 + \alpha - z} \right]^k, \ \forall k > 0.$$

is the generating function of the Negative Binomial pdf.

## II.3 THE MULTIVARIATE WEIBULL

By utilizing the positive stable distribution as a mixture on the hazard function of a Weibull distribution with shape parameter $\gamma$, Hougaard [18] derived a univariate Weibull distribution with a smaller shpae parameter. Hougaard [19] derived his multivariate Weibull using this same approach. specifically, he defined $Y_j = Z^{-1/\gamma} W_j$, $j = 1, \ldots, p$, where $Z$ is distributed as s positive stable distribution and $W_1, \ldots, W_p$ is iid Weibull with shape $\gamma$. The joint survival function for $p = 2$ is:

$$S(y_1, y_2) = e^{-[\phi_1 y_1^\gamma + \phi_2 y_2^\gamma]^\alpha}.$$

## II.4 THE EXPONENTIAL FAMILY TYPE

We consider the class of exponential family type probability distributions on the real line from McCullagh and Nelder [38]. The class is defined by the family of densities $\mathcal{G}$ with respect to the Lebesgue measure as follows

$$f(x; \theta, \varphi) = exp\left\{ \frac{\theta T(x) - b(\theta)}{a(\varphi)} + c(x, \varphi) \right\}, \tag{6}$$

where

- $f \in \mathcal{G}$.

- $\varphi$ is a constant scale parameter typically called nuisance parameter.

- $\theta$ is a location parameter.

- $a(\varphi)$ and $c(x, \varphi)$ are specific functions of the scale parameter.

- $b(\theta)$ and $T(x)$ are functions of the location parameter and variable $x$, respectively.

In fact, this exponential family density in (6) is a reformulation of the form given in McCullagh and Nelder [38] as they simplify $T(x)$ in (6) to simply $x$. Also, the expression (6) generalizes the exponential family type of distributions as described in Terbeche et al. [47] in the sense that:

- if $\varphi$ is known, then (6) is the linear exponential family with canonical parameter $\theta$.

- if $\varphi$ is unknown, then (6) may be used as a 2-parameter exponential family type.

As described in McCullagh and Nelder [38], this family includes the normal, exponential, gamma, Poisson types of distributions. In this setting,

$$U = U(\theta) = \frac{\partial \log L(\theta, x)}{\partial \theta} = \frac{\partial f(x, \theta)/\partial \theta}{f(x, \theta)} \tag{7}$$

is the score function. Note that:

- $E(U) = 0$

- $Var(U) = E(U^2) = -E(\partial U/\partial \theta) = I(\theta)$ also known as Fisher's information.

In the exponential family case as in (6),

$$l(\theta, \varphi, x) = \log L(\theta, \varphi, x) = \frac{\theta T(x) - b(\theta)}{a(\varphi)} + c(x, \varphi),$$

$$U = \frac{\partial l}{\partial \theta} = \frac{T(x) - \partial b(\theta)/\partial \theta}{a(\varphi)},$$

$$\text{and} \quad E(U) = 0 \Rightarrow E(T(x)) = \frac{\partial b(\theta)}{\partial \theta} = b'(\theta).$$

## II.5  LAPLACE TRANSFORM

The Laplace transform (the equivalent concept of moment generating function) provides a great deal of insight about the nature of a distribution. We first recall the definition.

**Definition II.5.1.** *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 [1] as:*

$$L_X(s) = Ee^{-sX} = \int_0^\infty e^{-sx} dF_X(x). \tag{8}$$

Here are some properties associated with the LST:

- Existence: the integral in (8) is with respect to the Lebesgue-Stieltjes integration in discrete and continuous case. In our cases of positive support distributions, (8) always exists. In fact $0 < L_X(s) \le 1$.

- $L_X(s)$ is infinitely differentiable, and $\dfrac{d^n L_X}{ds^n}(s)$ exist for all $n$.

- 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 rv's is obtained by taking the product of the LST of the individual r.v. For $X_1, \ldots, X_n$ independent rv's, then $X = \sum_{i=1}^n X_i$ has LST:

$$L_X(s) = E e^{-sX} = E \prod_{i=1}^n e^{sX_i} = \prod_{i=1}^n E e^{sX_i} = \prod_{i=1}^n L_{X_i}(s).$$

- Uniqueness: if $X_1$ and $X_2$ are two rv'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.

- The LST completely characterizes the distribution.

These transforms help in the computations and in the linear combinations of rv's associated with some distributions.

**Example II.5.1.**

For a gamma distribution $X \sim Ga(\mu, \lambda, \alpha)$ with pdf (2), its LST is given as:

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

By change of variables we get

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

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 II.5.2.**

Let $X_1, X_2, \cdots, X_n$, $n > 1$, be independent rv's distributed as $Ga(\mu_i, \lambda, \alpha_i)$ for $1 \leq i \leq n$. Then, using LST, it is easy to show that $X = X_1 + X_2 + \cdots X_n$ has density of $Ga(\sum_{i=1}^{n} \mu_i, \lambda, \sum_{i=1}^{n} \alpha_i)$. The result can also be found in Dudewicz and Mishra [13] page 277, and in Billingsley [5]. Notice that the simplicity of the result is due to the fact that the scale parameter is the same for all these gamma distributions.

In the remaining cases, we consider situations where the scale parameters are different. When two or more of the distributions have the same parameter $\lambda$, we can add them first to obtain another gamma distribution with the same parameter $\lambda$. Notice that the distribution of the sum of mutually independent gamma rv'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. See Mathai and Moschopoulos [35] and Mathai and Saxena [37] for more details.

**Example II.5.3.**

Suppose $X$ and $Y$ are independent discrete and positive support continuous distributions with probability mass function (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 II.5.3 in the case when $X$ is a Bernoulli rv with probability $p$, and $Y$ is a positive support distribution. In that case, we have that:

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

The sum of rv's will be of interest specially to verify the answers we will provide.

For two rv'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:

$$f_X(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{x-x_1} f(x_1, x_2) dx_1 dx_2.$$

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

**Theorem II.5.4.** *Assume that $X_1$ and $X_2$ are independent rv'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 [20]. □

However, a lot of work can be alleviated as the LST of $X$ in Theorem II.5.4 is given by: $L_X(t) = L_{X_1}(t) L_{X_2}(t)$, and is recognized in some distributional form.

**Example II.5.5.**

Consider $X$ to be the sum of two independent and identically distributed(iid) exponential type rv's with same scale parameter $\lambda$. Then $f_X(x) = \int_0^x f_1(t) f_2(x - t) dt = \int_0^x \lambda e^{-\lambda t} \lambda e^{-\lambda(x-t)} dt = \lambda^2 x e^{-\lambda x}$, which is a $Ga(0, \lambda, 2)$.

**Theorem II.5.6.** *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 $Ga(\mu, \lambda, \alpha)$ iff $\lambda_1 = \lambda_2$.*

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:

$$L_X(s) = e^{-s\mu} \left( \frac{\lambda}{\lambda + s} \right)^{\alpha}$$

$$f_X(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} (x - \mu)^{\alpha - 1} e^{-\lambda(x - \mu)}$$

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

For $X_i \sim Ga(\mu_i, \lambda_i, \alpha_i)$ for $i = 0, 1, \cdots, p$, based on the LST given in (8) and using the binomial theorem, we have the following:

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

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

In particular, $Cov(X_i, X_j) = a_i a_j Var(X_0)$, $i, j = 1, 2, \ldots, p$, where $X_i = a_i X_0 + Z_i$, $i = 1, 2, \ldots, p$, as in model in equation (1).

The LST has many helpful properties as discussed earlier. In particular, we use it for many of the derivations of the distributions of the latent variables. Other important concepts that will be helpful are the concept of infinitely divisibility and complete monotonicity. They have many applications in the theory of limit distributions for the sum of independent rv's. In general it is difficult to determine whether a given distribution is infinite divisible or not. We would like to consider what conditions are required for the pdf of the gamma distributions to be infinitely divisible.

We first give a notation and some definitions drawing from Feller [14] or Billingsley [5]. Let the symbol $\overset{d}{=}$ denote "equality in distribution".

**Definition II.5.2.** *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 $\sum_k X_{nk} \overset{d}{=} X$. 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 rv'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 [5] 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 $X_i$, $1 \leq i \leq n$.*

Two simple examples of infinitely divisible distributions are the Poisson distribution and the negative Binomial distribution. The Poisson r.v. $X$ which takes only nonnegative integer values with density function expressed as:

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \cdots, \quad, \lambda > 0.$$

Here the parameter $\lambda$, is the mean value of $X$. One may express $f(x)$ as:

$$f(x) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} \delta(x - n).$$

Its Laplace transform is then given by: $L_X(s) = e^{\lambda(e^{-s} - 1)}$.

**Definition II.5.3.** : *A function $\phi$ 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 II.5.7.** *$\phi$ is completely monotone iff $\phi$ is the Laplace transform of some measure.*

Proof: See Feller [14]. □

For two real valued functions $\phi_1$ and $\phi_2$ 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 [14]. Moreover, if $\phi$ is completely monotone on $[0, \infty)$ and $\phi(c) = 0$ for some $c > 0$, then $\phi$ must be identically zero on $[0, \infty)$.

## II.6  THE DIRAC DELTA FUNCTION

The Dirac delta function at the point $c \in \mathbb{R}$, is a point mass distribution denoted $\delta_c$. As in Abramowitz and Stegun [1], a r.v. $X$ has point mass $\delta_c$ distribution at $c$ if its probability mass function (pmf) is given by:

$$f(x|c) = \delta_c(x) = \delta(x-c) = 0 \quad \text{if} \quad x \neq c, \quad \text{and} \quad \int_{-\infty}^{\infty} f(x|c)dx = 1. \quad (9)$$

There have been several references on this function. The expressions in (9) can be thought as Lebesgue-Stieltjes integral and they have been studied by a number of authors such as Folland [15] and Rudin [44]. Such integrals are well developed topic in analysis, with applications in probability. For example, consider a real valued r.v. $X$ with $F$ as its cdf. Then, its expectation is: $E(g(X)) = \int_{-\infty}^{+\infty} g(x)F'(x)dx$.

However, if $F$ has discontinuities, or it is not differentiable at certain points, the above integral may not be valid. One way to avoid that situation is to consider the Stieltjes integral expression $E(g(X)) = \int_{-\infty}^{+\infty} g(x)dF(x)dx$, which always holds as long as $F$ is a proper cdf.

Despite its name, the Dirac's delta function is not a function in the classical sense. One reason for this is that because the function $f(x) = \delta(x)$, and $g(x) = 0$, are equal almost everywhere, yet their (Lebesgue) integrals are different. Another reason is that it is too 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 $(-\infty, \infty)$ such that for a given $f(x) \in \mathbf{D}$, the value of the functional is given by formula (2) in Khuri [26], Kallenberg [25], or Hunter and Nachtergaele [20]. 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 $\delta_x$ is defined by its action of a function $f(x) \in \mathbf{D}$ when used in integral as in formula (2) in Khuri [26]. Thus one should (never) not consider its value at $x$, i.e. the domain of $\delta$ is $\mathbf{D}$ and its values are given by formula (2) in Khuri [26]. As such, 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 1902.

So the Dirac delta function should be regarded as a distribution. As distribution, the Heaviside step function is an antiderivative of the Dirac distribution. The Heaviside step function, also called unit step function, see for example Abromowitz and Stegun [1], is a discontinuous function defined as

$$H(x) = \int_{-\infty}^{x} \delta(t)dt = \begin{cases} 0, & \text{if } x < 0; \\ 1, & \text{if } x > 0. \end{cases} \tag{10}$$

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 will 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. Pazman and Pronzato [43] used such function in their nonlinear settings. 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, as in Rudin [44]. It is important to see that the integral is simply a notational convenience, and not a true integral.

We should not confuse this above with the Dirac as a measure defined based on a fixed element $s$ of the space of interest, say $\mathbb{R}$. More precisely, the Dirac measure $\delta$ is given for any measurable set $E \in \mathcal{B}(\mathbb{R})$ by

$$\delta(E) = \begin{cases} 1 & s \in E \\ 0 & , s \notin E \end{cases}$$

and then $\int_{-\infty}^{\infty} f(x)d\delta(x) = f(s)$ for all continuous function $f$.

Here, differently from the Lebesgue measure which is translation invariant, it is not true that two intervals with the same endpoints necessarily have the same measure. $\delta([0,1]) = 1$ whereas $\delta((0,1)) = 0$. This is because $\delta$ is not absolutely continuous with respect to the Lebesgue measure. More details are given in Kallenberg [25], Williamson [48] or Shilov and Gurevich [45] to mention a few.

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:

- $\int_{-\infty}^{\infty} \delta(\alpha x) dx = \int_{-\infty}^{\infty} \delta(u) \frac{du}{|\alpha|} = \frac{1}{|\alpha|}, \ \forall \alpha \neq 0.$

- $\delta(\alpha x) = \frac{\delta(x)}{|\alpha|}, \ \forall \alpha \neq 0.$

- $\delta(x) = \lim_{a \to 0} \delta_a(x)$ 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: $H(x - a) = 1 - H(-x + a) = 1 - H(a - x)$ 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. More details and applications of the $\delta$-function can be found in Au and Tam [3]. So for a rv $X$ 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 $\delta_c$ are given by: $EX^k = c^k, \ Var(X) = 0$ and its characteristic function is given by $\phi(t) = e^{itc}$. The Dirac function provides a very helpful tool in mathematical statistics as it provided a unifying approach in the treatment of discrete and continuous distributions. We review two examples in each case below.

**Example II.6.1.**

Let $X_1 \sim Ga(2, \alpha)$, $X_2 \sim Ga(2, \beta)$ and $X_1, X_2$ are independent. Find the joint pdf of $(Y, Z)$, where $Y = \frac{X_1}{X_1 + X_2}$ and $Z = X_1 + X_2$ (Khuri [26]).

Let $f(x_1, x_2)$ be the joint pdf of $(X_1, X_2)$. Then joint pdf of $(Y, Z)$ is:

$$
\begin{aligned}
\tau(y, z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) \delta\left[\frac{x_1}{x_1 + x_2} - y\right] \delta(x_1 + x_2 - z)\, dx_1 dx_2 \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)2^{\alpha+\beta}} \int_{0}^{\infty} \int_{0}^{\infty} x_1^{\alpha-1} x_2^{\beta-1} e^{-\frac{(x_1+x_2)}{2}} \delta\left[\frac{x_1}{x_1+x_2} - y\right] \delta[x_1 - (z - x_2)]\, dx_1 dx_2 \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)2^{\alpha+\beta}} \int_{0}^{\infty} x_2^{\beta-1}\, dx_2 \int_{0}^{\infty} x_1^{\alpha-1} e^{-\frac{(x_1+x_2)}{2}} \delta\left[\frac{x_1}{x_1+x_2} - y\right] \delta[x_1 - (z - x_2)]\, dx_1 \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)2^{\alpha+\beta}} \int_{0}^{\infty} x_2^{\beta-1}(z - x_2)^{\alpha-1} \delta\left[\frac{z - x_2}{z} - y\right] e^{-\frac{z}{2}}\, dx_2 \\
&= \frac{e^{-\frac{z}{2}} z}{\Gamma(\alpha)\Gamma(\beta)2^{\alpha+\beta}} \int_{0}^{\infty} x_2^{\beta-1}(z - x_2)^{\alpha-1} \delta[x_2 - (z - zy)]\, dx_2 \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)2^{\alpha+\beta}} y^{\alpha-1}(1 - y)^{\beta-1} z^{\alpha+\beta-1} e^{-\frac{z}{2}}.
\end{aligned}
$$

**Example II.6.2.**

Let $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$ and $X, Y$ are independent. Find the pdf of $Z$, where $Z = \dfrac{nX}{mY}$ (Au and Tam [3]).

Let $f(x, y)$ be the joint pdf of $(X, Y)$. Then joint pdf of $(X, Y)$ is:

$$
f(x, y) = \frac{1}{2^{\frac{m+n}{2}}\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} x^{\frac{m}{2}-1} y^{\frac{n}{2}-1} e^{-\frac{x}{2}} e^{-\frac{y}{2}}
$$

Then the pdf of $Z$ is:

$$
\begin{aligned}
g(z) &= \frac{1}{2^{\frac{m+n}{2}}\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \int_{0}^{\infty} \int_{0}^{\infty} x^{\frac{m}{2}-1} y^{\frac{n}{2}-1} e^{-\frac{x}{2}} e^{-\frac{y}{2}} \delta\left[\frac{nx}{my} - z\right] dx\, dy \\
&= \frac{1}{2^{\frac{m+n}{2}}\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \int_{0}^{\infty} \frac{my}{n} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}\, dy \int_{0}^{\infty} x^{\frac{m}{2}-1} e^{-\frac{x}{2}} \delta\left[x - \frac{myz}{n}\right] dx \\
&= \frac{1}{2^{\frac{m+n}{2}}\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \int_{0}^{\infty} \frac{my}{n} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} e^{-\frac{myz}{2n}} \left(\frac{myz}{n}\right)^{\frac{m}{2}-1} dy.
\end{aligned}
$$

By change of variables

$$
\begin{aligned}
g(z) &= \frac{1}{2^{\frac{m+n}{2}}\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} z^{\frac{m}{2}-1} \left(\frac{2n}{n + mz}\right)^{\frac{m+n}{2}} \int_{0}^{\infty} u^{\frac{m+n}{2}-1} e^{-u}\, du \\
&= \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{z^{\frac{m}{2}-1}}{(n + mz)^{\frac{m+n}{2}}}.
\end{aligned}
$$

# CHAPTER III

# SUM OF IID RANDOM VARIABLES

Sum of independently normally distributed models are of interest in many settings and are described by many authors such as Dudewick and Mishra [13], and Casella and Berger [7]. Normal distributions are very valuable, but they have limitations, specially if the sample sizes are small, and/or the data do not take values in the negative range. As suggested by Hougaard [19], transformations to achieve normality should not be used unless other alternatives have been explored. For positive support data, other lifetime distributions, such as the Binomial, the gamma, the Weibull, have been suggested. The finite sum of non-normal independent distributions has been studied by many authors such as Moschopoulos [41], Mathai and Saxena [37] and Nadarajah and Kotz [42]. Applications for such distributional sum can be found in many areas. Mathai [34] gives example in storage capabilities. It is well recognized that the sum of independent gamma type distributions, the scale parameter being same for the individual distributions, again has a gamma distribution. Approximations are also suggested as the exact distribution of the sum is not always possible to get in a closed form, as suggested in Luke [31] or in Krisnaiah and Rao [28]. In such cases,the nonparametric approach is clearly preferred because of the difficulty in estimation of associated parameters. Recently, the product of Bernoulli and exponential distributions have received great attention. In fact, Iyer and Manjunath [22] considered such product in a reliability system. Marshall and Olkin [32] and Marshall and Olkin [33] used such a distribution for the modeling of events that can occur simultaneously. The idea of simultaneous occurrence of events has been shown to be very useful in many areas of science and it is famously known as the Marshall and Olkin property as in Carpenter et al. [6]. However, the sum of such independent distributions has not been fully described and in many cases, estimations and tests depend on the true distributional form of the data. In this chapter, we address the sum of independent products of Bernoulli and exponential distributions as a mixture of two types of distribution functions: the Dirac delta and gamma types. In the next section, we describe the product of Bernoulli and exponential type of distributions. Section 2 presents the exact distribution of the sum of independent such products. In section 3, we present the properties associated with the sum, and in section 4 we summarize the results in this chapter.

## III.1 PRODUCT OF BERNOULLI AND EXPONENTIAL DISTRIBU-TIONS

In this section, we give the definition of the product of Bernoulli and exponential distributions. We show the relationship between these distributions through a latent rv (rv) $Y$ whose cdf and pdf is represented in Figure 3 and which is independent to the independent variable. More precisely, let $X_1$ and $X_2$ be two exponential rv's with parameters $\lambda_1$ and $\lambda_2$, respectively, and under a linear relationship defined as: $X_2 = aX_1 + Y$, where $a$ is a fixed positive constant. That means that:

$$X_2 = \begin{cases} aX_1, & \text{with some probability } p, \\ aX_1 + Y, & \text{with probability } 1 - p, \end{cases}$$

where $Y$ has pdf and cdf described based on the LST as in example II.5.3.
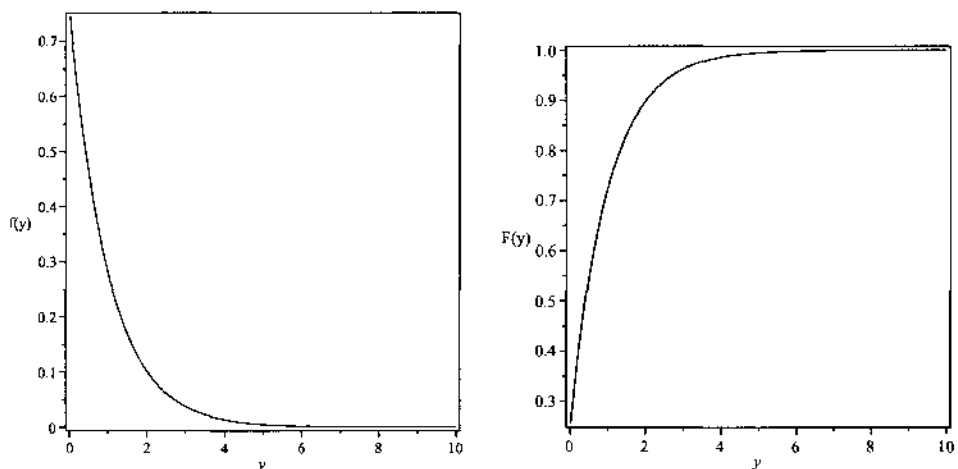


(a) The Probability Density Function.     (b) The Cumulative Distribution Function.

FIG. 3: Graphs of pdf and cdf of $Y$ when $a = 1$, $\lambda_1 = 4$, and $\lambda_2 = 1$.

This model gives independent coordinates when $a = 0$, and does not permit negative association as described in Iyer et al. [21]. The idea for such distribution is not new. The characterization described in which the marginal distributions are exponential was introduced by Marshall and Olkin [32], and has been studied by many authors such as Johnson et al. [24]. In that setting, the rv $Y$ is a product of a Bernoulli rv with parameter $p$ and an exponential rv with parameter $\lambda_2$. Its pdf and

cdf are given by:

$$f(y) = p\delta(y) + (1 - p)f_{X_2}(y)I(y > 0), \tag{11}$$

$$F(y) = pH(y) + (1 - p)F_{X_2}(y)I(y > 0). \tag{12}$$

where

- $p = P(X_2 = aX_1) = P(Y = 0) = \dfrac{a\lambda_2}{\lambda_1}$, $a\lambda_2 < \lambda_1$ is the probability of proportional occurrence between $X_1$ and $X_2$.

- $\delta(t)$ refers to the Dirac delta function, i.e.
  $\delta(t) = 0$, if $t \neq 0$, and $\displaystyle\int_{-\infty}^{+\infty} \delta(t)\,dt = 1$.

- $f_{X_2}(t) = \lambda_2 e^{-\lambda_2 t}, t > 0$, and $F_{X_2}(y) = \displaystyle\int_0^y f_{X_2}(t)\,dt$.

- $H(y)$ is the Heaviside function, the generalized anti-derivative of $\delta(y)$, i.e.
  $\delta(y) = \dfrac{dH(y)}{dy}$.

As described in Marshall and Olkin [32], the rv's $X_1$ and $X_2$ could represent the times to failures of two components in a parallel system. Instead of simultaneous failures, we adopt the case of proportional failures. A medical application is the analysis of two related types of diseases whose occurrence could be suggested to be simultaneous or proportional to each other, or the first and second period responses to a treatment order in the different clusters. The mean and the variance of (11) are given by: $\dfrac{1 - p}{\lambda_2}$ and $\dfrac{(1 - p)(1 + p)}{\lambda_2^2}$, respectively. In fact:

$$
\begin{aligned}
E(Y) &= \int_0^{+\infty} y f(y)\,dy, \\
&= p\int_0^{+\infty} y\delta(y)\,dy + (1 - p)\lambda_2 \int_0^{+\infty} y e^{-\lambda_2 y}\,dy = \frac{1 - p}{\lambda_2}.
\end{aligned}
$$

$$
\begin{aligned}
E(Y^2) &= \int_0^{+\infty} y^2 f(y)\,dy, \\
&= p\int_0^{+\infty} y^2 \delta(y)\,dy + (1-p)\lambda_2 \int_0^{+\infty} y^2 e^{-\lambda_2 y}\,dy = \frac{2(1-p)}{\lambda_2^2}, \\
V(Y) &= E(Y^2) - [E(Y)]^2, \\
&= \frac{(1-p)(1+p)}{\lambda_2^2}, \text{ and} \\
L_Y(s) &= \left[\frac{\lambda_2}{(\lambda_2 + s)}(1-p) + p\right] \text{ is its Laplace Stieltjes Transform(LST),} \quad (13) \\
M_Y(t) &= \left[\frac{\lambda_2}{\lambda_2 - t}(1-p) + p\right] \text{ is its Moment generating function(MGF).} \quad (14)
\end{aligned}
$$

We next derive the form of the distribution of the finite sum of such models.

## III.2 EXACT DENSITY OF THE SUM

The sum of independent rv's is described by discrete or n-fold convolution. It is possible to calculate the density of the sum in certain cases. Sum of independent Binomial with same probability of success is again a Binomial rv. Similar conclusion are available for Uniform, Poisson, and exponential distributions. The convolution of geometric distributions with same probability of success is a negative binomial distribution. We look at the combinations of the Bernoulli and exponential distributions.

As stated earlier, the sum of independent and identically distributed products of Bernoulli and exponential models are of interest in many settings. We present such sums, and present a practical way to estimate its parameters.

**Theorem III.2.1.** *Let* $Y_1, \dots, Y_n$ *be independent and identically distributed rvs with the pdf as in (11). Define* $S_n = Y_1 + \dots + Y_n$. *Then the distribution of* $S_n$ *can be written as a mixture of gamma and Dirac delta distributions with Binomial weights, i.e.*

$$
f_{S_n}(y) = \sum_{j=0}^{n} \binom{n}{n-j} p^{n-j}(1-p)^j f_{g_j}(y), \quad y \geq 0, \tag{15}
$$

*where* $f_{g_0}(t) = \delta(t)$, *and* $f_{g_j}(t) = \dfrac{\lambda_2^{\,j}}{\Gamma(j)} t^{j-1} e^{-\lambda_2 t}, t > 0,$ *for* $1 \leq j \leq n$.

*Proof.* The result is shown by induction. When $n = 1$, then $S_1 = Y_1$, and the result

is obvious. Consider two iid products of Bernoulli and exponential, $Y_1$ and $Y_2$:

$$Y_1 \sim f_{Y_1}(y) \quad = \quad p\delta(y) + (1-p)f_{X_2}(y)I(y > 0) \text{ and}$$

$$Y_2 \sim f_{Y_2}(y) \quad = \quad p\delta(y) + (1-p)f_{X_2}(y)I(y > 0).$$

Then $S_2 = Y_1 + Y_2$, has its pdf given by: $f_{S_2}(y) = \int_0^{+\infty} f_{Y_1}(w)f_{Y_2}(y-w)\, dw$, with
$f_{Y_1}(w) = p\delta(w) + (1-p)f_{X_2}(w)I(w > 0)$,
$f_{Y_2}(y-w) = p\delta(y-w) + (1-p)f_{X_2}(y-w)I(y-w > 0)$.
That is, $S_2$ is the convolution of $f_{Y_1}$ and $f_{Y_2}$, and

$$
\begin{aligned}
f_{Y_1}(w)f_{Y_2}(y-w) \quad = \quad & p^2\delta(w)\delta(y-w) + p(1-p)\delta(w)f_{X_2}(y-w)I(y-w > 0) \\
& + p(1-p)\delta(y-w)f_{X_2}(y)I(y > 0) \\
& + (1-p)^2 f_{X_2}(w)f_{X_2}(y-w)I(w > 0)I(y-w > 0).
\end{aligned}
$$

Then,

$$
\begin{aligned}
f_{S_2}(y) \quad = \quad & p^2\int_0^{+\infty}\delta(w)\delta(y-w)\, dw + p(1-p)\int_0^{+\infty}\delta(w)f_{X_2}(y-w)I(y-w > 0)\, dw \\
& + p(1-p)\int_0^{+\infty}\delta(y-w)f_{X_2}(w)I(w > 0)\, dw \\
& + (1-p)^2\int_0^{+\infty}f_{X_2}(w)f_{X_2}(y-w)I(w > 0)I(y-w > 0)\, dw \\
= \quad & A_1 + A_2 + A_3 + A_4, \text{ where}
\end{aligned}
$$

$$
\begin{aligned}
A_1 \quad = \quad & p^2\int_0^{+\infty}\delta(w)\delta(y-w)dw = p^2\delta(y), \\
A_2 \quad = \quad & p(1-p)\int_0^{+\infty}\delta(w)f_{X_2}(y-w)I(y-w > 0)\, dw \\
= \quad & p(1-p)\int_0^y f_{X_2}(y-w)\delta(w)\, dw = p(1-p)f_{X_2}(y), \\
A_3 \quad = \quad & p(1-p)\int_0^{+\infty}\delta(y-w)f_{X_2}(w)I(w > 0)\, dw \\
= \quad & p(1-p)\int_0^{+\infty}f_{X_2}(w)\delta(y-w)\, dw = p(1-p)f_{X_2}(y) = A_2\ , \\
A_4 \quad = \quad & (1-p)^2\int_0^{+\infty}f_{X_2}(w)f_{X_2}(y-w)I(w > 0)I(y-w > 0)\, dw \\
= \quad & (1-p)^2\lambda_2^2 e^{-\lambda_2 y}\int_0^y dw = (1-p)^2\lambda_2^2 y e^{-\lambda_2 y}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
f_{S_2}(y) &= p^2\delta(y) + 2p(1-p)f_{X_2}(y) + (1-p)^2\lambda_2^2 y e^{-\lambda_2 y}\\
&= p^2\delta(y) + 2p(1-p)\lambda_2 e^{-\lambda_2 y} + (1-p)^2\lambda_2^2 y e^{-\lambda_2 y}\\
&= p^{2-0}(1-p)^0\delta(y) + 2p^{2-1}(1-p)^1\frac{\lambda_2^1}{\Gamma(1)}y^{1-1}e^{-\lambda_2 y} + p^{2-2}(1-p)^2\frac{\lambda_2^2}{\Gamma(2)}y^{2-1}e^{-\lambda_2 y}\\
&= p^{2-0}(1-p)^0 f_{g_0}(y) + 2p^{2-1}(1-p)^1 f_{g_1}(y) + p^{2-2}(1-p)^2 f_{g_2}(y)\\
&= \sum_{j=0}^{2}\binom{2}{2-j}p^{2-j}(1-p)^j f_{g_j}(y).
\end{aligned}
$$

Therefore the result is true for $n = 2$. Assume the result is true for some $n \in \mathbb{N}$. We prove the result is valid for $n+1$. Let $S_{n+1} = S_n + Y_{n+1}$, where: $Y_{n+1} \sim f_{Y_{n+1}}(y) = p\delta(y) + (1-p)f_{X_2}(y)I(y > 0)$. Then,

$$f_{S_{n+1}}(y) = \int_0^{+\infty} f_{S_n}(w)f_{Y_{n+1}}(y-w)\,dw, \text{ and}$$

$$f_{S_n}(w) = \sum_{j=0}^{n}\binom{n}{n-j}p^{n-j}(1-p)^j f_{g_j}(w).$$

Since $f_{Y_{n+1}}(y-w) = p\delta(y-w) + (1-p)f_{X_2}(y-w)I(y-w > 0)$, we have that

$$
\begin{aligned}
f_{S_n}(w)f_{Y_{n+1}}(y-w) &= \sum_{j=0}^{n}\binom{n}{n-j}p^{n+1-j}(1-p)^j f_{g_j}(w)\delta(y-w)\\
&\quad + \sum_{j=0}^{n}\binom{n}{n-j}p^{n-j}(1-p)^{j+1}f_{g_j}(w)f_{X_2}(y-w)I(y-w > 0),
\end{aligned}
$$

and
$$
\begin{aligned}
f_{S_{n+1}}(y) &= \sum_{j=0}^{n}\binom{n}{n-j}p^{n+1-j}(1-p)^j \int_0^{+\infty} f_{g_j}(w)\delta(y-w)\,dw\\
&\quad + \sum_{j=0}^{n}\binom{n}{n-j}p^{n-j}(1-p)^{j+1}\int_0^{y} f_{g_j}(w)f_{X_2}(y-w)\,dw\\
&= \text{Part1} + \text{Part2},
\end{aligned}
$$

with

$$
\begin{aligned}
\text{Part1} &= \sum_{j=0}^{n}\binom{n}{n-j}p^{n+1-j}(1-p)^j \int_0^{+\infty} f_{g_j}(w)\delta(y-w)\,dw\\
&= \sum_{j=0}^{n}\binom{n}{n-j}p^{n+1-j}(1-p)^j f_{g_j}(y), \text{ and}
\end{aligned}
$$

$$\text{Part2} = \sum_{j=0}^{n} \binom{n}{n-j} p^{n-j}(1-p)^{j+1} \int_{0}^{y} f_{g_j}(w) f_{X_2}(y-w)\, dw$$

$$= \sum_{j=0}^{n} \binom{n}{n-j} p^{n-j}(1-p)^{j+1} \frac{\lambda_2^{j+1} e^{-\lambda_2 y}}{\Gamma(j)} \int_{0}^{y} w^{j-1}\, dw$$

$$= \sum_{j=0}^{n} \binom{n}{n-j} p^{n-j}(1-p)^{j+1} \frac{\lambda_2^{j+1} e^{-\lambda_2 y}}{\Gamma(j)} \frac{y^j}{j}$$

$$= \sum_{j=0}^{n} \binom{n}{n-j} p^{n-j}(1-p)^{j+1} \frac{\lambda_2^{j+1}}{\Gamma(j+1)} y^{j+1-1} e^{-\lambda_2 y}$$

$$= \sum_{j=0}^{n} \binom{n}{n-j} p^{n-j}(1-p)^{j+1} f_{g_{j+1}}(y).$$

So,

$$f_{S_{n+1}}(y) = \sum_{j=0}^{n} \binom{n}{n-j} p^{n+1-j}(1-p)^{j} f_{g_j}(y)$$

$$+ \sum_{j=0}^{n} \binom{n}{n-j} p^{n-j}(1-p)^{j+1} f_{g_{j+1}}(y)$$

$$= p^{n+1} f_{g_0}(y) + \sum_{j=1}^{n} \binom{n}{n-j} p^{n+1-j}(1-p)^{j} f_{g_j}(y)$$

$$+ \sum_{j=0}^{n-1} \binom{n}{n-j} p^{n-j}(1-p)^{j+1} f_{g_{j+1}}(y) + (1-p)^{n+1} f_{g_{n+1}}(y).$$

Using change of variables in the third term of above equation, we get:

$$f_{S_{n+1}}(y) = p^{n+1} f_{g_0}(y) + \sum_{j=1}^{n} \binom{n}{n-j} p^{n+1-j}(1-p)^{j} f_{g_j}(y)$$

$$+ \sum_{l=1}^{n} \binom{n}{n+1-l} p^{n+1-l}(1-p)^{l} f_{g_l}(y) + (1-p)^{n+1} f_{g_{n+1}}(y)$$

$$= p^{n+1} f_{g_0}(y) + \sum_{j=1}^{n} p^{n+1-j}(1-p)^{j} f_{g_j}(y) \left[ \binom{n}{n-j} + \binom{n}{n-j+1} \right]$$

$$+ (1-p)^{n+1} f_{g_{n+1}}(y)$$

$$= p^{n+1} f_{g_0}(y) + \sum_{j=1}^{n} \binom{n+1}{n+1-j} p^{n+1-j}(1-p)^{j} f_{g_j}(y) + (1-p)^{n+1} f_{g_{n+1}}(y)$$

$$= \sum_{j=0}^{n+1} \binom{n+1}{n+1-j} p^{n+1-j}(1-p)^{j} f_{g_j}(y).$$

Therefore the result is true for $n+1$. $\qquad \square$

The pdf of the sum can be thought of as a hyper gamma rv. Figure 4 gives a graphical display of the hyper gamma for $\lambda_1 = 4$, $\lambda_2 = 1$, $a = 1$, and $n = 10$.

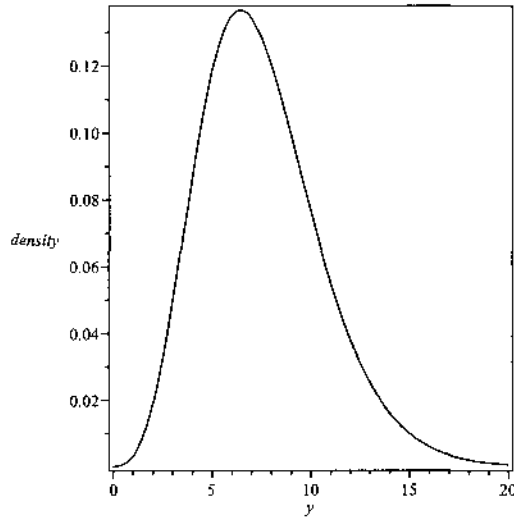This finite sum result is a simple and explicit expression not using infinite sums or



FIG. 4: The pdf of $S_n$.

hypergeometric functions or approximations, as proposed by many authors such as Mathai and Saxena [37]. Since there are other characterizations of distribution, our result can be confirmed in other ways (such as the Laplace transform).

## III.3 PROPERTIES OF THE RANDOM SUM

**Lemma III.3.1.** *Suppose $S_n$ is distributed as in (15). Then $P(S_n = 0) = p^n$.*

*Proof.*

$$
\begin{aligned}
P(S_n = 0) &= P(Y_1 + \ldots + Y_n = 0), \\
&= P(Y_1 = 0, \ldots, Y_n = 0), \\
&= P(Y_1 = 0) \ldots P(Y_n = 0), \\
&= \left(\frac{a\lambda_2}{\lambda_1}\right)^n.
\end{aligned}
$$

$\square$

We next give the general form and properties associated with its moments and the survival function, the LST and the MGF.

**Lemma III.3.2.** *Suppose $S_n$ is distributed as in (15). The cdf of $S_n$ can be expressed as:*

$$F_{S_n}(y) = \sum_{j=0}^{n} \binom{n}{n-j} p^{n-j}(1-p)^j F_{g_j}(y) \quad y > 0. \tag{16}$$

*Proof.* From the theory of Heaviside functions in (10), we have that

$$
\begin{aligned}
F_{S_n}(y) &= P(S_n < y) \\
&= \int_0^y f_{S_n}(t)\,dt \\
&= p^n \int_0^y \delta(t)\,dt + p^n \sum_{j=1}^{n} \binom{n}{n-j} \frac{(1-p)^j}{p^j} \int_0^y f_{g_j}(t)\,dt \\
&= p^n \int_0^y \delta(t)\,dt + p^n \sum_{j=1}^{n} \binom{n}{n-j} \frac{(1-p)^j}{p^j} \frac{\lambda_2^j}{\Gamma(j)} \int_0^y t^{j-1} e^{-\lambda_2 t}\,dt \\
&= p^n H(y) + p^n \sum_{j=1}^{n} \binom{n}{n-j} \frac{(1-p)^j}{p^j} F_{g_j}(y) \\
&= \sum_{j=0}^{n} \binom{n}{n-j} p^{n-j}(1-p)^j F_{g_j}(y),
\end{aligned}
$$

where

- $F_{g_0}(y) = H(y)$ is the Heaviside function, the generalized anti-derivative of $\delta(y)$, i.e. $\delta(y) = \dfrac{dH(y)}{dy}$.

- $F_{g_j}(y) = \displaystyle\int_0^y f_{g_j}(t)\,dt = \frac{\gamma(j,\,\lambda_2 y)}{\Gamma(j)} = \left[ 1 - \sum_{i=0}^{j-1} \frac{e^{-\lambda_2 y}(\lambda_2 y)^i}{i!} \right].$

□

**Lemma III.3.3.** *Suppose $S_n$ is distributed as in (15). The survival function of $S_n$ can be expressed as:*

$$P(S_n \geq y) = \sum_{j=1}^{n} \binom{n}{n-j} p^{n-j}(1-p)^j \sum_{i=0}^{j-1} \frac{e^{-\lambda_2 y}(\lambda_2 y)^i}{i!} - p^n H(y) \quad y > 0. \tag{17}$$

*Proof.* The result follows from the result in (16). Indeed,

$$
\begin{aligned}
F_{S_n}(y) &= p^n H(y) + \sum_{j=1}^{n} \binom{n}{n-j} p^{n-j}(1-p)^j \frac{\gamma(j, \lambda_2 y)}{\Gamma(j)}, \\
&= p^n H(y) + \sum_{j=1}^{n} \binom{n}{n-j} p^{n-j}(1-p)^j \left[ 1 - \sum_{i=0}^{j-1} \frac{e^{-\lambda_2 y}(\lambda_2 y)^i}{i!} \right] \\
&= p^n H(y) + 1 - \sum_{j=1}^{n} \binom{n}{n-j} p^{n-j}(1-p)^j \sum_{i=0}^{j-1} \frac{e^{-\lambda_2 y}(\lambda_2 y)^i}{i!}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
P(S_n \ge y) &= 1 - F_{S_n}(y) \\
&= \sum_{j=1}^{n} \binom{n}{n-j} p^{n-j}(1-p)^j \sum_{i=0}^{j-1} \frac{e^{-\lambda_2 y}(\lambda_2 y)^i}{i!} - p^n H(y).
\end{aligned}
$$

$\square$

**Lemma III.3.4.** *The $m^{th}$ moment of the above mixture is:*

$$
E[S_n^m] = \frac{1}{\lambda_2^m} \sum_{j=0}^{n} \binom{n}{n-j} p^{n-j}(1-p)^j \frac{\Gamma(j+m)}{\Gamma(j)}. \tag{18}
$$

*Proof.* We can see that:

$$
\begin{aligned}
E[S_n^m] &= \int_0^{+\infty} y^m f_{S_n}(y)\, dy \\
&= p^n \int_0^{+\infty} y^m \delta(y)\, dy + \int_0^{+\infty} y^m \sum_{j=1}^{n} \binom{n}{n-j} p^{n-j}(1-p)^j f_{g_j}(y)\, dy \\
&= \sum_{j=1}^{n} \binom{n}{n-j} p^{n-j}(1-p)^j \int_0^{+\infty} \frac{\lambda_2^j}{\Gamma(j)} y^{j+m-1} e^{-\lambda_2 y}\, dy \\
&= \frac{p^n}{\lambda_2^m} \sum_{j=1}^{n} \binom{n}{n-j} \frac{(1-p)^j}{p^j} \frac{\Gamma(j+m)}{\Gamma(j)}.
\end{aligned}
$$

$\square$

Result 1:

$$
E[S_n] = \frac{n(1-p)}{\lambda_2}.
$$

*Proof.* Take using (18) with $m = 1$.

Then,

$$
\begin{aligned}
E[S_n] &= \frac{p^n}{\lambda_2} \sum_{j=1}^{n} \binom{n}{n-j} \frac{(1-p)^j}{p^j} \frac{\Gamma(j+1)}{\Gamma(j)}, \\
&= \frac{p^n}{\lambda_2} \sum_{j=1}^{n} j \binom{n}{n-j} \frac{(1-p)^j}{p^j}, \\
&= \frac{np^n}{\lambda_2} \sum_{j=1}^{n} \binom{n-1}{j-1} \frac{(1-p)^j}{p^j}, \\
&= \frac{np^n}{\lambda_2} \sum_{j=1}^{n} \binom{n-1}{j-1} r^j = \frac{np^n r}{\lambda_2} \sum_{j=1}^{n} \binom{n-1}{j-1} r^{j-1} 1^{n-j}, \\
&= \frac{np^n r}{\lambda_2} (r+1)^{n-1}, \\
&= \frac{n(1-p)}{\lambda_2},
\end{aligned}
$$

with $r = \dfrac{1-p}{p}$. $\qquad\square$

Result 2:

$$
V[S_n] = \frac{n(1-p)(1+p)}{\lambda_2^2}.
$$

*Proof.* Take the result in (18) with $m = 2$. Then,

$$
\begin{aligned}
E[S_n^2] &= \frac{p^n}{\lambda_2^2} \sum_{j=1}^{n} \binom{n}{n-j} \frac{(1-p)^j}{p^j} \frac{\Gamma(j+2)}{\Gamma(j)}, \\
&= \frac{p^n}{\lambda_2^2} \sum_{j=1}^{n} \binom{n}{n-j} j(j+1) r^j, \quad \text{where} \quad r = \frac{1-p}{p}, \\
&= \frac{np^n}{\lambda_2^2} \sum_{j=1}^{n} \binom{n-1}{j-1} (j+1) r^j, \\
&= \frac{n(n-1)p^n}{\lambda_2^2} \sum_{j=2}^{n} \binom{n-2}{j-2} r^j + \frac{2np^n}{\lambda_2^2} \sum_{j=1}^{n} \binom{n-1}{j-1} r^j, \\
&= \frac{n(n-1)p^n r^2}{\lambda_2^2} \sum_{j=2}^{n} \binom{n-2}{j-2} r^{j-2} + \frac{2np^n r}{\lambda_2^2} \sum_{j=1}^{n} \binom{n-1}{j-1} r^{j-1}, \\
&= \frac{n(n-1)p^n r^2}{\lambda_2^2} (r+1)^{n-2} + \frac{2np^n r}{\lambda_2^2} (r+1)^{n-1}.
\end{aligned}
$$

$$
\begin{aligned}
E[S_n^2] &= \frac{n(n-1)}{\lambda_2^2}(1-p)^2 + \frac{2n}{\lambda_2^2}(1-p), \\
\text{and } V[S_n] &= E[Y^2] - E[Y]^2, \\
V[S_n] &= \frac{n(1-p)(1+p)}{\lambda_2^2}.
\end{aligned}
$$

$\square$

**Lemma III.3.5.** *The MGF is:*

$$
M_{S_n}(t) = \left[ \frac{\lambda_2}{\lambda_2 - t}(1-p) + p \right]^n. \tag{19}
$$

*Proof.*

$$
\begin{aligned}
M_{S_n}(t) &= \int_0^{+\infty} e^{ty} f_{S_n}(y)\, dy, \\
&= p^n \int_0^{+\infty} e^{ty} \delta(y)\, dy + \int_0^{+\infty} e^{ty} \sum_{j=1}^n \binom{n}{n-j} p^{n-j}(1-p)^j f_{g_j}(y)\, dy, \\
&= p^n + \sum_{j=1}^n \binom{n}{n-j} p^{n-j}(1-p)^j \int_0^{+\infty} \frac{\lambda_2^j}{\Gamma(j)} y^{j-1} e^{-(\lambda_2 - t)y}\, dy, \\
&= p^n + p^n \sum_{j=1}^n \binom{n}{n-j} \frac{(1-p)^j}{p^j} \frac{\lambda_2^j}{(\lambda_2 - t)^j}, \\
&= p^n \sum_{j=0}^n \binom{n}{n-j} \left[ \frac{\lambda_2}{(\lambda_2 - t)} \frac{(1-p)}{p} \right]^j, \\
&= \left[ \frac{\lambda_2}{(\lambda_2 - t)}(1-p) + p \right]^n.
\end{aligned}
$$

$\square$

**Lemma III.3.6.** *The LST is:*

$$
L_{S_n}(s) = \left[ \frac{\lambda_2}{\lambda_2 + s}(1-p) + p \right]^n. \tag{20}
$$

*Proof.*

$$
\begin{aligned}
L_{S_n}(t) &= \int_0^{+\infty} e^{-sy} f_{S_n}(y)\, dy, \\
&= p^n \int_0^{+\infty} e^{-sy} \delta(y)\, dy + \int_0^{+\infty} e^{-sy} \sum_{j=1}^n \binom{n}{n-j} p^{n-j}(1-p)^j f_{g_j}(y)\, dy, \\
&= p^n + \sum_{j=1}^n \binom{n}{n-j} p^{n-j}(1-p)^j \int_0^{+\infty} \frac{\lambda_2^j}{\Gamma(j)} y^{j-1} e^{-(\lambda_2 - t)y}\, dy, \\
&= p^n + p^n \sum_{j=1}^n \binom{n}{n-j} \frac{(1-p)^j}{p^j} \frac{\lambda_2^j}{(\lambda_2 + s)^j}, \\
&= p^n \sum_{j=0}^n \binom{n}{n-j} \left[ \frac{\lambda_2}{(\lambda_2 + s)} \frac{(1-p)}{p} \right]^j = \left[ \frac{\lambda_2}{(\lambda_2 + s)}(1-p) + p \right]^n.
\end{aligned}
$$

$\square$

The result in (19) and (20) confirm the expression of the product of $n$ identical distributions from (14) and (13).

## III.4  SUMMARY

Relationships among distributions have long fascinated many authors. Leemis and McQueston [30] describe many univariate distributions and state that their sum is asymptotically normal. However normality assumption has limitations and non asymptotic results can take a long time to approach. We have proposed the sum of a particular type of survival distribution, and we have given its exact density and properties.

# CHAPTER IV

# BIVARIATE MODELS

Bivariate models are the natural extension of the univariate models. They can later be generalized to higher level multivariate models. Applications using multivariate gamma can be found in Kotz et al. [27]. In this chapter, we introduce several bivariate models. If $X_1, X_2$ are Erlang, then $(X_1, X_2)$ is called the bivariate Erlang-Erlang model. Since two distributions will be of specified marginals, we will use the name of the marginal distributions to denote such bivariate models. We define the Erlang-Erlang bivariate model to be a model where the marginal densities are Erlang distributed and the two rv's are linearly related as in (1). First, the Erlang-Erlang bivariate model is introduced. Second, we introduce the Exponential-Exponential bivariate model, its estimated parameters and the application of this model to simulated and real-life data. Starting with exponential rv's and with the aide of transformation technique, we derive the univariate Weibull and use the univariate Weibull to derive our Weibull-Weibull bivariate model. Next we discuss the Erlang-Gamma bivariate model. We also introduce the Gamma-Gamma model. The result in Theorem III.2.1 will be critical in explaining all the forms of subsequent densities. $(X_1, X_2)$ is called the bivariate model.

## IV.1   ERLANG-ERLANG MODEL

**Theorem IV.1.1.** *Let $f_1$ and $f_2$ represent the marginal densities of two rv's $X_1$ and $X_2$ from (3) with the same scale parameter $\alpha$ with $\alpha \in \mathbb{N}$. More specifically, let $X_1 \sim Erlang(\lambda_1, \alpha)$ and $X_2 \sim Erlang(\lambda_2, \alpha)$ Then the joint probability density function of $(X_1, X_2)$ is given by:*

$$g(x_1, x_2) = \sum_{j=0}^{\alpha} \binom{\alpha}{\alpha - j} p^{\alpha - j}(1 - p)^j f_1(x_1) f_{g_j}(x_2 - ax_1) \qquad (21)$$

*where*

- *the rv's $X_1$ and $X_2$ are related as $X_2 = aX_1 + Z$, with $a$ a nonnegative fixed constant called the coefficient of linear relationship, and $Z$ an unknown random*

*variable independent of $X_1$*

- $p = P(X_2 = aX_1) = \frac{a\lambda_2}{\lambda_1}$.

- $f_{g_j}(t) = \dfrac{\lambda_2^j}{\Gamma(j)} t^{j-1} e^{-\lambda_2 t}, t > 0, \text{ for } 1 \leq j \leq n.$

- $f_{g_0}(t) = \delta(t)$ refers to the Dirac delta function, i.e. $\delta(t) = 0, \text{ if } t \neq 0,$ and $\displaystyle\int_{-\infty}^{+\infty} \delta(t)\, dt = 1.$

Proof: First we consider the Laplace transform of $X_2$,

$$
\begin{aligned}
L_{X_2}(s) &= E[e^{-sX_2}], \\
&= E[e^{-s(aX_1+Z)}] \\
&= E[e^{-saX_1}]E[e^{-sZ}] \text{ Using the independence of } X_1 \text{ and } Z, \\
\text{So } L_{X_2}(s) &= L_{X_1}(as)L_Z(s).
\end{aligned}
$$

Since we know the Laplace transforms of $X_1$ and $X_2$, we can now obtain the Laplace transform of $Z$. Therefore:

$$
\begin{aligned}
L_Z(s) &= \frac{L_{X_2}(s)}{L_{X_1}(as)}, \\
&= \left[\frac{\lambda_2}{\lambda_1}\frac{\lambda_1 + as}{\lambda_2 + s}\right]^\alpha, \\
&= \left[(1-p)\frac{\lambda_2}{\lambda_2 + s} + p\right]^\alpha \text{ where } p = \frac{a\lambda_2}{\lambda_1}.
\end{aligned}
$$

Since the Laplace transform uniquely determines the pdf of a rv (Feller [14], the unknown rv $Z$, is the sum of $\alpha$ independent rv's, each being the product of two independent rvs: a Bernoulli rv with mean $(1 - p)$ and an exponential rv with parameter $\lambda_2$, where, $p = \frac{a\lambda_2}{\lambda_1}$. The probability density function of $Z$ is:

$$
f_Z(z) = \sum_{j=0}^\alpha \binom{\alpha}{\alpha - j} p^{\alpha-j}(1 - p)^j f_{g_j}(z), \quad z \geq 0, \quad \text{by Theorem III.2.1}
$$

Using the independence of $X_1$ and $Z$, we have:

$$\begin{aligned}
f_{X_1,Z}(x_1, z) &= f_1(x_1)f_Z(z) \\
&= p^\alpha f_1(x_1)\delta(z) \\
&\quad + \sum_{j=1}^{\alpha} \binom{\alpha}{\alpha - j} p^{\alpha - j}(1 - p)^j f_{g_j}(z)f_1(x_1) \ .
\end{aligned}$$

$$\begin{aligned}
\text{Then} \quad g(x_1, x_2) &= \int_{-\infty}^{+\infty} f_{X_1,Z}(x_1, z)\delta(ax_1 + z - x_2)\, dz \\
&= \int_{-\infty}^{+\infty} p^\alpha f_1(x_1)\delta(z)\delta(ax_1 + z - x_2)\, dz + \\
&\quad \int_{-\infty}^{+\infty} \sum_{j=1}^{\alpha} \binom{\alpha}{\alpha - j} p^{\alpha - j}(1 - p)^j f_1(x_1)f_{g_k}(z)\delta(ax_1 + z - x_2)\, dz \\
&= \text{Part1} + \text{Part2},
\end{aligned}$$

with

$$\begin{aligned}
\text{Part1} &= \int_{-\infty}^{+\infty} p^\alpha f_1(x_1)\delta(z)\delta(ax_1 + z - x_2)\, dz \\
&= p^\alpha f_1(x_1) \int_{-\infty}^{+\infty} \delta(z)\delta(ax_1 + z - x_2)\, dz \\
&= p^\alpha f_1(x_1)\delta(x_2 - ax_1), \text{ and}
\end{aligned}$$

$$\begin{aligned}
\text{Part2} &= \int_{-\infty}^{+\infty} \sum_{j=1}^{\alpha} \binom{\alpha}{\alpha - j} p^{\alpha - j}(1 - p)^j f_1(x_1)f_{g_k}(z)\delta(ax_1 + z - x_2)\, dz \\
&= \sum_{j=1}^{\alpha} \binom{\alpha}{\alpha - j} p^{\alpha - j}(1 - p)^j f_1(x_1) \int_{-\infty}^{+\infty} f_{g_k}(z)\delta(ax_1 + z - x_2)\, dz \\
&= \sum_{j=1}^{\alpha} \binom{\alpha}{\alpha - j} p^{\alpha - j}(1 - p)^j f_1(x_1)f_{g_k}(x_2 - ax_1).
\end{aligned}$$

Putting together Part1 and Part 2, we obtain,

$$g(x_1, x_2) = \sum_{j=0}^{\alpha} \binom{\alpha}{\alpha - j} p^{\alpha - j}(1 - p)^j f_1(x_1)f_{g_j}(x_2 - ax_1).$$

$\square$

The common shape parameter assumption is quite common in applications for practice Figure 5 describes the joint pdf in (21) for $\lambda_1 = 4$, $\lambda_2 = 1$, $a = 1$, and $\alpha = 2$. The line of discontinuity can be seen and it describes the proportional occurrence case.
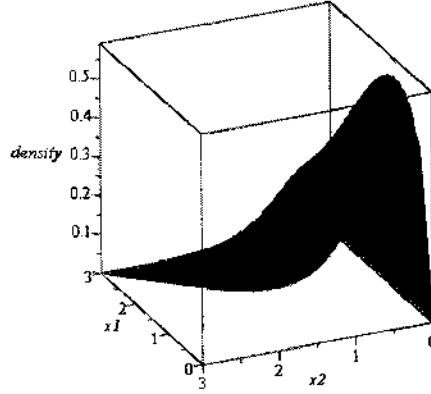
FIG. 5: The pdf of $(X_1, X_2)$.

**Theorem IV.1.2.** *Let the two rv's $X_1$ and $X_2$ be jointly distributed as (21). Then the JLST of $X_1$ and $X_2$ is given by:*

$$L_{X_1,X_2}(s_1, s_2) = \left[\frac{\lambda_1}{\lambda_1 + s_1 + as_2}\right]^\alpha \left[(1-p)\frac{\lambda_1}{\lambda_2 + s} + p\right]^\alpha.$$

Proof: The JLST of $X_1$ and $X_2$ is,

$$
\begin{aligned}
L_{X_1,X_2}(s_1, s_2) &= E[e^{-s_1 X_1 - s_2 X_2}] \\
&= E[e^{-s_1 X_1 - s_2(aX_1 + Z)}] \\
&= E[e^{-(s_1 + as_2)X_1}]E[e^{-s_2 Z}] \text{ Using the independence of } X_1 \text{ and } Z, \\
&= \left[\frac{\lambda_1}{\lambda_1 + s_1 + as_2}\right]^\alpha \left[(1-p)\frac{\lambda_1}{\lambda_2 + s} + p\right]^\alpha.
\end{aligned}
$$

$\square$

Based on (21), the random vector $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ is said to have a Erlang-Erlang bivariate distribution. Its pdf $g(x \mid \theta)$ is given as in (21). Let $\theta = (\alpha, \lambda_1, \lambda_2)$ be the unkonown parameter vector. The variables $X_1$ and $X_2$ originate from Erlang distributions with the same shape parameter $\alpha$, and scale parameters $\lambda_1$, and $\lambda_2$

respectively. Such pdf can be viewed as a finite mixture of $(\alpha + 1)$ component densities. It can also be written as:

$$g(x \mid \theta) = g(x_1, x_2 \mid \theta) = \sum_{j=0}^{\alpha} w_j g_j(x \mid \theta). \tag{22}$$

where

- $w_j = \binom{\alpha}{\alpha-j} p^{\alpha-j}(1-p)^j$, and $\sum_{j=0}^{\alpha} w_j = 1$,

- $g_j(x \mid \theta) = g_j(x_1, x_2 \mid \theta) = f_1(x_1) f_{g_j}(x_2 - ax_1)$, $j = 0, \ldots, \alpha$,

- $g_0(x \mid \theta) = g_0(x_1, x_2 \mid \theta) = f_1(x_1) \delta(x_2 - ax_1)$.

Consider a random sample of size $N \in \mathbb{N}$, bivariate vectors as in (2), $\mathbf{x} = (x_1, \ldots, x_N)^T$, where $x_i = (x_{i1}, x_{i2})^T$. The probability density function of $X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix}$ $i = 1, \ldots, N$ is given by:

$$\begin{aligned} g(x_i \mid \theta) &= g(x_{i1}, x_{i2} \mid \theta) \\ &= \sum_{j=0}^{\alpha} \binom{\alpha}{\alpha - j} p^{\alpha-j}(1-p)^j f_{i1}(x_{i1}) f_{g_j}(x_{i2} - ax_{i1}) , \ i = 1, \ldots, N. \end{aligned}$$

Therefore the likelihood function is given by:

$$L(\theta \mid \mathbf{x}) = \prod_{i=1}^{N} \prod_{j=o}^{\alpha} w_j^{z_{ij}} g_j(x_i \mid \theta)^{z_{ij}}, \tag{23}$$

$$z_{ij} = \begin{cases} 1, & \text{if } x_i \text{ comes from population } j, \\ 0, & \text{otherwise.} \end{cases}$$

Taking the logarithm of (23) gives the following expression:

$$l(\theta \mid \mathbf{x}) = \sum_{i=1}^{N} \sum_{j=o}^{\alpha} z_{ij} \log[w_j g_j(x_i \mid \theta)]. \tag{24}$$

First set

$$Q(\theta^{(t)} = E_{\theta=\theta^{(t)}}[l(\theta \mid \mathbf{x})] = \sum_{i=1}^{N} \sum_{j=o}^{\alpha} \alpha_{ij} \log[w_j g_j(x_i \mid \theta^{(t)})], \tag{25}$$

where

$$\alpha_{ij} = \alpha_{ij}^{(t)} = \alpha_{ij}(\theta^{(t)}) = \frac{w_j g_j(x_i \mid \theta^{(t)})}{\sum_{j=0}^{\alpha} w_j g_j(x_i \mid \theta^{(t)})}, \quad i = 1, \ldots, N, \quad j = 0, \ldots, \alpha.$$

The estimate of $\alpha_{ij} = \alpha_{ij}^{(t)} = \alpha_{ij}(\theta^{(t)})$ is given by

Changing the order of summation in (25), the goal is to maximize the following expression:

$$\sum_{i=1}^{N} \alpha_{ij}(\theta^{(t)}) \log[g_j(x_i \mid \theta(t))], \tag{26}$$

for each fixed $j = 0, \ldots, \alpha$, and get the new set of estimates of the parameters for $\theta^{(t+1)}$.

Let $LL_j = \sum_{i=1}^{N} \alpha_{ij} \log[g_j(x_i \mid \theta(t))]$. The estimates are obtained by dividing the above expression into two components, for $j = 0$ and for $j \geq 1$.
When $j = 0$,

$$
\begin{aligned}
LL_0 &= \sum_{i=1}^{N} \alpha_{i0} \log \left[ g_0(x_i \mid \theta^{(t)}) \right] \\
&= \log \prod_{i=1}^{N} \left[ g_0(x_i \mid \theta^{(t)}) \right]^{\alpha_{i0}} \\
&= \log \prod_{i=1}^{N} \left[ \frac{\lambda_{10}^{\alpha_0}}{\Gamma(\alpha_0)} x_{i1}^{\alpha_0-1} e^{-\lambda_{10} x_{i1}} \right]^{\alpha_{i0}} \\
&= \sum_{i=1}^{N} \alpha_{i0} \log \left[ \frac{\lambda_{10}^{\alpha_0}}{\Gamma(\alpha_0)} x_{i1}^{\alpha_0-1} e^{-\lambda_{10} x_{i1}} \right] \\
&= \alpha_0 \log \lambda_{10} \sum_{i=1}^{N} \alpha_{i0} - \log \Gamma(\alpha_0) \sum_{i=1}^{N} \alpha_{i0} + (\alpha_0 - 1) \sum_{i=1}^{N} \alpha_{i0} \log x_{i1} - \lambda_{10} \sum_{i=1}^{N} \alpha_{i0} x_{i1}.
\end{aligned}
$$

Taking the derivative of $LL_0$ with respect to $\alpha_0$ gives

$$\frac{\partial LL_0}{\partial \alpha_0} = \log \lambda_{10} \sum_{i=1}^{N} \alpha_{i0} - \frac{\Gamma'(\alpha_0)}{\Gamma(\alpha_0)} \sum_{i=1}^{N} \alpha_{i0} + \sum_{i=1}^{N} \alpha_{i0} \log x_{i1}.$$

Setting this equal to zero, we have:

$$\log\lambda_{10} - \frac{\Gamma'(\alpha_0)}{\Gamma(\alpha_0)} = \frac{-\sum_{i=1}^{N} \alpha_{i0}\log x_{i1}}{\sum_{i=1}^{N} \alpha_{i0}}.$$

Similarly taking the derivative of $LL_0$ with respect to $\lambda_{10}$ gives

$$\frac{\partial LL_0}{\partial \lambda_{10}} = \frac{\alpha_0}{\lambda_{10}} \sum_{i=1}^{N} \alpha_{i0} - \sum_{i=1}^{N} \alpha_{i0}x_{i1}.$$

Setting this equal to zero, we have:

$$\frac{\alpha_0}{\lambda_{10}} = \frac{\sum_{i=1}^{N} \alpha_{i0}x_{i1}}{\sum_{i=1}^{N} \alpha_{i0}}.$$

When $j > 0$,

$$
\begin{aligned}
LL_j &= \sum_{i=1}^{N} \alpha_{ij}\log\left[g_j(x_i \mid \theta^{(t)})\right] \\
&= \log \prod_{i=1}^{N} \left[g_j(x_i \mid \theta^{(t)})\right]^{\alpha_{ij}} \\
&= \log \prod_{i=1}^{N} \left[\frac{\lambda_{1j}^{\alpha_j}}{\Gamma(\alpha_j)}x_{i1}^{\alpha_j-1}e^{-\lambda_{1j}x_{i1}}\frac{\lambda_{2j}^{\alpha_j}}{\Gamma(j)}(x_{i2}-ax_{i1})^{j-1}e^{-\lambda_{2j}(x_{i2}-ax_{i1})}\right]^{\alpha_{ij}} \\
&= \sum_{i=1}^{N} \alpha_{ij}\log\left[\frac{\lambda_{1j}^{\alpha_j}\lambda_{2j}^{j}}{\Gamma(\alpha_j)\Gamma(j)}x_{i1}^{\alpha_j-1}(x_{i2}-ax_{i1})^{j-1}e^{-(\lambda_{1j}-\lambda_{2j}a)x_{i1}-\lambda_{2j}x_{i2}}\right] \\
&= \alpha_j\log\lambda_{1j}\sum_{i=1}^{N}\alpha_{ij} + j\log\lambda_2\sum_{i=1}^{N}\alpha_{ij} - \log\Gamma(\alpha_j)\sum_{i=1}^{N}\alpha_{ij} + (\alpha_j-1)\sum_{i=1}^{N}\alpha_{ij}\log x_{i1} \\
&\quad -(\lambda_{1j}-\lambda_{2j}a)\sum_{i=1}^{N}x_{i1}\alpha_{ij} - \lambda_{2j}\sum_{i=1}^{N}x_{i2}\alpha_{ij} + K,
\end{aligned}
$$

where $K$ represents a constant independent of $\alpha_j$, $\lambda_{1j}$, and $\lambda_{2j}$, whose value may change at each occurrence.

Taking the derivative of $LL_j$ with respect to $\alpha_j$ gives

$$\frac{\partial LL_j}{\partial \alpha_j} = \log\lambda_{1j}\sum_{i=1}^{N}\alpha_{ij} - \frac{\Gamma'(\alpha_j)}{\Gamma(\alpha_j)}\sum_{i=1}^{N}\alpha_{ij} + \sum_{i=1}^{N}\alpha_{ij}\log x_{i1}.$$

Setting this equal to zero, we have

$$\log\lambda_{1j} - \frac{\Gamma'(\alpha_j)}{\Gamma(\alpha_j)} = \frac{-\sum_{i=1}^{N}\alpha_{ij}\log x_{i1}}{\sum_{i=1}^{N}\alpha_{ij}}.$$

Taking the derivative of $LL_j$ with respect to $\lambda_{1j}$ gives

$$\frac{\partial LL_j}{\partial\lambda_{1j}} = \frac{\alpha_j}{\lambda_{1j}}\sum_{i=1}^{N}\alpha_{ij} - \sum_{i=1}^{N}\alpha_{ij}x_{i1}.$$

Setting this equal to zero, we have

$$\frac{\alpha_j}{\lambda_{1j}} = \frac{\sum_{i=1}^{N}\alpha_{ij}x_{i1}}{\sum_{i=1}^{N}\alpha_{ij}}.$$

Taking the derivative of $LL_j$ with respect to $\lambda_{2j}$ gives

$$\frac{\partial LL_j}{\partial\lambda_{2j}} = \frac{j}{\lambda_{2j}}\sum_{i=1}^{N}\alpha_{ij} + a\sum_{i=1}^{N}\alpha_{ij}x_{i1} - \sum_{i=1}^{N}\alpha_{ij}x_{i2}.$$

Setting this equal to zero, we have

$$\lambda_{2j} = \frac{j\sum_{i=1}^{N}\alpha_{ij}}{\sum_{i=1}^{N}\alpha_{ij}x_{i2} - a\sum_{i=1}^{N}\alpha_{ij}x_{i1}}.$$

## IV.2 EXPONENTIAL-EXPONENTIAL CASE

When $\alpha = 1$, the mixture distribution in (21) reduces to a mixture with weights $p$ and $(1-p)$. More specifically, when $X_1 \sim exponential(\lambda_1)$ and $X_2 \sim exponential(\lambda_2)$, the joint probability density function of $(X_1, X_2)$ is given by:

$$g(x_1, x_2) = p\lambda_1 e^{-\lambda_1 x_1}\delta(x_2 - ax_1) + (1-p)\lambda_1\lambda_2 e^{-\lambda_2 x_2}e^{-(\lambda_1 - a\lambda_2)x_1}I(x_2 > ax_1) \quad (27)$$

where the rv's $X_1$ and $X_2$ are related as in $X_2 = aX_1 + Z$.

Let $r$ index of proportional occurrence between the two events.

$$r = I(x_2 - ax_1) = \begin{cases} 1, & \text{if } x_2 - ax_1 = 0; \\ 0, & \text{if } x_2 - ax_1 > 0. \end{cases}$$

Then the bivariate exponential density in (27) can be written as:

$$g(x_1, x_2) = [p\lambda_1 e^{-\lambda_1 x_1}]^r [(1-p)\lambda_1 \lambda_2 e^{-\lambda_2 x_2} e^{-(\lambda_1 - a\lambda_2)x_1}]^{1-r}. \tag{28}$$

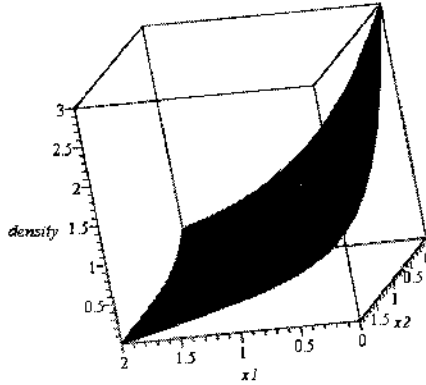Figure 6 describes the joint pdf in (27) for $\lambda_1 = 4$, $a = 1$, and $\lambda_2 = 1$.



FIG. 6: The joint pdf of $(X_1, X_2)$.

## IV.3 WEIBULL-WEIBULL MODEL

Consider two rv's $Y_1$, $Y_2$ from (5) with the same $\beta$ and $\lambda_1$, $\lambda_2$. Let these two rvs be related as $Y_2^\beta = aY_1^\beta + Z_1$, with $a$ a nonnegative fixed constant called the coefficient of linear relationship, and $Z_1$ is independent of $Y_1$. Then the joint density function of $(Y_1, Y_2)$ is given as (Carpenter et al. [6]):

$$h(y_1, y_2) = p\lambda_1 \beta y_1^{\beta-1} e^{-\lambda_1 y_1^\beta} \delta(y_1, y_2) \tag{29}$$
$$+ (1-p)\lambda_1 \lambda_2 \beta^2 y_1^{\beta-1} y_2^{\beta-1} e^{-\lambda_2 y_2^\beta} e^{-(\lambda_1 - a\lambda_2)y_1^\beta} I(y_1, y_2),$$

where

- $\delta(y_1, y_2) = \delta(y_2 - a^{\frac{1}{\beta}} y_1)$,

- $I(y_1, y_2) = I(y_2 > a^{\frac{1}{\beta}} y_1)$.

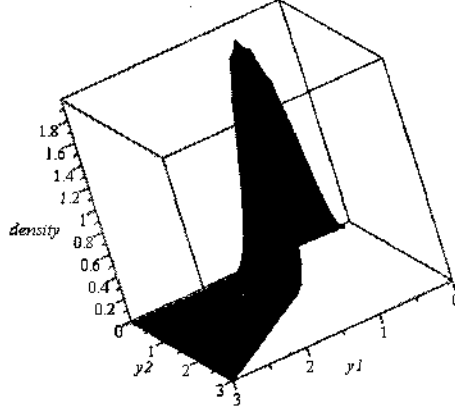Figure 7 describes the joint pdf in (29) for $\lambda_1 = 4$, $\lambda_2 = 1$, $a = 1$, and $\beta = 3$.



FIG. 7: The joint pdf of $(Y_1, Y_2)$.

Notation

Let $r$ denote the index of proportional occurrence between the two events.

$$r = I(y_2 - a^{\frac{1}{\beta}} y_1) = \begin{cases} 1, & \text{if } y_2 - a^{\frac{1}{\beta}} y_1 = 0; \\ 0, & \text{if } y_2 - a^{\frac{1}{\beta}} y_1 > 0. \end{cases}$$

Then the joint probability density function of $(Y_1, Y_2)$ can be written as:

$$h(y_1, y_2) = [p\lambda_1 \beta y_1^{\beta-1} e^{-\lambda_1 y_1^{\beta}}]^r [(1-p)\lambda_1 \lambda_2 \beta^2 y_1^{\beta-1} y_2^{\beta-1} e^{-\lambda_2 y_2^{\beta}} e^{-(\lambda_1 - a\lambda_2) y_1^{\beta}}]^{1-r} . \quad (30)$$

**Theorem IV.3.1.** *Suppose $(Y_{1i}, Y_{2i})$, $i = 1, \ldots, n$ is a random sample of size $n$ from (30). Then the joint maximum likelihood estimators $(\widehat{\lambda_1}, \widehat{\lambda_2}, \widehat{\beta})$ of $(\lambda_1, \lambda_2, \beta)$ are given by,*

$$\widehat{\lambda}_1 = \frac{a}{\overline{Y}_2^*} + \frac{n-k}{n\overline{Y}_1^*},$$

$$\widehat{\lambda}_2 = \frac{1}{\overline{Y}_2^*},$$

*and $\widehat{\beta}$ is the solution to*

$$\widehat{\lambda}_1 \sum_{i=1}^{n} Y_{1i}^{\widehat{\beta}} \log Y_{1i} + \widehat{\lambda}_2 \sum_{i=1}^{n} (1-r_i)[Y_{2i}^{\widehat{\beta}} \log Y_{2i} - a Y_{1i}^{\widehat{\beta}} \log Y_{1i}] = \frac{2n}{\widehat{\beta}}$$

$$-\frac{\sum_{i=1}^{n} r_i}{\widehat{\beta}} + \sum_{i=1}^{n} r_i \log Y_{1i} + \sum_{i=1}^{n} (1-r_i) \log(Y_{1i}Y_{2i}),$$

*where*

- $\overline{Y}_1^* = \frac{\sum_{i=1}^{n} Y_{1i}^{\widehat{\beta}}}{n}$ ,

- $\overline{Y}_2^* = \frac{\sum_{i=1}^{n} Y_{2i}^{\widehat{\beta}}}{n}$,

- $k = \sum_{i=1}^{n} r_i$.

Proof: The log-likelihood function is:

$$
\begin{aligned}
l(\lambda_1, \lambda_2, \beta) &= \log \prod_{i=1}^{n} h(y_{1i}, y_{2i}) \\
&= \sum_{i=1}^{n} \log[h(y_{1i}, y_{2i})] \\
&= \log(a\lambda_2\beta) \sum_{i=1}^{n} r_i + (\beta - 1) \sum_{i=1}^{n} r_i \log Y_{1i} - \lambda_1 \sum_{i=1}^{n} r_i Y_i^{\beta} \\
&\quad + \log(\lambda_1 - a\lambda_2) \sum_{i=1}^{n} (1 - r_i) + \log(\lambda_2\beta^2) \sum_{i=1}^{n} (1 - r_i) \\
&\quad + (\beta - 1) \sum_{i=1}^{n} (1 - r_i) \log(Y_{1i}Y_{2i}) - \lambda_2 \sum_{i=1}^{n} (1 - r_i) Y_{2i}^{\beta} \\
&\quad - (\lambda_1 - a\lambda_2) \sum_{i=1}^{n} (1 - r_i) Y_{1i}^{\beta}.
\end{aligned}
$$

Taking the partial derivative of $l(\lambda_1, \lambda_2, \beta)$ with respect to $\lambda_1$ gives

$$\frac{\partial l(\lambda_1, \lambda_2, \beta)}{\partial \lambda_1} = \frac{\sum_{i=1}^{n}(1 - r_i)}{(\lambda_1 - a\lambda_2)} - \sum_{i=1}^{n} Y_{1i}^{\beta}.$$

Taking the partial derivative of $l(\lambda_1, \lambda_2, \beta)$ with respect to $\lambda_2$ gives

$$\frac{\partial l(\lambda_1, \lambda_2, \beta)}{\partial \lambda_2} = \frac{\sum_{i=1}^{n} r_i}{\lambda_2} - \frac{a\sum_{i=1}^{n}(1 - r_i)}{(\lambda_1 - a\lambda_2)} + \frac{\sum_{i=1}^{n}(1 - r_i)}{\lambda_2}$$
$$- \sum_{i=1}^{n}(1 - r_i)Y_{2i}^{\beta} + a\sum_{i=1}^{n}(1 - r_i)Y_{1i}^{\beta}.$$

Taking the partial derivative of $l(\lambda_1, \lambda_2, \beta)$ with respect to $\beta$ gives

$$\frac{\partial l(\lambda_1, \lambda_2, \beta)}{\partial \beta} = \frac{\sum_{i=1}^{n} r_i}{\beta} + \sum_{i=1}^{n} r_i \log Y_{1i} - \lambda_1 \sum_{i=1}^{n} r_i Y_{1i}^{\beta} \log Y_{1i}$$
$$+ \frac{2\sum_{i=1}^{n}(1 - r_i)}{\beta} + \sum_{i=1}^{n}(1 - r_i) \log(Y_{1i}Y_{2i})$$
$$- \lambda_2 \sum_{i=1}^{n}(1 - r_i)Y_{2i}^{\beta} \log Y_{2i} - (\lambda_1 - a\lambda_2) \sum_{i=1}^{n}(1 - r_i)Y_{1i}^{\beta} \log Y_{1i}.$$

Setting $\dfrac{\partial l(\lambda_1, \lambda_2, \beta)}{\partial \lambda_1}$ equal to zero, we have

$$\widehat{\lambda}_1 - a\widehat{\lambda}_2 = \frac{\sum_{i=1}^{n}(1 - r_i)}{\sum_{i=1}^{n} Y_{1i}^{\widehat{\beta}}}. \tag{31}$$

Setting $\dfrac{\partial l(\lambda_1, \lambda_2, \beta)}{\partial \lambda_2}$ equal to zero, we have

$$\frac{n}{\widehat{\lambda}_2} = \sum_{i=1}^{n} Y_{2i}^{\widehat{\beta}} - \sum_{i=1}^{n} r_i Y_{2i}^{\widehat{\beta}} + a\sum_{i=1}^{n} r_i Y_{1i}^{\widehat{\beta}}. \tag{32}$$

Setting $\dfrac{\partial l(\lambda_1, \lambda_2, \beta)}{\partial \beta}$ equal to zero, we have

$$\widehat{\lambda}_1 \sum_{i=1}^{n} Y_{1i}^{\widehat{\beta}} \log Y_{1i} + \widehat{\lambda}_2 \sum_{i=1}^{n}(1 - r_i)[Y_{2i}^{\widehat{\beta}} \log Y_{2i} - aY_{1i}^{\widehat{\beta}} \log Y_{1i}] =$$
$$\frac{2n}{\widehat{\beta}} - \frac{\sum_{i=1}^{n} r_i}{\widehat{\beta}} + \sum_{i=1}^{n} r_i \log Y_{1i} + \sum_{i=1}^{n}(1 - r_i) \log(Y_{1i}Y_{2i}). \tag{33}$$

Solving for $\widehat{\lambda}_2$ in (32), we have

$$
\begin{aligned}
\frac{1}{\widehat{\lambda}_2} &= \frac{\sum_{i=1}^{n} Y_{2i}^{\widehat{\beta}}}{n} - \frac{1}{n} \sum_{i=1}^{n} r_i(Y_{2i}^{\widehat{\beta}} - aY_{1i}^{\widehat{\beta}}), \\
\widehat{\lambda}_2 &= \frac{1}{\overline{Y}_2^*} - \frac{k_1}{n}, \\
&= \frac{1}{\overline{Y}_2^*}.
\end{aligned}
$$

where

- $k_1 = \sum_{i=1}^{n} r_i(Y_{2i}^{\widehat{\beta}} - aY_{1i}^{\widehat{\beta}}) = 0.$

Solving for $\widehat{\lambda}_1$ in (31), and substituting the above value for $\widehat{\lambda}_2$ gives

$$
\widehat{\lambda}_1 = \frac{a}{\overline{Y}_2^*} + \frac{n-k}{n\overline{Y}_1^*}.
$$

## IV.4  GAMMA-ERLANG MODEL

**Theorem IV.4.1.** *Let $X_1 \sim Ga(\lambda_1, \alpha_1)$ and $X_2 \sim Ga(\lambda_2, \alpha_2)$ with $\alpha_1 \in \mathbb{N}, \alpha_2 \in \mathbb{R}$ $\lambda_1, \lambda_2 \in \mathbb{R}, \alpha_1, \alpha_2, \lambda_1, \lambda_2 \geq 0$ and $a\lambda_2 \leq \lambda_1, \alpha_1 \leq \alpha_2$ and $X_1, X_2$ are related as $X_2 = aX_1 + Z$, with $a$ a nonnegative fixed constant called the coefficient of linear relationship, and $Z$ an unknown random variable independent of $X_1$. Then the pdf of $Z$ is:*

$$
f_Z(z) = \sum_{j=0}^{\alpha_1} \binom{\alpha_1}{\alpha_1 - j} p^{\alpha_1 - j}(1-p)^j \frac{\lambda_2^{j+\alpha_2-\alpha_1}}{\Gamma(j+\alpha_2-\alpha_1)} z^{j+\alpha_2-\alpha_1} e^{-\lambda_2 z}, z \geq 0, \quad (34)
$$

*where*

- $p = \dfrac{a\lambda_2}{\lambda_1}.$

Proof: From the Laplace transforms we obtain,

$$
\begin{aligned}
L_Z(s) &= \frac{L_{X_2}(s)}{L_{X_1}(as)}, \\
&= \left[\frac{\lambda_2}{\lambda_2 + s}\right]^{\alpha_2}\left[\frac{\lambda_1 + as}{\lambda_1}\right]^{\alpha_1} \\
&= \left[\frac{\lambda_2}{\lambda_2 + s}\right]^{\alpha_2 - \alpha_1}\left[\frac{\lambda_2\,(\lambda_1 + as)}{\lambda_1\,(\lambda_2 + s)}\right]^{\alpha_1} \\
&= \left[\frac{\lambda_2}{\lambda_2 + s}\right]^{\alpha_2 - \alpha_1}\left[(1 - p)\frac{\lambda_2}{\lambda_2 + s} + p\right]^{\alpha_1}, \quad \text{where } p = \frac{a\lambda_2}{\lambda_1}.
\end{aligned}
$$

Therefore, $Z$ is the sum of two independent types of rv's:

- The first one is the sum of $\alpha_1$ independent rv's, each being the product of two independent rv's: a Bernoulli rv with mean $(1 - p)$ and an exponential rv with parameter $\lambda_2$, where, $p = \frac{a\lambda_2}{\lambda_1}$,

- and the second one is a gamma rv with scale $\lambda_2$ and shape $\alpha_2 - \alpha_1$.

Let $Z = D + G$, where:

$$
\begin{aligned}
h_D(w) &= \sum_{j=0}^{\alpha_1}\binom{\alpha_1}{\alpha_1 - j}p^{\alpha_1 - j}(1 - p)^j f_{g_j}(w), \quad w \geq 0, \quad \text{by Theorem III.2.1,} \\
h_G(g) &= \frac{\lambda_2^{\alpha_2 - \alpha_1}}{\Gamma(\alpha_2 - \alpha_1)}g^{\alpha_2 - \alpha_1 - 1}e^{-\lambda_2 g}, g > 0.
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
f_Z(z) &= \int_0^\infty h_D(w)h_G(z - w)\,dw, \\
&= \int_0^\infty p^{\alpha_1}\delta(w)h_G(z - w)\,dw + \int_0^\infty \sum_{j=1}^{\alpha_1}\binom{\alpha_1}{\alpha_1 - j}p^{\alpha_1 - j}(1 - p)^j f_{g_j}(w)h_G(z - w)\,dw, \\
&= \text{Part1} + \text{Part2},
\end{aligned}
$$

with

$$
\text{Part1} = p^{\alpha_1}\int_0^\infty \delta(w)h_G(z - w)\,dw.
$$

By change of variables,

$$\text{Part1} = p^{\alpha_1} \int_{-\infty}^{z} \delta(u - z) h_G(u)\, du,$$
$$= p^{\alpha_1} h_G(z), \text{ and}$$

$$\text{Part2} = \sum_{j=1}^{\alpha_1} \binom{\alpha_1}{\alpha_1 - j} p^{\alpha_1 - j}(1 - p)^j \int_0^{\infty} f_{g_j}(w) h_G(z - w)\, dw,$$

$$= \sum_{j=1}^{\alpha_1} \binom{\alpha_1}{\alpha_1 - j} p^{\alpha_1 - j}(1 - p)^j \frac{\lambda_2^{j + \alpha_2 - \alpha_1}}{\Gamma(j)} \frac{e^{-\lambda_2 z}}{\Gamma(\alpha_2 - \alpha_1)}$$
$$\int_0^{\infty} w^{j-1}(z - w)^{\alpha_2 - \alpha_1 - 1}\, dw,$$

$$= \sum_{j=1}^{\alpha_1} \binom{\alpha_1}{\alpha_1 - j} p^{\alpha_1 - j}(1 - p)^j \frac{\lambda_2^{j + \alpha_2 - \alpha_1}}{\Gamma(j)} \frac{e^{-\lambda_2 z}}{\Gamma(\alpha_2 - \alpha_1)}$$
$$z^{\alpha_2 - \alpha_1 + j - 2} \int_0^{\infty} \left(\frac{w}{z}\right)^{j-1} \left(1 - \frac{w}{z}\right)^{\alpha_2 - \alpha_1 - 1}\, dw.$$

By change of variables,

$$\text{Part2} = \sum_{j=1}^{\alpha_1} \binom{\alpha_1}{\alpha_1 - j} p^{\alpha_1 - j}(1 - p)^j \frac{\lambda_2^{j + \alpha_2 - \alpha_1}}{\Gamma(j)} \frac{e^{-\lambda_2 z}}{\Gamma(\alpha_2 - \alpha_1)}$$
$$z^{\alpha_2 - \alpha_1 + j - 1} \int_0^{1} u^{j-1}(1 - u)^{\alpha_2 - \alpha_1 - 1}\, du,$$

$$= \sum_{j=1}^{\alpha_1} \binom{\alpha_1}{\alpha_1 - j} p^{\alpha_1 - j}(1 - p)^j \frac{\lambda_2^{j + \alpha_2 - \alpha_1}}{\Gamma(j + \alpha_2 - \alpha_1)} z^{\alpha_2 - \alpha_1 + j - 1} e^{-\lambda_2 z}.$$

Putting together Part1 and Part2, we obtain,

$$f_Z(z) = p^{\alpha_1} h_G(z) +$$
$$\sum_{j=1}^{\alpha_1} \binom{\alpha_1}{\alpha_1 - j} p^{\alpha_1 - j}(1 - p)^j \frac{\lambda_2^{j + \alpha_2 - \alpha_1}}{\Gamma(j + \alpha_2 - \alpha_1)} z^{\alpha_2 - \alpha_1 + j - 1} e^{-\lambda_2 z}$$

$$= p^{\alpha_1} \frac{\lambda_2^{\alpha_2 - \alpha_1}}{\Gamma(\alpha_2 - \alpha_1)} z^{\alpha_2 - \alpha_1 - 1} e^{-\lambda_2 z} +$$
$$\sum_{j=1}^{\alpha_1} \binom{\alpha_1}{\alpha_1 - j} p^{\alpha_1 - j}(1 - p)^j \frac{\lambda_2^{j + \alpha_2 - \alpha_1}}{\Gamma(j + \alpha_2 - \alpha_1)} z^{\alpha_2 - \alpha_1 + j - 1} e^{-\lambda_2 z}$$

$$= \sum_{j=1}^{\alpha_1} \binom{\alpha_1}{\alpha_1 - j} p^{\alpha_1 - j}(1 - p)^j f_{g_{j + \alpha_2 - \alpha_1}}(z),$$

where

- $f_{g_j}(t) = \dfrac{\lambda_2^{\,j+\alpha_2-\alpha_1}}{\Gamma(j+\alpha_2-\alpha_1)} t^{j+\alpha_2-\alpha_1-1} e^{-\lambda_2 t}, \; t > 0, \text{ for } 1 \le j \le \alpha_1.$

- $f_{g_{\alpha_2-\alpha_1}}(t) = \dfrac{\lambda_2^{\,\alpha_2-\alpha_1}}{\Gamma(\alpha_2-\alpha_1)} t^{\alpha_2-\alpha_1-1} e^{-\lambda_2 t}, \; t > 0.$

$\square$

Figure 8 describes the joint pdf in (34) for $\lambda_1 = 4$, $\lambda_2 = 1$, $a = 1$, $\alpha_1 = 2$ and $\alpha_2 = 3$.
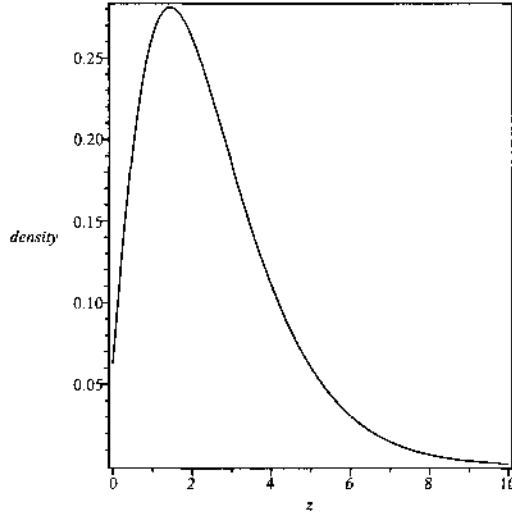


FIG. 8: The pdf of $Z$.

**Theorem IV.4.2.** *Let* $X_1 \sim Ga(\lambda_1, \alpha_1)$ *and* $X_2 \sim Ga(\lambda_2, \alpha_2)$ *with* $\alpha_1 \in \mathbb{N}, \alpha_2 \in \mathbb{R}$ $\lambda_1, \lambda_2 \in \mathbb{R}, \alpha_1, \alpha_2, \lambda_1, \lambda_2 \ge 0$ *and* $a\lambda_2 \le \lambda_1, \alpha_1 \le \alpha_2$ *and* $X_1, X_2$ *are related as* $X_2 = aX_1 + Z$, *with* $a$ *a nonnegative fixed constant called the coefficient of linear relationship, and* $Z$ *an unknown random variable independent of* $X_1$ *. Then the joint pdf of* $(X_1, X_2)$ *is:*

$$\tau_{X_1,X_2}(x_1, x_2) = \sum_{j=0}^{\alpha_1} \binom{\alpha_1}{\alpha_1 - j} p^{\alpha_1 - j}(1-p)^j f_{X_1}(x_1) f_{g_{j+\alpha_2-\alpha_1}}(x_2 - ax_1), \; x_2 - ax_1 \ge 0,$$

$$(35)$$

*where*

- $p = P(X_2 = aX_1) = \frac{a\lambda_2}{\lambda_1}$.

- $f_{g_j}(t) = \frac{\lambda_2^{j+\alpha_2-\alpha_1}}{\Gamma(j+\alpha_2-\alpha_1)} t^{j+\alpha_2-\alpha_1-1} e^{-\lambda_2 t}, t > 0, \text{ for } 1 \le j \le \alpha_1$.

- $f_{g_{\alpha_2-\alpha_1}}(t) = \frac{\lambda_2^{\alpha_2-\alpha_1}}{\Gamma(\alpha_2-\alpha_1)} t^{\alpha_2-\alpha_1-1} e^{-\lambda_2 t}, t > 0$.

- $f_{X_1}(x_1) = \frac{\lambda_1^{\alpha_1}}{\Gamma(\alpha_1)} x_1^{\alpha_1-1} e^{-\lambda_1 x_1}, x_1 > 0$.

The probability density function of $Z$ is:

$$f_Z(z) = \sum_{j=0}^{\alpha_1} \binom{\alpha_1}{\alpha_1-j} p^{\alpha_1-j}(1-p)^j f_{g_{j+\alpha_2-\alpha_1}}(z), \ z \ge 0, \ \text{(by theorem IV.4.1)}.$$

Using the independence of $X_1$ and $Z$, we have:

$$f_{X_1,Z}(x_1,z) = f_1(x_1) f_Z(z),$$

$$= \sum_{j=0}^{\alpha_1} \binom{\alpha_1}{\alpha_1-j} p^{\alpha_1-j}(1-p)^j f_{g_{j+\alpha_2-\alpha_1}}(z) f_{X_1}(x_1),$$

$$\text{Then} \quad \tau_{X_1,X_2}(x_1,x_2) = \int_{-\infty}^{+\infty} f_{X_1,Z}(x_1,z)\delta(ax_1+z-x_2)\,dz,$$

$$= \sum_{j=0}^{\alpha_1} \binom{\alpha_1}{\alpha_1-j} p^{\alpha_1-j}(1-p)^j f_{X_1}(x_1)$$

$$\int_{-\infty}^{+\infty} f_{g_{j+\alpha_2-\alpha_1}}(z)\delta[z-(x_2-ax_1)]\,dz,$$

$$= \sum_{j=0}^{\alpha_1} \binom{\alpha_1}{\alpha_1-j} p^{\alpha_1-j}(1-p)^j f_{X_1}(x_1) f_{g_{j+\alpha_2-\alpha_1}}(x_2-ax_1).$$

$\square$

Figure 9 describes the joint pdf in (35) for $\lambda_1 = 4$, $\lambda_2 = 1$, $a = 1$, $\alpha_1 = 2$ and $\alpha_2 = 3$.
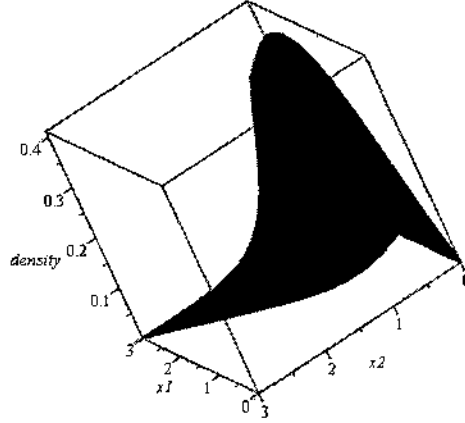
FIG. 9: The joint pdf of $(X_1, X_2)$..

### IV.4.1 Gamma-Gamma Model

**Theorem IV.4.3.** *Let $X_1 \sim Ga(\lambda_1, \alpha_1)$ and $X_2 \sim Ga(\lambda_2, \alpha_2)$ with $\alpha_1, \alpha_2, \lambda_1, \lambda_2 \in \mathbb{R}$ $\alpha_1, \alpha_2, \lambda_1, \lambda_2 \geq 0$, $a\lambda_2 \leq \lambda_1, \alpha_1 \leq \alpha_2$, $[\alpha_1 + 1] = N \leq \alpha_2$ and $X_1, X_2$ are related as $X_2 = aX_1 + Z$, with $a$ a nonnegative fixed constant called the coefficient of linear relationship, and $Z$ an unknown random variable independent of $X_1$. Then the pdf of $Z$ is:*

$$
\begin{aligned}
f_Z(z) &= \sum_{j=0}^{N} \binom{N}{N-j} p^{N-j} (1-p)^j \left(\frac{\lambda_1}{a}\right)^{\theta_1} \lambda_2^{\theta_2+j} \\
&\quad * e^{-(\frac{\lambda_1}{a})z} \frac{1}{\Gamma(\theta_1)\Gamma(\theta_2+j)} \\
&\quad * \int_0^z w^{\theta_2+j-1}(z-w)^{\theta_1-1} e^{-(\lambda_2-\frac{\lambda_1}{a})w} \, dw,
\end{aligned}
$$

*where*

- $N = [\alpha_1 + 1]$.

- $\theta_1 = N - \alpha_1$.

- $\theta_2 = \alpha_2 - N$.

Proof: From the Laplace transforms we obtain,

$$
\begin{aligned}
L_Z(s) &= \frac{L_{X_2}(s)}{L_{X_1}(as)} \\[2mm]
&= \left[\frac{\lambda_2}{\lambda_2 + s}\right]^{\alpha_2}\left[\frac{\lambda_1 + as}{\lambda_1}\right]^{\alpha_1} \\[2mm]
&= \left[\frac{\lambda_2}{\lambda_2 + s}\right]^{\alpha_2 - [\alpha_1 + 1]}\left[\frac{\lambda_1 + as}{\lambda_1}\right]^{\alpha_1 - [\alpha_1 + 1]}\left[\frac{\lambda_2\,(\lambda_1 + as)}{\lambda_1\,(\lambda_2 + s)}\right]^{[\alpha_1 + 1]} \\[2mm]
&= \left[\frac{\lambda_2}{\lambda_2 + s}\right]^{\alpha_2 - N}\left[\frac{\frac{\lambda_1}{a}}{\frac{\lambda_1}{a} + s}\right]^{\theta_1}\left[(1 - p)\frac{\lambda_2}{\lambda_2 + s} + p\right]^{N},
\end{aligned}
$$

where

- $N = [\alpha_1 + 1]$.

- $\theta_1 = N - \alpha_1$.

- $\theta_2 = \alpha_2 - N$.

- $p = \frac{a\lambda_2}{\lambda_1}$.

Therefore, $Z$ is the sum of three types of independent rv's

- $N$ independent rv's, each being the product of two independent rv's: a Bernoulli rv with mean $(1 - p)$ and an exponential rv with parameter $\lambda_2$.

- $Ga(\lambda_2, \theta_2)$ .

- $Ga\left(\frac{\lambda_1}{a}, \theta_1\right)$.

Let $Z = D + G_1 + G_2$, where:

$$
\begin{aligned}
h_D(w) &= \sum_{j=0}^{N}\binom{N}{N - j}p^{N-j}(1 - p)^j f_{g_j}(w), \quad w \geq 0, \quad \text{by Theorem III.2.1,} \\[2mm]
h_{G_1}(g) &= \frac{\lambda_2^{\theta_2}}{\Gamma(\theta_2)}g^{\theta_2 - 1}e^{-\lambda_2 g}, g > 0, \\[2mm]
h_{G_2}(v) &= \left(\frac{\lambda_1}{a}\right)^{\theta_1}\frac{1}{\Gamma(\theta_1)}v^{\theta_1 - 1}e^{-\frac{\lambda_1}{a}v}, v > 0.
\end{aligned}
$$

Now let $Z_1 = D + G_1$, and consider the convolution of the rv's $D$ and $G_1$.

Therefore:

$$
\begin{aligned}
\widehat{f_{Z_1}(z)} &= \int_0^\infty h_D(w) h_{G_1}(z - w)\, dw \\
&= \int_0^z p^N \delta(w) h_{G_1}(z - w)\, dw + \int_0^z \sum_{j=1}^N \binom{N}{N-j} p^{N-j}(1-p)^j f_{g_j}(w) h_{G_1}(z-w)\, dw \\
&= \text{Part1} + \text{Part2},
\end{aligned}
$$

with

$$
\text{Part1} = p^N \frac{\lambda_2^{\theta_2}}{\Gamma(\theta_2)} \int_0^z \delta(w)(z-w)^{\theta_2-1} e^{-\lambda_2(z-w)}\, dw.
$$

By change of variables,

$$
\begin{aligned}
\text{Part1} &= p^N \frac{\lambda_2^{\theta_2}}{\Gamma(\theta_2)} \int_0^z \delta(u-z) u^{\theta_2-1} e^{-\lambda_2 u}\, du \\
&= p^N \frac{\lambda_2^{\theta_2}}{\Gamma(\theta_2)} z^{\theta_2-1} e^{-\lambda_2 z} \quad \text{and}
\end{aligned}
$$

$$
\begin{aligned}
\text{Part2} &= \sum_{j=1}^N \binom{N}{N-j} p^{N-j}(1-p)^j \frac{\lambda_2^{j+\theta_2} e^{-\lambda_2 z}}{\Gamma(j)\Gamma(\theta_2)} \int_0^z w^{j-1}(z-w)^{\theta_2-1}\, dw \\
&= \sum_{j=1}^N \binom{N}{N-j} p^{N-j}(1-p)^j \frac{\lambda_2^{j+\theta_2} e^{-\lambda_2 z}}{\Gamma(j)\Gamma(\theta_2)} z^{j+\theta_2-2} \\
&\quad \int_0^z \left(\frac{w}{z}\right)^{j-1} \left(1 - \frac{w}{z}\right)^{\theta_2-1}\, dw.
\end{aligned}
$$

By change of variables,

$$
\begin{aligned}
\text{Part2} &= \sum_{j=1}^N \binom{N}{N-j} p^{N-j}(1-p)^j \frac{\lambda_2^{j+\theta_2} e^{-\lambda_2 z}}{\Gamma(j)\Gamma(\theta_2)} z^{j+\theta_2-1} \\
&\quad \int_0^1 u^{j-1}(1-u)^{\theta_2-1}\, du \\
&= \sum_{j=1}^N \binom{N}{N-j} p^{N-j}(1-p)^j \frac{\lambda_2^{j+\theta_2}}{\Gamma(j+\theta_2)} z^{j+\theta_2-1} e^{-\lambda_2 z}.
\end{aligned}
$$

Putting together Part1 and Part2, we obtain,

$$\widehat{f_{Z_1}(z)} = \sum_{j=0}^{N} \binom{N}{N-j} p^{N-j}(1-p)^j \frac{\lambda_2^{j+\theta_2}}{\Gamma(j+\theta_2)} z^{j+\theta_2-1} e^{-\lambda_2 z}.$$

Finally let $Z = Z_1 + G_2$, and consider the convolution of the rvs $Z_1$ and $G_2$, where:

$$\widehat{f_{Z_1}(w)} = \sum_{j=0}^{N} \binom{N}{N-j} p^{N-j}(1-p)^j \frac{\lambda_2^{j+\theta_2}}{\Gamma(j+\theta_2)} w^{j+\theta_2-1} e^{-\lambda_2 w}, w \geq 0, \text{ and}$$

$$h_{G_2}(v) = \left(\frac{\lambda_1}{a}\right)^{\theta_1} \frac{1}{\Gamma(\theta_1)} v^{\theta_1-1} e^{-\frac{\lambda_1}{a}v}, \ v > 0.$$

Therefore:

$$
\begin{aligned}
f_Z(z) &= \int_0^\infty \widehat{f_{Z_1}(w)}(w) h_{G_1}(z-w)\,dw \\
&= \int_0^z \sum_{j=0}^{N} \binom{N}{N-j} p^{N-j}(1-p)^j \frac{\lambda_2^{j+\theta_2}}{\Gamma(j+\theta_2)} w^{j+\theta_2-1} e^{-\lambda_2 w} \\
&\quad \left(\frac{\lambda_1}{a}\right)^{\theta_1} \frac{1}{\Gamma(\theta_1)} (z-w)^{\theta_1-1} e^{-\frac{\lambda_1}{a}(z-w)}\,dw \\
&= \sum_{j=0}^{N} \binom{N}{N-j} p^{N-j}(1-p)^j \left(\frac{\lambda_1}{a}\right)^{\theta_1} \lambda_2^{j+\theta_2} \\
&\quad \frac{e^{-(\frac{\lambda_1}{a})z}}{\Gamma(\theta_1)\Gamma(j+\theta_2)} \int_0^z w^{j+\theta_2-1}(z-w)^{\theta_1-1} e^{-(\lambda_2-\frac{\lambda_1}{a})w}\,dw.
\end{aligned}
$$

□

## IV.5 ESTIMATION

### IV.5.1 Estimation in the exponential case

Let $(X_{1i}, X_{2i}), i = 1..n$ represent a random sample from (28), the joint maximum likelihood estimators of $(\lambda_1, \lambda_2)$ are given by (Carpenter et al. [6]):

$$
\begin{aligned}
l(\lambda_1, \lambda_2) &= \log \prod_{i=1}^{n} f(x_{1i}, x_{2i}) \\
&= \sum_{i=1}^{n} \log[f(x_{1i}, x_{2i})] \\
&= \sum_{i=1}^{n} \log\{[p\lambda_1 e^{-\lambda_1 x_i}]^{r_i} \\
&\qquad [(1-p)\lambda_1\lambda_2 e^{-\lambda_2 y_i} e^{-(\lambda_1 - a\lambda_2)x_i}]^{1-r_i}\}, \\
l(\lambda_1, \lambda_2) &= \log(a\lambda_2)\sum_{i=1}^{n} r_i - \lambda_1 \sum_{i=1}^{n} r_i x_{1i} + \log(\lambda_1 - a\lambda_2)\lambda_2 \sum_{i=1}^{n}(1 - r_i) \\
&\qquad -\lambda_1 \sum_{i=1}^{n}(1 - r_i)x_{1i} - \lambda_2 \sum_{i=1}^{n} z_i,
\end{aligned}
$$

$$
\text{and} \quad \hat{\lambda}_1 = \frac{a}{\bar{x}_2} + \frac{n - k}{n\bar{x}_1}, \tag{36}
$$

$$
\hat{\lambda}_2 = \frac{1}{\bar{x}_2}, \tag{37}
$$

where

- $k = \sum_{i=1}^{n} I(x_{2i} - ax_{1i} = 0)$.
- $\bar{x}_1 = \frac{\sum_{i=1}^{n} x_{1i}}{n}$.
- $\bar{x}_2 = \frac{\sum_{i=1}^{n} x_{2i}}{n}$.

The variance covariance matrix of $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ is:

$$
\Sigma = \frac{1}{n} \begin{pmatrix} \lambda_1(\lambda_1 - a\lambda_2) + a^2\lambda_2^2 & a\lambda_2^2 \\ a\lambda_2^2 & \lambda_2^2 \end{pmatrix}. \tag{38}
$$

The traditional estimates are:

$$\hat{\lambda}_1^* = \frac{1}{\bar{x}_1},$$

$$\hat{\lambda}_2^* = \frac{1}{\bar{x}_2}.$$

The main computational cost is incurred by the estimate of the parameter $\lambda_1$, whose value depends on the that has been made on the mean of each sample data. The following example shows that

- Ignoring the dependence comes at the cost of the parameter estimation for $\lambda_1$ as its mean square error (MSE) increases.

- The proposed model estimation evidences their performance in simulated and real data.

**Example IV.5.1.**

Here the above mentioned estimators in the Exponential-Exponential case a simulation study is assessed. The data were generated from a BVE with $\lambda_1 = 4, \lambda_2 = 1$ and $a = 1$. The sample size of $n = 25$ was considered. Using the simulated data the parameters are estimated. The estimated parameters and the estimated variance-covariance matrix is given below.

$$\hat{\lambda} = \begin{pmatrix} 4.1615 \\ 1.0410 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 0.5992 & 0.0471 \\ 0.0471 & 0.0472 \end{pmatrix}.$$

As we can see, there is a close correspondence with its original parameters. Now a real data is considered.

**Example IV.5.2.**

In order to compare the Mean Squared Errors (MSEs) of $\hat{\lambda}_1$ and $\hat{\lambda}_1^*$, we have conducted a simulation study. The data were generated from a Bivariate Exponential (BVE) with $\lambda_1 = 1, \lambda_2 = 1$ and $a = 1$. The random sample size was $n = 25$. In this setting the correlation coefficient coefficient $\rho$ is same as $a$. We defined the % improvement in the MSE (%-imp) as

$$\% - imp = \frac{MSE(\hat{\lambda}_1^*) - MSE(\hat{\lambda}_1)}{MSE(\hat{\lambda}_1^*)} \times 100$$

TABLE 1: MSEs and the Percentage Improvement

| $a$ | MSE($\hat{\lambda}_1$) | MSE($\hat{\lambda}_1^*$) | %-imp |
|------|------------|------------|---------|
| 0.01 | 0.048484 | 0.048904 | 0.8588 |
| 0.05 | 0.045787 | 0.047861 | 4.3334 |
| 0.10 | 0.045554 | 0.049456 | 7.8898 |
| 0.20 | 0.045554 | 0.049456 | 7.8898 |
| 0.30 | 0.039644 | 0.049456 | 19.8399 |
| 0.40 | 0.038226 | 0.049456 | 22.7071 |
| 0.50 | 0.036344 | 0.049456 | 26.5125 |
| 0.60 | 0.036344 | 0.049456 | 26.5125 |
| 0.70 | 0.038701 | 0.049456 | 21.7466 |
| 0.80 | 0.041663 | 0.050368 | 17.2828 |
| 0.90 | 0.041663 | 0.050368 | 17.2828 |
| 0.99 | 0.048457 | 0.048796 | 0.6947 |

Table 1 shows that the perecent improvement increases up to $a = 0.6$ and then decreases. So ignoring the correlation of the data in the analysis has disadvantages when that correlation is between 0.10 and 0.90. When the correlation is close to 0 we can ignore the new technique as $X_2$ and $X_1$ are then independent. When the correlation is close to 1 we can also assume $X_2 = X_1$ and then $Z$ can be ignored.

**Example IV.5.3.**

We fit exponential-exponential bivariate model to the American Football League (AFL) data published by Csorgo and Welsh [10]. The joint MLE's, and the estimated asymptotic variance/covarince matrix of the MLE's (from 36, 37 and 38) are as follows:

$$\hat{\lambda} = \begin{pmatrix} 0.1217 \\ 0.0744 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 2.7 \times 10^{-4} & 1.3 \times 10^{-4} \\ 1.3 \times 10^{-4} & 1.3 \times 10^{-4} \end{pmatrix}.$$

The log-likelihood of the data is given in Figure 10.

**Example IV.5.4.**

We used data from a mammary cancer chemoprevention study (Carpenter et al. [6]) that was carried out to determine if a red wine extract suppresses the incidence
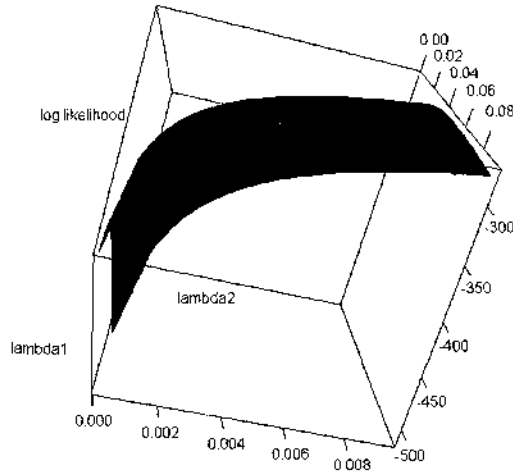
FIG. 10: The log-likelihood.

of dimethylbenzo(a)anthracene induced tumors in transgenic mice. In that study, the mice were randomly assigned two groups: the control group and the treated group. For each group, the number of tumors and the time to appearance of the tumors were recorded. This data is from an Exponential-Exponential bivariate model that was described in (28). Assuming the data follows an Exponential-Exponential bivariate model, we demonstrate that the joint MLE's, and the estimated asymptotic variance/covarince matrix of the MLE's (from 36, 37 and 38) are as follows:
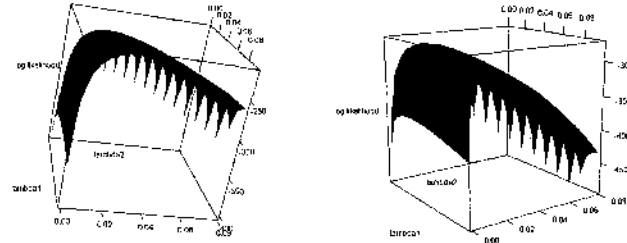
Control group:

$$\hat{\lambda} = \begin{pmatrix} 0.0739 \\ 0.0227 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 1.4 \times 10^{-1} & 1.7 \times 10^{-5} \\ 1.7 \times 10^{-5} & 1.7 \times 10^{-5} \end{pmatrix}.$$

Treated group:

$$\hat{\lambda} = \begin{pmatrix} 0.0379 \\ 0.0095 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 3.8 \times 10^{-5} & 3.1 \times 10^{-6} \\ 3.1 \times 10^{-6} & 3.1 \times 10^{-6} \end{pmatrix}.$$

The log-likelihood of the data is given in Figure 11. Notice that the log-likelihood of the treated group is below the log-likelihood of the control group. Such a difference

would not be easily seen if our proposed methodology was not used. In the next example, the methods as applied to another real data from kidney patients.



(a) The log-likelihood for the control group.

(b) The log-likelihood for the treated group.

FIG. 11: Graphs of the log-likelihood functions

## Example IV.5.5.

Here we consider the complete data from McGilchrist and Aisbett [39], assuming there is no censoring. The data set describes the recurrence times to infection at point of insertion of the catheter for kidney patients who are using portable dialysis equipment.We consider that the random sample data from bivariate Exponential-Exponential distribution. Our suggested model is fitted and the specifications for the coefficient of linear relationship is chosen to be 1 and no censored data. Here we fit an Exponential-Exponential bivariate model that was described in (28). The joint MLE's, and the estimated asymptotic variance/covarince matrix of the MLE's(from 36, 37 and 38) are as follows:

$$\hat{\lambda} = \begin{pmatrix} 0.0139 \\ 0.0049 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 3.9 \times 10^{-6} & 6.4 \times 10^{-7} \\ 6.4 \times 10^{-7} & 6.4 \times 10^{-7} \end{pmatrix}.$$

The log-likelihood of the data is given in Figure 12.

Typically, an analysis of disease types can lead to debate of the differences between associated mortalities, estimates of disease risks and variations. The data was further aggregated into disease types. To address such issues related to McGilchrist and Aisbett [39] kidney data, our model that allows estimates of risk parameters associated with the four types of diseases has been fitted. The four disease types are
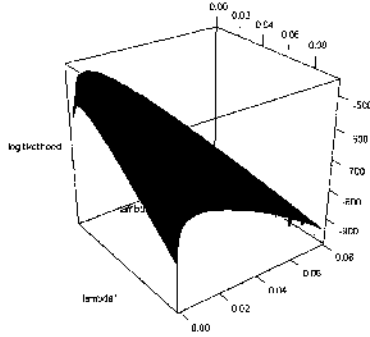
FIG. 12: The log-likelihood.

$0 = $ GN, $1 = $ AN, $2 = $ PKD, and $3 = $ other. Using our construction, each disease type model was fitted separately. The results are presented below. We observed conjugate property that shows differences between those diseases.

For disease type 0, the estimates and the variance-covariance are respectively:

$$\hat{\lambda} = \begin{pmatrix} 0.0143 \\ 0.0055 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 1.7 \times 10^{-5} & 3.4 \times 10^{-6} \\ 3.4 \times 10^{-6} & 3.4 \times 10^{-6} \end{pmatrix}.$$

For disease type 1, the estimates and the variance-covariance are respectively:

$$\hat{\lambda} = \begin{pmatrix} 0.0283 \\ 0.0071 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 5.4 \times 10^{-5} & 4.2 \times 10^{-6} \\ 4.2 \times 10^{-6} & 4.2 \times 10^{-6} \end{pmatrix}.$$

For disease type 2, the estimates and the variance-covariance are respectively:
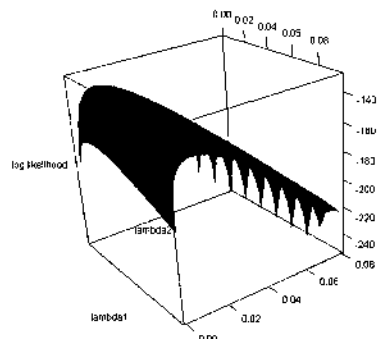
$$\hat{\lambda} = \begin{pmatrix} 0.0145 \\ 0.0038 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 4.2 \times 10^{-5} & 3.6 \times 10^{-6} \\ 3.6 \times 10^{-6} & 3.6 \times 10^{-6} \end{pmatrix}.$$

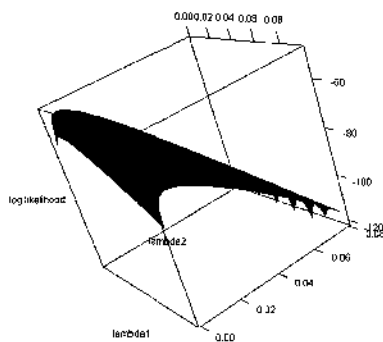For disease type 3, the estimates and the variance-covariance are respectively:

$$\hat{\lambda} = \begin{pmatrix} 0.0095 \\ 0.0038 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 5.3 \times 10^{-6} & 1.1 \times 10^{-6} \\ 1.1 \times 10^{-6} & 1.1 \times 10^{-6} \end{pmatrix}.$$
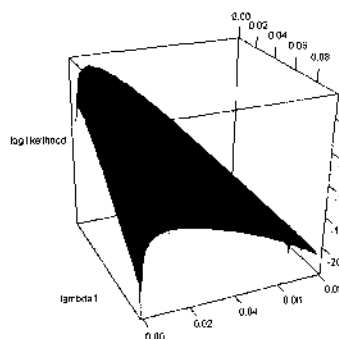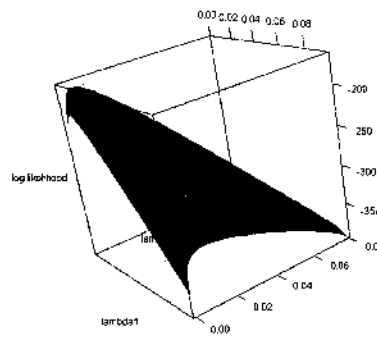
(a) The log-likelihood for the disease type 0.

(b) The log-likelihood for the disease type 1.

(c) The log-likelihood for the disease type 2.

(d) The log-likelihood for the disease type 3.

FIG. 13: Graphs of the log-likelihood functions

The log-likelihoods for the different disease types are given in Figure 13. Such figures along with the parameter estimates show that going from disease type 0 to 3, the likelihood becomes flattened. The estimated variance from our suggested model are smaller than the one proposed by McGilchrist and Aisbett [39]. Without use of prior distribution, the relationship between recurrence time to infection at point of insertion shows that there is substantial differences found and since maximum likelihood estimation was used variance is stable. Indeed, the disease type 0 is easier to cure among them and disease type 4 has no clear remedy.

### IV.5.2  Estimation in the Weibull-Weibull case

The weibull model is another very useful case to consider and is a natural extension of the exponential distribution. Let $(Y_{i1}, Y_{2i})$, $i = 1, \ldots, n$ is a random sample of size $n$ from (30). Then the joint maximum likelihood estimators $(\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\beta})$ of $(\lambda_1, \lambda_2, \beta)$ are given by,

$$
\widehat{\lambda}_1 = \frac{a}{\overline{Y}^*_2} + \frac{n-k}{n\overline{Y}^*_1},
$$

$$
\widehat{\lambda}_2 = \frac{1}{\overline{Y}^*_2},
$$

and $\widehat{\beta}$ is the solution to

$$
\widehat{\lambda}_1 \sum_{i=1}^{n} Y_{1i}^{\widehat{\beta}} \log Y_{1i} + \widehat{\lambda}_2 \sum_{i=1}^{n} (1 - r_i)[Y_{2i}^{\widehat{\beta}} \log Y_{2i} - a Y_{1i}^{\widehat{\beta}} \log Y_{1i}] = \frac{2n}{\widehat{\beta}}
$$

$$
- \frac{\sum_{i=1}^{n} r_i}{\widehat{\beta}} + \sum_{i=1}^{n} r_i \log Y_{1i} + \sum_{i=1}^{n} (1 - r_i) \log(Y_{1i} Y_{2i}),
$$

where

- $\overline{Y}^*_1 = \frac{\sum_{i=1}^{n} Y_{1i}^{\widehat{\beta}}}{n}$,

- $\overline{Y}^*_2 = \frac{\sum_{i=1}^{n} Y_{2i}^{\widehat{\beta}}}{n}$,

- $k = \sum_{i=1}^{n} r_i$.

The Hessian

$$H(\lambda_1, \lambda_2, \beta) = \left( \begin{array}{ccc} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{array} \right),$$

where

- $A_{11} = \dfrac{-\sum_{i=1}^{n}(1-r_i)}{(\lambda_1 - a\lambda_2)^2}$,

- $A_{12} = \dfrac{a \sum_{i=1}^{n}(1-r_i)}{(\lambda_1 - a\lambda_2)^2} = A_{21}$,

- $A_{13} = -\sum_{i=1}^{n} Y_{1i}^{\beta} \log Y_{1i} = A_{31}$,

- $A_{22} = \dfrac{-a^2 \sum_{i=1}^{n}(1-r_i)}{(\lambda_1 - a\lambda_2)^2} - \dfrac{n}{\lambda_2^2}$,

- $A_{23} = -\sum_{i=1}^{n}(1 - r_i)Y_{2i}^{\beta}\log Y_{2i} + a \sum_{i=1}^{n}(1 - r_i)Y_{1i}^{\beta}\log Y_{1i} = A_{32}$,

- $A_{33} = \dfrac{-\sum_{i=1}^{n} r_i}{\beta^2} - \lambda_1 \sum_{i=1}^{n} r_i Y_{1i}^{\beta}(\log Y_{1i})^2 - 2\dfrac{\sum_{i=1}^{n}(1-r_i)}{\beta^2}$
  $- \lambda_2 \sum_{i=1}^{n}(1 - r_i)Y_{2i}^{\beta}(\log Y_{2i})^2 - (\lambda_1 - a\lambda_2) \sum_{i=1}^{n}(1 - r_i)Y_{1i}^{\beta}(\log Y_{1i})^2$.

Fisher's information

$$I(\lambda_1, \lambda_2, \beta) = \left( \begin{array}{ccc} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{array} \right),$$

where

- $B_{11} = \dfrac{n}{\lambda_1(\lambda_1 - a\lambda_2)}$,

- $B_{12} = \dfrac{-na}{\lambda_1(\lambda_1 - a\lambda_2)} = B_{21}$,

- $B_{13} = \frac{n(1-\gamma-\log\lambda_1)}{\beta\lambda_1} = B_{31}$,

- $B_{22} = \frac{na^2}{\lambda_1(\lambda_1-a\lambda_2)} + \frac{n}{\lambda_2^2}$,

- $B_{23} = \frac{n\lambda_2(1-p)}{\beta^2}\left[\frac{1-\gamma-\log\lambda_2}{\lambda_2^2}\right] - \frac{an\lambda_1(1-p)}{\beta}\left[\frac{1-\gamma-\log\lambda_1}{\lambda_1^2}\right] = B_{32}$,

- $B_{33} = \frac{-na\lambda_2}{\lambda_1\beta^2} + \frac{n\lambda_2^2(1-p)}{6\beta^2}\left[\frac{c_1+6(\log\lambda_2)^2+c_2\log\lambda_2}{\lambda_2^2}\right] + \left[\frac{n\lambda_1^2}{\beta^2} - \frac{na\lambda_1\lambda_2(1-p)}{6\beta^2}\right]\left[\frac{c_1+6(\log\lambda_1)^2+c_2\log\lambda_1}{\lambda_1^2}\right]$:
  where

  $$- c_1 = \pi^2 + 6\gamma^2 - 12\gamma,$$

  $$- c_2 = 12(\gamma - 1).$$

## Example IV.5.6.

In order to assess the the above mentioned estimators in the Weibull-Weibull case a simulation study was performed. The data were generated from a Bivariate Weibull (BVW) with $\lambda_1 = 4, \lambda_2 = 1, \beta = 3$ and $a = 1$. The sample size of $n = 25$ was considered. The estimated parameters and the estimated variance-covariance matrix is given below.

MLE's and the estimated variance/covariance matrix.

$$\hat{\lambda} = \begin{pmatrix} 3.9485 \\ 1.1642 \\ 3.0838 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} 0.1191 & 0.0098 & 0.0090 \\ 0.0098 & 0.0099 & -0.0001 \\ 0.0090 & -0.0001 & 0.0114 \end{pmatrix}.$$

The estimated parameters and their variance/covariance imply that the implemented algorithm is very promising. So the likelihood based method is found to give substantial results compared to the independence assumption, reducing biased and applying to the model of interest.

# CHAPTER V

# SEQUENTIAL ESTIMATION

In section II.4, the exponential family type distribution is discussed and the estimation of the mean difference based on the linear model is provided. A sequential estimation procedure is adopted.

Consider some index set $I$, we now consider two classes of exponential families of rv's called $\mathbf{X} = (X_i)_{i \in I}$ and $\mathbf{Y} = (Y_i)_{i \in I}$ with densities

$$f(x_i; \theta) = exp[\theta T(x_i) - b(\theta) + S(x_i)], \tag{39}$$

and

$$f(y_i; \widetilde{\theta}) = exp[\widetilde{\theta} T(y_i) - \widetilde{b}(\widetilde{\theta}) + \widetilde{S}(y_i)]. \tag{40}$$

in the classes $\mathcal{G}_X$ and $\mathcal{G}_Y$ with the following linear relationship:

$$Y_i = aX_i + Z_i, \tag{41}$$

where $i \in I$, $a$ is a fixed positive constant, and $Z_i$'s are unknown rv's whose means are of interest.

The set $I$ is an index countable set that could be finite or infinite. The linear relation described in (41) of association of rv's is not new, but is still a challenging problem. In fact, many authors such as Carpenter et al. [6], Iyer et al. [21], Iyer and Manjunath [22] have suggested its use and importance in applications.

Our goal is to estimate the parameter

$$\lambda = E_{\widetilde{\theta}}[T(\mathbf{Y})] - aE_{\theta}[T(\mathbf{X})] , \tag{42}$$

with squared error loss. When $a = 1$, Equation (42) reduces to the difference between two dependent exponential family of distributions. The dependence concept is the innovation here as in many cases independence is assumed, even if it is known that there is great cost associated with that independence assumption.

## V.1  SEQUENTIAL ANALYSIS

We use the sequential estimation procedure to estimate the mean of the difference of two exponential families distributions with conjugate priors of the gamma or

Bernoulli or Poisson types. This procedure helps address the problem in the small sample size case, maintaining a high power . The approach we use is Bayesian and we assume that $\pi_1(\theta)$ and $\pi_2(\widetilde{\theta})$ are the conjugate priors given by:

$$\pi_1(\theta) \propto \exp[t(\mu_1\theta - b(\theta))], \quad \text{and} \quad \pi_2(\widetilde{\theta}) \propto \exp[s(\mu_2\widetilde{\theta} - \widetilde{b}(\widetilde{\theta}))].$$

This is not a new idea as Diaconis and Ylvisaker [11] adopted this alternative to the maximum likelihood estimation regarding the parameter $\theta$ as a rv with prior distribution, and the inference was based on the posterior distribution. They used this setting in the exponential family with conjugate prior distribution of the parameter $\theta$ given as:

$$\pi(\theta) = \frac{\exp\{t(\mu\theta - \phi(\theta))\}}{\int \exp\{t(\mu\theta - \phi(\theta))\}d\theta}, \tag{43}$$

where

- $\theta \in \Theta$.

- $t$ can be thought as prior sample size.

- $\mu$ is the mean parameter (See also Annis [2]).

In that regard, we obtain that $\mu_1 = E_{\pi_1}[b'(\theta)]$ and $\mu_2 = E_{\pi_2}[\widetilde{b}'(\widetilde{\theta})]$ are prior estimators of $E_\theta[T(\mathbf{X})]$ and $E_{\widetilde{\theta}}[T(\mathbf{Y})]$, respectively.

Hence following the idea by Terbeche et al. [47], the Bayes estimate of $\lambda$, based on a r.s. of size $n$ of $X_1, X_2, \ldots, X_n$ of $\mathbf{X}$, and $Y_1, Y_2, \cdots, Y_n$ of $\mathbf{Y}$ is given by:

$$\begin{aligned}
\widehat{\lambda} &= \widehat{\lambda}(\mathbf{X}, \mathbf{Y}) = \widehat{\lambda}(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \\
&= E[\lambda | X_1, \ldots, X_n, Y_1, \ldots, Y_n] \\
&= E[\widetilde{b}'(\widetilde{\theta}) | Y_1, \ldots, Y_n] - aE[b'(\theta) | X_1, \ldots, X_n],
\end{aligned}$$

where

$$E[b'(\theta) | X_1, \ldots, X_n] = \frac{n\bar{T}_n^{\mathbf{X}} + t\mu_1}{n + t}, \tag{44}$$

and

$$E[\tilde{b}'(\tilde{\theta})|Y_1,\ldots,Y_n] = \frac{n\bar{T}_n^{\mathbf{Y}} + s\mu_2}{n + s}, \tag{45}$$

with $\quad \bar{T}_n^{\mathbf{X}} = \dfrac{T(X_1) + \ldots + T(X_n)}{n}$; and $\quad \bar{T}_n^{\mathbf{Y}} = \dfrac{T(Y_1) + \ldots + T(Y_n)}{n}$.

Hence,

$$\widehat{\lambda} = \frac{n\bar{T}_n^{\mathbf{Y}} + s\mu_2}{n + s} - a\frac{n\bar{T}_n^{\mathbf{X}} + t\mu_1}{n + t}. \tag{46}$$

The asymptotic estimate for the parameter as $n \longrightarrow \infty$ is

$$\widehat{\lambda} = \bar{T}_n^{\mathbf{Y}} - a\bar{T}_n^{\mathbf{X}}. \tag{47}$$

A criteria for stopping the estimation of $\lambda$ is developed.
When $t = s$,

$$\widehat{\lambda} = \frac{n(\bar{T}_n^{\mathbf{Y}} - a\bar{T}_n^{\mathbf{X}}) + t(\mu_2 - a\mu_1)}{n + t} \tag{48}$$

$$= \frac{n}{n + t}(\bar{T}_n^{\mathbf{Y}} - a\bar{T}_n^{\mathbf{X}}) + \frac{t}{n + t}(\mu_2 - a\mu_1). \tag{49}$$

When $t = s = n$,

$$\widehat{\lambda} = \frac{(\bar{T}_n^{\mathbf{Y}} - a\bar{T}_n^{\mathbf{X}}) + (\mu_2 - a\mu_1)}{2}. \tag{50}$$

In the sequential analysis, the sample size is not predetermined. Hence, a natural question is to ask is when is the sample size large enough to make conclusions.

## V.2 STOPPING RULES

The Bayes risk of the estimate $\widehat{\lambda}$ of $\lambda$ with respect to the prior $\pi(\theta)$ in (43) is given by:

$r(\theta, \widehat{\lambda}) = E[R(\theta, \widehat{\lambda})], \quad$ where

$R(\theta, \widehat{\lambda}) = E[L(\theta, \widehat{\lambda})] \quad$ and $\quad L(\theta, \widehat{\lambda}) = (\lambda - \widehat{\lambda})^2$ is the loss function.

In this setting, the Bayes risk is given by:

$$
\begin{aligned}
r(\pi_1, \pi_2) &= r\big(\widehat{\lambda}(\mathbf{X}, \mathbf{Y})\big) \\
&= E_{(\mathbf{X}, \mathbf{Y})}\Big[ E_{\lambda|(\mathbf{X}, \mathbf{Y})}\big(\widehat{\lambda}(\mathbf{X}, \mathbf{Y}) - \lambda\big)^2 \Big] \\
&= E_{(\mathbf{X}, \mathbf{Y})}\Big[ Var\big(\lambda|(\mathbf{X}, \mathbf{Y})\big) \Big] \\
&= E_{(\mathbf{X}, \mathbf{Y})}\Big[ Var\big(\widetilde{b}'(\widetilde{\theta}) - ab'(\theta)|(\mathbf{X}, \mathbf{Y})\big) \Big] \\
&= E_{(\mathbf{X}, \mathbf{Y})}\Big[ Var\big(\widetilde{b}'(\widetilde{\theta})\big) + a^2 Var\big(b'(\theta)\big) - 2a\rho\sqrt{Var\big(\widetilde{b}'(\widetilde{\theta})\big)}\sqrt{Var\big(b'(\theta)\big)} \Big],
\end{aligned}
$$

and the upper bound is achieved using idea of Equation (4) in Terbeche et al. [47]. It is given by

$$
r(\pi_1, \pi_2) = E_{\mathbf{Y}}\Bigg[ E_{\widetilde{\theta}|\mathbf{Y}}|\frac{\widetilde{b}''(\widetilde{\theta})}{n+s}| \Bigg] + a^2 E_{\mathbf{X}}\Bigg[ E_{\theta|\mathbf{X}}|\frac{b''(\theta)}{n+t}| \Bigg], \tag{51}
$$

with equality achieved in (51) when $\rho = corr(\widetilde{b}'(\widetilde{\theta}), b'(\theta)) = corr(\mathbf{X}, \mathbf{Y})$ is minimized.

Considering the loss function

$$
L(\lambda, \widehat{\lambda}, n) = (\lambda - \widehat{\lambda})^2 + cn, \tag{52}
$$

where $c$ can be looked at as the cost of sampling and the decision rule $\Delta = (\tau, \delta)$ where $\tau = \tau_n(\mathbf{x}, \mathbf{y})$ is the stopping rule and $\delta = \delta_n(\mathbf{x}, \mathbf{y})$ is the decision rule, we have that the Bayes risk to minimize from a suitable sample size $n$ obtained sequentially, is given by:

$$
\begin{aligned}
r(\tau, \pi_1, \pi_2) &= E_{(\mathbf{X}, \mathbf{Y}, \tau)}\Bigg[ \frac{U_n}{n+t} + \frac{V_n}{n+s} - 2a\rho\sqrt{Var(\widetilde{b}'(\widetilde{\theta}))}\sqrt{Var(b'(\theta))} + cn \Bigg] \\
&= E_{(\mathbf{Y}, \tau)}\Bigg[ \frac{U_n}{n+t} \Bigg] + E_{(\mathbf{X}, \tau)}\Bigg[ \frac{V_n}{n+s} \Bigg] \\
&\quad + E_{(\mathbf{X}, \mathbf{Y}, \tau)}\Bigg[ - 2a\rho\sqrt{Var(\widetilde{b}'(\widetilde{\theta}))}\sqrt{Var(b'(\theta))} + cn \Bigg],
\end{aligned}
$$

where $U_n = E_{\mathbf{Y}, \tau}[\widetilde{b}''(\widetilde{\theta})]$ and $V_n = E_{\mathbf{X}, \tau}[b''(\theta)]$.

Using ideas in Terbeche et al. [47] to achieve the upper bound in (51), the stopping rule criteria can be expressed as follows:

If $U_n \leq c(n+t)^2$ or if $V_n \leq c(n+s)^2$, then take another pair of observations. Otherwise, stop the collection process. That is the estimation of the difference of the two exponential family can be evaluated from the available informative sample. In other words, the stopping variable is defined by the quantity

$$ n \;\geq\; min\left\{ \sqrt{\frac{U_n}{c}} - t, \sqrt{\frac{V_n}{c}} - s \right\}. \tag{53} $$

In order to study the optimized stopping rule in (53) and its efficiency, a numerical simulation technique is provided in the next section. We consider two exponentially related distributions with gamma priors.

## V.3  SIMULATION

We have described a methodology to compare the mean difference between two exponential distributions that are linearly related. In this section, we show an example of a simulation data of the related bivariate exponential distribution with the different values of the correlations $\rho$.
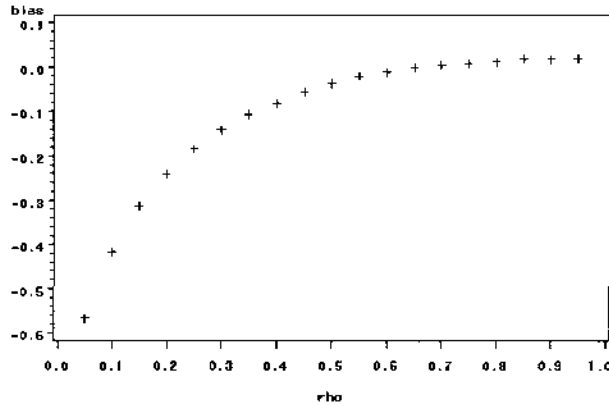


FIG. 14: Graph of the bias from rho for c= 0.

Since we consider two dependent rv's, we create one exponential rv, and create the other one with the desired correlation $\rho$. We generate sample data of size 50. We assume a coefficient of linear relationship $a = 1$ of simultaneous occurrence as described in Marshall and Olkin [32], and $c = 0$ and $c = 0.25$ in (52) over 5000 runs.
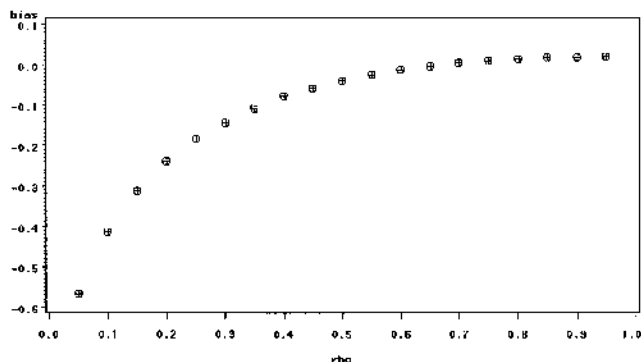
The simulation was carried out using SAS®.



FIG. 15: Graph of the bias from $\rho$ for c= 0.25.

The results of the two figures show that data does not need to be large to achieve convergence. The pattern is the same regardless of the number of runs. Figures 14 and 15 give the bias of the mean difference for $c = 0$ and $c = 0.25$, respectively. The convergence is justified by the maximal error allowed to reach based on the stopping rule. The algorithm performs very well even when the sample size is small, showing great robustness.

The resulting plot of the bias is very helpful in explaining the effectiveness of the estimator. When the correlation is present, this new estimator should be considered. Furthermore, the choice of the cost of re-sampling $c$ does not affect significantly in the error estimation. Setting $c = 0.25$ as in Figure 15 shows the same trend as for Figure 14. The risk is then minimized considerably when the correlation is significant.

## V.4 SUMMARY

The proposed sequential parametric procedure in the estimation of the difference of two exponential distributions is quite useful and relevant. This sequential estimation for the bivariate distributions of the exponential type families is used to get estimate of the mean difference. It is more efficient in terms of bias.

# CHAPTER VI

# CONCLUSION

In this dissertation, we have developed methods for the linearly related type events that have simultaneous/proportional occurrence. To this purpose, we have applied a family of methods known as the Marshall and Olkin procedure which was originally proposed by Marshall and Olkin [32] for tracking systems that can evolve simultaneously in a linear relation. We demonstrate the usefulness of these methods for occurrence of events using a non-zero probability of simultaneous occurrence. Since the event dynamics are identical to other phenomena such as in action/reaction, we have applied these methods to McGilchrist and Aisbett [39] epidemiology case and found results that were not apparent in previously proposed models. Our results support the use of linear relationship in describing related events, due to its relative simplicity and comparative ease of implementation.

Also the subject of this dissertation is to describe the dependencies. Dependencies which are then described in terms of mixing coefficients. We show that there exists a finite mixing sequence quantifying the (relationship) dependence between the two rv's. We investigate the nature of the algorithm and use such result to suggest a robust version for classification. We have proposed new results on fitting a bivariate gamma distribution that is easy to work with because of its analytic form and the one to one correspondence between its parameters. To assess significant risk effects in disease/event associations studies have used models with limited capabilities and no real justification in their usage. Consequently, those results are problematic. The use of joint density function in the linear form after including a latent rv may be a remedy. We then present a robust alternative that avoids multiple testing on model Parameter estimates based on the proposed bivariate model. Such distributions are useful. Our analysis will result in better decisions for various related events.

There is scope for extension of the methods described in this dissertation for the sum independent such distributions. One choice is to consider the case where the probability of proportional occurrence is not the same or to derive the limiting distribution of the sum. Our approach is semi parametric, and is based on known individual related distributions with an unknown component. The field of applications

of such models is wider than the failure time described in the introduction. This approach is also valuable in investigating the relation from bivariate to multivariate models. $\square$

# BIBLIOGRAPHY

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

[2] Annis, D. H., (2007), *A note on quasi-likelihood for exponential families*, Stat. and Prob. Letters, Vol. 77, pp 431-437.

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

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

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

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

[7] Casella, G. and Berger, R.L. (2001) *Statistical Inference*, Duxbury Press $2^{nd}$ edition.

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

[9] Crane, J., Pearce, N., Burgess, C., Woodman, K., Robson, B. and Beasley, R., (1992) *Markers of risk of asthma death or readmission in the 12 months following a hospital admission for asthma*, Int J Epidemiol; 21: pp. 737-44

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

[11] Diaconis, P. and Ylvisaker, D. (1979), *Conjugate Priors for Exponential Families*, Ann. Statist. Vol 7, pp 269-281.

[12] Dirac, P.A.M. (1902) *The Principles of Quantum Mechnics*, Oxford: Clardon Press.

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

[14] Feller, W. (1966) *An Introduction to Probability Theory and Its Applications*, Volume II, Wiley Series in Probability and Mathematical Statistics.

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

[16] Gaver, D.P. (1970) *Multivariate Gamma Distribution Generated by Mixture*, Sankhya, Series A, Vol.32, No.1, pp 123-126.

[17] Holgate, S. Bousquet, J., Wenzel, S., Fox, H., Liu, J. and Castellsague, J. (2001), *Efficacy Efficacy of Omalizumab, an Anti-immunoglobulin E Antibody in Patients With Allergic Asthma at High Risk of Serious Asthma-related Morbidity and Mortality*, Curr Med Res Opin 17(4), pp. 233-240

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

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

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

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

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

[23] Johnson, N. and Kotz, S. (1970) *Distributions in Statistics*, Vol.1 Houghton Mifflin, Boston

[24] Johnson, R.A. , Evans, J.W., and Green, D.W. (1999) *Some Bivariate Distributions for Modeling the strength properties of Lumber*, US Department of Agriculture, Forest Service, Forest Products laboratory, Res. Pap. pp. 11

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

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

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

[28] Krisnaiah, P.R. and Rao, M.M. (1961) *Remarks on a Multivariate Gamma Distribution*, The American Mathematical Monthly, Vol. 68, No. 4. pp 342-346.

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

[30] Leemis, L.M. and McQueston J.T. (2008), *Univariate Distribution Relationships,*The American Statistician, V.62, 1, pp. 4553.

[31] Luke, L.Y. (1975) *Mathematical Functions and their approximations*, New York, Academic Press.

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

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

[34] Mathai, A.M. (1982), *Storage capacity of a dam with gamma types inputs*, Ann. Inst. Statist. Math., Vol. 34 (3), A, pp. 591-597

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

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

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

[38] Mccullagh, P. and Nelder, J.A. (1999), *Generalized Linear Models, 2$^{nd}$* edition, Chapman & Hall/CRC.

[39] McGilchrist, C.A. and Aisbett, C.W. (1991), *Regression with Frality in Survival Analysis*, Biometrics, Vol. 47, pp. 461-466.

[40] Moran, P.A.P. (1970) *The Methodolgy of Rain Making Experiments*, Review of the International Statistical Institute, Vol. 38, No.1, pp. 105-119.

[41] 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

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

[43] 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

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

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

[46] Steel, S. J., Le Roux, N.J. *A Representation of a Bivariate Gamma Extension*, Communication in Statistics Theory and Methods, Vol. 16, No. 1, pp 293-305.

[47] Terbeche, M., Oluyede, B.O. and Barbour, A. (2005) *On sequential and fixed designs for estimation with comparisons and applications*, SORT 29 (2), pp 217-234.

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

# VITA

S. H. Sathish Indika

Department of Department of Mathematics and Statistics

Old Dominion University

Norfolk, VA 23529

Education

College of William & Mary, VA, M.S., Computer Science, 2006.

Clemson University, SC, M.S., Mathematical Sciences-Operations Research, 2004.

New Mexico Institute of Mining & Technology, NM,M.S., Mathematics, 2002.

University of Colombo, Sri Lanka, B.S., Mathematics, 1997.

Publications

1. Sinkovskaya, E., Berkley E., Horton S., Cooper J., Indika S., and Abuhamad A. (2010), *Fetal Cardiac Axis in the First and Early Second Trimester of Pregnancy*, Accepted for publication by the American Journal of Obstetrics & Gynecology.

2. Mohr, M.R., Indika S.H.S., and Hood A.F. (2010), *The Utility of Clinical Photographs in Dermatopathology Diagnosis: A Survey Study*, Accepted for publication by the Archives of Dermatology.

3. Indika, S.H.S., Diawara N., and King R. (2010), *Efficient semi-parametric estimation for marginal correlated small sample data*, Accepted for publication by the Proceedings of Neural, Parallel & Scientific Computations.

4. Indika, S.H.S., Diawara N. (2010), *Linear Dependency for the Difference in Exponential Regression*, Accepted for publication in Involve.

5. Freund B., Petrakos D., Indika S.H.S. (2009), *Error Specific Driving Restriction Allow Continued driving and Time to Transition to Non-Driver Status*, Technical Report, American Alzheimer's Association.

6. Indika S.H.S., Shier D.R. (2006), *Optimization Models for Scheduling of Jobs*, Journal of Research of the National Institute of Standards and Technology, Vol. 111(2), pp.103-111.

7. Indika S.H.S., Hossain A. (2005), *Rayleigh Parameter Estimation using Progressive Type-II Right Censored Data*, Journal of Statistical Studies, Vol. 25, pp.43-53.

Typeset using LaTeX.