2014-01-01

A New Bivariate Distribution With Applications

Sathya Amarasekara
University of Texas at El Paso, truesathya@gmail.com

A NEW BIVARIATE DISTRIBUTION WITH APPLICATIONS

SATHYA AMARASEKARA

Department of Mathematical Sciences

APPROVED:

Panagis Moschopoulos, chair, Ph.D.

Xiaogang Su, Ph.D.

James Michael Wood, Ph.D.

Charles H. Ambler, Ph.D.
Dean of the Graduate School

to my

FAMILY

with love

A NEW BIVARIATE DISTRIBUTION WITH APPLICATIONS

by

SATHYA AMARASEKARA

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Mathematical Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

AUGUST 2014

Acknowledgements

It is my pleasure to express my gratitude to the people who have been influential in the successful completion of this thesis project. There were many who helped and encouraged me toward the success of this thesis project.

First of all I would like to express my special gratitude to my advisor, Dr. Panagis Moschopoulos for his advice, encouragement and support, which have been a tremendous help for me to write this thesis. I am feeling so honored to be the student of the most experienced Statistics professor of the Mathematical Sciences department at The University of Texas at El Paso.

I also extend my appreciation to the other Advisory Committee members, Dr. xiaogang Su from the Mathematical Sciences Department and Dr. James Michael Wood from the Psychology Department, both at The University of Texas at El Paso, for their valuable comments on this work.

I also wish to thank Dr. Ori Rosen, Dr. Amy Wagler, Dr. Joan Staniswalis, Dr. Behzad Djafari-Rouhani and Dr. Lawrence Lesser from the Mathematical Sciences Department at The University of Texas at El Paso for stimulating my knowledge. Also I would like to thank the Department of Mathematical Sciences at UTEP for giving me the opportunity to study statistics, and to perform this work.

Finally, I would like to express my heart-felt gratitude to my dear wife Mrs. Dulanjie Prabuddhika, my family in Sri Lanka and friends in El Paso; I appreciate the support and encouragement you gave throughout my academic life.

Abstract

We construct a bivariate distribution of (X, Y) by assuming that the conditional distribution of Y given X is a two-parameter Gamma $(s, \nu(x))$, where the scale parameter depends on X. If X is assumed to be any distribution, then clearly the joint distribution is well defined in all cases. We will study this new distribution by developing all possible properties, moments, shapes and estimation of parameters with a view towards applications of the distribution to data from many random number generation methods.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The distribution in this thesis is directly related to bivariate distributions obtained from conditional specifications. In this method we assume that the conditional distribution of $Y|X$ and $X|Y$ have certain form. Let $f_X(x)$ and $f_Y(y)$ be the marginal densities of X and Y and $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$ be the conditional distributions respectively. Then the equation

$$f_X(x)\, f_{Y|X}(y|x) = f_Y(y)\, f_{X|Y}(x|y) \tag{1.1}$$

under some compatibility conditions defines a bivariate distribution $f_{X,Y}(x,y)$. If the conditionals are assumed to be gamma with rate parameter only depending on the conditional variables then several models can result from (1.1), see Arnold et al. (1999). One of the models in this reference can be deduced from (1.1) by following Moschopoulos and Staniswalis (1994). The resulting bivariate distribution in this case is

$$f(x,y) = \tfrac{(\beta\delta)^r\gamma^s}{\theta_{r,s}(\delta)\Gamma(s)} x^{r-1}y^{s-1}e^{-(\beta x+\gamma y+\beta\gamma\delta xy)}, \qquad r,\beta,s,\delta>0, x,y>0 \tag{1.2}$$

$$\theta_{r,s}(\delta) = \int_0^\infty t^{r-1}(1+t)^{-s}e^{\frac{-t}{\delta}}\,dt$$

This distribution has been extensively studied by Arnold et al. (1999), Arnold (2008) and Moschopoulos and Sha (2005). The distribution we study here is a simpler form that cannot be obtained by letting β or γ to be zero in (1.2).

We assume that the conditional distribution of Y given X is a two-parameter Gamma$(s, \nu(x))$, that is assuming that the rate parameter depends on X. If X is assumed to have any distribution, then clearly the joint distribution is well defined in all cases. Moshin et. al.

(2012) has used this construction with $\nu(x)$ being the inverse function of X alone, to obtain a bivariate distribution that is applicable in Hydrology.

We assume a linear function for $\nu(x)$. As it turns out, the resulting distribution is

$$f(x,y) = \frac{\beta^r}{\Gamma(r)}\frac{(\delta)}{\Gamma(s)}x^{s+r-1}y^{s-1}e^{-(\beta x+\delta xy)}, \qquad r,\beta,s,\delta > 0, x,y > 0. \tag{1.3}$$

This distribution has not been studied yet. This distribution has gamma and type-2 beta marginals respectively. The distribution is quite simple, considering that other distributions arising from similar arguments are quite complicated. In this thesis we study this new distribution. We develop all possible properties, particular moments, shapes and estimation of parameters with a view towards applications of the distribution to data from many random number generation methods.

The derivation of the new bivariate distribution and its properties will be illustrated in chapter 2. We derived the marginal distribution of Y which is a beta of the second kind and the conditional distribution of $X|Y$ which is a gamma distribution. Moreover, the Product moments function of X and Y and the correlation of the X and Y presented in here. The relationship between beta of type two and beta of type one distributions can be found in this chapter. The estimation of the parameters is carried out using maximum likelihood estimation method and method of moments estimation method. Chapter 2 also includes the properties of maximum likelihood estimators and derivation of fisher information matrix.

Chapter 3 introduces methods to simulate the bivariate distribution using different random number generation methods. This chapter explains direct simulation method of a bivariate distribution when marginal and conditional distributions are given, binomial simulation method by easily utilizing the bivariate density and copula method by making use of marginal distributions. Chapter 3 consist figures of the new bivariate distributions with respective parameter selections.

Chapter 4 proposes a finite mixture of the new bivariate distribution in an attempt to model the data precisely in the presence of positively skewed subpopulations. This chapter demonstrates the proposed methods in a finite mixture model with two new bivariate distributions as subpopulations. This chapter provides details about the estimation of the parameters using Expectation Maximization (EM) algorithm method with the derivations and its implementations.

Chapter 5 performs the analysis of data. The derived properties and methods of the new bivariate model were implemented and verified in this chapter. In this process results for a wide range of parameter values and for different sample sizes were considered in simulation data and used to check whether the properties and methods hold for any of the combinations of parameter values and sample sizes.

Future research will focus applications to real data using the new bivariate distribution and its finite mixture models.

Chapter 2

A New Bivariate Distribution

In this section we derive the new bivariate distribution. This distribution is generated as a compound gamma. The gamma random variable X with shape parameter r and rate parameter β has density

$$f(x|r,\beta) = \tfrac{\beta^r}{\Gamma(r)} x^{r-1} e^{-\beta x}, \qquad r, \beta > 0, x > 0. \tag{2.1}$$

Then we consider a random variable Y such that the conditional density of Y given $X = x$ is a gamma with shape s and rate $\nu(x)$

$$f(y|s,\nu(x)) = \tfrac{(\nu(x))^s}{\Gamma(s)} y^{s-1} e^{-\nu(x)y}, \qquad s, \nu(x) > 0, x, y > 0. \tag{2.2}$$

The joint distribution of X and Y is obtained from (2.1) and (2.2) as

$$f(x,y) = \tfrac{\beta^r}{\Gamma(r)} \tfrac{(\nu(x))^s}{\Gamma(s)} x^{r-1} y^{s-1} e^{-(\beta x + \nu(x) y)}, \qquad r, \beta, s, \nu(x) > 0, x, y > 0. \tag{2.3}$$

The density in (2.3) yields various forms of bivariate distributions depending on the choice of $\nu(X)$.

1. If $\nu(x) = \delta x + \gamma$ then (2.3) becomes a special form of the following bivariate distribution:

$$f(x,y) = \tfrac{(\beta\delta)^r \gamma^s}{\theta_{r,s}(\delta)\Gamma(s)} x^{r-1} y^{s-1} e^{-(\beta x + \gamma y + \beta\gamma\delta xy)}, \qquad r, \beta, s, \delta > 0, x, y > 0, \tag{2.4}$$

$$\theta_{r,s}(\delta) = \int_0^\infty t^{r-1} (1+t)^{-s} e^{\frac{-t}{\delta}} dt.$$

This distribution is the bivariate Gamma distribution presented in Moschopoulos and Staniswalis (1994). As it turn out, the conditional distribution of X given Y is also gamma. This distribution arises from specifying the conditional distributions for a set of models arising from gamma conditionals as Arnold Castillo and Sarabia (1999).

2. If $\nu(x) = \delta/x$ then (2.3) becomes the bivariate distribution (Mohsin et al. (2012)):

$$f(x,y) = \tfrac{\beta^r}{\Gamma(r)}\tfrac{\delta^s}{\Gamma(s)}x^{r-s-1}y^{s-1}e^{-(\beta x+\delta\frac{y}{x})}, \qquad r,\beta,s,\delta > 0, x,y > 0. \tag{2.5}$$

   This distribution is used in Hydrology applications.

3. Our choice is $\nu(x) = \delta x$. Then, the new bivariate distribution is:

$$f(x,y) = \tfrac{\beta^r}{\Gamma(r)}\tfrac{\delta^s}{\Gamma(s)}x^{s+r-1}y^{s-1}e^{-(\beta x+\delta xy)}, \qquad r,\beta,s,\delta > 0, x,y > 0. \tag{2.6}$$

This distribution is clearly a particular case of (2.4) and it is considerably a simpler form. In the following, we derive the marginal distribution of Y and the conditional distribution of X given $Y = y$.

$$\begin{aligned}
f_Y(y|r,\beta,s,\delta) &= \int_0^\infty \frac{\delta^s}{\Gamma(\mathrm{s})}\frac{\beta^r}{\Gamma(r)}y^{s-1}x^{s+r-1}e^{-(\delta y+\beta)x}dx \\
&= \frac{\delta^s\beta^r}{\Gamma(s)\Gamma(r)}y^{s-1}\int_0^\infty x^{s+r-1}e^{-(\delta y+\beta)x}dx \\
&= \frac{\delta^s\beta^r}{\Gamma(s)\Gamma(r)}y^{s-1} \times \frac{\Gamma(s+r)}{(\delta y+\beta)^{s+r}} \\
&= \frac{\delta^s\beta^r\Gamma(s+r)}{\Gamma(s)\Gamma(r)}y^{s-1}(\delta y+\beta)^{-(s+r)}
\end{aligned}$$

$$f_Y(y|r,\beta,s,\delta) = \left(\tfrac{\delta}{\beta}\right)^s \tfrac{\Gamma(s+r)}{\Gamma(s)\Gamma(r)}\tfrac{y^{s-1}}{\left(1+\frac{\delta}{\beta}y\right)^{s+r}}, \qquad r,\beta,s,\delta > 0, x,y > 0 \tag{2.7}$$

This distribution is a special case of a generalized beta type II (GB2) (Tadikamalla (1980)). If $V \sim \text{GB2}(s, r, p, q)$, then the density is given by

$$f_V(v; s, r, p, q) = \left(\frac{p}{q}\right) \frac{\Gamma(s+r)}{\Gamma(s)\Gamma(r)} \frac{\left(\frac{v}{q}\right)^{sp-1}}{\left(1+\left(\frac{v}{q}\right)^{p}\right)^{s+r}}. \tag{2.8}$$

Note that if we take $p = 1$ and $q = \beta/\delta$, the above yields a special form of a generalized beta type II which is known as beta of the second kind (McDonald (1984)).

$$f_V(v; s, r, 1, \frac{\beta}{\delta}) = \left(\frac{\delta}{\beta}\right) \frac{\Gamma(s+r)}{\Gamma(s)\Gamma(r)} \frac{\left(\frac{\delta v}{\beta}\right)^{s-1}}{\left(1+\left(\frac{\delta v}{\beta}\right)^{1}\right)^{s+r}}$$

Thus, the marginal of Y in (2.7) is beta of the second kind with parameters r, β, s, δ.

The mean and variance of Y are obtained from (2.7)

$$E(Y) = \left(\tfrac{\beta}{\delta}\right)\left(\tfrac{s}{r-1}\right) \text{ and } V(Y) = \left(\tfrac{\beta}{\delta}\right)\left(\tfrac{s(s+r-1)}{(r-2)(r-1)^2}\right);\ r > 2.$$

The conditional distribution of X given $Y = y$ can be derived as:

$$\begin{aligned} f\left(x|s, r, \beta, \mu\left(y\right)\right) &= \frac{f\left(x, y\right)}{f\left(y|r, s, \beta, \delta\right)} \\ &= \frac{\frac{\beta^r}{\Gamma(r)}\frac{\delta^s}{\Gamma(s)} x^{s+r-1} y^{s-1} e^{-(\beta x+\delta x y)}}{\left(\frac{\delta}{\beta}\right)^{s} \frac{\Gamma(s+r)}{\Gamma(s)\Gamma(r)} \frac{y^{s-1}}{\left(1+\frac{\delta}{\beta}y\right)^{s+r}}} \\ &= \frac{\beta^{s+r}}{\Gamma(s+r)} x^{s+r-1}\left(1+\frac{\delta}{\beta}y\right)^{s+r} e^{-(\beta x+\delta x y)} \\ &= \frac{(\beta+\delta y)^{s+r}}{\Gamma(s+r)} x^{s+r-1} e^{-(\beta+\delta y)x}. \end{aligned}$$

It's clear that the conditional density of X given $Y = y$ is a gamma with shape $s + r$ and scale $\mu\left(y\right) = \beta + \delta y$. Hence it is a special form of a bivariate distribution with gamma conditionals mentioned by Moschopoulos and Staniswalis (1994). Thus we can give a gamma conditional specification for the new bivariate distribution such that Y given

$X = x$ is a gamma with shape s and scale $\nu(x) = \delta x$ and X given $Y = y$ is a gamma with shape $s + r$ and scale $\mu(y) = \beta + \delta y$.

Now we consider the product moment function of X and Y. The $(i, j) - th$ the product moments of (X, Y) in (2.6) is

$$
\begin{aligned}
E(X^i Y^j) &= E_X[E_{Y|X}(X^i Y^j)] \\
&= E_X[X^i E_{Y|X}(Y^j)] \\
&= E_X[X^i \int_0^\infty y^j \frac{(\delta x)^s}{\Gamma(s)} y^{s-1} e^{-\delta x y} dy] \\
&= E_X[\frac{x^{i+s}\delta^s}{\Gamma(s)} \frac{\Gamma(s+j)}{(\delta x)^{s+j}} \\
&= \frac{\Gamma(s+j)}{\Gamma(s)\delta^j} E_X[x^{i-j}] \\
&= \frac{\Gamma(s+j)}{\Gamma(s)} \delta^{-j} \int_0^\infty x^{i-j} \frac{\beta^r}{\Gamma(r)} x^{r-1} e^{-\beta x} dx \\
&= \frac{\Gamma(s+j)}{\Gamma(s)} \frac{1}{\Gamma(r)} \beta^r \delta^{-j} \frac{\Gamma(i-j+r)}{\beta^{i-j+r}}
\end{aligned}
$$

$$
E(X^i Y^j) = \tfrac{\Gamma(s+j)}{\Gamma(s)} \tfrac{1}{\Gamma(r)} \tfrac{\beta^{j-i}}{\delta^j} \Gamma(i-j+r) \qquad i - j + r > 0 \ . \tag{2.9}
$$

For $i = 1$, $j = 1$ we get

$$
E(XY) = \frac{\Gamma(s+1)}{\Gamma(s)\Gamma(r)} \frac{1}{\delta^1} \Gamma(r) = \frac{s}{\delta}
$$

and since

$$
E(X) = \tfrac{r}{\beta}, \qquad E(Y) = \left(\tfrac{\beta}{\delta}\right)\left(\tfrac{s}{r-1}\right)
$$

we have

$$
\mathrm{Cov}(X, Y) = E(XY) - \mu_X \mu_Y = \frac{s}{\delta} - \left(\frac{r}{\beta}\right)\left(\frac{\beta}{\delta}\right)\left(\frac{s}{r-1}\right),
$$

thus

$$
\mathrm{Cov}(X, Y) = \frac{(-1)s}{\delta(1-r)}. \tag{2.10}
$$

Now,

$$\sigma_X = \tfrac{\sqrt{r}}{\beta}, \qquad \sigma_Y = \left(\tfrac{\beta}{\delta}\right)\sqrt{\tfrac{s(s+r-1)}{(r-2)(r-1)^2}}$$

$$\sigma_X\sigma_Y = \sqrt{\frac{r}{\delta^2}\frac{s(s+r-1)}{(r-s)(r-1)^2}}. \tag{2.11}$$

We get the correlation as

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X\sigma_Y} = \frac{\frac{(-1)s}{\delta(1-r)}}{\sqrt{\frac{r}{\delta^2}\frac{s(s+r-1)}{(r-s)(r-1)^2}}}$$
$$= (-1)\sqrt{\frac{s(r-2)}{r(s+r-1)}}. \tag{2.12}$$

We now note that by taking $w = \frac{\delta}{\beta}y$ and substituting to (2.7) we obtain the beta prime distribution (also known as the standard form of pearson type VI distribution (Keeping (1962)) and inverted beta distribution or beta distribution of the second kind (Johnson et al. (1995)) as

$$w = \frac{\delta}{\beta}y,$$
$$\frac{dg^{-1}(w)}{dw} = \frac{\beta}{\delta},$$
$$f_W(w) = f_Y(g^{-1}(w))\left|\frac{dg^{-1}(w)}{dw}\right|$$
$$= \left(\frac{\delta}{\beta}\right)^s \frac{\Gamma(s+r)}{\Gamma(s)\Gamma(r)}\frac{(\frac{\beta}{\delta}w)^{s^{-1}}}{(1+w)^{s+r}}\left(\frac{\beta}{\delta}\right),$$
$$f(w|r,s) = \tfrac{\Gamma(s+r)}{\Gamma(s)\Gamma(r)}\tfrac{w^{s-1}}{(1+w)^{s+r}}, \qquad r,s>0, w>0. \tag{2.13}$$

We can transform the above to a beta type-I distribution by a simple transformation. We let $z = \dfrac{\frac{\delta}{\beta}y}{1+\frac{\delta}{\beta}y}$, so the transformed probability density function will take the form of

$$f(z) = \tfrac{1}{B(r,s)}z^{s-1}(1-z)^{r-1}, \qquad r,s>0, 0<z<1. \tag{2.14}$$

2.1 Maximum Likelihood Estimation

Here we estimate the four parameters of the new bivariate distribution by using the maximum likelihood method. The log likelihood function of (2.6) can be expressed as

$$\log L(\beta,\delta,r,s) = n\log\varphi + (r+s-1)\sum_{i=1}^{n}\log x_i + (s-1)\sum_{i=1}^{n}\log y_i - \beta\sum_{i=1}^{n}x_i - \delta\sum_{i=1}^{n}x_iy_i \quad (2.15)$$

where

$$\log\varphi = r\log\beta + s\log\delta - \log\Gamma(r) - \log\Gamma(s).$$

The partial derivatives of (2.15) with respective to β, δ, r and s are

$$\frac{\partial \log L(\beta,\delta,r,s)}{\partial\beta} = n\frac{r}{\beta} - \sum_{i-1}^{n} x_i,$$
$$\frac{\partial \log L(\beta,\delta,r,s)}{\partial\delta} = \frac{ns}{\delta} - \sum_{i-1}^{n} x_iy_i,$$
$$\frac{\partial \log L(\beta,\delta,r,s)}{\partial r} = n[\log\beta - \psi(r)] + \sum_{i-1}^{n}\log x_i,$$
$$\frac{\partial \log L(\beta,\delta,r,s)}{\partial s} = n[\log\delta - \psi(s)] + \sum_{i-1}^{n}\log x_i + \sum_{i-1}^{n}\log y_i.$$

Where $\psi(Z)$ is the logarithm derivative of the gamma function given by $\psi(Z) = \frac{d\log\Gamma(Z)}{dZ}$, $Z > 0$. Equating the above expression to zero we get the following maximum likelihood equations:

$$\frac{\hat{r}}{\hat{\beta}} - \frac{\sum_{i=1}^{n} x_i}{n} = 0, \quad (2.16)$$

$$\frac{\hat{s}}{\hat{\delta}} - \frac{\sum_{i=1}^{n} x_iy_i}{n} = 0, \quad (2.17)$$

$$-\log\hat{\beta} + \psi(\hat{r}) - \frac{\sum_{i=1}^{n}\log x_i}{n} = 0, \quad (2.18)$$

$$-\log\hat{\delta} + \psi(\hat{s}) - \frac{\sum_{i=1}^{n}\log x_iy_i}{n} = 0. \quad (2.19)$$

Solving the above nonlinear equations numerically we get estimated values $\hat{\beta}, \hat{\delta}, \hat{r}$ and $\hat{s}$.

2.1.1 Consistency of maximum likelihood estimators

Let $X_1, X_2, \ldots,$ be *iid* $f(x|\theta)$, and let $L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta)$ be the likelihood function. Let $\hat{\theta}$ denote the MLE of θ. For every $\epsilon > 0$ and every $\theta \in \Theta$,

$$\lim_{n\to\infty} P_\theta\left(\left|\hat{\theta} - \theta\right| \geq \epsilon\right) = 0.$$

That is, $\hat{\theta}$ is a consistent estimator of θ.

2.1.2 Asymptotic efficiency of maximum likelihood estimators

Let $X_1, X_2, \ldots,$be *iid* $f(x|\theta)$, and let $\hat{\theta}$ denote the MLE of θ,

$$\sqrt{n}\left[\hat{\theta} - \theta\right] \underset{\longrightarrow}{d} N\left[0, v(\theta)\right].$$

where $v(\theta)$ is the Cramér-Rao Lower Bound and also the inverse of the Fisher information matrix. That is, $\hat{\theta}$ is a consistent and asymptotically efficient estimator of θ.

2.2 Fisher information matrix

The Fisher information matrix $\mathcal{F}$ required to compute the variances of the asymptotic distributions of the maximum likelihood estimates. The matrix $\mathcal{F}$ with corresponding entries can be expressed as

$$\mathcal{F} = n\begin{bmatrix} \frac{r}{\beta^2} & 0 & -\frac{1}{\beta} & 0 \\ 0 & \frac{s}{\delta^2} & 0 & -\frac{1}{\delta} \\ -\frac{1}{\beta} & 0 & \frac{\partial}{\partial r}\psi(r) & 0 \\ 0 & -\frac{1}{\delta} & 0 & \frac{\partial}{\partial s}\psi(s) \end{bmatrix}_{4\times 4}.$$

By taking the inverse of the Fisher information matrix one can calculate the variance co-variances matrix and then obtain variance and covariances of the maximum likelihood estimates.

The elements of the Fisher information matrix can be found by taking the second partial derivatives of the log likelihood function. All the second derivatives of (2.15) take the form

$$\frac{\partial^2 \log L(\beta,\delta,r,s)}{\partial u \partial v} = \frac{n}{\varphi}\frac{\partial^2 \varphi}{\partial u \partial v} - \frac{n}{\varphi^2}\frac{\partial \varphi}{\partial v}.\frac{\partial \varphi}{\partial u}.$$

Since all the second derivatives are constants the elements of $\mathcal{F}$ basically are the negative of the derivatives. Hence we can obtain the elements of $\mathcal{F}$ as

$$\begin{aligned}
E\left(-\frac{\partial \log L(\beta,\delta,r,s)}{\partial \beta^2}\right) &= \frac{nr}{\beta^2},\\
E\left(-\frac{\partial \log L(\beta,\delta,r,s)}{\partial \beta \partial \delta}\right) &= 0,\\
E\left(-\frac{\partial \log L(\beta,\delta,r,s)}{\partial \beta \partial r}\right) &= -\frac{n}{\beta},\\
E\left(-\frac{\partial \log L(\beta,\delta,r,s)}{\partial \beta \partial s}\right) &= 0,\\
E\left(-\frac{\partial^2 \log L(\beta,\delta,r,s)}{\partial \delta^2}\right) &= \frac{ns}{\delta^2},\\
E\left(-\frac{\partial^2 \log L(\beta,\delta,r,s)}{\partial \delta \partial r}\right) &= 0,\\
E\left(-\frac{\partial^2 \log L(\beta,\delta,r,s)}{\partial \delta \partial s}\right) &= -\frac{n}{\delta},\\
E\left(-\frac{\partial^2 \log L(\beta,\delta,r,s)}{\partial r^2}\right) &= \frac{n\partial\psi(r)}{\partial r},\\
E\left(-\frac{\partial^2 \log L(\beta,\delta,r,s)}{\partial r \partial s}\right) &= 0,\\
E\left(-\frac{\partial^2 \log L(\beta,\delta,r,s)}{\partial s^2}\right) &= \frac{n\partial\psi(s)}{\partial s}.
\end{aligned}$$

2.3 Method of moments estimation

In previous sections, we derived estimators using the method of maximum likelihood estimation. Here, we consider estimation of the four parameters by the method of moments. Let's denote the first two sample moments of X and Y as m_1^X, m_2^X and m_1^Y, m_2^Y respectively. Then let's take the corresponding population moments of X and Y as μ_1^X, μ_2^X and μ_1^Y, μ_2^Y.

The sample moments are

$$m_1^X = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X},$$

$$m_2^X = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

and

$$m_1^Y = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y},$$

$$m_2^Y = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

where $\bar{X}$ and $\bar{Y}$ are sample means of X and Y respectively.

The population moments are:

$$\mu_1^X = E(X),$$

$$\mu_2^X = E(X^2)$$

and

$$\mu_1^Y = E(Y),$$

$$\mu_2^Y = E(Y^2).$$

Specializing the moments to our distribution, the moments can be derived by using $(i,j)-th$ the product moments of (X,Y) given in (2.9) for $0 \leq i \leq 2$ and $0 \leq j \leq 2$.

For $i = 1$, $j = 0$ we obtain

$$\mu_1^X = E(X)$$
$$= \frac{\Gamma(s+0)}{\Gamma(s)}.\frac{1}{\Gamma(r)}.\frac{\beta^{0-1}}{\delta^0}(\Gamma(1-0+r)) = \frac{r}{\beta}.$$

For $i = 2$ $j = 0$

$$\mu_2^X = E(X^2)$$
$$= \frac{\Gamma(s+0)}{\Gamma(s)}.\frac{1}{\Gamma(r)}.\frac{\beta^{0-2}}{\delta^0}(\Gamma(2-0+r)) = \frac{\Gamma(r+2)}{\Gamma(r)}.\frac{1}{\beta^2} = \frac{r(r+1)}{\beta^2}.$$

For $i = 0$, $j = 1$

$$\mu_1^Y = E(Y)$$
$$= \frac{\Gamma(s+1)}{\Gamma(s)}.\frac{1}{\Gamma(r)}.\frac{\beta^{1-0}}{\delta^1}(\Gamma(0-1+r))$$
$$= \frac{s\Gamma(s)}{\Gamma(s)}.\frac{\Gamma(r-1)}{\Gamma(r)}.\frac{\beta}{\delta} = \frac{s\beta}{\delta(r-1)}.$$

For $i = 0$ $j = 2$

$$\mu_2^Y = E(Y^2)$$
$$= \frac{\Gamma(s+2)}{\Gamma(s)}.\frac{1}{\Gamma(r)}.\frac{\beta^{2-0}}{\delta^2}(\Gamma(0-2+r))$$
$$= \frac{(s+1)s\Gamma(s)}{\Gamma(s)}.\frac{\Gamma(r-2)}{(r-1)(r-2)\Gamma(r-2)}.\frac{\beta^2}{\delta^2},$$
$$= \left(\frac{\beta}{\delta}\right)^2 \frac{s(s+1)}{(r-2)(r-1)}.$$

We now find the method of moments estimators by equating the first two sample moments of X and Y to the corresponding two population moments of X and Y, and solving the resulting system of simultaneous equations.

By equating $m_1^X = \frac{1}{n}\Sigma_{i=1}^n x_i$ with $\mu_1^X = \frac{r}{\beta}$ we have

$$\bar{X} = \frac{\tilde{r}}{\tilde{\beta}} \tag{2.20}$$

and from $m_2^X = \frac{1}{n}\Sigma_{i=1}^n x_i^2$ with $\mu_2^X = \frac{r(r+1)}{\beta^2}$ give us

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{\tilde{r}(\tilde{r}+1)}{\tilde{\beta}^2} \tag{2.21}$$

since, (2.20) and (2.21) simultaneous equations only contain $\tilde{r}$ and $\tilde{\beta}$, first we can solve for $\tilde{r}$ and $\tilde{\beta}$.

From (2.20) we have

$$\tilde{r} = \overline{X}\tilde{\beta},$$

then by substituting $\tilde{r}$ in (2.21)

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} X_i^2 &= \frac{\tilde{r}(\tilde{r}+1)}{\tilde{\beta}^2} \\
\frac{1}{n}\sum_{i=1}^{n} X_i^2 &= \frac{\overline{X}\tilde{\beta}(\overline{X}\tilde{\beta}+1)}{\tilde{\beta}^2} \\
\frac{1}{n}\sum_{i=1}^{n} X_i^2 &= \frac{\overline{X}^2\tilde{\beta}^2}{\tilde{\beta}^2} + \frac{\overline{X}\tilde{\beta}}{\tilde{\beta}^2} \\
\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 &= \frac{\overline{X}}{\tilde{\beta}} \\
\frac{1}{n}\sum_{i=1}^{n}(X_i^2 - \overline{X})^2 &= \frac{\overline{X}}{\tilde{\beta}}
\end{aligned}$$

$$
\begin{aligned}
\tilde{\beta} &= \frac{\overline{X}}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{x})^2} \\
\tilde{\beta} &= \frac{\overline{X}}{\sigma_X^2}
\end{aligned}
$$

and

$$
\tilde{r} = \frac{\overline{X}^2}{\sigma_X^2}.
$$

Second, we can find $\tilde{s}$ and $\tilde{\delta}$ estimators by equating m_1^Y, m_2^Y and μ_1^Y, μ_2^Y correspondingly.

Letting $m_1^Y = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y} = \mu_1^Y = \frac{s\beta}{\delta(r-1)}$, we obtain

$$
\overline{Y} = \left(\frac{\tilde{\beta}}{\tilde{\delta}}\right)\left(\frac{\tilde{s}}{\tilde{r}-1}\right), \tag{2.22}
$$

then from $m_2^y = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 = \mu_2^y = \left(\frac{\beta}{\delta}\right)^2 \frac{s(s+1)}{(r-1)(r-2)}$ we obtain

$$
\frac{1}{n}\sum_{i=1}^{n} Y_i^2 = \left(\frac{\tilde{\beta}}{\tilde{\delta}}\right)^2 \frac{\tilde{s}(\tilde{s}+1)}{(\tilde{r}-1)(\tilde{r}-2)}. \tag{2.23}
$$

Using (2.22) and (2.23) we can solve for $\tilde{s}$, $\tilde{\delta}$, by rearranging (2.23) as

$$
\frac{1}{n}\sum_{i=1}^{n} Y_i^2 = \left(\frac{\tilde{\beta}}{\tilde{\delta}}\right)^2 \left(\frac{\tilde{s}}{\tilde{r}-1}\right)^2 \frac{(\tilde{r}-1)(\tilde{s}+1)}{\tilde{s}(\tilde{r}-2)}.
$$

Substituting from (2.22) we have

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} Y_i^2 &= \overline{Y}^2 \frac{(\tilde{r}-1)(\tilde{s}+1)}{(\tilde{r}-2)\tilde{s}}, \\
\frac{1}{n}\sum_{i=1}^{n} Y_i^2 &= \frac{\overline{Y}^2(\tilde{r}-1)}{(\tilde{r}-2)} \frac{(\tilde{s}+1)}{\tilde{s}},
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n}.\frac{\sum_{i=1}^{n} Y_i^2}{\overline{Y}^2}.\frac{(\tilde{r}-2)}{(\tilde{r}-1)} &= 1+\frac{1}{\tilde{s}}, \\
\frac{1}{n}.\frac{\sum_{i=1}^{n} Y_i^2}{\overline{Y}^2}.\frac{(\tilde{r}-2)}{(\tilde{r}-1)} - 1 &= \frac{1}{\tilde{s}}, \\
\frac{\frac{1}{n}\sum_{i=1}^{n} Y_i^2(\tilde{r}-2) - \overline{Y}^2(\tilde{r}-1)}{\overline{Y}^2(\tilde{r}-1)} &= \frac{1}{\tilde{s}}, \\
\frac{(\tilde{r}-2)\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2\right) - \overline{Y}^2}{\overline{Y}^2(\tilde{r}-1)} &= \frac{1}{\tilde{s}}, \\
\frac{(\tilde{r}-2)\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 - \overline{Y}^2}{\overline{Y}^2(\tilde{r}-1)} &= \frac{1}{\tilde{s}}, \\
\frac{\overline{Y}^2(\tilde{r}-1)}{(\tilde{r}-2)\sigma_y^2 - \overline{Y}^2} &= \frac{1}{\tilde{s}}.
\end{aligned}
$$

Since $\tilde{r} = \frac{\overline{X}^2}{\sigma_X^2}$

$$
\begin{aligned}
\tilde{s} &= \frac{\overline{Y}^2\left(\frac{\overline{X}^2}{\sigma_X^2} - 1\right)}{\left(\frac{\overline{X}^2}{\sigma_X^2} - 2\right)\sigma_Y^2 - \overline{Y}^2}, \\
\tilde{s} &= \frac{\overline{Y}^2\left(\overline{X}^2 - \sigma_X^2\right)}{\left(\overline{X}^2 - 2\sigma_X^2\right)\sigma_Y^2 - \overline{Y}^2\sigma_X^2}, \\
\tilde{s} &= \frac{\overline{Y}^2\left(\overline{X}^2 - \sigma_X^2\right)}{\overline{X}^2\sigma_Y^2 - 2\sigma_X^2\sigma_Y^2 - \overline{Y}^2\sigma_X^2}.
\end{aligned}
$$

Now from (2.22)

$$
\overline{Y} = \left(\frac{\tilde{\beta}}{\tilde{\delta}}\right)\left(\frac{\tilde{s}}{\tilde{r}-1}\right)
$$

we can obtain

$$
\tilde{\delta} = \left(\frac{\tilde{\beta}}{\overline{Y}}\right)\left(\frac{\tilde{s}}{\tilde{r}-1}\right),
$$

$$\tilde{\delta} = \left(\frac{\frac{\overline{X}}{\sigma_X^2}}{\overline{Y}}\right) \frac{\overline{Y}^2\left(\overline{X}^2 - \sigma_X^2\right)}{\left(\frac{\overline{X}}{\sigma_X^2} - 1\right)\left(\overline{X}^2\sigma_Y^2 - 2\sigma_X^2\sigma_Y^2 - \overline{Y}^2\sigma_X^2\right)},$$

$$\tilde{\delta} = \frac{\overline{XY}\left(\overline{X}^2 - \sigma_X^2\right)}{\left(\overline{X}^2 - \sigma_X^2\right)\left(\overline{X}^2\sigma_Y^2 - 2\sigma_X^2\sigma_Y^2 - \overline{Y}^2\sigma_X^2\right)},$$

$$\tilde{\delta} = \frac{\overline{XY}}{\left(\overline{X}^2\sigma_Y^2 - 2\sigma_X^2\sigma_Y^2 - \overline{Y}^2\sigma_X^2\right)}.$$

Thus, the method of moments estimators are

$$\tilde{\beta} = \frac{\overline{X}}{\sigma_X^2}, \tag{2.24}$$

$$\tilde{\delta} = \frac{\overline{XY}}{\left(\overline{X}^2\sigma_Y^2 - 2\sigma_X^2\sigma_Y^2 - \overline{Y}^2\sigma_X^2\right)}, \tag{2.25}$$

$$\tilde{r} = \frac{\overline{X}^2}{\sigma_X^2}, \tag{2.26}$$

$$\tilde{s} = \frac{\overline{Y}^2\left(\overline{X}^2 - \sigma_X^2\right)}{\overline{X}^2\sigma_Y^2 - 2\sigma_X^2\sigma_Y^2 - \overline{Y}^2\sigma_X^2}. \tag{2.27}$$

Chapter 3

Simulation of The New Bivariate Distribution

Simulation of random samples from the bivariate distribution is useful to examine the behavior of the bivariate distribution and to investigate the estimation methods. Simulation is a numerical technique for conducting experiments on the Computer. If we basically think about statistical simulations it is a way to simulate random events, such that simulated outcomes approximate the population of interest. It is a flexible methodology that we can use to analyze the behavior of the present statistical model. By performing simulations and analyzing the results, we can gain an understanding of how the new bivariate model behaves and how our findings can be applied in this context.

In subsequent sections we give more detailed descriptions of simulation methods for the new bivariate distribution. We will mainly focus on bivariate distributions which are marginally specified and conditionally specified. Also we will give more simulation methods for conditionally specified models.

Some work on the simulations of the conditionally specified models has been performed in Arnold and Strauss (1988) and in in Arnold et al. (1992). In Arnold and Strauss (1988) they observed a straightforward rejection sampling simulation technique and in Arnold et al. (1992) they performed importance sampling which allows to forget about the normalizing constant.

Most of the time in Statistics the Monte Carlo method is been used to simulate the statistical models. Thus we make use of the Monte Carlo simulation method to simulate the new bivariate distribution.

Monte Carlo simulation can be identified as a computer experiment involving random sampling from probability distributions. It is a method of estimating the value of an unknown quantity by making use of the principles of inferential statistics. In brief the guiding principle of inferential statistics is that a random sample tends to exhibit the same properties as the population from which it is drawn.

By following the Monte Carlo simulation method we will generate S independent data sets using a random number generation method of interest. Then we will compute the numerical value of the estimators for each data set. If S is large enough, the summary statistics across the computed estimator values should be good approximations to the true sampling properties of the estimator of interest.
The random number generation methods that we are interested in are;

- Direct method,
- Binomial method,
- Copula method.

A detailed explanation about these methods can be found in the subsequent subsections.

3.1 Direct method

It is very intuitive to show that for bivariate distributions with a gamma marginal and a gamma conditionals or any other specified marginals or conditionals it will be easy to use the direct method.

In marginal and conditional specification we first generate variable X corresponding to its marginal density. Then we generate variable Y corresponding to its conditional density which condioned on the generated X variable.

Thus the new bivariate class can be easily simulated by generating random numbers using a gamma distribution and then by conditioning on those generated values we can generate the other random variable by another conditionally specified gamma distributed random numbers as follows.

$$\begin{aligned} X &\sim \text{Gamma}\,(r, \beta) \\ Y|X &\sim \text{Gamma}\,(s, \delta x) \end{aligned}$$

In conditional specification we will require to take a slightly different approach. It will require to start with a one random number generation using a marginal distribution, which followed by a sequence of random number generate iterations using conditional specifications.

$$\begin{aligned} X &\sim \text{Gamma}\,(r, \beta) \\ Y|X &\sim \text{Gamma}\,(s, \delta x) \\ X|Y &\sim \text{Gamma}\,(r, \beta + \delta y) \end{aligned}$$

3.2 The binomial method

The binomial method can be implemented as a general way of simulating any bivariate distribution of interest. This only requires the density of the bivariate distribution. It is a very flexible method. In this section we will focus on implementing this method to the new bivariate distribution.

We can consider $\mathcal{R}(\triangle_1, \triangle_2) = \{R_{ij}\}_{i=1,\infty}^{j=1,\infty}$ to denote a partition of the support of the bivariate distribution where each

$$R_{ij} = \{(x, y) \ s.t. \ x \in ((i-1)\triangle_1, i\triangle_1] \ and \ y \in ((j-1)\triangle_2, j\triangle_2]\}.$$

Hence each element in the partition $\mathcal{R}$ is a rectangle with area $\triangle_1\triangle_2$ and is centered at the point

$$(g_{1i}, g_{2j}) = ((i-0.5)\triangle_1, (j-0.5)\triangle_2).$$

By letting n_{ij} to denote the number of observations in the sample falling in R_{ij}, we can consider random binomial numbers in help of calculating n_{ij} with success probability being calculated at each R_{ij} as ($f(x,y)$ denotes the new bivariate distribution);

$$P((X,Y) \in R_{ij}) = \triangle_1\triangle_2 f(g_{1i}, g_{2j}). \tag{3.1}$$

By taking as many as (g_{1i}, g_{2j}) for n_{ij} times at each R_{ij} we can simulate the new bivariate distribution.

As an alternative way we can generate two uniform random numbers U and V in the intervals $((i-1)\triangle_1, i\triangle_1)$ and in $((j-1)\triangle_2, j\triangle_2)$ respectively for n_{ij} times at each R_{ij}. In this way we can give more randomness to the simulated data.

3.3 The copula method

In this chapter we demonstrate how to generate random numbers from the new bivariate distribution using the copula formula. We can use copula even when the individual variables are from different distributions. In our case it is gamma and beta of the second kind.

Copula is a multivariate distribution for which the marginal distribution of each variable is uniform. The copula can be used to describe the dependence between the random variables.

It can be difficult to generate random numbers with dependence when the particular distributions are not from standard distributions. Moreover, one can make inputs independent, which is not always sensible where it may lead us to wrong conclusions. Copulas are useful to generate dependent variables. It provides ways to model correlated multivariate data.

Here we construct the bivariate distribution by specifying marginal distributions, and choose a Gaussian copula to provide a correlation structure between variables.

Let's go through the steps performed by copula to generate random numbers. First, we generate random numbers from a bivariate normal distribution. The two variables are statistical dependent. Second, by applying the normal cumulative density function (cdf) to the generated random variables we get a uniform random variable on the interval $[0, 1]$. To show this we use the theorem on the probability integral transformation. Let X have continuous cdf $F_X(x)$ and define the random variable U as $U = F_X(X)$. Then U is uniformly distributed on $[0, 1]$, that is, $P(U \leq u) = u, \quad 0 < u < 1.$

$$
\begin{aligned}
F_U(u) &= P(U \leq u) \\
&= P(F_X(X) \leq u), \\
&= P(F_X^{-1}(F_X(X)) \leq F_X^{-1}(u)), \\
&= P(X \leq F_X^{-1}(u)), \\
&= F_X(F_X^{-1}(u)), \\
&= u,
\end{aligned}
$$

proving that U has a uniform distribution, at the end points we have $P(U \leq u) = 1$ for $u \geq 1$ and $P(U \leq u) = 0$ for $u \leq 0$. This will hold iff the F_X is strictly increasing. Hence $U = F_Z(z)$.

In fact, as we stated earlier in theorem probability integral transformation, one can consider other bivariate distributions and take cdf to obtain random variables which is uniform on the interval $[0, 1]$. For example, another approach in copula is generating pairs of values from t distribution.

Third, by taking the inverse transformation, in other words inverse cdf of the distribution on a uniform random variable in the $[0, 1]$ interval, results in a random variable whose distribution is what we took inverse cdf from. The proof is essentially the reverse of the above proof for the probability integral transformation.

These steps can be applied to each variable of a standard bivariate normal, creating dependent random variables with arbitrary marginal distributions. Because the transformation works on each variable separately, the two resulting variables do not necessary to have the same marginal distributions.

The statistical software does not allow to take the beta of the second kind at once. Hence, we deploy the fact that beta distributions can be transformed to beta of the second kind in order to obtain the marginal of Y. It can be shown that by taking the following transformation one can obtain the marginal of Y. When $B \sim Beta(s, r)$, $0 \leq B \leq 1$ and by

$$Y = \frac{\beta}{\delta} . \left(\frac{B}{1 - B} \right) .$$

3.4 The parameter values

It is important to focus on a range of parameter values to determine whether the derived properties and methods of the new bivariate distribution work in all possible cases. We can think about many parameter combinations and many sample sizes, but it is important to come up with a feasible number of parameter combinations to work with. In the process of determining parameters, we considered many parameter values and observed their shapes and contour plots to come up with the suitable parameter values. The figures B.1, B.2, B.3 and B.4 are plotted using MATLAB program for selected parameter values.

Figures B.1, B.2, B.3 and B.4 represent the shapes of the new bivariate distribution with the selected parameter sets of ($\beta = 3, \delta = 2.5, r = 5, s = 4$), ($\beta = 2.5, \delta = 1.5,\ r = 7, s = 6.5$), ($\beta = 2, \delta = 0.5, r = 9, s = 9$) and ($\beta = 0.7, \delta = 0.07, r = 11, s = 10$) respectively. These parameters could achieve a range of different domains and ranges for the new bivariate distribution. For these sets of parameter values we considered random number generation in Chapter 5. The direct random number generation method was used to simulate the new bivariate distributions. The simulated bivariate distributions were plotted and presented in the Appendix.

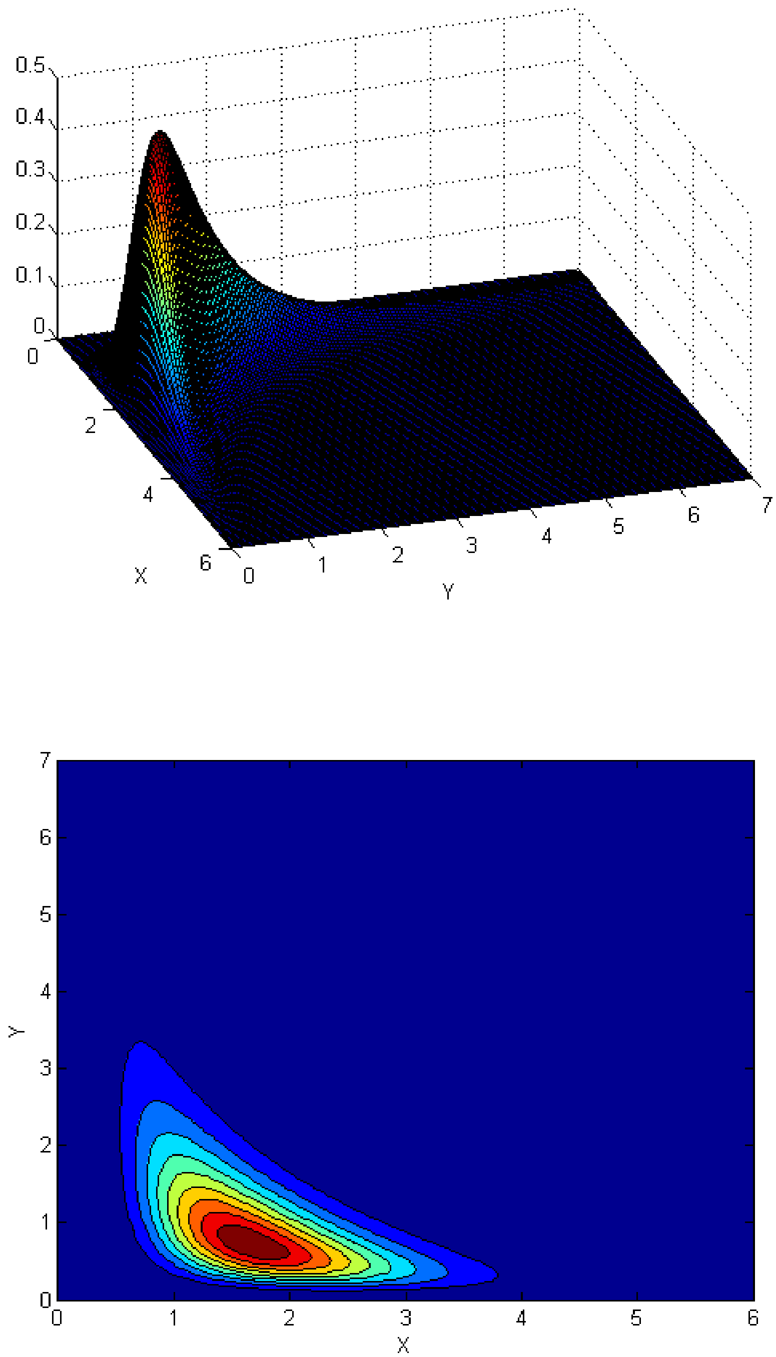

Figure 3.1: 3D plot and Contour of the new bivariate distribution 1

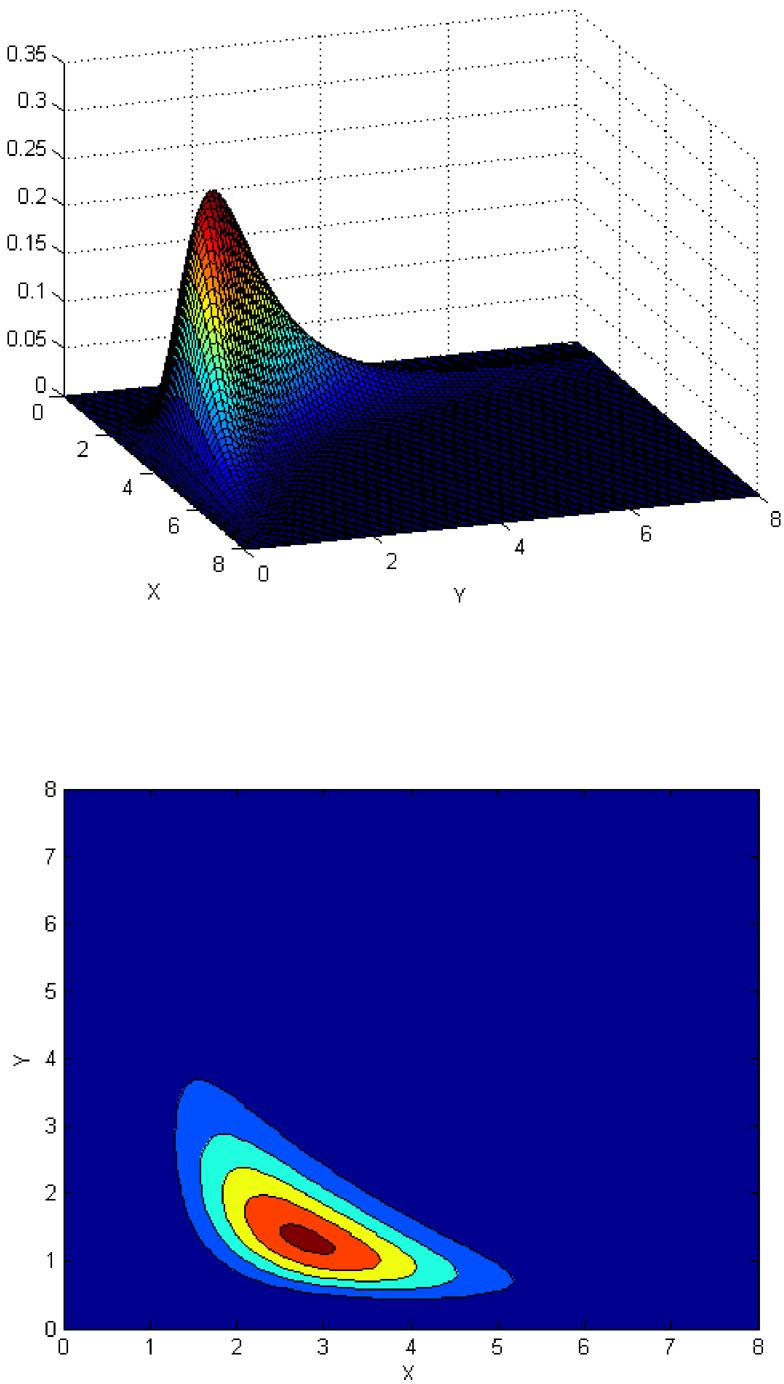

Figure 3.2: 3D plot and Contour of the new bivariate distribution 2

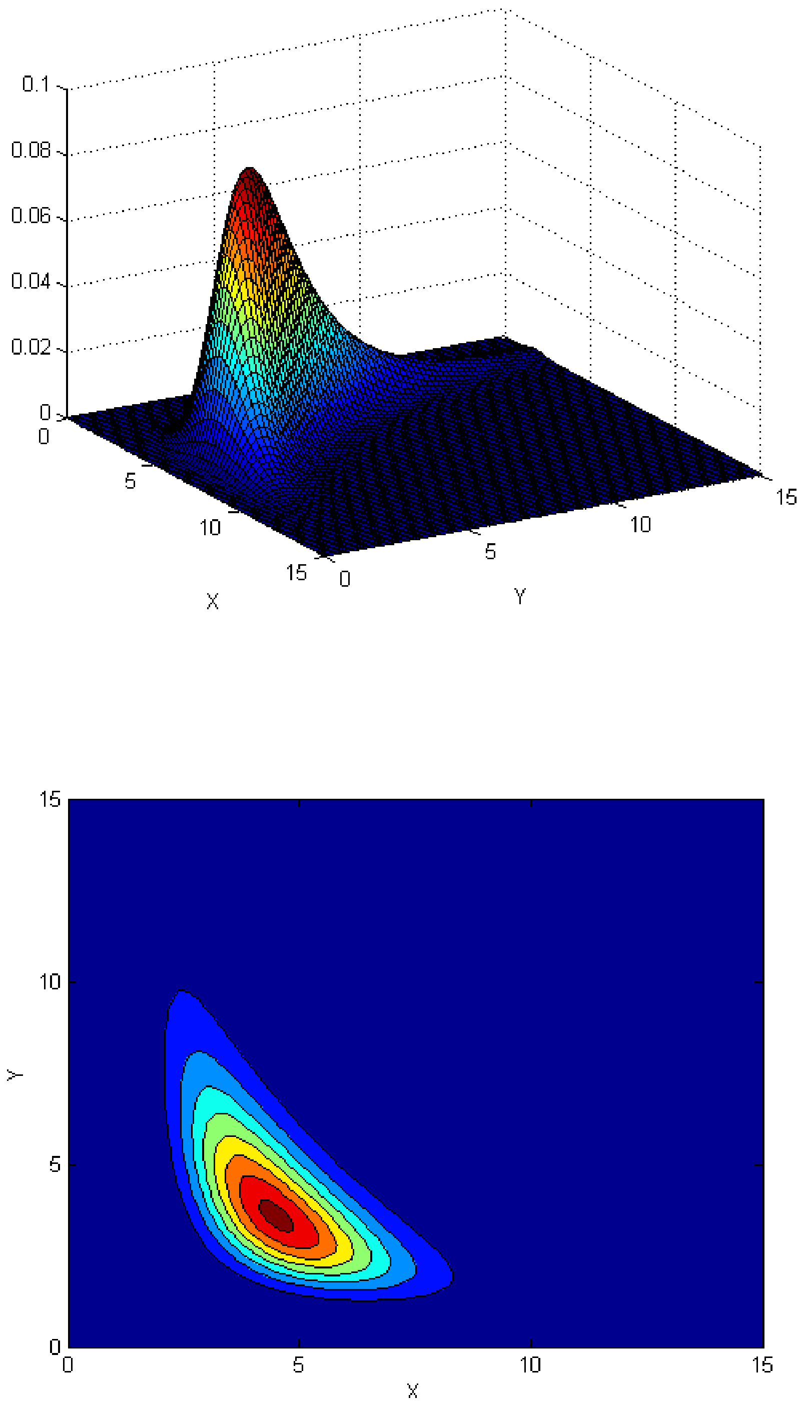

Figure 3.3: 3D plot and Contour of the new bivariate distribution 3

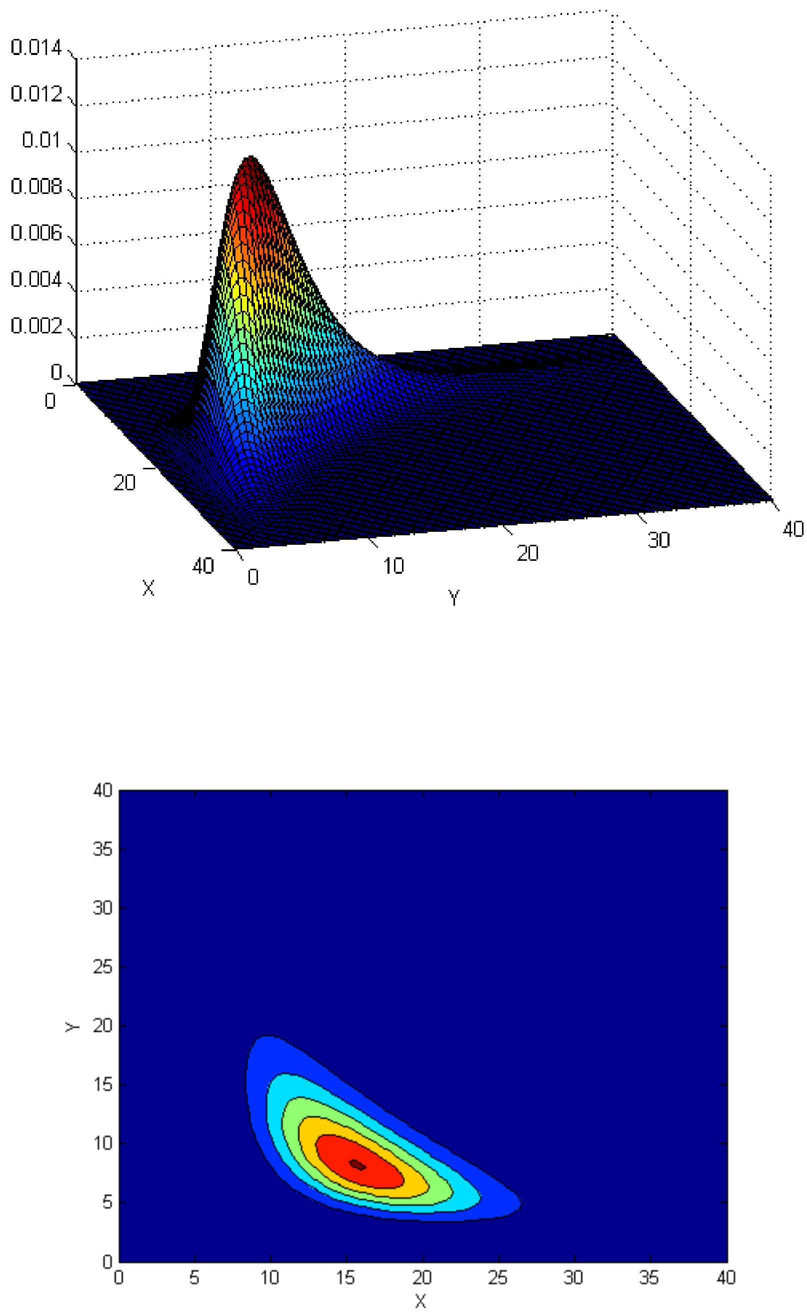

Figure 3.4: 3D plot and Contour of the new bivariate distribution 4

Chapter 4

Finite Mixture of Bivariate Distributions

We discussed the new bivariate distribution and its applications in the previous chapters. Here we focus on the finite mixtures of bivariate distributions. Finite mixture models have been heavily used and discussed in the last two decades and implemented in various fields for more than 100 years with the starting point of mixture models running up to Newcomb (1886) and Pearson (1894).

The parameters of finite mixture models are estimated in the context of maximum likelihood and Bayesian methods respectively.

Finite mixture models is a flexible method of modeling, it has continued to receive increasing attention over the years with the introduction of the EM algorithm by Dempster et al. (1977).

Comprehensive reviews on mixture models can be found in monographs by Everitt and Hand (1981), Titterington et al. (1985), McLachlan and Basford (1988), Lindsay (1995), Böhning (1999), McLachlan and Peel (2000), Frühwirth-Schnatter (2006), and Mengersen et al. (2011),

In finite mixture models the attention has focused on the use of bivariate normal components to model bivariate data. Nevertheless, in many instances the tails of the normal

distribution are often shorter than required (Peel and Mclachlan 2000). Also, finite mixture models of normal distribution may not be satisfactory when observed data is non-normal. In these circumstance our focus is that when observed data has long positive tails to find an alternative mixture model.

In this chapter, we propose mixtures of the new bivariate distribution in an attempt to model the data precisely in the presence of positively skewed subpopulations.

Very little work has done in the context of bivariate gamma mixture distribution and associate shape alike bivariate distributions. In certain situations bivariate gamma mixture models have been outperformed finite mixtures of normal distributions (Jones et al. 2000). The bivariate gamma mixtures will be more effective than normal mixtures in situations the observed data have longer positive tails.

4.1 General form of finite mixture model

We consider a random sample of size n which contains $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, where $\mathbf{X}_i$ $(i = 1, 2, \ldots, n)$ is a d-dimensional random vector with probability density function on $\mathbb{R}^d$. We can use $\mathbf{X}$ to denote the entire random sample. Thus we let $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T, \ldots, \mathbf{X}_n^T)^T$.

Letting the probability density function of $\mathbf{X}_i$ as

$$f\left(\mathbf{X}_i = \mathbf{x}_i\right) = f_{\mathbf{X}_i}\left(\mathbf{x}_i\right) = P\left(\mathbf{x}_i\right) \tag{4.1}$$

we suppose the probability density function of $\mathbf{X}_i$ can be written in the following form.

$$P\left(\mathbf{x}_i|\omega\right) = \sum_{j=1}^{c} \pi_j p_j\left(\mathbf{x}_i|\theta_j\right) \tag{4.2}$$

The vector ω consists of all the unknown parameters of the mixture model. The ω can be viewed as $\omega = (\pi_1, \pi_2, \ldots, \pi_c, \theta_1, \theta_2, \ldots, \theta_c)$.

We refer to $P(\mathbf{x}_i|\omega)$ as a c-component finite mixture density and functions $p_1(\mathbf{x}_i|\theta_1), p_2(\mathbf{x}_i|\theta_2), \ldots, p_c(\mathbf{x}_i|\theta_c)$ as component densities of the mixture.

The θ_j is the vector which consists of the unknown parameters for the j^{th} component density of the mixture.

The π_j's are the mixing proportions or weights. The mixing proportions $\pi_1, \pi_2, \ldots, \pi_c$ are nonnegative values that are constrained sum to one. Hence

$$\sum_{j=1}^{c} \pi_j = 1 \tag{4.3}$$

so that

$$0 \leq \pi_j \leq 1; \; (j = 1, \ldots, c).$$

From the c component finite mixture density $P(\mathbf{x}_i|\omega)$ we can derive the c component finite mixture distribution $F(\mathbf{x}_i)$.

We consider the number of components "c" as fixed in this context. In most cases the value c is unknown and should be inferred using data in hand.

We can simply take the mixing proportions as $pr\,(component\ j)$. To represent the mixing proportions in a formal manner we consider a categorical random variable Y_i which corresponds to $\mathbf{X}_i$. Letting Y_i to take the values $1, \ldots, c$ we can assign $\mathbf{X}_i$ with label components. We can define

$$\pi_j = pr\{Y_i = j\}.$$

The component densities $p_1(\mathbf{x}_i|\theta_1), p_2(\mathbf{x}_i|\theta_2), \ldots, p_c(\mathbf{x}_i|\theta_c)$ of the mixture can be viewed as

$$p_j(\mathbf{x}_i|\theta_j) = f_{\mathbf{x}_i|Y_i}(\mathbf{x}_i|y_i = j).$$

We can also label components by replacing categorical variable Y_i by a c dimensional component label vector $\mathbf{Y}_i$. If y_{ki} is the k^{th} element of $\mathbf{y}_i$ and if $\mathbf{x}_i$ is drawn from the j^{th} component, then we have

$$y_{ki}\begin{cases} 1 & if\ k=j \\ 0 & ow \end{cases} \quad ;\ k=1,\ldots,c.$$

By considering the assigned mixing proportions $\pi_1, \pi_2, \ldots, \pi_c$ as success probabilities for corresponding categories, we can show that $\mathbf{Y}_i$ distributed as a multinomial distribution as follows:

$$f_{\mathbf{Y}_i}(\mathbf{y}_i) = \pi_1^{y_{1i}}\pi_2^{y_{2i}}\ldots\pi_c^{y_{ci}}.$$

Figure 4.1 shows a mixture model of a two component univariate distributions.

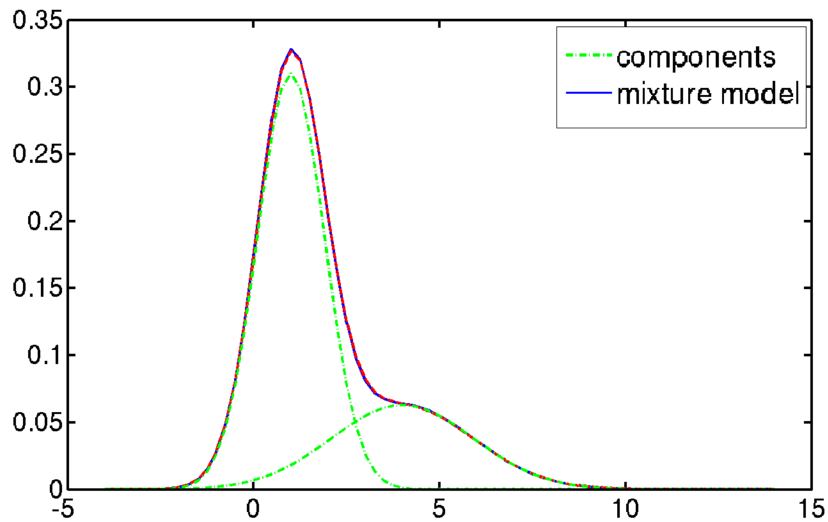

Figure 4.1: Univariate finite mixture model

Figure 4.2 shows a mixture model of a two component univariate distributions.

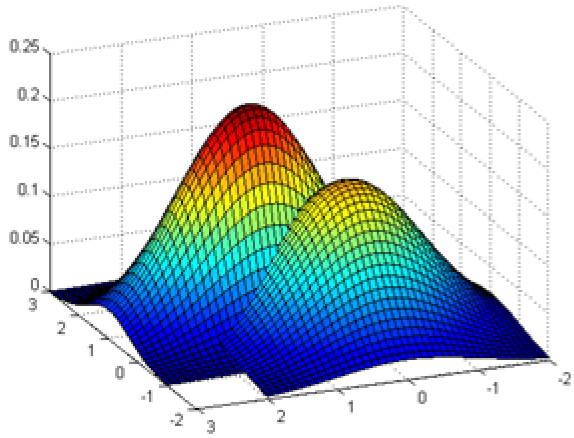

Figure 4.2: Bivariate finite mixture model

A complete explanation can be found in McLachlan and Peel (2004).

4.2 Incomplete data and complete data

We denoted the observed data vector as $\mathbf{x} = (\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_n^T)^T$. As we already discussed earlier each of these observations can be assigned a component label. In the finite mixture framework it is very important to infer the component the observations are drawn from.

Therefore in the context of mixture models $\mathbf{X}, \mathbf{Y}$ are considered as complete data, while the observed data are considered as incomplete data.

For a graphical overview the observed data can be viewed as in Figure 4.3.

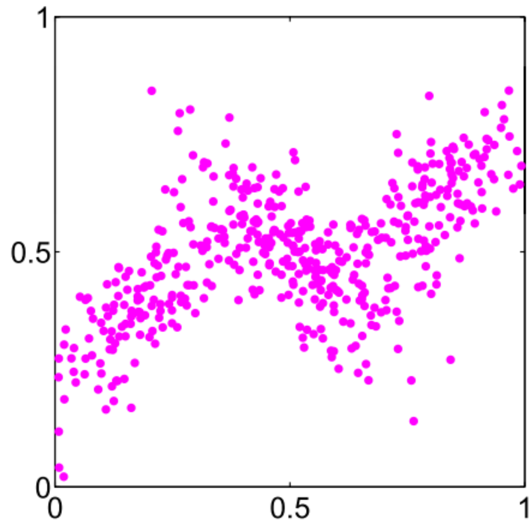

Figure 4.3: Incomplete Data

After we assign labels for each observations the complete data can be illustrated as in Figure 4.4. Here we used different colors to distinguish components.

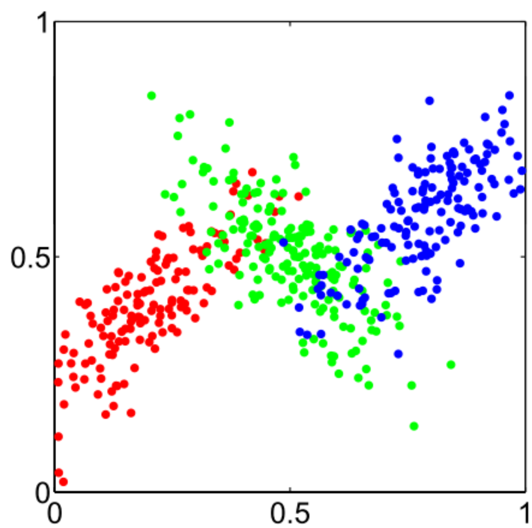

Figure 4.4: Complete Data

4.3 Maximum likelihood estimation for mixture models

Upon the introduction of the EM algorithm the maximum likelihood (ML) estimation method has been widely adopted in finding the estimators of mixture models.

We consider a density function $p(\mathbf{x}|\omega)$ with a set of parameters ω and a random sample $\mathbf{x} = (\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_n^T)^T$ from $p(\mathbf{x}|\omega)$.

The likelihood function is

$$L(\omega|\mathbf{x}) = \prod_{i=1}^{n} p(\mathbf{x}_i|\omega).$$

To find the maximum likelihood estimates of ω, which is ω^* such that:

$$\omega^* = \max_{\omega} L(\omega/X)$$

We usually maximize $\log L(\omega|X)$ instead of maximizing $L(\omega|X)$. This will make the computation of ω^* analytically simpler.

Then the log likelihood function for mixture model in Equation (4.2) is

$$\begin{aligned}
\log L(\Theta|X) &= \log \prod_{i=1}^{N} P(x_i|\Theta) \\
&= \sum_{i=1}^{n} \log \left(\sum_{j=1}^{c} \pi_j p_j (x_i|\theta_j) \right).
\end{aligned}$$

Since the log likelihood is the log of a sum it is difficult to calculate the maximum likelihood estimators. One can introduce the missing data for the above incomplete log likelihood and

make it simpler by taking the complete data likelihood in Equation (4.4).

$$\log(L(\omega|X,Y)) = \log p(X,Y|\omega) \tag{4.4}$$

Thus, the missing information which is the component number can be introduced by assuming the existence of a random variable Y such that $Y = \{y_1, \dots, y_N\}$. Here $y_i \in 1, \dots, M$ for each i, where $y_i = l$ if the i^{th}sample was drawn by the l^{th}mixture component.

Upon the introduction of missing data we have the complete data log likelihood for mixture models as

$$\begin{aligned}\log(L(\omega|X,Y)) &= \log p(X,Y|\omega)\\ &= \sum_{i=1}^{N} \log(p(x_i|y_i)p(y_i))\end{aligned}$$

$$\log(L(\omega|X,Y)) = \sum_{i=1}^{N} \log(\pi_{y_i} p_{y_i}(x_i|\theta_{y_i})). \tag{4.5}$$

Next we illustrates the use of Expectation Maximization (EM) algorithm to proceed with finding the maximum likelihood estimators of the above mixture model.

4.4 EM algorithm for mixture models

In situations where the data is incomplete or has missing values the EM algorithm can be used to find the maximum likelihood estimators of the underlying model parameters.

As we noted in the previous section in the context of mixture modeling the observed data are viewed as incomplete. We first discuss the general EM framework and implement it in the context of mixture modeling.

The EM algorithm proceeds in two steps, namely expectation step and maximization step. EM algorithm is guaranteed to converge to the respective maximum likelihood estimators (Dempster et al. (1977), Wu (1983)). This replaces one difficult likelihood maximization with a sequence of easier maximization whose limit is the answer to the original problem.

E step takes the conditional expectation value of the complete data log likelihood $\log p(X, Y|\omega)$ with respective to the unknown data Y given the observed data X and the current parameter estimates ω^k, which can be written as

$$Q(\omega, \omega^k) = E\left\{\log p\left(X, Y|\omega\right)|X, \omega^k\right\}. \tag{4.6}$$

M step requires to maximize the expectation we computed in the E step. That is, we find

$$\omega^{k+1} = \max_{\omega} Q(\omega, \omega^k).$$

The $Q(\omega, \omega^k)$ should be simple to make our calculations easy, in order to do to make it simple we follow the following. In the Equation (4.6) X and ω^k are constants and ω are the new parameters that we optimize to maximize $Q(\omega, \omega^k)$. Since Y is a random variable, we can take its distribution $f\left(y|X, \omega^k\right)$ and rewrite the Equation (4.6) as

$$E\left\{\log p\left(X, Y|\omega\right)|X, \omega^k\right\} = \int_{\mathbf{y}\in\Upsilon} \log p\left(X, \mathbf{y}|\omega\right) f\left(\mathbf{y}|X, \omega^k\right) d\mathbf{y}. \tag{4.7}$$

Here Υ denote space of values y can take. We can specify the π_j as prior probabilities of each mixture component, so that $\pi_j = p(component\, j)$. From Bayes rule we have

$$\begin{aligned} p(y_i|x_i, \omega^k) &= \frac{\pi^k_{y_i} p_{y_i}(x_i|\theta^k_{y_i})}{p(x_i|\Theta^g)} \\ &= \frac{\pi^k_{y_i} p_{y_i}(x_i|\theta^k_{y_i})}{\sum_{j=1}^{M} \pi^k_j p_j(x_i|\theta^k_j)} \end{aligned}$$

and

$$p(y|X, \omega^k) = \prod_{h=1}^{n} p\left(y_h|x_h, \omega^k\right).$$

Now we can consider the Equation 4.6 such that

$$
\begin{aligned}
Q(\omega,\omega^k) &= \sum_{y\in\Upsilon} \log\left(L\left(\omega|X,\mathbf{y}\right)\right) p\left(\mathbf{y}|X,\omega^k\right) \\
&= \sum_{y\in\Upsilon}\sum_{i=1}^{n} \log\left(\pi_{y_i} p_{y_i}\left(x_i|\theta_{y_i}\right)\right) \prod_{h=1}^{n} p\left(y_h|x_h,\omega^k\right) \\
&= \sum_{y_1=1}^{c}\sum_{y_2=1}^{c}\cdots\sum_{y_n=1}^{c}\sum_{i=1}^{n} \log\left(\pi_{y_i} p_{y_i}\left(x_i|\theta_{y_i}\right)\right) \prod_{h=1}^{n} p\left(y_h|x_h,\omega^k\right) \\
&= \sum_{y_1=1}^{c}\sum_{y_2=1}^{c}\cdots\sum_{y_n=1}^{c}\sum_{i=1}^{n}\sum_{j=1}^{c} \delta_{j,y_i} \log\left(\pi_j p_j\left(x_i|\theta_j\right)\right) \prod_{h=1}^{n} p\left(y_h|x_h,\omega^k\right) \\
&= \sum_{j=1}^{c}\sum_{i=1}^{n} \log\left(\pi_j p_j\left(x_i|\theta_j\right)\right) \sum_{y_1=1}^{c}\sum_{y_2=1}^{c}\cdots\sum_{y_n=1}^{c} \delta_{j,y_i} \prod_{h=1}^{n} p\left(y_h|x_h,\omega^k\right) \\
&= \sum_{j=1}^{c}\sum_{i=1}^{n} \log\left(\pi_j p_j\left(x_i|\theta_j\right)\right) \left(\sum_{y_1=1}^{c}\cdots\sum_{y_{i-1}=1}^{c}\sum_{y_{i+1}=1}^{c}\cdots\sum_{y_n=1}^{c}\prod_{h=1,h\neq i}^{n} p\left(y_h|x_h,\omega^k\right)\right) p\left(j|x_i,\omega^k\right) \\
&= \sum_{j=1}^{c}\sum_{i=1}^{n} \log\left(\pi_j p_j\left(x_i|\theta_j\right)\right) \prod_{h=1,h\neq i}^{n} \left(\sum_{y_h=1}^{c} p\left(y_h|x_h,\omega^k\right)\right) p\left(j|x_i,\omega^k\right) \\
&= \sum_{j=1}^{c}\sum_{i=1}^{n} \log\left(\pi_j p_j\left(x_i|\theta_j\right)\right) p\left(j|x_i,\omega^k\right)
\end{aligned}
$$

and simplify to the form of

$$
Q(\omega,\omega^k) = \sum_{j=1}^{c}\sum_{i=1}^{n} \log(\pi_j) p(j|x_i,\omega^k) + \sum_{j=1}^{c}\sum_{i=1}^{n} \log(p_j(x_i|\theta_j)) p(j|x_i,\omega^k). \tag{4.8}
$$

In the Equation (4.8) E step contains two independent terms with π_j and θ_j respectively. Therefore we can maximize π_j and θ_j separately. For more explanation see Bilmes (1998).

4.5 Finite mixtures of bivariate distribution and parameter estimation

First let's start with finding an expression for π_j. Since

$$\sum_{j=1}^{c} \pi_j = 1$$

and if y_i values are known, then

$$\pi_j = \frac{\#\{y_i = j\}}{n}.$$

By the conditional expectation of the y_i

$$\begin{aligned} E\left(y_i | x_i, \omega^k\right) &= pr\left(y_i = 1 | x_i, \omega^k\right) \\ &= \frac{\pi_{y_i}^k p_{y_i}(x_i | \theta_{y_i}^k)}{\sum_{j=1}^{c} \pi_j^k p_j(x_i | \theta_j^k)}. \end{aligned}$$

Thus,

$$\pi_j^{k+1} = \frac{\sum_{i=1}^{n} \pi_{y_i}^k p_{y_i}(x_i | \theta_{y_i}^k)}{\sum_{j=1}^{c} \pi_j^k p_j(x_i | \theta_j^k)}. \tag{4.9}$$

Assume two components mixture model of the bivariate distribution. When

$$p\left((x_1, x_2) | \omega\right) = \sum_{i=1}^{2} \pi_i p_i\left((x_1, x_2) | \theta_i\right)$$

we have

$$p_j\left((x_1, x_2) | r_j, s_j, \beta_j, \delta_j\right) = \frac{\beta_j^{r_j}}{\Gamma(r_j)} \frac{\delta_j^{s_j}}{\Gamma(s_j)} x_1^{s_j + r_j - 1} x_2^{s_j - 1} e^{-(\beta_j x_1 + \delta_j x_1 x_2)}.$$

By taking log

$$\log\left(p_j\left(\left(x_1,x_2\right)|r_j,s_j,\beta_j,\delta_j\right)\right)=r_j\log\left(\beta_j\right)-\log\Gamma\left(r_j\right)+s_j\log\left(\delta_j\right)-\log\Gamma\left(s_j\right)$$
$$+(r_j+s_j-1)\log\left(x_1\right)+\left(s_j-1\right)\log\left(x_2\right)-\left(\beta_jx_1+\delta_jx_1x_2\right). \tag{4.10}$$

By the substitution using (4.10) in to (4.8), we can obtain

$$Q(\omega,\omega^k)=\sum_{j=1}^{2}\sum_{i=1}^{n}\log(\pi_j)p(j|x_i,\omega^k)+\sum_{j=1}^{2}\sum_{i=1}^{n}\log(p_j(x_i|\theta_j))p(j|x_i,\omega^k)$$

$$=\sum_{j=1}^{2}\sum_{i=1}^{n}\log(\pi_j)p(j|\left(x_{1i},x_{2i}\right),\omega^k)+\sum_{j=1}^{2}\sum_{i=1}^{n}\left[r_j\log\left(\beta_j\right)-\log\Gamma\left(r_j\right)+s_j\log\left(\delta_j\right)\right]p(j|\left(x_{1i},x_{2i}\right),\omega^k)$$

$$+\sum_{j=1}^{2}\sum_{i=1}^{n}\left[-\log\Gamma\left(s_j\right)+\left(r_j+s_j-1\right)\log\left(x_{1i}\right)+\left(s_j-1\right)\log\left(x_{2i}\right)\right]p(j|\left(x_{1i},x_{2i}\right),\omega^k)$$

$$-\sum_{j=1}^{2}\sum_{i=1}^{n}\left[\beta_jx_{1i}+\delta_jx_{1i}x_{2i}\right]p(j|\left(x_{1i},x_{2i}\right),\omega^k). \tag{4.11}$$

To find an expression for r_j we can take $\frac{\partial}{\partial r_j}Q(\omega,\omega^k)=0$

$$\frac{\partial}{\partial r_j}Q(\omega,\omega^k)=\sum_{i=1}^{n}\left[\log\left(\beta_j\right)-\psi\left(r_j\right)+\log\left(x_{1i}\right)\right]p\left(j|\left(x_{1i},x_{2i}\right),\omega^k\right)\underset{=}{set}\,0,$$

$$\sum_{i=1}^{n}\left[\log\left(\beta_j\right)+\log\left(x_{1i}\right)\right]p\left(j|\left(x_{1i},x_{2i}\right),\omega^k\right)=\sum_{i=1}^{n}\psi\left(r_j\right)p\left(j|\left(x_{1i},x_{2i}\right),\omega^k\right),$$

$$\psi\left(r_j^{k+1}\right)=\frac{\sum_{i=1}^{n}\left(\log\left(\beta_j^k\right)+\log\left(x_{1i}\right)\right)p\left(j|\left(x_{1i},x_{2i}\right),\omega^k\right)}{\sum_{i=1}^{n}p\left(j|\left(x_{1i},x_{2i}\right),\omega^k\right)},$$

$$r_j^{k+1}=\psi^{-1}\left(\frac{\sum_{i=1}^{n}\left(\log\left(\beta_j^k\right)+\log\left(x_{1i}\right)\right)p\left(j|\left(x_{1i},x_{2i}\right),\omega^k\right)}{\sum_{i=1}^{n}p\left(j|\left(x_{1i},x_{2i}\right),\omega^k\right)}\right),$$

for s_j, $\frac{\partial}{\partial s_j} Q(\omega, \omega^k) = 0$

$$\frac{\partial}{\partial s_j} Q(\omega, \omega^k) = \sum_{i=1}^{n} \left[\log\left(\delta_j\right) - \psi\left(s_j\right) + \log\left(x_{1i}\right) + \log\left(x_{2i}\right)\right] p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right) \underset{=}{set}\ 0,$$

$$\psi\left(s_j^{k+1}\right) = \frac{\sum_{i=1}^{n} \left[\log\left(\delta_j^k\right) + \log\left(x_{1i}\right) + \log\left(x_{2i}\right)\right] p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}{\sum_{i=1}^{n} p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)},$$

$$s_j^{k+1} = \psi^{-1}\left(\frac{\sum_{i=1}^{n} \left[\log\left(\delta_j^k\right) + \log\left(x_{1i}\right) + \log\left(x_{2i}\right)\right] p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}{\sum_{i=1}^{n} p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}\right),$$

for β_j, $\frac{\partial}{\partial \beta_j} Q(\omega, \omega^k) = 0$

$$\frac{\partial}{\partial \beta_j} Q(\omega, \omega^k) = \sum_{i=1}^{n} \left[\frac{r_j}{\beta_j} - x_{1i}\right] p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right) \underset{=}{set}\ 0,$$

$$\beta_j^{k+1} = \frac{\sum_{i=1}^{n} r_j^k p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}{\sum_{i=1}^{n} x_{1i} p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)},$$

and for δ_j, $\frac{\partial}{\partial \delta_j} Q(\omega, \omega^k) = 0$

$$\frac{\partial}{\partial \delta_j} Q(\omega, \omega^k) = \sum_{i=1}^{n} \left[\frac{s_j}{\delta_j} - x_{1i}x_{2i}\right] p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right) \underset{=}{set}\ 0,$$

$$\delta_j^{k+1} = \frac{\sum_{i=1}^{n} s_j^k p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}{\sum_{i=1}^{n} x_{1i}x_{2i} p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}.$$

In summary now we have the estimates of the new parameters in terms of the old parameters as

$$\pi_j^{k+1} = \frac{\sum_{i=1}^{n} \pi_{y_i}^k p_{y_i}(x_i|\theta_{y_i}^k)}{\sum_{j=1}^{c} \pi_j^k p_j(x_i|\theta_j^k)},$$

$$\beta_j^{k+1} = \frac{\sum_{i=1}^{n} r_j^k p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}{\sum_{i=1}^{n} x_{1i} p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)},$$

$$
\begin{aligned}
\delta_j^{k+1} &= \frac{\sum_{i=1}^{n} s_j^k p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}{\sum_{i=1}^{n} x_{1i} x_{2i} p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}, \\
r_j^{k+1} &= \psi^{-1}\left(\frac{\sum_{i=1}^{n}\left(\log\left(\beta_j^k\right) + \log\left(x_{1i}\right)\right) p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}{\sum_{i=1}^{n} p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}\right), \\
s_j^{k+1} &= \psi^{-1}\left(\frac{\sum_{i=1}^{n}\left[\log\left(\delta_j^k\right) + \log\left(x_{1i}\right) + \log\left(x_{2i}\right)\right] p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}{\sum_{i=1}^{n} p\left(j|\left(x_{1i}, x_{2i}\right), \omega^k\right)}\right).
\end{aligned}
$$

The EM algorithm can be continued by using the newly derived parameters as the old parameters in the proceeding iterations accordingly. Here, the above equations perform both expectation step and maximization step simultaneously.

Chapter 5

Analysis of Data

In this chapter we utilized the derived properties and methods of the new bivariate model on the simulation data. We obtained results for a wide range of parameter values and for different sample sizes to verify whether the methods hold for any of the combinations of parameter values and sample sizes.

The maximum likelihood estimation and method of moments estimation are used to estimate the parameters for the simulated data. Bias of the estimators, mean squared error and relative efficiency measures are used for the analysis of estimation methods. The figures of the new bivariate distributions for different sets of parameters with respective simulated illustrations are presented in the Appendix B.

5.1 Bias

The bias of an estimator is the difference between the expected value of an estimator and the true value of the parameter being estimated.

5.2 Mean Squared Error

The mean squared error of an estimator is the average of the squares of the difference between the estimator and what is estimated. The mean squared error is an important measure used to evaluate the performance of an estimator. Mean squared error is a risk function; it is the expected value of the squared error loss or quadratic loss.

5.3 Efficiency

The efficiency of an unbiased estimator is the ratio between the minimum possible variance for an unbiased estimator (the inverse of the fisher information) and its actual variance. The efficiency is a value which is less than or equal to one. If the efficiency of an unbiased estimator attains one for all values of the parameter, then the estimator called efficient or minimum variance unbiased estimator.

5.4 Asymptotic Efficiency

Some estimators attain efficiency asymptotically; these estimators are called asymptotically efficient estimators. Sometimes maximum likelihood estimators are asymptotically efficient estimators.

5.5 Relative Efficiency

The relative efficiency of two estimators is the ratio between their mean squared error values in the process of estimating the same parameter value. The relative efficiency can take values greater than zero. If the relative efficiency is inbetween zero and one then the method of estimation we considered on top of the ratio is more efficient than the method of estimation considered in the bottom of the ratio, if the relative efficiency is greater than one it is the other way around.

The parameter combinations we considered are $(\beta = 3, \delta = 2.5, r = 5, s = 4)$, $(\beta = 2.5, \delta = 1.5, r = 7, s = 6.5)$, $(\beta = 2, \delta = 0.5, r = 9, s = 9)$ and $(\beta = 0.7, \delta = 0.07, r = 11, s = 10)$. For each set of these parameter values 50, 100, 200 and 1000 sizes of random numbers were generated using the direct random number generation method and the mean over the estimators, biases, mean squared error values, relative efficiency values and correlations were computed and used for analysis. We generated S independent data sets in other words

it is the number of random number generation runs to use Monte Carlo simulation method. Here the number of independent data sets S is 2000.

Tables 5.1, 5.2, 5.3 and 5.4 contains the estimated values of the parameters, biases of the estimators and mean square error (MSE) values. These tables show the biases for maximum likelihood estimators are relatively low compared to method of moment estimators. Irrespective to the method of estimation the biases of the estimators are getting small as the size of the sample is increasing. It is noticeable, while maximum likelihood estimators have very small biases to deal with, method of moment estimators illustrate large biases; this getting worse for small sample sizes where method of moment estimators produce unacceptably large biases. When the sample size is 200 and more, maximum likelihood estimators almost all the time attain the respective parameter value for one decimal point; note that almost all parameter values in this context is only up to one decimal point.

If we consider the mean squared error values in the tables 5.1, 5.2, 5.3 and 5.4; it is an evident that method of moments estimators produce really large mean squared error values compared those to the maximum likelihood estimators. Thus the fact that performances of maximum likelihood estimators are better than the method of moment estimators is being articulated by mean squared error values. Moreover, as one could expect the mean squared error values are getting smaller as the size of the sample increases; these mean squared values are still not getting small enough for method of moment estimators to compete with maximum likelihood estimators.

Table 5.1: Estimation of Parameters 1

Sample size (n)	Parameters	Maximum likelihood estimator			Method of moment estimator		
		Values	Bias	MSE	Values	Bias	MSE
50	$\beta = 3$	3.1713	0.1713	0.4725	3.1442	0.1442	0.5448
	$\delta = 2.5$	2.6592	0.1592	0.3359	9.4152	6.9152	325145.9
	$r = 5$	5.2686	0.2686	1.2185	5.2693	0.2693	1.4146
	$s = 4$	4.2284	0.2284	0.7386	14.5271	10.5271	750082
100	$\beta = 3$	3.0956	0.0956	0.2220	3.0841	0.0841	0.2595
	$\delta = 2.5$	2.5680	0.0680	0.1507	4.0142	1.5142	241.9439
	$r = 5$	5.1467	0.1467	0.5616	5.1280	0.1280	0.6687
	$s = 4$	4.1005	0.1005	0.3399	6.2875	2.2875	557.3893
200	$\beta = 3$	3.0341	0.0341	0.1023	3.0271	0.0271	0.1195
	$\delta = 2.5$	2.5437	0.0437	0.0712	3.7926	1.2926	243.8261
	$r = 5$	5.0567	0.0567	0.2653	5.0453	0.0453	0.3153
	$s = 4$	4.0614	0.0614	0.1590	5.9547	1.9547	538.1533
1000	$\beta = 3$	3.0103	0.0103	0.0182	3.0084	0.0084	0.0230
	$\delta = 2.5$	2.5016	0.0016	0.0128	2.7639	0.2639	0.7412
	$r = 5$	5.0164	0.0164	0.0460	5.0131	0.0131	0.0598
	$s = 4$	4.0018	0.0018	0.0291	4.4074	0.4074	1.7422

Table 5.2: Estimation of Parameters 2

Sample size (n)	Parameters	Maximum likelihood estimator			Method of moment estimator		
		Values	Bias	MSE	Values	Bias	MSE
50	$\beta = 2.5$	2.6471	0.1471	0.3306	2.6188	0.1188	0.3496
	$\delta = 1.5$	1.5936	0.0936	0.1188	-0.7124	-2.2124	27093.7600
	$r = 7$	7.3802	0.3802	2.3828	7.3018	0.3018	2.5436
	$s = 6.5$	6.8817	0.3817	2.0384	-2.5257	-9.0257	415542.1
100	$\beta = 2.5$	2.5707	0.0707	0.1375	2.5604	0.0604	0.1510
	$\delta = 1.5$	1.5393	0.0393	0.0510	3.1030	1.6030	1606.8440
	$r = 7$	7.1836	0.1836	1.0123	7.1550	0.1550	1.1227
	$s = 6.5$	6.6587	0.1587	0.8741	13.0849	6.5849	26142.6800
200	$\beta = 2.5$	2.5294	0.0294	0.0706	2.5236	0.0236	0.0809
	$\delta = 1.5$	1.5214	0.0214	0.0251	2.3895	0.8895	413.8296
	$r = 7$	7.0744	0.0744	0.5198	7.0581	0.0581	0.6012
	$s = 6.5$	6.5910	0.0910	0.4398	10.1724	3.6724	6992.9710
1000	$\beta = 2.5$	2.5086	0.0086	0.0124	2.5066	0.0066	0.0150
	$\delta = 1.5$	1.5011	0.0011	0.0049	1.5685	0.0685	0.1117
	$r = 7$	7.0236	0.0236	0.0914	7.0181	0.0181	0.1117
	$s = 6.5$	6.5050	0.0050	0.0843	6.7868	0.2868	1.9567

Table 5.3: Estimation of Parameters 3

Sample size (n)	Parameters	Maximum likelihood estimator			Method of moment estimator		
		Values	Bias	MSE	Values	Bias	MSE
50	$\beta = 2$	2.1059	0.1059	0.2073	2.0799	0.0799	0.2135
	$\delta = 0.5$	0.5342	0.0342	0.0139	0.4379	-0.0621	104.815
	$r = 9$	9.4568	0.4568	3.8747	9.3403	0.3403	4.008
	$s = 9$	9.5776	0.5776	4.229	7.8852	-1.1148	31136.43
100	$\beta = 2$	2.0572	0.0572	0.0883	2.0458	0.0458	0.094
	$\delta = 0.5$	0.5139	0.0139	0.0057	0.6807	0.1807	5.8046
	$r = 9$	9.2474	0.2474	1.7001	9.1964	0.1964	1.8247
	$s = 9$	9.2317	0.2317	1.7521	12.1481	3.1481	1721.077
200	$\beta = 2$	2.0299	0.0299	0.0441	2.0226	0.0226	0.0488
	$\delta = 0.5$	0.5055	0.0055	0.0027	0.5920	0.0920	0.2222
	$r = 9$	9.1276	0.1276	0.8467	9.0949	0.0949	0.9459
	$s = 9$	9.0956	0.0956	0.8543	10.6058	1.6058	72.1360
1000	$\beta = 2$	2.0075	0.0075	0.0082	2.0054	0.0054	0.0093
	$\delta = 0.5$	0.5007	0.0007	0.0005	0.5159	0.0159	0.0074
	$r = 9$	9.0323	0.0323	0.1562	9.0228	0.0228	0.1790
	$s = 9$	9.0143	0.0143	0.1549	9.2801	0.2801	2.2575

Table 5.4: Estimation of Parameters 4

Sample size (n)	Parameters	Maximum likelihood estimator			Method of moment estimator		
		Values	Bias	MSE	Values	Bias	MSE
50	$\beta = 0.7$	0.7421	0.0421	0.0262	0.7320	0.0320	0.0263
	$\delta = 0.07$	0.0745	0.0045	0.0003	0.1223	0.0523	1.5497
	$r = 11$	11.6435	0.6435	6.0701	11.4845	0.4845	6.1119
	$s = 10$	10.6044	0.6044	5.2583	17.4138	7.4138	31660.2100
100	$\beta = 0.7$	0.7207	0.0207	0.0116	0.7166	0.0166	0.0123
	$\delta = 0.07$	0.0720	0.0020	0.0001	0.0866	0.0166	0.0097
	$r = 11$	11.3158	0.3158	2.7343	11.2510	0.2510	2.9143
	$s = 10$	10.2600	0.2600	2.2114	12.2757	2.2757	181.1949
200	$\beta = 0.7$	0.7101	0.0101	0.0053	0.7079	0.0079	0.0057
	$\delta = 0.07$	0.0708	0.0008	0.0001	0.0779	0.0079	0.0008
	$r = 11$	11.1478	0.1478	1.2604	11.1124	0.1124	1.3666
	$s = 10$	10.1163	0.1163	1.0564	11.0892	1.0892	15.6049
1000	$\beta = 0.7$	0.7013	0.0013	0.0010	0.7006	0.0006	0.0011
	$\delta = 0.07$	0.0702	0.0002	0.0000	0.0715	0.0015	0.0001
	$r = 11$	11.0187	0.0187	0.2332	11.0077	0.0077	0.2641
	$s = 10$	10.0321	0.0321	0.1951	10.2087	0.2087	2.0184

We were discussing as the sample size is 200 or more, maximum likelihood estimators almost all the time attain the respective parameter value for an accuracy of one decimal point; the respective parameter values also is up to one decimal point. This can be identified as asymptotic unbiasedness of the maximum likelihood estimators. The efficiencies of the maximum likelihood estimators are as in the tables 5.5, 5.6, 5.7 and 5.8. We can observe the efficiencies of the maximum likelihood estimators are very close to one. Therefore the maximum likelihood estimators are asymptotically efficient in this context.

Table 5.5: Efficiency of maximum likelihood estimators 1

Efficiency of maximum likelihood estimators				
Parameters	Sample size (n)			
	50	100	200	1000
$\beta = 3$	0.9399	0.9331	0.9439	1.0385
$\delta = 2.5$	0.9521	0.9455	0.9776	1.0284
$r = 5$	0.9109	0.9213	0.9154	1.0313
$s = 4$	0.9661	0.9433	0.9826	1.0179

Table 5.6: Efficiency of maximum likelihood estimators 2

Efficiency of maximum likelihood estimators				
Parameters	Sample size (n)			
	50	100	200	1000
$\beta = 2.5$	0.9291	1.0219	0.9411	1.0453
$\delta = 1.5$	0.9475	0.9841	0.9641	0.9459
$r = 7$	0.9310	1.0076	0.9290	1.0369
$s = 6.5$	0.9543	0.9945	0.9581	0.9547

Table 5.7: Efficiency of maximum likelihood estimators 3

Efficiency of maximum likelihood estimators				
Parameters	Sample size (n)			
	50	100	200	1000
$\beta = 2$	0.9216	1.0145	0.9729	1.0124
$\delta = 0.5$	0.9151	0.9685	0.9741	1.0163
$r = 9$	0.9422	1.0068	0.9676	1.0140
$s = 9$	0.9099	0.9682	0.9439	1.0129

Table 5.8: Efficiency of maximum likelihood estimators 4

Efficiency of maximum likelihood estimators				
Parameters	Sample size (n)			
	50	100	200	1000
$\beta = 0.7$	0.9151	0.9411	0.9893	1.0255
$\delta = 0.07$	0.8897	0.9604	0.9562	0.9998
$r = 11$	0.9317	0.9438	0.9738	1.0119
$s = 10$	0.8909	0.9508	0.9496	1.0036

The tables 5.9, 5.10, 5.11 and 5.12 shows the relative efficiency values between maximum likelihood estimators and method of moment estimators. Since the relative efficiency values are very small and sometime very close to zero it is an evident to claim maximum likelihood estimators are optimal than the method of moment estimators. The relative efficiency values are also a good indication that maximum likelihood estimation is better method to use to estimate the parameters of the new bivariate distribution.

Table 5.9: Relative efficiency of estimators 1

Relative efficiency of estimators				
Parameters	Sample size (n)			
	50	100	200	1000
$\beta = 3$	0.8673	0.8554	0.8562	0.7918
$\delta = 2.5$	0.0000	0.0006	0.0003	0.0172
$r = 5$	0.8614	0.8398	0.8412	0.7694
$s = 4$	0.0000	0.0006	0.0003	0.0167

Table 5.10: Relative efficiency of estimators 2

Relative efficiency of estimators				
Parameters	Sample size (n)			
	50	100	200	1000
$\beta = 2.5$	0.9456	0.9107	0.8719	0.8302
$\delta = 1.5$	4.38359E-06	3.17389E-05	6.07086E-05	0.0438
$r = 7$	0.9368	0.9017	0.8646	0.8180
$s = 6.5$	4.90552E-06	3.34349E-05	6.28855E-05	0.0431

Table 5.11: Relative efficiency of estimators 3

Relative efficiency of estimators				
Parameters	Sample size (n)			
	50	100	200	1000
$\beta = 2$	0.9707	0.9388	0.9020	0.1673
$\delta = 0.5$	0.0001	0.0009	0.0122	0.0023
$r = 9$	0.9667	0.9317	0.8951	0.1651
$s = 9$	0.0001	0.0010	0.0118	0.0021

Table 5.12: Relative efficiency of estimators 4

Relative efficiency of estimators				
Parameters	Sample size (n)			
	50	100	200	1000
$\beta = 0.7$	0.9955	0.9443	0.9244	0.8867
$\delta = 0.07$	0.0002	0.0116	0.0661	0.0960
$r = 11$	0.9932	0.9383	0.9223	0.8828
$s = 10$	0.0002	0.0122	0.0677	0.0966

Furthermore, lets look at the estimated sample correlations from both maximum likelihood and method of moments estimation methods. The tables 5.13, 5.14, 5.15 and 5.16 show the respective correlation values. We know from earlier observations that the maximum likelihood estimators are performing better than method of moments estimators. Therefore, as one could expect the correlations which are estimated from maximum likelihood are very close to the original correlations than those estimated by method of moments estimators. When sample size increases the estimated correlation from the maximum likelihood estimation method attain the original correlation.

Table 5.13: Estimated sample correlations 1

Parameters($\beta = 3, \delta = 2.5, r = 5, s = 4$)		
Correlation -0.5477		
Sample size	Estimated correlations	
	Maximum Likelihood	Method of moments
50	-0.5556	-0.6914
100	-0.5514	-0.6068
200	-0.5499	-0.5995
1000	-0.5478	-0.5609

Table 5.14: Estimated sample correlations 2

Parameters ($\beta = 2.5, \delta = 1.5, r = 7, s = 6.5$)		
Correlation -0.6094		
Sample size	Estimated correlations	
	Maximum Likelihood	Method of moments
50	-0.6151	NA
100	-0.6117	-0.7000
200	-0.6110	-0.6702
1000	-0.6094	-0.6156

Table 5.15: Estimated sample correlations 3

Parameters($\beta = 2, \delta = 0.5, r = 9, s = 9$)		
Correlation -0.6417		
Sample size	Estimated correlations	
	Maximum Likelihood	Method of moments
50	-0.6179	-0.6471
100	-0.6434	-0.6836
200	-0.6422	-0.6651
1000	-0.6416	-0.6461

Table 5.16: Estimated sample correlations 4

Parameters ($\beta = 0.7, \delta = 0.07, r = 11, s = 10$)		
Correlation -0.6396		
Sample size	Estimated correlations	
	Maximum Likelihood	Method of moments
50	-0.6429	-0.7180
100	-0.6407	-0.6694
200	-0.6400	-0.6549
1000	-0.6399	-0.6428

We could observe that maximum likelihood is immune to any combinations of parameter values and sample sizes in giving satisfactory results while method of moment estimators is susceptible in these contexts.

References

[1] Arnold, B. C. (1987). *Bivariate distributions with Pareto conditionals.* Statistics and probability letters, 5(4), 263-266.

[2] Arnold, Barry C. (2008). *Enrique Castillos Contributions to Conditional Specification.* Advances in Mathematical and Statistical Modeling. Birkhäuser Boston. 3-19.

[3] Arnold, B. C., Castillo, E., and Sarabia, J. M. (1999). *Conditional specification of statistical models.* Springer.

[4] Arnold, B. C., Castillo, E., and Sarabia, J. M. (2004). *Compatibility of partial or complete conditional probability specifications.* Journal of statistical planning and inference, 123(1), 133-159.

[5] Arnold, B. C., and Strauss, D. (1988). *Bivariate distributions with exponential conditionals.* Journal of the American Statistical Association, 83(402), 522-527.

[6] Arnold, B. C., and Strauss, D. J. (1991). *Bivariate distributions with conditionals in prescribed exponential families.* Journal of the Royal Statistical Society. Series B (Methodological), 365-375.

[7] Arnold, B. C., and Press, S. J. (1989). *Compatible conditional distributions.*Journal of the American Statistical Association, 84(405), 152-156.

[8] Bilmes, J. A. (1998). *A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models.* International Computer Science Institute, 4(510), 126.

[9] Böhning, D. (2000). *Computer-assisted analysis of mixtures and applications.* Technometrics, 42(4), 442-442.

[10] Casella, G., and Berger, R. L. (1990). *Statistical inference* (Vol. 70). Belmont, CA: Duxbury Press.

[11] Castillo, E., and Galambos, J. (1987). *Bivariate distributions with normal conditionals.* Proceedings of the International Association of Science and Technology for Development, 59-62.

[12] Chen, H. Y. (2010). *Compatibility of conditionally specified models.* Statistics and probability letters, 80(7), 670-677.

[13] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal statistical Society, 39(1), 1-38.

[14] Everitt, B. S., and Hand, D. J. (1981). *Finite mixture distributions.* Springer.

[15] Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models: Modeling and Applications to Random Processes.* Springer.

[16] Ip, E. H., and Wang, Y. J. (2009). *Canonical representation of conditionally specified multivariate discrete distributions.* Journal of Multivariate Analysis,100(6), 1282-1290.

[17] Johnson, Norman L., Samuel Kotz, and N. Balakrishnan (1995). *Continuous univariate distributions , vol. 2Wiley.* New York.

[18] Jones, G., Lai, C. D., and Rayner, J. C. W. (2000). *A bivariate gamma mixture distribution.* Communications in Statistics-Theory and Methods, 29(12), 2775-2790.

[19] Keeping, E. S. (1962). *Introduction to statistical inference (Vol. 26).* DoverPublications. com.

[20] Lindsay, B. G. (1995, January). *Mixture models: theory, geometry and applications.* In NSF-CBMS regional conference series in probability and statistics (pp. i-163). Institute of Mathematical Statistics and the American Statistical Association.

[21] McDonald, J. B. (1984). *Some generalized functions for the size distribution of income.* Econometrica: Journal of the Econometric Society, 647-663.

[22] McLachlan, G. J., and Basford, K. E. (1988). *Mixture models. Inference and applications to clustering.* Statistics: Textbooks and Monographs, New York: Dekker, 1988, 1.

[23] McLachlan, G. J., and Peel, D. (2004). *Finite Mixture Models.* John Wiley & Sons.

[24] Mengersen, K., Robert, C., and Titterington, M. (2011). *Mixtures: estimation and applications (Vol. 896).* John Wiley and Sons.

[25] Mohsin, M., Gebhardt, A., Pilz, J., and Spck, G. (2013). *). A new bivariate Gamma distribution generated from functional scale parameter with application to drought data.* Stochastic Environmental Research and Risk Assessment, 27(5), 1039-1054.

Moschopoulos, P., and Sha, N. (2005). *Bayesian inference of scale parameters in exponential family using conditionally specified priors.* Communications in Statistics-Theory and Methods 34.2: 303-318.

[26] Moschopoulos, P., and Staniswalis, J. G. (1994). *Estimation given conditionals from an exponential family.* The American Statistician, 48(4), 271-275.

[27] Newcomb, S. (1886). *A generalized theory of the combination of observations so as to obtain the best result.* American Journal of Mathematics (1886): 343-366.

[28] Pearson, K. (1894). *Contributions to the mathematical theory of evolution.* Philosophical Transactions of the Royal Society of London. A, 71-110.

[29] Peel, D., and McLachlan, G. J. (2000). *Robust mixture modelling using the t distribution.* Statistics and computing, 10(4), 339-348.

[30] Tadikamalla, P. R. (1980). *A look at the Burr and related distributions.* International Statistical Review/Revue Internationale de Statistique, 337-344.

[31] Titterington, D. M., Smith, A. F., and Makov, U. E. (1985). *Statistical analysis of finite mixture distributions (Vol. 7).* New York: Wiley.

[32] Wang, Y. J., and Ip, E. H. (2008). *Conditionally specified continuous distributions.* Biometrika, 95(3), 735-746.

[33] Wu, C. J. (1983). *On the convergence properties of the EM algorithm.* The Annals of statistics, 95-103.

Appendix A

R codes

```
b <- 0.7
d <- 0.07
r <- 11
s <- 10
n <- 1000

Par <- c(b,d,r,s)

set.seed(12)

install.packages("nleqslv")
library("nleqslv")

MLEmat <- NULL

methodOMmat <- NULL

ptm <- proc.time()

for (i in 1:2000){
X <- rgamma(n, shape = r, scale = 1/b )
```

```
deltaX <- d*X

out <- NULL

for(i in 1:length(X)){
dx <- deltaX[i]
Y <- rgamma(1, shape = s, scale = 1/dx)
BIV <- cbind(X[i],Y)
out <- rbind(out,BIV)
}

X <- out[,1]
Y <- out[,2]

MLEeq <- function(x) {
y <- numeric(4)
b <- x[1]
d <- x[2]
r <- x[3]
s <- x[4]
y[1] <- (r/b)-(sum(X)/n)
y[2] <- (s/d)-(sum(X*Y)/n)
y[3] <- log(b)-digamma(r)+(sum(log(X))/n)
y[4] <- log(d)-digamma(s)+(sum(log(X*Y))/n)
y
}

# Starting vector with values for the parameters
```

```
stVct <- c(b,d,r,s)
n <- dim(out)[1]

# *************************************************
# PARAMETER ESTIMATION BY MAXIMUM LIKELIHOOD METHOD
# *************************************************

#=======================
# THE PACKAGE "NLEQSLV"
#=======================

# solves MLEq's with Newton method
MLE <- nleqslv(stVct, MLEeq , method="Newton", global="none",
control=list(trace=1,stepmax=1))
MLE <- MLE$x

MLEmat <- cbind(MLEmat,MLE)

# *****************************************
# PARAMETER ESTIMATION BY METHOD OF MOMENTS
# *****************************************

bMom <- (mean(X)/var(X))
rMom <- (mean(X)^2)/var(X)
sMom <- (mean(Y)^2*(mean(X)^2-var(X)))/(mean(X)^2*var(Y)
-2*var(X)*var(Y)-mean(Y)^2*var(X))
dMom <- (mean(X)*mean(Y))/(mean(X)^2*var(Y)-2*var(X)*var(Y)
-mean(Y)^2*var(X))
```

```
methodOM <- c(bMom,dMom,rMom,sMom)

methodOMmat <- cbind(methodOMmat,methodOM)
}


# Maximum likelihood estimates

aveMLE <- apply(MLEmat,1,mean)
OutMLEave <- cbind(aveMLE,Par)

OutMLEave

mse_b <- (sum(((MLEmat-Par)^2)[1,]))/dim(MLEmat)[2]
mse_d <- (sum(((MLEmat-Par)^2)[2,]))/dim(MLEmat)[2]
mse_r <- (sum(((MLEmat-Par)^2)[3,]))/dim(MLEmat)[2]
mse_s <- (sum(((MLEmat-Par)^2)[4,]))/dim(MLEmat)[2]

mse_b
mse_d
mse_r
mse_s


# Method of moments estimates

aveMOM <- apply(methodOMmat,1,mean)
OutMOMave <- cbind(aveMOM,Par)
```

```
OutMOMave

mse_b_mom <- (sum(((methodOMmat-Par)^2)[1,]))/dim(methodOMmat)[2]
mse_d_mom <- (sum(((methodOMmat-Par)^2)[2,]))/dim(methodOMmat)[2]
mse_r_mom <- (sum(((methodOMmat-Par)^2)[3,]))/dim(methodOMmat)[2]
mse_s_mom <- (sum(((methodOMmat-Par)^2)[4,]))/dim(methodOMmat)[2]

mse_b_mom
mse_d_mom
mse_r_mom
mse_s_mom


# Fisher information matrix

sl <- as.matrix(aveMLE)
fisher <- matrix(0,nrow=4,ncol=4)
fisher[1,1] <- n*(sl[3,1]/(sl[1,1])^2)
fisher[2,2] <- n*(sl[4,1]/(sl[2,1])^2)
fisher[3,3] <- n*trigamma(sl[3,1])
fisher[4,4] <- n*trigamma(sl[4,1])
fisher[1,3] <- (-n/sl[1,1])
fisher[2,4] <- (-n/sl[2,1])

library(lattice)
library(Matrix)
```

```
FishMat <- forceSymmetric(fisher)
Fishe <- solve(FishMat)
Fishe

varMLE <- apply(MLEmat,1,var)
varMOM <- apply(methodOMmat,1,var)


# Efficiency of maximum likelihood estimators

ef_b <- (Fishe[1,1]/varMLE[1])
ef_d <- (Fishe[2,2]/varMLE[2])
ef_r <- (Fishe[3,3]/varMLE[3])
ef_s <- (Fishe[4,4]/varMLE[4])

ef_b
ef_d
ef_r
ef_s


# Relative efficiency of estimators

ef_ml_mom_b <- (mse_b/mse_b_mom)
ef_ml_mom_d <- (mse_d/mse_d_mom)
ef_ml_mom_r <- (mse_r/mse_r_mom)
ef_ml_mom_s <- (mse_s/mse_s_mom)
```

```
ef_ml_mom_b
ef_ml_mom_d
ef_ml_mom_r
ef_ml_mom_s


# Calculations of correlations

RHO <- (-1)*sqrt(s*(r-2)/(r*(s+r-1)))
RHO

RHOmle <- (-1)*sqrt(aveMLE[4]*(aveMLE[3]-2)/(aveMLE[3]*
(aveMLE[4]+aveMLE[3]-1)))
RHOmle

RHOmom <- (-1)*sqrt(aveMOM[4]*(aveMOM[3]-2)/(aveMOM[3]*
(aveMOM[4]+aveMOM[3]-1)))
RHOmom
```

Appendix B

Additional plots

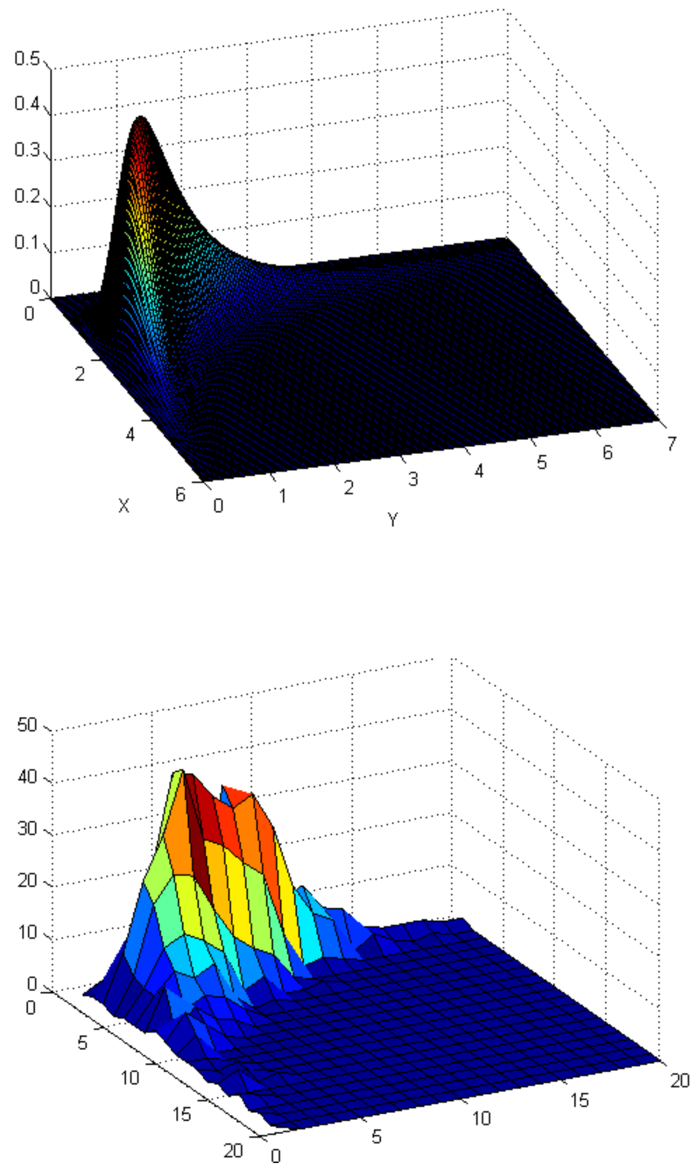

Figure B.1: 3D plot and simulation of the new bivariate distribution 1

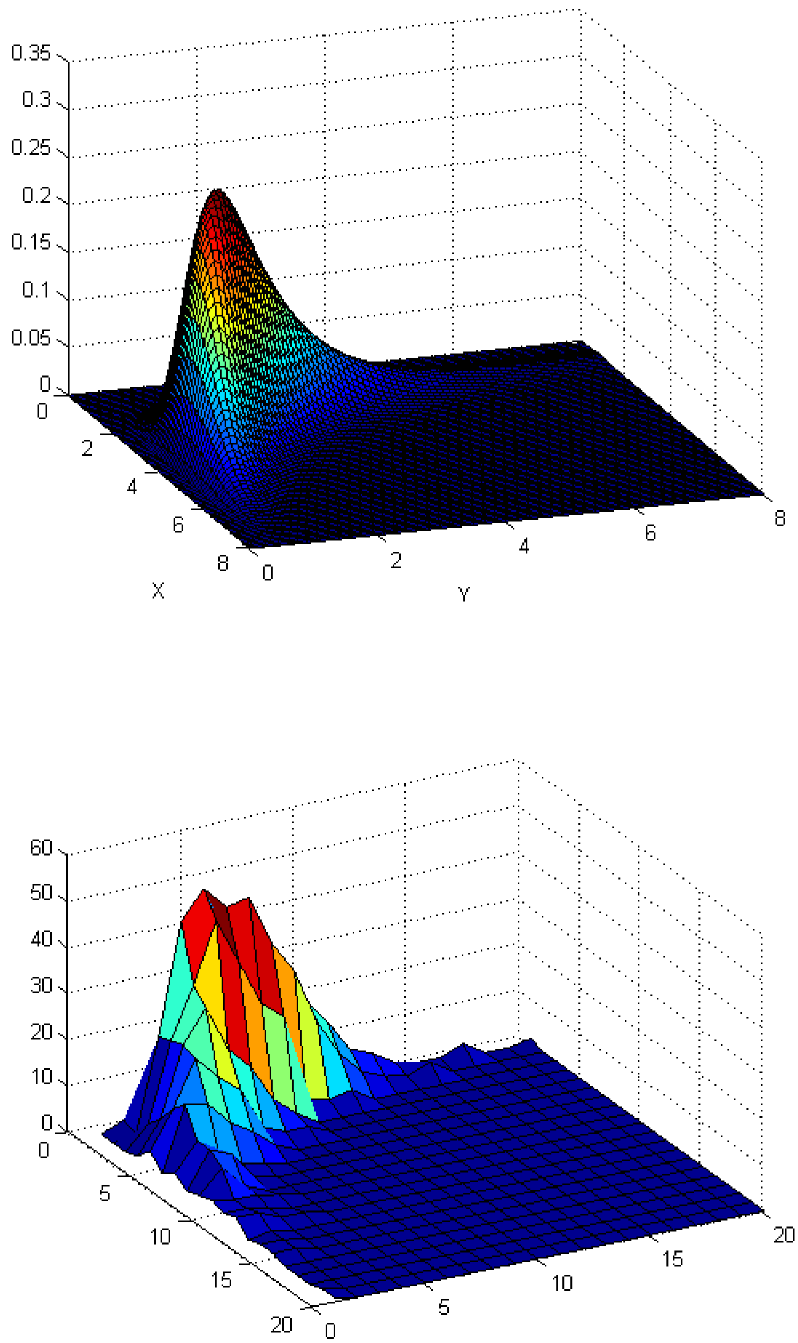

Figure B.2: 3D plot and simulation of the new bivariate distribution 2

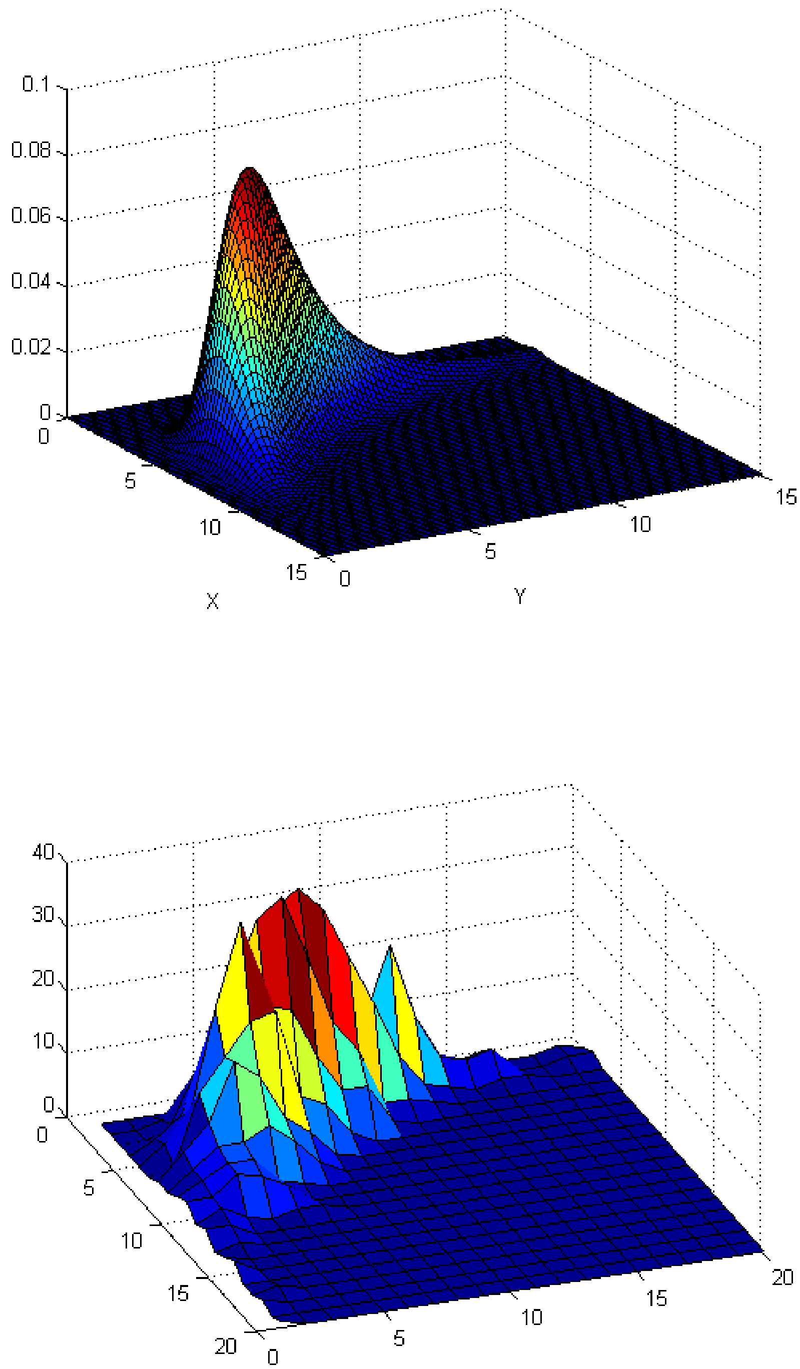

Figure B.3: 3D plot and simulation of the new bivariate distribution 3

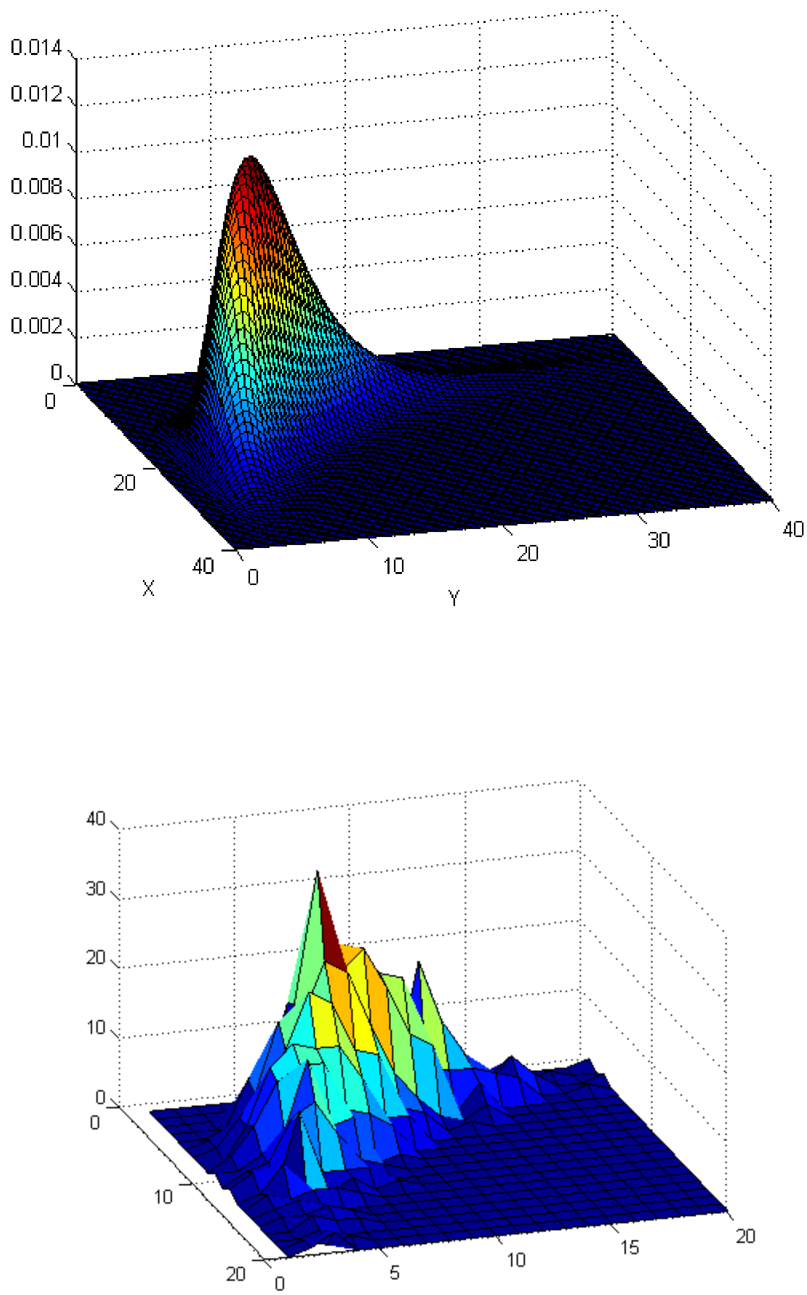

Figure B.4: 3D plot and simulation of the new bivariate distribution 4

Curriculum Vitae

Sathya Amarasekara was born in Sri Lanka on March 22, 1986. He is the third son of Mr. Mahinda Amarasekara and Mrs. Ivonne Rukmani Amarasekara. After graduating from Dharmapala Vidyalaya High School, Pannipitiya, Colombo in 2006, he entered into University of Colombo (UOC) in the fall of 2007 where he pursued a bachelor's degree in Physical Sciences. He graduated with honors after successful completion of the degree. At UOC, Sathya showed active involvement in extra curriculum activities and leadership skills.

In the Fall of 2012, he entered the Graduate School of The University of Texas at El Paso. While pursuing his master's degree in Statistics, he worked as a Teaching Assistant until May 2014. He is currently working as a Research Assistant at the Centre for Institutional Evaluation, Research and Planning (CIERP) since June 2014.

Present address: 806 West Yandell Dr, Apt:02
El Paso, Texas 79902