MAXIMUM LIKELIHOOD ANALYSIS FOR BIVARIATE EXPONENTIAL DISTRIBUTIONS

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
OKYERE EBENEZER
aus
Agona Nsaba, Ghana

Göttingen 2007

First Examiner:	Prof. Dr. Manfred Denker
Second Examiner:	Prof. Dr. Jochen Mau
Day of oral exams:	31 July 2007

Abstract

A generalization of Marshall-Olkin(1967) bivariate exponential model is proposed and the existence, uniqueness and asymptotic distributional properties of the maximum likelihood estimators are studied. The classical Marshall-Olkin model is a mixture of an absolutely continuous and a singular component, that concentrates its mass on the line $x = y$. In this paper, I generalize Marshall-Olkin's results considering a distribution with concentrate positive mass on the line $x = \mu y$. Some simulation results to compare the two models are presented.

I also derive an extension of Marshall-Olkin (1967) model for any function which is continuous and twice continuously differentiable in some open dense domain. This extension gives class of models some of it have exponential marginals. We derive its asymptotic normalities.

I model the first mixed moments of bivariate exponential models whose marginals are also exponential using the method of generalized linear models.

Contents

1 Introduction and Motivation **7**

2 Bivariate Exponential Distributions **13**

2.1 Derivation of Marshall-Olkin BVE Model 14

2.2 Some Examples of Bivariate Exponential Distributions 15

2.2.1 Gumbel's BVE 15

2.2.2 Moran's BVE 16

2.2.3 Freund's Model 16

2.2.4 Block-Basu's ACBVE Model 17

2.3 Marshall-Olkin model revisited 17

2.4 Copula 21

2.4.1 Bivariate Marshall-Olkin Copulas 21

2.4.2 Comparison of the bivariate exponential with the case of independence 22

3 Generalized Marshall-Olkin model **25**

3.1 The generalized Marshall-Olkin model 25

3.2 Maximum Likelihood Estimation 29

3.3 Asymptotic Properties 33

3.3.1 Consistency 34

3.3.2 Asymptotic Normality 37

3.3.3 Asymptotic Efficiency 42

3.3.4 Simulation results 43

4 Model Extension **45**

4.1 Another extension of Marshall-Olkin model 47

4.2 Maximum Likelihood Estimation 49

4.3 Asymptotic Properties 52

4.3.1 Consistency 52

4.3.2 Asymptotic Normality 55

5 Multivariate Generalized Linear Models 61

5.1 GLM for Minimum model 62

BIBLIOGRAPHY 71

Chapter 1

Introduction and Motivation

The idea behind maximum likelihood estimation (MLE) is to determine the parameters that maximize the probability (likelihood) of the sample data. From a statistical point of view, the method of maximum likelihood is considered to be more robust (with some exceptions) and yields estimators with good statistical properties. In other words, MLE methods are versatile and apply to most models and to different types of data. In addition, they provide efficient methods for quantifying uncertainty through confidence bounds. As a general method of estimation it was first introduced by Fisher (1912) in a short paper and he made further developments in a series of papers. MLE has many optimal properties in estimation: sufficiency (complete information about the parameter of interest contained in its MLE estimator); consistency(true parameter value that generated the data recovered asymptotically, i.e. for data of sufficiently large samples); efficiency(lowest-possible variance of parameter estimates achieved asymptotically); and parametrization invariance(same MLE solution obtained independent of the parametrization used).

The consistency of a maximum likelihood estimator has been established under very general conditions by Wald (1949) and Wolfowitz (1949). Conditions needed for it to be asymptotically efficient, that is , consistent and asymptotically normal with variance equal to the Cramer-Rao lower bound has been treated by several authors. Typical conditions are given by Cramer (1946), Dugue (1937), Gurland (1954), Kulldorf (1957). Chanda (1954) generalizes a result by Cramer (1946) and proves, under some regularity conditions stated in Chanda (1954), that there exists a unique solution of the likelihood equations which is consistent and asymptotically normally distributed. Using the same conditions Peters and Walker (1978) show that there is a unique strongly consistent solution of the likelihood equations, which locally maximizes the log-likelihood functions. Consistency problems have been studied in many particular cases: see Jewell (1982), Hathaway (1985), Pfanzagl (1988), Leroux (1992) , Van De Geer (2003) and Atienza et al. (2007). Authors such as, for example Le Cam (1955) and Bahadur (1960) discussed large sample estimation in a more general context. Daniel (1961) developed an argument showing that it is possible to deduce asymptotic efficiency from a much weaker set of assumptions concerning the behaviour of the density function.

Generalized linear models(GLMs) were introduced by Nelder and Wedderburn(1972), as a means of unifying a number of classical statistical models such as normal-theory linear models and analysis of variance, logistic regression, Poisson regression and log linear models for contingency tables. The unification extends to the method of inference, known as analysis of deviance, which generalizes the analysis of variance for normal models. GLMs have unified regression methodology for a wide variety of discrete, continuous, and censored responses that can be assumed to be independent.

A main feature of GLMs is the presence of a linear predictor, which is built from explanatory variables. This linear predictor is linked to the mean response by a so called link function, which may take various forms. Many ideas of linear regression carry over to this wider class of models. An important extension of GLMs is the incorporation of nonparametric parts in the predictor. The parametric model assumes that variables enter the model in the form of a linear predictor in non- and semiparametric regression techniques, however, this assumption is weakened when the covariates are allowed to have unspecified functional form.

An important consideration is that (generalized)linear models are easily understood and can be summarized and communicated to others in a straightforward manner. In addition, parameter estimates from these models can be used to predict or classify new cases simply and readily.

GLMs as described for example by Nelder and Wedderburn (1972) and McCullagh and Nelder (1989) are regression models to analyze continuous or discrete response variables. The association between the response variable and the covariables is given by the so-called link function. GLM assume that the observations are independent and do not consider any correlation between the outcome of the n observations. Marginal models, conditional models and random effects models are extensions of the GLM for correlated data.

There are many publications on these models, like, Gibbons and Hedeker (1997), Heagerty (1999), Heagerty and Zeger (1996), Hedeker and Gibbons (1994), Zeger and Karim (1991), Molenberghs and Lesaffre (1994), Lipsitz and Ibrahim (1996), Daniels and Zhao (2003), Zeger and Qaqish (1988), Zeger, Diggle and Yasui(1990), Zeger(1988) and others. In the marginal model, the primary interest of the analysis is to model the marginal expectation of the response variable given the covariables. Here, the correlation-or more general the association-between the outcome variables is modeled separately and is regarded as nuisance parameter. The major goal is to investigate the effect of the covariables in the population on the response variable. Including the correlation structure in estimating the effects mainly yields different variance estimation. Marginal models have been introduced first by Zeger, Liang and Self(1985), Liang and Zeger(1986).

Exponential distributions have been introduced in a rich literature as a simple model for statistical analysis of lifetimes. There is an extensive literature on the construction of bivariate models, for example, Gumbel (1960), Freund (1961), Block and Basu (1974) and so on. Marshall-Olkin (1967) proposed a multivariate extension of exponential distributions which is much of interest in both theoretical developments and applications.

The physical motivation for the bivariate exponential distribution due to Marshall-Olkin (1967) is common in engineering applications. This model has received the most attention in describing the statistical dependence of components in a 2-component system and in developing statistical inference procedures. Statistical inferences for scale parameters have been considered by many authors. For example, Arnold (1968) and Bemis, Bain and Higgins (1972) derived estimators for the scale parameters. Bhattacharyya and Johnson (1971) and Proschan and Sullo (1976) studied the existence, uniqueness and asymptotic distributional properties of the maximum likelihood estimators.

Objectives of Research: The objectives of this thesis are to

1. generalized Marshall-Olkin (1967) bivariate exponential model and derive its asymptotic normalities. The classical Marshall-Olkin model is a mixture of an absolutely continuous and a singular component, that concentrates its mass on the line $x = y$. In this paper, we generalize Marshall-Olkin's results considering a distribution with concentrate positive mass on the line $x = \mu y$.

2. derive an extension of Marshall-Olkin model for any function which is continuous and twice continuously differentiable in some open dense domain. This extension gives class of models some of it have exponential marginals.

3. model the first mixed moments of bivariate exponential models whose marginals are also exponential using the method of generalized linear models.

As already stated in the objectives, we propose a BVE distribution which is a generalization of Marshall-Olkin model with concentrate positive mass on the line $x = \mu y$ and derive its asymptotic normalities. One important characteristic of this model is that, there is a late failure of one component when a "big shock" strikes both components simultaneously as against the case of Marshall-Olkin's model where both components failures simultaneously when they are strike by the "big shock".

In Marshall-Olkin (1967), the authors derived a multivariate exponential distribution from the points of view designed to indicate the applicability of the distribution. Two of these derivations are based on "shock models" and one is based on the requirement that residual life is independent of Age. In their paper Marshall-Olkin (1967b), the distribution of joint waiting times in a bivariate Poisson process was investigated. They gave several ways of definitions to "joint waiting time". Some of these lead to the Marshall-Olkin BVE model with a joint survival function given as:

$$\overline{F}(x,y) = \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x,y)\}\,;\, x,y \geq 0\,;\lambda_1 > 0\,,\, \lambda_2 > 0\,,\, \lambda_3 \geq 0$$

but others lead to a joint survival function given as:

$$\begin{aligned}\overline{F}(x,y;\vartheta) &= \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \max[x, y + \min(x,\vartheta)]\}\,;\, \vartheta \geq 0\,;\, x,y \geq 0 \\ &= \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \max[x, y + \min(x,-\vartheta)]\}\,;\, \vartheta < 0\,;\, x,y \geq 0\end{aligned}$$

which is the generalization of it. The parameter ϑ is called a "shift" parameter, though it is not simply a location parameter. The course when $\vartheta = 0$ the above equation reduces to Marshall-Olkin BVE model.

Hyakutake (1990) proposed a bivariate distribution having location parameters which is also a generalization of Marshall-Olkin BVE model. The joint survival function is

$$\begin{aligned}\overline{F}(x,y) &= \exp\{-\lambda_1(x-\pi_1) - \lambda_2(y-\pi_2) - \lambda_3 \max[(x-\pi_1, y-\pi_2)]\}; \\ & \quad x > \pi_1\,,\, y > \pi_2\,,\, \lambda_1 > 0, \lambda_2 > 0, \lambda_3 \geq 0\end{aligned}$$

π_1 and π_2 are location parameters. The case where $\pi_1 = \pi_2 = 0$ we have Marshall-Olkin BVE model. He then derived a two-stage procedure of constructing a fixed-size confidence region for the location parameters and the procedure was applied to the ranking and selected problems. The author proposed two-step procedures of testing a hypothesis on a structure of location parameters. None of these two authors examined the asymptotic distributional properties of the maximum likelihood estimators.

The uniqueness and asymptotic properties of the maximum likelihood estimators of Marshall-Olkin BVE model were studied by Bhattacharyya and Johnson (1971) and Proschan and Sullo (1976). Bhattacharyya and Johnson (1971) and Proschan and Sullo (1976) proved the uniqueness of properties of MLE by splitting the negative of the matrix of the second partial derivatives of log likelihood (Hessian matrix) into positive definite matrix and positive semi-definite. They then concluded that the Hessian matrix is negative definite and thus, the log likelihood is strictly concave. In this dissertation, similar method will be used to prove for the uniqueness but we will also show that, the Hessian matrix is negative definite for any vector. Bhattacharyya and Johnson (1971) used the strong consistency property of MLE (c.f. Rao 1965, page 300) to deduce that for large sample, the maximum likelihood estimator is the unique root of the likelihood equation and that the maximum likelihood estimator converges to the true parameter with probability 1. Proschan and Sullo (1976) proved consistency and asymptotic normality by showing that the information matrix of the sample is positive definite and so Cramer-Rao regularity conditions are satisfied. We will prove consistency by considering the behavior of the log likelihood taking at all points on the surface on the sphere with center at a certain true point and with some radius. We will show that for any sufficiently small radius the probability tends to 1 that log likelihood at all points on the surface is less than that at the true point. This will mean that the log likelihood has a local maximum in the interior of the sphere. This will then follow that for any radius, with probability tending to 1 for large sample size, the likelihood equation have a solution within the sphere. Bhattacharyya and Johnson (1971) stated that since the likelihood function satisfies Cramer conditions (c.f. Rao 1965, page 299) asymptotic normality follows. In proving asymptotic normality, we will use results from the prove of consistency that the expectation of the first derivative of the likelihood function is zero so that we can then claim that $\frac{1}{\sqrt{n}}l'$ is asymptotically normal with expectation $\mathbf{0}$ and covariance matrix $\mathbf{\Pi}$, from this the results follow.

Marshall-Olkin (1967) characterize a bivariate distribution, assuming that it has exponential marginals and the following functional equation holds: $\overline{F}(s_1+t, s_2+t) = \overline{F}(s_1, s_2)\overline{F}(t,t)$ which represents a particular type of lack of memory property. This distribution is a mixture of an absolutely continuous and a singular component, that concentrates its mass on the line $x = y$. Muliere and Scarsini (1987) generalize this results by considering a lack-of-memory-property functional equation which involves operations different than the addition: $\overline{F}(s_1 * t, s_2 * t) = \overline{F}(s_1, s_2)\overline{F}(t,t)$ and analogous equations for the marginals. The authors considered an associative, binary operation $*$. They obtained a class of bivariate distributions whose marginals are not necessarily exponential; their form depends on the associative operation. These distributions concentrate positive mass on the line $x = y$ like Marshall-Olkin's one. They also examined some properties of these distributions. In this dissertation, we give the prove of some of the properties. As another form of model extension, we a bivariate exponential function which depends on some function $\varphi(x, y)$, where this function $\varphi(x, y)$ is continuous and twice continuously differentiable in some open dense domain $G = G_\phi \subset \mathbb{R}^2$. We obtain a class of bivariate distributions each of whose marginals depends on the structure of $\varphi(x, y)$. With the assumption that ϕ'' vanishes to make the equation easy to solve, we derive the MLE and examine the asymptotic normalities.

Chapter 2

Bivariate Exponential Distributions

In this chapter we will take a look at the derivation of Marshall-Olkin bivariate exponential model and some examples of bivariate exponential distributions.

Definition 2.0.1 *A bivariate exponential model (BVE) is defined as a two-dimensional distribution function with exponentially distributed one-dimensional marginals.*

If F is a bivariate distribution function, let

$$\overline{F}(x,y) = 1 + F(x,y) - F(x,\infty) - F(\infty,y) \qquad (x,y) \in \mathbb{R}^2,$$

where $F(x,\infty) = \lim_{y\to\infty} F(x,y)$ and similarly for $F(\infty,y)$, We begin noticing the well known characterization of a bivariate distribution function.

Lemma 2.0.1 *Let F_1 and F_2 be one-dimensional distribution functions. Then a function $F : \mathbb{R}^2 \to [0,1]$ is a distribution function with these marginal distributions if and only if the following conditions hold:*

$$\overline{F}(x,-\infty) = 1 - F_1(x) =: \overline{F_1}(x) \qquad ; \qquad \overline{F}(-\infty,y) = 1 - F_2(y) =: \overline{F_2}(y) \qquad x,y \in \mathbb{R}$$

$$\overline{F}(x,\infty) = \overline{F}(\infty,y) = \overline{F}(\infty,\infty) = 0 \qquad x,y \in \mathbb{R}$$

$$\overline{F}(-\infty,-\infty) = 1$$

$$\overline{F}(x_2,y_2) - \overline{F}(x_2,y_1) - \overline{F}(x_1,y_2) + \overline{F}(x_1,y_1) \geq 0 \qquad x_1 \leq x_2;\ y_1 \leq y_2$$

Lemma 2.0.2 *If a bivariate distribution function F has mixed partial derivatives in a domain $G \subset \mathbb{R}^2$, then its probability P is absolutely continuous on G with respect to the Lebesgue measure on G and $P_G = P(\cdot \cap G)$ has density*

$$f_G(x,y) = \frac{\partial^2 \overline{F}}{\partial x \partial y}(x,y) \qquad (x,y) \in G.$$

Corollary 2.0.1 *If $L \subset \mathbb{R}^2$ is a line and if F has mixed partial derivatives in $G = \mathbb{R}^2 \setminus L$, then the probability P associated to F has a decomposition $P = P_{\lambda^2} + P_L$, where P_{λ^2} is a measure on G with density f_G with respect to the two-dimensional Lebesgue measure λ^2 and where P_L is a measure on L.*

Definition 2.0.2 *A bivariate random variable (X, Y) is said to have the loss of memory property (LMP) iff*

$$\overline{F}(x_1 + y, x_2 + y) = \overline{F}(x_1, x_2)\overline{F}(y, y)\,;\; x_1, x_2, y \geq 0 \tag{2.1}$$

where $\overline{F}(x, y) = P(X > x, Y > y)$.

2.1 Derivation of Marshall-Olkin BVE Model

Fatal Shock Model: Marshall and Olkin's (1967) "fatal shock" model assumes that the components of a two-component system die after receiving a shock which is always fatal. Independent Poisson processes $S_1(t; \lambda_1)$, $S_2(t; \lambda_2)$, $S_3(t; \lambda_3)$ govern the occurrence of shocks. Events in the process $S_1(t; \lambda_1)$ are shocks to component 1, events in the process $S_2(t; \lambda_2)$ are shocks to component 2, and events in the process $S_3(t; \lambda_3)$ are shocks to both components. The joint survival distribution (X, Y) of the components 1 and 2 is

$$\begin{aligned}
\overline{F}(x, y) &= P(X > x\,,\, Y > y) \\
&= P\{S_1(x; \lambda_1) = 0\,,\, S_2(y; \lambda_2) = 0\,,\, S_3(\max(x, y)\,; \lambda_3 = 0)\} \\
&= \exp[-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x, y)] \;\; ; x \geq 0, y \geq 0
\end{aligned} \tag{2.2}$$

Nonfatal Shock Model: Let assume that the shocks from the three sources are not necessarily fatal. Instead a shock from source 1 causes the failure of component 1 with probability q_1, a shock from source 2 causes the failure of component 2 with probability q_2. Also, a shock from source 3 causes the failure

1. of both components, with probability q_{11}
2. of component 1 only, with probability q_{10}
3. of component 2 only, with probability q_{01}
4. of neither component, with probability q_{00}

where $q_{11} + q_{10} + q_{01} + q_{00} = 1$. We assume that each shock represents an independent opportunity for failure.

Then the joint survival probability for X, the life length of component 1, and for Y, the life length of component 2, may be written according to Barlow and Proschan (1975) as

$$\begin{aligned} P[X > x, Y > y] &= \left\{\sum_{k=0}^{\infty} e^{-\lambda_1 x}\frac{(\lambda_1 x)^k}{k!}(1-q_1)^k\right\} \\ &\times \left\{\sum_{l=0}^{\infty} e^{-\lambda_2 y}\frac{(\lambda_2 y)^l}{l!}(1-q_2)^l\right\} \\ &\times \left\{\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\left[e^{-\lambda_3 x}\frac{(\lambda_3 x)^m}{m!}q_{00}^m\right]\right. \\ &\times \left.\left[e^{-\lambda_3 (y-x)}\frac{(\lambda_3 (y-x))^n}{n!}(q_{00}+q_{10})^n\right]\right\} \end{aligned}$$

when $0 \le x \le y$. Summing series and simplifying, using the fact that

$$\sum_{k=0}^{\infty} e^{-\lambda_1 x}\frac{(\lambda_1 x)^k}{k!}(1-q_1)^k = e^{-\lambda_1 x}e^{\lambda_1 x(1-q)}$$

we obtain

$$P[X > x, Y > y] = \exp\{-x[\lambda_1 q_1 + \lambda_3 q_{10}] - y[\lambda_2 q_2 + \lambda_3(1 - q_{00} - q_{10})]\}.$$

For $0 \le y \le x$, by symmetry,

$$P[X > x, Y > y] = \exp\{-x[\lambda_1 q_1 + \lambda_3(1 - q_{00} - q_{10})] - y[\lambda_2 q_2 + \lambda_3 q_{01})]\}$$

combining the two survival probabilities, we have the BVE

$$P[X > x, Y > y] = \exp\{-\lambda_1^* x - \lambda_2^* y - \lambda_3^* \max(x, y)\}$$

where $\lambda_1^* = \lambda_1 q_1 + \lambda_3 q_{10}$, $\lambda_2^* = \lambda_2 q_2 + \lambda_3 q_{01}$, $\lambda_3^* = \lambda_3 q_{11}$.

2.2 Some Examples of Bivariate Exponential Distributions

2.2.1 Gumbel's BVE

Gumbel (1960) studied the bivariate exponential distribution, given by the joint distribution

$$\overline{F}(x, y) = 1 - \exp\{-x\} - \exp\{-y\} + \exp\{-x - y - \delta xy\}\,;\ x, y > 0, 0 \le \delta \le 1. \quad (2.3)$$

The marginal probabilities are exponential.

2.2.2 Moran's BVE

Moran (1967), considered a class of distributions with positive correlation ω, and marginal distributions of gamma type whose index parameter is any positive integral multiple of 1/2. He considered the case where the marginal distributions are negative exponentials. Let

$$X = U_1^2 + U_2^2 \quad Y = U_3^2 + U_4^2$$

where U_1, U_3 are jointly distributed normally with zero means, variances 1/2 and correlation $\omega\,(0 \leq \omega \leq 1)$. Here U_2, and U_4 are independent of (U_1, U_3) but have the same joint distribution. The joint probability distribution is of the form

$$F(x, y) = \sum_{n=0}^{\infty} \omega^{2n} C_n(x, y)$$

where C_n has a Fourier transform

$$C_n(x, y) = \sum_{j=0}^{n} n!/(r!\,(n-r)!)(-1)^j/j!\,x^j e^{-x} \sum_{k=0}^{n} n!/(k!\,(n-k)!)(-1)^k/k!\,y^k e^{-y}.$$

2.2.3 Freund's Model

Freund(1961) suggested a bivariate distribution based on a model where two components share a common load. Suppose that X and Y are random variables representing the lifetimes of two components 1 and 2. The respective density functions(when both components are in operation) are

$$f_1(x) = \lambda_1 \exp\{-\lambda_1\}\,;\ x > 0$$

$$f_2(y) = \lambda_2 \exp\{-\lambda_2\}\,;\ y > 0$$

for $\lambda_1, \lambda_2 > 0$, then component 1 and component 2 are dependent in that a failure of either component changes the parameter of the life distribution of the other component. Thus when component 1 fails, the parameter of component 2 becomes λ_2', when component 2 fails, the parameter for component 1 becomes λ_1'. There is no other dependence. The joint survival distribution for the two components is given as:

$$\begin{aligned} \overline{F}(x, y) &= \frac{\lambda_1}{\breve{\lambda} - \lambda_2'}[\exp(-\lambda_2' y - (\breve{\lambda} - \lambda_2')x)] + \frac{\lambda_2 - \lambda_2'}{\breve{\lambda} - \lambda_2'}\exp(-\breve{\lambda} y)\,;\ x < y \\ &= \frac{\lambda_2}{\breve{\lambda} - \lambda_1'}[\exp(-\lambda_1' x - (\breve{\lambda} - \lambda_1')y)] + \frac{\lambda_1 - \lambda_1'}{\breve{\lambda} - \lambda_1'}\exp(-\breve{\lambda} x)\,;\ y < x \end{aligned} \tag{2.4}$$

where $\breve{\lambda} = \lambda_1 + \lambda_2$. The marginal distributions are in general not exponential.

2.2.4 Block-Basu's ACBVE Model

Block and Basu(1974) proposed absolutely continuous bivariate exponential ACBVE which is absolutely continuous part of Marshall-Olkin (1967) also a proper sub-family of Freund(1961) with joint distribution given by

$$\begin{aligned} \overline{F}(x,y) &= \frac{\lambda}{(\lambda_1+\lambda_2)} \exp[-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x,y)] \\ &- \frac{\lambda_3}{(\lambda_1+\lambda_2)} \exp[-\lambda \max(x,y)]\,;\; x,y>0. \end{aligned} \tag{2.5}$$

The marginal distributions are given as:

$$\begin{aligned} \overline{F}_1(x) &= \frac{\lambda}{\lambda_1+\lambda_2} \exp[-(\lambda_1+\lambda_3)x] \\ &- \frac{\lambda_3}{\lambda_1+\lambda_2} \exp(-\lambda x)\,;\; x>0 \\ \overline{F}_2(y) &= \frac{\lambda}{\lambda_1+\lambda_2} \exp[-(\lambda_2+\lambda_3)y] \\ &- \frac{\lambda_3}{\lambda_1+\lambda_2} \exp(-\lambda y)\,;\; y>0 \end{aligned}$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_3$. The above marginals are not exponential. Thus ACBVE model is not a special case of the BVE since the BVE must have exponential marginals and ACBVE does not.

2.3 Marshall-Olkin model revisited

In this section, we will introduce a different model of the Marshall-Olkin (1967) bivariate exponential distribution (BVE).

Let (X, Y) be bivariate random variable. We propose a BVE for (X, Y) to be of the form

$$\begin{aligned} \overline{F}^M_{\lambda}(x,y) &= P(X>x, Y>y) \\ &= \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \min(x,y)\}\,;\; x,y>0, \underline{\lambda} \in \Lambda^+ \end{aligned} \tag{2.6}$$

where

$$\underline{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$$

and the parameter space is

$$\Lambda^+ = \{\underline{\lambda} : 0 \le \lambda_i < \infty, i = 1,2, \lambda_3 < 0; \lambda_3 + \lambda_j > 0, j = 1,2\}. \tag{2.7}$$

Lemma 2.3.1 *1. Λ^+ is convex*

2. *For every* $\lambda \in \Lambda^+$ *;*

$$F_\lambda^M(x,y) = \overline{F}_\lambda^M(x,y) + F_1^M(x) + F_2^M(y) - 1$$

is a bivariate distribution function on $\mathbb{R}^2$*.*

Proof: The proof of (1) is obvious from the definition of convexity. For the proof of (2), we show that $\overline{F}_\lambda^M(x,y)$ defines a distribution function if the conditions of Lemma 2.0.1 are satisfied. Clearly, by definition,

$$\lim_{x\to\infty} \overline{F}_\lambda^M(x,y) = \lim_{y\to\infty} \overline{F}_\lambda^M(x,y) = 0$$

We have

$$\overline{F}_\lambda^M(x,-\infty) = \overline{F}_\lambda^M(x,0) = \exp[-\lambda_1 x]$$

and

$$\overline{F}_\lambda^M(-\infty,y) = \overline{F}_\lambda^M(0,y) = \exp[-\lambda_2 y].$$

Finally, let $x_1 \le x_2 \le y_1 \le y_2$. Then

$$\begin{aligned} & \overline{F}(x_2,y_2) - \overline{F}(x_2,y_1) - \overline{F}(x_1,y_2) + \overline{F}(x_1,y_1) \\ = \; & \left(\exp[-(\lambda_1+\lambda_3)x_2] - \exp[-(\lambda_1+\lambda_3)x_1]\right)\left(\exp[-\lambda_2 y_2] - \exp[-\lambda_2 y_1]\right) \ge 0. \end{aligned}$$

The following results can easily be checked

1. $\overline{F}_\lambda^M(x,0) = 0$ when $x \to \infty$
2. $\overline{F}_\lambda^M(0,y) = 0$ when $y \to \infty$
3. $\overline{F}_\lambda^M(0,0) = 1$
4. $\overline{F}_\lambda^M(\infty,\infty) = 0$

So, from above, we have

$$\begin{aligned} F_\lambda^M(0,0) &= 1 + \overline{F}_\lambda^M(0,0) - \overline{F}_\lambda^M(0,0) - \overline{F}_\lambda^M(0,0) \\ &= 1+1-1-1 \\ &= 0 \end{aligned}$$

also,

$$\begin{aligned} F_\lambda^M(\infty,\infty) &= 1 + \overline{F}_\lambda^M(\infty,\infty) - \overline{F}_\lambda^M(\infty,0) - \overline{F}_\lambda^M(0,\infty) \\ &= 1+0-0-0 \\ &= 1. \qquad \square \end{aligned}$$

The minimum of X and Y given as

$$\begin{aligned} P(\min(X,Y) > x) &= P(X > x, Y > x) \\ &= \exp\{(-\lambda_1 - \lambda_2 - \lambda_3)x\}\,;\ x > 0 \end{aligned}$$

that is the minimum of X and Y is exponential with parameter $(\lambda_1 + \lambda_2 + \lambda_3)$. Let the partial derivative $\frac{\partial^2 \overline{F}(x,y)}{\partial x \partial y}$ exist almost everywhere, then the joint density $f(x,y) \geq 0$ as Lemma 2.0.2, is defined as

$$f(x,y) = \frac{\partial^2 \overline{F}(x,y)}{\partial x \partial y}\,;\ x\,,y > 0$$

so the joint density for eqn. 2.6 is given as

$$f(x,y) = \begin{cases} \lambda_2(\lambda_1 + \lambda_3)\,\overline{F}_\lambda^M(x,y)\,;\ x < y\,;\ x\,,y > 0 \\ \lambda_1(\lambda_2 + \lambda_3)\,\overline{F}_\lambda^M(x,y)\,;\ y < x\,;\ x\,,y > 0 \\ -\lambda_3\,\overline{F}_\lambda^M(x,x)\,;\ x = y\ > 0 \end{cases} \tag{2.8}$$

Because $P(X = Y)$ is not equal to zero, the function $f(x,y)$ may be considered to be a density for the minimum model, if it is understood that the first two terms are densities with respect to two-dimensional Lebesque measure and the third is a density with respect to one-dimensional Lebesque measure. The conditional probability distribution $P(X > x \mid Y = y)$ is derived (cf.Barlow and Proschan, 1975, page 132) by differentiating $\overline{F}(x,y)$ w.r.t. y (evaluated at value y) and divided it by the pdf of Y. Thus

$$P(X > x \mid Y = y) = \begin{cases} \lambda_2^{-1}(\lambda_2 + \lambda_3)\exp\{-\lambda_1 x - \lambda_3 y\}\,;\ y < x\,;\ x\,,y > 0 \\ \exp\{-(\lambda_1 + \lambda_3)x\}\,;\ y > x\,;\ x\,,y > 0 \end{cases} \tag{2.9}$$

Lemma 2.3.2 *The family of distributions given by* $\overline{F}_\lambda^M : \lambda \in \Lambda^+$ *is exactly the family of distributions given by the Marshall-Olkin model.*

Proof Using the identity

$$x + y = \max(x,y) + \min(x,y)$$

we can rewrite $\overline{F}_\lambda^M$ $(\lambda \in \Lambda^+)$ as

$$\overline{F}_\lambda^M(x,y) = \exp[-(\lambda_1 + \lambda_3)x - (\lambda_2 + \lambda_3)y + \lambda_3 \max(x,y)].$$

Reparametrising $\eta_i = \lambda_i + \lambda_3$ $(i = 1,2)$ and $\eta_3 = -\lambda_3$ shows the claim. □

We consider a model similar to eqn. 2.6 which was proposed by Marshall and Olkin in 1967. Let $\overline{F}_{mo}(x,y)$ denote Marshall-Olkin's model.

$$\overline{F}_{mo}(x,y) = \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x,y)\}\,;\ x,y > 0 \tag{2.10}$$

Theorem 2.3.1 *If* $\overline{F}_{mo}(x,y)$ *is* $BVE(\lambda_1, \lambda_2, \lambda_3)$ *and* $\lambda = \lambda_1 + \lambda_2 + \lambda_3$*, then*

$$\overline{F}_{mo}(x,y) = \frac{\lambda_1 + \lambda_2}{\lambda}\overline{F}_a(x,y) + \frac{\lambda_3}{\lambda_1 + \lambda_2}\overline{F}_s(x,y)$$

where

$$\overline{F}_s(x,y) = \exp[-\lambda \max(x,y)]$$

is a singular distribution, and

$$\overline{F}_a(x,y) = \frac{\lambda}{\lambda_1 + \lambda_2}\exp[-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x,y)] - \frac{\lambda_3}{\lambda_1 + \lambda_2}\exp[-\lambda \max(x,y)]$$

is absolutely continuous.

Proof of Theorem 2.3.1 The proof can be found in Marshall-Olkin (1967, page 34 and 35). □

A singular distribution is of the fact that its mixed second partial derivative is zero where $x \neq y$, and the absolutely continuous is from the fact that its mixed second partial derivative is a density. In the case of the BVE, the presence of a singular part is a reflection of the fact that if X and Y are BVE, then $X = Y$ with positive probability, whereas the line $x = y$ has two-dimensional Lebesgue measure zero.

The marginal distributions for X and Y are given by

$$\overline{F}_{mo1}(x) = \exp\{-(\lambda_1 + \lambda_3)x\}\,; x > 0$$

and the corresponding pdf is

$$f_{mo1}(x) = (\lambda_1 + \lambda_3)\exp\{-(\lambda_1 + \lambda_3)x\}\,;\; x > 0$$

similarly, the marginal distribution of Y is

$$\overline{F}_{mo2}(y) = \exp\{-(\lambda_2 + \lambda_3)y\}\,;\; y > 0$$

and the pdf is

$$f_{mo2}(y) = (\lambda_y + \lambda_3)\exp\{-(\lambda_y + \lambda_3)y\}\,;\; y > 0$$

The minimum of X and Y for Marshall-Olkin is again exponential with parameter λ (cf. Marshall-Olkin, 1967, page 37). The conditional probability distribution $P(X > x \mid Y = y)$ for Marshall-Olkin's model is

$$P(X > x \mid Y = y) = \begin{cases} \lambda_2(\lambda_2 + \lambda_3)^{-1}\exp\{-(\lambda_1 + \lambda_3)x + \lambda_3 y\}\,;\; y < x\,;\; x\,, y > 0 \\ \exp\{-\lambda_1 x\}\,;\; y > x\,;\; x\,, y > 0 \end{cases} \tag{2.11}$$

2.4 Copula

The copula concept is used frequently in survival analysis and actuarial sciences.

Definition 2.4.1 *A 2-dimensional copula C is the joint distribution function*

$$C : [0\,,\,1]^2 \to [0\,,\,1]$$

of a vector (U,V) of a uniform (0,1) random variables, that is

$$C(u,v) \;=\; P(U \leq u\,,\, V \leq v),\;\; u,v \in [0,1].$$

The following theorem, which was first proved by Sklar in 1959, states that for any joint distribution function H there exists a copula C that "couples" H to its marginal distribution functions G_1 and G_2. Before we can state the result, we first recall that a distribution function G is non-decreasing with $\lim_{x\to-\infty} G(x) = 0$ and $\lim_{x\to\infty} G(x) = 1$. As a distribution function G does not need to be strictly increasing, we define its quasi-inverse as

$$G^{(-1)}(t) = \inf\{x : G(x) > t\}.$$

If G is strictly increasing then the quasi-inverse is just the ordinary inverse.

Theorem 2.4.1 {**Sklar 1996**} *Let H be a joint distribution function with marginals G_1, G_2. Then there exists a copula C such that*

$$H(x,y) \;=\; C(G_1(x), G_2(y)).$$

Let $\bar{H} = \{(G_1(x), G_2(y)) :\; x, y \in \mathbb{R}\}$, then for any $(u,v) \in \bar{H}, C$ is given by

$$C(u,v) \;=\; H(G_1^{(-1)}(u), G_2^{(-1)}(v))$$

In particular, if G_1, G_2 are continuous then C is unique; otherwise C is uniquely determined on $RanG_1 \times RanG_2$, where $RanG$ denotes range of G. Conversely, if C is a copula and G_1, G_2 are distribution functions then the function H defined above is a joint distribution function with marginals G_1, G_2

2.4.1 Bivariate Marshall-Olkin Copulas

In the Marshall-Olkin model, the times till the event occurs which kills component 1 only, 2 only or both the components is modeled by independent exponential random variables T_1, T_2 and T_3 with parameters λ_1, λ_2 and λ_3 respectively. Then $X = \min\{T_1, T_3\}$ and

$Y = \min\{T_2, T_3\}$ and the probability that component 1 survives beyond time x and component 2 beyond y is given by

$$P(X > x, Y > y) = \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x, y)\}\,;\; x\,,\, y > 0$$

using the fact that $x + y - \min\{x, y\}$ we learn that

$$\begin{aligned} P(X > x, Y > y) &= \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x, y)\}\,;\; x, y > 0 \\ &= \exp\Big\{-(\lambda_1 + \lambda_3)x - (\lambda_2 + \lambda_3)y + \lambda_3 \min\{x, y\}\Big\} \\ &= P(X > x)P(Y > y) \min\{\exp(\lambda_3 x), \exp(\lambda_3 y)\} \end{aligned}$$

let $\alpha_1 = \lambda_3/(\lambda_1 + \lambda_3)$ and $\alpha_2 = \lambda_3/(\lambda_2 + \lambda_3)$, then

$$\exp(\lambda_3\, x) = \overline{F}_{mo1}(x)^{-\alpha_1}\ ,\ \exp(\lambda_3\, y) = \overline{F}_{mo2}(y)^{-\alpha_2}$$

and hence the the survival copula of (X, Y) is given by

$$\hat{C}(u, v) = u, v \min(u^{-\alpha_1}\,,\, v^{-\alpha_2}) = \min(u^{1-\alpha_1} v\,,\, uv^{1-\alpha_2}).$$

This construction leads to a copula family given by

$$\begin{aligned} C_{\alpha_1,\alpha_2}(u, v) &= \min(u^{1-\alpha_1} v\,,\, uv^{1-\alpha_2}) \\ &= u^{1-\alpha_1} v\,,\, u^{\alpha_1} \geq v^{\alpha_2} \\ &= uv^{1-\alpha_2}\,,\, u^{\alpha_1} \leq v^{\alpha_2}. \end{aligned}$$

This family is known as the Marshall-Olkin family.

2.4.2 Comparison of the bivariate exponential with the case of independence

It is common practice in reliability theory to assume the components of a system have independent life lengths. It is of interest to see the effect of this assumption when in fact the lives have a BVE distribution.

Let us suppose the marginal distributions are known to be given by

$$\overline{F}_1(x) = \exp\{-\lambda_1 x\}\,,\, \overline{F}_2(y) = \exp\{-\lambda_2 y\}$$

suppose that we operate under the assumption that the joint distribution $F(x, y)$ is $F_1(x)F_2(y)$, when in fact, $\overline{F}(x, y)$ is given by eqn. 2.6. Clearly, the difference

$$\overline{F}(x, y) - \overline{F}_1(x)\overline{F}_2(y) = \exp\{-\lambda_1 x - \lambda_2 y\}(\exp\{-\lambda_3 \min(x, y)\} - 1)$$

is negative for larger values of x and y, so the probability that both items survive is lesser.

On the other hand, using $\overline{F}_{mo}(x,y)$ in eqn. 2.10, the difference

$$\overline{F}_{mo}(x,y) - \overline{F}_{mo1}(x)\overline{F}_{mo2}(y) = \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x,y)\}(1 - \exp\{-\lambda_3 \min(x,y)\})$$

is positive for all x and y, so the probability that both items survive is actually greater than the assumption of independence would lead us to believe.

Lemma 2.4.1 *If (X,Y) is distributed as the minimum model of eqn. 2.6, then*

1. $P(Y < X) = (\lambda_2 + \lambda_3)\,/\,(\lambda_1 + \lambda_2 + \lambda_3)$
2. $P(X < Y) = (\lambda_1 + \lambda_3)\,/\,(\lambda_1 + \lambda_2 + \lambda_3)$
3. $P(X = Y) = -\lambda_3\,/\,(\lambda_1 + \lambda_2 + \lambda_3)$

Proof The proof of point (1) is as follows:

$$\begin{aligned}
P(Y < X) &= \int_0^\infty \int_0^x \lambda_1(\lambda_2+\lambda_3)\exp\{-\lambda_1 x - (\lambda_2+\lambda_3)y\}dx\,dy \\
&= \int_0^\infty \lambda_1(\lambda_2+\lambda_3)\exp\{-\lambda_1 x\}\Big[\int_0^x \exp\{-(\lambda_2+\lambda_3)y\}\,dy\Big]dx \\
&= \lambda_1\Big[\int_0^\infty \exp\{-\lambda_1 x\,dx\} - \int_0^\infty \exp\{-(\lambda_1+\lambda_2+\lambda_3)x\,dx\,\}\Big] \\
&= (\lambda_2+\lambda_3)\,/\,(\lambda_1+\lambda_2+\lambda_3).
\end{aligned}$$

The proof of point (2) follows from the symmetric property of the distribution.

We proof point(3) as follows:

$$\begin{aligned}
P(X = Y) &= -\lambda_3 \int_0^\infty \exp\{-(\lambda_1+\lambda_2+\lambda_3)\,x\}dx \\
&= -\lambda_3/(\lambda_1+\lambda_2+\lambda_3)\Big[\exp\{-(\lambda_1+\lambda_2+\lambda_3)x\}\Big]_\infty^0 \\
&= -\lambda_3/(\lambda_1+\lambda_2+\lambda_3). \qquad \square
\end{aligned}$$

Lemma 2.4.2 *If (X,Y) is distributed as the Marshall-Olkin model of eqn. 2.10, then*

1. $P(Y < X) = \lambda_2\,/\,(\lambda_1 + \lambda_2 + \lambda_3)$
2. $P(X < Y) = \lambda_1\,/\,(\lambda_1 + \lambda_2 + \lambda_3)$
3. $P(X = Y) = \lambda_3\,/\,(\lambda_1 + \lambda_2 + \lambda_3)$

Proof The proof of point (1) is as follows:

$$\begin{aligned}
P(X < Y) &= \int_0^\infty \int_0^y \lambda_1(\lambda_2+\lambda_3)\exp\{-\lambda_1 x - (\lambda_2+\lambda_3)y\}dx\,dy \\
&= \lambda_1(\lambda_2+\lambda_3)\Big[\int_0^\infty \exp\{-(\lambda_2+\lambda_3)y\}\,dy\Big]\int_0^y \exp\{-(\lambda_1)x\,dx\} \\
&= (\lambda_2+\lambda_3)\Big[\int_0^\infty \exp\{(-\lambda_2+\lambda_3)y\,dy\} - \int_0^\infty \exp\{-(\lambda_1+\lambda_2+\lambda_3)y\,dy\,\}\Big] \\
&= \lambda_1\,/\,(\lambda_1+\lambda_2+\lambda_3)
\end{aligned}$$

The proof of point (2) follows from the symmetric property of the distribution.

We proof point(3) as follows:

$$\begin{aligned} P(X = Y) &= 1 - P(X < Y) - P(X > Y) \\ &= 1 - \lambda_1 / (\lambda_1 + \lambda_2 + \lambda_3) - \lambda_2 / (\lambda_1 + \lambda_2 + \lambda_3) \\ &= \lambda_3/(\lambda_1 + \lambda_2 + \lambda_3). \qquad \square \end{aligned}$$

Chapter 3

Generalized Marshall-Olkin model

3.1 The generalized Marshall-Olkin model

There is an extensive literature on the generalization of Marshall-Olkin bivariate exponential model, for example, Marshall-Olkin(1967b) proposed that there are several ways to define " joint waiting time". Some of these lead to the bivariate exponential distribution previously obtained by the authors, but other lead to a generalization of it. Hyakutake (1990) proposed a Marshall-Olkin BVE distribution having location parameters. Ryu (1993) extended Marshall-Olkin's BVE such that it is absolutely continuous and need not be memoryless.

Lemma 3.1.1 *For $\lambda_i \geq 0$, $i = 1, 2, 3$, and $\mu \geq 0$ the function*

$$\overline{F}_\lambda(x, y) = 1_{\mathbb{R}^2_+}(x, y) \exp[-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x, \mu y)]$$

defines a bivariate exponential model where the marginals have parameters $\lambda_1 + \lambda_3$ and $\lambda_2 + \mu\lambda_3$.

Proof: The function $\overline{F}_\lambda(x, y)$ defines bivariate exponential model if both marginal distributions are exponential, cf. Johnson and Kotz (1972, page 260). Let $\overline{F}_1(x) = \overline{F}(x, 0)$ and $\overline{F}_2(y) = \overline{F}(0, y)$ denotes the marginal distributions of X and Y respectively. Then

$$\begin{aligned}\overline{F}_1(x) &= \exp\{-\lambda_1 x - \lambda_2(0) - \lambda_3 \max(x, \mu(0))\}\,;\, x > 0 \\ &= \exp\{-(\lambda_1 + \lambda_3)x\}\,;\, x > 0\end{aligned}$$

which is exponential with parameter $(\lambda_1 + \lambda_3)$. Also,

$$\begin{aligned}\overline{F}_2(y) &= \exp\{-\lambda_1(0) - \lambda_2 y - \lambda_3 \max((0), \mu y)\}\,;\, y > 0 \\ &= \exp\{-(\lambda_2 + \mu\lambda_3)y\}\,;\, y > 0\end{aligned}$$

which is exponential with parameter $(\lambda_2 + \mu\lambda_3)$. Hence the function $\overline{F}_\lambda(x, y)$ defines bivariate exponential model. $\square$

Corollary 3.1.1 *Let $\lambda = (\lambda_1, \lambda_2, \lambda_3, \mu)$. Then F_λ restricted to the domain $G = \mathbb{R}_+^2 \setminus \{(x, y) : x = \mu y\}$ has a density with respect to $dxdy$ given by*

$$f_G(x, y) = \lambda_1(\lambda_2 + \mu\lambda_3)\overline{F}_\lambda(x, y)1_{\{x<\mu y\}} + \lambda_2(\lambda_1 + \lambda_3)\overline{F}_\lambda(x, y)1_{\{x>\mu y\}}$$

and the measure on the line $L = \{(x, y) : x = \mu y\}$ is given by the density

$$f_L(x, y) = \frac{\lambda_3}{\sqrt{1 + \mu^2}} \exp -(\lambda_1 + \lambda_2/\mu + \lambda_3)x$$

with respect to the measure $\sqrt{1 + \mu^2}dx$ on L.

Proof:

$$\overline{F}(u, v) = \overline{F}_G + \overline{F}_L.$$

Now,

$$\begin{aligned} \int_u \int_v f_G(x, y)dxdy &= \int_u \int_v \lambda_1(\lambda_2 + \mu\lambda_3)\overline{F}_\lambda(x, y)1_{\{x<\mu y\}} \\ &+ \lambda_2(\lambda_1 + \lambda_3)\overline{F}_\lambda(x, y)1_{\{x>\mu y\}}dxdy. \end{aligned}$$

When $x < \mu y$, we have

$$B(u, v) = \lambda_1(\lambda_2 + \mu\lambda_3) \int_v^\infty \int_u^{\mu v} e^{-\lambda_1 x-(\lambda_2+\mu\lambda_3)y}\, dxdy$$

$$= \lambda_1(\lambda_2 + \mu\lambda_3) \int_v^\infty e^{-(\lambda_2+\mu\lambda_3)y}dy \int_u^{\mu v} e^{-\lambda_1 x}\, dx$$

$$= \exp\{-\lambda_1 u - \lambda_2 v - \lambda_3\mu v\} - \exp\{-\lambda_1\mu v - \lambda_2 v - \lambda_3\mu v\}.$$

When $x > \mu y$, we have

$$\begin{aligned} C(u, v) &= \lambda_2(\lambda_1 + \lambda_3) \int_u^\infty \int_v^{\frac{u}{\mu}} e^{-(\lambda_1+\lambda_3)x-\lambda_2 y}\, dxdy \\ &= \lambda_2(\lambda_1 + \lambda_3) \int_u^\infty e^{-(\lambda_1+\lambda_3)x}dx \int_v^{\frac{u}{\mu}} e^{-\lambda_2 y}\, dy \\ &= \exp\{-\lambda_1 u - \lambda_2 v - \lambda_3 u\} - \exp\{-\lambda_1 u - \lambda_2 u\mu^{-1} - \lambda_3 u\}. \end{aligned}$$

If $x < \mu y$,

$$\begin{aligned} \overline{F}_G(u, v) &= B(u, v) + C(\mu v, v) \\ &= e^{-\lambda_1\mu-\lambda_2 v-\lambda_3\mu v} - e^{-\lambda_1\mu v-\lambda_2 v-\lambda_3\mu v}. \end{aligned}$$

If $x > \mu y$,

$$\begin{aligned}\overline{F}_G(u,v) &= B(u,\frac{u}{\mu}) + C(u,v) \\ &= e^{-\lambda_1 u-\lambda_2 v-\lambda_3 u} - e^{-\lambda_1 u-\lambda_2 \frac{u}{\mu}-\lambda_3 u}.\end{aligned}$$

since,

$$\overline{F}_L(u,v) = \exp\{-(\lambda_1 + \frac{\lambda_2}{\mu} + \lambda_3)u\} \ ; \ u = \mu v$$

so we have that

$$\overline{F}(u,v) = \overline{F}_G + \overline{F}_L. \qquad \square$$

Definition 3.1.1 *A bivariate random vector (X,Y) satisfies the generalized Marshall-Olkin model if its distribution function is given by some*

$$F_\lambda(x,y) = 1_{\mathbb{R}_+^2}(x,y)\exp[-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x,\mu y)] \tag{3.1}$$

where $\lambda = (\lambda_1,\lambda_2,\lambda_3,\mu) \in \Lambda = \mathbb{R}_+^4$.

The joint density function for eqn. 3.1 is

$$f(x,y) = \begin{cases} \lambda_1(\lambda_2+\mu\lambda_3)\overline{F}_\lambda(x,y)\,; \ x < \mu y\,; \, x\,,y > 0 \\ \lambda_2(\lambda_1+\lambda_3)\overline{F}_\lambda(x,y)\,; \ x > \mu y\,; \, x\,,y > 0 \\ \frac{\lambda_3}{\sqrt{1+\mu^2}}\exp\{-(\lambda_1+\lambda_2/\mu+\lambda_3)x\}\,; \ x = \mu y\,; \, x\,,y > 0 \end{cases} \tag{3.2}$$

Remark 3.1.1 *There is a simple interpretation of this model: Looking at implants, once the side represented by Y survives the other side, the survival time has a different rate only at some later time (if $\mu < 1$).*

Lemma 3.1.2 *If (X,Y) is distributed as the generalized Marshall-Olkin model of eqn. 3.1 then*

1. $P(X < \mu Y) = \frac{\mu\lambda_1}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)}$
2. $P(\mu Y < X) = \frac{\lambda_2}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)}$
3. $P(X = \mu Y) = \frac{\mu\lambda_3}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)}$
4. $P(\min(X,\mu Y) \geq t) = \exp\{-(\lambda_1+\lambda_2/\mu+\lambda_3)t\}$

Proof. The proof of point No.1 is as follows:

$$\begin{aligned}
P(X < \mu Y) &= \lambda_1(\lambda_2+\mu\lambda_3)\int_{y=0}^{\infty}\int_{x=0}^{\mu y}\exp\{-\lambda_1 x\}\exp\{-(\lambda_2+\mu\lambda_3)y\}\,dx\,dy \\
&= \lambda_1(\lambda_2+\mu\lambda_3)\int_{y=0}^{\infty}\exp\{-(\lambda_2+\mu\lambda_3)y\}\,dy\int_{x=0}^{\mu y}\exp\{-\lambda_1 x\}dx \\
&= (\lambda_2+\mu\lambda_3)\int_{y=0}^{\infty}\exp\{-(\lambda_2+\mu\lambda_3)y\}\,[1-\exp(-\lambda_1\mu y)]\,dy \\
&= (\lambda_2+\mu\lambda_3)\Big[\int_{y=0}^{\infty}\exp\{-(\lambda_2+\mu\lambda_3)y\}\,dy - \int_{y=0}^{\infty}\exp\{-(\mu\lambda_1+\lambda_2+\mu\lambda_3)y\}\,dy\Big] \\
&= \frac{\mu\lambda_1}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)}.
\end{aligned}$$

The proof of point No.2 is as follows:

$$\begin{aligned}
P(\mu Y < X) &= P(Y < \mu^{-1}X) \\
&= \lambda_2(\lambda_1+\lambda_3)\int_{x=0}^{\infty}\int_{y=0}^{x/\mu}\exp\{-\lambda_2 y\}\exp\{-(\lambda_1+\lambda_3)x\}\,dy\,dx \\
&= \lambda_2(\lambda_1+\lambda_3)\int_{x=0}^{\infty}\exp\{-(\lambda_1+\lambda_3)x\}\,dx\int_{y=0}^{x/\mu}\exp\{-\lambda_2 y\}dy \\
&= (\lambda_1+\lambda_3)\int_{x=0}^{\infty}\exp\{-(\lambda_1+\lambda_3)x\}\,[1-\exp(-\lambda_2\,x/\mu)]\,dx \\
&= (\lambda_1+\lambda_3)\Big[\int_{y=0}^{\infty}\exp\{-(\lambda_1+\lambda_3)x\}\,dx - \int_{x=0}^{\infty}\exp\{-(\lambda_1+\lambda_2\,x/\mu+\lambda_3)x\}\,dx\Big] \\
&= \frac{\lambda_2}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)}.
\end{aligned}$$

The proof of point No.3 is as follows:

$$\begin{aligned}
P(X = \mu Y) &= 1 - P(\mu Y < X) - P(X < \mu Y) \\
&= \frac{\mu\lambda_3}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)}.
\end{aligned}$$

To proof point No.4 we have that

$$\begin{aligned}
P(\min(X,\mu Y)\geq t) &= P(X\geq t\,,\,Y\geq\frac{t}{\mu}) = F(t,\frac{t}{\mu}) \\
&= \exp\{-(\lambda_1+\lambda_2/\mu+\lambda_3)t\} \qquad \square
\end{aligned}$$

This implies that $\min(X,\mu Y)$ is distributed according to the exponential distribution with parameter $\{\lambda_1+\lambda_2/\mu+\lambda_3\}$. Then

$$E(\min(X,\mu Y)) = \frac{\mu}{(\lambda_1\mu+\lambda_2+\lambda_3\mu)}. \qquad \square$$

3.2 Maximum Likelihood Estimation

This section is devoted to the derivation of the maximum likelihood estimation for the parameters of the generalized Marshall-Olkin model.

To begin with notice the following fact. If a sequence $(X_1, Y_1), ..., (X_n, Y_n)$ of independent identically distributed random vectors from the distribution F_λ is observed, then

$$n_3 = \sharp\{1 \le i \le n : X_i = \mu Y_i\} \ge 2$$

holds for at most one parameter μ, and as $n \to \infty$,

$$P(n_3 \ge 2) \to 1.$$

This suggests the estimator for μ:

$$\widehat{\mu} = \frac{X_i}{2Y_i} + \frac{X_j}{2Y_j},$$

where i and j are chosen to satisfy

$$\left|\frac{X_i}{Y_i} - \frac{X_j}{Y_j}\right| = \min\left\{\left|\frac{X_l}{Y_l} - \frac{X_k}{Y_k}\right| : 1 \le l < k \le n\right\}.$$

Note that $\widehat{\mu} = \mu$ for large n almost surely.

We let $n_1 = \sharp\{i : X_i < \mu Y_i\}$, $n_2 = \sharp\{i : \mu Y_i < X_i\}$ and $n_3 = \sharp\{i : X_i = \mu Y_i\}$, whence $n_1 + n_2 + n_3 = n$. Also, let $Z = (X, Y)$. The conditional likelihood function for the generalized Marshall-Olkin model for a random sample of size n of pairs $z_i = (x_i, y_i)$ for $1 \le i \le n$, conditioned that μ is fixed and $\underline{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$, is given by

$$\begin{aligned} l(\underline{\lambda}) &= \prod_{i=1}^{n} f(x_i, y_i) \\ &= \exp\{-\lambda_1 \sum_{i=1}^{n} x_i - \lambda_2 \sum_{i=1}^{n} y_i - \lambda_3 \sum_{i=1}^{n} \max(x_i, \mu y_i)\} \\ &\quad [\lambda_2(\lambda_1 + \lambda_3)]^{n_2} [\lambda_1(\lambda_2 + \mu\lambda_3)]^{n_1} \left(\frac{\lambda_3}{\sqrt{1+\mu^2}}\right)^{n_3} \end{aligned} \tag{3.3}$$

the log likelihood is

$$\begin{aligned} \mathbf{l}(\underline{\lambda}) &= \log l(\underline{\lambda}) \\ &= -\lambda_1 \sum_{i=1}^{n} x_i - \lambda_2 \sum_{i=1}^{n} y_i - \lambda_3 \sum_{i=1}^{n} \max(x_i, \mu y_i) + n_2 \log[\lambda_2(\lambda_1 + \lambda_3)] \\ &+ n_1 \log[\lambda_1(\lambda_2 + \mu\lambda_3)] + n_3 \log \frac{\lambda_3}{\sqrt{1+\mu^2}}. \end{aligned} \tag{3.4}$$

For brevity's sake we write $\mathbf{l}$ for $\mathbf{l}(\underline{\lambda})$. The partial derivative of $\mathbf{l}$ with respect to λ_1 is given by

$$\partial\mathbf{l}/\partial\lambda_1 = n_1/\lambda_1 + n_2/(\lambda_1+\lambda_3) - \sum_{i=1}^{n} x_i$$

setting $\partial\mathbf{l}/\partial\lambda_1 = 0$ gives the following likelihood equation

$$n_1/\lambda_1 + n_2/(\lambda_1+\lambda_3) = \sum_{i=1}^{n} x_i$$

The partial derivative of $\mathbf{l}$ with respect to λ_2 is given as

$$\partial\mathbf{l}/\partial\lambda_2 = n_2/\lambda_2 + n_1/(\lambda_2+\mu\lambda_3) - \sum_{i=1}^{n} y_i$$

setting $\partial\mathbf{l}/\partial\lambda_2 = 0$ gives the following likelihood equation

$$n_2/\lambda_2 + n_1/(\lambda_2+\mu\lambda_3) = \sum_{i=1}^{n} y_i$$

The partial derivative of $\mathbf{l}$ with respect to λ_3 is given as

$$\partial\mathbf{l}/\partial\lambda_3 = n_3/\lambda_3 + n_2/(\lambda_1+\lambda_3) + \mu n_1/(\lambda_2+\mu\lambda_3) - \sum_{i=1}^{n} \max(x_i, \mu y_i)$$

setting $\partial\mathbf{l}/\partial\lambda_3 = 0$ gives the following likelihood equation

$$n_3/\lambda_3 + n_2/(\lambda_1+\lambda_3) + \mu n_1/(\lambda_2+\mu\lambda_3) = \sum_{i=1}^{n} \max(x_i, \mu y_i)$$

Definition 3.2.1 *Given μ, any value $\hat{\underline{\lambda}}(z)$ that maximizes the likelihood function*

$$l(z \mid \hat{\underline{\lambda}}(z)) = \sup_{\underline{\lambda}\in\Lambda^+} l(z \mid \underline{\lambda})$$

is called a maximum likelihood estimate (m.l.e) of $\underline{\lambda}$.

Hence, the likelihood equations are

$$\left\{\begin{array}{l} n_1/\lambda_1 + n_2/(\lambda_1+\lambda_3) = \sum_{i=1}^{n} x_i \\ n_2/\lambda_2 + n_1/(\lambda_2+\mu\lambda_3) = \sum_{i=1}^{n} y_i \\ n_3/\lambda_3 + n_2/(\lambda_1+\lambda_3) + \mu n_1/(\lambda_2+\mu\lambda_3) = \sum_{i=1}^{n} \max(x_i, \mu y_i) \end{array}\right\} \tag{3.5}$$

The second partial derivatives of the log likelihood (Hessian matrix) on Λ is given by

$$\begin{aligned} Q &= \nabla^2\mathbf{l}(\underline{\lambda}) = (\partial^2\mathbf{l}/\partial\lambda_i\partial\lambda_j)_{i,j=1,2,3} \qquad (3.6) \\ &= -\begin{pmatrix} \frac{n_2}{(\lambda_1+\lambda_3)^2} + \frac{n_1}{\lambda_1^2} & 0 & \frac{n_2}{(\lambda_1+\lambda_3)^2} \\ 0 & \frac{n_1}{(\lambda_2+\mu\lambda_3)^2} + \frac{n_2}{\lambda_2^2} & \frac{\mu n_1}{(\lambda_2+\mu\lambda_3)^2} \\ \frac{n_2}{(\lambda_1+\lambda_3)^2} & \frac{\mu n_1}{(\lambda_2+\mu\lambda_3)^2} & \frac{n_3}{\lambda_3^2} + \frac{n_2}{(\lambda_1+\lambda_3)^2} + \frac{\mu^2 n_1}{(\lambda_2+\mu\lambda_3)^2} \end{pmatrix} \end{aligned}$$

Letting

$$C = \begin{pmatrix} n_1/\lambda_1^2 & 0 & 0 \\ 0 & n_2/\lambda_2^2 & 0 \\ 0 & 0 & n_3/\lambda_3^2 \end{pmatrix} \tag{3.7}$$

and

$$\begin{aligned} D[a,b] &= \begin{pmatrix} a & 0 & a \\ 0 & b & \mu b \\ a & \mu b & a+\mu^2 b \end{pmatrix} \\ &= D[\,n_2/(\lambda_1+\lambda_3)^2\,,\, n_1/(\lambda_2+\mu\lambda_3)^2] \end{aligned} \tag{3.8}$$

then

$$Q = -\{C + D[\,n_2/(\lambda_1+\lambda_3)^2\,,\, n_1/(\lambda_2+\mu\lambda_3)^2]\}.$$

The existence and uniqueness properties of MLE are given in the following theorem.

Theorem 3.2.1 *Let $(X_i, Y_i); i = 1, \ldots, n$, $n \in \mathbb{N}$ be independent identically distributed (i.i.d.) sequence with cumulative distribution function (c.d.f) F_λ given by eqn. 3.1, the generalized Marshall-Olkin model, conditioned on μ with parameters $\underline{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in \Lambda$ cf. eqn. 3.1. Let $\mathbb{R} = \{(x_i, y_i) \ldots, (x_n, y_n) : \exists i : x_i < y_i, \exists j : y_j < x_i, \exists k : y_k = x_i\} \subset \mathbb{R}^{2n}$. Then for all $(x_i, y_i); i = 1, \ldots, n;\ n \in \mathbb{R}$. The MLE $\widehat{\lambda}$ for λ exists and is uniquely determined by eqn. 3.5.*

Proof of Thm. 3.2.1 Let $l(\underline{\lambda})$ denote the likelihood function of the generalized Marshall-Olkin model as derived in eqn. 3.3. The Hessian Q of the likelihood is given as eqn. 3.6 where n_1, n_2 and n_3 denotes $\sharp\{i : X_i < \mu Y_i\}$, $\sharp\{i : \mu Y_i < X_i\}$ and $\sharp\{i : X_i = \mu Y_i\}$ respectively. The likelihood function $l(\underline{\lambda})$ is twice differentiable. The negative Hessian $(-Q)$ can be written as $C + D$ where C and D are defined in eqn. 3.7 and eqn. 3.8. resp. The matrix C is positive definite because:

1. The first entry $n_1/\lambda_1^2 > 0$.

2. The determinant of the matrix
$$\begin{pmatrix} n_1/\lambda_1^2 & 0 \\ 0 & n_2/\lambda_2^2 \end{pmatrix} > 0.$$

3. The determinant of the whole matrix
$$\begin{pmatrix} n_1/\lambda_1^2 & 0 & 0 \\ 0 & n_2/\lambda_2^2 & 0 \\ 0 & 0 & n_3/\lambda_3^2 \end{pmatrix} > 0$$

The matrix D is positive semi-definite because:

1. The first entry $n_2/(\lambda_1+\lambda_3)^2 > 0$

2. The determinant of the matrix

$$\begin{pmatrix} n_2/(\lambda_1+\lambda_3)^2 & 0 \\ 0 & n_1/(\lambda_2+\mu\lambda_3)^2 \end{pmatrix} > 0$$

   and

3. The determinant of the whole matrix

$$= \begin{pmatrix} n_2/(\lambda_1+\lambda_3)^2 & 0 & n_2/(\lambda_1+\lambda_3)^2 \\ 0 & n_1/(\lambda_2+\mu\lambda_3)^2 & \mu n_1/(\lambda_2+\mu\lambda_3)^2 \\ n_2/(\lambda_1+\lambda_3)^2 & \mu n_1/(\lambda_2+\mu\lambda_3)^2 & n_2/(\lambda_1+\lambda_3)^2+\mu^2 n_1/(\lambda_2+\mu\lambda_3)^2 \end{pmatrix} = 0$$

and for any vector θ, we have

$$\begin{aligned} \theta^T(-Q)\theta &= \theta^T\,C\,\theta+\theta^T\,D\,\theta \\ &= \;>0+\;\geq\;0 \end{aligned}$$

In order to show that this inequality is strict, for any $\theta\neq 0$ write

$$\theta^T(-Q)\theta=\theta^T\,C\,\theta+\theta^T\,D\,\theta.$$

If $\theta=(\theta_1,\theta_2,\theta_3)^T$ with θ_1 or $\theta_2\neq 0$ then strict positivity from

$$\theta^T\,C\,\theta=\frac{n_1}{\lambda_1^2}\,\theta_1^2+\frac{n_2}{\lambda_2^2}\,\theta_2^2\;>\;0$$

and positive semi-definiteness of D. If $\theta=(\theta_1,\theta_2,\theta_3)^T$ and $\theta_1=\theta_2=0$ and $\theta_3\neq 0$, then

$$\theta^T\,C\,\theta=\frac{n_3}{\lambda_3^2}\,\theta_3^2\;>\;0$$

and

$$\theta^T\,D\,\theta=\Big(\frac{n_2}{(\lambda_1+\lambda_3)^2}+\frac{\mu^2 n_1}{(\lambda_2+\mu\lambda_3)^2}\Big)\,\theta_3^2\;>\;0. \qquad \square$$

So $-Q$ is positive definite, hence Q is negative definite. By Thms. 3.2 and 4.2 of Mangarsin (1969, pages 89 and 91) the likelihood function $l(\underline{\lambda})$ is strictly concave on Λ. Hence $l(\underline{\lambda})$ must have a unique maximum on Λ given by the roots of the $\nabla\mathbf{l}(\underline{\lambda})=0$.

Definition 3.2.2 *A statistic $\Im=\Im(Z_1,Z_2,\ldots,Z_n)$ is said to be sufficient for a parameter $\underline{\lambda}\in\Lambda$ if conditional probability function*

$$P\{Z_1=z_1,Z_2=z_2,\ldots,Z_n=z_n\,|\,\Im(Z_1,Z_2,\ldots,Z_n)=v\}$$

does not depend on $\underline{\lambda}$.

To find the sufficient statistic for the exponential distribution given by eqn.3.1, we use the general factorization theorem for sufficiency.

Theorem 3.2.2 {The General Factorization Theorem for Sufficiency} *Let $P_{\underline{\lambda}}$ be a family of probability measures and let $P_{\underline{\lambda}}$ admit a probability density $p_{\underline{\lambda}} = dP_{\underline{\lambda}}/d\psi$ with respect to a σ- finite measure ψ. Then $\Im$ is sufficient for $P_{\underline{\lambda}}$ (or simply), if and only if there exist non-negative measurable functions $g_{\underline{\lambda}}[\Im(Z_1, Z_2, \ldots, Z_n)]$ and $h(Z_1, Z_2, \ldots, Z_n)$ such that*

$$P_{\underline{\lambda}}(Z_1, Z_2, \ldots, Z_n) = g_{\underline{\lambda}}[\Im(Z_1, Z_2, \ldots, Z_n)]h(Z_1, Z_2, \ldots, Z_n)$$

Using the above theorem, we write likelihood function in the form

$$\begin{aligned} f_{\underline{\lambda}}(x_i, y_i) &= h(X,Y)g_{\underline{\lambda}}(\Im(X,Y)) \\ &= [\lambda_2(\lambda_1+\lambda_3)]^{n_2}[\lambda_1(\lambda_2+\mu\lambda_3)]^{n_1}\left(\frac{\lambda_3}{\sqrt{1+\mu^2}}\right)^{n_3} \\ &\quad \exp\{-\lambda_1\sum_{i=1}^{n} x_i - \lambda_2\sum_{i=1}^{n} y_i - \lambda_3\sum_{i=1}^{n}\max(x_i, \mu y_i)\} \end{aligned}$$

where

$$h(X,Y) = [\lambda_2(\lambda_1+\lambda_3)]^{n_2}[\lambda_1(\lambda_2+\mu\lambda_3)]^{n_1}\left(\frac{\lambda_3}{\sqrt{1+\mu^2}}\right)^{n_3}$$

and

$$g_{\underline{\lambda}}(\Im(X,Y)) = \exp\{-\lambda_1\sum_{i=1}^{n} x_i - \lambda_2\sum_{i=1}^{n} y_i - \lambda_3\sum_{i=1}^{n}\max(x_i, \mu y_i)\}$$

Hence,

$$\{n_1\,,\, n_2\,,\, \sum_{i=1}^{n} x_i\,,\, \sum_{i=1}^{n} y_i\,,\, \sum_{i=1}^{n}\max(x_i, \mu y_i)\}$$

or equivalently,

$$\{n_1\,,\, n_2\,,\, \sum_{i=1}^{n}\max(x_i, \mu y_i)\,,\, \sum_{i=1}^{n}\min(x_i, \mu y_i)\,,\, \sum_{i=1}^{n}\max(x_i, \mu y_i) - \sum_{i=1}^{n}\min(x_i, \mu y_i)\}$$

are jointly sufficient statistics. Bemis et. al.(1972)and Bhattacharyya and Johnson (1971) obtained the similar results for Marshall-Olkin BVE using factorization criterion but not for the case of μ.

3.3 Asymptotic Properties

This section is devoted to the study of the asymptotic properties of the MLE for the parameters of the generalized Marshall-Olkin model given μ.

3.3.1 Consistency

Theorem 3.3.1 *For every $n \in \mathbb{N}$, let $Z_i^n = (X_i^n, Y_i^n); i = 1, \ldots, n$ be i.i.d sequence with c.d.f. F_λ given by eqn. 3.1, the generalized Marshall-Olkin model , with parameters $\underline{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in \Lambda$ cf. eqn. 3.1. Let $\hat{\underline{\lambda}}_n$ denote the MLE for $\underline{\lambda}_n$ based on $Z_1^n, \ldots, Z_n^n; n \in \mathbb{N}$. Then for each $\lambda \in \Lambda$ and for a fixed μ, $\hat{\lambda}_n$ converges stochastically to λ under the law F_λ. The consequence the MLE is consistent. For $n \to \infty$*

$$\hat{\lambda}_n \xrightarrow{p} \lambda$$

that is

$$\lim_{n\to\infty} P_\lambda\Big(|\hat{\underline{\lambda}}_n - \underline{\lambda}| \le \epsilon\Big) = 1$$

Proof Let $l(\underline{\lambda})$ denote the likelihood function of the generalized Marshall-Olkin model as derived in eqn. 3.1. We are considering a set

$$C_\delta = \{\underline{\lambda} \in \Lambda : \|\underline{\lambda} - \tilde{\underline{\lambda}}\| \le \delta\}.$$

where $\delta > 0$ and $\tilde{\underline{\lambda}} = \lambda$ in Thm. 3.3.1 is fixed. Let the notation $\wp C_\delta$ denotes the boundary of C_δ. We want to show that $\forall\, \delta > 0$

$$\lim_{n\to\infty} P_{\tilde{\underline{\lambda}}}\Big(l_n(\underline{\lambda}) < l_n(\tilde{\underline{\lambda}}); \forall\, \lambda \in \wp C_\delta\Big) = 1 \Rightarrow \lim_{n\to\infty} P_{\tilde{\underline{\lambda}}}\Big(|\hat{\underline{\lambda}}_n - \tilde{\underline{\lambda}}| \le \delta\Big) = 1$$

Let

$$A_n = \{\omega \in \Omega : \ l_n(\lambda, z(\omega)) < l_n(\tilde{\lambda}) \ : \forall\, \lambda \in \wp C_\delta\}$$
$$B_n = \{|\hat{\underline{\lambda}}_n - \tilde{\underline{\lambda}}| \le \delta\}$$

Known

$$\lim_{n\to\infty} P_{\tilde{\underline{\lambda}}}(A_n) = 1\,, \text{claim} \lim_{n\to\infty} P_{\tilde{\underline{\lambda}}}(B_n) = 1$$

if we would know that $A_n \subset B_n$ then

$$P_{\tilde{\underline{\lambda}}}(A_n) \le P_{\tilde{\underline{\lambda}}}(B_n)$$

so

$$1 \ge \lim_{n\to\infty} P_{\tilde{\underline{\lambda}}}(B_n) \ge \lim_{n\to\infty} P_{\tilde{\underline{\lambda}}}(A_n) = 1$$

To show that $A_n \subset B_n$, let $\omega \in A_n$, this implies

$$l_n(\lambda; z_1(\omega), \ldots, z_n(\omega)) < l_n(\tilde{\lambda}|z_1(\omega), \ldots, z_n(\omega)) \ : \forall\, \lambda \in \wp C_\delta)$$

this implies, the maximum of $l_n(\lambda; z(\omega))$, is attained in the interior of $C_\delta \Rightarrow l_n(.; z(\omega))$ has a zero for l' in C_δ. l' has only one zero implies MLE lies in C_δ implies $\hat{\lambda} \in C_\delta$, implies

$$|\hat{\lambda}_n - \tilde{\lambda}| \le \delta \Rightarrow \omega \in B_n$$

The next point is to proof that $\lim_{n\to\infty} P_{\tilde{\underline{\lambda}}}\Big(l_n(\underline{\lambda}) \;<\; l_n(\tilde{\underline{\lambda}}); \forall\, \lambda \,\in \wp C_\delta\Big) = 1$ we proof this by first showing that

$$\lim_{n\to\infty} \frac{1}{n}(\underline{\lambda} - \tilde{\underline{\lambda}})^T \dot{\mathbf{l}}(\lambda)_{\lambda=\tilde{\lambda}} \;=\; E_{\tilde{\lambda}}[\dot{\mathbf{l}}]_{\lambda=\tilde{\lambda}} = 0$$

where $\dot{\mathbf{l}} = \frac{\partial \log f}{\partial \lambda}$ and that

$$\lim_{n\to\infty} \frac{1}{n}(\underline{\lambda} - \tilde{\underline{\lambda}})^T \ddot{\mathbf{l}}(\lambda)(\underline{\lambda} - \tilde{\underline{\lambda}})_{\lambda=\tilde{\lambda}} \;=\; E_{\tilde{\lambda}}[\ddot{\mathbf{l}}]_{\lambda=\tilde{\lambda}} < 0$$

The partial derivative of $\mathbf{l}$ with respect to λ_1 is given as

$$\partial \mathbf{l}/\partial\lambda_1 = n_1/\lambda_1 + n_2/(\lambda_1+\lambda_3) - \sum_{i=1}^{n} x_i$$

dividing both sides by n, we obtain

$$\begin{aligned}
\frac{1}{n}\frac{\partial \mathbf{l}}{\partial\lambda_1} &= \frac{1}{n}(n_1/\lambda_1 + n_2/(\lambda_1+\lambda_3) - \sum_{i=1}^{n} x_i) \\
&= \frac{n_1}{n\lambda_1} + \frac{n_2}{n(\lambda_1+\lambda_3)} - \frac{\sum_{i=1}^{n} x_i}{n} \\
\lim_{n\to\infty}\frac{1}{n}\frac{\partial \mathbf{l}}{\partial\lambda_1} &= \frac{1}{\lambda_1}P(\mu Y \;>\; X\,) + \frac{1}{(\lambda_1+\lambda_3)}P(\,X \;>\; \mu Y\,) \;-\; E(X) \\
&= \frac{1}{\lambda_1}\,\frac{\mu\lambda_1}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)} + \frac{1}{(\lambda_1+\lambda_3)}\,\frac{\lambda_2}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)} - \frac{1}{(\lambda_1+\lambda_3)} \\
&= 0
\end{aligned}$$

The partial derivative of $\mathbf{l}$ with respect to λ_2 is given as

$$\partial \mathbf{l}/\partial\lambda_2 = n_2/\lambda_2 + n_1/(\lambda_2+\mu\lambda_3) - \sum_{i=1}^{n} y_i$$

dividing both sides by n, we obtain

$$\begin{aligned}
\frac{1}{n}\frac{\partial \mathbf{l}}{\partial\lambda_2} &= \frac{1}{n}(n_2/\lambda_2 + n_1/(\lambda_2+\mu\lambda_3) - \sum_{i=1}^{n} y_i) \\
&= \frac{n_2}{n\lambda_2} + \frac{n_1}{n(\lambda_2+\mu\lambda_3)} - \frac{\sum_{i=1}^{n} y_i}{n} \\
\lim_{n\to\infty}\frac{1}{n}\frac{\partial \mathbf{l}}{\partial\lambda_2} &= \frac{1}{\lambda_2}P(\mu Y \;<\; X\,) + \frac{1}{(\lambda_2+\mu\lambda_3)}P(\,X \;<\; \mu Y\,) \;-\; E(Y) \\
&= \frac{1}{\lambda_2}\,\frac{\lambda_2}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)} + \frac{1}{(\lambda_2+\mu\lambda_3)}\,\frac{\mu\lambda_1}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)} - \frac{1}{(\lambda_2+\mu\lambda_3)} \\
&= 0
\end{aligned}$$

The partial derivative of $\mathbf{l}$ with respect to λ_3 is given as

$$\partial \mathbf{l}/\partial\lambda_3 = n_3/\lambda_3 + n_2/(\lambda_1+\lambda_3) + \mu n_1/(\lambda_2+\mu\lambda_3) - \sum_{i=1}^{n} \max(x_i, \mu y_i)$$

dividing both sides by n, we obtain

$$\begin{aligned}
\frac{1}{n}\frac{\partial \mathbf{l}}{\partial \lambda_3} &= \frac{1}{\lambda_3}(\frac{n_3}{n}) + \frac{1}{(\lambda_1+\lambda_3)}(\frac{n_2}{n}) + \frac{\mu}{(\lambda_2+\mu\lambda_3)}(\frac{n_1}{n}) - (\frac{\sum_{i=1}^{n}\max(x_i,\mu y_i)}{n}) \\
\lim_{n\to\infty}\frac{1}{n}\frac{\partial \mathbf{l}}{\partial \lambda_3} &= \frac{1}{\lambda_3}P(\,X\,=\,\mu Y\,) + \frac{1}{(\lambda_1+\lambda_3)}P(\,X\,>\,\mu Y\,) + \frac{\mu}{(\lambda_2+\mu\lambda_3)}P(\,X\,<\,\mu Y\,) \\
&- \;E(X)\,-\,E(\,\mu Y\,)\,-\,E(\min(X,\mu Y)) \\
&= (\frac{\mu}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)}) + \frac{1}{(\lambda_1+\lambda_3)}(\frac{\lambda_2}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)}) + (\frac{\mu}{(\lambda_2+\mu\lambda_3)}) \\
&\times \frac{\mu\lambda_1}{(\mu\lambda_1+\lambda_2+\mu\lambda_3)} - (\frac{1}{(\lambda_1+\lambda_3)}) - (\frac{\mu}{(\lambda_2+\mu\lambda_3)}) - \frac{\mu}{(\lambda_1\mu+\lambda_2+\lambda_3\mu)} \\
&= 0
\end{aligned}$$

this shows that $E_{\tilde{\lambda}}[\dot{\mathbf{l}}]_{\lambda=\tilde{\lambda}} = 0$.

Now, by taking a third order Taylor expansion around $\underset{\sim}{\tilde{\lambda}}$, we have

$$\begin{aligned}
\frac{1}{n}(\mathbf{l}(\underset{\sim}{\lambda}) - \mathbf{l}(\underset{\sim}{\tilde{\lambda}})) &= \frac{1}{n}(\underset{\sim}{\lambda} - \underset{\sim}{\tilde{\lambda}})^T\dot{\mathbf{l}}(\underset{\sim}{\tilde{\lambda}}) \\
&+ \frac{1}{2}(\underset{\sim}{\lambda} - \underset{\sim}{\tilde{\lambda}})^T\Big(\frac{1}{n}\ddot{\mathbf{l}}(\underset{\sim}{\tilde{\lambda}})\Big)(\underset{\sim}{\lambda} - \underset{\sim}{\tilde{\lambda}}) \\
&+ \frac{1}{6}\frac{1}{n}\sum_{r=1}^{k}\sum_{s=1}^{k}\sum_{t=1}^{k}(\underset{\sim}{\lambda}_r - \underset{\sim}{\tilde{\lambda}}_r)(\underset{\sim}{\lambda}_s - \underset{\sim}{\tilde{\lambda}}_s)(\underset{\sim}{\lambda}_t - \underset{\sim}{\tilde{\lambda}}_t)\Big\{\gamma_{rst}(Z_i)H_{rst}(Z_i)\Big\} \\
&= S_1 + S_2 + S_3
\end{aligned}$$

where

$$S_1 = \frac{1}{n}(\underset{\sim}{\lambda} - \underset{\sim}{\tilde{\lambda}})^T\dot{\mathbf{l}}(\underset{\sim}{\tilde{\lambda}})$$

$$S_2 = \frac{1}{2}(\underset{\sim}{\lambda} - \underset{\sim}{\tilde{\lambda}})^T\Big(\frac{1}{n}\ddot{\mathbf{l}}(\underset{\sim}{\tilde{\lambda}})\Big)(\underset{\sim}{\lambda} - \underset{\sim}{\tilde{\lambda}})$$

$$S_3 = \frac{1}{6}\frac{1}{n}\sum_{r=1}^{k}\sum_{s=1}^{k}\sum_{t=1}^{k}(\underset{\sim}{\lambda}_r - \underset{\sim}{\tilde{\lambda}}_r)(\underset{\sim}{\lambda}_s - \underset{\sim}{\tilde{\lambda}}_s)(\underset{\sim}{\lambda}_t - \underset{\sim}{\tilde{\lambda}}_t)\Big\{\gamma_{rst}(Z_i)H_{rst}(Z_i)\Big\}$$

we make assumptions that $0 \le |\gamma_{rst}(z)| < 1$ and $|\frac{\partial^3 \log f}{\partial\lambda_r\partial\lambda_s\partial\lambda_t}| < H_{rst}(Z_i)$. We have seen that

$$S_1 \xrightarrow{p} 0$$

The Hessian Q is negative definite so the second term S_2 is negative with probability tending to 1. S_1 and S_3 are small compared to S_2 so the

$$\sup_{\underset{\sim}{\lambda}\in C_\delta}(S_1 + S_2 + S_3) < 0$$

Thus, for n large enough,

$$\frac{1}{n}(\mathbf{l}(\underset{\sim}{\lambda}) - \mathbf{l}(\underset{\sim}{\tilde{\lambda}})) < 0 \qquad \square$$

this completes the proof.

3.3.2 Asymptotic Normality

Theorem 3.3.2 *Let $Z_n = (X_n, Y_n); (n \geq 1)$ be independent identically distributed (i.i.d.) random variables with cumulative distribution function (c.d.f) F_λ given by eqn. 3.1, the generalized Marshall-Olkin model , with parameters $\underline{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in \Lambda$. Then for a fixed μ, the MLE $\widehat{\lambda}$ for λ is asymptotically normal*

$$\mathbb{N}\,(\,\underline{\lambda}\,,\,\Sigma(\underline{\lambda})^{-1}\,)$$

where

$$\Sigma(\underline{\lambda}) = \mathbf{\Pi}\Sigma_0^{-1}.$$

Proof Let $\delta > 0$ and set

$$G_n = \{(z_1, \ldots, z_n) \in \mathbb{R}^{2n} : |\,\underline{\lambda}\, - \,\widehat{\lambda}(z_1, \ldots, z_n)\,| \, < \, \delta$$

If δ is small enough, then for $(z_1, \ldots, z_n) \in G_n$, $\widehat{\lambda}$ is unique. Then , by consistency,

$$P_{\underline{\lambda}}(\,G_n\,) \;\to\; 1 \text{ as } n \;\to\; \infty$$

hence for $\epsilon > 0$ there is $n_0 \in \mathbb{N}$ such that

$$P_{\underline{\lambda}}(\,G_n^c\,) \;<\; \epsilon \;\forall\, n \;\geq\; n_0$$

It follows that for $\;n \;\geq\; n_0$

$$\begin{aligned} &|\,P_{\underline{\lambda}}(\{\omega : \widehat{\lambda}(z_1(\omega), \ldots, z_n(\omega)) \;\leq\; \underline{t}\} \cap \{z_1, \ldots, z_n \in G_n\}) \\ - \quad & P_{\underline{\lambda}}(\{\omega : \widehat{\lambda}(z_1(\omega), \ldots, z_n(\omega)) \leq \underline{t}\})\,| \;\leq\; P_{\underline{\lambda}}(\,G_n^c\,) \;<\; \epsilon \end{aligned}$$

Note that $\underline{t} \in \mathbb{R}^3$ and $\underline{s} \leq \underline{t}$ means $\underline{s}_i \leq \underline{t}_i\,(i = 1, 2, 3)$. The likelihood function l can be expressed as a Taylor series by

$$l'(\lambda) \;=\; l'(\underline{\lambda}) + (\lambda \;-\; \underline{\lambda})^T l''(\underline{\lambda}^*)$$

for some value of $\underline{\lambda}^* \in B(\underline{\lambda}, \delta)$, the δ-ball around $\underline{\lambda}$. So if $z_1, \ldots, z_n \in G_n$, $n \geq n_0$

$$l'(\widehat{\lambda}(z_1, \ldots, z_n\,)) \;=\; 0$$

and

$$l'(z_1, \ldots, z_n\,, \underline{\lambda}) \;=\; -\,(\widehat{\lambda}(z_1, \ldots, z_n\,) - \underline{\lambda})l''(z_1, \ldots, z_n\,, \,\underline{\lambda}^*)$$

by eqn. (3.4)

$$\begin{aligned} l(z_1, \ldots, z_n\,, \underline{\lambda}) \quad &= \quad \log\; l(z_1, \ldots, z_n\,, \underline{\lambda}) \\ &= \quad -\lambda_1 \sum_{i=1}^{n} x_i - \lambda_2 \sum_{i=1}^{n} y_i - \lambda_3 \sum_{i=1}^{n} \max(x_i, \mu y_i) + n_2 \log[\lambda_2(\lambda_1 + \lambda_3)] \\ &+ \quad n_1 \log[\lambda_1(\lambda_2 + \mu\lambda_3)] + n_3 \log \frac{\lambda_3}{\sqrt{1 + \mu^2}} \end{aligned}$$

so

$$
\begin{aligned}
l'(z_1,\ldots,z_n\,,\underline{\lambda}) &= \begin{pmatrix} \partial \mathbf{l}/\partial\lambda_1 \\ \partial \mathbf{l}/\partial\lambda_2 \\ \partial \mathbf{l}/\partial\lambda_3 \end{pmatrix} \\
&= \left\{ \begin{array}{l} n_1/\lambda_1 + n_2/(\lambda_1+\lambda_3) - \sum_{i=1}^n x_i \\ n_2/\lambda_2 + n_1/(\lambda_2+\mu\lambda_3) - \sum_{i=1}^n y_i \\ n_3/\lambda_3 + n_2/(\lambda_1+\lambda_3) + \mu n_1/(\lambda_2+\mu\lambda_3) - \sum_{i=1}^n \max(x_i,\mu y_i) \end{array} \right\}
\end{aligned}
$$

where $E\,l' = 0$ as calculated in the proof of theorem 3.3.1 (consistency). It follows that $\frac{1}{\sqrt{n}} l'(z_1,\ldots,z_n\,,\,\underline{\lambda})$ is asymptotically normal with expectation $\mathbf{0}$ and covariance matrix $\mathbf{\Pi} = (\pi_{ij})_{1\le i,j\le 3}$. We calculate this matrix as follows:

$$var(X) = 1/(\lambda_1+\lambda_3)^2$$

$$var(Y) = 1/(\lambda_2+\mu\lambda_3)^2$$

$$var(\min(X,\mu Y)) = \mu^2/(\lambda_1\mu+\lambda_2+\lambda_3\mu)^2$$

$$
\begin{aligned}
var(\max(X,\mu Y)) &= var(X+\mu Y - \min(X,\mu Y)) \\
&= var(X) + \mu^2 var(Y) + var(\min(X,\mu Y)) - 2cov(X,\mu Y) \\
&- 2cov(X,\min(X,\mu Y)) - 2cov(Y,\min(X,\mu Y)).
\end{aligned}
$$

Now,

$$cov(X,\mu Y) = \mu E(XY) - \mu E(X)E(Y) = 0$$

$$cov(X,\min(X,\mu Y)) = \int\int x\min(x,\mu y)dF dx dy - E(X)E(\min(X,\mu Y))$$

$$
\begin{aligned}
\int\int x\min(x,\mu y)dF dx dy &= \lambda_1(\lambda_2+\mu y)\int_{y=0}^{\infty}\int_{x=0}^{\mu y} x^2 e^{-\lambda_1 x} e^{-(\lambda_2+\mu\lambda_3)y} dx dy \\
&+ \lambda_2(\lambda_1+\lambda_3)\mu \int_{x=0}^{\infty}\int_{y=0}^{x/\mu} xy e^{-(\lambda_1+\lambda_3)x} e^{-\lambda_2 y} dx dy \\
&+ \lambda_3 \int_{x=0}^{\infty} x^2 e^{-(\lambda_1+\lambda_2/\mu+\lambda_3)x} dx \\
&= \frac{2}{\lambda_1^2} - \frac{2\mu^2(\lambda_2+\mu\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} - \frac{2\mu(\lambda_2+\mu\lambda_3)}{\lambda_1(\lambda_1\mu+\lambda_2+\mu\lambda_3)^2} \\
&- \frac{2(\lambda_2+\mu\lambda_3)}{\lambda_1^2(\lambda_1\mu+\lambda_2+\mu\lambda_3)} + \frac{\mu\lambda_2(\lambda_1+\lambda_3)}{\lambda_2^2(\lambda_1+\lambda_3)^2} - \frac{2\mu^3(\lambda_1+\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} \\
&- \frac{\mu^2(\lambda_1+\lambda_3)}{\lambda_2(\lambda_1\mu+\lambda_2+\mu\lambda_3)^2} + \frac{2\lambda_3\mu^3}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3}
\end{aligned}
$$

hence,

$$\begin{aligned} cov(X, \min(X, \mu Y)) &= \frac{2}{\lambda_1^2} - \frac{2\mu^2(\lambda_2+\mu\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} - \frac{2\mu(\lambda_2+\mu\lambda_3)}{\lambda_1(\lambda_1\mu+\lambda_2+\mu\lambda_3)^2} \\ &- \frac{2(\lambda_2+\mu\lambda_3)}{\lambda_1^2(\lambda_1\mu+\lambda_2+\mu\lambda_3)} + \frac{\mu\lambda_2(\lambda_1+\lambda_3)}{\lambda_2^2(\lambda_1+\lambda_3)^2} - \frac{2\mu^3(\lambda_1+\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} \\ &- \frac{\mu^2(\lambda_1+\lambda_3)}{\lambda_2(\lambda_1\mu+\lambda_2+\mu\lambda_3)^2} + \frac{2\lambda_3\mu^3}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} - \frac{1}{(\lambda_1+\lambda_3)}\frac{\mu}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)} \end{aligned}$$

$$cov(Y, \min(X, \mu Y)) = \int\int y\min(x,\mu y)dFdxdy - E(Y)E(\min(X,\mu Y))$$

$$\begin{aligned} \int\int y\min(x,\mu y)dFdxdy &= \lambda_1(\lambda_2+\mu\lambda)\int_{y=0}^{\infty}\int_{x=0}^{\mu y} xye^{-\lambda_1 x}e^{-(\lambda_2+\mu\lambda_3)y}dxdy \\ &+ \mu\lambda_2(\lambda_1+\lambda_3)\int_{x=0}^{\infty}\int_{y=0}^{x/\mu} y^2e^{-\lambda_2 y}e^{-(\lambda_1+\lambda_3)x}dxdy \\ &+ \mu\lambda_3\int_{y=0}^{\infty} y^2e^{-(\lambda_1+\lambda_2/\mu+\lambda_3)y}dy \\ &= \frac{1}{\lambda_1(\lambda_2+\mu\lambda_3)} - \frac{(\lambda_2+\mu\lambda_3)}{\lambda_1(\lambda_1\mu+\lambda_2+\mu\lambda_3)^2} - \frac{2\mu(\lambda_2+\mu\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} \\ &- \frac{2\mu^2(\lambda_1+\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} - \frac{2\mu^2(\lambda_1+\lambda_3)}{\lambda_2(\lambda_1\mu+\lambda_2+\mu\lambda_3)^2} + \frac{2\mu}{\lambda_2^2} \\ &- \frac{2\mu^2(\lambda_1+\lambda_3)}{\lambda_2^2(\lambda_1\mu+\lambda_2+\mu\lambda_3)} + \frac{2\lambda_3\mu^4}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} \end{aligned}$$

hence,

$$cov(Y, \min(X, \mu Y)) = \frac{1}{\lambda_1(\lambda_2+\mu\lambda_3)} - \frac{(\lambda_2+\mu\lambda_3)}{\lambda_1(\lambda_1\mu+\lambda_2+\mu\lambda_3)^2} - \frac{2\mu(\lambda_2+\mu\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3}$$

$$-\frac{2\mu^2(\lambda_1+\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} - \frac{2\mu^2(\lambda_1+\lambda_3)}{\lambda_2(\lambda_1\mu+\lambda_2+\mu\lambda_3)^2} + \frac{2\mu}{\lambda_2^2}$$

$$-\frac{2\mu^2(\lambda_1+\lambda_3)}{\lambda_2^2(\lambda_1\mu+\lambda_2+\mu\lambda_3)} + \frac{2\lambda_3\mu^4}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} - \frac{1}{(\lambda_2+\mu\lambda_3)}\frac{\mu}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)}$$

hence the $var(\max(X, \mu Y))$ is calculated from the above results. The next is

$$cov(X, \max(X, \mu Y)) = \int\int x\max(x,\mu y)dFdxdy - E(X)E(\max(X,\mu Y))$$

$$\begin{aligned}
\int\int x\max(x,\mu y)dFdxdy &= \lambda_2(\lambda_1+\lambda_3)\int_{x=0}^{\infty}\int_{y=0}^{x/\mu} x^2 e^{-(\lambda_1+\lambda_3)x}e^{-\lambda_2 y}dydx \\
&+ \mu\lambda_1(\lambda_2+\mu\lambda_3)\int_{y=0}^{\infty}\int_{x=0}^{\mu y} xye^{-\lambda_1 x}.e^{-(\lambda_2+\mu\lambda_3)y}dxdy \\
&+ \lambda_3\int_0^{\infty} x^2 e^{-(\lambda_1+\lambda_2/\mu+\lambda_3)x}dx \\
&= \frac{2}{(\lambda_1+\lambda_3)^2} - \frac{2\mu^3(\lambda_1+\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)} - \frac{\mu(\lambda_2+\mu\lambda_3)}{\lambda_1(\lambda_1\mu+\lambda_2+\mu\lambda_3)^2} \\
&- \frac{2\mu^2(\lambda_2+\mu\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} + \frac{\mu}{\lambda_1(\lambda_2+\mu\lambda_3)} + \frac{2\lambda_3\mu^3}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3}
\end{aligned}$$

hence the $cov(X,\max(X,\mu Y))$ is

$$\begin{aligned}
cov(X,\max(X,\mu Y)) = \frac{2}{(\lambda_1+\lambda_3)^2} - \frac{2\mu^3(\lambda_1+\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)} - \frac{\mu(\lambda_2+\mu\lambda_3)}{\lambda_1(\lambda_1\mu+\lambda_2+\mu\lambda_3)^2} \\
-\frac{2\mu^2(\lambda_2+\mu\lambda_3)}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} + \frac{\mu}{\lambda_1(\lambda_2+\mu\lambda_3)} + \frac{2\lambda_3\mu^3}{(\lambda_1\mu+\lambda_2+\mu\lambda_3)^3} \\
-\frac{1}{(\lambda_1+\lambda_3)}\{(\frac{1}{(\lambda_1+\lambda_3)}) + (\frac{\mu}{(\lambda_2+\mu\lambda_3)}) - \frac{\mu}{(\lambda_1\mu+\lambda_2+\lambda_3\mu)}\}
\end{aligned}$$

similarly,

$$\begin{aligned}
cov(Y,\max(X,\mu Y)) = \frac{1}{\lambda_2(\lambda_1+\lambda_3)} - \frac{2\mu^2(\lambda_1+\lambda_3)}{(\lambda_1\mu+\lambda_2+\lambda_3\mu)^2} + \frac{2\mu}{(\lambda_2+\mu\lambda_3)^2} \\
-\frac{2\mu(\lambda_2+\mu\lambda_3)}{(\lambda_1\mu+\lambda_2+\lambda_3\mu)^3} + \frac{2\lambda_3\mu^4}{(\lambda_1\mu+\lambda_2+\lambda_3\mu)^3} \\
-\frac{\mu}{(\lambda_2+\mu\lambda_3)}\{(\frac{1}{(\lambda_1+\lambda_3)}) + (\frac{\mu}{(\lambda_2+\mu\lambda_3)}) - \frac{\mu}{(\lambda_1\mu+\lambda_2+\lambda_3\mu)}\}.
\end{aligned}$$

If we denote the covariance matrix by $\mathbf{\Pi} = (\pi_{ij})_{1\le i,j\le 3}$ then

$$\pi_{11} = var(X)\,,\ \pi_{12} = \pi_{21} = 0\,,\ \pi_{33} = var(\max(X,\mu Y))$$

$$\pi_{22} = var(Y)\,,\ \pi_{13} = cov(X,\max(X,\mu Y))\,,\ \pi_{23} = cov(Y,\max(X,\mu Y))$$

since

$$P(\,(Z_1\ldots Z_n\,)\in G_n\,)\to 1 \text{ as } n\to\infty$$

$\frac{1}{\sqrt{n}}l'(Z_1,\ldots,Z_n\,,\underline{\lambda})$ and $-\frac{1}{\sqrt{n}}(\hat{\lambda}-\underline{\lambda})^T l''(\underline{\lambda}^*)$ are equivalent. Now,

$$\frac{1}{n}l''(Z_1,\ldots,Z_n\,,\underline{\lambda}^*) \;=\; \frac{1}{n}U(Z_1,\ldots,Z_n) \;\xrightarrow{p}\; \Sigma_0$$

using Q from eqn. 3.6 with $\lambda_i = \lambda_i^*$ we work Σ_0 as follows:

$$\begin{aligned}
\tau_{11} = E(-\frac{\partial^2 l}{\partial \lambda_1^{*2}}) &= \lim_{n\to\infty} \frac{1}{n}\frac{\partial^2 l}{\partial \lambda_1^{*2}} \\
&= \frac{1}{(\lambda_1^* + \lambda_3^*)^2}\frac{n_2}{n} + \frac{1}{\lambda_1^{*2}}\frac{n_1}{n} \\
&= \frac{1}{(\lambda_1^* + \lambda_3^*)^2} P(\mu Y < X) + \frac{1}{\lambda_1^{*2}} P(X < \mu Y) \\
&= \frac{\lambda_2^*}{(\lambda_1^* + \lambda_3^*)^2 \Phi} + \frac{\mu}{\lambda_1^* \Phi}
\end{aligned}$$

where $\Phi = (\lambda_1^* \mu + \lambda_2^* + \lambda_3^* \mu)$

$$\begin{aligned}
\tau_{22} = E(-\frac{\partial^2 l}{\partial \lambda_2^{*2}}) &= \lim_{n\to\infty} \frac{1}{n}\frac{\partial^2 l}{\partial \lambda_2^{*2}} \\
&= \frac{1}{(\lambda_2^* + \mu\lambda_3^*)^2}\frac{n_1}{n} + \frac{1}{\lambda_2^{*2}}\frac{n_2}{n} \\
&= \frac{1}{(\lambda_2^* + \mu\lambda_3^*)^2} P(X < \mu Y) + \frac{1}{\lambda_2^{*2}} P(\mu Y < X) \\
&= \frac{\mu\lambda_1^*}{(\lambda_2^* + \mu\lambda_3^*)^2 \Phi} + \frac{1}{\lambda_2^* \Phi}
\end{aligned}$$

$$\begin{aligned}
\tau_{33} = E(-\frac{\partial^2 l}{\partial \lambda_3^{*2}}) &= \lim_{n\to\infty} \frac{1}{n}\frac{\partial^2 l}{\partial \lambda_3^{*2}} \\
&= \frac{1}{(\lambda_1^* + \lambda_3^*)^2}\frac{n_2}{n} + \frac{\mu^2}{(\lambda_2^* + \mu\lambda_3^*)^2}\frac{n_1}{n} + \frac{1}{\lambda_3^{*2}}\frac{n_3}{n} \\
&= \frac{1}{(\lambda_1^* + \lambda_3^*)^2} P(\mu Y < X) + \frac{\mu^2}{(\lambda_2^* + \mu\lambda_3^*)^2} P(X < \mu Y) + \frac{1}{\lambda_3^{*2}} P(\mu Y = X) \\
&= \frac{\lambda_2^*}{(\lambda_1^* + \lambda_3^*)^2 \Phi} + \frac{\mu^3 \lambda_1^*}{(\lambda_2^* + \mu\lambda_3^*)^2 \Phi} + \frac{\mu}{\lambda_3^* \Phi}
\end{aligned}$$

$$\tau_{12} = \tau_{21} = E(-\frac{\partial^2 l}{\partial \lambda_1^* \partial \lambda_2^*}) = 0$$

$$\begin{aligned}
\tau_{13} = E(-\frac{\partial^2 l}{\partial \lambda_1^* \partial \lambda_3^*}) &= \lim_{n\to\infty} \frac{1}{n}\frac{\partial^2 l}{\partial \lambda_1^* \partial \lambda_3^*} \\
&= \frac{1}{(\lambda_1^* + \lambda_3^*)^2}\frac{n_2}{n} \\
&= \frac{1}{(\lambda_1^* + \lambda_3^*)^2} P(\mu Y < X) \\
&= \frac{\lambda_2^*}{(\lambda_1^* + \lambda_3^*)^2 \Phi}
\end{aligned}$$

$$
\begin{aligned}
\tau_{23} = E(-\frac{\partial^2 l}{\partial\lambda_2^*\partial\lambda_3^*}) &= \lim_{n\to\infty}\frac{1}{n}\frac{\partial^2 l}{\partial\lambda_2^*\partial\lambda_3^*} \\
&= \frac{\mu}{(\lambda_2^*+\mu\lambda_3^*)^2}\frac{n_1}{n} \\
&= \frac{\mu}{(\lambda_2^*+\mu\lambda_3^*)^2}P(X<\mu Y) \\
&= \frac{\mu\lambda_1^*}{(\lambda_2^*+\mu\lambda_3^*)^2\Phi}
\end{aligned}
$$

consequently the information matrix has the form

$$
\Sigma_0 = \begin{pmatrix} \tau_{11} & 0 & \tau_{13} \\ 0 & \tau_{22} & \tau_{23} \\ \tau_{31} & \tau_{32} & \tau_{33} \end{pmatrix}
$$

it follows that

$$
-\frac{1}{\sqrt{n}}l'(Z_1,\ldots,Z_n\,,\underline{\lambda})\Sigma_0^{-1}
$$

is equivalent to $\sqrt{n}(\hat{\lambda}-\underline{\lambda})$ hence the claim with

$$
\Sigma(\underline{\lambda}) \;=\; \mathbf{\Pi}\Sigma_0^{-1}. \qquad \square
$$

3.3.3 Asymptotic Efficiency

The efficiency of a method is measured in terms of the ratio of the Cramer-Rao lower bounds of the individual estimates to the sum of the mean squared errors (MSE) of the individual estimates. That is, Eff. $=\mathrm{tr}(I_n^{-1})/\sum \mathrm{MSE}(\hat{\lambda}_i)$. Also the asymptotic relative efficiencies are based on the ratio of the traces of the appropriate asymptotic covariance matrices. We define trace of asymptotic relative efficiency as

$$
\text{tr. ARE} = \mathrm{tr}(I_n^{-1})/\mathrm{tr}\sum \mathrm{MSE}(\hat{\lambda}_i) \tag{3.9}
$$

and

$$
\text{Generalized}\quad \text{ARE} = |(I_n^{-1})|/|\sum \mathrm{MSE}(\hat{\lambda}_i)| \tag{3.10}
$$

Arnold (1968) and Bemis et. al. (1972) have used the criteria eqn. 3.9 while Bhattacharyya and Johnson (1971) used eqn. 3.10 to study the asymptotic efficiencies of BVE estimators relative to the MLE, eqn. 3.9 comparing asymptotic average variances and eqn. 3.10 asymptotic generalized variance. From Arnold(1968, page 850), the information matrix for Marshall-Olkin model is given as

$$\mathrm{I}_n = \tfrac{n}{\lambda}\begin{pmatrix} a+c & 0 & a \\ 0 & b+d & b \\ a & b & a+b+e \end{pmatrix}$$

where $\lambda = \lambda_1 + \lambda_2 + \lambda_3$, $a = \lambda_2(\lambda_1 + \lambda_3)^{-2}$, $b = \lambda_1(\lambda_2 + \lambda_3)^{-2}$, $c = \lambda_1^{-1}$, $d = \lambda_2^{-1}$ and $e = \lambda_3^{-1}$. Let Σ_{mo} denote the dispersion matrix of the limiting normal distribution of Marshall-Olkin BVE estimator, $\hat{\underline{\lambda}}_n^{mo}$, and $\Sigma_{mo} = (I_n)^{-1}$. Let Σ_g denote the asymptotic dispersion matrix of $\hat{\underline{\lambda}}_n$, the MLE of the generalized Marshall-Olkin model, and $\Sigma_g = (\Sigma_0)^{-1}$. We define the trace of asymptotic relative efficiency tr. ARE of Σ_{mo} relative to Σ_g as

$$\text{tr. ARE} = \frac{\text{trace }(\Sigma_g)}{\text{trace }(\Sigma_{mo})}$$

Using the simulated results, the following values were obtained for tr. ARE:

In the case of 500 simulations with $\mu = 1.33$,

$$\text{tr. ARE} = \frac{\text{trace }(\Sigma_g)}{\text{trace }(\Sigma_{mo})} = \frac{0.16671}{0.28902} = 0.57681$$

In the case of 1000 simulations with $\mu = 0.067$,

$$\text{tr. ARE} = \frac{\text{trace }(\Sigma_g)}{\text{trace }(\Sigma_{mo})} = \frac{1.04712}{0.96362} = 1.08665$$

The results show that as the sample size increases and with a decrease in μ, the information matrices of both models are approximately the same and the efficiency of the MLE'S is almost 1.

3.3.4 Simulation results

The $S - Plus^{®}$(2001) programming language was used to conduct the statistical simulations. The aim was to compare the estimates of Marshall-Olkin's model with that of the generalised Marshall-Olkin model. The data were simulated using the marginal distributions of the respective models. Sample sizes of 500 and 1000 were simulated.

From the table 3.1, it can be deduced that the estimates for Marshall-Olkin's model seems better than that of the Generalized Marshall-Olkin model in both simulations. Generalized Marshall-Olkin model shows some good results as the sample size increases but with a decrease in μ. When the sample is small then we have $n_3 = \sharp\{i : X_i = \mu Y_i\}$ to be less in that of Generalized Marshall-Olkin model than of classical Marshall-Olkin model. This can explain the discrepancy in the results.

Initial value	Marshall-Olkin		Gen. Marshall-Olkin	
	500	1000	500($\mu = 1.33$)	1000($\mu = 0.067$)
$\lambda_1 = 0.100$	0.0667	-	0.1178	-
$\lambda_2 = 0.200$	0.1675	-	0.1611	-
$\lambda_3 = 0.300$	0.3627	-	0.1699	-
$\lambda_1 = 0.250$	-	0.185	-	0.213
$\lambda_2 = 0.360$	-	0.318	-	0.290
$\lambda_3 = 0.420$	-	0.521	-	0.166

Table 3.1: Comparison of Marshall-Olkin ML estimates with Gen. Marshall-Olkin ML estimates

Chapter 4

Model Extension

We generalize the BVE's to any given function. Muliere and Scarsini (1987) characterized a class of bivariate Marshall-Olkin type distribution that generalize the Marshall-Olkin exponential distribution through functional equation involving binary associative operations. These classes of bivariate distributions and their marginal distributions and their form depends on the associative operation. They concentrate mass on the line $x = y$ as in the case of bivariate exponential distribution introduced by Marshall-Olkin(1967). It also should be noted that Marshall and Olkin treated the case of the binary relation $x * y = x + y$ in 1967. That is they considered a lack-of-memory -property-type functional equation which involves addition:

$$\overline{F}(x_1 * y_1, x_2 * y_2) = \overline{F}(x_1 + y_1, x_2 + y_2) = \overline{F}(x_1, x_2)\overline{F}(y_1, y_2)$$

A binary operation $*$ over real numbers is said to be associative if $(x * y) * h = x * (y * h)$. The operation $*$ is said to have an identity element $\mathbf{e}$ if $x * \mathbf{e} = x$. Let $*$ be a binary associative operation with an identity element $\mathbf{e} \in \mathbb{R}$. It is known (see Aczél 1966) that there is a strictly monotonically increasing continuous function $g : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying

$$x * y = g^{-1}(g(x) + g(y)).$$

Assuming there is an identity $\mathbf{e}$ for the relation, and the relation is associative and reducible, then any continuous solution of the equation

$$H(s * t) = H(s)H(t)$$

is of the form

$$H(s) = \exp(-\lambda g(s)).$$

Here we extend this concept by fixing some $\mu > 0$. The case $\mu = 1$ will lead to the results of Muliere and Scarsini. If $\mu \neq 1$ assume in addition that

$$\frac{s * t}{\mu} = \frac{s}{\mu} * \frac{t}{\mu}.$$

Suppose that the survival function $S(x, y)$ satisfies the functional equations

$$S(x * t, y * t/\mu) \;=\; S(x, y)S(t, t/\mu) \tag{4.1}$$

$$S_1(x * t) = S_1(x)S_1(t)\ ,\ \ S_1(x) = S(x, \mathbf{e}) \tag{4.2}$$

and

$$S_2(y * t) = S_2(y)S_2(t)\ ,\ \ S_2(y) = S(\mathbf{e}, y) \tag{4.3}$$

for all $x, y, t \geq \mathbf{e}$. Example 1: If $x * y = x + y$, then $g(x) = x$ and

$$S(x, y) = \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \max(x, y)\}\,;\ x, y > 0.$$

This is the Marshall-Olkin distribution.

Example 2: If $x * y = xy$, then $g(x) = \log x$ and

$$S(x, y) = x^{-\lambda_1} y^{-\lambda_2} (\max(x, y))^{-\lambda_3}.$$

This is the bivariate Pareto distribution over the set $(1, \infty) \times (1, \infty)$.

Example 3: If $x * y = (x^\alpha + y^\alpha)^{1/\varsigma}$, then $g(x) = x^\alpha$ and

$$S(x, y) = \exp\{-\lambda_1 x^\alpha - \lambda_2 y^\alpha - \lambda_3 \max(x^\alpha, y^\alpha)\}\,;\, x, y > 0$$

This is the bivariate Weibull distribution (cf. Marshall and Olkin(1967)); Moeschberger (1974). We extend the result in the following theorem.

Theorem 4.0.3 *The only continuous solutions of the equations (4.1)–(4.3) are of the form*

$$\overline{F}(u, v) = \exp[-\lambda_1 g(u) - \lambda_2 g(\mu v) - \lambda_3 g(\max(u, \mu v))], \tag{4.4}$$

where $\lambda_i > 0$.

Proof: We first note from eqn. 4.3 that $F_1(x) = \overline{F}(x, \mathbf{e})$ satisfies $\overline{F}_1(x*y) = \overline{F}_1(x)\overline{F}_1(y)$, whence has the form

$$\overline{F}_1(x) = \exp[-\theta_1 g(x)].$$

Likewise, the function $\overline{F}_0(x) = \overline{F}(\mathbf{e}, x/\mu)$ satisfies

$$\overline{F}_0(x*y) = \overline{F}(\mathbf{e}, \frac{x*y}{\mu}) = \overline{F}(\mathbf{e}, \frac{x}{\mu} * \frac{y}{\mu}) = \overline{F}_0(x)\overline{F}_0(y),$$

hence

$$\overline{F}_1(x) = \overline{F}_0(\mu x) = \exp[-\theta_2 g(\mu x)].$$

According to our assumptions,

$$\overline{F}(x*t, \frac{x*t}{\mu}) = \overline{F}(x,x)\overline{F}(t, t/\mu).$$

Defining $H(s) = \overline{F}(s, s/\mu)$, we find $H(s*t) = H(s)H(t)$ and so

$$H(x) = \exp[-\lambda g(x)].$$

It follows that

$$\overline{F}(s*t, \mathbf{e} * \frac{t}{\mu}) = \overline{F}(x, \mathbf{e})\overline{F}(t, t/\mu) = \exp[-\lambda g(t) - \theta_1 g(s)].$$

Putting $u = s*t$ and $t = \mu v$ we arrive at

$$\overline{F}(u,v) = \exp[-\lambda g(\mu v) - \theta_1[g(u) - g(\mu v)]],$$

where we used the fact that $g(u) = g(s) + g(\mu v)$. On the other hand

$$\overline{F}(\mathbf{e} * u, s * \frac{u}{\mu}) = \overline{F}(\mathbf{e}, s)\overline{F}(u, u/\mu) = \exp[-\lambda g(u) - \theta_2 g(\mu s)].$$

Putting $\mu v = \mu s * u$ and using $g(\mu v) = g(\mu s) + g(u)$ we obtain

$$\overline{F}(u,v) = \exp[-\lambda g(u) - \theta_2[g(\mu v) - g(u)]].$$

Setting $\lambda_1 = \lambda - \theta_2$, $\lambda_2 = \lambda - \theta_1$ and $\lambda_3 = \theta_1 + \theta_2 - \lambda$ it follows that

$$\overline{F}(u,v) = \exp[-\lambda_1 g(u) - \lambda_2 g(\mu v) - \lambda_3 g(\max(u, \mu v))]. \qquad \square$$

4.1 Another extension of Marshall-Olkin model

Let $\varphi(x, y)$ be function which is continuous and twice continuously differentiable in some open dense domain $G = G_\phi \subset \mathbb{R}^2$. $\overline{F}$ is a two-dimensional distribution function so the mixed partial derivative gives a density. We generalize our joint distribution by changing the notation from $\overline{F}(.,.)$ to $\overline{F}_{\varphi}(.,.)$. The resulting bivariate distribution is given by:

$$\begin{aligned}\overline{F}_{\varphi}(x,y) &= P(X > x, Y > y)\\ &= \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \varphi(x,y)\}\,;\; x,y>0,\; \underline{\lambda} \in \bar{\Lambda}\end{aligned} \tag{4.5}$$

We call eqn. (4.5) the φ model. We always assume that this defines a bivariate distribution function. Note that the marginals need not to be exponential. In case where $\varphi(x,y) = \max(x,y)$ we get the Marshall-Olkin's models with exponential marginals. See below. The parameter space is obtained by taken a negative partial derivative with respect to x and y:

$$-\partial\overline{F}_{\varphi}(x,y)/\partial x \;=\; -\Big[-\lambda_1 - \lambda_3 \partial\varphi(x,y)/\partial x\Big]\overline{F}_{\varphi}(x,y) \;\geq\; 0$$

so $\lambda_1 + \lambda_3 \partial\varphi(x,y)/\partial x \;\geq\; 0, \forall\, x,y$

$$-\partial\overline{F}_{\varphi}(x,y)/\partial y \;=\; -\Big[-\lambda_2 - \lambda_3 \partial\varphi(x,y)/\partial y\Big]\overline{F}_{\varphi}(x,y) \;\geq\; 0$$

so $\lambda_2 + \lambda_3 \partial\varphi(x,y)/\partial y \;\geq\; 0, \forall\, x,y.$

It follows that the density is

$$\frac{\partial^2 F_{\phi}}{\partial x \partial y}(u,v) = \big(\lambda_3 \phi''(u,v) + (\lambda_1 + \lambda_3 \phi_x(u,v))(\lambda_2 + \lambda_3 \phi_y(u,v))\big)\,\overline{F}_{\phi}(u,v)\,;$$

thus,

$$\bar{\Lambda} = \Big\{\; (\lambda_1,\lambda_2,\lambda_3) : \lambda_3 \phi''(x,y) + (\lambda_1 + \lambda_3 \phi_x(x,y))(\lambda_2 + \lambda_3 \phi_y(x,y)) \;\geq\; 0\,;\, x,y>0 \;\Big\} \tag{4.6}$$

when $\varphi(x,y) = \min(x,y)$, we have

$$\partial\varphi(x,y)/\partial y = \begin{cases} 0\,;\ \text{if } x \,<\, y; \\ 1\,;\ \text{if } x \,>\, y\,; \end{cases}$$

In that case $\varphi'_y = 1$, we define the joint density function in this case as $g_{\varphi}(x,y) \geq 0$ as:

$$g_{\varphi}(x,y) = \begin{cases} \lambda_2(\lambda_1 + \lambda_3)\overline{F}_{\varphi}\,;\; x<y\,;\; x,y>\,0, \\ \lambda_1(\lambda_2 + \lambda_3)\overline{F}_{\varphi}\,;\; y<x\,;\; x,y>\,0, \end{cases} \tag{4.7}$$

where φ'_x and φ'_y denotes $\partial\varphi(x,y)/\partial x$ and $\partial\varphi(x,y)/\partial y$ respectively. The case where

$$\overline{F}_{\varphi}(x,y) = \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 xy\}\,;\; x,y>0$$

this is recognizable as Gumbel's type 1 bivariate exponential distribution (cf. Gumbel, 1960) but in this case, $\lambda_1, \lambda_2 > 0$ and $0 < \lambda_3 \leq 1$.
We define the joint density function on G_ϕ for the general case as

$$f_\varphi(x,y) = \left(\lambda_3\phi''(x,y) + (\lambda_1 + \lambda_3\varphi_x(x,y))(\lambda_2 + \lambda_3\varphi_y(x,y))\right)\overline{F}_\varphi(x,y). \tag{4.8}$$

Let $\overline{F}_{\varphi(1)}$ and $\overline{F}_{\varphi(2)}$ denotes the marginal distributions of X and Y respectively.

$$\begin{aligned}\overline{F}_{\varphi(1)}(x) &= \overline{F}_\varphi(x,0)\\ &= \exp\{-\lambda_1 x - \lambda_3\varphi(x,0)\}\end{aligned}$$

the density for X is given as

$$f_{\varphi(1)}(x) = (\lambda_1 + \lambda_3\varphi_x)\exp\{-\lambda_1 x - \lambda_3\varphi(x,0)\}$$

and

$$E(X) = (\lambda_1 + \lambda_3\varphi_x)^{-1}$$

likewise, the marginal of Y is

$$\overline{F}_{\varphi(2)}(y) = \exp\{-\lambda_2 y - \lambda_3\varphi(0,y)\}$$

the corresponding density is

$$f_{\varphi(2)}(y) = (\lambda_2 + \lambda_3\varphi_y)\exp\{-\lambda_2 y - \lambda_3\varphi(0,y)\}$$

and

$$E(Y) = (\lambda_2 + \lambda_3\varphi_y)^{-1}.$$

4.2 Maximum Likelihood Estimation

In this section we will derive the MLE for eqn. 4.8. when assuming that ϕ'' vanishes. The likelihood function of a sample of size n of pairs $z_i = (x_i, y_i)$ for $1 \leq i \leq n$ individuals is

$$\begin{aligned}L^\varphi(\underline{\lambda}) &= \prod_{i=1}^n f_\varphi(x_i,y_i)\\ &= \exp\{-\lambda_1\sum_{i=1}^n x_i - \lambda_2\sum_{i=1}^n y_i - \lambda_3\sum_{i=1}^n \varphi(x_i,y_i)\}\\ &\quad \prod_{i=1}^n[\lambda_3\phi''(x_i,y_i) + (\lambda_1 + \lambda_3\varphi_x(x_i,y_i))(\lambda_2 + \lambda_3\varphi_y(x_i,y_i))].\end{aligned} \tag{4.9}$$

The log likelihood function is

$$\begin{aligned}\log L^{\varphi}(\underline{\lambda}) &= l^{\varphi} \\ &= -\lambda_1 \sum_{i=1}^{n} x_i - \lambda_2 \sum_{i=1}^{n} y_i - \lambda_3 \sum_{i=1}^{n} \varphi(x_i, y_i) \\ &+ \sum_{i=1}^{n} \log\Big[\lambda_3 \phi''(x_i, y_i) + (\lambda_1 + \lambda_3 \varphi_x(x_i, y_i))(\lambda_2 + \lambda_3 \varphi_y(x_i, y_i)).\Big] \quad (4.10)\end{aligned}$$

The likelihood equations obtained by taken partial derivatives of l^{φ} with respect to $\lambda_i's$ and setting to zero are given as:

$$\sum_{i=1}^{n}\Big\{(\lambda_2 + \lambda_3\varphi_y(x_i, y_i))/(\lambda_1 + \lambda_3\varphi_x(x_i, y_i))(\lambda_2 + \lambda_3\varphi_y(x_i, y_i))\Big\} = \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n}\Big\{(\lambda_1 + \lambda_3\varphi_x(x_i, y_i))/(\lambda_1 + \lambda_3\varphi_x(x_i, y_i))(\lambda_2 + \lambda_3\varphi_y(x_i, y_i))\Big\} = \sum_{i=1}^{n} y_i$$

$$\begin{aligned}\sum_{i=1}^{n} \Big\{&[\varphi_x(x_i, y_i)(\lambda_2 + \lambda_3\varphi_y(x_i, y_i)) + \varphi_y(x_i, y_i)(\lambda_1 + \lambda_3\varphi_x(x_i, y_i))] \\ &/(\lambda_1 + \lambda_3\varphi_x(x_i, y_i))(\lambda_2 + \lambda_3\varphi_y(x_i, y_i))\Big\} = \sum_{i=1}^{n} \varphi(x_i, y_i)\end{aligned}$$

from the assumption of ϕ''. Hence, the likelihood equations are

$$\begin{cases} \sum_{i=1}^{n}\Big\{(\lambda_2 + \lambda_3\varphi_y(x_i, y_i))/(\lambda_1 + \lambda_3\varphi_x(x_i, y_i))(\lambda_2 + \lambda_3\varphi_y(x_i, y_i))\Big\} = \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n}\Big\{(\lambda_1 + \lambda_3\varphi_x(x_i, y_i))/(\lambda_1 + \lambda_3\varphi_x(x_i, y_i))(\lambda_2 + \lambda_3\varphi_y(x_i, y_i))\Big\} = \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} \Big\{[\varphi_x(x_i, y_i)(\lambda_2 + \lambda_3\varphi_y(x_i, y_i)) + \varphi_y(x_i, y_i)(\lambda_1 + \lambda_3\varphi_x(x_i, y_i))] \\ /(\lambda_1 + \lambda_3\varphi_x(x_i, y_i))(\lambda_2 + \lambda_3\varphi_y(x_i, y_i))\Big\} = \sum_{i=1}^{n} \varphi(x_i, y_i) \end{cases} \quad (4.11)$$

For brevity we omit $\sum_{i=1}^{n}$ and write $\varphi_x = \varphi_x(x_i, y_i)$ and $\varphi_y = \varphi_y(x_i, y_i)$. The Hessian matrix on $\bar{\Lambda}$ is given by

$$Q_{\varphi} = \nabla^2 l^{\varphi} = (\partial^2 l^{\varphi}/\partial\lambda_i\partial\lambda_j)_{i,j=1,2,3} \quad (4.12)$$

$$= -\begin{pmatrix} 1/(\lambda_1 + \lambda_3\varphi_x)^2 & 0 & \varphi_x/(\lambda_1 + \lambda_3\varphi_x)^2 \\ 0 & 1/(\lambda_2 + \lambda_3\varphi_y)^2 & \varphi_y/(\lambda_2 + \lambda_3\varphi_y)^2 \\ \varphi_x/(\lambda_1 + \lambda_3\varphi_x)^2 & \varphi_y/(\lambda_2 + \lambda_3\varphi_y)^2 & \varphi_x^2/(\lambda_1 + \lambda_3\varphi_x)^2 + \varphi_y^2/(\lambda_2 + \lambda_3\varphi_y)^2 \end{pmatrix}$$

The existence and uniqueness properties of MLE are given in the following theorem.

Theorem 4.2.1 *Let $(X_i, Y_i); i = 1, \ldots, n$, $n \in \mathbb{N}$ be independent identically distributed (i.i.d.) sequence with cumulative distribution function (c.d.f) F_φ given by eqn. 4.5, with parameters $\underline{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in \bar{\Lambda}$ cf. eqn. 4.6. Let $\mathbb{R} = \{(x_i, y_i) \ldots, (x_n, y_n) : \exists i : x_i < y_i, \exists j : y_j < x_i, \exists k : y_k = x_i\} \subset \mathbb{R}^{2n}$. Then for all $(x_i, y_i); i = 1, \ldots, n$; $n \in \mathbb{R}$. The MLE $\widehat{\lambda}$ for λ exists and is uniquely determined by eqn. 4.11.*

Proof of Thm. 4.2.1 Let $l^\varphi(\underline{\lambda})$ denote the likelihood function of the model as derived in eqn. 4.10. The Hessian Q_φ of the likelihood is given as eqn. 4.12. The likelihood function $l^\varphi(\underline{\lambda})$ is twice differentiable. We want to show that the negative Hessian $(-Q_\varphi)$ is positive definite :

1. The first entry $1/(\lambda_1 + \lambda_3\varphi_x)^2 > 0$.

2. The determinant of the matrix

$$\begin{pmatrix} 1/(\lambda_1 + \lambda_3\varphi_x)^2 & 0 \\ 0 & 1/(\lambda_2 + \lambda_3\varphi_y)^2 \end{pmatrix} > 0$$

   and

3. The determinant of the whole matrix

$$\begin{aligned} &\begin{pmatrix} 1/(\lambda_1 + \lambda_3\varphi_x)^2 & 0 & \varphi_x/(\lambda_1 + \lambda_3\varphi_x)^2 \\ 0 & 1/(\lambda_2 + \lambda_3\varphi_y)^2 & \varphi_y/(\lambda_2 + \lambda_3\varphi_y)^2 \\ \varphi_x/(\lambda_1 + \lambda_3\varphi_x)^2 & \varphi_y/(\lambda_2 + \lambda_3\varphi_y)^2 & \varphi_x^2/(\lambda_1 + \lambda_3\varphi_x)^2 + \varphi_y^2/(\lambda_2 + \lambda_3\varphi_y)^2 \end{pmatrix} \\ = \quad & \frac{1}{(\lambda_1 + \lambda_3\varphi_x)^2}\left\{\frac{1}{(\lambda_2 + \lambda_3\varphi_y)^2}\left(\frac{\varphi_x^2}{(\lambda_1 + \lambda_3\varphi_x)^2} + \frac{\varphi_y^2}{(\lambda_2 + \lambda_3\varphi_y)^2}\right) - \frac{\varphi_y^2}{(\lambda_2 + \lambda_3\varphi_y)^4}\right\} \\ - \quad & \frac{1}{(\lambda_2 + \lambda_3\varphi_y)^2}\frac{\varphi_x^2}{(\lambda_1 + \lambda_3\varphi_x)^4}. \end{aligned}$$

   In order to show that $-Q_\varphi$ is positive definite, for any $\theta \neq 0$ write If $\theta = (\theta_1, \theta_2, \theta_3)^T$ with θ_1 or $\theta_2 \neq 0$ then strict positivity from

$$\theta^T(-Q_\varphi)\,\theta = \frac{1}{(\lambda_1 + \lambda_3\varphi_x)^2}\,\theta_1^2 + \frac{1}{(\lambda_2 + \lambda_3\varphi_y)^2}\,\theta_2^2 > 0.$$

   If $\theta = (\theta_1, \theta_2, \theta_3)^T$ and $\theta_1 = \theta_2 = 0$ and $\theta_3 \neq 0$, then

$$\theta^T(-Q_\varphi)\,\theta = \left[\varphi_x^2/(\lambda_1 + \lambda_3\varphi_x)^2 + \varphi_y^2/(\lambda_2 + \lambda_3\varphi_y)^2\right]\theta_3^2 > 0.$$

   So $-Q_\varphi$ is positive definite, hence Q_φ is negative definite. By Thms. 3.2 and 4.2 of Mangarsin (1969, pages 89 and 91) the likelihood function $l^\varphi(\underline{\lambda})$ is strictly concave on $\bar{\Lambda}$. Hence $l^{\varphi}(\underline{\lambda})$ must have a unique maximum on $\bar{\Lambda}$ given by the roots of the $\nabla \mathbf{l}^{\varphi}(\underline{\lambda}) = 0$. □

Using theorem 3.2.2, the joint sufficient statistic for eqn. 4.5. is

$$\{n_1\,,\, n_2\,,\, \sum_{i=1}^{n} x_i\,,\, \sum_{i=1}^{n} y_i\,,\, \sum_{i=1}^{n} \varphi(x_i, y_i)\}$$

4.3 Asymptotic Properties

In this section we shall deal with the asymptotic properties of the MLE for the parameters of eqn. 4.5.

4.3.1 Consistency

Theorem 4.3.1 *For every $n \in \mathbb{N}$, let $Z_i^n = (X_i^n, Y_i^n); i = 1, \ldots, n$ be i.i.d sequence with c.d.f. F_{φ} given by , eqn. 4.5, with parameters $\underline{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in \bar{\Lambda}$ cf. eqn. 4.6. Let $\hat{\underline{\lambda}}_n^{\varphi}$ denote the MLE for $\underline{\lambda}$ based on $Z_1^n, \ldots, Z_n^n; n \in \mathbb{N}$. Then for each $\lambda \in \bar{\Lambda}$, $\hat{\lambda}_n^{\varphi}$ converges stochastically to λ under the law F_{φ}. The consequence the MLE is consistent. For $n \to \infty$*

$$\hat{\lambda}_n^{\varphi} \xrightarrow{p} \lambda$$

that is

$$\lim_{n\to\infty} P_{\lambda}\Big(|\hat{\underline{\lambda}}_n^{\varphi} - \underline{\lambda}| \le \epsilon\Big) = 1$$

Proof Let $l^{\varphi}(\underline{\lambda})$ denote the likelihood function of the model as derived in eqn. 4.11. We are considering a set

$$C_{\delta}^{\varphi} = \{\underline{\lambda} \in \bar{\Lambda} : \|\underline{\lambda} - \tilde{\underline{\lambda}}^{\varphi}\| \le \delta\}.$$

where $\delta > 0$ and $\tilde{\underline{\lambda}}^{\varphi}$ is fixed. Let the notation $\wp C_{\delta}^{\varphi}$ denotes the boundary of C_{δ}^{φ}. We want to show that $\forall\, \delta > 0$

$$\lim_{n\to\infty} P_{\tilde{\underline{\lambda}}}^{\varphi}\Big(l_n(\underline{\lambda})^{\varphi} < l_n(\tilde{\underline{\lambda}}^{\varphi}); \forall\, \lambda \in \wp C_{\delta}^{\varphi}\Big) = 1 \Rightarrow \lim_{n\to\infty} P_{\tilde{\underline{\lambda}}}^{\varphi}\Big(|\hat{\underline{\lambda}}_n^{\varphi} - \tilde{\underline{\lambda}}^{\varphi}| \le \delta\Big) = 1$$

Let

$$A_n^{\varphi} = \{\kappa \in \Omega :\, l_n^{\varphi}(\lambda, z(\kappa)) < l_n^{\varphi}(\tilde{\lambda}^{\varphi}) \,: \forall\, \lambda \in \wp C_{\delta}^{\varphi}\}$$
$$B_n^{\varphi} = \{|\hat{\underline{\lambda}}_n^{\varphi} - \tilde{\underline{\lambda}}^{\varphi}| \le \delta\}$$

Known

$$\lim_{n\to\infty} P_{\tilde{\lambda}}^{\varphi}(A_n^{\varphi}) = 1\,, \text{claim} \lim_{n\to\infty} P_{\tilde{\lambda}}^{\varphi}(B_n^{\varphi}) = 1$$

if we would know that $A_n^{\varphi} \subset B_n^{\varphi}$ then

$$P_{\tilde{\lambda}}^{\varphi}(A_n^{\varphi}) \le P_{\tilde{\lambda}}^{\varphi}(B_n^{\varphi})$$

so

$$1 \geq \lim_{n\to\infty} P^{\varphi}_{\tilde{\lambda}}(B^{\varphi}_n) \geq \lim_{n\to\infty} P_{\tilde{\lambda}^{\varphi}}(A^{\varphi}_n) = 1$$

To show that $A^{\varphi}_n \subset B^{\varphi}_n$, let $\kappa \in A^{\varphi}_n$, this implies

$$l^{\varphi}_n(\lambda; z_1(\kappa), \ldots, z_n(\kappa)) < l^{\varphi}_n(\tilde{\lambda}|z_1(\kappa), \ldots, z_n(\kappa)) : \forall\, \lambda \in \wp C^{\varphi}_{\delta})$$

this implies, the maximum of $l^{\varphi}_n(\lambda; z(\kappa))$, is attained in the interior of $C^{\varphi}_{\delta} \Rightarrow$ $l^{\varphi}_n(.; z(\kappa))$ has a zero for $l^{\varphi\,\prime}$ in C^{φ}_{δ}. $l^{\varphi\,\prime}$ has only one zero implies MLE lies in C^{φ}_{δ} implies $\hat{\lambda}^{\varphi} \in C^{\varphi}_{\delta}$, implies

$$|\hat{\lambda}^{\varphi}_n - \tilde{\lambda}^{\varphi}| \leq \delta \Rightarrow \kappa \in B^{\varphi}_n$$

The next point is to proof that

$$\lim_{n\to\infty} P^{\varphi}_{\tilde{\underline{\lambda}}}\Big(l^{\varphi}_n(\underline{\lambda}) < l^{\varphi}_n(\tilde{\underline{\lambda}}); \forall\, \lambda \in \wp C^{\varphi}_{\delta}\Big) = 1$$

we proof this by first showing that

$$\lim_{n\to\infty} \frac{1}{n}(\underline{\lambda} - \tilde{\underline{\lambda}}^{\varphi})^T \dot{\mathbf{l}}^{\varphi}(\lambda)_{\lambda=\tilde{\lambda}^{\varphi}} = E^{\varphi}_{\tilde{\lambda}}[\dot{\mathbf{l}}^{\varphi}]_{\lambda=\tilde{\lambda}^{\varphi}} = 0$$

where $\dot{\mathbf{l}}^{\varphi} = \frac{\partial \log f}{\partial \lambda}$ and that

$$\lim_{n\to\infty} \frac{1}{n}(\underline{\lambda} - \tilde{\underline{\lambda}})^T \ddot{\mathbf{l}}^{\varphi}(\lambda)(\underline{\lambda} - \tilde{\underline{\lambda}})_{\lambda=\tilde{\lambda}} = E_{\tilde{\lambda}}[\ddot{\mathbf{l}}^{\varphi}]_{\lambda=\tilde{\lambda}} < 0$$

The partial derivative of $\mathbf{l}^{\varphi}$ with respect to λ_1 is given as

$$\partial \mathbf{l}^{\varphi}/\partial\lambda_1 = n/(\lambda_1 + \lambda_3\varphi_x) - \sum_{i=1}^{n} x_i$$

dividing both sides by n, we obtain

$$\begin{aligned}
\frac{1}{n}\frac{\partial \mathbf{l}^{\varphi}}{\partial \lambda_1} &= \frac{1}{n}(1/(\lambda_1 + \lambda_3\varphi_x) - \sum_{i=1}^{n} x_i) \\
&= \frac{1}{(\lambda_1 + \lambda_3\varphi_x)} - \frac{\sum_{i=1}^{n} x_i}{n} \\
\lim_{n\to\infty} \frac{1}{n}\frac{\partial \mathbf{l}^{\varphi}}{\partial \lambda_1} &= \frac{1}{(\lambda_1 + \lambda_3\varphi_x)} - E(X) \\
&= \frac{1}{(\lambda_1 + \lambda_3\varphi_x)} - \frac{1}{(\lambda_1 + \lambda_3\varphi_x)} \\
&= 0
\end{aligned}$$

It can same way be shown that

$$\lim_{n\to\infty} \frac{1}{n}\frac{\partial \mathbf{l}^{\varphi}}{\partial \lambda_2} = 0$$

The partial derivative of $\mathbf{l}^{\varphi}$ with respect to λ_3 is given as

$$\partial \mathbf{l}^{\varphi}/\partial\lambda_3 = \Big[\varphi_x(\lambda_2+\lambda_3\varphi_y)+\varphi_y(\lambda_1+\lambda_3\varphi_x)\Big]/(\lambda_1+\lambda_3\varphi_x)(\lambda_2+\lambda_3\varphi_y) - \sum_{i=1}^{n}\varphi(x_i,y_i)$$

dividing both sides by n, we obtain

$$\begin{aligned}\frac{1}{n}\frac{\partial \mathbf{l}^{\varphi}}{\partial\lambda_3} = \frac{1}{n}\Bigg(\Big[\varphi_x(\lambda_2 &+ \lambda_3\varphi_y)+\varphi_y(\lambda_1+\lambda_3\varphi_x)\Big]/(\lambda_1+\lambda_3\varphi_x)(\lambda_2+\lambda_3\varphi_y) \\ &- \sum_{i=1}^{n}\varphi(x_i,y_i)\Bigg) \\ \lim_{n\to\infty}\frac{1}{n}\frac{\partial \mathbf{l}^{\varphi}}{\partial\lambda_3} = \lim_{n\to\infty}\frac{1}{n}\Bigg(\Big[\varphi_x(\lambda_2 &+ \lambda_3\varphi_y)+\varphi_y(\lambda_1+\lambda_3\varphi_x)\Big]/(\lambda_1+\lambda_3\varphi_x)(\lambda_2+\lambda_3\varphi_y)\Bigg) \\ - E(\varphi(X,Y)) &= 0\end{aligned}$$

where we have assumed the weak law of large numbers.

Now, by taking a third order Taylor expansion around $\underline{\tilde{\lambda}}$, we have

$$\begin{aligned}\frac{1}{n}(\mathbf{l}^{\varphi}(\underline{\lambda})-\mathbf{l}^{\varphi}(\underline{\tilde{\lambda}})) &= \frac{1}{n}(\underline{\lambda}-\underline{\tilde{\lambda}})^T\dot{\mathbf{l}}^{\varphi}(\underline{\tilde{\lambda}}) \\ &+ \frac{1}{2}(\underline{\lambda}-\underline{\tilde{\lambda}})^T\Big(\frac{1}{n}\ddot{\mathbf{l}}^{\varphi}(\underline{\tilde{\lambda}})\Big)(\underline{\lambda}-\underline{\tilde{\lambda}}) \\ &+ \frac{1}{6}\frac{1}{n}\sum_{r=1}^{k}\sum_{s=1}^{k}\sum_{t=1}^{k}(\underline{\lambda}_r-\underline{\tilde{\lambda}}_r)(\underline{\lambda}_s-\underline{\tilde{\lambda}}_s)(\underline{\lambda}_t-\underline{\tilde{\lambda}}_t)\Big\{\gamma_{rst}(Z_i)H_{rst}(Z_i)^{\varphi}\Big\} \\ &= S_1^{\varphi}+S_2^{\varphi}+S_3^{\varphi}\end{aligned}$$

where

$$S_1^{\varphi} = \frac{1}{n}(\underline{\lambda}-\underline{\tilde{\lambda}})^T\dot{\mathbf{l}}^{\varphi}(\underline{\tilde{\lambda}})$$

$$S_2^{\varphi} = \frac{1}{2}(\underline{\lambda}-\underline{\tilde{\lambda}})^T\Big(\frac{1}{n}\ddot{\mathbf{l}}^{\varphi}(\underline{\tilde{\lambda}})\Big)(\underline{\lambda}-\underline{\tilde{\lambda}})$$

$$S_3^{\varphi} = \frac{1}{6}\frac{1}{n}\sum_{r=1}^{k}\sum_{s=1}^{k}\sum_{t=1}^{k}(\underline{\lambda}_r-\underline{\tilde{\lambda}}_r)(\underline{\lambda}_s-\underline{\tilde{\lambda}}_s)(\underline{\lambda}_t-\underline{\tilde{\lambda}}_t)\Big\{\gamma_{rst}(Z_i)H_{rst}(Z_i)^{\varphi}\Big\}$$

we make some assumptions that $0 \le |\gamma_{rst}(z)| < 1$ and $|\frac{\partial^3 \log f_{\varphi}}{\partial\lambda_r\partial\lambda_s\partial\lambda_t}| < H_{rst}(Z_i)^{\varphi}$. We have seen that

$$S_1^{\varphi} \overset{p}{\longrightarrow} 0$$

The Hessian Q is negative definite from Thm. 4.2.1, so the second term S_2^{φ} is negative with probability tending to 1. S_1^{φ} and S_3^{φ} are small compared to S_2^{φ}

so the

$$\sup_{\underline{\lambda} \in C_\delta^\varphi} (S_1^\varphi + S_2^\varphi + S_3^\varphi) < 0.$$

Thus, for n large enough,

$$\frac{1}{n}(\mathbf{l}^\varphi(\underline{\lambda}) - \mathbf{l}^\varphi(\tilde{\underline{\lambda}})) < 0 \qquad \square$$

this completes the proof.

4.3.2 Asymptotic Normality

Theorem 4.3.2 *Let $Z_n = (X_n, Y_n); (n \geq 1)$ be independent identically distributed (i.i.d.) random variables with cumulative distribution function (c.d.f) F_φ given by , eqn. 4.5, with parameters $\underline{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in \bar{\Lambda}$. Then the MLE $\widehat{\lambda}$ for λ is asymptotically normal*

$$\mathbb{N}\,(\,\underline{\lambda},\, \Sigma^\varphi(\underline{\lambda})^{-1}\,)$$

where

$$\Sigma^\varphi(\underline{\lambda}) = \mathbf{\Pi}^\varphi \Sigma_0^{\varphi -1}.$$

Proof Let $\delta > 0$ and set

$$G_n^\varphi = \{(z_1, \ldots, z_n) \in \mathbb{R}^{2n} : |\,\underline{\lambda} - \widehat{\lambda}(z_1, \ldots, z_n)\,| < \delta$$

If δ is small enough, then for $(z_1, \ldots, z_n) \in G_n^\varphi$, $\widehat{\lambda}$ is unique. Then , by consistency,

$$P_{\underline{\lambda}}(\,G_n^\varphi\,) \to 1 \text{ as } n \to \infty$$

hence for $\epsilon > 0$ there is $n_0 \in \mathbb{N}$ such that

$$P_{\underline{\lambda}}(\,G_n^{\varphi c}\,) < \epsilon \;\forall n \geq n_0$$

It follows that for $n \geq n_0$

$$\begin{aligned} &|\,P_{\underline{\lambda}}(\{\omega' : \widehat{\lambda}(z_1(\omega'), \ldots, z_n(\omega')) \leq \underline{t'}\} \cap \{z_1, \ldots, z_n \in G_n^\varphi\}) \\ - \quad &P_{\underline{\lambda}}(\{\omega' : \widehat{\lambda}(z_1(\omega'), \ldots, z_n(\omega')) \leq \underline{t'}\})\,| \leq P_{\underline{\lambda}}(\,G_n^{\varphi c}\,) < \epsilon \end{aligned}$$

Note that $\underline{t'} \in \mathbb{R}^3$ and $\underline{s'} \leq \underline{t'}$ means $\underline{s'}_i \leq \underline{t'}_i\,(i = 1, 2, 3)$. The likelihood function l^φ can be expressed as a Taylor series by

$$l^{\varphi\,\prime}(\lambda) = l^{\varphi\,\prime}(\underline{\lambda}) + (\lambda - \underline{\lambda})^T l^{\varphi\,\prime\prime}(\underline{\lambda}^*)$$

for some value of $\underline{\lambda}^* \in B(\underline{\lambda}, \delta)$, the δ-ball around $\underline{\lambda}$. So if $z_1, \ldots, z_n \in G_n^{\varphi}, n \geq n_0$

$$l^{\varphi\,\prime}(\widehat{\lambda}(z_1, \ldots, z_n)) = 0$$

and

$$l^{\varphi\,\prime}(z_1, \ldots, z_n, \lambda) = -(\widehat{\lambda}(z_1, \ldots, z_n) - \underline{\lambda}) l^{\varphi\,\prime\prime}(z_1, \ldots, z_n, \underline{\lambda}^*)$$

by eqn. (4.10)

$$\begin{aligned} \log L^{\varphi}(\underline{\lambda}) &= l^{\varphi} \\ &= -\lambda_1 \sum_{i=1}^{n} x_i - \lambda_2 \sum_{i=1}^{n} y_i - \lambda_3 \sum_{i=1}^{n} \varphi(x_i, y_i) \\ &+ \sum_{i=1}^{n} \log\Big[\lambda_3 \phi''(x_i, y_i) + (\lambda_1 + \lambda_3 \varphi_x(x_i, y_i))(\lambda_2 + \lambda_3 \varphi_y(x_i, y_i))\Big] \end{aligned}$$

so

$$l^{\varphi\,\prime}(z_1, \ldots, z_n, \lambda) = \begin{pmatrix} \partial \mathbf{l}^{\varphi}/\partial \lambda_1 \\ \partial \mathbf{l}^{\varphi}/\partial \lambda_2 \\ \partial \mathbf{l}^{\varphi}/\partial \lambda_3 \end{pmatrix}$$

$$\sum_{i=1}^{n} \Big\{ (\lambda_2 + \lambda_3 \varphi_y(x_i, y_i)) / (\lambda_1 + \lambda_3 \varphi_x(x_i, y_i))(\lambda_2 + \lambda_3 \varphi_y(x_i, y_i)) \Big\} = \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} \Big\{ (\lambda_1 + \lambda_3 \varphi_x(x_i, y_i)) / (\lambda_1 + \lambda_3 \varphi_x(x_i, y_i))(\lambda_2 + \lambda_3 \varphi_y(x_i, y_i)) \Big\} = \sum_{i=1}^{n} y_i$$

$$\begin{aligned} \sum_{i=1}^{n} \Big\{ [\varphi_x(x_i, y_i)(\lambda_2 &+ \lambda_3 \varphi_y(x_i, y_i)) + \varphi_y(x_i, y_i)(\lambda_1 + \lambda_3 \varphi_x(x_i, y_i))] \\ &/ \quad (\lambda_1 + \lambda_3 \varphi_x(x_i, y_i))(\lambda_2 + \lambda_3 \varphi_y(x_i, y_i)) \Big\} = \sum_{i=1}^{n} \varphi(x_i, y_i) \end{aligned}$$

where $E\, l^{\varphi\,\prime} = 0$ from theorem 4.3.1 (consistency). It follows that $\frac{1}{\sqrt{n}} l^{\varphi\,\prime}(z_1, \ldots, z_n, \underline{\lambda})$ is asymptotically normal with expectation $\mathbf{0}$ and covariance matrix $\mathbf{\Pi}^{\varphi} = (\pi_{ij})^{\varphi}_{1 \leq i,j \leq 3}$. We calculate this matrix as follows:

$$var(X) = 1/(\lambda_1 + \lambda_3 \varphi_x)^2$$

$$var(Y) = 1/(\lambda_2 + \lambda_3\varphi_y)^2$$

$$\begin{aligned}\partial \mathbf{l}^{\varphi}/\partial\lambda_3 &= \sum_{i=1}^{n}\Big\{[\varphi_x(x_i,y_i)(\lambda_2+\lambda_3\varphi_y(x_i,y_i)) + \varphi_y(x_i,y_i)(\lambda_1+\lambda_3\varphi_x(x_i,y_i))] \\ &/ \quad (\lambda_1+\lambda_3\varphi_x(x_i,y_i))(\lambda_2+\lambda_3\varphi_y(x_i,y_i))\Big\} = \sum_{i=1}^{n}\varphi(x_i,y_i)\end{aligned}$$

$$var((\varphi(X,Y)) = E((\varphi(X,Y))^2 - E^2((\varphi(X,Y))$$

but

$$E((\varphi(X,Y)) = E(\partial \mathbf{l}^{\varphi}/\partial\lambda_3) = 0$$

from the proof of consistency

$$\begin{aligned}-\frac{1}{\sqrt{n}}\frac{\partial \mathbf{l}^{\varphi}}{\partial\lambda_3} &= -\frac{1}{n}\Big\{[\varphi_x(x_i,y_i)(\lambda_2+\lambda_3\varphi_y(x_i,y_i)) + \varphi_y(x_i,y_i)(\lambda_1+\lambda_3\varphi_x(x_i,y_i))] \\ &\quad /(\lambda_1+\lambda_3\varphi_x(x_i,y_i))(\lambda_2+\lambda_3\varphi_y(x_i,y_i))\Big\}\end{aligned}$$

now

$$E((\varphi(X,Y))^2 = E(\partial \mathbf{l}^{\varphi}/\partial\lambda_3)^2$$

therefore

$$\begin{aligned}\Big(-\frac{1}{\sqrt{n}}\frac{\partial \mathbf{l}^{\varphi}}{\partial\lambda_3)}\Big)^2 &= \frac{1}{n}\Big\{[\varphi_x(x_i,y_i)(\lambda_2+\lambda_3\varphi_y(x_i,y_i)) + \varphi_y(x_i,y_i)(\lambda_1+\lambda_3\varphi_x(x_i,y_i))] \\ &\quad /(\lambda_1+\lambda_3\varphi_x(x_i,y_i))(\lambda_2+\lambda_3\varphi_y(x_i,y_i))\Big\}^2\end{aligned}$$

now,

$$\sum_{i=1}^{n} var((\varphi(X,Y)) = \sum_{i=1}^{n} E\Big(-\frac{1}{\sqrt{n}}\frac{\partial \mathbf{l}^{\varphi}}{\partial\lambda_3)}\Big)^2 = n\,E\Big(-\frac{1}{\sqrt{n}}\frac{\partial \mathbf{l}^{\varphi}}{\partial\lambda_3)}\Big)^2$$

therefore

$$\begin{aligned}var((\varphi(X,Y)) &= \Big\{[\varphi_x(x_i,y_i)(\lambda_2+\lambda_3\varphi_y(x_i,y_i)) + \varphi_y(x_i,y_i)(\lambda_1+\lambda_3\varphi_x(x_i,y_i))] \\ &\quad /(\lambda_1+\lambda_3\varphi_x(x_i,y_i))(\lambda_2+\lambda_3\varphi_y(x_i,y_i))\Big\}^2\end{aligned}$$

Now,

$$\begin{aligned}cov(X,\varphi(X,Y)) &= E(X\varphi(X,Y)) - E(X)E(\varphi(X,Y)) \\ &= E(X\varphi(X,Y)) \; - \; \{1/(\lambda_1+\lambda_3\varphi_x)\}\{\lim_{n\to\infty}\frac{1}{n}\Bigg(\Big[\varphi_x(\lambda_2+\lambda_3\varphi_y) \\ &+ \; \varphi_y(\lambda_1+\lambda_3\varphi_x)\Big]/(\lambda_1+\lambda_3\varphi_x)(\lambda_2+\lambda_3\varphi_y)\Bigg)\} \\ &= E(X\varphi(X,Y))\end{aligned}$$

also,

$$\begin{aligned} cov(Y,\varphi(X,Y)) &= E(Y\varphi(X,Y)) - E(Y)E(\varphi(X,Y)) \\ &= E(Y\varphi(X,Y)) \; - \; \{1/(\lambda_2+\lambda_3\varphi_y)\}\{\lim_{n\to\infty}\frac{1}{n}\Bigg(\Big[\varphi_x(\lambda_2+\lambda_3\varphi_y) \\ &+ \varphi_y(\lambda_1+\lambda_3\varphi_x)\Big]/(\lambda_1+\lambda_3\varphi_x)(\lambda_2+\lambda_3\varphi_y)\Bigg)\} \\ &= E(Y\varphi(X,Y)) \end{aligned}$$

If we denote the covariance matrix by $\mathbf{\Pi}^{\varphi} = (\pi_{ij})^{\varphi}_{1\le i,j\le 3}$ then

$$\pi^{\varphi}_{11} = var(X)\,,\; \pi^{\varphi}_{12} = \pi^{\varphi}_{21} = 0\,,\; \pi^{\varphi}_{33} = var(\varphi)$$

$$\pi^{\varphi}_{22} = var(Y)\,,\; \pi^{\varphi}_{13} = cov(X,\varphi)\,,\; \pi^{\varphi}_{23} = cov(Y,\varphi)$$

since

$$P(\,(Z_1\ldots Z_n\,)\in G^{\varphi}_n\,)\to 1 \text{ as } n\to\infty$$

$\frac{1}{\sqrt{n}}l^{\varphi\,\prime}(Z_1,\ldots,Z_n\,,\underline{\lambda})$ and $-\frac{1}{\sqrt{n}}(\hat{\lambda}-\underline{\lambda})^T l^{\varphi\,\prime\prime}(\underline{\lambda}^*)$ are equivalent. Now,

$$\frac{1}{n}l^{\varphi\,\prime\prime}(Z_1,\ldots,Z_n\,,\underline{\lambda}^*) \;=\; \frac{1}{n}U^{\varphi}(Z_1,\ldots,Z_n) \;\xrightarrow{p}\; \Sigma^{\varphi}_0$$

using Q_{φ} from eqn. 4.12 with $\lambda_i = \lambda^*_i$ we work Σ^{φ}_0 as follows:

$$\tau_{11} = E(-\frac{\partial^2 l^{\varphi}}{\partial\lambda^{*2}_1}) \;=\; 1/(\lambda_1+\lambda_3\varphi_x)^2$$

$$\tau_{22} = E(-\frac{\partial^2 l^{\varphi}}{\partial\lambda^{*2}_2}) \;=\; 1/(\lambda_2+\lambda_3\varphi_y)^2$$

$$\tau_{33} = E(-\frac{\partial^2 l^{\varphi}}{\partial\lambda^{*2}_3}) \;=\; \varphi^2_x/(\lambda_1+\lambda_3\varphi_x)^2 + \varphi^2_y/(\lambda_2+\lambda_3\varphi_y)^2$$

$$\tau_{12} = \tau_{21} = E(-\frac{\partial^2 l^{\varphi}}{\partial\lambda^*_1\partial\lambda^*_2}) = 0$$

$$\tau_{13} = E(-\frac{\partial^2 l^{\varphi}}{\partial\lambda^*_1\partial\lambda^*_3}) \;=\; \varphi_x/(\lambda_1+\lambda_3\varphi_x)^2$$

$$\tau_{23} = E(-\frac{\partial^2 l^{\varphi}}{\partial\lambda^*_2\partial\lambda^*_3}) \;=\; \varphi_y/(\lambda_2+\lambda_3\varphi_y)^2$$

consequently the information matrix has the form

$$\Sigma_0^\varphi = \begin{pmatrix} \tau_{11} & 0 & \tau_{13} \\ 0 & \tau_{22} & \tau_{23} \\ \tau_{31} & \tau_{32} & \tau_{33} \end{pmatrix}$$

it follows that

$$-\frac{1}{\sqrt{n}} l^{\varphi\,\prime}(Z_1, \ldots, Z_n\,, \underline{\lambda}) \Sigma_0^{\varphi-1}$$

is equivalent to $\sqrt{n}(\hat{\lambda} - \underline{\lambda})$ hence the claim with

$$\Sigma^\varphi(\underline{\lambda}) \;=\; \mathbf{\Pi}^\varphi \Sigma_0^{\varphi-1}. \qquad \square$$

We define the trace of asymptotic relative efficiency tr. A.R.E. of Σ_{mo} relative to Σ_φ as

$$tr.A.R.E. = \frac{trace\ \Sigma_\varphi}{trace\ \Sigma_{mo}}$$

where $\Sigma_\varphi = \Sigma_0^{\varphi-1}$.

Chapter 5

Multivariate Generalized Linear Models

The generalized linear model (GLM) introduced by Nelder and Wedderburn(1972) and (McCullagh and Nelder 1989) neatly synthesizes many of the most statistical techniques for the analysis of both continuous and discrete data in a unified conceptual and methodological framework. It permits the adaptation of procedures for model building and model checking, originally developed for normal theory of linear regression, for use in a much wider setting.

Definition 5.0.1 *A family of distributions P_θ of a q- dimensional random variable Y, $\theta \in \Theta \subset \mathbb{R}^q$, which have densities*

$$f(y|\theta,\nu) = \exp\{[y\theta' - b(\theta)]/\nu + c(y,\nu)]\} \tag{5.1}$$

($c(.) \geq 0$ measurable,) with respect to a σ-finite measure is called a natural exponential family with natural parameter θ. $\nu > 0$ is a nuisance or dispersion parameter.

The GLM specification according to Fahrmeir and Kaufmann (1985) has the following three components:

i. The random component specifies the probability distribution of the response variables. Specifically, it states that the components of Y have (probability mass function)pmf or pdf from an exponential family of distributions. We let $l(\theta,\nu;y)$ denotes the logarithm of the likelihood function. Since

$$E(\partial l/\partial \theta) = 0$$

$$E(\partial^2 l/\partial \theta^2) + E(\partial l/\partial \theta)^2 = 0$$

and

$$\partial l/\partial \theta = \{y - b'(\theta)\}/\nu$$
$$\partial^2 l/\partial \theta^2 = -b''(\theta)/\nu$$

it follows that $E(Y) = \mu = b''(\theta)$, and $var(Y) = b''(\theta)\nu$. The function $b''(\theta)$ depends on the canonical parameter, and hence on μ, and is called the variance function, denoted by $V(\mu)$.

ii. The systematic component specifies a linear predictor $\eta = X\beta = \sum_{i=1}^{p} x_i\beta_i$, as a function of explanatory variables $X_1, \dots, X_k$ and unknown parameters.

iii. The link function $g(.)$ provides a functional relationship between the systematic component and the expectation of the response in the random component, $\eta = g(\mu)$.

If $g(\mu_i) = \mu_i$, i.e., $\eta_i = \mu_i$, $i = 1, \dots, n$, we call $g(.)$ the identity link function.

5.1 GLM for Minimum model

We apply the method of GLM to the Marshall-Olkin model with parameters of the minimum model. We start by evaluate the EXY for the joint distribution X, Y. To enable us determine what happens at the other halves of the plane, we exclude the case where $X = Y$. We are considering positive random variables, hence the Laplace transform (moment generating function) exists. This transform is given by

$$\psi(s,t) = \int_0^\infty \int_0^\infty \exp\{-sx - ty\} dF_\lambda(x,y) \tag{5.2}$$

$$\overline{F}_\varphi(x,y) = \exp\{-\lambda_1 x - \lambda_2 y - \lambda_3 \varphi(x,y)\};\ x, y > 0$$

The corresponding joint density is given as

$$f(x,y) = \begin{cases} \lambda_2(\lambda_1 + \lambda_3)\overline{F}_\varphi;\ x < y \\ \lambda_1(\lambda_2 + \lambda_3)\overline{F}_\varphi;\ y < x \end{cases} \tag{5.3}$$

using eqn. 5.3 in eqn. 5.2, we have

$$\begin{aligned} \psi(s,t) &= \int\int_{x>y} \exp\{-sx - ty\}\lambda_1(\lambda_2 + \lambda_3)\overline{F}_\varphi dxdy \\ &+ \int\int_{x<y} \exp\{-sx - ty\}\lambda_2(\lambda_1 + \lambda_3)\overline{F}_\varphi dxdy \\ &= I + II \end{aligned} \tag{5.4}$$

Now considering

$$\int\int_{x>y} \exp\{-sx-ty\}\lambda_1(\lambda_2+\lambda_3)\overline{F}_{\varphi}dxdy$$

For

$$\begin{aligned}
I &= \lambda_1(\lambda_2+\lambda_3)\int_{x=0}^{\infty}\int_{y=0}^{x} e^{-(\lambda_1+s)x}\Big\{e^{-(\lambda_2+\lambda_3+t)y}\Big\}dydx \\
&= \lambda_1(\lambda_2+\lambda_3)\int_{x=0}^{\infty} e^{-(\lambda_1+s)x}\Big\{\int_{y=0}^{x} e^{-(\lambda_2+\lambda_3+t)y}dy\Big\}dx \\
&= \lambda_1(\lambda_2+\lambda_3)\int_{x=0}^{\infty} e^{-(\lambda_1+s)x}\Big(-\frac{e^{-(\lambda_2+\lambda_3+t)y}}{(\lambda_2+\lambda_3+t)}\Big)_0^x dx \\
&= \frac{\lambda_1(\lambda_2+\lambda_3)}{(\lambda_2+\lambda_3+t)}\int_{x=0}^{\infty} e^{-(\lambda_1+s)x}\Big(1-e^{-(\lambda_2+\lambda_3+t)x}\Big)dx \\
&= \frac{\lambda_1(\lambda_2+\lambda_3)}{(\lambda_2+\lambda_3+t)}\int_{x=0}^{\infty}\Big(e^{-(\lambda_1+s)x}-e^{-(\lambda+s+t)x}\Big)dx\ ;\ \lambda=\lambda_1+\lambda_2+\lambda_3 \\
&= \frac{\lambda_1(\lambda_2+\lambda_3)}{(\lambda_2+\lambda_3+t)}\Big(-\frac{e^{-(\lambda_1+s)x}}{(\lambda_1+s)}+\frac{e^{-(\lambda+s+t)x}}{(\lambda+s+t)}\Big)_0^{\infty} \\
&= \frac{\lambda_1(\lambda_2+\lambda_3)}{(\lambda_2+\lambda_3+t)}\Big(\frac{1}{(\lambda_1+s)}-\frac{1}{(\lambda+s+t)}\Big) \\
&= \frac{\lambda_1(\lambda_2+\lambda_3)}{(\lambda_1+s)(\lambda+s+t)}
\end{aligned}$$

By symmetry,

$$II=\frac{\lambda_2(\lambda_1+\lambda_3)}{(\lambda_2+t)(\lambda+s+t)}$$

Hence, the moment generating function

$$\psi(s,t)=\frac{\lambda_1(\lambda_2+\lambda_3)}{(\lambda_1+s)(\lambda+s+t)}+\frac{\lambda_2(\lambda_1+\lambda_3)}{(\lambda_2+t)(\lambda+s+t)} \tag{5.5}$$

differentiating eqn. 5.5 with respect to s and t and setting $s=t=0$, we have

$$E(XY)=\frac{(2\lambda_1+2\lambda_3)}{\lambda^3}+\frac{(\lambda_1+\lambda_3)}{\lambda^2\lambda_2}+\frac{(2\lambda_2+2\lambda_3)}{\lambda^3}+\frac{(\lambda_2+\lambda_3)}{\lambda^2\lambda_1}$$

which upon simplification yields;

$$E(XY)=\frac{\lambda^2(\lambda_1+\lambda_2)+2\lambda_1\lambda_2\lambda_3}{\lambda_1\lambda_2\lambda^3} \tag{5.6}$$

Let ρ be the correlation between X and Y, then we can say that

$$EXY=\rho\frac{1}{\lambda_1}\frac{1}{\lambda_2}$$

Let ρ be the correlation between X and Y, then we can say that

$$EXY=\rho\frac{1}{\lambda_1}\frac{1}{\lambda_2}$$

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2 f_1(x)dx \\ &= \int_0^\infty x^2 \lambda_1 \exp(-\lambda_1 x)\, dx \\ &= 2/\lambda_1^2 \end{aligned} \tag{5.7}$$

similarly,

$$E(Y^2) = 2/\lambda_2^2$$

By Cauchy-Schwarz inequality

$$0 \le EXY \le \sqrt{EX^2 EY^2} = 2/(\lambda_1 \lambda_2)$$

thus the correlation EXY is in the range $[0, 2/(\lambda_1\lambda_2)]$.

We consider the model (X_{ij}, Y_{ij}) for $1 \le i \le n$ and $1 \le j \le n_i$ where the random vectors denote the lifetimes of $j-th$ pair organs under the $i-th$ experimental condition. The joint survival distribution of the (X_{ij}, Y_{ij}) for fixed i is denoted by $\overline{F}_{\lambda(i)}$ with parameters $\lambda_1(i), \lambda_2(i)$ and $\lambda_3(i)$. We will denote its corresponding joint density function by $f_{\lambda(i)}(x_i, y_i)$. We will assume for simplicity that only $\lambda_3(i)$ depends on i. Thus,

$$\overline{F}_{\lambda(i)}(x_i, y_i) \;=\; \exp\{-\lambda_1 x_i - \lambda_2 y_i - \lambda_3(i)\varphi_i(x_i, y_i)\}$$

and we consider the model (omitting the index j for simplicity)

$$\begin{aligned} EX_iY_i &= C(\lambda_1, \lambda_2, \lambda_3(i)) \int\int x_i y_i \exp\{-\lambda_1 x_i - \lambda_2 y_i - \lambda_3(i)\varphi_i(x_i, y_i)\} dxdy \\ &= \rho_i \frac{1}{\lambda_1 \lambda_2} \\ &= \frac{1 - e^{-\alpha - \beta\, t_i}}{1 + e^{-\alpha - \beta\, t_i}} \frac{2}{\lambda_1 \lambda_2}\,;\ \beta\ >\ 0 \end{aligned} \tag{5.8}$$

where $\frac{1-e^{-\alpha-\beta\, t_i}}{1+e^{-\alpha-\beta\, t_i}}$ is a link function with predictor $\alpha + \beta t_i$, α and β unknown parameters, t_i models the $i-th$ experimental condition. We have assumed that the model is non-linear in its parameters. When $\varphi_i(x_i, y_i) = \min(x_i, y_i)$, we have from eqn. 5.6 that

$$E(X_iY_i) = \frac{\lambda(i)^2(\lambda_1 + \lambda_2) + 2\lambda_1\lambda_2\lambda(i) - 2\lambda_1\lambda_2(\lambda_1 + \lambda_2)}{\lambda_1\lambda_2\lambda(i)^3}$$

so

$$\frac{1 - e^{-\alpha - \beta\, t_i}}{1 + e^{-\alpha - \beta\, t_i}} \;=\; \frac{\lambda(i)^2(\lambda_1 + \lambda_2) + 2\lambda_1\lambda_2\lambda(i) - 2\lambda_1\lambda_2(\lambda_1 + \lambda_2)}{2\,\lambda(i)^3}$$

Multiplying this equation by $\lambda(i)^3$ and differentiating both side in β yields

$$\frac{d\lambda(i)}{d\beta}\Big(2\lambda(i)(\lambda_1+\lambda_2)+\lambda_1\lambda_2\Big) = 6\lambda(i)^2\frac{1-e^{-\alpha-\beta t_i}}{1+e^{-\alpha-\beta t_i}}\frac{d\lambda(i)}{d\beta}+2\lambda(i)^3\frac{2t_ie^{-\alpha-\beta t_i}}{(1+e^{-\alpha-\beta t_i})^2}.$$

We want to determine the maximum likelihood estimator for β. Note that the model is given by

$$\mathbf{l}(x,y) = \log l(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{n_i} f_{\lambda(i)}(x_{ij},y_{ij})$$

hence differentiating with respect to β and using eqn. 3.5, for the case $\mu = 1$, yields

$$\frac{d\mathbf{l}(x,y)}{d\beta} = \sum_{i=1}^{n}\left(\frac{n_2(i)}{\lambda_1+\lambda_3(i)}+\frac{n_1(i)}{\lambda_2+\lambda_3(i)}+\frac{n_3(i)}{\lambda_3(i)}-\sum_{j=1}^{n_i}\max\{x_{ij},y_{ij}\}\right)\frac{d\lambda_3(i)}{d\beta},$$

where $n_k(i)$ denote the numbers n_k for the $i-th$ sample. Using $\lambda(i) = \lambda_1 + \lambda_2 + \lambda_3(i)$ we can rewrite this expression as

$$\frac{d\mathbf{l}(x,y)}{d\beta} = \sum_{i=1}^{n}\left(\frac{n_2(i)}{\lambda(i)-\lambda_2}+\frac{n_1(i)}{\lambda(i)-\lambda_1}+\frac{n_3(i)}{\lambda(i)-\lambda_1-\lambda_2}-\sum_{j=1}^{n_i}\max\{x_{ij},y_{ij}\}\right)\frac{d\lambda(i)}{d\beta}.$$

This is the maximum likelihood equation to determine the maximum likelihood estimator $\widehat{\beta}$ for the parameter β. In order to show that this is a.s. well defined we calculate the second derivative:

$$\begin{aligned}\frac{d^2\mathbf{l}(x,y)}{d\beta^2} &= -\sum_{i=1}^{n}\left(\frac{n_2(i)}{(\lambda(i)-\lambda_2)^2}+\frac{n_1(i)}{(\lambda(i)-\lambda_1)^2}+\frac{n_3(i)}{(\lambda(i)-\lambda_1-\lambda_2)^2}\right)\left(\frac{d\lambda(i)}{d\beta}\right)^2\\ &+ \sum_{i=1}^{n}\left(\frac{n_2(i)}{\lambda(i)-\lambda_2}+\frac{n_1(i)}{\lambda(i)-\lambda_1}+\frac{n_3(i)}{\lambda(i)-\lambda_1-\lambda_2}-\sum_{j=1}^{n_i}\max\{x_{ij},y_{ij}\}\right)\\ &\times \frac{d^2\lambda(i)}{d\beta^2}.\end{aligned}$$

It has been shown in the proof of Theorem 3.3.1 that

$$E\left(\frac{n_2(i)}{\lambda(i)-\lambda_2}+\frac{n_1(i)}{\lambda(i)-\lambda_1}+\frac{n_3(i)}{\lambda(i)-\lambda_1-\lambda_2}-\sum_{j=1}^{n_i}\max\{x_{ij},y_{ij}\}\right) = 0$$

hence (assuming that all $n_k(i)/n$ stay bounded as $n\to\infty$)

$$\lim_{n\to\infty}\frac{1}{n}\left(\frac{n_2(i)}{\lambda(i)-\lambda_2}+\frac{n_1(i)}{\lambda(i)-\lambda_1}+\frac{n_3(i)}{\lambda(i)-\lambda_1-\lambda_2}-\sum_{j=1}^{n_i}\max\{x_{ij},y_{ij}\}\right) = 0 \quad a.s.$$

Therefore the second derivative is a.s. negative and the maximum likelihood estimator $\widehat{\beta}$ is well defined.

Theorem 5.1.1 *The MLE , $\widehat{\beta}$, of β is consistent and asymptotically normal.*

Proof: We first show that the classical Cramer-Rao cf. Cramer (1946, pages 500 and 501) conditions are satisfied. For each recall that we consider only the case when β is unknown, all other parameters are known. It can be shown in a similar way as above that $\frac{d^k \log f_{\lambda(i)}(x,y)}{d\beta^k}$ exists for $k = 1, 2, 3$ and $\frac{df_{\lambda(i)}(x,y)}{d\beta}$ bounded by integrable function. Since

$$\frac{d^2 f_{\lambda(i)}}{d\beta^2} = \frac{d^2 f_{\lambda(i)}}{d\lambda(i)^2}\left(\frac{d\lambda(i)}{d\beta}\right)^2 + \frac{df_{\lambda(i)}}{d\lambda(i)}\frac{d^2\lambda(i)}{d^2\beta} \tag{5.9}$$

it follows that it is bounded by integrable function for all β. There is a function H satisfying

$$\left|\frac{d^3 \log f_{\lambda(i)}(x,y)}{d\beta^3}\right| < H(x,y) \tag{5.10}$$

with

$$\int H(x,y) f_{\lambda(i)}(x,y)dxdy < \infty \text{ independent of } \lambda(i) \tag{5.11}$$

Expand the function $\frac{d \log f_{\lambda(i)}(x,y)}{d\beta}$ in a Taylor's series around the point $\beta = \beta_0$, where β_0 denotes the unknown true value of the parameter.

$$\begin{aligned}(\frac{d\log f_\lambda}{d\beta})_\beta &= (\frac{d\log f_\lambda}{d\beta})_{\beta_0} + (\beta - \beta_0)(\frac{d}{d\beta}(\frac{d\log f_\lambda}{d\beta}))_{\beta_0} \\ &+ \frac{1}{2}(\beta-\beta_0)^2(\frac{d^2}{d\beta^2}(\frac{d\log f_\lambda}{d\beta}))_{[\gamma(\beta-\beta_0)+\beta_0]}\end{aligned}$$

where $|\gamma(x, y, \beta, \beta_0)| < 1$. Summation of this relation over the n observations and division by n the likelihood equation may be written as

$$\frac{1}{n}\frac{d\log l_\lambda(x,y)}{d\beta} = B_0 + (\beta-\beta_0)B_1 + \frac{1}{2}(\beta-\beta_0)^2 B_2 \tag{5.12}$$

$$B_0 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}(\frac{d\log f_\lambda(x_{ij}, y_{ij})}{d\beta})_{\beta_0}$$

$$B_1 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}(\frac{d}{d\beta}[\frac{d\log f_\lambda(x_{ij}, y_{ij})}{d\beta}])_{\beta_0}$$

$$B_2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n_i}(\frac{d^2}{d\beta^2}(\frac{d\log f_\lambda(x_{ij}, y_{ij})}{d\beta}))_{[\gamma(\beta-\beta_0)+\beta_0]}$$

now, since $E\frac{d\log f_{\lambda(i)}(x_{ij},y_{ij})}{d\beta} = 0$ we obtain

$$\sigma^2 = -E((\frac{d}{d\beta}[\frac{d\log f_{\lambda(i)}(x_{ij},y_{ij})}{d\beta}])_{\beta_0}) = \int_{-\infty}^{\infty}\Big(\frac{d^2 f}{d\beta^2} - (\frac{1}{f}(\frac{d\,f}{d\beta})^2)\Big)_{\beta_0} dxdy < \infty$$

By use of strong law of large numbers, we find that B_0 converges to zero a.s, B_1 to σ^2, while B_2 converges in probability to the non-negative value $EH(x,y) < M$, where M depends on β_0 and $0 < M < \infty$. Given two positive quantities, δ and ϵ, it is then possible to find a number $n_1 = n_1(\delta,\epsilon,\beta_0)$ such that for any $n > n_1$

$$P_1 = P(|B_0| \geq \delta^2) < \frac{1}{3}\epsilon$$

$$P_2 = P(B_1 \geq -\frac{1}{2}\sigma^2) < \frac{1}{3}\epsilon$$

$$P_3 = P(|B_2| \geq 2M) < \frac{1}{3}\epsilon$$

Let Υ denote the set of all points for which the inequalities

$$|B_0| < \delta^2 \,,\; B_1 < -\frac{1}{2}\sigma^2 \,,\; |B_2| < 2M$$

are simultaneously satisfied. Let Υ^* denotes that complementary set of Υ, we have

$$P(\Upsilon^*) \leq P_1 + P_2 + P_3 < \epsilon$$

and hence

$$P(\Upsilon) > 1 - \epsilon$$

if $n > n_1$. Let δ be sufficiently small so that the parameter values $\beta = \beta_0 \pm \delta$. We then have

$$\frac{1}{n}(\frac{d\log l_\lambda(x_{ij},y_{ij})}{d\beta})_{\beta=\beta_0\pm\delta} = B_0 \pm B_1\delta + \frac{1}{2}\delta^2 B_2.$$

For every point in Υ

$$|B_0 + \frac{1}{2}\delta^2 B_2| < \delta^2(1+M)$$

and , if $\delta < \frac{1}{2}\sigma^2/(1+M)$,

$$|\pm\delta B_1| > \frac{1}{2}\sigma^2\delta > \delta^2(1+M)$$

which shows that the sign of the expression $B_0 \pm B_1\delta + \frac{1}{2}\delta^2 B_2$ is determined by the sign of its second term. So that we have $(\frac{d\log l_\lambda(x_{ij},y_{ij})}{d\beta}) > 0$ for $\beta = \beta_0 - \delta$, and $(\frac{d\log l_\lambda(x_{ij},y_{ij})}{d\beta}) < 0$ for $\beta = \beta_0 + \delta$, further by eqn. 5.9 the function $(\frac{d\log l_\lambda(x_{ij},y_{ij})}{d\beta})$ is bounded by integrable function for all β. We can therefore conclude that, when δ is sufficiently small, the root $\widehat{\beta}_n$ of the likelihood equation exists and lies between $\beta_0 - \delta$ and $\beta_0 + \delta$ for every point in Υ. For all $n > n_1$ we have

$$P(|\widehat{\beta}_n - \beta_0| < \delta) \geq P(\Upsilon) > 1 - \epsilon$$

which completes the proof of consistency. Insertion of $\widehat{\beta}_n$ in eqn. 5.12, the likelihood equation for β gives

$$B_0 + (\widehat{\beta}_n - \beta_0)B_1 + \frac{1}{2}(\widehat{\beta}_n - \beta_0)^2 (B_2)_{\beta=\widehat{\beta}_n} = 0 \tag{5.13}$$

from which we obtain

$$\begin{aligned} \sqrt{n}(\widehat{\beta}_n - \beta_0) &= \Big[\frac{1}{\sigma^2\sqrt{n}} \sum_{i=1}^{n} (\frac{d \log f_{\lambda(i)}(x_{ij}, y_{ij})}{d\beta})_{\beta=\beta_0}\Big] \\ &/ \Big[\frac{-B_1}{\sigma^2} - \frac{(\widehat{\beta}_n - \beta_0)(B_2)_{\beta=\widehat{\beta}_n}}{2\sigma^2}\Big] \end{aligned} \tag{5.14}$$

In eqn. 5.14, the denominator of the fraction on the right-hand side converges in probability to unity, while the numerator is asymptotically normal with mean zero and variance $\frac{1}{\sigma^2}$. It then follows from convergence theorem by (Cramer, page 254) that $\sqrt{n}(\widehat{\beta}_n - \beta_0)$ is asymptotically normal with mean zero and variance $\frac{1}{\sigma^2}$. □

List of Notations

X	Random variable
Y	Random variable
$F(x, y)$	Bivariate distribution function of continuous random variables X and Y
$\overline{F}(x, y)$	Joint survival distribution function for X and Y
$\overline{F}_1(x)$	Marginal distribution function of X
$\overline{F}_2(y)$	Marginal distribution function of Y
$F_1(x)$	One-dimensional distribution function of X
$F_2(y)$	One-dimensional distribution function of Y
$\overline{F}_{mo}(x, y)$	Joint survival distribution function of Marshall-Olkin's BVE
$\overline{F}_{mo1}(x)$	Marginal distribution function of X in the Marshall-Olkin's BVE
$\overline{F}_{mo2}(y)$	Marginal distribution function of Y in the Marshall-Olkin's BVE
$\overline{F}^M_\lambda(x, y)$	Joint survival distribution function of the minimum model
$\overline{F}^M_\lambda(x)$	Marginal distribution of X in the minimum model
$\overline{F}^M_\lambda(y)$	Marginal distribution of Y in the minimum model
$f(x, y)$	Joint density function for X and Y
$f_{mo1}(x)$	Probability density function of X in the Marshall-Olkin's BVE
$f_{mo2}(y)$	Probability density function of X in the Marshall-Olkin's BVE
$\overline{F}_\lambda(x, y)$	Joint survival distribution function of the Generalized Marshall-Olkin model
$f_G(x, y)$	Joint density function of the Generalized Marshall-Olkin model on the domain G
$f_L(x, y)$	Joint density function of the Generalized Marshall-Olkin model on the line L
$\overline{F}_\varphi(x, y)$	Joint survival distribution function of the φ model
Λ^+	Parameter space for the minimum model
Λ	Parameter space for the Generalized Marshall-Olkin model
$\bar{\Lambda}$	Parameter space for the φ model

Bibliography

[1] Arnold, B.C. (1968). Parametric Estimation for a Multivariate Exponential Distribution. *J. Amer. Statist. Assoc.* **63**, 848-852.

[2] Atienza, N., Garcia-Heras, J., Munoz-Pichardo, J.M. and Villa, R. (2007). On the Consistency of MLE in the finite Mixture Models of Exponential Families. *J. Statist. Plann. Inf.* **137**, 496-505.

[3] Bahadur, R. R. (1960) "On the Asymptotic Efficiency of Tests and Estimates," *Sankhya* **22** No. 3-4, 229-252.

[4] Barlow, R.E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing Probability Models. *Holt, Rinehart and Winston, New York.*

[5] Bemis, B.M.(1971). Some Statistical Inference for the Bivariate Exponential Distribution. *Unpublished Ph.D dissertation, University of Missouri-Rolla.*

[6] Bemis, B.M., Bain, L. J., and Higgins, J. J.(1972). Estimation and Hypothesis Testing for the Parameters of a Bivariate Exponential Distribution. *J. Amer. Statist. Assoc.* **67**, 927-929.

[7] Bhattacharyya, G.K. and Johnson, R.A. (1971). Maximum Likelihood Estimation and Hypothesis Testing in the Bivariate Exponential Model of Marshall and Olkin. *Technical Report No. 276, Department of Statistics, University of Wisconsin.*

[8] Block, H. W. and Basu, A. P. (1974). A Continuous Bivariate Exponential Extension. *J. Amer. Statist. Assoc.* **69**, 1031-1037.

[9] Chanda, S. (1954). A notes on the consistency and maxima of the roots of the likelihood equations. *Biometrika.* **41**, 56-61.

[10] Cramer, H. (1946). Mathematical Methods of Statistics. *Princeton University Press.*

[11] Daniels, H. E. (1961). The asymptotic efficiency of a maximum likelihood estimator. *Proc. Fourth Berkley Symp. Math. Statist and Prob.* **1**, 151-164.

[12] Daniels, M. J. and Zhao, Y. D. (2003). Modelling the Random Effects Covariance Matrix in Longitudinal Data. *Statist. in Medicine.* **22**, 1631-1647.

[13] Dugue, J. L. (1937). Application des propiétés de la limite au sens du calcul des probabilités à l'étude de diverses questions d'estimation. *Journ. de l'Éc. Polytechn.* , 305.

[14] Fahrmeir, L. and Kaufmann, H. (1985). Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models. *Ann. Statist.* **13, No. 1**, 342-368.

[15] Fisher, R. A. (1912). On an absolute criterion for fitting frequency curves. *Mess. Of Math.* **41**, 155.

[16] Freund, J. E. (1961). A Bivariate Extension of the Exponential Distribution. *J. Amer. Statist. Assoc.* **56**, 971-977.

[17] Gibbons, R. D. and Hedeker, D. (1997). Random Effects Probit and Logistic Regression Models for three-level data. *Biometrics.* **53**, 1527-1537.

[18] Gumbel, E. J.(1960). A Bivariate Extension of the Exponential Distribution. *J. Amer. Statist. Assoc.* **56**, 971-977.

[19] Gurland, J. (1954). " On regularity conditions for maximum likelihood estimators". *Skand. Aktuar.* **37**, 71-76.

[20] Hathaway, R. J.,. (1985). A constrained formulation of maximum likelihood estimation for normal mixture distributions. *Ann. Statist.* **13**, 795-800.

[21] Heagerty, P. J. (1999). Marginally Specified Logistic-Normal Models for Longitudinal Binary Data. *Biometrics.* **55**, 688-698.

[22] Heagerty, P. J. and Zeger, S. L. (1996). Marginal Regression Models for Clustered Ordinal Measurements. *J. Amer. Statist. Assoc.* **91**, 1024-1036.

[23] Hedeker, D. and Gibbons, R. (1994). A Random- Effects Ordinal Regression Model for Multilevel Analysis. *Biometrics.* **50**, 933-944.

[24] Hyakutake, H. (1990). Statistical Inference on location parameters of Bivariate Exponential Distributions. *Hiroshima Math. J.* **20**, 525-547.

[25] Jewell, N. P., (1982). Mixtures of Exponential Distributions. *Ann. Statist.* **10(2)**, 479-484.

[26] Johnson, N. L. and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. *Wiley, New York.*

[27] Kulldorf, G. (1949). " On the conditions for consistency and asymptotic efficiency of maximum likelihood estimators". *Skand. Aktuar.* **40**, 130-144.

[28] Le Cam, L. (1955). " On Some Asymptotic Theory of Estimation and Testing Hypotheses". *Proc. Third Berkley Symp. Math. Statist and Prob.* **1**, 129-156.

[29] Leroux, S. (1992). Consistent Estimation of a Mixing Distribution. *Ann. Statist.* **20(3)**, 1350-1360.

[30] Liang, K. Y. and Zeger, S. L. (1986). Longitudinal Data Analysis using Generalized Linear Models. *Biometrika* **73**, 13-22.

[31] Lipsitz, S. R., and Ibrahim, J. G. (1996). A Conditional Model for Incomplete Covariates in Parametric Regression Models. *Biometrika* **83**, 916-922.

[32] Mangasarin, O. L. (1969). Nonlinear Programming New York: *McGraw-Hill Book Co.*

[33] Marshall, A.W, and Olkin, I.(1967). A Multivariate Exponential Distribution. *J. Amer. Statist. Assoc.* **62**, 30-44.

[34] Marshall, A.W, and Olkin, I.(1967(b)). A Generalized Bivariate Exponential Distribution. *J. Appl. Prob.* **4**, 291-302.

[35] McCullagh, P. and Nelder, J. A.(1989). Generalized Linear Models. *2nd ed. London: Chapman and Hall.*

[36] Moeschberger, M. L. (1974). Life Test Under Dependent Competing Causes of Failure. *Technometrics* **16**, 39-47.

[37] Molenberghs, G. and Lesaffre, E. (1994). Marginal Modeling of Correlated Ordinal Data using a Multivariate Plackett Distribution. *J. Amer. Statist. Assoc.* **89**, 633-644.

[38] Moran, P. A. P. (1967). Testing for Correlation between non- negative variates. *Biometrika.* **54**, 385-3941.

[39] Muliere, P. and Scarsini, M. (1987). Characterization of a Marshall-Olkin type distributions. *Ann. Inst. Statist. Math.* **39**, 429-441.

[40] Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models. *J. Roy. Statist. Soc. Ser. A***135**, 370-384.

[41] Peters, B. C., Walker, H. F.,(1978). An Iterative Procedure for obtaining Maximum-Likelihood Estimates of the parameters for a Mixture of Normal Distributions. *SIAM J. Appl. Math.* **35**, 362-378.

[42] Pfanzagl, S. (1988). Consistency and Maximum Likelihood Estimators for certain Nonparametric Ffamilies, in particular: Mixtures. *J. Statist. Plann. Inf.***19**, 137-158.

[43] Platz, O. (1984). A Markov Model for Common-Cause Failures. *Reliability Engineering.* **9**, 25-31.

[44] Proschan, F. and Sullo, P. (1976). Estimating the Parameters of a Multivariate Exponential Distribution. *J. Amer. Statist. Assoc.* **71**, 465-471.

[45] Rao, C. R. (1973). Linear Statistical Inference and Its Applications. *New York: Wiley. 2nd Edition.*

[46] Ryu, K. (1993). An Extension of Marshall-Olkin's Bivariate Exponential Distribution. *J. Amer. Statist. Assoc.* **88**, 1458-1465.

[47] Sklar, A.(1996). "Random variables, distribution functions, and copulas-a personal look backward and forward," in Distributions with Fixed Marginals and Related Topics, *ed. L. Rueschendorff, B. Schweizer, and M. Taylor Institute of Mathematical Statistics, Hayward, CA. 1-14.*

[48] Van De Geer, S. (2003). Asymptotic theory for maximum likelihood in nonparametric mixture models. *Comput. Statist. Data Anal.* **41**, 453-464.

[49] Wald, A. (1949). Notes on the consistency of the maximum likelihood estimates. *Ann. Math. Statist.* **20**, 595-600.

[50] Wilks, S. S. (1962). Mathematical Statistics. *New York: Wiley. 2nd Edition.*

[51] Wolfowitz, J. (1949). " On Wald's proof of the Consistency of the Maximum Likelihood Estimates". *Ann. Math. Statist.* **20**, 601-602.

[52] Zeger, S. L. (1988). A Regression Model for Time Series of Counts. *Biometrika.* **75**, 621-629.

[53] Zeger, S. L., Diggle, P. J., and Yasui, Y. (1990). Marginal Regression Models for Time Series. *Inst. for Mathematics and Its Appl, Time Series Workshop, Preprint.*

[54] Zeger, S. L., Liang, K.-Y. and Self, S. G. (1985). The Analysis of Binary Longitudinal Data with time independent covariates. *Biometrika.* **72**, 31-38.

[55] Zeger, S. L. and Karim, M. R. (1991). Generalized Linear Models with random Effects: A Gibbs Sampling Approach. *J. Amer. Statist. Assoc.* **86**, 79-86.

[56] Zeger, S. L. and Qaqish, B. (1988). Markov Regression Models for Time Series: A Quasi-Likelihood Approach. *Biometrics.* **44**, 1019-1031.

Curriculum Vitae:

Name	OKYERE Ebenezer
Date/Place of birth	1/9/1967, Agona Nsaba, Ghana
Nationality	Ghanaian
Sept. 1980 - June 1985	G.C.E. Ordinary Levels, Nsaba Sec. School
Sept. 1985 - June 1987	G.C.E. Advanced Levels, Swedru Sec. School
Sept. 1990 - June 1994	BSc(Hons) Mathematics and Statistics(Combined), Univ. of Ghana, Legon
April 2001 - May 2003	MSc. Mathematics, Georg-August University, Goettingen
Sept. 2003 - July 2007	PhD in Statistics, Georg-August University, Goettingen

ACKNOWLEDGMENTS

I am extremely grateful to my advisors, Prof. Dr. Manfred Denker and Prof. Dr. Jochem Mau for all their help and guidance during the course of my graduate studies. I am especially thankful to Prof. Dr. Mau for providing a wonderful work environment here at the Institute for Statistics in Medicine (ISM), Duesseldorf. The autonomy and the resources provided to me have been unmatched. I have the highest appreciation for the manner in which he has encouraged and supported me at all times. I am equally indebted to Prof. Dr. Denker for his unflinching confidence in me and for all the valuable time he spared from his busy schedule for the innumerable discussions we had. His wisdom and humour have been a source of motivation and amazement for me. Without their constant support, motivation and useful suggestions, this project would never have taken its present shape.I would like to thank Prof. Dr. Brunner, Prof. Dr. Schlather, Prof. Patterson and Prof. Dr. Waack for being part of the examination committee.

I would also like to thank my colleagues of ISM, Dr. Yong, Pharm, Li, Feuersenger, Dr. Drabik and Jahavel for creating a wonderful work atmosphere. I also thank my friends, Dr. and Mrs. Kwabena Adusei-Poku, Daplah family and Christiana Essiaw for making my stay at Duesseldorf and Germany as a whole one of the most enjoyable and memorable experiences of my life.

Finally, words cannot describe the love, support, advice and encouragement that I have received from my family (Mr. and Mrs. Okyere, Fred, George, Alex, Sarah, Gladys and Norah). My wife, Connie, has always been a constant source of energy in my life. I thank her for her unfailing and loving support and for her boundless patience during the course of this work. As for my parents, Mr. and Mrs. Okyere, their love, support, prayers and motivation have been the very backbone of my existence. All that I have achieved today is a result of the innumerable silent sacrifices they have made over the past decades.