Title

Inference for the Bivariate and Multivariate Hidden Truncated Pareto(type II) and Pareto(type IV) Distribution and Some Measures Of Divergence Related To Incompatibility of Probability Distribution.

Author

Ghosh, Indranil

Publication Date

2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Inference for the Bivariate and Multivariate Hidden Truncated Pareto(type II) and Pareto(type IV) Distribution and Some Measures of Divergence Related to Incompatibility of Probability Distribution

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Indranil Ghosh

August 2011

Dissertation Committee:
Prof. Barry C. Arnold, Chairperson
Prof. D.V. Gokhale
Prof. Aman Ullah

The Dissertation of Indranil Ghosh is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I would like to offer my sincere gratitude to my advisor Dr.Arnold for his continuous encouragement, patience and guidance during my research. Dr.Arnold has been instrumental in ensuring my academic, professional, financial, and moral well being ever since I joined the Ph.D program in Statistics in our very nice and supportive department. In every sense, none of this work would have been possible without him. I had enjoyed an incredible experience working with him. Also I would like to offer my sincere thanks to my other two committee members Dr.Ullah and Dr.Gokhale for their constructive criticism in a positive way and their invaluable intellectual input which has helped me a lot in preparing and also improving the content of my dissertation, thanks a lot.

Far too many people to mention individually have assisted in so many ways during my work at UC Riverside. They all have my sincere gratitude. In particular, I would like to thank Dr.Jeske, our department chair for his continued support showering upon me despite by far the toughest economic crisis and also providing me all the care academically that a graduate student may expect; my deepest gratitude. Also I would like to thank Dr.Linda Penas for her encouragement and invaluable suggestions not only to be a good teacher inside a class but also to become an effective member in a teaching community and for facilitating the use of the labs. I wish to thank the Statistics department stuff; specially Perla and Paula for all their support. A penultimate thank-you goes to my wonderful parents. For always being there when I needed them most, and never once complaining about how infrequently I visit, they deserve far more credit than I can ever give them. My final, and most heartfelt, acknowledgment must go to my would be wife Debdutta. Debdutta has supported me diligently, and successfully, for the last few months when the situation was very demanding, living far away from me. Her support, encouragement, and companionship has turned my journey through graduate school into a pleasing experience. For all that, and for being everything I am not, she has my everlasting love.

ABSTRACT OF THE DISSERTATION

Inference for the Bivariate and Multivariate Hidden Truncated Pareto(type II) and Pareto(type IV) Distribution and Some Measures of Divergence Related to Incompatibility of Probability Distribution

by

Indranil Ghosh

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, August 2011
Prof. Barry C. Arnold, Chairperson

Consider a discrete bivariate random variable (X, Y) with possible values $x_1, x_2, ..., x_I$ for X and $y_1, y_2, ..., y_J$ for Y. Further suppose that the corresponding families of conditional distributions, for X given values of Y and of Y for given values of X are available. We specifically consider those situations where the above mentioned conditional distributions are not compatible. In such a case we seek a joint probability matrix (P), say, that is minimally incompatible with the given conditional distributions. The Kullback-Leibler Information function provides a convenient measure of (pseudo) distance between two distributions. However we will use a more general measure which is called

the "power divergence criterion" which includes the Kullback-Leibler Function as a special case. We will, along with this measure, also consider some other measures of diversity which are widely used in the field of Information Theory. Using all these measures we have developed algorithms for finding the joint probability matrix which is minimally incompatible with the given conditionals. We will also propose here some alternative measures of compatibility related to discrepancy. Our main objective here is to find among the various measures of discrepancy (for example the power divergence test statistic, modified Renyi's measure etc.), along with the proposed measures, which one give us the minimally incompatible distribution with a faster convergence rate. This topic will be discussed in detail in chapter 1.

Next we consider an alternative approach to determine whether or not any two given matrices (say, A and B) with non negative elements (where for the matrix A, the column sums add up to one and for the matrix B, the row sums add up to one) are compatible in the sense that there exists a joint probability matrix for which has the columns and rows, respectively, of A and B as its conditional distributions. We formulate the above problem as a homogeneous and consistent set of equations and consider the LPP (Linear Programming Problem) approach to solve for the unknown quantities. Furthermore we will also discuss briefly, under the condition of compatibility, how can we find some of the elements of the two given conditional matrices A and B, in case they are unknown. This is the subject matter for chapter 2.

Next we consider the hidden truncation paradigm for the bivariate Pareto (type II) distribution when one variable is subject to hidden truncation from above. We consider the classical method of estimation and a reasonable testing procedure for the truncation parameter and also the other parameters involved in the model along with an application of the above mentioned model to a real life data. We will also focus on the estimation procedure under the Bayesian paradigm. This will be the subject matter in chapter 3.

In chapter 4 we will consider the hidden truncation paradigm for a bivariate Pareto (type IV) distribution where both the marginals as well as the conditionals are again members of the Pareto (type IV) family. Here also we will consider inference for such a distribution under the classical approach as well as the Bayesian approach along with an application to a real life data.

Next we will consider a possible extension of hidden truncation concept to the multivariate case for the Pareto (type II) family. In particular we will focus on single variable truncation as well as more than one variable truncation. We will discuss about the tractability of such type of models in the context of estimation and testing for the parameters involved in the model. In particular we focus on a trivariate Pareto (type II) set-up and we will consider estimation procedures under both the classical and Bayesian paradigm. Two specific situations are dealt with: (1) when one of the concomitant variables is truncated from above and (2) when more than one variable is truncated from above. This material is discussed in detail in chapter 5. In chapter 6 we provide a general discussion of the topics which are covered in chapter 1 through 5, together with a brief discussion of potential future work.

Contents

List of Figures

List of Tables

Chapter 1

Study of incompatibility of bivariate discrete conditional probability distributions

1.1 Introduction

In an effort to specify bivariate probability models, one is frequently obstructed by an inability to visualize the implications of assuming that a given bivariate family of densities will contain a member which will adequately describe the given phenomenon. One of the main difficulties encountered while using probabilistic models to solve real-life problems is the selection of an appropriate model to reflect the reality being observed. One possibility consists of selecting one of the well-known parametric families of distributions to approximately fit the observed data. However, those so called "well known" families are too simple in the sense that they depend only on a limited number of parameters, and may not be adequate to model the observed phenomena. In such a situation one might consider the idea of conditional specification. Specification of joint distributions by means of conditional densities has received considerable attention over the years by authors such as Dawid (1979, 1980), and Gelman and Speed (1993, 1999). Arnold et al. (1999) have discussed this problem for a wide range of families of distributions including exponential families. Such models can be useful in situations such as model building in classical statistical settings and in the elicitation and consideration and construction of multiparameter prior distributions in a Bayesian framework. One of the problems of defining joint densities by specifying their conditionals is the compatibility problem. For example, one possible approach to the specification of the distribution of a two-dimensional random variable (X, Y) involves presenting both families of conditionals (X given Y and of Y given X). However the consistency of both the conditional distributions must be checked (see Arnold and Press, 1989), to determine if any joint distribution exists with them as its conditional distributions. Several alternative approaches exist in the literature with regard to the problem of determining the possible compatibility of two families of conditional distributions (Arnold and Press, 1989; Arnold and Gokhale

1994). Also in addition to it, the problem of determination of most nearly compatible (ε-compatible, as introduced by authors Arnold and Gokhale (1994, 1998)) has been addressed. Moreover, the compatibility problem also arises when there exists more than one expert participating in the model selection process, or when we have partial information about conditional probabilities.

We consider the question of compatibility and near compatibility of given families of conditional distributions in the finite discrete case. In the finite discrete case, there exists a variety of compatibility conditions (Arnold, Castillo and Sarabia 1999). Based on those conditions the above mentioned authors have provided a broad spectrum of alternative ways of measuring discrepancy between incompatible conditionals. In addition, they have made suggestions of alternative ways in which most nearly compatible distributions can be defined in incompatible cases. In this chapter we focus on the measurement of incompatibility in situations when we are given two families of conditional distributions under the discrete set up which are not compatible. How can we find a distribution P that is, in some sense, minimally incompatible with the given conditional specifications? Such questions are of interest from a Bayesian viewpoint in the context of elicitation of joint prior distributions. For example in the case of a two-dimensional parameter $\underline{\delta}$, our well informed expert might give conditional probabilities for δ_1 given particular choices of values for δ_2 and conditional probabilities for δ_2 for given particular choices of values for δ_1. If our expert is human, it is quite possible that the collection of conditional probabilities thus elicited might be incompatible. A suitable choice of prior to use in subsequent analysis might then be that joint distribution $f(\delta_1, \delta_2)$ that is the least at variance with the given elicited conditional probabilities. More generally, we might think of obtaining partial or complete conditional specification from more than one expert. Such information would most likely lack consistency and, again a minimally discrepant distribution might be sought. The Kullback-Leibler information function provides a convenient discrepancy measure in such settings. As we shall see, not only does it provide a discrepancy measure but, using it, a straightforward algorithm can be described which will result in the most nearly compatible distribution. We will mention some alternatives to Kullback-Leibler measure and we will also consider the relative performance of all those measures.

The remainder of this chapter is organized in the following way. In Section 2, we will consider the concept of compatible distributions. In Section 3, we consider the conditions for compatibility with some examples. In Section 4, a detailed discussion on the existing methods for finding minimally compatible distribution has been provided. In Section 5, we focus on considering some measures of divergence which are already in the literature (Kullback 1959; Renyi 1961). In Section 6, we will propose some new measures of divergence for checking compatibility under the condition that the given two families of conditional probability distributions are incompatible along with an iterative study for all the measures of divergence. In Section 7, some comments have been made about all those divergence measures mentioned earlier.

1.2 Compatible distributions

Let us consider a two dimensional random vector (X, Y) with possible values $x_1, x_2, \ldots, x_I$ and $y_1, y_2, \ldots, y_J$, for X and Y respectively. Further let A and B denote two $(I \times J)$ matrices with non-negative elements and with at least one positive entry in each row and each column. We make the assumption that A has columns which sum to 1 while B has rows which add up to 1. Note that throughout this chapter we will always assume that A has columns summing to 1 and B has rows summing to 1, whenever they will appear. Then A and B are said to form a compatible conditional specification for the distribution of (X, Y) if there exists some $(I \times J)$ matrix P with non-negative entries

p_{ij} and with $\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij} = 1$, such that

$a_{ij} = \frac{p_{ij}}{p_{.j}}, \forall(i,j)$ and $b_{ij} = \frac{p_{ij}}{p_{i.}}, \forall(i,j)$, where $p_{i.} = \sum_{j=1}^{J} p_{ij}$ and $p_{.j} = \sum_{i=1}^{I} p_{ij}$. If such a matrix P exists then, if we assume that

$p_{ij} = P(X = x_i, Y = y_j), \forall(i,j)$,

we will have $a_{ij} = P(X = x_i | Y = y_j), \forall(i,j)$,

and

$b_{ij} = P(Y = y_j | X = x_i), \forall(i,j)$.

Thus A and B are compatible if there exists a joint distribution (P) which has the columns and rows respectively of A and B as its conditional distributions.

One obvious requirement for compatibility is that A and B should have identical incidence sets. The incidence set of a matrix A is

$$\{(i,j) : a_{ij} > 0\},$$

the set of locations of non-zero entries in the matrix. We shall denote the common incidence set by $N = N^A = N^B$ and will usually assume that A and B have this common incidence property. Otherwise they are incompatible.

1.3 Compatibility conditions

Conditions for compatibility are listed in the following theorems:

Theorem 1 (Arnold and Press 1989). *Supposing that A and B have identical incidence sets then they are compatible if and only if either of the following two conditions hold:*

- *There exist non negative vectors*

  $$\underline{\tau} = (\tau_1, \tau_2, \ldots, \tau_I) \textit{ and } \underline{\eta} = (\eta_1, \eta_2, \ldots, \eta_J)$$

  such that $\eta_j a_{ij} = \tau_i b_{ij}, \forall (i,j)$.

  In the case of compatibility, the vectors $\underline{\tau}$ and $\underline{\eta}$ can readily be interpreted as being proportional to the marginal distributions of X and Y respectively.

- *There exist vectors $\underline{u}$ and $\underline{v}$ of appropriate dimension for which $d_{ij} = \frac{a_{ij}}{b_{ij}} = u_i v_j$, $\forall (i,j) \in N$.*

If $N = (1, 2, \ldots, I) \times (1, 2, \ldots, J)$, i.e., if all the entries in A and B are positive, then we have the following theorem given by (Arnold and Gokhale (1994)).

Theorem 2 (Arnold,Gokhale 1994).

- *A and B are compatible iff they have identical uniform marginal representations (UMRs) (Mosteller 1968).*

- *A and B are compatible iff all cross product ratios of A are identical to those of B.*

Note: Some restrictions on the common incidence set of A and B is necessary for the above theorem. For example if we consider

$$A = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}, \text{ and } B = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 0 & 1/3 & 2/3 \\ 2/3 & 0 & 1/3 \end{pmatrix},$$

then it may be verified here that A and B have equal cross product ratios(there are no positive 2×2 submatrices)and have identical uniform marginal representations but A and B are not compatible.

Compatibility of A and B of course does not confirm existence of a unique compatible matrix P. The simplest sufficient condition is positivity (i.e., $a_{ij} > 0$ and $b_{ij} > 0$, $\forall (i, j)$.) For example we can illustrate by examples how things can change when A and B contain zero elements.

Example[1]:No cross-product ratios and incompatible

Let us consider,

$$A = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \\ 1/2 & 0 & 1/2 \end{pmatrix}, \text{ and } B = \begin{pmatrix} 1/3 & 3/4 & 0 \\ 0 & 1/4 & 3/5 \\ 2/3 & 0 & 2/5 \end{pmatrix}.$$

No compatible P exists in this case.

Example[2]: (UMRs do not exist but compatible)

$$A = \begin{pmatrix} 0 & 1/3 & 0 \\ 1 & 0 & 1 \\ 0 & 2/3 & 0 \end{pmatrix}, \text{ and } B = \begin{pmatrix} 0 & 1 & 0 \\ 1/5 & 0 & 4/5 \\ 0 & 1 & 0 \end{pmatrix}.$$

So the joint probability distribution in this case is given by

$$P = \begin{pmatrix} 0 & 1/8 & 0 \\ 1/8 & 0 & 4/8 \\ 0 & 2/8 & 0 \end{pmatrix}.$$

Example[3]:(UMRs exist and are equal, but incompatible).

$$A = \begin{pmatrix} 0 & 1/3 & 0 \\ 1 & 1/3 & 1/2 \\ 0 & 1/3 & 1/2 \end{pmatrix}, \text{ and } B = \begin{pmatrix} 0 & 1 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/5 & 4/5 \end{pmatrix}.$$

Here we have

$$UMR(A) = UMR(B) = \begin{pmatrix} 0 & 1/3 & 0 \\ 1/3 & 0 & 0 \\ 0 & 0 & 1/3 \end{pmatrix}.$$

1.4 Existing methods for finding minimally Incompatible distributions

1.4.1 Idea of minimal incompatibility

The idea behind the concept of minimal incompatibility of two conditional distributions can be described by the following two concepts.

1.4.2 ε-compatibility

Suppose that we do not insist on precise compatibility. Instead, suppose that we wish to have p_{ij} approximately consistent with two given conditional probability matrices A and B. Let us introduce a weight matrix W which will represent the relative importance of accuracy in determining p_{ij} for each (i, j). For a given weight matrix W which might be uniform, i.e., we might choose $w_{ij} = 1, \forall(i, j)$ if all pairs (i, j) were equally important to us, we then consider the following strategies, written as non-linear and linear programming problems:

- **First Method**:

Find a matrix P, with $p_{ij} \geq 0, \forall(i,j)$ such that

$$\begin{aligned} |p_{ij} - a_{ij} \sum_{i=1}^{I} p_{ij}| \leq \varepsilon w_{ij}, \forall(i,j) \in N \\ |p_{ij} - b_{ij} \sum_{j=1}^{J} p_{ij}| \leq \varepsilon w_{ij}, \forall(i,j) \in N, \end{aligned} \tag{1.1}$$

with the linear constraint $\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} = 1$.

- **Second Method**:

  Seek two probability vectors $\underline{\eta}$ and $\underline{\tau}$ such that

$$|a_{ij}\eta_j - b_{ij}\tau_i| \leq \varepsilon w_{ij}; \forall(i,j), \sum_j \eta_j = 1, \sum_i \tau_i = 1, \tag{1.2}$$

  and $\tau_i \geq 0$, $\eta_j \geq 0$, $\forall(i,j) \in N$.

- **Third Method**:

  Find one (marginal) probability vector $\underline{\tau} \geq \underline{0}$ such that

$$|a_{ij} \sum_{k=1}^{I} b_{kj}\tau_k - b_{ij}\tau_i| \leq \varepsilon w_{ij}; \forall(i,j) \in N, \sum_i \eta_i = 1, \tag{1.3}$$

  and $\tau_i \geq 0, \forall i$.

- The above methods motivate three different concepts of ε -compatibility.

- If we use method 1, and if A and B are ε -compatible, then the matrix P^* which satisfies (1.1), will be most nearly compatible.

If we use method 2, and if A and B are ε -compatible, then a reasonable choice for a most nearly compatible matrix P^* will be

$$P^* = \frac{a_{ij}\eta_j^* + b_{ij}\tau_i^*}{2},$$

where $\underline{\eta}^*, \underline{\tau}^*$ satisfy (1.2). Finally if we use method 3 and if A and B are ε -compatible, then a plausible choice for a most nearly compatible P^* will be $P^* = (b_{ij}\tau_i^*)$, where τ_i^* satisfies (1.3).

1.4.3 Kullback-Leibler Measure of Incompatibility

As mentioned earlier our main focus is concentrated on situations in which the conditional specifications are incompatible. The very first choice of a measure of discrepancy which was suggested by Arnold and Gokhale (1994) is the Kullback-Leibler Information function which provides a convenient measure of pseudo distance between distributions. In the case of an incompatible set-up we are actually looking for a matrix P with non-negative elements that add up to 1 and which has conditionals as close as possible to those given by A and B. So we are seeking $P_{I\times J} = (p_{ij})$ with $\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij} = 1$ and with $\frac{p_{ij}}{\sum_{i=1}^{I} p_{ij}} \approx a_{ij}, \forall(i,j)$, and $\frac{p_{ij}}{\sum_{j=1}^{J} p_{ij}} \approx b_{ij}, \forall(i,j)$.

Using the above mentioned measure, a reasonable strategy for selection of a minimally incompatible P is to minimize the following objective function

$$\sum_{i=1}^{I}\sum_{j=1}^{J} b_{ij}\log(\frac{b_{ij}p_{i.}}{p_{ij}}) + \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}\log(\frac{a_{ij}p_{.j}}{p_{ij}}). \tag{1.4}$$

It is convenient here to introduce a new matrix

$$C = A + B,$$

with elements $c_{ij} = a_{ij} + b_{ij}$. In order to ensure that a unique minimizing choice of P exists for the objective function in equation (1.1), it is necessary to have some assumptions about the incidence set of the matrix C. For example C must not be block diagonal. Also it is assumed that some powers of C, perhaps C itself have all elements strictly positive. Next we have the following theorem due to Arnold and Gokhale (1998).

Theorem 3 (Arnold and Gokhale (1998))**.** *Denote by P^* the choice of P which minimizes equation(1.1). Then P^* must satisfy the following system of equations:*

$$\frac{p^*_{ij}}{p^*_{i.}} + \frac{p^*_{ij}}{p^*_{.j}} = a_{ij} + b_{ij}, i = 1, 2, \ldots, I, and j = 1, 2, \ldots, J.$$

The above expression is obtained by differentiating equation(1.1) with respect to p_{ij} *using a Lagrangian multiplier for the linear constraint* $\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij} = 1.$

An iterative algorithm to solve the above equation is proposed as follows

$$p_{ij}^{n+1} = \frac{\frac{a_{ij}+b_{ij}}{\frac{1}{p_{i.}^{n}}+\frac{1}{p_{.j}^{n}}}}{\sum_{i=1}^{I}\sum_{j=1}^{J}\left[\frac{a_{ij}+b_{ij}}{\frac{1}{p_{i.}^{n}}+\frac{1}{p_{.j}^{n}}}\right]}. \tag{1.5}$$

For an initial guess $p_{ij}^{(0)} = \frac{1}{IJ}$, $i = 1, 2, \ldots, I$ *and* $j = 1, 2, \ldots, J$. *The process is convergent.*

1.4.4 Power Divergence Statistic as a measure of divergence

Here we consider under the discrete set up the power divergence statistic to select a minimally incompatible P, from two given conditional probability matrices.

1.4.5 Why should we consider this?

A divergence measure between two probability distributions is such that given any two probability distributions $\underline{p}$ and $\underline{q}$ (which are of the same dimension), it will return a measure of the similarity or distance between them and it is non negative. Again we know that a diversity index can be considered to measure the divergence between the population distribution $\underline{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$ and the uniform distribution $(\frac{1}{k}, \ldots, \frac{1}{k})$, where an index closer to zero represents a wider divergence from the uniform distribution. A natural generalization, when considered in this way, is to define a measure of the divergence between two general distributions. However this concept was first considered by Kullback in his directed divergence measure (1959). This form was actually considered by (Arnold and Gokhale (1994, 1998)) while considering minimum incompatibility via the Kullback -Leibler criterion. It is of the form

$$K(\underline{p} : \underline{q}) = \sum_{i=1}^{k} p_i \log_2(\frac{p_i}{q_i}),$$

where $\underline{p}$ and $\underline{q}$ are two discrete probability distributions defined on the $(k-1)$ dimensional simplex $\Delta_k = \underline{\pi} : \pi_i \geq 0; i = 1, .., k; \sum_{i=1}^{k} \pi_i = 1$. Here we adopt the convention that $p_i \log_2(\frac{p_i}{q_i}) = 0$, when $p_i = 0$ and for any $0 \leq q_i \leq 1$. However a family of power Divergence measures indexed by $\lambda \in \mathbb{R}$ for $\underline{p} = (p_1, p_2, \ldots, p_k)$, $\underline{q} = (q_1, q_2, \ldots, q_k)$ is defined as

$$I^{\lambda}(\underline{p} : \underline{q}) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^{k} p_i[(\frac{p_i}{q_i})^{\lambda} - 1],$$

and we adopt the convention that whenever $q_i = 0$, then $p_i = 0$. Next considering the fact that a matrix can be written as an array of column vectors, we define the power divergence statistic for the matrices A and B in the following way:

$$
\begin{aligned}
& D_1 \\
& = I^{\lambda}(p_{ij} : a_{ij}p_{.j}) + I^{\lambda}(p_{ij} : b_{ij}p_{i.}) \\
& = \frac{1}{\lambda(\lambda+1)}[\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij}((\frac{p_{ij}}{a_{ij}p_{.j}})^{\lambda} - 1) + \sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij}((\frac{p_{ij}}{b_{ij}p_{i.}})^{\lambda} - 1)],
\end{aligned}
\tag{1.6}
$$

where the matrices A and B have been defined earlier and $\lambda \in \mathbb{R}$ is a parameter.

Note: The power-divergence family is undefined for $\lambda = -1$ or $\lambda = 0$. However if we define these two cases by the continuous limits of D_1 as $\lambda \to -1$ and $\lambda \to 0$, then D_1 is continuous in λ.

The name power divergence derives from the fact that the statistic D_1 measures the divergence of p_{ij} from $(a_{ij}p_{.j})$ and $(b_{ij}p_{i.})$ through a weighted sum of powers of the term $(\frac{p_{ij}}{a_{ij}p_{.j}})$ and $(\frac{p_{ij}}{b_{ij}p_{i.}})$,$\forall(i,j) \in N$. We want to minimize D_1 with respect to $\sum\sum_{(i,j)\in N} p_{ij} = 1$.

Using the Lagrangian multiplier for the constraint $\sum\sum_{(i,j)\in N} p_{ij} = 1$, the optimal value of p_{ij} will be a solution to the following equation:

$$p_{ij} = \frac{1}{(\sum\sum_{(i,j)\in N}([(\frac{1}{a_{ij}p_{.j}})^{\lambda} + (\frac{1}{b_{ij}p_{i.}})^{\lambda}]^{\frac{1}{\lambda}})^{-1})^{-1}((\frac{1}{a_{ij}p_{.j}})^{\lambda} + (\frac{1}{b_{ij}p_{i.}})^{\lambda})^{\frac{1}{\lambda}}}.$$

Since the function D_1 is strictly convex, the value of p_{ij} as obtained really corresponds to a minimum.

1.4.6 Iterative Algorithm

In this case we consider the following iterative algorithm

$$p_{ij}^{(n+1)} = \frac{1}{(\sum\sum_{(i,j)\in N}([(\frac{1}{a_{ij}p_{.j}^{(n)}})^{\lambda} + (\frac{1}{b_{ij}p_{i.}^{(n)}})^{\lambda}]^{\frac{1}{\lambda}})^{-1})^{-1}((\frac{1}{a_{ij}p_{.j}^{(n)}})^{\lambda} + (\frac{1}{b_{ij}p_{i.}^{(n)}})^{\lambda})^{\frac{1}{\lambda}}}, \tag{1.7}$$

$n = 0, 1, 2, \ldots$

For an initial guess we consider $p_{ij}^{(0)} = \frac{1}{IJ}, \forall(i,j) \in N$. We consider the following stopping rule for the convergence of our iterative algorithm:$|D_1^{(n)} - D_1^{(n+1)}| \leq 10^{-5}$. Furthermore the iterative procedure in all the examples investigated has been found to be convergent.

1.4.7 What about the choice of λ

In the power divergence statistic, λ is a parameter that can take on any real value. A natural question that arises here is: what should be the optimum choice of λ? There are some conflicting recommendations regarding which value of λ results in the optimal test statistic. In all our examples of iterative study wherever we used this, it has been found that whenever we consider the value of λ other than $\frac{2}{3}$ our iterative procedure although it converges has a rate of convergence that is very slow. For example when we consider λ=0.2,0.3, and 0.5 for our iterative study, for the divergence measure D_1 our iterative procedure converges at n=20, 27 and 34 respectively. For negative choices of λ the value of D_1 is quite big and moreover the resulting matrix is not a probability matrix. However when we consider $\lambda = \frac{2}{3}$, for the measure D_1 our iterative procedure converges quite rapidly as has been demonstrated in all the examples later on.

1.4.8 Properties of the Power Divergence Statistic:

- Nonnegativity:

  A natural requirement for a measure of divergence is that it take only positive values and that it increases as $\underline{p}$ and $\underline{q}$ "diverge". In particular, for the power divergence we have $I^{\lambda}(\underline{p}_i : \underline{q}_i) \geq 0$, with equality iff $p_i = q_i, \forall i$. This result follows from the strict convexity of the function $\theta(x) = \frac{x^{\lambda+1}-1}{\lambda(\lambda+1)}$ and by Jensen's inequality.

- Permutation invariance:

  The value of the power divergence is not affected by a simultaneous and equivalent reordering of the discrete probability masses in both of the distributions; i.e.,

$I^\lambda(\underline{p}:\underline{q}) = I^\lambda(\underline{p'}:\underline{q'})$, where $\underline{p'} = (p_{a_1}, p_{a_2}, \ldots, p_{a_k}), \underline{q'} = (q_{a_1}, \ldots, q_{ak})$, and $(a_1, \ldots, a_k)$ is an arbitrary permutation of the natural order $(1, 2, \ldots, k)$.

- Continuity:

  $I^\lambda(\underline{p}:\underline{q})$ is a continuous function in each of its arguments, in other words small changes in the probability distributions under comparison will result in only small changes in the power divergence.

- This function reduces to usual Kullback-Leibler measure of divergence in the limit as $\lambda \to 0$.

Next we consider certain other measures of divergence which are as follows

1.4.9 Modified Renyi's measure of divergence

Renyi (1961) proposed the following measure of divergence (of order α) which is defined as

$$R^\alpha(\underline{p}:\underline{q}) = (\alpha-1)^{-1}\log_2[\sum_{i=1}^{k} p_i^\alpha q_i^{1-\alpha}], \alpha \neq 1.$$

Using additive property of Renyi's measure we propose the following measure with a slight modification (we consider the natural log) as follows

$$D_2 = R^\alpha = (\alpha - 1)^{-1}[\sum\sum_{(i,j)\in N} (a_{ij}p_{.j})^{-1} \log(\frac{p_{ij}}{a_{ij}p_{.j}})^\alpha + (b_{ij}p_{i.})^{-1} \log(\frac{p_{ij}}{b_{ij}p_{i.}})^\alpha]. \tag{1.8}$$

Note: Nadarajah and Zografos (2003, 2005) provide review of Renyi's entropy for different univariate and k-variate random variables.

Minimizing this function using Lagrangian multiplier for the constraint $\sum\sum_{(i,j)\in N} p_{ij} = 1$, the optimal value of p_{ij}, will be a solution to the following equation:

$$p_{ij} = \frac{\frac{1}{a_{ij}p_{.j}} + \frac{1}{b_{ij}p_{i.}}}{\sum\sum_{(i,j)\in N}(\frac{1}{a_{ij}p_{.j}} + \frac{1}{b_{ij}p_{i.}})}.$$

Also note that

$$\left.\frac{\partial^2 R^\alpha}{\partial (p_{ij})^2}\right|_{p_{ij}} > 0,$$

which implies that the estimate of p_{ij} as obtained really gives a minimum.

1.4.10 Renyi's measure of divergence is a limiting case of Kullback-Leibler information criterion

Here we will show that Renyi's measure is in fact a special case of Kullback-Leibler measure of discrepancy. We will show the result for the continuous case. The proof for the discrete case will follow the same logic with only the integration being replaced by summation. First of all recall that Renyi's measure of divergence is given by

$$\begin{aligned} R^{\alpha}(f_1 : f_2) &= \frac{1}{\alpha-1}\log\int_S [f_1^{\alpha}(x) f_2^{1-\alpha}(x)]dx \\ &= \frac{1}{\alpha-1}\log\int_S [f_1(x)(\frac{f_1(x)}{f_2(x)})^{\alpha-1}]dx. \end{aligned} \tag{1.9}$$

However next we consider

$$\begin{aligned} \lim_{\alpha\to 1} R^{\alpha}(f_1 : f_2) &= \lim_{\alpha\to 1}\frac{1}{\alpha-1}\log\int_S [f_1(x)(\frac{f_1(x)}{f_2(x)})^{\alpha-1}]dx \\ &= \lim_{\alpha\to 1}\frac{\int_S f_1(x)(\frac{f_1(x)}{f_2(x)})^{\alpha-1}\log(\frac{f_1(x)}{f_2(x)})dx}{\int_S f_1(x)(\frac{f_1(x)}{f_2(x)})^{\alpha-1}dx} \\ &= \frac{\int_S f_1(x)\log(\frac{f_1(x)}{f_2(x)})\lim_{\alpha\to 1}(\frac{f_1(x)}{f_2(x)})^{\alpha-1}dx}{\int_S f_1(x)\lim_{\alpha\to 1}(\frac{f_1(x)}{f_2(x)})^{\alpha-1}dx} \\ &= \int_S f_1(x)\log(\frac{f_1(x)}{f_2(x)})dx \\ &= K(f_1 : f_2), \end{aligned} \tag{1.10}$$

assuming the limit under the integral sign is valid and applying La Hospital's rule. Hence the proof where $K(:)$ is the Kullback-Leibler measure of divergence.

1.4.11 Iterative algorithm:

Here we consider the following scheme: $p_{ij}^{(n+1)} = \frac{\frac{1}{a_{ij}p_{.j}^{(n)}}+\frac{1}{b_{ij}p_{i.}^{(n)}}}{\sum\sum_{(i,j)\in N}(\frac{1}{a_{ij}p_{.j}^{(n)}}+\frac{1}{b_{ij}p_{i.}^{(n)}})}$, with the initial guess as $p_{ij}^0 = \frac{1}{IJ}, \forall(i,j) \in N$. We consider the following stopping rule for the convergence of our iterative algorithm: $|D_2^{(n)} - D_2^{(n+1)}| \leq 10^{-5}$. In this case also our process is convergent. Simulation with an incompatible set-up is shown in later sections.

1.5 χ^2 Divergence criterion:

In this case the statistic (D_3, say) is given by

$$D_3 = \sum\sum_{(i,j)\in N}((\frac{p_{ij}}{a_{ij}p_{.j}}-1)^2)a_{ij}p_{.j}+\sum\sum_{(i,j)\in N}((\frac{p_{ij}}{b_{ij}p_{i.}}-1)^2)b_{ij}p_{i.}. \quad (1.11)$$

Using the same technique as before i.e., minimizing D_3 using Lagrangian multiplier for the constraint $\sum\sum_{(i,j)\in N} p_{ij} = 1$, the optimal value of p_{ij} will be a solution to the following equation

$$p_{ij} = \frac{1}{[\frac{1}{a_{ij}p_{.j}} + \frac{1}{b_{ij}p_{i.}}][\sum\sum_{(i,j)\in N}[\frac{1}{a_{ij}p_{.j}} + \frac{1}{b_{ij}p_{i.}}]^{-1}]}, \forall(i,j) \in N. \quad (1.12)$$

Furthermore

$$\left.\frac{\partial^2 D_3}{\partial (p_{ij})^2}\right|_{p_{ij}} > 0,$$

which again implies that the optimal value of p_{ij} as obtained in (1.12) provides the minimum for the function as earlier.

1.5.1 Iterative Algorithm:

In this case for the iterative algorithm we consider

$$p_{ij}^{(n+1)} = \frac{1}{[\frac{1}{a_{ij}p_{.j}^{(n)}} + \frac{1}{b_{ij}p_{i.}^{(n)}}][\sum\sum_{(i,j)\in N}[\frac{1}{a_{ij}p_{.j}^{(n)}} + \frac{1}{b_{ij}p_{i.}^{(n)}}]^{-1}]}, n = 0, 1, 2, \ldots,$$

with the same initial guess as before i.e., $p_{ij}^0 = \frac{1}{IJ}\forall(i,j) \in N$. We consider the following stopping rule for the convergence of our iterative algorithm: $|D_3^{(n)} - D_3^{(n+1)}| \leq 10^{-5}$. In this case also our iterative procedure is convergent.

1.6 Proposed new measures of divergence:

1. First of all we consider a (pseudo)distance measure of the form:

$$D_4 = \sum\sum_{(i,j)\in N}[(\frac{2p_{ij}}{a_{ij}p_{.j} + b_{ij}p_{i.}} - 1)^2]^\lambda, \quad (1.13)$$

   where $\lambda > 0$, is a constant.

   Note that if in this case A and B are compatible, then $D_4 = 0$ and vice versa and also this measure is non-negative. Next minimizing D_4 with the linear constraint that $\sum\sum_{(i,j)\in N} p_{ij} = 1$, the optimal value of p_{ij} will be a solution to the following equation

$$p_{ij} = \frac{(a_{ij}p_{.j} + b_{ij}p_{i.})^{1-\frac{1}{\lambda}}}{\sum\sum_{(i,j)\in N}[(a_{ij}p_{.j} + b_{ij}p_{i.})^{1-\frac{1}{\lambda}}]}.$$

Based on the above optimal value an iterative algorithm could be

$$p_{ij}^{(n+1)} = \frac{(a_{ij}p_{.j}^{(n)} + b_{ij}p_{i.}^{(n)})^{1-\frac{1}{\lambda}}}{\sum\sum_{(i,j)\in N}[(a_{ij}p_{.j}^{(n)} + b_{ij}p_{i.}^{(n)})^{1-\frac{1}{\lambda}}]}, n = 0, 1, \ldots.$$

With the initial choice $p_{ij}^{(0)} = \frac{1}{IJ}, \forall(i,j) \in N$. It can be shown that $\frac{\partial^2}{\partial p_{ij}^2} D_4\Big|_{p_{ij}} > 0$, which means that the optimal value of p_{ij} really gives a minimum. For our iterative study we have considered $\lambda = \frac{2}{3}$. The reason behind considering this particular choice is that with this choice of λ, our iterative procedure converges to the minimally incompatible P faster than for any other assumed values. We have considered the values of $\lambda = 0.1, 0.3, 0.5, \text{and} 0.9$. For each of these choices we have performed our iterative study, and the resultant most nearly compatible P although not much different in comparison to those which we have obtained in all of our examples tried, but the rate at which it converges to that P is really slow. Also our iterative procedure is convergent. We consider the following as our convergence criteria for the above iterative algorithm $|D_4^{(n)} - D_4^{(n+1)}| \leq 10^{-5}$.

2. Next we consider a divergence measure of the following form:

$$D_5 = \sum \sum_{(i,j)\in N} [(p_{ij} - a_{ij}p_{.j})^2 a_{ij}p_{.j}] + \sum \sum_{(i,j)\in N} [(p_{ij} - b_{ij}p_{i.})^2 b_{ij}p_{i.}] \tag{1.14}$$

Again note that if A and B are compatible then $D_5 = 0$ and vice versa. In addition this measure is non-negative. Minimizing D_5 under the linear constraint the optimal value of p_{ij} will be a solution to the following equation:

$$p_{ij} = \frac{(a_{ij}p_{.j})^2 + (b_{ij}p_{i.})^2}{\sum\sum_{(i,j)\in N}[(a_{ij}p_{.j})^2 + (b_{ij}p_{i.})^2]},$$

and also it is easy to verify that $\frac{\partial^2}{\partial p_{ij}^2} D_5 \Big|_{p_{ij}} > 0$.

So that an iterative scheme would be

$$p_{ij}^{(n+1)} = \frac{(a_{ij}p_{.j}^{(n)})^2 + (b_{ij}p_{i.}^{(n)})^2}{\sum\sum_{(i,j)\in N}[(a_{ij}p_{.j}^{(n)})^2 + (b_{ij}p_{i.}^{(n)})^2]},$$

for $n = 0, 1, 2, \ldots$ With the same initial choice $p_{ij}^{(0)} = \frac{1}{IJ}, \forall (i,j) \in N$. We consider the following as our convergence criteria for the above iterative algorithm $|D_5^{(n)} - D_5^{(n+1)}| \leq 10^{-5}$.

In this case our procedure is also convergent.

3. Now we consider the divergence measure of the following form:

$$D_6 = \sum \sum_{(i,j)\in N} [\sqrt{p_{ij}} - \sqrt{a_{ij}p_{.j}}]^2 + \sum \sum_{(i,j)\in N} [\sqrt{p_{ij}} - \sqrt{b_{ij}p_{i.}}]^2, \tag{1.15}$$

Note that if the two matrices A and B are compatible then $D_6 = 0$ and vice versa. Also D_6 is nonnegative. Minimizing D_5 under the linear constraint the optimal value of p_{ij} will be a solution to the following equation:

$$p_{ij} = \frac{(\frac{1}{\sqrt{a_{ij}p_{.j}}+\sqrt{b_{ij}p_{i.}}})^2}{\sum\sum_{(i,j)\in N}[(\frac{1}{\sqrt{a_{ij}p_{.j}}+\sqrt{b_{ij}p_{i.}}})^2]},$$

So that an iterative scheme would be

$$p_{ij}^{(n+1)} = \frac{(\frac{1}{\sqrt{a_{ij}p_{.j}^{(n)}}+\sqrt{b_{ij}p_{i.}^{(n)}}})^2}{\sum\sum_{(i,j)\in N}[(\frac{1}{\sqrt{a_{ij}p_{.j}^{(n)}}+\sqrt{b_{ij}p_{i.}^{(n)}}})^2]},$$

for $n = 0, 1, 2, \ldots$ With the same initial choice $p_{ij}^{(0)} = \frac{1}{IJ}, \forall (i,j) \in N$. We consider the following as our convergence criteria for the above iterative algorithm $|D_6^{(n)} - D_6^{(n+1)}| \leq 10^{-5}$.

In this case our procedure is also convergent.

1.6.1 Iterative study

- Let us first consider some examples with $I = 3$ and $J = 3$ and with the following choices of A and B:

1. Incompatible of type 1 (a case in which all the elements are strictly positive)

   Let

$$A = \begin{pmatrix} 1/5 & 2/7 & 3/8 \\ 3/5 & 2/7 & 1/8 \\ 1/5 & 3/7 & 1/2 \end{pmatrix},$$

   and

$$B = \begin{pmatrix} 1/6 & 1/3 & 1/2 \\ 1/2 & 1/3 & 1/6 \\ 1/8 & 3/8 & 1/2 \end{pmatrix}.$$

   In this case the two matrices A and B are not compatible since if we consider the lower right cross product ratios of the two matrices A and B, we have for the matrix B, the lower right cross product $= \frac{\frac{1}{3}\frac{1}{2}}{\frac{1}{6}\frac{3}{8}} = 2$, while for the matrix A, the lower right cross product $= \frac{\frac{2}{7}\frac{1}{2}}{\frac{3}{7}\frac{1}{8}} = \frac{8}{3}$. Thus A and B are not compatible.

   (a) Under the Power divergence criterion D_1, We obtained $D_1^{(10)} = 0.0002416142$ and for this particular value of D_1, the corresponding minimally incompatible P is given by is given

by

$$P = \begin{pmatrix} 0.05086816 & 0.1699624 & 0.05264803 \\ 0.09217352 & 0.1019318 & 0.14278151 \\ 0.14184063 & 0.0523794 & 0.19541450 \end{pmatrix},$$

and we stopped since in this case $|D_1^{(10)} - D_1^{(11)}| \leq 10^{-5}$, and also $|P^{(10)} - P^{(11)}| \leq 10^{-5}$.

(b) Under the Modified Renyi's divergence criterion D_2, we obtained $D_2^{(11)} = 0.0002447811$, and for this particular value of D_2, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.05123717 & 0.1603623 & 0.05054213 \\ 0.09211853 & 0.1102115 & 0.1405672 \\ 0.14099155 & 0.0523794 & 0.2015902 \end{pmatrix},$$

and we stopped since in this case $|D_2^{(11)} - D_2^{(12)}| \leq 10^{-5}$, and also $|P^{(11)} - P^{(12)}| \leq 10^{-5}$.

(c) Under the χ^2 divergence criterion, i.e., under D_3, we obtained $D_3^{(12)} = 0.00013453267$, and for this particular value of D_3, the corresponding minimally incompatible P is given

by

$$P = \begin{pmatrix} 0.05156237 & 0.17013456 & 0.053196463 \\ 0.09260173 & 0.10306895 & 0.14113752 \\ 0.14198767 & 0.05230725 & 0.19137217 \end{pmatrix},$$

and we stopped since in this case $|D_3^{(12)} - D_3^{(13)}| \leq 10^{-5}$, and also $|P^{(12)} - P^{(13)}| \leq 10^{-5}$.

(d) Now under D_4, we obtained $D_4^{(15)} = 0.00007405324$, and for this particular value of D_4, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.05424209 & 0.17194484 & 0.05481822 \\ 0.09251337 & 0.09775432 & 0.14024400 \\ 0.14389799 & 0.05068330 & 0.19390186 \end{pmatrix},$$

and we stopped since in this case $|D_4^{(15)} - D_4^{(16)}| \leq 10^{-5}$, and also $|P^{(15)} - P^{(16)}| \leq 10^{-5}$.

(e) For D_5, we obtained $D_5^{(15)} = 0.000115798$, and for this particular value of D_5, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.0515650 & 0.1721786 & 0.0527697 \\ 0.0926017 & 0.1030674 & 0.1421476 \\ 0.1409875 & 0.0523072 & 0.1923749 \end{pmatrix},$$

and we stopped since in this case $|D_5^{(15)} - D_5^{(16)}| \leq 10^{-5}$, and also $|P^{(15)} - P^{(16)}| \leq 10^{-5}$.

(f) Furthermore under D_6, we obtained $D_6^{(16)} = 0.0000032561$, and for this particular value of D_6, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.05177626 & 0.17205234 & 0.0528665 \\ 0.09258417 & 0.10333889 & 0.1420280 \\ 0.14089836 & 0.05236851 & 0.1920870 \end{pmatrix},$$

and we stopped since in this case $|D_6^{(16)} - D_6^{(17)}| \leq 10^{-5}$, and also $|P^{(16)} - P^{(17)}| \leq 10^{-5}$.

2. Incompatible of type 2 (a case in which there are some zeros in the same position(s) for both the matrices)

Let us consider two matrices A and B of the following forms:

$$A = \begin{pmatrix} 0 & 1/3 & 0 \\ 1 & 1/3 & 1/2 \\ 0 & 1/3 & 1/2 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 0 & 1 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/5 & 4/5 \end{pmatrix}.$$

Again by the same argument as before the two matrices A and B are incompatible.

(a) Then under D_1, we obtained $D_1^{(9)} = 0.000176543$, and for this particular value of D_1, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.0000000 & 0.1466426 & 0.0000000 \\ 0.1590778 & 0.2054299 & 0.1104262 \\ 0.0000000 & 0.1482233 & 0.2302003 \end{pmatrix},$$

and we stopped since in this case $|D_1^{(9)} - D_1^{(10)}| \le 10^{-5}$, and also $|P^{(9)} - P^{(10)}| \le 10^{-5}$.

(b) However under the modified Renyi's measure of divergence, i.e., under D_2, we obtained $D_2^{(12)} = 0.000453816$, and for this particular value of D_2, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.0000000 & 0.1529851 & 0.0000000 \\ 0.1648741 & 0.1934395 & 0.1317364 \\ 0.0000000 & 0.1530584 & 0.2039064 \end{pmatrix},$$

and we stopped since in this case $|D_2^{(12)} - D_2^{(13)}| \leq 10^{-5}$, and also $|P^{(12)} - P^{(13)}| \leq 10^{-5}$.

(c) Then under D_3, the $\varepsilon = 0.007263714$ compatible distribution is given by

$$P = \begin{pmatrix} 0.0000000 & 0.17712336 & 0.0000000 \\ 0.1349069 & 0.19947534 & 0.1547621 \\ 0.0000000 & 0.09877285 & 0.2349595 \end{pmatrix},$$

and it is achieved at $n = 3$ stage of iteration.

(d) Then under D_4, we obtained $D_4^{(14)} = 0.000189767$, and for this particular value of D_4, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.0000000 & 0.1515781 & 0.0000000 \\ 0.1683712 & 0.1946174 & 0.1295497 \\ 0.0000000 & 0.1526528 & 0.2132307 \end{pmatrix},$$

and we stopped since in this case $|D_4^{(14)} - D_4^{(15)}| \leq 10^{-5}$, and also $|P^{(14)} - P^{(15)}| \leq 10^{-5}$.

(e) While under the divergence measure D_5, we obtained $D_5^{(15)} = 0.000523176$, and for this particular value of D_5, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.0000000 & 0.1496455 & 0.0000000 \\ 0.1643222 & 0.1927815 & 0.1305459 \\ 0.0000000 & 0.1673558 & 0.1953493 \end{pmatrix},$$

and we stopped since in this case $|D_5^{(15)} - D_5^{(16)}| \leq 10^{-5}$, and also $|P^{(15)} - P^{(16)}| \leq 10^{-5}$.

(f) Under D_6, we obtained $D_6^{(13)} = 0.00006321$, and for this particular value of D_6, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.0000000 & 0.1377437 & 0.0000000 \\ 0.1539082 & 0.2015677 & 0.1184920 \\ 0.0000000 & 0.1480436 & 0.2402449 \end{pmatrix},$$

and we stopped since in this case $|D_6^{(13)} - D_6^{(14)}| \leq 10^{-5}$, and also $|P^{(13)} - P^{(14)}| \leq 10^{-5}$.

- Again let us consider some examples with $I = 4$ and $J = 4$ and with the following choices of A and B:

1. Incompatible of type 1 (a case in which all the elements are strictly positive)

   Let

$$A = \begin{pmatrix} 3/10 & 3/10 & 2/10 & 2/10 \\ 1/5 & 1/10 & 1/5 & 3/10 \\ 2/5 & 1/5 & 3/10 & 2/5 \\ 1/10 & 2/5 & 3/10 & 1/10 \end{pmatrix},$$

   and

$$B = \begin{pmatrix} 1/5 & 3/10 & 2/5 & 1/10 \\ 3/10 & 1/5 & 1/5 & 3/10 \\ 3/5 & 1/10 & 1/5 & 1/10 \\ 1/5 & 2/5 & 1/5 & 1/5 \end{pmatrix}.$$

   Here also as before the two matrices A and B are not compatible since if we consider the upper right cross product ratios of the two matrices A and B, we have for the matrix A, the upper right cross product $= \frac{(\frac{2}{10})(\frac{3}{10})}{(\frac{2}{10})(\frac{1}{5})} = \frac{3}{2}$, while for the matrix B, the upper right cross product $= \frac{(\frac{2}{5})(\frac{3}{10})}{(\frac{1}{5})(\frac{1}{10})} = 6$. Thus A and B are not compatible.

   (a) Under the Power divergence criterion D_1, we obtained $D_1^{(10)} = 0.0002143704$ and for this particular value of D_1, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.07704966 & 0.06688648 & 0.14051643 & 0.03330258 \\ 0.08339375 & 0.03279953 & 0.03554727 & 0.08354802 \\ 0.08655587 & 0.05554226 & 0.07060513 & 0.05502597 \\ 0.03774765 & 0.06954348 & 0.04433518 & 0.02760073 \end{pmatrix},$$

and we stopped since in this case $|D_1^{(10)} - D_1^{(11)}| \leq 10^{-5}$, and also $|P^{(10)} - P^{(11)}| \leq 10^{-5}$.

(b) Under the Modified Renyi's divergence criterion D_2, we obtained $D_2^{(8)} = 0.0001410993$ and for this particular value of D_2, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.07604231 & 0.06688648 & 0.14051643 & 0.03330258 \\ 0.08339375 & 0.03279953 & 0.03554727 & 0.08354802 \\ 0.08655423 & 0.05554021 & 0.07060513 & 0.05502349 \\ 0.03774684 & 0.06954323 & 0.04433488 & 0.02861562 \end{pmatrix},$$

and we stopped since in this case $|D_2^{(8)} - D_2^{(9)}| \leq 10^{-5}$, and also $|P^{(8)} - P^{(9)}| \leq 10^{-5}$.

(c) However under the χ^2 divergence criterion, i.e., under D_3, we obtained $D_3^{(9)} = 0.00001768364$ and for this particular value of D_3, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.07023404 & 0.07858238 & 0.06903101 & 0.02853365 \\ 0.06796916 & 0.03265322 & 0.04667173 & 0.05455597 \\ 0.17033167 & 0.04156970 & 0.07271910 & 0.04446242 \\ 0.04098023 & 0.09951178 & 0.05845184 & 0.02374210 \end{pmatrix},$$

and we stopped since in this case $|D_3^{(9)} - D_3^{(10)}| \leq 10^{-5}$, and also $|P^{(9)} - P^{(10)}| \leq 10^{-5}$.

(d) Under D_4, we obtained $D_4^{(12)} = 0.002070486$ and for this particular value of D_4, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.07353635 & 0.05889123 & 0.12842186 & 0.03112296 \\ 0.07324347 & 0.02819919 & 0.03873810 & 0.08597404 \\ 0.07811256 & 0.05683596 & 0.08120368 & 0.07152065 \\ 0.03870103 & 0.06705061 & 0.06140633 & 0.02704200 \end{pmatrix},$$

and we stopped since in this case $|D_4^{(12)} - D_4^{(13)}| \leq 10^{-5}$, and also $|P^{(12)} - P^{(13)}| \leq 10^{-5}$.

(e) For D_5, we obtained $D_5^{(10)} = 0.00001845390$ and for this particular value of D_5, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.07496105 & 0.06485755 & 0.13969506 & 0.03282546 \\ 0.08113641 & 0.03309287 & 0.03563895 & 0.08205221 \\ 0.08914853 & 0.05454113 & 0.07193831 & 0.05855724 \\ 0.03696451 & 0.06784473 & 0.04870969 & 0.02803630 \end{pmatrix},$$

and we stopped since in this case $|D_5^{(10)} - D_5^{(11)}| \leq 10^{-5}$, and also $|P^{(10)} - P^{(11)}| \leq 10^{-5}$.

(f) Furthermore under D_6, we obtained $D_6^{(8)} = 0.00001609678$ and for this particular value of D_6, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.07292820 & 0.06273524 & 0.13917061 & 0.03242250 \\ 0.07897186 & 0.03333501 & 0.03546013 & 0.08041667 \\ 0.09114661 & 0.05365939 & 0.07318725 & 0.06237551 \\ 0.03630846 & 0.06608111 & 0.05323884 & 0.02856262 \end{pmatrix},$$

and we stopped since in this case $|D_6^{(8)} - D_6^{(9)}| \leq 10^{-5}$, and also $|P^{(8)} - P^{(9)}| \leq 10^{-5}$.

2. Incompatible of type 2 (a case in which there are some zeros in the same position (s) for both the matrices) Let us consider

two matrices A and B of the following form:

$$A = \begin{pmatrix} 0 & 2/10 & 1/10 & 1/5 \\ 1/5 & 2/5 & 0 & 3/10 \\ 1/2 & 0 & 3/5 & 1/2 \\ 3/10 & 2/5 & 3/10 & 0 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 0 & 1/10 & 2/5 & 1/2 \\ 3/10 & 3/10 & 0 & 2/5 \\ 2/5 & 0 & 3/10 & 3/10 \\ 3/5 & 3/10 & 1/10 & 0 \end{pmatrix}.$$

It can be easily checked that here also the matrices A and B are incompatible as not all the cross product ratios are equal for both the matrices.

(a) Under the Power divergence criterion D_1, we obtained $D_1^{(10)} =$ 0.0002363084 and for this particular value of D_1, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0 & 0.05515792 & 0.1042738 & 0.09847155 \\ 0.03798426 & 0.09140565 & 0 & 0.10173017 \\ 0.05837866 & 0 & 0.1207913 & 0.05239492 \\ 0.08346835 & 0.09363649 & 0.1023069 & 0 \end{pmatrix},$$

and we stopped since in this case $|D_1^{(10)} - D_1^{(11)}| \leq 10^{-5}$, and also $|P^{(10)} - P^{(11)}| \leq 10^{-5}$.

(b) Under the Modified Renyi's divergence criterion D_2, we obtained $D_2^{(14)} = 0.0002363084$ and for this particular value of D_2, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0 & 0.05515792 & 0.1042738 & 0.09847155 \\ 0.03798426 & 0.09140565 & 0 & 0.10173017 \\ 0.05837866 & 0 & 0.1207913 & 0.05239492 \\ 0.08346835 & 0.09363649 & 0.1023069 & 0 \end{pmatrix},$$

and we stopped since in this case $|D_2^{(14)} - D_2^{(15)}| \leq 10^{-5}$, and also $|P^{(14)} - P^{(15)}| \leq 10^{-5}$.

(c) Under the χ^2 divergence criterion, i.e., under D_3, we obtained $D_3^{(11)} = 0.00001768364$ and for this particular value of D_3, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0 & 0.04514829 & 0.10173578 & 0.08144188 \\ 0.03252637 & 0.06651628 & 0 & 0.09081290 \\ 0.06417685 & 0.00000000 & 0.14749101 & 0.04978037 \\ 0.07848409 & 0.08351710 & 0.1042117 & 0 \end{pmatrix},$$

and we stopped since in this case $|D_3^{(11)} - D_3^{(12)}| \leq 10^{-5}$, and also $|P^{(11)} - P^{(12)}| \leq 10^{-5}$.

(d) Now under D_4, we obtained $D_4^{(10)} = 0.00002014352$ and for this particular value of D_4, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0 & 0.04103676 & 0.09801774 & 0.07469942 \\ 0.03471148 & 0.07678929 & 0 & 0.08466566 \\ 0.05602149 & 0 & 0.19132350 & 0.07820885 \\ 0.06706973 & 0.07883477 & 0.11862132 & 0 \end{pmatrix},$$

and we stopped since in this case $|D_4^{(10)} - D_4^{(11)}| \leq 10^{-5}$, and also $|P^{(10)} - P^{(11)}| \leq 10^{-5}$.

(e) For D_5 we obtained $D_5^{(15)} = 0.00006752265$ and for this particular value of D_5, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0 & 0.05416158 & 0.1089195 & 0.10710785 \\ 0.03450341 & 0.07926110 & 0 & 0.09049601 \\ 0.06069293 & 0 & 0.1329969 & 0.05628083 \\ 0.08369955 & 0.08598767 & 0.1058927 & 0 \end{pmatrix},$$

and we stopped since in this case $|D_5^{(15)} - D_5^{(16)}| \leq 10^{-5}$, and also $|P^{(15)} - P^{(16)}| \leq 10^{-5}$.

(f) Furthermore under D_6 we obtained $D_6^{(14)} = 0.01305660$ and for this particular value of D_6, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0 & 0.05005653 & 0.1097232 & 0.11128045 \\ 0.03227245 & 0.06889548 & 0 & 0.08181862 \\ 0.06548797 & 0 & 0.1440695 & 0.06263719 \\ 0.08703722 & 0.07864499 & 0.1080764 & 0 \end{pmatrix},$$

and we stopped since in this case $|D_6^{(14)} - D_6^{(15)}| \leq 10^{-5}$, and also $|P^{(14)} - P^{(15)}| \leq 10^{-5}$.

- Let us consider an example with $I = 3$ and $J = 4$ and with the following choices of A and B:

  1. Incompatible of type 1

     Let

$$A = \begin{pmatrix} 1/7 & 2/7 & 4/7 & 1/4 \\ 1/2 & 1/4 & 3/7 & 1/7 \\ 3/7 & 1/7 & 2/7 & 4/7 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 1/6 & 2/7 & 1/3 & 1/6 \\ 2/7 & 1/12 & 1/2 & 1/7 \\ 1/4 & 1/6 & 2/7 & 1/3 \end{pmatrix}.$$

It can be easily checked that here also the matrices A and B are incompatible as not all the cross product ratios are equal for both the matrices.

(a) Under the Power divergence criterion D_1, we obtained $D_1^{(9)} = 0.0001934688$ and for this particular value of D_1, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.04013147 & 0.04133851 & 0.1217852 & 0.04013147 \\ 0.08021275 & 0.08262430 & 0.0405695 & 0.08021275 \\ 0.15683188 & 0.04036863 & 0.1189616 & 0.15683188 \end{pmatrix},$$

and we stopped since in this case $|D_1^{(9)} - D_1^{(10)}| \leq 10^{-5}$, and also $|P^{(9)} - P^{(10)}| \leq 10^{-5}$.

(b) Under the Modified Renyi's divergence criterion D_2, we obtained $D_2^{(12)} = 0.00001463646$ and for this particular value of D_2, the corresponding minimally incompatible P is given

by

$$P = \begin{pmatrix} 0.12321281 & 0.08303605 & 0.04108691 & 0.12321281 \\ 0.06905963 & 0.04897125 & 0.13816718 & 0.06905963 \\ 0.04695743 & 0.14765297 & 0.06262588 & 0.04695743 \end{pmatrix},$$

and we stopped since in this case $|D_2^{(12)} - D_2^{(13)}| \leq 10^{-5}$, and also $|P^{(12)} - P^{(13)}| \leq 10^{-5}$.

(c) Under the χ^2 divergence criterion, i.e., under D_3, we obtained $D_3^{(14)} = 0.0001843705$ and for this particular value of D_3, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.04010651 & 0.04128553 & 0.12156161 & 0.04010651 \\ 0.08025643 & 0.08261707 & 0.04054269 & 0.08025643 \\ 0.15698329 & 0.04037408 & 0.11892656 & 0.15698329 \end{pmatrix},$$

and we stopped since in this case $|D_3^{(14)} - D_3^{(15)}| \leq 10^{-5}$, and also $|P^{(14)} - P^{(15)}| \leq 10^{-5}$.

(d) Now under D_4, we obtained $D_4^{(17)} = 0.00002486131$ and for this particular value of D_4, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.03950049 & 0.04258211 & 0.11978602 & 0.03950049 \\ 0.07938223 & 0.08557522 & 0.04012137 & 0.07938223 \\ 0.15661401 & 0.04220805 & 0.11873378 & 0.15661401 \end{pmatrix},$$

and we stopped since in this case $|D_4^{(17)} - D_4^{(18)}| \leq 10^{-5}$, and also $|P^{(17)} - P^{(18)}| \leq 10^{-5}$.

(e) For D_5, we obtained $D_5^{(11)} = 0.0001901572$ and for this particular value of D_5, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.04011167 & 0.04133195 & 0.12161477 & 0.04011167 \\ 0.08024703 & 0.08268831 & 0.04055023 & 0.08024703 \\ 0.15688426 & 0.04041425 & 0.11891455 & 0.15688426 \end{pmatrix},$$

and we stopped since in this case $|D_5^{(11)} - D_5^{(12)}| \leq 10^{-5}$, and also $|P^{(11)} - P^{(12)}| \leq 10^{-5}$.

(f) Furthermore under D_6, we obtained $D_6^{(13)} = 0.0001305660$ and for this particular value of D_6, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0.04006222 & 0.04131085 & 0.12121682 & 0.04006222 \\ 0.08034045 & 0.08283992 & 0.04051391 & 0.08034045 \\ 0.15701657 & 0.04049440 & 0.11878562 & 0.15701657 \end{pmatrix},$$

and we stopped since in this case $|D_6^{(13)} - D_6^{(14)}| \leq 10^{-5}$, and also $|P^{(13)} - P^{(14)}| \leq 10^{-5}$.

2. Incompatible of type 2 Let us consider two matrices A and B of the following forms: $A = \begin{pmatrix} 0 & 2/7 & 7/12 & 1/4 \\ 1/2 & 0 & 5/12 & 2/4 \\ 1/2 & 5/7 & 0 & 1/4 \end{pmatrix}$ and
$B = \begin{pmatrix} 0 & 1/6 & 3/6 & 2/6 \\ 1/5 & 0 & 2/5 & 2/5 \\ 4/17 & 9/17 & 0 & 4/17 \end{pmatrix}$
It can be easily checked that here also matrices A and B are incompatible as not all the cross product ratios are equal for both the matrices.

   (a) However under the Power divergence criterion D_1, we obtained $D_1^{(10)} = 0.00002722612$ and for this particular value of D_1, the corresponding minimally incompatible P is given

by

$$P = \begin{pmatrix} 0 & 0.05852942 & 0.1625191 & 0.08997387 \\ 0.07462162 & 0 & 0.1314126 & 0.15039219 \\ 0.07903001 & 0.17384650 & 0 & 0.07967465 \end{pmatrix},$$

and we stopped since in this case $|D_1^{(10)} - D_1^{(11)}| \leq 10^{-5}$, and also $|P^{(10)} - P^{(11)}| \leq 10^{-5}$.

(b) Under the Modified Renyi's divergence criterion D_2, we obtained $D_2^{(8)} = 0.00001765243$ and for this particular value of D_2, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0 & 0.05852942 & 0.1625191 & 0.08997387 \\ 0.07462162 & 0 & 0.1314126 & 0.15039219 \\ 0.07903001 & 0.17384650 & 0 & 0.07967465 \end{pmatrix},$$

and we stopped since in this case $|D_2^{(8)} - D_2^{(9)}| \leq 10^{-5}$, and also $|P^{(8)} - P^{(9)}| \leq 10^{-5}$.

(c) However under the χ^2 divergence criterion, i.e., under D_3, we obtained $D_3^{(12)} = 0.0002579339$ and for this particular value of D_3, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0 & 0.0589141 & 0.1625083 & 0.09074997 \\ 0.07465284 & 0 & 0.1315608 & 0.15056431 \\ 0.07874187 & 0.1729022 & 0 & 0.07940566 \end{pmatrix},$$

and we stopped since in this case $|D_3^{(12)} - D_4^{(13)}| \leq 10^{-5}$, and also $|P^{(12)} - P^{(13)}| \leq 10^{-5}$.

(d) Now under D_4, we obtained $D_4^{(10)} = 0.0002014352$ and for this particular value of D_4, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0 & 0.0617139 & 0.1633906 & 0.08587921 \\ 0.07817756 & 0 & 0.1272337 & 0.15207774 \\ 0.08089520 & 0.1719500 & 0 & 0.07868216 \end{pmatrix},$$

and we stopped since in this case $|D_4^{(10)} - D_5^{(11)}| \leq 10^{-5}$, and also $|P^{(10)} - P^{(11)}| \leq 10^{-5}$.

(e) For D_5 the $D_5^{(13)} = 0.00002579339$ and for this particular value of D_5, the corresponding minimally incompatible P

is given by

$$P = \begin{pmatrix} 0.00000000 & 0.0589141 & 0.1625083 & 0.09074997 \\ 0.07465284 & 0.0000000 & 0.1315608 & 0.15056431 \\ 0.07874187 & 0.1729022 & 0.0000000 & 0.07940566 \end{pmatrix},$$

and we stopped since in this case $|D_5^{(13)} - D_6^{(14)}| \leq 10^{-5}$, and also $|P^{(13)} - P^{(14)}| \leq 10^{-5}$.

(f) Furthermore under D_6 we obtained $D_6^{(15)} = 0.0002668771$ and for this particular value of D_6, the corresponding minimally incompatible P is given by

$$P = \begin{pmatrix} 0 & 0.05906419 & 0.1611827 & 0.09070318 \\ 0.07520482 & 0 & 0.1312231 & 0.15042760 \\ 0.07940207 & 0.17338110 & 0 & 0.07941121 \end{pmatrix},$$

and we stopped since in this case $|D_6^{(15)} - D_6^{(16)}| \leq 10^{-5}$, and also $|P^{(15)} - P^{(16)}| \leq 10^{-5}$.

1.7 Comments on the behavior of the divergence measures

Since Mahalanobis (1936) introduced the concept of distance between two probability distributions, several coefficients have been suggested in statistical literature to reflect the fact that some probability distributions are "closer together" than others and consequently that it may be "easier to distinguish" between a pair of distributions which are "far from each other" than between those which are closer. Such coefficients have been variously called measures of distance between two distributions (Adhikari and Joshi, 1956), measures of separation (Rao, 1949, 1954), measures of discriminatory information (Chernoff, 1952, Kullback, 1959) and measures of variation-distance (Kolmogorov, 1963). Many of the currently used tests, such as likelihood ratio, the Chi-square, the score and Wald tests, can in fact be shown to be defined in terms of appropriate distance measures.

While the cited coefficients have not all been introduced for exactly the same purpose, they all possess the common property of increasing as the two distributions under study are "far from each other". In this chapter, a coefficient with this property has been called a divergence measure between two probability distributions.

In this chapter we focussed our attention in finding the optimum value of p_{ij} based on various measures of divergence when the given two conditional probability matrices A and B are incompatible. Furthermore we have done a comparative study based on an iterative algorithm provided for each of the divergence measures mentioned in this chapter.

Based on our iterative study it has been found that the performance of all the measures of divergence is similar. We have obtained almost the same minimally incompatible P, under the same convergence criterion set for all of them. This has been found to be true in both the situations where the two conditional probability matrices A and B have all the elements strictly positive and also when they have zeros appearing in the same position. To this end we can say that in a search for a most nearly compatible P in an incompatible set-up, one can use any of those measures. However among them the performance of the Kullback-Leibler measure of divergence may be preferred over the remaining divergence measures since it achieves the same minimally incompatible P but at a faster rate of convergence as can be easily verified with all the examples tried in this chapter.

A careful and detailed study on the use of several measures of divergence, distinct from those treated in this chapter, in the problem of finding a minimally compatible P, under an incompatible set-up has been made by authors such as Arnold et.al (1998, 2001), Arnold and Gokhale (1998). A rigorous proof of the convergence of the iterative algorithms introduced in those papers, as well those discussed in the present chapter, is not yet available.

Chapter 2

An alternative approach to compatibility for the two conditional probability matrices under the discrete set-up

2.1 Introduction

Specification of joint distributions by means of conditional densities has received considerable attention in the last decade or so. Possible applications may be found in the area of model building in classical statistical settings and in the elicitation and construction of multiparameter prior distributions in Bayesian scenarios. For example suppose that $\underline{X} = (X_1, X_2, \cdots, X_k)$ is a k-dimensional random vector taking on values in the finite range set $\underline{\mathcal{X}}_1 \times \underline{\mathcal{X}}_2 \times \cdots \times \underline{\mathcal{X}}_k$ where $\underline{\mathcal{X}}_i$ denote the possible values of $X_i, i = 1, 2, \cdots, k$. Efforts to ascertain an appropriate distribution for $\underline{X}$ frequently involve acceptance or rejection of a series of bets about the stochastic behavior of $\underline{X}$. Let us consider that in this situation we are facing a question of whether or not to accept with odds 4 to 1 a bet that X_1 is equal to 1. Then if we accept the bet then it puts a bound on the probability that X=1.

The basic problem is most easily visualized in the finite discrete case. Several alternative approaches exist in the literature with regard to the problem of determination of the possible compatibility of two families of conditional distributions (Arnold and Press, 1989; Arnold and Gokhale, 1994; Cacoullos and Papageorgiou, 1995; Wesolowski, 1995). In addition, the problem of determining most nearly compatible distributions, in the absence of compatibility, has been addressed (Arnold and Gokhale 1998; Arnold, Castillo and Sarabia (1999, 2001)). In this chapter, focussed on the finite discrete case, we take a closer look at the compatibility problem, viewing it as a problem involving linear equations in restricted domains. The issue of near compatibility is also discussed using the concept of ε-compatibility which is mentioned in chapter 1. Furthermore we also focus our attention on situations where we have incomplete (or partial) information on (either or both) of the two conditional probability matrices A and B, under the compatible set-up.

In particular we transform the problem of compatibility to a linear programming problem and we derive conditions for compatibility based on the rank of a matrix D, whose elements are functions of the two conditional probability matrices A and B. It will be found that the problem of compatibility, with our approach, is reduced to a large extent in the sense that we are left with only finding a solution to a set of IJ equations in $(I-1)$ unknowns with non negativity constraints, where I and J are the dimensions of the matrices A and B. However in this chapter we will mainly focus on cases in which $I = 2, 3, 4$ and $J = 2, 3$. Detailed discussion will be provided for these cases and compatibility in higher dimensions will be discussed later on. The remainder of this chapter is organized in the following way. In Section 2, we will discuss the concept of compatibility for any two conditional probability matrices A and B. In Section 3, we will discuss in detail how the problem of compatibility can be looked upon as a linear programming problem. In Section 4, we will discuss the alternative approach to compatibility for the (2×2) and (3×3) cases. In Section 5, we will discuss the problem of compatibility when we have incomplete specification of matrices A or B or both.

2.2 Compatibility

Let A and B be two $(I \times J)$ matrices with non-negative elements such that $\sum_{i=1}^{I} a_{ij} = 1$, $\forall j = 1, \ldots, J$ and $\sum_{j=1}^{J} b_{ij} = 1$, $\forall i = 1, 2, \ldots, I$. Without loss of generality it can be assumed that $I \leq J$. Matrices A and B are said to form a compatible conditional specification for the distribution of (X, Y) if there exists some $(I \times J)$ matrix P with non-negative entries p_{ij} and with $\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} = 1$, such that for every (i, j)

$$a_{ij} = \frac{p_{ij}}{p_{.j}},$$

and

$$b_{ij} = \frac{p_{ij}}{p_{i.}},$$

where $p_{.j} = \sum_{j=1}^{J} p_{ij}$ and $p_{i.} = \sum_{i=1}^{I} p_{ij}$. If such a matrix P exists then, if we assume that

$$p_{ij} = P(X = x_i, Y = y_j),$$

$i = 1, 2, \cdots, I$, $j = 1, 2, \cdots, J$, we will have

$$a_{ij} = P(X = x_i | Y = y_j),$$

$i = 1, 2, \cdots, I$, $j = 1, 2, \cdots, J$, and

$$b_{ij} = P(Y = y_j | X = x_i),$$

$i = 1, 2, \cdots, I,\ j = 1, 2, \cdots, J.$

Equivalently A and B are compatible if there exist stochastic vectors $\underline{\tau} = (\tau_1, \tau_2, \cdots, \tau_J)$, and $\underline{\eta} = (\eta_1, \eta_2, \cdots, \eta_I)$ such that

$$a_{ij}\tau_j = b_{ij}\eta_i,$$

for every(i, j). In the case of compatibility, $\underline{\eta}$ and $\underline{\tau}$ can be readily interpreted as the resulting marginal distributions of X and Y, respectively. For any probability vector $\eta = (\eta_1, \eta_2, \ldots, \eta_I)$, $p_{ij} = b_{ij}\eta_i$ is a probability distribution on the IJ cells. So the conditional probability matrix denoted by A, and it's elements (a_{ij}) will be given by

$$a_{ij} = \frac{p_{ij}}{\sum_{s=1}^{I} p_{sj}} = \frac{b_{ij}\eta_i}{\sum_{s=1}^{I} b_{sj}\eta_s}, \tag{2.1}$$

for every(i, j).

If A and B are compatible then

$$a_{ij} \sum_{s=1}^{I} b_{sj}\eta_s = b_{ij}\eta_i.$$

Then we have

$$\tau_j = \sum_{s=1}^{I} b_{ij}\eta_s,$$

$\forall j = 1, \ldots, J$. In that case expressions given in (2.1) can be written as

$$a_{ij} \sum_{s=1}^{I} b_{sj}\eta_s - b_{ij}\eta_i = 0.$$

In matrix notation the above can be written as

$$D\underline{\eta} = 0, \tag{2.2}$$

where D is a matrix of dimension $IJ \times I$ and the above equation is a homogeneous system of IJ equations in I unknowns η_i . Through well known matrix operations (such as left-multiplication by non-singular matrices) it's rows can be reduced to at most I rows with non-zero elements (the so called "Row Echelon form"). Now let this reduced system be denoted by $D_r\underline{y} = 0$ where $\underline{y} = (y_1, y_2, \ldots, y_I)'$. Matrices A and B are compatible if the system $D_r y = 0$ has a solution y^* of non-negative elements with at least one positive element. If such a y^* exists it can be scaled to arrive at a probability vector. However A and B are not compatible if the only solution with non-negative elements of $D_r y = 0$ is the null vector. In order to examine whether or not such a solution y^* of $D_r y = 0$ exists (especially when I is large), the methodology of linear programming may be used. Specifically, consider the problem of maximizing the objective function $\sum_i y_i$, subject to (a) the non-negativity constraints $\sum_{i=1}^{I} y_i \geq 0$, (b) the equality constraints

$D_r y = 0$, and (c) the constraint $\sum_i y_{i=1}^I \leq 1$. If the maximum of the objective function is positive, then the corresponding optimizing vector is y^*, which can be scaled into the probability vector η^*. If the maximum is 0, then A and B are not compatible. First of all we will consider some results related to compatibility of two matrices A and B and later on we will discuss an alternative approach to compatibility.

2.3 Compatibility of two matrices A and B

We know that if the matrices A and B are compatible then $a_{ij}p_{.j} = b_{ij}p_{i.}$, for every $i = 1, 2 \cdots, I; j = 1, 2, \cdots J$. Equivalently we can write

$$a_{ij}\sum_{s=1}^{I} p_{sj} - b_{ij}\sum_{k=1}^{J} p_{ik} = 0,$$

for every $i = 1, 2 \cdots, I; j = 1, 2, \cdots J$, which is again can be written as

$$a_{ij}[p_{1j} + p_{2j} + \cdots + p_{ij} + \cdots + p_{Ij}] - b_{ij}[p_{i1} + p_{i2} + \cdots + p_{ij} + \cdots + p_{iJ}] = 0$$

for every $i = 1, 2 \cdots, I; j = 1, 2, \cdots J$.

In matrix notation the above system of linear equations can be written as

$$C\underline{p} = 0,$$

where C contains elements calculated from those of A and B and is a matrix of dimension $IJ \times IJ$ and $\underline{p}^{(IJ\times 1)} = (p_{11}, p_{12}, \cdots, p_{IJ})^T$. We need to show that the solution space Ω, (say) of these equations is given by $(I-M)\underline{z}$, where M is an idempotent matrix and $\underline{z}^{(IJ\times 1)}$ is any arbitrary vector of dimension $IJ \times 1$.

Proof: We may consider $M = C^-C$, where C^- is the g-inverse of the matrix C and we have considered the g-inverse of C since rank$(C^{IJ\times IJ}) < IJ$. Next observe that

$$\begin{aligned} M &= C^-C \\ M^2 &= C^-CC^-C \\ &= C^-C \\ &= M, \end{aligned}$$

which follows from the definition of g-inverse since $CM = CC^-C = C$.

Hence each of the IJ columns $\underline{h}_1, \underline{h}_2, \ldots, \underline{h}_{IJ}$ of $(I - M)$ are orthogonal to the rows of C. But

$$(I - M)^2 = I - M - M + M^2 = I - M,$$

since $M^2 = M$. And

$$\text{rank}(I - M) = \text{tr}(I - M) = \text{tr}(I) - tr(M) = IJ - r,$$

where $r = \text{rank}(C) = \text{rank(M)}$ and $tr \equiv trace$. So only $(IJ - r)$ of the column vectors $\underline{h}_1, \underline{h}_2, \ldots, \underline{h}_{IJ}$ are linearly independent and without loss of generality we may consider them to be

$$\underline{h}_1, \underline{h}_2, \ldots, \underline{h}_{IJ-r}.$$

Again since C is an $(IJ \times IJ)$ matrix of rank r, it's rows are IJ-vectors and therefore we can find at most $(IJ-r)$ linearly independent vectors orthogonal to them and $\underline{h}_1, \underline{h}_2, \ldots, \underline{h}_{IJ-r}$ is one such set. If there is any other vector orthogonal to the rows of C, it must be a linear combination of $\underline{h}_1, \underline{h}_2, \ldots, \underline{h}_{IJ-r}$. Now since $C\underline{p} = 0$, $\underline{p}$ is orthogonal to the rows of C and so any vector $\underline{p}$ satisfying $C\underline{p} = 0$ must be a linear combination of $\underline{h}_1, \underline{h}_2, \ldots, \underline{h}_{IJ-r}$.

But equivalently we can say that $\underline{p}$ will be a combination of $\underline{h}_1, \underline{h}_2, \cdots, \underline{h}_{IJ}$, because $\underline{h}_{IJ-r+1}, \underline{h}_{IJ-r+2}, \ldots, \underline{h}_{IJ}$ are combinations of $\underline{h}_1, \underline{h}_2, \cdots, \underline{h}_{IJ-r}$. Hence $\underline{p}$ must be of the form

$$\begin{aligned}
\underline{p} &= z_1\underline{h}_1 + z_2\underline{h}_2 + z_3\underline{h}_3 + \ldots + z_{IJ}\underline{h}_{IJ} \\
&= (\underline{h}_1, \ldots, \underline{h}_{IJ})z \\
&= (I - M)\underline{z},
\end{aligned}$$

for some choices of $\underline{z} = (z_1, \cdots, z_{IJ})'$.

Conversely if the above holds then

$$\begin{aligned} CM &= C(I-M)\underline{z} \\ &= (C-CM)\underline{z} \\ &= 0, \end{aligned}$$

because of the fact that $CM = C$. Hence the proof.

Next we have the following two propositions:

1. If there exists one vector in Ω such that all its elements are non-negative and at least one element is strictly positive then A and B are compatible.

   Proof: We have from our earlier result the solution space as

   $$\Omega = (I-M)\underline{z}, \text{where} M = C^{-}C.$$

   Now suppose $\underline{z} = (z_1, z_2, \ldots, z_J)'$, where $z_u = (z_{1u}, \ldots, z_{Iu}).'$ Let us consider $z_i \geq 0$ and (say) $z_j > 0$, where $j \in i$, meaning at least one of the z_i's is strictly zero. Then we can write $W = \sum_{u=1}^{J} z_{Iu}$. Obviously $W > 0$. So we can rewrite $\underline{z}$ as

   $$\begin{aligned} \underline{z}_{new} &= (\frac{z_1}{W}, \frac{z_2}{W}, \ldots, \frac{z_J}{W}) \\ &= (p_1, p_2, \ldots, p_J)', \end{aligned}$$

which is a valid probability distribution since $\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{z_{ij}}{W} = 1$. Hence the proof.

2. If every vector in Ω has non-positive elements then A and B are not compatible.

   Proof: In this situation suppose that we have as before $\underline{z}^{(IJ\times 1)} = (\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_J)'$, with $\underline{z}_i \geq 0$, $\forall i = 1(1)J$. But if we consider the trivial situation or the possibility of every $\underline{z}_i = 0$ then all the elements of $\underline{z}^{(IJ\times 1)}$ are zero which implies that it can not be a valid probability distribution.

Next observe that since $\sum\sum p_{ij}$, is a linear function in p_{ij}, the problem is equivalent to the following linear programming (LP) problem: Maximize $f(\underline{p}) = \sum\sum p_{ij}$, subject to $C\underline{p} = 0$ and $1 \geq p_{ij} \geq 0\ \forall(i,j)$.

Theorem 4. *In the above LP problem max $f(\underline{p}) > 0$ if and only if A and B are compatible.*

Proof: Note that if max $f(\underline{p}) > 0$ then at least one of the element in $\underline{p} = (p_{11}, \ldots, p_{IJ})'$ is strictly positive so we may consider $p_{uv} > 0$, where $u \in i, v \in j$ and $p_{ij} \geq 0, \forall u \neq i, v \neq j$. Then obviously we have

$$max f(\underline{p}) = \sum_i \sum_j p_{ij} = \sum_{i\neq u}\sum_{j\neq v} p_{ij} + p_{uv} > 0.$$

Next note that we can write

$$\underline{p}$$

$$= (\underbrace{p_{11}, p_{12}, \cdots, p_{1J}}, \underbrace{p_{21}, p_{22}, \cdots, p_{2J}}, \cdots, \underbrace{p_{I1}, p_{I2}, \cdots, p_{IJ}}).$$

Then we can rewrite $\underline{p}$ as

$$\begin{aligned} \underline{p} &= (\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_I)' \\ &= (\frac{\underline{p}_1}{max f(p)}, \frac{\underline{p}_2}{max f(p)}, \ldots, \frac{\underline{p}_I}{max f(p)}). \end{aligned}$$

Then it becomes a valid probability distribution and then it follows from our previous result that A and B are compatible.

2.4 An alternative approach to compatibility

First we will consider the following theorem:

Theorem 5. *For any two given conditional probability matrices A and B of dimension $(I \times J)$, they are compatible if rank$(D^{(IJ \times I)}) \leq I - 1$, with equality when there exists a unique solution for the unknown $\eta_i, \forall i$.*

Proof: Note that rank$(D^{(IJ \times I)}) \leq \min(IJ, I) = I$. Now when D has full rank, i.e., rank$(D) = I$, then the only solution to the equation $D\underline{\eta} = 0$ is the null vector(trivial solution). So matrices A and B are incompatible.

Next if we have $\text{rank}(D^{(IJ \times I)}) \leq I - 1$, the number of equations $(IJ) >$ number of unknowns $(I - 1)$, so we must have a non-trivial solution. If the non-trivial solution is positive then it can be appropriately scaled to arrive at a probability vector $\underline{\eta}^*$. Hence the two matrices A and B will be compatible. However in this case the system of equations is not homogeneous and at most we will have $(I - 1)$ solutions.

Again when $\text{rank}(D) = I - 1$, then we have $(I - 1)$ unknowns, (subject to the linear constraint $\sum_{i=1}^{I} \eta_i = 1$) and $(I - 1)$ equations (excluding the redundant equations from the total set of IJ equations) and the system of linear equations is homogeneous so that there exists a unique solution. This completes the proof.

Note: This theorem will be in particular useful in situations when the two conditional matrices A and B have zeroes as elements appearing in the same position in them and we can not guarantee the existence of a compatible matrix P by the cross product ratio criterion.

Next we will discuss this in the context of compatibility for (2×2), (3×3), and (4×3) cases with examples.

2.4.1 Compatibility in (2×2) case

We will provide the outline of the proof for $I = 2, J = 2$ and similar argument will follow in higher dimensions.

Proof: In a (2×2) case the D-matrix is given by

$$D = \begin{pmatrix} b_{11}(a_{11}-1) & a_{11}b_{21} \\ b_{12}(a_{12}-1) & a_{12}b_{22} \\ a_{21}b_{11} & b_{21}(a_{21}-1) \\ a_{22}b_{12} & b_{22}(a_{22}-1) \end{pmatrix}.$$

Equivalently we can write D as (because $a_{11}+a_{21}=1$ and $a_{12}+a_{22}=1$)

$$D = \begin{pmatrix} -b_{11}(a_{21}) & a_{11}b_{21} \\ -b_{12}(a_{22}) & a_{12}b_{22} \\ a_{21}b_{11} & -b_{21}(a_{11}) \\ a_{22}b_{12} & -b_{22}(a_{12}) \end{pmatrix}.$$

Next we consider the elementary row-transformations

- (new(row 3)=row 1+row 3
- new(row 4)=row4+row2

So that our D matrix reduces to

$$D = \begin{pmatrix} -b_{11}(a_{21}) & a_{11}b_{21} \\ -b_{12}(a_{22}) & a_{12}b_{22} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

However in a (2×2) case if A and B are compatible then cprs(A)=cprs(B), so that we can write

$$a_{12}a_{21}b_{22}b_{11} = a_{11}a_{22}b_{21}b_{12}. \tag{2.3}$$

Again we apply the elementary row transformations

- new(row1)=row $1\times(a_{12}b_{22})$
- new(row2)=row $2\times(a_{11}b_{21})$.

So that our D-matrix reduces to

$$D = \begin{pmatrix} -b_{11}(a_{21})a_{12}b_{22} & a_{11}b_{21}a_{12}b_{22} \\ -b_{12}(a_{22})a_{11}b_{21} & a_{12}b_{22}a_{11}b_{21} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Now because of equation (2.3) and by applying the elementary row transformation new(row 2)=row 2-row 1 our D-matrix reduces to

$$D = \begin{pmatrix} -b_{11}(a_{21})a_{12}b_{22} & a_{11}b_{21}a_{12}b_{22} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Hence rank(D)=1.

Next we have to show that rank(D) is bigger than 1 if A and B are not compatible and we will show here that in a (2×2) case rank(D)=2.

Proof: It actually follows from our previous result where after some elementary row-transformation our D-matrix reduces to

$$D = \begin{pmatrix} -b_{11}(a_{21})a_{12}b_{22} & a_{11}b_{21}a_{12}b_{22} \\ -b_{12}(a_{22})a_{11}b_{21} & a_{12}b_{22}a_{11}b_{21} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

When A and B are incompatible (2.3) does not hold so the rows of D are not proportional. Therefore rank(D) = 2. This completes our proof.

2.4.2 Examples of (2×2) case

First of all we will consider a situation where the two matrices A and B are compatible with all the elements strictly positive.

- Suppose we have

$$A = \begin{pmatrix} 1/4 & 2/3 \\ 3/4 & 1/3 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 1/3 & 2/3 \\ 3/4 & 1/4 \end{pmatrix}.$$

In this case the resulting D-matrix is given by

$$D = \begin{pmatrix} -0.2500000 & 0.1875000 \\ -0.2222222 & 0.1666667 \\ 0.2500000 & -0.1875000 \\ 0.2222222 & -0.1666667 \end{pmatrix}.$$

In this case rank(D)=1. So A and B are compatible as can be verified by checking the cross product ratios of A and B. The solution for the joint probability distribution in this case is given by

$$P = \begin{pmatrix} 1/7 & 2/7 \\ 3/7 & 1/7 \end{pmatrix}.$$

- Next we consider two matrices A and B of the following forms:

$$A = \begin{pmatrix} 1/7 & 3/4 \\ 6/7 & 1/4 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 2/5 & 3/5 \\ 3/8 & 5/8 \end{pmatrix}.$$

The resulting D-matrix in this case is given by

$$D = \begin{pmatrix} -0.3428571 & 0.05357143 \\ -0.1500000 & 0.46875000 \\ 0.3428571 & -0.05357143 \\ 0.1500000 & -0.46875000 \end{pmatrix}.$$

In this case rank(D)=2, so A and B are incompatible.

2.4.3 Compatibility in (3×3) case

1. First of all let us consider a compatible case of type 1, where by type 1, we mean a case in which all the elements of the matrices A and B are strictly positive.

   Let

$$A = \begin{pmatrix} 1/5 & 2/7 & 3/8 \\ 3/5 & 2/7 & 1/8 \\ 1/5 & 3/7 & 1/2 \end{pmatrix},$$

   while B has the following form:

$$B = \begin{pmatrix} 1/6 & 1/3 & 1/2 \\ 1/2 & 1/3 & 1/6 \\ 1/8 & 3/8 & 1/2 \end{pmatrix}.$$

In this case the corresponding D-matrix is given by

$$D^{(9\times3)} = \begin{pmatrix} -0.13333333 & 0.10000000 & 0.0250000 \\ -0.23809524 & 0.09523810 & 0.1071429 \\ -0.31250000 & 0.06250000 & 0.1875000 \\ 0.10000000 & -0.20000000 & 0.0750000 \\ 0.09523810 & -0.23809524 & 0.1071429 \\ 0.06250000 & -0.14583333 & 0.0625000 \\ 0.03333333 & 0.10000000 & -0.1000000 \\ 0.14285714 & 0.14285714 & -0.2142857 \\ 0.25000000 & 0.08333333 & -0.2500000 \end{pmatrix}.$$

Note that in this case rank(D)=2 and hence A and B are compatible. Now solving the equation as mentioned earlier the solution for the marginal of X is given by $\underline{\eta} = (0.3, 0.3, 0.4)$.

2. Next consider a compatible case of type 2, where by type 2, we mean a case in which some of the elements of the two matrices A and B are zeros which appear in the same positions in both A and B. Suppose that we have A and B of the following forms:

$$A = \begin{pmatrix} 1/3 & 0 & 2/3 \\ 0 & 1/2 & 1/3 \\ 2/3 & 1/2 & 0 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 1/3 & 0 & 2/3 \\ 0 & 1/2 & 1/2 \\ 2/3 & 1/3 & 0 \end{pmatrix}.$$

In this case the corresponding D-matrix is given by

$$D^{(9\times3)} = \begin{pmatrix} -0.2222222 & 0.0000000 & 0.2222222 \\ 0.0000000 & 0.0000000 & 0.0000000 \\ -0.2222222 & 0.3333333 & 0.0000000 \\ 0.0000000 & 0.0000000 & 0.0000000 \\ 0.0000000 & -0.2500000 & 0.1666667 \\ 0.2222222 & -0.3333333 & 0.0000000 \\ 0.2222222 & 0.0000000 & -0.2222222 \\ 0.0000000 & 0.2500000 & -0.1666667 \\ 0.0000000 & 0.0000000 & 0.0000000 \end{pmatrix}.$$

In this case rank(D)=2 and indeed A and B are compatible. Solving the equation we get the solution for the marginal of X given by $\underline{\eta} = (0.375, 0.250, 0.375)$. The joint probability distribution in this case is given by

$$P = \begin{pmatrix} 0.125 & 0.000 & 0.250 \\ 0.000 & 0.125 & 0.125 \\ 0.250 & 0.125 & 0.000 \end{pmatrix}.$$

3. Next we consider an incompatible case of type 1 in which A and B have the following forms:

$$A = \begin{pmatrix} 0.2 & 0.3 & 0.1 \\ 0.1 & 0.4 & 0.4 \\ 0.7 & 0.3 & 0.5 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 0.2 & 0.1 & 0.7 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}.$$

In this case the resulting D-matrix is given by

$$D^{(9\times 3)} = \begin{pmatrix} -0.16 & 0.06 & 0.02 \\ -0.07 & 0.12 & 0.12 \\ -0.63 & 0.03 & 0.05 \\ 0.02 & -0.27 & 0.01 \\ 0.04 & -0.24 & 0.16 \\ 0.28 & -0.18 & 0.20 \\ 0.14 & 0.21 & -0.03 \\ 0.03 & 0.12 & -0.28 \\ 0.35 & 0.15 & -0.25 \end{pmatrix}.$$

In this case $\mathrm{rank}(D) = 3$ and hence A and B are not compatible.

4. Next we consider an incompatible case of type 2 in which A and B have the following forms

$$A = \begin{pmatrix} 0 & 1/3 & 0 \\ 1 & 1/3 & 1/2 \\ 0 & 1/3 & 1/2 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 0 & 1 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/5 & 4/5 \end{pmatrix}.$$

In this case the resulting D-matrix is given by

$$D^{(9\times3)} = \begin{pmatrix} 0.0000000 & 0.0000000 & 0.00000000 \\ -0.6666667 & 0.1666667 & 0.06666667 \\ 0.0000000 & 0.0000000 & 0.00000000 \\ 0.0000000 & 0.0000000 & 0.00000000 \\ 0.3333333 & -0.3333333 & 0.06666667 \\ 0.0000000 & -0.1250000 & 0.40000000 \\ 0.0000000 & 0.0000000 & 0.00000000 \\ 0.3333333 & 0.1666667 & -0.13333333 \\ 0.0000000 & 0.1250000 & -0.40000000 \end{pmatrix}.$$

In this case $\mathrm{rank}(D) = 3$ and hence A and B are not compatible.

5. Again let us consider a compatible case of type 1 where $I = 4$ and $J = 3$ in which A and B have the following forms

$$A = \begin{pmatrix} 1/9 & 2/8 & 1/7 \\ 5/9 & 1/8 & 3/7 \\ 1/9 & 4/8 & 2/7 \\ 2/9 & 1/8 & 1/7 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 1/4 & 2/4 & 1/4 \\ 5/9 & 1/9 & 3/9 \\ 1/7 & 4/7 & 2/7 \\ 2/4 & 1/4 & 1/4 \end{pmatrix}.$$

In this case the resulting D-matrix is given by

$$D^{(9\times 3)} = \begin{pmatrix} -0.22222222 & 0.06172840 & 0.01587302 & 0.05555556 \\ -0.37500000 & 0.02777778 & 0.14285714 & 0.06250000 \\ -0.21428571 & 0.04761905 & 0.04081633 & 0.03571429 \\ 0.13888889 & -0.24691358 & 0.07936508 & 0.27777778 \\ 0.06250000 & -0.09722222 & 0.07142857 & 0.03125000 \\ 0.10714286 & -0.19047619 & 0.12244898 & 0.10714286 \\ 0.02777778 & 0.06172840 & -0.12698413 & 0.05555556 \\ 0.25000000 & 0.05555556 & -0.28571429 & 0.12500000 \\ 0.07142857 & 0.09523810 & -0.20408163 & 0.07142857 \\ 0.05555556 & 0.12345679 & 0.03174603 & -0.38888889 \\ 0.06250000 & 0.01388889 & 0.07142857 & -0.21875000 \\ 0.03571429 & 0.04761905 & 0.04081633 & -0.21428571 \end{pmatrix}.$$

In this case $\text{rank}(D) = 3$ and hence A and B are not compatible.

2.4.4 Proof that rank(D)=2 when A and B are compatible in a (3×3) case

The form of the D-matrix in a (3×3) case is given by

$$D^{(9\times 3)} = \begin{pmatrix} b_{11}(a_{11}-1) & a_{11}b_{21} & a_{11}b_{31} \\ b_{12}(a_{12}-1) & a_{12}b_{22} & a_{12}b_{32} \\ b_{13}(a_{13}-1) & a_{13}b_{23} & a_{13}b_{33} \\ a_{21}b_{11} & b_{21}(a_{21}-1) & a_{21}b_{31} \\ a_{22}b_{12} & b_{22}(a_{22}-1) & a_{22}b_{32} \\ a_{23}b_{13} & b_{23}(a_{23}-1) & a_{23}b_{33} \\ a_{31}b_{11} & a_{31}b_{21} & b_{31}(a_{31}-1) \\ a_{32}b_{12} & a_{32}b_{22} & b_{32}(a_{32}-1) \\ a_{33}b_{13} & a_{33}b_{23} & b_{33}(a_{33}-1) \end{pmatrix}.$$

However if matrices A and B are compatible then all possible cross product ratio of A are equal to the corresponding cross product ratios of B. First of all we apply the following elementary row operations

- new(row1)=row 1+row 4+row 7
- new(row2)=row 2+row 5+row 8
- new(row3)=row 3+row 6+row 9.

So that matrix D reduces to

$$D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ a_{21}b_{11} & b_{21}(a_{21}-1) & a_{21}b_{31} \\ a_{22}b_{12} & b_{22}(a_{22}-1) & a_{22}b_{32} \\ a_{23}b_{13} & b_{23}(a_{23}-1) & a_{23}b_{33} \\ a_{31}b_{11} & a_{31}b_{21} & b_{31}(a_{31}-1) \\ a_{32}b_{12} & a_{32}b_{22} & b_{32}(a_{32}-1) \\ a_{33}b_{13} & a_{33}b_{23} & b_{33}(a_{33}-1) \end{pmatrix}.$$

Next consider the following elementary row and column operations:

- new(row4)=$\frac{\text{row4}}{a_{21}}$
- new(row5)=$\frac{\text{row5}}{a_{22}}$
- new(row6)=$\frac{\text{row4}}{a_{23}}$
- new(row7)=$\frac{\text{row7}}{a_{31}}$
- new(row8)=$\frac{\text{row8}}{a_{32}}$
- new(row9)=$\frac{\text{row4}}{a_{33}}$
- new(col4)=col 4+col 5+col 6
- new(col7)=col 7+col 8+col 9.

So that the D matrix has the form:

$$D^{(9\times 3)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1-(\frac{b_{21}}{a_{21}}+\frac{b_{22}}{a_{22}}+\frac{b_{23}}{a_{23}}) & 1 \\ b_{12} & b_{22}(1-\frac{1}{a_{22}}) & b_{32} \\ b_{13} & b_{23}(1-\frac{1}{a_{23}}) & b_{33} \\ 1 & 1 & 1-(\frac{b_{31}}{a_{31}}+\frac{b_{32}}{a_{32}}+\frac{b_{33}}{a_{33}}) \\ b_{12} & b_{22} & b_{32}(1-\frac{1}{a_{32}}) \\ b_{13} & b_{23} & b_{33}(1-\frac{1}{a_{33}}) \end{pmatrix}.$$

Then we consider the following

- new(row 5)=row 5+row 6,
- new(row 8)=row 8+row 9
- new(row 5)=row 5-row 8
- new(row 4)=row 4-row 7
- new(row 6)=row 6-row 8.

Then our D matrix reduces to

$$D^{(9\times 3)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -(\frac{b_{21}}{a_{21}} + \frac{b_{22}}{a_{22}} + \frac{b_{23}}{a_{23}}) & (\frac{b_{31}}{a_{31}} + \frac{b_{32}}{a_{32}} + \frac{b_{33}}{a_{33}}) \\ 0 & -(\frac{b_{22}}{a_{22}} + \frac{b_{23}}{a_{23}}) & (\frac{b_{32}}{a_{32}} + \frac{b_{33}}{a_{33}}) \\ 0 & -\frac{b_{23}}{a_{23}} & \frac{b_{33}}{a_{33}} \\ 1 & 1 & 1 - (\frac{b_{31}}{a_{31}} + \frac{b_{32}}{a_{32}} + \frac{b_{33}}{a_{33}}) \\ -b_{11} & -b_{21} & -b_{31} - (\frac{b_{32}}{a_{32}} + \frac{b_{33}}{a_{33}}) \\ b_{13} & b_{23} & b_{33}(1 - \frac{1}{a_{33}}) \end{pmatrix}.$$

Again we consider new (row 4)=row 4-row 5, new(row5)=row 5-row 6, so that our D matrix reduces to

$$D^{(9\times 3)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -\frac{b_{21}}{a_{21}} & \frac{b_{31}}{a_{31}} \\ 0 & -\frac{b_{22}}{a_{22}} & \frac{b_{32}}{a_{32}} \\ 0 & -\frac{b_{23}}{a_{23}} & \frac{b_{33}}{a_{33}} \\ 1 & 1 & 1-(\frac{b_{31}}{a_{31}}+\frac{b_{32}}{a_{32}}+\frac{b_{33}}{a_{33}}) \\ -b_{11} & -b_{21} & -b_{31}-(\frac{b_{32}}{a_{32}}+\frac{b_{33}}{a_{33}}) \\ b_{13} & b_{23} & b_{33}(1-\frac{1}{a_{33}}) \end{pmatrix}.$$

Note that in this case we have rank($D^{(9\times 3)}$)$\leq \min(9,3) = 3$. However for our matrix D, the determinants of all possible submatrices of order (3×3) are zero and hence rank($D^{(9\times 3)}$)<3. Now we know that the rank of a matrix is the highest order non-vanishing determinant. Let us consider the determinant of any submatrix of order (2×2),

$$B = \begin{pmatrix} -b_{11} & -b_{21} \\ b_{13} & b_{23} \end{pmatrix}.$$

The determinant for the matrix B is given by

$$\begin{aligned} \det(B) &= -b_{11}b_{23} + b_{13}b_{21} \\ &\neq 0. \end{aligned} \tag{2.4}$$

So we have rank(D) $= I - 1 = 2$, which follows from the definition of the rank of a matrix. Hence A and B are compatible iff rank(D)=$I-1$. However if A and B are not compatible then the rows of A are not proportional to the rows of B which implies that rank(D) > 2. Hence the proof.

2.5 Study of compatibility under incomplete specification of A or B or both

In this section we will consider the problem of compatibility of two conditional probability matrix A and B under the discrete set-up when one or more than one element either in A or in B is unknown. In particular we will discuss in detail for the (2×3) cases and we will consider two different situations which are listed as follows:

- More than one element is unknown only in A.
- More than one element is unknown in both A and B.

Our objective here is to investigate what happens to the compatibility condition when we have situations as listed above.

2.5.1 Compatibility when only elements of A are unknown

1. Let us consider $I = 2$ and $J = 3$ and assume that only two elements of A are unknown while all the elements of B are known. We denote the $(i, j)^{-th}$ unknown element of A by α_{ij}. Suppose that we have

$$A = \begin{pmatrix} a_{11} & \alpha_{12} & a_{13} \\ a_{21} & \alpha_{22} & a_{23} \end{pmatrix},$$

and

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix}.$$

Note: Here we assume that all the elements corresponding to matrices A and B are strictly positive. Also A has elements such that column sums are equal to one and B has elements such that the row sums are equal to one. So that we have

$$\alpha_{11} + \alpha_{22} = 1.$$

We know that the problem of compatibility can be reduced to (in matrix notation as): $D\underline{\eta} = 0$, where D-has elements computed from the matrices A and B. In this case we have two constraints $\alpha_{12} + \alpha_{22} = 1$ and $\eta_1 + \eta_2 = 1$. So the set of equations involving α_{12} and α_{22} which are sufficient to find the unknown values (the remaining equations will be redundant), from equation (2.2), will be

$$b_{12}(\alpha_{12} - 1)\eta_1 + \alpha_{12}b_{22}\eta_2 = 0 \tag{2.5}$$

$$b_{22}(\alpha_{22} - 1)\eta_2 + \alpha_{22}b_{12}\eta_1 = 0 \tag{2.6}$$

$$b_{13}(a_{13} - 1)\eta_1 + a_{13}b_{23}\eta_2 = 0. \tag{2.7}$$

Now because of the constraint $\eta_1 + \eta_2 = 1$, we get from equation (2.7),

$$\eta_1 = \frac{a_{13}b_{23}}{a_{13}b_{23} + b_{13}(1 - a_{13})}.$$

Again substituting the value of η_1 in equation (2.7) and using the constraint $\alpha_{12} + \alpha_{22} = 1$, we get the value of α_{22}, as

$$\alpha_{22} = \frac{b_{22}(1 - \eta_1)}{b_{22}(1 - \eta_1) + b_{12}\eta_1} = \frac{b_{22}b_{13}(1 - a_{13})}{b_{22}b_{13}(1 - a_{13}) + b_{22}b_{13}a_{13}}.$$

Also the unknown value of α_{12} will be $\alpha_{12} = 1 - \alpha_{22}$.

Let us consider an example. Suppose we have two matrices A and

B which are given as follows:

$$A = \begin{pmatrix} 1/5 & \alpha_{12} & 3/4 \\ 4/5 & \alpha_{22} & 1/4 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 1/6 & 2/6 & 3/6 \\ 4/6 & 1/6 & 1/6 \end{pmatrix}.$$

Now if we are given the information that A and B are compatible then the values of α_{12} and α_{12} will be given by

$$\alpha_{22} = \frac{b_{22}b_{13}(1-a_{13})}{b_{22}b_{13}(1-a_{13}) + b_{22}b_{13}a_{13}} = \frac{\frac{1}{6}\frac{3}{6}(1-\frac{3}{4})}{\frac{1}{6}\frac{3}{6}(1-\frac{3}{4}) + \frac{2}{6}\frac{1}{6}\frac{3}{4}} = \frac{1}{3}.$$

So $\alpha_{12} = 1 - \frac{1}{3} = \frac{2}{3}$.

Note that these are the unique choices for the unknown elements in the matrix A for which the above matrices are compatible.

2. Next we consider the situation where $I = 3$ and $J = 3$ and as before denoting the unknown values of the matrix A by α_{ij}, in the $(i,j)^{-th}$ position we have (with all elements of B known). Suppose

that we have

$$A = \begin{pmatrix} a_{11} & \alpha_{12} & a_{13} \\ a_{21} & \alpha_{22} & a_{23} \\ a_{31} & \alpha_{32} & a_{33} \end{pmatrix},$$

and

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}.$$

So that the linear constraints in this case are as follows (considering the fact that the column sums of the matrix A are each equal to one and $\eta_i, i = 1, 2, 3$ are the marginal probability vectors corresponding to X)

$$\alpha_{12} + \alpha_{22} + \alpha_{32} = 1 \tag{2.8}$$

$$\eta_1 + \eta_2 + \eta_3 = 1. \tag{2.9}$$

Then according to the compatibility condition we will have $D\underline{\eta} = 0$ if matrices A and B are compatible. However the D- matrix in this case will be

$$\begin{bmatrix} b_{11}(a_{11}-1) & a_{11}b_{21} & a_{11}b_{31} \\ b_{12}(\alpha_{12}-1) & \alpha_{12}b_{22} & \alpha_{12}b_{32} \\ b_{13}(a_{13}-1) & a_{13}b_{23} & a_{13}b_{33} \\ a_{21}b_{11} & b_{21}(a_{21}-1) & a_{21}b_{31} \\ \alpha_{22}b_{12} & b_{22}(\alpha_{22}-1) & \alpha_{22}b_{32} \\ a_{23}b_{13} & b_{23}(a_{23}-1) & a_{23}b_{33} \\ a_{31}b_{11} & a_{31}b_{21} & b_{31}(a_{31}-1) \\ \alpha_{32}b_{12} & \alpha_{32}b_{22} & b_{32}(\alpha_{32}-1) \\ a_{33}b_{13} & a_{33}b_{23} & b_{33}(a_{33}-1) \end{bmatrix}.$$

So the set of linear equations to which must be solved find the unknown $\eta_i's$ as well as the unknown $\alpha_{ij}'s$ will be (from the above D-matrix)

$$b_{11}(a_{11}-1)\eta_1 + a_{11}b_{21}\eta_2 + a_{11}b_{31}\eta_3 = 0 \tag{2.10}$$

$$b_{13}(a_{13}-1)\eta_1 + a_{13}b_{23}\eta_2 + a_{13}b_{33}\eta_3 = 0 \tag{2.11}$$

$$b_{11}a_{21}\eta_1 + (a_{21}-1)b_{21}\eta_2 + a_{21}b_{31}\eta_3 = 0 \tag{2.12}$$

$$b_{12}(\alpha_{12}-1)\eta_1 + \alpha_{12}b_{22}\eta_2 + \alpha_{12}b_{32}\eta_3 = 0 \tag{2.13}$$

$$b_{12}\alpha_{22}\eta_1 + (\alpha_{22}-1)b_{22}\eta_2 + \alpha_{22}b_{32}\eta_3 = 0 \tag{2.14}$$

$$b_{12}\alpha_{32}\eta_1 + \alpha_{32}b_{22}\eta_2 + (\alpha_{32}-1)b_{32}\eta_3 = 0. \tag{2.15}$$

Solving the above set of equations with those constraints in (2.8-2.9), we get the following expressions for the unknowns:

$$\eta_1 = d_{22} \tag{2.16}$$

$$\eta_2 = \frac{d_{11}}{d_{12}} = d_{13} \tag{2.17}$$

$$\eta_3 = 1 - d_{22} - d_{13} \tag{2.18}$$

$$\alpha_{12} = \frac{b_{12}d_{22}}{b_{12}d_{22} + b_{22}d_{13} + b_{32}d_{23}} \tag{2.19}$$

$$\alpha_{22} = \frac{b_{12}d_{13}}{b_{12}d_{22} + b_{12}d_{13} + b_{32}d_{23}} \tag{2.20}$$

$$\alpha_{32} = \frac{b_{32}d_{23}}{b_{12}d_{22} + b_{22}d_{13} + b_{32}d_{23}}, \tag{2.21}$$

where

$$d_{11} = a_{11}b_{31}[a_{13}b_{33} - b_{13}(a_{13} - 1)] + a_{13}b_{33}[b_{11}(a_{11} - 1) - a_{11}b_{31}], \tag{2.22}$$

and

$$\begin{aligned} d_{12} = {} & [(a_{11}b_{21} - a_{11}b_{31})(b_{13}(a_{13} - 1) - a_{13}b_{33})] \\ & -[(a_{13}b_{23} - a_{13}b_{33})(b_{11}(a_{11} - 1) - a_{11}b_{31})] \end{aligned} \tag{2.23}$$

Finally

$$d_{22} = \frac{d_{13}(a_{11}b_{31} - a_{11}b_{21}) - a_{11}b_{31}}{b_{11}(a_{11} - 1) - a_{11}b_{31}}. \tag{2.24}$$

In particular let us consider a situation where we have the following choices for the two matrices A and B given by

$$A = \begin{pmatrix} 1/6 & \alpha_{12} & 1/4 \\ 1/3 & \alpha_{22} & 7/16 \\ 3/6 & \alpha_{32} & 5/16 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 1/7 & 2/7 & 4/7 \\ 2/5 & 2/5 & 1/5 \\ 1/4 & 1/4 & 2/4 \end{pmatrix}.$$

Using the set of equations in (2.22-2.24), we get the following: $d_{11} = -0.002976190$, $d_{12} = -0.0142857$, $d_{13} = \frac{d_{11}}{d_{12}} = 0.208334$, $d_{22} = 0.2916667$. So the unknown elements for the matrix A will be (from (2.19-2.21))

$$\alpha_{12} = 0.2857165, \alpha_{22} = 0.222248, \alpha_{32} = 0.4920587.$$

Here also note that these are the unique choices for which the two given matrices A and B are compatible.

2.5.2 Compatibility when some elements both in A and B are unknown

Suppose that we have a situation here where both in A and in B some of the elements are unknown and we denote the unknown elements of the matrix A by α_{ij} and unknown elements of the matrix B by β_{ij}. First we consider the situation when $I = 2$ and $J = 3$ with

$$\begin{bmatrix} a_{11} & \alpha_{12} & a_{11} \\ a_{21} & \alpha_{22} & a_{23} \end{bmatrix},$$

and

$$B = \begin{pmatrix} b_{11} & \beta_{12} & \beta_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix}.$$

So that in this case we have same constraint on the unknown elements α_{ij} and for the β_{ij} we have the following restriction:

$$b_{11} + \beta_{12} + \beta_{13} = 1 \tag{2.25}$$

$$\alpha_{12} + \alpha_{22} = 1 \tag{2.26}$$

$$\eta_1 + \eta_2 = 1. \tag{2.27}$$

However in this case we will have the following set of equations (

those involving the unknowns and excluding the redundant equations)

$$b_{11}(a_{11}-1)\eta_1 + a_{11}b_{21}\eta_2 = 0 \tag{2.28}$$

$$\beta_{12}(\alpha_{12}-1)\eta_1 + \alpha_{12}b_{22}\eta_2 = 0 \tag{2.29}$$

$$b_{22}(\alpha_{22}-1)\eta_2 + \alpha_{22}\beta_{12}\eta_1 = 0 \tag{2.30}$$

$$\beta_{13}(a_{13}-1)\eta_1 + a_{13}b_{23}\eta_2 = 0. \tag{2.31}$$

Using the constraint in equation (2.25), we get from equation (2.31),

$$\beta_{13} = \frac{(1-\eta_1)a_{13}b_{23}}{(1-a_{13})\eta_1}, \tag{2.32}$$

and from equation (2.28), using the constraint in (2.27), we get

$$\eta_1 = \frac{a_{11}b_{21}}{a_{11}b_{21} + b_{11}(1-a_{11})}.$$

Substituting the above expression of η_1 in equation (2.31), we get (after little algebra),

$$\beta_{13} = \frac{b_{11}b_{23}a_{21}a_{13}}{a_{11}a_{23}b_{21}}. \tag{2.33}$$

Hence the value of β_{12} will be

$$\beta_{12} = 1 - b_{11} - \frac{b_{11}b_{23}a_{21}a_{13}}{a_{11}a_{23}b_{21}}. \tag{2.34}$$

Substituting this in equation (2.29), we get

$$
\begin{aligned}
\alpha_{12} &= \frac{b_{21}a_{11}\beta_{12}}{b_{21}a_{11}b_{22}+b_{21}a_{11}\beta_{12}} \\
&= \frac{(a_{11}a_{23}b_{21}-b_{11}b_{23}a_{21}a_{13})a_{11}b_{21}}{(a_{11}a_{23}b_{21}-b_{11}b_{23}a_{21}a_{13})a_{11}b_{21}+b_{11}b_{22}a_{21}a_{11}a_{23}b_{21}}. \quad (2.35)
\end{aligned}
$$

Because of the constraint we can find the unknown value of α_{22}, which will be

$$\alpha_{22} = 1 - \alpha_{12}.$$

Let us consider a specific example. Consider matrices A and B which are of the form:

$$A = \begin{pmatrix} 1/5 & \alpha_{12} & 1/2 \\ 4/5 & \alpha_{22} & 1/2 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 1/6 & \beta_{12} & \beta_{13} \\ 2/5 & 2/5 & 1/5 \end{pmatrix}.$$

Here if we are given the information that A and B are compatible, then the choices of the unknown values of $\alpha'_{ij}s$ and $\beta'_{ij}s$ will be given by

$$\beta_{13} = \frac{b_{11}b_{23}a_{21}a_{13}}{a_{11}a_{23}b_{21}} = \frac{1}{3}.$$

So that

$$\beta_{12} = 1 - \frac{1}{6} - \frac{1}{3} = \frac{1}{2}.$$

Again

$$\eta_1 = \frac{\frac{1}{5}\frac{2}{5}}{\frac{1}{5}\frac{2}{5} + \frac{1}{6}\frac{4}{5}} = \frac{3}{8}.$$

So that

$$\begin{aligned}\alpha_{12} &= \frac{\beta_{12}\eta_1}{\beta_{12}\eta_1 + b_{22}(1-\eta_1)} \\ &= \frac{3}{7}.\end{aligned}$$

Hence

$$\alpha_{22} = 1 - \alpha_{12} = \frac{4}{7}.$$

Furthermore note that in this case also these are the unique choices for the unknown values for which matrices A and B are compatible.

2.6 Compatibility in the general Case

Here we will discuss the problem of compatibility when the dimension of the two matrices A and B is of the order $(I \times J)$. However we will consider here the situation where there are two elements unknown only in A. While for the matrix B all the elements are known. Let us consider that in matrix A in the l_1^{-th} column $(1 \leq l_1 \leq J)$, any two elements are unknown and that they appear in i_1^{-th} and i_2^{-th} rows (without loss of generality the unknown elements can be assumed to be in positions (1,1) and (2,1) by relabelling), where $(1 \leq (i_1, i_2) \leq I)$. So that (since column sums of A add up to 1), we can write, considering the unknown elements to be denoted by α'_{ij}s

$$\alpha_{i_1 l_1} + \alpha_{i_2 l_1} + \sum_{k \neq i_1, i_2} a_{kj} = 1, \forall (j, l_1) = 1(1)J. \tag{2.36}$$

while for B all the row sums add up to 1, and all the elements are known

$$\sum_{j=1}^{J} b_{ij} = 1, \forall i = 1, 2, \cdots, I.$$

Now if we have the information that the matrices A and B are compatible, then from the equation (for any i_1^{-th} row and l_1^{-th} column), we can write

$$\alpha_{i_1 l_1} [\sum_{s=1}^{I} b_{s l_1} \eta_s] - b_{i_1 l_1} \eta_{i_1} = 0 \tag{2.37}$$

$$\Rightarrow \alpha_{i_1 l_1} [b_{i_1 l_1} \eta_{i_1} + \sum_{\substack{s=1 \\ s \neq i_1}} b_{s l_1} \eta_s] - b_{i_1 l_1} \eta_{i_1} = 0. \tag{2.38}$$

Hence the unknown value of $\alpha_{i_1 l_1}$ will be given by

$$\alpha_{i_1 l_1} = \frac{b_{i_1 l_1} \eta_{i_1}}{b_{i_1 l_1} \eta_{i_1} + \sum_{\substack{s=1 \\ s \neq i_1}} b_{s l_1} \eta_s}.$$

Again because of (2.43), we can write

$$\alpha_{i_2 l_1} = 1 - \sum_{k \neq i_1, i_2} a_{kj} - \alpha_{i_1 l_1} = 1 - \sum_{k \neq i_1, i_2} a_{kj} - \frac{b_{i_1 l_1} \eta_{i_1}}{b_{i_1 l_1} \eta_{i_1} + \sum_{\substack{s=1 \\ s \neq i_1}} b_{s l_1} \eta_s}.$$

However the solution for the η_i' s can be obtained by using any set of $(I-1)$ equations since $\sum_{i=1}^{I} \eta_i = 1$.

2.7 Concluding remarks

Our search for a compatible P in terms of equations subject to inequality constraints is based on the fact that we really need to find one compatible marginal, say that corresponding to the random variable X, and we consider the fact that when this is combined with B it will give us P. Compatible conditional and marginal specifications of distributions are of fundamental importance in modeling scenarios. Moreover in Bayesian prior elicitation contexts, inconsistent conditional specifications are to be expected. In such situations interest will center on most nearly compatible distributions which we have discussed in the previous chapter. In the finite discrete case a variety of compatibility conditions can be derived. In this chapter we have discussed in detail the problem of compatibility in context as mentioned earlier by identifying it as a programming problem and have developed a rank based criterion. We have shown that the rank of the matrix (whose elements are constructed from the two given matrices A and B) under compatibility will be $I-1$, for a (2×2) and (3×3) case, along with the proof for any dimension $(I\times J)$. A significant amount of the material of this chapter draws heavily on Arnold, Castillo and Sarabia (1999) and Arnold and Gokhale (1998). Also we have provided a small amount of discussion on the topic of compatibility when we have missing elements in either A or in B or in both. It has been observed that for a given A and B, under compatibility, the choices for the missing elements are unique. However for a general case where the dimension of the matrix D is $(IJ\times I)$, the strategy that we have mentioned in this chapter will be quite challenging in identifying the solution for the unknown elements either in any of the conditional probability matrices A and B or in both of them. Also when we have elements missing in A and B in different positions then also our procedure will result in solving a set of IJ number of equations which is cumbersome and quite difficult to handle. In such a situation one is suggested to consider the concept of compatibility under rank one criterion as proposed by Arnold, Castillo and Sarabia (2001). One interesting question that may arise here is how can we extend the above technique under compatibility when there exists more than two conditional matrices in the discrete case i.e; if we are given three arrays A and B and C where

- A is the conditional probability matrix of (say) X given Y and Z.
- B is the conditional probability matrix of (say) Y given X and Z.
- C is the conditional probability matrix of (say) X given X and Y.

Furthermore what would happen in this situation (under compatibility) when some of the elements are unknown in either any of A or B or C or in all of them.

Chapter 3

Classical and Bayesian inference for a hidden truncated bivariate $P(II)$ distribution

3.1 Introduction

The use of the Pareto distribution as a model for various socio-economic phenomena dates back to the late nineteenth century. Pareto's distributions and their close relatives and generalizations provide a very flexible family of fat-tailed distributions which may be used to model income distributions as well as a wide variety of other social and economic distributions. An extensive historical survey of the use of these models in the context of income distributions may be found in, for example, Arnold (1983). Authors such as Pareto (1897) and Zipf (1949) have asserted law-like adherence to a Pareto model. In addition, a variety of data sets, support the assertion that, certainly in the upper tail, a Pareto distribution provides a satisfactory model. A specific example is the distribution of annual per capita income of adult males in the USA.

It is quite remarkable that the development of inference procedures under the classical approach for Pareto distributions and their close relatives has been quite limited. A lot of emphasis has been given to the classical Pareto distribution because of its simplicity and analytical tractability. However as one progresses to study more complex Pareto distributions all those available nice distributional properties disappear. Some authors have assumed one or more than one parameter to be known in order to make some progress. If all the parameters are assumed unknown,then such well known techniques as maximum-likelihood and the method of moments will involve iterative solution of systems of distinctly non-linear equations. At this point it is quite important for us to introduce families of distribution which qualify as univariate and bivariate Pareto (II) distribution and which from now onwards will be denoted by $P(II)(\mu, \sigma, \alpha)$ and $P^2(II)(\underline{\mu}, \underline{\sigma}, \underline{\alpha})$ respectively whenever they appear in our discussion in this chapter. Following Arnold (1983), we write $X \sim P(II)(\mu, \sigma, \alpha)$, if it has the following survival function:

$$P(X > x) = [1 + (\frac{x-\mu}{\sigma})]^{-\alpha}, x \geq \mu,$$

where μ, the location parameter, is real, and σ, the scale parameter is positive. Next we write $(X, Y) \sim P^2(II)(\underline{\mu}, \underline{\sigma}, \underline{\alpha})$, if it has the following joint survival function:

$$P(X > x, Y > y) = [1 + (\frac{x-\mu_1}{\sigma_2}) + (\frac{y-\mu_2}{\sigma_2})]^{-\alpha},$$

where μ_1, μ_2 and σ_1, σ_2 are the location and scale parameters for X and Y respectively and α, is the index of inequality.

Since $X_{1:n}$ is a consistent estimate of μ in the $P(II)$ family, one may consider setting $\mu = X_{1:n}$ and solving the resulting simplified likelihood equation to get approximate likelihood estimates. For the method of moment estimation we will consider the approach of Arnold and Laguna (1977) of using sample fractional moments for our proposed model described later on. In the quartile estimation procedure we will follow the approach suggested by Quandt (1966).

Inference procedures for Pareto populations under the Bayesian paradigm are not well developed either. In this context one may consider earlier works by authors such as Arnold and Press (1982, 1989) where they discussed selection of proper prior distributions which do not lead to an anomalous posterior distribution for classical Pareto data. However not many results are available for Pareto distributions other than the classical one in the Bayesian framework.

3.2 Why the hidden truncation $P(II)$ model?

Among the family of Pareto distributions, the $P(II)$ model needs further investigation simply because of the fact that within our hierarchy of generalized Pareto distributions, it is the $P(II)$ family which can be viewed as the log-logistic family of distributions and thus is a viable competitor of log-normal distribution, which is a popular model for the distribution of income, wealth etc. Now we envisage a situation where we assume that an individual's actual income is divided into two parts, one is the observable reported income and the other one is the unreported income. We consider the case in which both the reported and unreported income have a Paretian distribution and in particular they can well be explained by a $P(II)$ model. Then one might be interested to know what is the structure of average reported income for those individuals who have an unreported income not exceeding a certain level. Models of such types can be explained by a hidden truncation paradigm where we observe one variable only when it is subject to hidden truncation from above with respect to one or more covariables.

In this chapter we focus our attention on estimation of the parameters of a bivariate $P(II)$ distribution when hidden truncation is applied to one of the variables with the only restriction that the truncation point will be greater than the location parameter of the truncated variable. We do not consider hidden truncation from below because in that situation our resulting density will be again a member in the same family of distribution with only reparametrization of the parent model. Specifically here we consider the hidden truncation paradigm when one variable is subject to hidden truncation from above for a bivariate $P(II)$ distribution and inference procedures for such models.

In this chapter we consider both classical and Bayesian methods of estimation.

The remainder of this chapter is organized in the following way. In Section 2, we will discuss the usefulness of the Hidden Truncated $P(II)$ model. In Section 3, we will consider the concept of hidden truncation. In Section 4, we will consider the hidden truncated (from above) density for the bivariate $P(II)$ model. In Section 5, we will consider the method of moment estimation using fractional moments. In Section 6, we will consider the maximum likelihood estimation of the parameters. In Section 7, we will consider the estimation of the parameters using sample quartiles. In Section 8, we will consider an application of our model to a real life data set. In Section 9, we will consider the estimation of all the parameters by all the above mentioned estimation procedures using a simulation study. In Section 10, we will consider the estimation of all the parameters under the Bayesian paradigm both with the choice of independent and dependent priors (except for the truncation parameter) for all the parameters. In Section 11, we will consider the likelihood ratio test for the truncation parameter and will also report the large sample distribution of the likelihood ratio test statistic. In Section 12, we will consider the asymptotic distribution of the smallest order statistic when the samples are drawn from a hidden truncated bivariate $P(II)$ model. The properties of the maximum likelihood estimators are discussed in detail in the appendix.

3.3 Hidden truncation

We consider a two dimensional absolutely continuous random vector (X, Y). We might focus on the conditional distribution of X given $Y \in M$ where M is a Borel set in $\mathbb{R}$. Indeed we could write

$$f_{X|Y \in M}(x) = f_X(x) \frac{P(Y \in M | X = x)}{P(Y \in M)}. \tag{3.1}$$

However, we will concentrate on hidden truncation of one of the following three forms only

(i) Lower truncation,where $M = (c, \infty)$.

(ii) Upper truncation,where $M = (-\infty, c)$.

(iii) Two sided truncation,where $M = (a, b]$.

For upper truncation (equivalently, truncation from above) at c, in which observations are only available for X's whose concomitant variable Y is less than c, equation (3.1) reduces to

$$f_{c-}(x) = f_X(x)\frac{P(Y \leq c|X = x)}{P(Y \leq c)}. \tag{3.2}$$

Models of this type are thus characterized by

(i) $f_X(x)$, the density assumed for X.

(ii) The conditional density of Y given X, $f_{Y|X}(y|x)$.

(iii) The specified value c.

However models of this type also may depend on other parameters in addition to c. First of all we consider the $P(II)$ case in which both the location parameters of X and Y are zero while both the scale parameters are equal to one. In that case the joint survival function of X and Y is given by

$$P(X > x, Y > y) = (1 + x + y)^{-\alpha}, x \geq 0, y \geq 0. \tag{3.3}$$

The corresponding distribution function is given by $F(x, y) = 1-(1+x)^{-\alpha}-(1+y)^{-\alpha}+(1+x+y)^{-\alpha}, x \geq 0, y \geq 0$. Therefore the corresponding joint density is given by $f(x,y) = \alpha(\alpha+1)(1+x+y)^{-(\alpha+2)}I(x \geq 0, y \geq 0)$.

The conditional density of Y for each fixed $X = x$ will be

$$
\begin{aligned}
f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\
&= \frac{(\alpha+1)}{(1+x)}(1+\frac{y}{(1+x)})^{-(\alpha+2)}I(y \geq 0). \qquad (3.4)
\end{aligned}
$$

Hence in this case we can write

$$Y|X = x \sim P(II)(0,(1+x),(\alpha+1)).$$

Next we define a set function defined for each $\delta \in \mathbb{R}^+$ which is as follows

$$\gamma_\delta(B) = \frac{\alpha}{\delta}\int_B (1+\frac{y}{\delta})^{-(\alpha+1)}dy, \qquad (3.5)$$

where B is any Borel set. Since $x \in \mathbb{R}^+$, so also $1+x \in \mathbb{R}^+$, so that we may define

$$\gamma_{1+x}(B) = \frac{\alpha}{\delta}\int_B (1+\frac{y}{1+x})^{-(\alpha+1)}dy. \qquad (3.6)$$

So that the hidden truncation model that we have earlier can be written as

$$f_{Y|X \in B}(x) = f_X(x) \frac{\gamma_{1+x}(B)}{\gamma_1(B)}. \tag{3.7}$$

3.4 Hidden truncated density for the bivariate Pareto model

We first consider a bivariate $P(II)$ distribution where both the marginals and also both the conditionals are members of a $P(II)$ family. The joint survival function of such a bivariate density is given by

$$P(X > x, Y > y) = [1 + (\frac{x - \mu_1}{\sigma_1}) + (\frac{y - \mu_2}{\sigma_2})]^{-\alpha}, x \geq \mu_1, y \geq \mu_2, \tag{3.8}$$

where μ_1, σ_1, μ_2, σ_2 are the location and scale parameters for X and Y respectively and α is the index of inequality. In this case we write

$$(X, Y) \sim P^2(II)(\underline{\mu}, \underline{\sigma}, \alpha).$$

Note that the correlation between X and Y is positive provided $\alpha > 2$, specifically

$$\text{corr}(X, Y) = \frac{1}{(\alpha - 1)(\alpha - 2)}.$$

The joint density of (X, Y) is given by

$$
\begin{aligned}
&f_{X,Y}(x,y)\\
&= \frac{\partial}{\partial x}\frac{\partial}{\partial y}[P(X > x, Y > y)]\\
&= \frac{\alpha(\alpha+1)}{\sigma_1\sigma_2}[1 + (\frac{x-\mu_1}{\sigma_1}) + (\frac{y-\mu_2}{\sigma_2})]^{-(\alpha+2)} I(x \geq \mu_1, y \geq \mu_2). \quad (3.9)
\end{aligned}
$$

So the conditional density of Y for each fixed $X = x$ will be given by

$$
\begin{aligned}
&f_{Y|X}(y|x)\\
&= \frac{f_{X,Y}(x,y)}{f_X(x)}\\
&= \frac{(\alpha+1)}{\sigma_2(1 + (\frac{x-\mu_1}{\sigma_1}))}[1 + \frac{(\frac{y-\mu_2}{\sigma_2})}{1 + (\frac{x-\mu_1}{\sigma_1})}]^{-(\alpha+2)} I(y \geq \mu_2). \quad (3.10)
\end{aligned}
$$

Let us first consider the situation in which X is observed only if Y is greater than some positive value b. In that case the hidden truncated density of X given that $Y \geq b$ with the condition that $b > \mu_2$ is given by

$$
f^{HT}_{X|Y\geq b}(x) = f_X(x)\frac{P(Y \geq b|X = x)}{P(Y \geq b)} I(x \geq b). \quad (3.11)
$$

Now in our case

$$P(Y \geq b|X = x) = \int_b^\infty f_{Y|X}(y|x)dy = \frac{1}{(1 + \frac{(\frac{b-\mu_2}{\sigma_2})}{1+(\frac{x-\mu_1}{\sigma_1})})^{\alpha+1}},$$

while

$$P(Y \geq b) = (1 + \frac{b - \mu_2}{\sigma_2})^{-\alpha}.$$

Hence on substitution in equation (3.11) we get the following

$$f^{HT}_{X|Y \geq b}(x) = \frac{\alpha}{\sigma_1(1 + (\frac{b-\mu_2}{\sigma_2}))}(1 + \frac{\frac{x-\mu_1}{\sigma_1}}{1 + (\frac{b-\mu_2}{\sigma_2})})^{-(\alpha+1)} I(x \geq \mu_1). \quad (3.12)$$

From above it is quite clear that the conditional distribution of X given $Y \geq b$ is again a member of a $P(II)$ family i.e., $X|Y \geq b \sim P(II)(\mu_1 + 2\frac{\sigma_1(1+\frac{b-\mu_2}{\sigma_2})}{\alpha-1}, \sigma_1^* = \sigma_1(1 + (\frac{b-\mu_2}{\sigma_2})), \alpha)$.

So it is obvious that with lower truncation there is no augmentation in the model.

In contrast let us consider the situation in which X is observed only if Y is less than some positive value c. Motivation for considering such a model may arise from any one of the following situations:

(1) Suppose that we have a bivariate population (X, Y) where X and Y represents the income of an individual derived from salary and other sources respectively. In such a situation if we have reasons to believe that the two components X and Y are

positively correlated and represent the income components of the individual then we may consider a bivariate Pareto distribution in such a situation. Then one might be interested to know what is the average salary structure for those individuals who have income from other sources (for example agriculture, housing income etc.) not exceeding a certain amount to be specified by the investigator and the salary data is available to us if and only if it is truncated from above by one (or more than one) of the covariables.

(2) In a departmental store, the store manager would like to know what is the average sales of a newly launched beauty product for those customers whose expenditure on food products does not exceed a certain amount to be specified by him and our data is available in that format as mentioned in earlier example.

In that case the truncated density of X given that $Y \leq c$ with the condition that $c > \mu_2$ is given by:

$$
\begin{aligned}
& f^{HT}_{X|Y\leq c}(x) \\
&= \tfrac{\alpha}{\sigma_1(1-(1+(\frac{c-\mu_2}{\sigma_2}))^{-\alpha})}[(1+(\tfrac{x-\mu_1}{\sigma_1}))^{-(\alpha+1)} \\
&- (1+(\frac{(x-\mu_1)}{\sigma_1})+(\frac{c-\mu_2}{\sigma_2}))^{-(\alpha+1)}]I(x \geq \mu_1). \qquad (3.13)
\end{aligned}
$$

We usually estimate μ_1 by $X_{1:n}$, where $X_{1:n} = \min_{1\le i\le n} X_i$, and so we can subtract and assume that $\mu_1 = 0$. From here after whenever it appears we will follow the above convention about μ_1. Also for notational simplicity let us consider

$$\psi(\alpha,\theta) = 1 - (1 + (\frac{c-\mu_2}{\sigma_2}))^{-\alpha} = 1 - (1+\theta)^{-\alpha},$$

where $\theta = \frac{c-\mu_2}{\sigma_2}$, and $\theta > 0$, then our density reduces to

$$f^{HT}_{X|Y\le\theta}(x) = \frac{\alpha}{\sigma_1\psi(\alpha,\theta)}[(1+\frac{x}{\sigma_1})^{-(\alpha+1)} - (1+\frac{x}{\sigma_1}+\theta)^{-(\alpha+1)}]I(x\ge 0). \tag{3.14}$$

Let us consider the hazard rate function for our density. The hazard rate function is given by

$$h_{T|Y\le\theta}(t) = \frac{f^{HT}_{T|Y\le\theta}(t)}{S_{T|Y\le\theta}(t)},$$

where

$$\begin{aligned} S_{T|Y\le\theta}(t) &= P_{T|Y\le\theta}(T>t) \\ &= \int_t^\infty \frac{\alpha}{\sigma_1\psi(\alpha,\theta)}[(1+\frac{u}{\sigma_1})^{-(\alpha+1)} - (1+\frac{u}{\sigma_1}+\theta)^{-(\alpha+1)}]du \\ &= \frac{1}{\psi(\alpha,\theta)}[(1+\frac{t}{\sigma_1})^{-\alpha} - (1+\frac{t}{\sigma_1}+\theta)^{-\alpha}]. \qquad (3.15) \end{aligned}$$

So the hazard rate function in our cases given by

$$h_{T|Y\leq\theta}(t) = \frac{\frac{\alpha}{\sigma_1\psi(\alpha,\theta)}[(1+\frac{t}{\sigma_1})^{-(\alpha+1)} - (1+\frac{t}{\sigma_1}+\theta)^{-(\alpha+1)}]}{\frac{1}{\psi(\alpha,\theta)}[(1+\frac{t}{\sigma_1})^{-\alpha} - (1+\frac{t}{\sigma_1}+\theta)^{-\alpha}]}$$

$$= \frac{\alpha[(1+\frac{t}{\sigma_1})^{-(\alpha+1)} - (1+\frac{t}{\sigma_1}+\theta)^{-(\alpha+1)}]}{\sigma_1[(1+\frac{t}{\sigma_1})^{-\alpha} - (1+\frac{t}{\sigma_1}+\theta)^{-\alpha}]}. \tag{3.16}$$

Next we consider for graphical reference for different choices of the truncation parameter (θ) but with the same choices for the other parameters (i.e., with shape=4, scale=1) the plot of the hazard rate function of a hidden truncated $P(II)$ density.

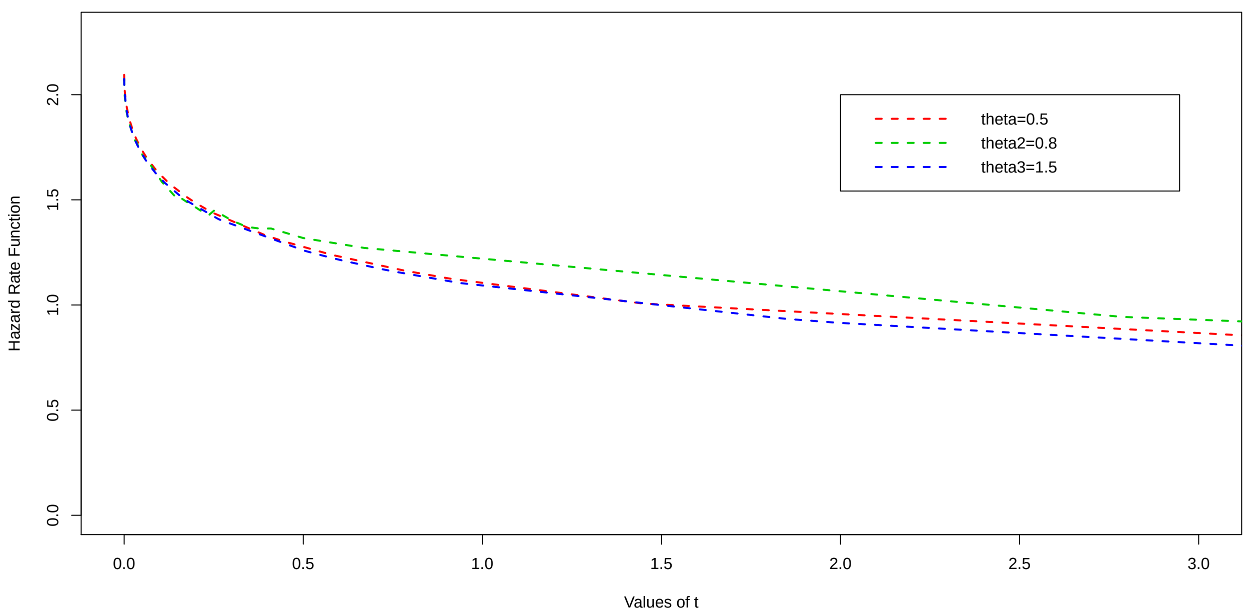


Figure 3.1: Hazard Rate Function Plot for different choices of the truncation parameter (θ).

3.5 Estimation of the Parameters using Fractional Moments

First of all we will consider for any real $r(\geq 1)$,

$$\begin{aligned}
E[(\frac{X}{\sigma_1})^r] &= \frac{\alpha}{\psi(\alpha,\theta)}\int_0^{\infty}(\frac{x}{\sigma_1})^r f^{HT}(x)dx \\
&= \frac{\alpha}{\psi(\alpha,\theta)\sigma_1}\int_0^{\infty}(\frac{x}{\sigma_1})^r[(1+(\frac{x}{\sigma_1}))^{-(\alpha+1)}-(1+\theta+(\frac{x}{\sigma_1}))^{-(\alpha+1)}]dx \\
&= I_1 - I_2, \qquad (3.17)
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \frac{\alpha}{\psi(\alpha,\theta)\sigma_1}\int_0^{\infty}(\frac{x}{\sigma_1})^r[(1+(\frac{x}{\sigma_1}))^{-(\alpha+1)}]dx \\
&= \frac{\alpha}{\psi(\alpha,\theta)}\int_0^{\infty}(u)^r(1+u)^{-(\alpha+1)}du \\
&= \frac{\alpha}{\psi(\alpha,\theta)}B(r+1,\alpha-r),
\end{aligned}$$

valid for $\alpha > r$ where $u = 1 + \frac{x}{\sigma_1}$.

Similarly

$$\begin{aligned}
I_2 &= \frac{\alpha}{\psi(\alpha,\theta)\sigma_1}\int_0^\infty (\frac{x}{\sigma_1})^r[(1+\theta+\frac{x}{\sigma_1})^{-(\alpha+1)}]dx \\
&= \frac{\alpha}{\psi(\alpha,\theta)}\int_0^\infty [u^r(1+u+\theta)^{-(\alpha+1)}]du \\
&= \frac{\alpha}{\psi(\alpha,\theta)}(1+\theta)^{r-\alpha}\int_0^\infty [w^r(1+w)^{-(\alpha+1)}]dw \\
&= \frac{\alpha}{\psi(\alpha,\theta)}(1+\theta)^{r-\alpha}B(r+1,\alpha-r),
\end{aligned}$$

valid for$\alpha > r$, where $u = 1 + \frac{x}{\sigma_1}$, and $w = \frac{u}{1+\theta}$.

So that we have for any $(r \geq 1)$,

$$E[(\frac{X}{\sigma_1})^r] = \frac{\alpha}{\psi(\alpha,\theta)}B(r+1,\alpha-r)[1-(1+\theta)^{r-\alpha}]. \tag{3.18}$$

Now substituting $r = \frac{1}{2}, \frac{1}{3}, 1$ we get the following

$$E[(\frac{X}{\sigma_1})^{\frac{1}{2}}] = \frac{\alpha}{\psi(\alpha,\theta)}B(\frac{1}{2}+1,\alpha-\frac{1}{2})[1-(1+\theta)^{\frac{1}{2}-\alpha}]. \tag{3.19}$$

However

$$\begin{aligned}
B(\frac{1}{2}+1,\alpha-\frac{1}{2}) &= \frac{\Gamma\left(\frac{3}{2}\right)\Gamma\left(\alpha-\frac{1}{2}\right)}{\Gamma\left(\alpha+1\right)} \\
&= \frac{\frac{1}{2}\sqrt{\pi}\Gamma\left(\alpha-\frac{1}{2}\right)}{\Gamma\left(\alpha+1\right)}.
\end{aligned}$$

Hence

$$E[X^{\frac{1}{2}}] = \frac{\alpha\sqrt{\pi\sigma_1}}{2\psi(\alpha,\theta)} \frac{\Gamma\left(\alpha - \frac{1}{2}\right)}{\Gamma\left(\alpha+1\right)}[1-(1+\theta)^{\frac{1}{2}-\alpha}]. \tag{3.20}$$

In the same way we also have

$$E[X^{\frac{1}{3}}] = \frac{\alpha(\sigma_1)^{\frac{1}{3}}\Gamma\left(\alpha - \frac{1}{3}\right)\Gamma\left(\frac{4}{3}\right)}{\psi(\alpha,\theta)\Gamma\left(\alpha+1\right)}[1-(1+\theta)^{\frac{1}{3}-\alpha}]. \tag{3.21}$$

and

$$E(X) = \frac{\sigma_1}{(\alpha-1)\psi(\alpha,\theta)}[1-(1+\theta)^{1-\alpha}]. \tag{3.22}$$

Next for the method of moment estimation we first define the following quantities:

- $M_1 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X_{1:n})$.
- $M_{\frac{1}{2}} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X_{1:n})^{\frac{1}{2}}$.
- $M_{\frac{1}{3}} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X_{1:n})^{\frac{1}{3}}$.

Next we consider the following (after equating the sample moments with the corresponding population moments)

$$\text{(say)}, M_{a_1} = \frac{M_{\frac{1}{3}}}{M_{\frac{1}{2}}} = \frac{2\Gamma\left(\frac{4}{3}\right)\Gamma\left(\alpha-\frac{1}{3}\right)}{\sqrt{\pi}(\sigma_1)^{\frac{1}{6}}\Gamma\left(\alpha-\frac{1}{2}\right)}\left[\frac{(1+\theta)^{\alpha}-(1+\theta)^{\frac{1}{3}}}{(1+\theta)^{\alpha}-(1+\theta)^{\frac{1}{2}}}\right]. \tag{3.23}$$

Also

$$\text{(say)}, M_{a_2} = \frac{M_{\frac{1}{3}}}{M_1} = \frac{\Gamma\left(\frac{4}{3}\right)\Gamma\left(\alpha-\frac{1}{3}\right)}{(\sigma_1)^{\frac{2}{3}}\Gamma(\alpha)}(\alpha-1)\left[\frac{(1+\theta)^{\alpha}-(1+\theta)^{\frac{1}{3}}}{(1+\theta)^{\alpha}-(1+\theta)^{\frac{1}{2}}}\right]. \tag{3.24}$$

and

$$\text{(say)}, M_{a_3} = \frac{M_{\frac{1}{2}}}{M_1} = \frac{\sqrt{\pi}\Gamma\left(\alpha-\frac{1}{2}\right)(\alpha-1)}{2\Gamma(\alpha)\sqrt{\sigma_1}}\left[\frac{(1+\theta)^{\alpha}-(1+\theta)^{\frac{1}{2}}}{(1+\theta)^{\alpha}-(1+\theta)}\right]. \tag{3.25}$$

So in this case we have three equations which are as follows

$$\begin{aligned} &2M_{a_3}\Gamma(\alpha)\sqrt{\sigma_1}((1+\theta)^{\alpha}-(1+\theta)) \\ &-\sqrt{\pi}\Gamma\left(\alpha-\frac{1}{2}\right)(\alpha-1)((1+\theta)^{\alpha}-(1+\theta)^{\frac{1}{2}})=0. \end{aligned} \tag{3.26}$$

Also

$$\begin{aligned} &M_{a_1}[\sqrt{\pi}(\sigma_1)^{\frac{1}{6}}\Gamma\left(\alpha-\frac{1}{2}\right)((1+\theta)^{\alpha} \qquad -(1+\theta)^{\frac{1}{2}}) \\ &-2(\Gamma\left(\frac{4}{3}\right)\Gamma\left(\alpha-\frac{1}{3}\right))((1+\theta)^{\alpha}-(1+\theta)^{\frac{1}{3}})=0, \end{aligned} \tag{3.27}$$

and

$$M_{a_2}(\sigma_1)^{\frac{2}{3}}\Gamma\left(\alpha\right)\left((1+\theta)^{\alpha}-(1+\theta)\right)$$
$$-\,(\Gamma\left(\frac{4}{3}\right)\Gamma\left(\alpha-\frac{1}{3}\right))(\alpha-1)((1+\theta)^{\alpha}-(1+\theta)^{\frac{1}{3}})=0. \qquad (3.28)$$

Note that in this case the only assumption that we need is $\alpha > 1$.

3.6 Estimation of the parameters using maximum likelihood

In this case the likelihood function is given by

$$\begin{aligned} L(\alpha,\sigma_1,\theta) &= \prod_{i=1}^{n}[\frac{\alpha}{\sigma_1\psi(\alpha,\theta)}((1+\frac{X_i}{\sigma_1})^{-(\alpha+1)}-(1+\frac{X_i}{\sigma_1}+\theta)^{-(\alpha+1)})] \\ &= \frac{\alpha^n}{(\sigma_1\psi(\alpha,\theta))^n}\prod_{i=1}^{n}[(1+\frac{X_i}{\sigma_1})^{-(\alpha+1)}(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})] \qquad (3.29) \end{aligned}$$

Equivalently the log-likelihood function is given by

$$\log L(\alpha,\sigma_1,\theta) = n\log\alpha - n\log\sigma_1 - n\log\psi(\alpha,\theta)$$
$$-\,(\alpha+1)\sum_{i=1}^{n}\log(1+\frac{X_i}{\sigma_1})+\sum_{i=1}^{n}\log(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}). \quad (3.30)$$

The likelihood equation for all the parameters are obtained by differentiating the likelihood function. For α we have

$$\frac{\partial}{\partial \alpha}[\log L(\alpha, \sigma_1, \theta)] = 0,$$

$$\text{i.e.,} \frac{n}{\alpha} - n\frac{\psi'(\alpha,\theta)}{\psi(\alpha,\theta)} - \sum_{i=1}^{n}[\log(1+\frac{X_i}{\sigma_1}) - \frac{(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}}{(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})}\log(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})] = 0. \quad (3.31)$$

Here $\psi(\alpha, \theta) = 1-(1+\theta)^{-\alpha}$. So that $\frac{\partial}{\partial \alpha}[\psi(\alpha, \theta)] = (1+\theta)^{-\alpha}\log(1+\theta)$. Also we have

$$\frac{\partial}{\partial \sigma_1}[\log L(\alpha, \sigma_1, \theta)] = 0.$$

Equivalently we can write

$$\text{i.e.,} -\frac{n}{\sigma_1} + (\alpha+1)\sum_{i=1}^{n}[\frac{X_i}{(1+\frac{X_i}{\sigma_1})(\sigma_1)^2}] + (\alpha+1)\sum_{i=1}^{n}[\frac{(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+2)}}{(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})}(\frac{\theta(\frac{X_i}{\sigma_1})}{(\sigma_1+X_i)^2})] = 0. \quad (3.32)$$

Furthermore we have

$$\frac{\partial}{\partial \theta}[\log L(\alpha, \sigma_1, \theta)] = 0,$$

which is equivalent to

$$
\begin{aligned}
& -n\frac{\psi'(\alpha,\theta)}{\psi(\alpha,\theta)}+(\alpha+1)\sum_{i=1}^{n}[\frac{1}{(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})} \\
& (1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+2)}(1+\frac{X_i}{\sigma_1})^{-1}]=0. \qquad (3.33)
\end{aligned}
$$

Note that here $\psi(\alpha,\theta)=1-(1+\theta)^{-\alpha}$, so that $\frac{\partial}{\partial\theta}[\psi(\alpha,\theta)]=\alpha(1+\theta)^{-(\alpha+1)}$.

3.7 Estimation based on the Sample Quartiles

Here first of all we will consider the three quartiles (namely first, second and third) calculated from the truncated density. Let ξ_p be any p-th order quantile ($p \in (0,1)$), then

$$P[X \leq \xi_p]=p. \qquad (3.34)$$

So in our case we have

$$
\begin{aligned}
P[X \le \xi_p] &= \frac{\alpha}{\psi(\alpha,\theta)\sigma_1}\int_0^{\xi_p} f^{HT}(x)dx \\
&= \frac{\alpha}{\psi(\alpha,\theta)\sigma_1}\int_0^{\xi_p} [(1+(\frac{x}{\sigma_1}))^{-(\alpha+1)} - (1+\theta+(\frac{x}{\sigma_1}))^{-(\alpha+1)}]dx \\
&= I_3 - I_4,
\end{aligned}
$$

where

$$
\begin{aligned}
I_3 &= \frac{\alpha}{\psi(\alpha,\theta)\sigma_1}\int_0^{\xi_p} [(1+(\frac{x}{\sigma_1}))^{-(\alpha+1)}]dx \\
&= \frac{1}{\psi(\alpha,\theta)}[1-(1+\frac{\xi_p}{\sigma_1})^{-\alpha}],
\end{aligned}
$$

and

$$
\begin{aligned}
I_4 &= \frac{\alpha}{\psi(\alpha,\theta)\sigma_1}\int_0^{\xi_p} (1+\theta+(\frac{x}{\sigma_1}))^{-(\alpha+1)}dx \\
&= \frac{1}{\psi(\alpha,\theta)}[(1+\theta)^{-\alpha} - (1+\theta+\frac{\xi_p}{\sigma_1})^{-\alpha}].
\end{aligned}
$$

So that we have

$$
\frac{1}{\psi(\alpha,\theta)}[1-(1+\frac{\xi_p}{\sigma_1})^{-\alpha} - ((1+\theta)^{-\alpha} - (1+\theta+\frac{\xi_p}{\sigma_1})^{-\alpha})] = p. \quad (3.35)
$$

Equivalently we can write

$$1-(1+\frac{\xi_p}{\sigma_1})^{-\alpha}-(1+\theta)^{-\alpha}+(1+\theta+\frac{\xi_p}{\sigma_1})^{-\alpha}=p\psi(\alpha,\theta). \qquad (3.36)$$

Our estimates are then obtained by equating 3 sample quantiles , denoted by $\hat{\xi}_p$ to the corresponding population quantiles. In particular by considering successively $p=\frac{1}{4},\frac{1}{2},\frac{3}{4}$ we have the following three equations:

$$1-(1+\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{-\alpha}-(1+\theta)^{-\alpha}+(1+\theta+\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{-\alpha}=\frac{\psi(\alpha,\theta)}{4}. \qquad (3.37)$$

$$1-(1+\frac{\hat{\xi}_{\frac{1}{2}}}{\sigma_1})^{-\alpha}-(1+\theta)^{-\alpha}+(1+\theta+\frac{\hat{\xi}_{\frac{1}{2}}}{\sigma_1})^{-\alpha}=\frac{\psi(\alpha,\theta)}{2}, \qquad (3.38)$$

and

$$1-(1+\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{-\alpha}-(1+\theta)^{-\alpha}+(1+\theta+\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{-\alpha}=\frac{3\psi(\alpha,\theta)}{4}. \qquad (3.39)$$

So from equation (3.38) and equation (3.40) we get

$$(1+\theta+\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{-\alpha}+(1+\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{-\alpha}-(1+\theta+\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{-\alpha}-(1+\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{-\alpha}=\frac{\psi(\alpha,\theta)}{2}. \quad (3.40)$$

Again from equation (3.38) and equation (3.49) we get

$$(1+\theta+\frac{\hat{\xi}_{\frac{1}{2}}}{\sigma_1})^{-\alpha}+(1+\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{-\alpha}-(1+\theta+\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{-\alpha}-(1+\frac{\hat{\xi}_{\frac{1}{2}}}{\sigma_1})^{-\alpha}=\frac{\psi(\alpha,\theta)}{4}. \quad (3.41)$$

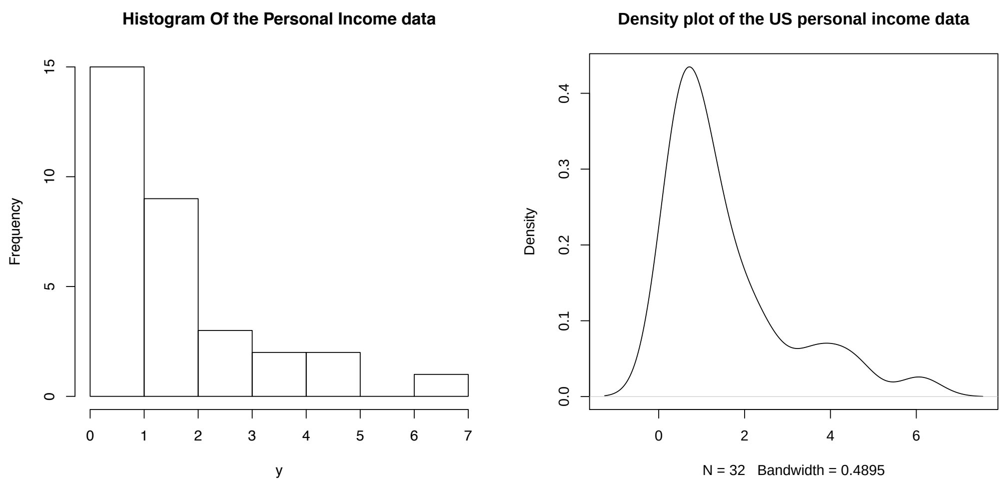
Also from equation (3.39) and equation (3.40) we get

$$(1+\theta+\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{-\alpha}+(1+\frac{\hat{\xi}_{\frac{1}{2}}}{\sigma_1})^{-\alpha}-(1+\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{-\alpha}-(1+\theta+\frac{\hat{\xi}_{\frac{1}{2}}}{\sigma_1})^{-\alpha}=\frac{\psi(\alpha,\theta)}{4}. \quad (3.42)$$

The above three non-linear equations needs to be solved for α, σ, and θ.

3.8 Application of a hidden truncated $P(II)$ model to a real life data set

Typically, in income modeling, the available data are grouped. We consider the US data on personal income (data source: US Census Bureau, 2008) from which we select the number of individuals (considered as percentages to the total population) corresponding to different levels of income. So this data set is also grouped. We argue at this point that the data is subject to hidden truncation because an individual's income is not always correctly specified or it may be unreported. For example the number of people in the income category (say, from 27,500



(a) Histogram of US personal income data 2008 (b) Density of US personal income data 2008

Figure 3.2: Histogram and Density plot for the US personal income data (in percentage) 2008.

to 29,999 (in USD)) as percentages is 2.71 which is not the exact percentage figure as some people in this income group may have earnings more than that from some outside sources which may not be reported.

Before that let us consider the histogram and density plot of the data (given below). In this case the density plot has been drawn by smoothing the histogram. So based on our data we get the following estimates of the parameters for a hidden truncated bivariate $P(II)$ model:

- Estimation based on fractional method of moments:

$$\sigma = 1.007523, \theta = 1.521450, \alpha = 4.050130.$$

- Estimation based on Quartile method:

$$\sigma = 1.011014, \theta = 1.506410, \alpha = 3.999089.$$

- Estimation based on Maximum Likelihood method:

$$\sigma = 0.992756, \theta = 1.5163710, \alpha = 3.9879389.$$

Moreover in this case the standard Kolmogorv-Smirnov goodness of fit test statistic tells us that indeed the fit is good.

So the nature of our data can well be explained by a hidden truncated bivariate $P(II)$ distribution with the following choice of the parameters (approximately):

$$\alpha = 4, \theta = 1.5, \sigma = 1.$$

3.9 Estimation of the parameters using a simulation study

We at first consider a simple situation where we specify all the parameter values and then generate samples (of various sizes) from our hidden truncated density.

3.9.1 Sample generation from the truncated density

For our simulation study we consider the following choices of the parameters , $\alpha = 4, \mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 2, c = 3$. So that

$\theta = \frac{c-\mu_2}{\sigma_2} = 1.5$, and $\psi(\alpha, \theta) = 1 - (1 + (\frac{c-\mu_2}{\sigma_2}))^{-\alpha} = 0.9774$, and our density reduces to

$$f_{X|Y \leq \theta=1.5}(x) = \frac{4}{0.9774}[(1+x)^{-5} - (1+x+1.5)^{-5}]I(0 \leq x \leq \infty). \quad (3.43)$$

In the following tables we illustrate for different sample sizes the estimated values of the parameters using the three methods of estimation

n	$\hat{\sigma}_1$	$\hat{\theta}$	$\hat{\alpha}$
50	0.9978038	1.5162621	3.9754118
100	1.021526	1.507272	4.009601
200	1.011014	1.506410	3.999089

Table 3.1: Estimates of the parameters using the sample quartile method.

n	$\hat{\sigma}_1$	$\hat{\theta}$	$\hat{\alpha}$
50	0.9540499	1.2938447	3.9383091
100	0.9718807	1.5124330	3.8979598
200	0.9378616	1.4897723	3.9351534

Table 3.2: Estimates of the parameters using the maximum likelihood method.

n	$\hat{\sigma}_1$	$\hat{\theta}$	$\hat{\alpha}$
50	0.9943157	1.5172427	3.9654113
100	1.132798	1.508072	4.199970
200	0.9243199	1.4979715	3.9981881

Table 3.3: Estimates of the parameters using the fractional method of moments

3.9.2 Comments on the output of the simulation study

In this small simulation study we observe that for all the estimation procedures the estimate of σ_1 is not so good when the sample size $n = 200$, for all the estimation procedures (except the Sample Quartile method). Also in the case of estimating θ with a sample size of 50 there is appreciable error in the estimates. This scenario, however is not true for the estimates of other parameters. From the output one can easily observe that our estimation is exceptionally good when we consider the quartile method in which with increase in the sample size our estimates are more close to the true value of the parameter, a desirable property of any estimation procedure. However the performance of the maximum likelihood estimation is not that good. One of the several possible reason might be that in situations where we do not have much knowledge about the nature of our likelihood function (specifically the dependence structure among the parameters), the true values for the parameters that we started for our simulation study, those values might not be the optimum values of the parameters for which our likelihood function attains it's maximum. So we can say from above simulation study that in estimating parameters for such models, precise estimation is not possible for estimating the scale parameter using the maximum likelihood method. However a more extensive simulation study is required to determine whether the anomaly of the results in the second table is artifactual.

3.10 Bayesian inference for the hidden truncated bivariate $P(II)$ model

3.10.1 Bayesian analysis with independent priors

As before we have our density of the form

$$f^{HT}_{X|Y\leq c}(x) = \frac{\alpha}{\sigma_1 \psi(\alpha,\theta)}[(1+\frac{x}{\sigma_1})^{-(\alpha+1)} - (1+\frac{x}{\sigma_1}+\theta)^{-(\alpha+1)}]I(x>0). \tag{3.44}$$

3.10.2 Sample and Prior information

We consider a random sample of size n from above density. Next we propose the following choice of independent priors for the three parameters which are as follows:

(a)Prior for α

$$f(\alpha) = \frac{1}{(1+\alpha)^2} I(\alpha > 0). \tag{3.45}$$

(b)Prior for θ

$$f(\theta) = \frac{1}{(1+\theta)^2} I(\theta > 0). \tag{3.46}$$

(c)Prior for σ_1

$$f(\sigma_1) = \frac{1}{(1+\sigma_1)^2} I(\sigma_1 > 0). \tag{3.47}$$

Note that we consider in this case independent (only mildly informative) priors as compared to locally uniform priors for all the parameters.

3.10.3 Posterior distribution of the parameters

In this case our likelihood function is given by

$$\begin{aligned}
&L(\alpha, \sigma_1, \theta | \underline{X} = \underline{x}) \\
&= \prod_{i=1}^{n} [\frac{\alpha}{\sigma_1 \psi(\alpha, \theta)}((1 + \frac{x_i}{\sigma_1})^{-(\alpha+1)} - (1 + \frac{x_i}{\sigma_1} + \theta)^{-(\alpha+1)})] \\
&= \frac{\alpha^n}{(\sigma_1 \psi(\alpha, \theta))^n} \prod_{i=1}^{n} [(1 + \frac{x_i}{\sigma_1})^{-(\alpha+1)}(1 - (1 + \frac{\theta}{(1 + \frac{x_i}{\sigma_1})})^{-(\alpha+1)})].
\end{aligned}$$

So the joint posterior of the three parameters is given by

$$f(\alpha, \sigma_1, \theta | \underline{X} = \underline{x}) = \frac{L(\alpha, \sigma_1, \theta | \underline{X} = \underline{x}) f(\alpha) f(\theta) f(\sigma_1)}{A}. \tag{3.48}$$

Where A is the normalizing constant which is given by

$$
\begin{aligned}
A &= \iiint_{(\mathbb{R}^+)^3} \Pi(\alpha, \sigma_1, \theta | \underline{X} = \underline{x}) f(\alpha) f(\theta) f(\sigma_1) f(\alpha, \sigma_1, \theta | \underline{X} = \underline{x}) d\sigma_1 d\theta d\alpha \\
&= \iiint_{(\mathbb{R}^+)^3} \frac{\alpha^n}{(\sigma_1 \psi(\alpha, \theta))^n} \prod_{i=1}^{n} [(1 + \frac{x_i}{\sigma_1})^{-(\alpha+1)} . \\
&\qquad (1 - (1 + \frac{\theta}{(1 + \frac{x_i}{\sigma_1})})^{-(\alpha+1)})]((1+\theta)(1+\alpha)(1+\sigma_1))^{-2} d\sigma_1 d\theta d\alpha \\
&= \iiint_{(\mathbb{R}^+)^3} \zeta(\alpha, \sigma_1, \theta) d\sigma_1 d\theta d\alpha,
\end{aligned}
\tag{3.49}
$$

where $\zeta(\alpha, \sigma_1, \theta)$ is given by

$$
\zeta(\alpha, \sigma_1, \theta) = \frac{\alpha^n}{(\sigma_1 \psi(\alpha, \theta))^n} \prod_{i=1}^{n} [(1 + \frac{x_i}{\sigma_1})^{-(\alpha+1)} (1 - (1 + \frac{\theta}{(1 + \frac{x_i}{\sigma_1})})^{-(\alpha+1)})] \\
((1+\theta)(1+\alpha)(1+\sigma_1))^{-2}. \tag{3.50}
$$

So the posterior density of α is given by

$$
f(\alpha | \underline{X} = \underline{x}) = \frac{\int_0^\infty \int_0^\infty \zeta(\alpha, \sigma_1, \theta) d\theta d\sigma_1}{A}. \tag{3.51}
$$

Also the posterior density of σ_1 is given by

$$
f(\sigma_1 | \underline{X} = \underline{x}) = \frac{\int_0^\infty \int_0^\infty \zeta(\alpha, \sigma_1, \theta) d\theta d\alpha}{A}. \tag{3.52}
$$

Finally the posterior distribution of θ is given by

$$f(\theta|\underline{X} = \underline{x}) = \frac{\int_0^\infty \int_0^\infty \zeta(\alpha, \sigma_1, \theta) d\theta d\sigma_1}{A}. \tag{3.53}$$

3.10.4 Comment on the choice of Priors and also on the Posterior distribution

Suppose that we have some specific information(in the form of prior belief) about any one of the parameters(say α) that it can take any values between (say) 1 and 2. Then a reasonable choice of prior distribution for α would be any kind of flat prior, (say) a uniform distribution with the support (1,2). This will reduce the complexity in the posterior analysis. Although the use of informative and /or dependent priors will increase the complexity in our analysis, but still one may want to consider them simply because of the fact that for analytically intractable models like ours, the corresponding posterior analysis can be efficiently performed by Markov Chain Monte Carlo (MCMC) algorithm which is specifically designed for complicated models. In our case we consider Metropolis-Hastings algorithm which is a general term for a family of Markov chain simulation methods that are useful for drawing samples from Bayesian posterior distributions.

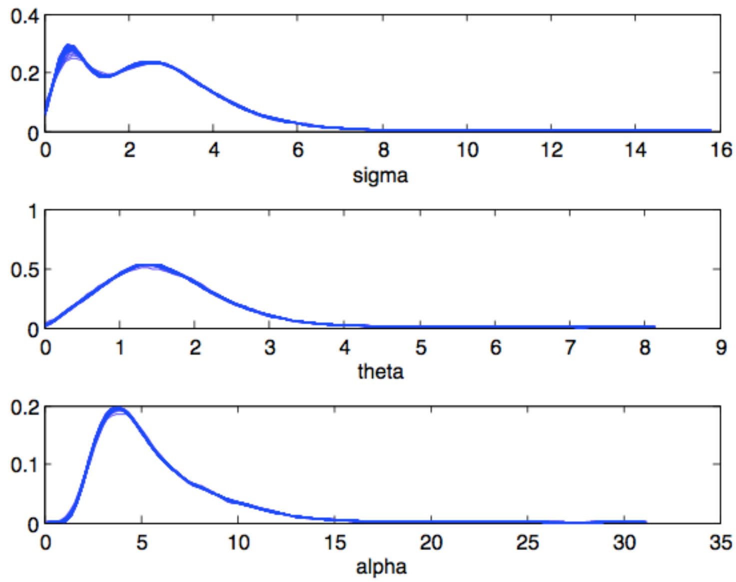
3.10.5 Posterior simulation study

First of all we draw random samples of size 500 and 1000 from our density for a particular choice ($\alpha = 4, \sigma_1 = 1, \theta = 1.5$) of the parameters. For the jumping distribution we consider the gamma distribution but with different shape and scale parameters. Here we consider independent priors for each of the parameters as mentioned in (10.2). The posterior analysis is based on the posterior modes and also the posterior means for each of those three parameters. Below we provide for our MCMC simulation study, various choices for the initial values of the parameters, and the starting distribution for all the parameters under study: Initial choices of the parameters: $\alpha = 2$, $\sigma_1 = 1.1$, $\theta = 1.4$. Jumping distribution of the parameters :

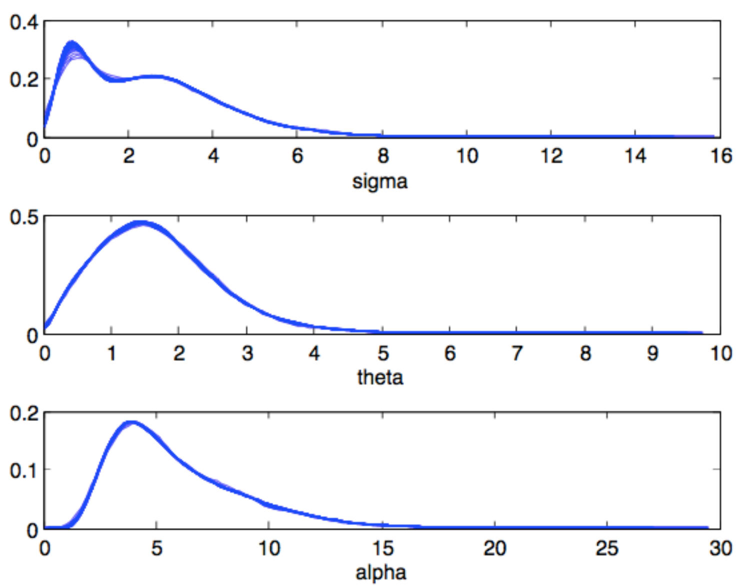
- $\alpha \sim \Gamma(18.9/5, 7/5)$.
- $\sigma_1 \sim \Gamma(5.3/2, 1.8/2)$.
- $\theta_1 \sim \Gamma(8/3, 1.8/3)$.

In the following tables the posterior modes and the posterior means of all the parameters are displayed.

The posterior density plots for the three parameter based on sample of sizes 500 and 1000 are displayed in Figure 3.3(a) and Figure 3.3(b) respectively.



(a) Posterior Density of the parameters for $n = 500$



(b) Posterior Density of the parameters for $n = 1000$

Figure 3.3: Posterior density for all the parameters for different choices of the sample size

n	Mode(α)	Mode(σ_1)	Mode(θ)
100	3.9521	0.9417	1.4383
500	4.0443	0.9372	1.4607
1000	4.0123	0.9673	1.4874

Table 3.4: Bayesian estimates of the parameters using the posterior mode

n	Mean(α)	Mean(σ_1)	Mean(θ)
100	3.9782	0.9537	1.4427
500	4.0391	0.9493	1.4576
1000	3.9857	0.9623	1.4636

Table 3.5: Bayesian estimates of the parameters using the posterior mean

3.10.6 Comment on the posterior simulation study

In our case instead of running the MCMC (Markov Chain Monte Carlo) for one time we consider running a multiple sequence of chains starting from 10,000 simulations up to 15,000 simulations with a increase of 100 in each chain length. The reason for that is that the key to Markov Chain simulation is to create a Markov process whose stationary distribution is a specified $p(\underline{\beta}|data)$, (which in our case is the posterior distribution) and run the simulation long enough so that the distribution of the current draws is close enough to the stationary distribution. From our simulation study, we can not differentiate which one among the posterior means or the posterior modes are uniformly better than the other. From the output in Table(3.4) and in Table(3.5), it may be said that one can use either of them.

3.10.7 Bayesian inference with dependent priors for shape and scale parameter

We have the density which is of the form

$$f^{HT}_{X|Y\leq\theta}(x) = \frac{\alpha}{\sigma_1\psi(\alpha,\theta)}[(1+\frac{x}{\sigma_1})^{-(\alpha+1)} - (1+\frac{x}{\sigma_1}+\theta)^{-(\alpha+1)}]I(x\geq 0), \tag{3.54}$$

where $\psi(\alpha,\theta) = 1-(1+\theta)^{-\alpha}$.

However if we consider $\tau = \frac{1}{\sigma_1}$, where τ is the precision parameter then our density reduces to

$$f^{HT}_{X|Y\leq\theta}(x) = \frac{\alpha\tau}{\psi(\alpha,\theta)}[(1+x\tau)^{-(\alpha+1)} - (1+x\tau+\theta)^{-(\alpha+1)}]I(x\geq 0). \tag{3.55}$$

We will first consider the situation where the prior information for the truncation parameter θ will be independent of the prior information for α and τ. However for the prior information of α and τ, we consider the following:

- The conditional density of α given τ is a Gamma distribution with shape parameter $= \xi_1(\tau)$ and intensity parameter $= \lambda_1(\tau)$. So that the conditional density of α given τ will be

$$f(\alpha|\tau) \propto \alpha^{\xi_1(\tau)-1}\exp(-\alpha\lambda_1(\tau)I(\alpha > 0).$$

- The conditional density of τ given α is a $P(II)$ distribution with inequality parameter $= \lambda_2(\alpha)$ and precision parameter $= 1$. So that the conditional density of τ given α will be

$$f(\tau|\alpha) \propto \lambda_2(\alpha)(1+\tau)^{-(\lambda_2(\alpha)+1)} I(\tau > 0).$$

- While for θ we consider a diffuse prior of the form $f(\theta) \propto \frac{1}{\theta} I(\theta > 0)$.

Note that both the conditional densities of α given τ and also τ given α are members of exponential families. So the joint density (alternatively the joint prior) will be of the form (Arnold Castillo and Sarabia, 1999)

$$\begin{aligned} f(\alpha, \tau) \propto & \exp[-\alpha(c_{11} + c_{12} - c_{13} \log(1+\tau) \\ & + (c_{21} + c_{22} - c_{23} \log(1+\tau)) \log \alpha \\ & + c_{31} + c_{32} - c_{33} \log(1+\tau))] I(\alpha > 0) I(\tau > 0), \end{aligned} \tag{3.56}$$

where $c_{ij}, \forall (i, j = 1, 2, 3)$ are the hyperparameters of the joint distribution.

So the marginal density of α will be given by

$$
\begin{aligned}
f(\alpha) &\propto \int_0^\infty f(\alpha,\tau)d\tau \\
&= \exp[c_{31}+c_{32}-\alpha(c_{11}+c_{12})+(c_{21}+c_{22})\log\alpha]\cdot \\
&\qquad\qquad \int_0^\infty \exp[(\alpha c_{13}-c_{23}\log\alpha-c_{33})\log(1+\tau)]d\tau \\
&= \exp[c_{31}+c_{32}-\alpha(c_{11}+c_{12})+(c_{21}+c_{22})\log\alpha]\times \\
&\int_0^\infty (1+\tau)^{-(-\alpha c_{13}+c_{23}\log\alpha+c_{33})}d\tau \\
&= \frac{\exp[c_{31}+c_{32}-(c_{11}+c_{12})\alpha]\alpha^{c_{21}+c_{22}}}{(1+\alpha c_{13}-c_{23}\log\alpha-c_{33}))} I(\alpha>0). \qquad (3.57)
\end{aligned}
$$

Similarly the marginal density of τ given α will be of the form

$$
\begin{aligned}
f(\tau) &\propto \int_0^\infty f(\alpha,\tau)d\alpha \\
&= \exp[c_{31}+c_{32}-c_{33}\log(1+\tau)]A_{11}I(\tau>0), \qquad (3.58)
\end{aligned}
$$

where

$$
\begin{aligned}
A_{11} &= \int_0^\infty \alpha^{c_{21}+c_{22}-c_{23}\log(1+\tau)}\exp[-\alpha(c_{11}+c_{12}-c_{13}\log(1+\tau))]d\alpha \\
&= \frac{(c_{11}+c_{12}-c_{13}\log(1+\tau))^{-(c_{21}+c_{22}-c_{23}\log(1+\tau)+1)}}{\Gamma(c_{21}+c_{22}-c_{23}\log(1+\tau)+1)}.
\end{aligned}
$$

Again the conditional density of τ given α will be (for each fixed

$\alpha \in \mathbb{R}^+$),

$$f(\tau|\alpha) \propto \alpha^{-c_{23}\log(1+\tau)}(1 + c_{23}\log\alpha - c_{33} - \alpha c_{13})$$
$$\times \exp[-\alpha(c_{13} - c_{33})\log(1+\tau)]I(\tau > 0). \tag{3.59}$$

While the conditional density of α given τ will be(for each fixed $\tau \in \mathbb{R}^+$),

$$f(\alpha|\tau)$$
$$\propto \Gamma(c_{21} + c_{22} - c_{23})\log(1+\tau) + 1)\exp[-\alpha(c_{11} + c_{12})$$
$$+ c_{33}(\alpha + 1)\log(1+\tau)]$$
$$\times ((c_{11} + c_{12} + c_{13})\log(1+\tau))^{c_{21}+c_{22}-c_{23})\log(1+\tau)+1}\alpha^{c_{21}+c_{22}-c_{23}\log(1+\tau)}I(\alpha > 0). \tag{3.60}$$

3.10.8 Assessment of proper choices for the hyperparameters

The assessment of hyperparameters for the hidden truncated Pareto model will be achieved in a manner similar to that described in Arnold and Press (1983). We will consider

- Matching conditional moments and percentiles. However conditional moments corresponding to the density useful for this assessment are listed below:

– From the conditional density of α given τ we can derive

$$
\begin{aligned}
&E(\alpha|\tau_i)\\
&= \frac{1}{\Delta_i}\int_0^\infty \alpha^{c_{21}+c_{22}-c_{23}\log(1+\tau)+1}\exp[-\alpha(c_{11}+c_{12})\\
&+ c_{33}(\alpha+1)\log(1+\tau)]\\
&= \frac{\Gamma(c_{21}+c_{22}-c_{23})\log(1+\tau)+2)}{\Delta_i((c_{11}+c_{12}+c_{13})\log(1+\tau))^{-(c_{21}+c_{22}-c_{23}\log(1+\tau)+2)}},
\end{aligned}
$$

where $\Delta_i = \Gamma(c_{21}+c_{22}-c_{23}\log(1+\tau_i+1))((c_{11}+c_{12}+c_{13})\log(1+\tau))^{c_{21}+c_{22}-c_{23}\log(1+\tau)+1}$.

So that after some algebraic simplification

$$
E(\alpha|\tau_i) = \frac{(c_{21}+c_{22}-c_{23}\log(1+\tau_i)+1)}{(c_{11}+c_{12}+c_{13}\log(1+\tau_i))}, i = 1, 2, \ldots.
$$

– In general we will have (for any $r \geq 1$)

$$
E(\alpha^r|\tau_i) = \frac{\Gamma(c_{21}+c_{22}-c_{23}\log(1+\tau)+r)}{\Delta_i(c_{11}+c_{12}+c_{13}\log(1+\tau_i))^{-(c_{21}+c_{22}-c_{23}\log(1+\tau_i)+r)}}. \tag{3.61}
$$

In particular the conditional variance will be given by

$$
\begin{aligned}
&\sigma^2(\alpha|\tau_i)\\
&= \frac{(c_{21}+c_{22}-c_{23}\log(1+\tau_i)+2)(c_{21}+c_{22}-c_{23}\log(1+\tau_i+1))}{(c_{11}+c_{12}+c_{13}\log(1+\tau_i))^2}\\
&- (\frac{c_{21}+c_{22}-c_{23}\log(1+\tau_i+1)}{c_{11}+c_{12}+c_{13}\log(1+\tau_i)})^2.
\end{aligned} \tag{3.62}
$$

- The conditional percentile function for τ given $\alpha = \alpha_j, j = 1, 2, \ldots$

  If $\varsigma_p(\tau|\alpha_j)$ be the conditional percentile function then we can write

$$P(\tau \leq \varsigma_p(\tau|\alpha_j)|\alpha_j) = p$$
$$(c_{23}\log\alpha - c_{33} - c_{13}\alpha)\int_0^{\varsigma_p(\tau|\alpha_j)} \alpha_j^{-c_{23}\log(1+\tau)}d\tau = p.$$

  However the integral

$$\begin{aligned}\int_0^{\varsigma_p(\tau|\alpha_j)} \alpha_j^{-c_{23}\log(1+\tau)}d\tau &= \left(\frac{1}{c_{23}\log\alpha_j}\right)\int_{\alpha_j^{-c_{23}\log(1+\varsigma_p(\tau|\alpha_j))}}^{\varsigma_p(\tau|\alpha_j)} u^{\frac{1-(c_{13}-c_{33})\alpha}{\delta(\alpha)}}du \\ &= \frac{(1-\alpha^{-c_{23}\log(1+\varsigma_p(\tau|\alpha_j))})^{1+\frac{1-(c_{13}-c_{33})\alpha}{\delta(\alpha)}}}{(c_{23}\log\alpha_j)1+\frac{1-(c_{13}-c_{33})\alpha}{\delta(\alpha)}}.\end{aligned}$$

  So that we have after some algebraic simplification

$$\varsigma_p(\tau|\alpha_j) = [1 - (\frac{p}{\varsigma_1(\alpha_j)})^{\frac{1}{\varsigma_2(\alpha_j)}}]^{\frac{1}{\delta(\alpha_j)}-1}, \qquad (3.63)$$

  for each $j = 1, 2, \ldots$ and where $\varsigma_1(\alpha_j) = \frac{(c_{23}\log\alpha_j - c_{33} - c_{13}\alpha_j)}{(1+\frac{1-(c_{13}-c_{33})\alpha_j}{\delta(\alpha)})c_{23}\log\alpha_j}$ and $\varsigma_2(\alpha_j) = 1 + \frac{1-(c_{13}-c_{33})\alpha_j}{\delta(\alpha_j)}$.

3.11 Likelihood ratio test of the truncation parameter

In our case we want to test whether there has been a truncation or not. In other words we wish to test whether the parameter θ is finite. Before we proceed further, we reparameterize our density by considering $\beta = \frac{1}{\theta}$, in which case our density reduces to

$$f^{HT}_{X|Y\leq c}(x) = \frac{\alpha}{\sigma_1\psi(\beta)}[(1+\frac{x}{\sigma_1})^{-(\alpha+1)}-(1+\frac{x}{\sigma_1}+\frac{1}{\beta})^{-(\alpha+1)}]I(x>0), \quad (3.64)$$

where

$$\psi(\beta) = 1-(1+\frac{1}{\beta})^{-\alpha}.$$

Now we can rewrite our problem as follows: We want to test

$$H_0 : \beta = 0$$

against

$$H_a : \beta > 0$$

where H_0 and H_a denote the null hypothesis and the alternative hypothesis respectively. Note that under the null i.e., $\beta = 0$, our density reduces to a simple $P(II)$ model which is given by:

$$f(x) = (\frac{\alpha}{\sigma_1})[(1+\frac{x}{\sigma_1})^{-(\alpha+1)}]I(x \geq 0).$$

So under the null hypothesis our likelihood function (for a random sample of size n drawn from above density) is given by

$$L(\alpha, \sigma_1, \theta) = \prod_{i=1}^{n}[(\frac{\alpha}{\sigma_1})(1+\frac{x_i}{\sigma_1})^{-(\alpha+1)}]. \quad (3.65)$$

Equivalently the log-likelihood function is given by:

$$\log L(\alpha, \sigma_1) = n \log \alpha - n \log \sigma_1 - (\alpha + 1) \sum_{i=1}^{n} \log(1 + \frac{x_i}{\sigma_1}). \qquad (3.66)$$

The corresponding likelihood equation for α is:

$$\frac{\partial}{\partial \alpha}[\log L(\alpha, \sigma_1)] = 0,$$

$$\Rightarrow \frac{n}{\alpha} - \sum_{i=1}^{n}[\log(1 + \frac{x_i}{\sigma_1})] = 0. \qquad (3.67)$$

The second likelihood equation is:

$$\frac{\partial}{\partial \sigma_1}[\log L(\alpha, \sigma_1, \theta)] = 0. \qquad (3.68)$$

Equivalently we can write

$$-\frac{n}{\sigma_1} + (\alpha + 1) \sum_{i=1}^{n}[\frac{1}{(1 + \frac{x_i}{\sigma_1})} \frac{x_i}{(\sigma_1)^2}] = 0.$$

So under the null the supremum of the likelihood is given by

$$Sup_{H_0} L(\alpha, \sigma_1) = \prod_{i=1}^{n}[\frac{(\hat{\alpha})}{\hat{\sigma}_1}(1 + \frac{x_i}{\hat{\sigma}_1})^{-(\hat{\alpha}+1)}],$$

where $\hat{\alpha}$ and $\hat{\sigma}_1$ denotes the estimates of the parameters under the null hypothesis. Again under the alternative our likelihood equation is given by

$$
\begin{aligned}
L(\alpha,\sigma_1,\beta) &= \prod_{i=1}^{n}[\frac{\alpha}{\sigma_1\psi(\beta)}((1+\frac{x_i}{\sigma_1})^{-(\alpha+1)}-(1+\frac{x_i}{\sigma_1}+\frac{1}{\beta})^{-(\alpha+1)})] \\
&= \frac{\alpha^n}{(\sigma_1\psi(\beta))^n}\prod_{i=1}^{n}[(1+\frac{x_i}{\sigma_1})^{-(\alpha+1)}(1-(1+\frac{1}{\beta(1+\frac{x_i}{\sigma_1})^{-(\alpha+1)}}))].
\end{aligned}
$$

Equivalently the log-likelihood function is given by:

$$
\begin{aligned}
&\log L(\alpha,\sigma_1,\beta) \\
&= n\log\alpha - n\log\sigma_1 - n\log\psi(\beta) - (\alpha+1)\sum_{i=1}^{n}[\log(1+\frac{x_i}{\sigma_1}) \\
&+ \log(1-(1+\frac{1}{\beta(1+\frac{x_i}{\sigma_1})})^{-(\alpha+1)})].
\end{aligned}
$$

So the first of the three likelihood equations is:

$$
\frac{\partial}{\partial\alpha}[\log L(\alpha,\sigma_1,\beta)] = 0,
$$

$$
\Rightarrow \frac{n}{\alpha} - \sum_{i=1}^{n}[\log(1+\frac{x_i}{\sigma_1}) + \frac{(1+\frac{1}{\beta(1+\frac{x_i}{\sigma_1})})^{-(\alpha+1)}}{(1-(1+\frac{1}{\beta(1+\frac{x_i}{\sigma_1})})^{-(\alpha+1)})}\log\alpha] = 0.
$$

Also we have

$$
\frac{\partial}{\partial\sigma_1}[\log L(\alpha,\sigma_1,\beta)] = 0.
$$

Equivalently we can write

$$-\frac{n}{\sigma_1}+(\alpha+1)\sum_{i=1}^{n}[\frac{x_i}{(1+\frac{x_i}{\sigma_1})\sigma_1^2}]$$
$$+(\alpha+1)\sum_{i=1}^{n}[(\frac{\frac{x_i(1+\frac{1}{\beta})}{(\sigma_1+x_i)^2}}{(1-(1+\frac{1}{\beta(1+\frac{x_i}{\sigma_1})})^{-(\alpha+1)})})(\frac{x_i+\sigma_1(1+\frac{1}{\beta})}{\sigma_1+x_i})^{-(\alpha+2)}]=0.$$

Furthermore we have

$$\frac{\partial}{\partial\beta}[\log L(\alpha,\sigma_1,\beta)]=0,$$

$$\Rightarrow -n\frac{\psi'(\beta)}{\psi(\beta)}-\beta^{-2}(\alpha+1)\sum_{i=1}^{n}[\frac{(1+\frac{1}{\beta(1+\frac{x_i}{\sigma_1})^{-(\alpha+2)}})}{(1-(1+\frac{1}{\beta(1+\frac{x_i}{\sigma_1})})^{-(\alpha+1)})}(1+\frac{x_i}{\sigma_1})^{-1}]=0.$$

Note that here $\psi(\beta)=1-(1+\frac{1}{\beta})^{-\alpha}$, so that $\frac{\partial}{\partial\beta}[\psi(\beta)]=-\alpha(1+\frac{1}{\beta})^{-(\alpha+1)}\beta^{-2}$.

So the supremum of the likelihood under the alternative is given by

$$Sup_{H_0UH_a}L((\alpha,\sigma_1,\beta))=\prod_{i=1}^{n}[\frac{\hat{\alpha}}{\hat{\sigma_1}\hat{\psi}(\beta)}((1+\frac{x_i}{\hat{\sigma_1}})^{-(\hat{\alpha}+1)}-(1+\frac{x_i}{\hat{\sigma_1}}+\frac{1}{\hat{\beta}})^{-(\hat{\alpha}+1)})].$$

Hence the likelihood ratio for the above test is given by

$$\lambda=\frac{Sup_{H_0}L(\alpha,\sigma_1)}{Sup_{H_0UH_a}L((\alpha,\sigma_1,\beta))}.$$

Equivalently the log of the likelihood ratio is given by

$$\log \lambda = \log(Sup_{H_0} L(\alpha, \sigma_1)) - \log(Sup_{H_0 U H_a} L((\alpha, \sigma_1, \beta))).$$

3.11.1 Asymptotic property of the ML estimate of θ and the large sample distribution of the likelihood ratio statistic

First of all we consider the following:

- we will use for notational simplicity $\underline{\xi} = (\alpha, \sigma_1, \beta)$ and our parameter space $\Omega = \mathbb{R}^{+^3} \subset \mathbb{R}^3$, and from here onwards we will consider the following $l(\underline{\xi}|\underline{x}) = \sum_{i=1}^{n} \log(L(\underline{\xi})) = \sum_{i=1}^{n} \log(f(x_i; \alpha, \sigma_1, \beta))$.
- We note that the true value of the parameter β, (i.e., $H_0 : \beta = 0 = \beta_0$, say) is on the boundary of Ω.

Next we also consider the following regularity conditions which are listed below:

- The parameter space Ω has finite dimension 3 and we consider one of the parameters β to be on the boundary.
- $f^*(x_i; \alpha', \sigma_1', \beta') = f^*(x_i; \alpha, \sigma_1, \beta)$ if and only if $\alpha' = \alpha, \sigma_1' = \sigma_1, \beta' = \beta$ $\forall (\alpha', \sigma_1', \beta', \alpha, \sigma_1, \beta) \in \Omega$.
- We consider the almost sure existence of the first three derivatives on the intersection of neighborhoods of the true parameter value and Ω. In our case since the true value which is β_0, is on the bound-

ary so the derivatives of $l(\underline{\xi}|\underline{x})$ has to be taken from appropriate side

- There exists an open subset of ω of $\mathbb{R}^3$ containing β_0 such that for all x for which $f^*(x_i;\alpha,\sigma_1,\beta) > 0$, has all three derivatives w.r.t all $\underline{\xi} \in \omega$, and

$$|\frac{\partial^3}{\partial\alpha\partial\sigma_1\partial\beta}\log[f^*(x;\alpha,\sigma_1,\beta)]| \le M(x),$$

$\forall((\alpha,\sigma_1,\beta))\text{and} f^*(x;\alpha,\sigma_1,\beta) > 0$ and $m = E_{\beta_0}[M(x)] < \infty$.

- The expectation of $n^{-1}I(\underline{\xi})$ is assumed to be positive definite on the neighborhood of β_0 and at β_0 is equal to the variance covariance matrix of $n^{-\frac{1}{2}}U(\beta_0)$, where $U(.)$ is the first derivative of the log-likelihood function.

Then according to Self and Liang (1987), pp. 607 (Theorem 3), we may consider the following asymptotic representation of the likelihood which is as follows:

$$Sup_{\xi\in C_{\Omega}-\beta_0}[-(Z-\xi)'I(\beta_0)(Z-\xi)] - Sup_{\xi\in C_{\Omega_0}-\beta_0}[-(Z-\xi)'I(\beta_0)(Z-\xi)]. \tag{3.69}$$

where Z has a multivariate Gaussian distribution with mean 0 and variance covariance matrix $I^{-1}(\beta_0)$ and C_{Ω} and C_{Ω_0} are non-empty cones approximating Ω_0, a subset of Ω where β_0 lies and Ω_1, which the complement of Ω_0. Mimicking Self and Liang (1987) we may write alternatively equation(3.69) as

$$inf_{\xi \in \tilde{C}_0} \parallel \tilde{Z} - \xi \parallel^2 - inf_{\xi \in \tilde{C}} \parallel \tilde{Z} - \xi \parallel^2, \tag{3.70}$$

with

- $\tilde{C} = (\tilde{\xi} : \tilde{\xi} = \Lambda^{\frac{1}{2}} P^T \xi \ \forall \xi \in C_{\Omega} - \beta_0)$,
- $\tilde{C}_0 = (\tilde{\xi} : \tilde{\xi} = \Lambda^{\frac{1}{2}} P^T \xi \ \forall \xi \in C_{\Omega_0} - \beta_0)$.

and where $\tilde{Z}$ has a multivariate Gaussian distribution with mean 0 and identity covariance matrix and $P \Lambda P^T$ represents the spectral decomposition of $I(\beta_0)$.

However in our case we have

$$\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$$

where Ω_i' s are open intervals in $\mathbb{R}^+$ and we have a situation here for which $\tilde{C} = [0, \infty) \times \mathbb{R}^{+^2}$ and $\tilde{C}_0 = 0 \times \mathbb{R}^{+^2}$, and because of that our equation (3.71) reduces to

$$\tilde{Z_1}^2 I(\tilde{Z}_1 > 0). \tag{3.71}$$

So the asymptotic distribution of $-2 \log \lambda$ is $50 : 50$ mixture of a χ_2^0 and χ_2^1 distribution.

3.12 Asymptotic distribution of the smallest order statistic

We have as before our density which is of the form

$$f^{HT}_{X|Y\leq\theta}(x) = \frac{\alpha}{\sigma_1\psi(\alpha,\theta)}[(1+\frac{x}{\sigma_1})^{-(\alpha+1)} - (1+\frac{x}{\sigma_1}+\theta)^{-(\alpha+1)}]I(x\geq 0).$$

So the corresponding distribution function is given by

$$F(x) = 1 - \frac{(1+\frac{x}{\sigma_1})^{-\alpha} - (1+\frac{x}{\sigma_1}+\theta)^{-\alpha}}{\psi(\alpha,\theta)}, x\geq 0.$$

So for a random sample of size n from above density the distribution of the smallest order statistic ($X_{1:n} = min_{i=1}^{n} X_i$) will be

$$f(x_{1:n}) = \frac{n\alpha}{(\psi(\alpha,\theta))^n}[(1+\frac{x_{1:n}}{\sigma_1})^{-\alpha} - (1+\frac{x_{1:n}}{\sigma_1}+\theta)^{-\alpha}]^{n-1}[(1+\frac{x_{1:n}}{\sigma_1})^{-(\alpha+1)} - (1+\frac{x_{1:n}}{\sigma_1}+\theta)^{-(\alpha+1)}]I(x_{1:n}\geq 0). \quad (3.72)$$

Let us consider the following

$$P(X_{1:n} > \frac{x_{1:n}}{n}) = [\frac{(1+\frac{x_{1:n}}{n\sigma_1})^{-\alpha} - (1+\frac{x_{1:n}}{n\sigma_1}+\theta)^{-\alpha}}{\psi(\alpha,\theta)}]^n, \quad (3.73)$$

using the expression for the distribution function as given earlier. Next consider the quantity

$$(1+\frac{x_{1:n}}{n\sigma_1})^{-\alpha}-(1+\frac{x_{1:n}}{n\sigma_1}+\theta)^{-\alpha}$$
$$=[1-\alpha(\frac{x_{1:n}}{n\sigma_1})+\frac{\alpha(\alpha+1)}{2}(\frac{x_{1:n}}{n\sigma_1})^2-\cdots]$$
$$-[(1+\theta)^{-\alpha}-\alpha(1+\theta)^{-\alpha+1}(\frac{x_{1:n}}{n\sigma_1})+\frac{\alpha(\alpha+1)}{2}(1+\theta)^{-\alpha+2}(\frac{x_{1:n}}{n\sigma_1})^2-\cdots]$$
$$=1-(1+\theta)^{-\alpha}-\alpha(1-(1+\theta)^{-\alpha+1})(\frac{x_{1:n}}{n\sigma_1})+\mathrm{o}(n^{-2})$$
$$=\psi(\alpha,\theta)-\alpha(1-(1+\theta)^{-\alpha+1})(\frac{x_{1:n}}{n\sigma_1})+\mathrm{o}(n^{-2}).$$

So on substitution in equation (3.73), we get

$$P(X_{1:n}>\frac{x_{1:n}}{n})=[1-\frac{\alpha(1-(1+\theta)^{-\alpha+1})}{\sigma_1\psi(\alpha,\theta)}\frac{x_{1:n}}{n}+\mathrm{o}(n^{-2})]^n.$$

Hence we may consider

$$\begin{aligned}\lim_{n\to\infty}P(X_{1:n}>\frac{x_{1:n}}{n}) &= \lim_{n\to\infty}[1-\frac{\alpha(1-(1+\theta)^{-\alpha+1})}{\sigma_1\psi(\alpha,\theta)}\frac{x_{1:n}}{n}+\mathrm{o}(n^{-2})]^n\\ &= \exp[-Ax_{1:n}], x_{1:n}\geq 0,\end{aligned}$$

where $A=\frac{\alpha(1-(1+\theta)^{-\alpha+1})}{\sigma_1\psi(\alpha,\theta)}$. So the asymptotic distribution of $X_{1:n}$ is exponential with the intensity parameter$=\frac{\alpha(1-(1+\theta)^{-\alpha+1})}{\sigma_1\psi(\alpha,\theta)}$, when the samples are drawn from a density of the form in (3.15).

3.13 Concluding remarks

Precise inference for the parameters of hidden truncation models is still an area where much work has to be done so that it becomes a powerful tool for the statisticians to analyze data which has been subject to hidden truncation with respect to one or more covariable(s). The hidden truncated $P(II)$ model does not belong to the exponential families of densities. As a result, essentially no reduction in complexity of the data can be obtained by invoking sufficiency arguments. Maximum likelihood estimation will necessarily be performed numerically as we have done, but unfortunately, the likelihood functions associated with these types of models often do not have easily identified modes because of the unavailability of analytical expressions for the maximum likelihood estimates. Arnold and Beaver (2002) provided a careful discussion for such models and associated inference procedures. Data sets involving hidden truncation are quite common nowadays and to deal with such types of data efficiently we do need a strong theory. The associated Bayesian analysis that is reported gives us a hint of how to deal with such types of models under the Bayesian framework. However the possibility of "multiple" hidden truncations cannot be ignored. One can easily imagine situations in which observations are made only if all of several other covariables attain certain critical levels which we have considered under the multivariate setting. To this end the tractability and also the applicability of such models are the two main concerns before we engage ourself in developing all the necessary theoretical work.

Appendix

Large sample property of the likelihood estimates

Before we discuss the large sample behavior of the likelihood estimates it is necessary for us to consider the information matrix which is by definition the expected value of the Hessian matrix (i.e., minus the second derivative of the log likelihood function at $\hat{\underline{\eta}}$ given the data $\underline{x}$, where the Hessian matrix is given by

$$H(\underline{\eta}) = -\frac{\partial}{\partial\underline{\eta}(\partial\underline{\eta})^T}[\log L(\underline{\eta}|\underline{X} = \underline{x})], \tag{3.74}$$

where "T" stands for transpose and $\underline{\eta} = (\alpha, \sigma_1, \theta)$.

So in our case the observed information matrix is given by

$$H(\underline{\eta}) = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}.$$

(say), where the elements of the Hessian matrix are given later.

Comment on the asymptotic efficiency of the ML estimates

Here we observe that in our case $\underline{X}|Y \leq \theta$ is independently and identically distributed with a density which satisfies the following:

- The parameter space $\Omega = \underline{\eta} \in \mathbb{R}^{+3}$, where $\mathbb{R}^{+}$ denotes the positive part of the real line, contains an open set ω, say, of which the true parameter $\underline{\eta}_0$ is an interior point.

- For all realizations from our density and for $\underline{\eta}_0$

$$|\frac{\partial^3}{\partial\alpha\partial\sigma_1\partial\theta}[\log f_{\underline{X}|Y\leq\theta}(\underline{x}|\underline{\eta})]| \leq M(\underline{x})\forall\underline{\eta} \in \omega, \tag{3.75}$$

  and $m = E_{\underline{\eta_0}}[M(\underline{x})] < \infty$.

Then with probability $\to$1 as $n \to \infty$, there exists $\hat{\underline{\eta}} = \hat{\underline{\eta}}(\underline{x})$ a solution of the likelihood equations such that

- $\hat{\underline{\eta}}_j$ is consistent,

- $\sqrt{n}(\hat{\underline{\eta}}_j - \underline{\eta})$ is asymptotically normal with mean vector zero and covariance matrix $[H(\underline{\eta})]^{-1}$,

$$\sqrt{n}(\hat{\underline{\eta}}_j - \underline{\eta}) \xrightarrow{\mathcal{L}} N(0, [H(\underline{\eta}_j)]_{jj}^{-1}), \tag{3.76}$$

where for $j = 1, 2, 3$, $\underline{\eta}_j$ would mean α, σ_1, θ respectively. In this chapter FIM will always mean Fisher Information Matrix whenever it appears. We have in our case a complicated situation similar to that described in Cox (1958), for which it is much easier to compute the observed FIM than expected FIM and also it tends to agree more closely with Bayesian and fiducial analysis. Next the elements of the Hessian matrix are

$$
\begin{aligned}
A_{11} &= \frac{\partial^2}{\partial \alpha^2}[\log L(\underline{\eta}|\underline{X} = \underline{x})] \\
&= \frac{n}{\alpha^2} - n\left[\frac{-(1-(1+\theta)^{-\alpha})(1+\theta)^{-\alpha}(\log(1+\theta))^2 - ((1+\theta)^{-\alpha}\log(1+\theta))^2}{(1-(1+\theta)^{-\alpha})^2}\right] \\
&\quad + \sum_{i=1}^{n}\Bigg[\Big(\log\Big(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})}\Big)\Big)\times \\
&\quad \Big(1-\Big(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})}\Big)^{-(\alpha+1)}\Big)^{-2}\Big(-\Big(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})}\Big)^{-2(\alpha+1)}\log\Big(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})}\Big) \\
&\quad + \log\Big(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})}\Big)\Big(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})}\Big)^{-2(\alpha+1)}\Big)\Bigg].
\end{aligned}
$$

Again

$$A_{12} = \frac{\partial^2}{\partial\alpha\partial\sigma_1}[\log L(\underline{\eta}|\underline{X} = \underline{x})]$$

$$= \sum_{i=1}^{n}[\frac{X_i}{(\sigma_1)^2(1+\frac{X_i}{\sigma_1})}] + \sum_{i=1}^{n}[\theta(\frac{X_i}{\sigma_1})(\sigma_1+X_i)^{-2}\frac{(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+2)}}{(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}})]$$

$$+ (\alpha+1)\sum_{i=1}^{n}[\theta(\frac{X_i}{\sigma_1})(\sigma_1+X_i)^{-2}$$

$$-(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})^{-1}\log(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}$$

$$+ (1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(2\alpha+3)})].$$

Also

$$A_{13} = \frac{\partial^2}{\partial\alpha\partial\theta}[\log L(\underline{\eta}|\underline{X} = \underline{x})]$$

$$= -n[\frac{-(1-(1+\theta)^{-\alpha})(\alpha(\alpha+1))((1+\theta)^{-\alpha+2}) - (\alpha(1+\theta)^{-\alpha+1})^2}{(1-(1+\theta)^{-\alpha})^2}]$$

$$+ \sum_{i=1}^{n}[(1+\frac{X_i}{\sigma_1})^{-1}(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+2)}(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})^{-1}]$$

$$+ (\alpha+1)\sum_{i=1}^{n}[(1+\frac{X_i}{\sigma_1})^{-1}\log(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})^{-2}\times$$

$$(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+2)}(2(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}-1)].$$

Similarly

$$A_{21} = \frac{\partial^2}{\partial\sigma_1\partial\alpha}[\log L(\underline{\eta}|\underline{X} = \underline{x})]$$

$$= \sum_{i=1}^{n}[\frac{X_i}{(\sigma_1)^2(1+\frac{X_i}{\sigma_1})^2}] + \sum_{i=1}^{n}[\frac{(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+2)}}{1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}}\theta X_i(\sigma_1+X_i)^{-2}]$$

$$+ \sum_{i=1}^{n}[\log(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})(\alpha+1)(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})^{-2}$$

$$(\theta X_i(\sigma_1+X_i)^{-2})(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+2)}(1-2(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})].$$

Again

$$A_{22} = \frac{\partial^2}{\partial(\sigma_1)^2}[\log L(\underline{\eta}|\underline{X} = \underline{x})]$$

$$= -\frac{2n}{(\sigma_1)^3} + (\alpha+1)\sum_{i=1}^{n}[\frac{(-3X_i-2\sigma_1)X_i}{(\sigma_1)^2(1+\frac{X_i}{\sigma_1})^2}]$$

$$+\theta(\alpha+1)\sum_{i=1}^{n}[\frac{(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+2)}}{(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})}(\frac{(-\frac{X_i^2}{\sigma_1^3})(\sigma_1+X_i)^2-2(\sigma_1+X_i)\frac{X_i}{\sigma_1}}{((\sigma_1+X_i)^2)^2})].$$

$$A_{23} = \frac{\partial^2}{\partial\sigma_1\partial\theta}[\log L(\underline{\eta}|\underline{X} = \underline{x})]$$

$$= (\alpha+1)\sum_{i=1}^{n}[\frac{(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+2)}}{1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}}(1+\frac{X_i}{\sigma_1})^{-2}\frac{X_i}{(\sigma_1)^2}$$

$$+(1+\frac{X_i}{\sigma_1})^{-1}(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})^{-2}((\frac{\theta X_i}{(\sigma_1)^2})(-((\alpha+1)(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(2\alpha+3)})$$

$$-(\alpha+2)(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+3)}))].$$

Furthermore

$$A_{31} = \frac{\partial^2}{\partial\theta\partial\alpha}[\log L(\underline{\eta}|\underline{X} = \underline{x})]$$

$$= -n[\frac{-(1-(1+\theta)^{-\alpha})((1+\theta)^{-2(\alpha+1)}\log(1+\theta)\alpha) - (\alpha(1+\theta)^{-\alpha})^2}{(1-(1+\theta)^{-\alpha})^2}]$$

$$+\sum_{i=1}^{n}[\frac{(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}}{1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}}(1+\frac{X_i}{\sigma_1})^{-1}(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-1}]$$

$$+\sum_{i=1}^{n}[\log(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})((1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{\alpha})(1+\frac{X_i}{\sigma_1})^{-1}(\alpha+1)$$

$$((-(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}))$$

$$-((1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(2\alpha+4)}(\alpha+1)(1+\frac{X_i}{\sigma_1})^{-1}))].$$

Again

$$A_{32} = \frac{\partial^2}{\partial\theta\partial\sigma_1}[\log L(\underline{\eta}|\underline{X} = \underline{x})]$$

$$= (\alpha+1)[\sum_{i=1}^{n}[(\frac{\frac{X_i}{\sigma_1}}{(\sigma_1+X_i)^2})(\frac{(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}}{1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}})]$$

$$+\theta[(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})^{-2}(1+\frac{X_i}{\sigma_1})^{-1}(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+3)}((-(\alpha+2)\times$$

$$(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}))-((\alpha+1))(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})]].$$

$$A_{33} = \frac{\partial^2}{\partial\theta^2}[\log L(\underline{\eta}|\underline{X} = \underline{x})]$$

$$= -n[\frac{-(1-(1+\theta)^{-\alpha})(\alpha(\alpha+1))((1+\theta)^{-(\alpha+2)})-(\alpha(1+\theta)^{-(\alpha+1)})^2}{(1-(1+\theta)^{-\alpha})^2}]$$

$$-(\alpha+1)\sum_{i=1}^{n}[(1+\frac{X_i}{\sigma_1})^{-2}(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)})^{-2}(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+3)}\times$$

$$(((\alpha+2)(1-(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}))+((\alpha+1)(1+\frac{\theta}{(1+\frac{X_i}{\sigma_1})})^{-(\alpha+1)}))].$$

Chapter 4

Classical and Bayesian inference for a hidden truncated bivariate $P(IV)$ distribution

4.1 Introduction

The $P(IV)$ family was suggested by Arnold and Laguna (1976), as well as by Ord (1975) and Cronin (1977, 1979). It is to be noted that most of the distribution theory regarding the $P(IV)$ distribution can be obtained by using available results for the Burr distributions as mentioned in Johnson and Kotz (1970). In particular, the Burr (XII) family can readily be recognized as a $P(IV)$ family with a suitable identification of the parameters. However there are also various other types of characterization of the $P(IV)$ distribution in the literature. Among them one could consider earlier works by Dubey (1968) and Harris and Singpurwalla (1969), who arrived at the $P(IV)$ distribution via a mixture of Weibull random variables. Later Singh and Maddala(1976) were led to the $P(IV)$ model using an argument involving decreasing failure rates. However we will follow the nomenclature used by Arnold (1983) and use the name Pareto (IV) for this family of distributions, abbreviated as $P(IV)$. The $P(IV)$ model is one of the most general families of Pareto distributions endowed with two shape parameters together with location and scale parameters. However as mentioned in a previous chapter, as we proceed to more and more complex models for Pareto distributions all the available "nice" results for the classical Pareto model disappear. So a valid question is whether or not it is possible to provide some efficient estimation strategies in situations in which, under a bivariate $P(IV)$ model, the variable of interest (say, X) is only observable if and only if the unobserved co-variable (say, Y) is truncated from above. As in the previous chapter (chapter 3) we are considering estimation procedures under the hidden truncation paradigm, but this time for a bivariate $P(IV)$ model. In addition, Bayesian estimation of the parameters of such models will be considered. This is the subject matter of this chapter. Again, as was mentioned in chapter 3, inference procedures for the $P(IV)$ model, both from the classical and Bayesian viewpoints have been severely restricted due to a lack of simplicity and analytical tractability of the model. Inference techniques for $P(IV)$ populations have typically assumed several of the parameters to be known. An exception is encountered in Harris and Singpurwalla (1969) where the likelihood equations for the full model were derived.

If all the parameters are assumed unknown, then such well known techniques such as the method of moments and the maximum likelihood method will involve iterative solution of systems of distinctly non-linear equations which for the $P(IV)$ model are more complicated than for the $P(II)$ model. Since $X_{1:n}$ is a consistent estimate of μ in the $P(IV)$ family, one may consider setting $\mu = X_{1:n}$ and solving the resulting simplified equations for the method of moments and quartile estimation. For the method of moments estimation we will consider the approach of Arnold and Laguna (1977) who used sample fractional moments. In the quartile estimation procedure we will follow the approach suggested by Quandt (1966).

4.2 Why the hidden truncation $P(IV)$ model

First of all note that the $P(IV)$ family provides a convenient vehicle for computing distributional results for the three more specialized Pareto families. Each of them may be identified as special cases of the Pareto (IV) family as follows: $P(I)(\sigma, \alpha) = P(IV)(\sigma, \sigma, 1, \alpha)$, $P(II)(\mu, \sigma, \alpha) =$

$P(IV)(\mu, \sigma, 1, \alpha)$ and $P(III)(\mu, \sigma, \gamma) = P(IV)(\mu, \sigma, \gamma, 1)$. So among the family of Pareto distributions, the $P(IV)$ model deserves further attention simply because of the fact that within our hierarchy of generalized Pareto distributions, it is the $P(IV)$ family which might be well adapted to modeling reliability problems and in addition it is a viable competitor for other popular models for income distributions, for example the log-logistic and log-normal models. Next we envisage a situation where we have data on an individual's income corresponding to different segments of time, and we assume that they can be well described by a $P(IV)$ model. One might be interested in the distribution of the most recent income values corresponding to income values recorded in earlier time period not exceeding a certain level. Models of such types can be explained by a hidden truncation paradigm where we observe one variable only when it is subject to hidden truncation from above with respect to one or more covariables.

In this chapter we focus our attention on estimation of all the parameters of a bivariate $P(IV)$ distribution when hidden truncation is applied to one of the variables with the only restriction that the truncation point will be greater than the location parameter of the truncated variable. As was observed in chapter 3, hidden truncation from below did not augment the $P(II)$ model. The same phenomenon occurs with the $P(IV)$ model also, since the resulting density will be again a member in the same family of distributions with only reparametrization of the parent model. In contrast, when we consider the hidden truncation paradigm when one variable is subject to hidden truncation from above for a bivariate $P(IV)$ distribution, model augmentation is observed. Inference procedures for such hidden truncation models are considered in this chapter both from the classical and the Bayesian perspectives.

The remainder of the chapter is organized in the following way. In section 3, we will develop the hidden truncation density of a bivariate $P(IV)$ model when one of the variable is subject to hidden truncation from above with the only restriction that the truncation point is at least greater than the location parameter of the truncated variable. In section 4, we will consider the method of fractional moments for estimating all the parameters involved in the model. In section 5, we will consider the quartile method of estimation. In section 6, we will consider a simulation study and will report the estimated values of all the parameters using both of the estimation strategies. In section 7, we will consider a real life situation where a small data set is analyzed by invoking the hidden truncated bivariate $P(IV)$ model as an application of our model. In section 8, we derive the asymptotic distribution of the smallest order statistic when the samples are drawn from a hidden truncated bivariate $P(IV)$ distribution. In section 9, we will consider the estimation of all the parameters under the Bayesian paradigm where we will consider posterior simulation results based on a simulation study together with comments related to posterior distribution convergence along with the choice of the parameters of the jumping distribution.

4.3 Hidden truncated bivariate $P(IV)$ model

The survival function corresponding to a bivariate $P(IV)$ distribution is given by

$$P(X > x, Y > y) = [1 + (\frac{(x-\mu_1)}{\sigma_1})^{\frac{1}{\delta_1}} + (\frac{(y-\mu_2)}{\sigma_2})^{\frac{1}{\delta_2}}]^{-\alpha}, x \geq \mu_1, y \geq \mu_2, \tag{4.1}$$

where $\mu_1,\ \mu_2,\ \sigma_1,\ \sigma_2,\ \delta_1,\ \delta_2$ are the location, scale and inequality parameter for X and Y respectively and α is the shape parameter. If (X, Y) has a survival function of the form in (4.1), then we will write

$$(X, Y) \sim \text{bivariate Pareto}(IV)(\mu_1, \mu_2, \sigma_1, \sigma_2, \delta_1, \delta_2, \alpha).$$

So the corresponding joint density of (X, Y) will be given by

$$\begin{aligned}
f(x,y) =& \frac{\partial}{\partial x}\frac{\partial}{\partial y}[P(X > x, Y > y)] \\
=& \frac{\alpha(\alpha+1)}{\sigma_1\sigma_2\delta_1\delta_2}(\frac{x-\mu_1}{\sigma_1})^{\frac{1}{\delta_1}-1}(\frac{y-\mu_2}{\sigma_2})^{\frac{1}{\delta_1}-1} \\
& \times [1 + (\frac{x-\mu_1}{\sigma_1})^{\frac{1}{\delta_1}} + (\frac{y-\mu_2}{\sigma_2})^{\frac{1}{\delta_1}}]^{-(\alpha+2)} I(x \geq \mu_1, y \geq \mu_2).
\end{aligned} \tag{4.2}$$

Note that for the above mentioned density both the marginal densities as well as the conditional densities are again members of the $P(IV)$ family with suitable choice of the parameters. In particular we may list them as follows:

- X $\sim P(IV)(\mu_1, \sigma_1, \delta_1, \alpha)$
- Y $\sim P(IV)(\mu_2, \sigma_2, \delta_2, \alpha)$
- For each possible values of $x \in (\mu_1, \infty)$, the conditional density of Y given $X = x$ will be

$$Y|X = x \sim P(IV)(\mu_2, \sigma_2^*, \delta_2, \alpha),$$

  where $\sigma_2^* = \sigma_2(1 + \frac{x-\mu_1}{\sigma_1})^{\frac{1}{\delta_1}}$.

- For each possible values of $y \in (\mu_2, \infty)$, the conditional density of

X given $Y = y$ will be

$$X|Y = y \sim P(IV)(\mu_1, \sigma_1^*, \delta_1, \alpha),$$

where $\sigma_1^* = \sigma_1(1 + \frac{y-\mu_2}{\sigma_2})^{\frac{1}{\delta_2}}$.

Also note that the correlation between X and Y is positive provided $\delta_2 - \delta_1 - \alpha > 2$, specifically

$$Corr(X, Y) = \alpha(\alpha + 1)B(\delta_2 + 2, \alpha - \delta_2)B(\delta_1 + 2, \delta_2 - \alpha - \delta_1 - 2).$$

At first we consider the situation where both $\mu_1 = \mu_2 = 0$, after replacing them by their consistent sample estimates $X_{1:n}, Y_{1:n}$ respectively where $X_{1:n} = min_{1\leq i\leq n} X_i$ and $Y_{1:n} = min_{1\leq i\leq n} Y_i$. The hidden truncated density of X given $Y \leq c$, for any positive value of c will be given by

$$f_{X|Y\leq c}(x) = \frac{f_X(x)P(Y < c|X = x)}{P(Y \leq c)} I(x \geq 0). \tag{4.3}$$

In our case for each fixed $X = x$,

$$P(Y \leq c|X = x) = 1 - [1 + \frac{\frac{c}{\sigma_2}}{1 + (\frac{x}{\sigma_1})^{\frac{1}{\delta_1}}}]^{-(\alpha+1)},$$

while

$$P(Y \leq c) = 1 - (1 + (\frac{c}{\sigma_2})^{\frac{1}{\delta_2}})^{-\alpha}.$$

So the hidden truncated density of X given $Y \leq c$ is given by

$$f^{HT}_{X|Y \leq c}(x) = \frac{\alpha x^{\frac{1}{\delta_1}-1}}{\sigma_1^{\delta_1}\delta_1}\left[\frac{(1+(\frac{x}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)}(1-[1+\frac{(\frac{c}{\sigma_2})^{\frac{1}{\delta_2}}}{1+(\frac{x}{\sigma_1})^{\frac{1}{\delta_1}}}]^{-(\alpha+1)})}{1-(1+(\frac{c}{\sigma_2})^{\frac{1}{\delta_2}})^{-\alpha}}\right] I(x \geq 0). \tag{4.4}$$

For notational simplicity we write $\frac{c}{\sigma_2} = \theta$ and $\Psi(\alpha, \theta, \delta_2) = 1 - (1 + \theta^{\frac{1}{\delta_2}})^{-\alpha}$.

So our resulting density in this case is given by

$$\begin{aligned} & f^{HT}_{X|Y \leq \theta}(x) \\ & = \frac{\alpha}{\sigma_1 \delta_1 \Psi(\alpha, \theta, \delta_2)}(\frac{x}{\sigma_1})^{\frac{1}{\delta_1}-1}[((1+\frac{x}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)} \\ & - (1+\theta^{\frac{1}{\delta_2}}+(\frac{x}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)}] I(x \geq 0). \end{aligned} \tag{4.5}$$

For graphical reference we will consider the plot of the hazard rate function which is given by

$$h_{T|Y \leq \theta}(t) = \frac{f^{HT}_{T|Y \leq \theta}(t)}{S_{T|Y \leq \theta}(t)},$$

where

$$
\begin{aligned}
& S_{T|Y\leq\theta}(t) \\
&= P_{T|Y\leq\theta}(T > t) \\
&= \int_t^{\infty} \frac{\alpha}{\sigma_1\delta_1\Psi(\alpha,\theta,\delta_2)}(\frac{x}{\sigma_1})^{\frac{1}{\delta_1}-1}[((1+\frac{u}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)} \\
&- (1+\theta^{\frac{1}{\delta_2}}+(\frac{u}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)}]du \\
&= \frac{1}{\Psi(\alpha,\theta,\delta_2)}[((1+\frac{t}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha} - (1+\theta^{\frac{1}{\delta_2}}+(\frac{t}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha}]. \qquad (4.6)
\end{aligned}
$$

In Figure (4.1) we consider for fixed choices of the parameters $\alpha = 4, \sigma_1 = 1, \delta_1 = 2, \delta_2 = 3$, the plot of the hazard rate function for different choices of the truncation parameter θ.

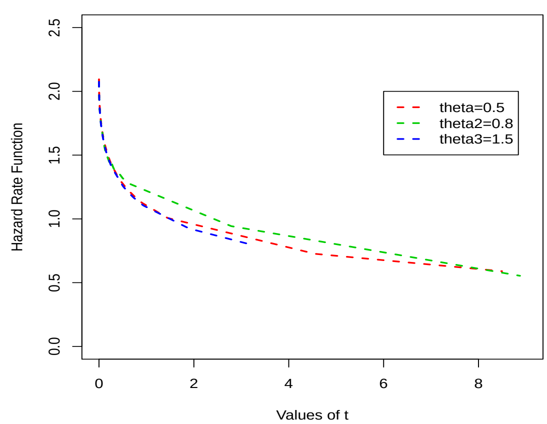


Figure 4.1: Hazard rate function for a Hidden truncated Pareto(type IV)density with $\alpha = 4$, $\sigma = 1$, $\delta_1 = 2$, $\delta_2 = 3$.

4.4 Fractional method of moments estimation

We will consider use of the fractional method of moments for estimating all the parameters of the hidden truncated bivariate $P(IV)$ model. We first consider for any $r \geq 1$,

$$
\begin{aligned}
& E[X^r] \\
&= \int_0^\infty x^r f_{X|Y\leq\theta}(x)dx \\
&= \int_0^\infty x^r (\frac{x}{\sigma_1})^{\frac{1}{\delta_1}-1} \frac{\alpha}{\sigma_1\delta_1\Psi(\alpha,\theta,\delta_2)}[((1+\frac{x}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)} \\
&- (1+\theta^{\frac{1}{\delta_2}}+(\frac{x}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)}]dx \\
&= \frac{\alpha\sigma_1^r}{\Psi(\alpha,\theta,\delta_2)}[(u-1)^{r\delta_1}(u^{-(\alpha+1)} - (u+\theta^{\frac{1}{\delta_2}})^{-(\alpha+1)})] \\
&= \frac{\alpha\sigma_1^r}{\Psi(\alpha,\theta,\delta_2)}[\int_1^\infty u^{-(\alpha+1)}(u-1)^{r\delta_1}du \\
&- \int_1^\infty (u-1)^{r\delta_1}(u+\theta^{\frac{1}{\delta_2}})^{-(\alpha+1)}du] \\
&= \frac{\alpha\sigma_1^r}{\Psi(\alpha,\theta,\delta_2)}[I_1 - I_2], \qquad (4.7)
\end{aligned}
$$

by considering

$$1+(\frac{x}{\sigma_1})^{\frac{1}{\delta_1}} = u,$$

say, where

$$
\begin{aligned}
I_1 &= \int_1^\infty u^{-(\alpha+1)}(u-1)^{r\delta_1} \\
&= \int_0^\infty t^{r\delta_1}(1+t)^{-(\alpha+1)}dt \\
&= B(r\delta_1+1, \alpha - r\delta_1), \qquad (4.8)
\end{aligned}
$$

by setting $u-1=t$ in which $B(m,n)$ is the Beta function with parameters m and n. Similarly,

$$
\begin{aligned}
I_2 &= \int_1^\infty (u-1)^{r\delta_1}(u+\theta^{\frac{1}{\delta_2}})^{-(\alpha+1)}du \\
&= (1+\theta^{\frac{1}{\delta_2}})^{r\delta_1-\alpha}B(r\delta_1+1, \alpha-r\delta_1). \qquad (4.9)
\end{aligned}
$$

Hence on substitution in (4.7) we get,

$$
E[X^r] = \frac{\alpha\sigma_1^r}{\Psi(\alpha,\theta,\delta_2)}[B(r\delta_1+1,\alpha-r\delta_1)(1-(1+\theta^{\frac{1}{\delta_2}})^{r\delta_1-\alpha})]
$$

Next for the method of moment estimation we first define the following quantities based on sample observations for a random sample of size n :

- $M_1 = \frac{1}{n-1}\sum_{i=1}^n (X_i - X_{1:n})$.
- $M_{\frac{1}{2}} = \frac{1}{n-1}\sum_{i=1}^n (X_i - X_{1:n})^{\frac{1}{2}}$.

- $M_{\frac{1}{3}} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X_{1:n})^{\frac{1}{3}}$.
- $M_{\frac{1}{4}} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X_{1:n})^{\frac{1}{4}}$.
- $M_{\frac{1}{5}} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X_{1:n})^{\frac{1}{5}}$.

Next we consider the following (after equating the sample moments with the corresponding population moments)

$$\frac{M_1}{M_{\frac{1}{2}}} = \frac{\sigma_1[B(\delta_1+1, \alpha-\delta_1)(1-(1+\theta^{\frac{1}{\delta_2}})^{\delta_1-\alpha})]}{\sigma_1^{\frac{1}{2}}[B(\frac{\delta_1}{2}+1, \alpha-\frac{\delta_1}{2})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{2}-\alpha})]}, \tag{4.10}$$

and

$$\frac{M_{\frac{1}{2}}}{M_{\frac{1}{3}}} = \frac{\sigma_1^{\frac{1}{2}}[B(\frac{\delta_1}{2}+1, \alpha-\frac{\delta_1}{2})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{2}-\alpha})]}{\sigma_1^{\frac{1}{3}}[B(\frac{\delta_1}{3}+1, \alpha-\frac{\delta_1}{3})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{3}-\alpha})]}. \tag{4.11}$$

Again

$$\frac{M_{\frac{1}{3}}}{M_{\frac{1}{4}}} = \frac{\sigma_1^{\frac{1}{3}}[B(\frac{\delta_1}{3}+1, \alpha-\frac{\delta_1}{3})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{3}-\alpha})]}{\sigma_1^{\frac{1}{4}}[B(\frac{\delta_1}{4}, \alpha-\frac{\delta_1}{4})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{4}-\alpha})]}. \tag{4.12}$$

Also

$$\frac{M_1}{M_{\frac{1}{4}}} = \frac{\sigma_1[B(\delta_1+1, \alpha-\delta_1)(1-(1+\theta^{\frac{1}{\delta_2}})^{\delta_1-\alpha})]}{\sigma_1^{\frac{1}{4}}[B(\frac{\delta_1}{4}, \alpha-\frac{\delta_1}{4})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{4}-\alpha})]}, \tag{4.13}$$

and

$$\frac{M_{\frac{1}{2}}}{M_{\frac{1}{5}}} = \frac{\sigma_1^{\frac{1}{2}}[B(\frac{\delta_1}{2}+1, \alpha-\frac{\delta_1}{2})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{2}-\alpha})]}{\sigma_1^{\frac{1}{5}}[B(\frac{\delta_1}{5}, \alpha-\frac{\delta_1}{5})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{5}-\alpha})]}. \tag{4.14}$$

So in this situation we have five equations which we can write equivalently as:

- $M_1[B(\frac{\delta_1}{2}+1,\alpha-\frac{\delta_1}{2})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{2}-\alpha})] = \sigma_1^{\frac{1}{2}} M_{\frac{1}{2}}[B(\delta_1+1,\alpha-\delta_1)(1-(1+\theta^{\frac{1}{\delta_2}})^{\delta_1-\alpha})]$.

- $M_{\frac{1}{2}}[B(\frac{\delta_1}{3}+1,\alpha-\frac{\delta_1}{3})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{3}-\alpha})] = \sigma_1^{\frac{1}{6}} M_{\frac{1}{3}}[B(\frac{\delta_1}{2}+1,\alpha-\frac{\delta_1}{2})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{2}-\alpha})]$.

- $M_{\frac{1}{3}}[B(\frac{\delta_1}{4},\alpha-\frac{\delta_1}{4})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{4}-\alpha})] = \sigma_1^{\frac{1}{12}} M_{\frac{1}{4}}[B(\frac{\delta_1}{3}+1,\alpha-\frac{\delta_1}{3})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{3}-\alpha})]$.

- $M_1[B(\frac{\delta_1}{4},\alpha-\frac{\delta_1}{4})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{4}-\alpha})] = \sigma_1^{\frac{3}{4}} M_{\frac{1}{4}}[B(\delta_1+1,\alpha-\delta_1)(1-(1+\theta^{\frac{1}{\delta_2}})^{\delta_1-\alpha})]$.

- $M_{\frac{1}{2}}[B(\frac{\delta_1}{5},\alpha-\frac{\delta_1}{5})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{5}-\alpha})] = \sigma_1^{\frac{3}{10}} M_{\frac{1}{5}}[B(\frac{\delta_1}{2}+1,\alpha-\frac{\delta_1}{2})(1-(1+\theta^{\frac{1}{\delta_2}})^{\frac{\delta_1}{2}-\alpha})]$.

4.5 Quartile method of estimation

Here we consider the quartile method of estimation. If ξ_p is the p-th order quantile ($p \in (0,1)$), then

$$P(X \leq \xi_p) = p.$$

So for our density we have

$$
\begin{aligned}
P[X \leq \xi_p] &= \frac{\alpha}{\Psi(\alpha,\theta,\delta_2)\sigma_1}\int_0^{\xi_p} f^{HT}(x)dx \\
&= \frac{\alpha}{\Psi(\alpha,\theta)\sigma_1}\int_0^{\xi_p} (\frac{x}{\sigma_1})^{\frac{1}{\delta_1}-1}[((1+\frac{x}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)} \\
&- (1+\theta^{\frac{1}{\delta_2}}+(\frac{x}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)}]dx \\
&= \frac{1}{\Psi(\alpha,\theta,\delta_2)[1-(1+\frac{\xi_p}{\sigma_1})^{-\alpha}]-[(1+\theta)^{-\alpha}-(1+\theta+\frac{\xi_p}{\sigma_1})^{-\alpha}]},
\end{aligned}
$$

So that we have

$$
\frac{1}{\Psi(\alpha,\theta,\delta_2)[1-(1+\frac{\xi_p}{\sigma_1})^{-\alpha}]-[(1+\theta)^{-\alpha}-(1+\theta+\frac{\xi_p}{\sigma_1})^{-\alpha}]} = p. \quad (4.15)
$$

Equivalently we can write

$$
[1-(1+\frac{\xi_p}{\sigma_1})^{-\alpha}]-[(1+\theta)^{-\alpha}-(1+\theta+\frac{\xi_p}{\sigma_1})^{-\alpha}] = p\Psi(\alpha,\theta,\delta_2). \quad (4.16)
$$

Our estimates are then obtained by equating 5 sample quantiles, denoted by $\hat{\xi}_p$ to the corresponding population quantiles. In particular by considering successively $p = \frac{1}{2}, \frac{3}{4}, \frac{1}{2}, \frac{1}{5}, \frac{1}{6}$, we have the following five equations:

$$
[1-(1+\frac{\hat{\xi}_{\frac{1}{2}}}{\sigma_1})^{-\alpha}]-[(1+\theta)^{-\alpha}-(1+\theta+\frac{\hat{\xi}_{\frac{1}{2}}}{\sigma_1})^{-\alpha}] = \frac{\Psi(\alpha,\theta,\delta_2)}{2}. \quad (4.17)
$$

$$[1-(1+\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{-\alpha}]-[(1+\theta)^{-\alpha}-(1+\theta+\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{-\alpha}]=\frac{3\Psi(\alpha,\theta,\delta_2)}{4}. \quad (4.18)$$

$$[1-(1+\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{-\alpha}]-[(1+\theta)^{-\alpha}-(1+\theta+\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{-\alpha}]=\frac{\Psi(\alpha,\theta,\delta_2)}{4}. \quad (4.19)$$

$$[1-(1+\frac{\hat{\xi}_{\frac{1}{5}}}{\sigma_1})^{-\alpha}]-[(1+\theta)^{-\alpha}-(1+\theta+\frac{\hat{\xi}_{\frac{1}{5}}}{\sigma_1})^{-\alpha}]=\frac{\Psi(\alpha,\theta,\delta_2)}{5}. \quad (4.20)$$

$$[1-(1+\frac{\hat{\xi}_{\frac{1}{6}}}{\sigma_1})^{-\alpha}]-[(1+\theta)^{-\alpha}-(1+\theta+\frac{\hat{\xi}_{\frac{1}{6}}}{\sigma_1})^{-\alpha}]=\frac{\Psi(\alpha,\theta,\delta_2)}{6}. \quad (4.21)$$

So from equation (4.19) and equation (4.17) we get,

$$\begin{aligned}1-2(1+(\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha}-(1+\theta)^{-\alpha}+(1+(\frac{\hat{\xi}_{\frac{1}{2}}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha}\\+2(1+(\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{\frac{1}{\delta_1}}+\theta)^{-\alpha}+2(1+(\frac{\xi_{\frac{1}{2}}}{\sigma_1})^{\frac{1}{\delta_1}}+\theta)^{-\alpha}=0\end{aligned}$$

Also from equation (4.18) and equation (4.17) we get,

$$\begin{aligned}1-3(1+(\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha}-(1+\theta)^{-\alpha}-2(1+(\frac{\hat{\xi}_{\frac{1}{2}}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha}\\+3(1+(\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{\frac{1}{\delta_1}}+\theta)^{-\alpha}-2(1+(\frac{\hat{\xi}_{\frac{1}{2}}}{\sigma_1})^{\frac{1}{\delta_1}}+\theta)^{-\alpha}=0.\end{aligned}$$

Also we have from equation (4.20) and equation (4.18),

$$11 - 15(1 + (\frac{\hat{\xi}_{\frac{1}{5}}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha} - 11(1+\theta)^{-\alpha} + 4(1 + (\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha}$$
$$- 4(1 + (\frac{\hat{\xi}_{\frac{3}{4}}}{\sigma_1})^{\frac{1}{\delta_1}} + \theta)^{-\alpha} + 15(1 + (\frac{\hat{\xi}_{\frac{1}{5}}}{\sigma_1})^{\frac{1}{\delta_1}} + \theta)^{-\alpha} = 0.$$

Similarly from equation (4.21) and equation (4.20) we get,

$$1 - 6(1 + (\frac{\hat{\xi}_{\frac{1}{6}}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha} - (1+\theta)^{-\alpha} + 5(1 + (\frac{\hat{\xi}_{\frac{1}{5}}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha}$$
$$+ 6(1 + \frac{\hat{\xi}_{\frac{1}{6}}}{\sigma_1} + \theta)^{-\alpha} - 5(1 + \frac{\hat{\xi}_{\frac{1}{5}}}{\sigma_1} + \theta)^{-\alpha} = 0.$$

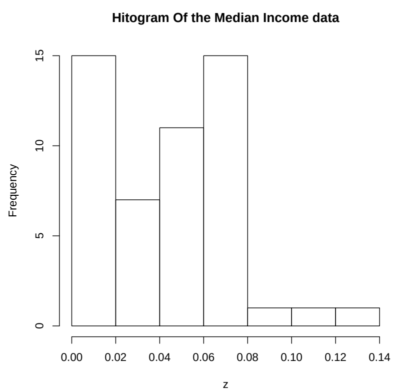
Finally we have from equation (4.21) and equation (4.19),

$$1 - 3(1 + (\frac{\hat{\xi}_{\frac{1}{6}}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha} - (1+\theta)^{-\alpha} + 2(1 + (\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha}$$
$$+ 3(1 + (\frac{\hat{\xi}_{\frac{1}{6}}}{\sigma_1})^{\frac{1}{\delta_1}} + \theta)^{-\alpha} - 2(1 + (\frac{\hat{\xi}_{\frac{1}{4}}}{\sigma_1})^{\frac{1}{\delta_1}} + \theta)^{-\alpha} = 0.$$

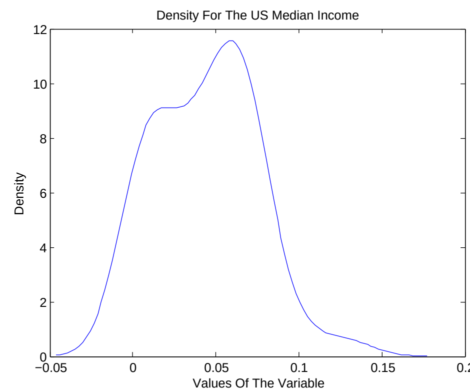
4.6 Application of the hidden truncated $P(IV)$ model to a real life data set

We consider the US income data available in the form of 2 year median income (data source: US Census Bureau, 2006-2009) and in particular median income for 2006-2007 and 2008-2009 across all the 50 states in US and want to study whether the change (in percentage) in the median income can be explained by a hidden truncated $P(IV)$ model.

We argue at this point that the data is subject to hidden truncation because in those median income figures there might be instances of unreported income and sometimes an individual's income from other sources is not properly reported.



(a) Histogram of US Median Income



(b) Density of US Median Income

Figure 4.2: Histogram and Density plot for the US Median Income.

First, let us consider the histogram and density plot of the data (displayed in Figure (4.2)). In this case the density plot has been drawn by smoothing the histogram. Based on the available data, we get the following estimates of the parameters for a hidden truncated bivariate $P(IV)$ model:

- Estimation based on fractional method of moments:

$$\hat{\sigma}_1 = 1.942376, \hat{\theta} = 2.518713, \hat{\alpha} = 3.384127, \hat{\delta}_1 = 2.545628, \hat{\delta}_2 = 3.14.$$

- Estimation based on quartile method:

$$\hat{\sigma}_1 = 1.915384, \hat{\theta} = 2.529446, \hat{\alpha} = 3.416816, \hat{\delta}_1 = 2.591568, \hat{\delta}_2 = 3.16.$$

Moreover, in this case the standard Kolmogorv-Smirnov goodness of fit test statistic tells us that indeed the fit is good.

So the nature of our data can well be explained by a hidden truncated bivariate $P(IV)$ distribution with the following choice of the parameters (approximately):

$$\sigma_1 = 1.9, \theta = 2.5, \alpha = 3.4, \delta_1 = 2.5, \delta_2 = 3.1.$$

Note that the estimated value of θ, which is far from zero, does indicates that the data has been subjected to hidden truncation.

4.7 Estimation of the parameters using a simulation study

4.7.1 Sample generation from the truncated density

First we draw a random sample of size n (we consider $n = 50, 100, 200$) from the hidden truncated bivariate $P(IV)$ distribution for the following particular choices of the parameters:

$$\alpha = 3, \sigma_1 = 2, \sigma_2 = 2, \delta_1 = 2, \delta_2 = 3, \mu_1 = 0, \mu_2 = 0, c = 2.$$

So that $\theta = \frac{c-\mu_2}{\sigma_2} = 1.5$ and $\Psi(\alpha, \theta, \delta_2) = 1 - (1 + \theta^{\frac{1}{\delta_2}})^{-\alpha} = 0.8605353$.

So that our density reduces to

$$f^{HT}_{X|Y\leq 1.5}(x) = \frac{3}{4}(\frac{x}{2})^{\frac{1}{2}-1}[\frac{(1+(\frac{x}{2})^{\frac{1}{2}})^{-4}(1-[1+\frac{1.5^{\frac{1}{3}}}{1+(\frac{x}{2})^{\frac{1}{2}}}]^{-4})}{0.8605353}]I(x \geq 0). \tag{4.22}$$

The estimates of all the parameters using both the quartile and fractional moment estimation methods are displayed in Table (4.1).

Estimates of the p arameters						
	n	$\hat{\sigma_1}$	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\delta_1}$	$\hat{\delta_2}$
Quartile method	50	2.914573	1.767512	2.191324	1.889380	3.096597
	100	1.879070	1.461497	2.899593	1.908241	3.026104
	200	1.915689	1.467915	2.871755	1.970107	3.028901
Fractional Method Of Moments	50	1.655359	1.654185	2.368763	1.904561	3.143666
	100	1.997621	1.757678	2.859842	1.968523	2.976938
	200	1.979760	1.649014	2.705301	2.040245	3.127698

Table 4.1: Estimates of the parameters using the quartile and the fractional method of moments.

4.7.2 Comment on the simulation study

From the simulation study for various choices of sample sizes (n=50, 100, 200), we observe is that for sample size n=50, the estimates of the parameters using the quartile method are not good when compared with those obtained using the fractional method of moments (except for the estimated value of δ_2. One can also observe that with the increase in the sample size precise estimation of all the parameters under both estimation procedures has not been achieved simultaneously. For instance when the sample size is either 100 or 200, the estimated value of σ_1 under the fractional method of moments is reasonably good as compared to quartile method, while for the parameter θ, the scenario is just the opposite. Moreover the estimated values of σ_1 are far away from the true value under the quartile method. Also the estimated values of θ for different sample sizes are very close to the true value. So overall we can not make a general recommendation. In terms of the relative performance of the two estimation strategies, one is not always better than the other. Furthermore a valid question that is worth mentioning, is why the popular maximum likelihood estimation procedure has not been used here. The answer to this question is that the results using maximum likelihood were not that promising. One reasonable explanation could be that for the complicated model that we have, it is really difficult to get some idea about the dependence structure among the parameters involved in our model thereby implying that we do not have much information about the likelihood surface. So we do not know exactly what are the optimum choices of the parameters for which the likelihood function will attain it's maximum. A more extensive study is required in this direction to find the real cause.

4.8 Asymptotic distribution of the smallest order statistic

We have as before our density which is of the form

$$\begin{aligned} & f^{HT}_{X|Y\leq\theta}(x) \\ & = \frac{\alpha}{\sigma_1\delta_1\Psi(\alpha,\theta,\delta_2)}(\frac{x}{\sigma_1})^{\frac{1}{\delta_1}-1}[((1+\frac{x}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)} \\ & - (1+\theta^{\frac{1}{\delta_2}}+(\frac{x}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)}]I(x\geq 0). \end{aligned}$$

So the corresponding distribution function is given by

$$F(x) = 1 - \frac{((1+\frac{x}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha} - (1+\theta^{\frac{1}{\delta_2}}+(\frac{x}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha}}{\Psi(\alpha,\theta,\delta_2)}, x\geq 0.$$

So for a random sample of size n from the above density the distribution function of the smallest order statistic ($X_{1:n} = min_{1\leq i\leq n}X_i$) will be

$$F(x_{1:n}) = 1 - [\frac{((1+\frac{x_{1:n}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha} - (1+\theta^{\frac{1}{\delta_2}}+(\frac{x_{1:n}}{\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha}}{\Psi(\alpha,\theta,\delta_2)}]^n, x\geq 0.$$

Let us consider the following

$$P(X_{1:n} > \frac{x_{1:n}}{n}) = [\frac{((1+\frac{x_{1:n}}{n\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha} - (1+\theta^{\frac{1}{\delta_2}}-(\frac{x_{1:n}}{n\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha}}{\Psi(\alpha,\theta,\delta_2)}]^n, \quad (4.23)$$

using the expression for the distribution function as given earlier. Next consider the quantity

$$((1+\frac{x_{1:n}}{n\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha} - (1+\theta^{\frac{1}{\delta_2}} - (\frac{x_{1:n}}{n\sigma_1})^{\frac{1}{\delta_1}})^{-\alpha} \tag{4.24}$$

$$= [1-\alpha(\frac{x_{1:n}}{n\sigma_1})^{\frac{1}{\delta_1}} + \frac{\alpha(\alpha+1)}{2}(\frac{x_{1:n}}{n\sigma_1})^{\frac{2}{\delta_1}} - \cdots]$$

$$-[(1+\theta^{\frac{1}{\delta_2}})^{-\alpha} - \alpha(1+\theta^{\frac{1}{\delta_2}})^{-\alpha+1}(\frac{x_{1:n}}{n\sigma_1})^{\frac{1}{\delta_1}}$$

$$+ \frac{\alpha(\alpha+1)}{2}(1+\theta^{\frac{1}{\delta_2}})^{-\alpha+2}(\frac{x_{1:n}}{n\sigma_1})^{\frac{2}{\delta_1}} - \cdots]$$

$$= 1-(1+\theta^{\frac{1}{\delta_2}})^{-\alpha} - \alpha(1-(1+\theta^{\frac{1}{\delta_2}})^{-\alpha+1})(\frac{x_{1:n}}{n\sigma_1})^{\frac{1}{\delta_1}} + \mathrm{o}(n^{-2})$$

$$= \Psi(\alpha,\theta,\delta_2) - \alpha(1-(1+\theta^{\frac{1}{\delta_2}})^{-\alpha+1})(\frac{x_{1:n}}{n\sigma_1})^{\frac{1}{\delta_1}} + \mathrm{o}(n^{-2}). \tag{4.25}$$

So on substitution in equation (4.24), we get

$$P(X_{1:n} > \frac{x_{1:n}}{n}) = [1 - \frac{\alpha(1-(1+\theta^{\frac{1}{\delta_2}})^{-\alpha+1})}{\sigma_1^{\frac{1}{\delta_1}}\Psi(\alpha,\theta,\delta_2)}(\frac{x_{1:n}}{n})^{\frac{1}{\delta_1}} + \mathrm{o}(n^{-2})]^n,$$

provided

$$0 < \delta_1 < 2.$$

Hence we can write

$$\begin{aligned}\lim_{n\to\infty} P(X_{1:n} > \frac{x_{1:n}}{n}) &= \lim_{n\to\infty}[1 - \frac{\alpha(1-(1+\theta^{\frac{1}{\delta_2}})^{-\alpha+1})}{\sigma^{\frac{1}{\delta_1}1}\Psi(\alpha,\theta,\delta_2)}(\frac{x_{1:n}}{n})^{\frac{1}{\delta_1}} + \mathrm{o}(n^{-2})]^n \\ &= \exp[-Bx_{1:n}],\ x_{1:n} \geq 0,\end{aligned}$$

where $B = \frac{\alpha(1-(1+\theta^{\frac{1}{\delta_2}})^{-\alpha+1})}{\sigma_1^{\delta_1}\Psi(\alpha,\theta,\delta_2)}$. So the asymptotic distribution of $X_{1:n}$ is exponential with the intensity parameter$=\frac{\alpha(1-(1+\theta^{\frac{1}{\delta_2}})^{-\alpha+1})}{\sigma_1^{\delta_1}\Psi(\alpha,\theta,\delta_2)}$, when the samples are drawn from a density of the form in (4.5).

4.9 Bayesian inference for the hidden truncated $P(IV)$ model

For a random sample of size n from the hidden truncated bivariate $P(IV)$ distribution our likelihood function takes the following form:

$$\begin{aligned}
&L(\alpha, \sigma_1, \theta, \delta_1, \delta_2) \\
&= \prod_{i=1}^{n} \frac{\alpha}{\sigma_1 \delta_1 \Psi(\alpha, \theta, \delta_2)} (\frac{x_i}{\sigma_1})^{\frac{1}{\delta_1}-1} [((1 + \frac{x_i}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)} \\
&- (1 + \theta^{\frac{1}{\delta_2}} + (\frac{x_i}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)}] \\
&= (\frac{\alpha}{\sigma_1 \delta_1 \Psi(\alpha, \theta, \delta_2)})^n \prod_{i=1}^{n} (\frac{x_i}{\sigma_1})^{\frac{1}{\delta_1}-1} [((1 + \frac{x_i}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)} \\
&- (1 + \theta^{\frac{1}{\delta_2}} + (\frac{x_i}{\sigma_1})^{\frac{1}{\delta_1}})^{-(\alpha+1)}].
\end{aligned}$$

4.9.1 Sample and prior information

In the beginning we consider the following choice of independent priors for the parameters in the model:

- Prior for α

  $f(\alpha) = \frac{1}{(1+\alpha)^2} I(\alpha > 0)$.

- Prior for σ_1

  $f(\sigma_1) = \frac{1}{(1+\sigma_1)^2} I(\sigma_1 > 0)$.

- Prior for θ

  $f(\theta) = \frac{1}{(1+\theta)^2} I(\theta > 0).$

- Prior for δ_1

  $f(\delta_1) = \frac{1}{(1+\delta_1)^2} I(\delta_1 > 0).$

- Prior for δ_2

  $f(\delta_2) = \frac{1}{(1+\delta_2)^2} I(\delta_2 > 0).$

So the joint posterior density will be given by

$$
\begin{aligned}
&f(\alpha, \sigma_1, \delta_1, \delta_2, \theta | \underline{X} = \underline{x}) \\
&= A^{-1} \prod_{i=1}^{n} \Big(\frac{1}{(1+\alpha)(1+\sigma_1)(1+\theta)(1+\delta_1)(1+\delta_2)}\Big)^2 \frac{\alpha}{\sigma_1 \delta_1 \Psi(\alpha, \theta, \delta_2)} \Big(\frac{x_i}{\sigma_1}\Big)^{\frac{1}{\delta_1}-1} \\
&\qquad \times \Big[\Big(\big(1 + \frac{x_i}{\sigma_1}\big)^{\frac{1}{\delta_1}}\Big)^{-(\alpha+1)} - \Big(1 + \theta^{\frac{1}{\delta_2}} + \big(\frac{x_i}{\sigma_1}\big)^{\frac{1}{\delta_1}}\Big)^{-(\alpha+1)}\Big] \\
&\qquad \times I(\alpha > 0, \sigma_1 > 0, \theta > 0, \delta_1 > 0, \delta_2 > 0),
\end{aligned}
\tag{4.26}
$$

where A is the normalizing constant which is given by

$$
\begin{aligned}
A = &\int_{R_1} \cdots \int \prod_{i=1}^{n} \Big(\frac{1}{(1+\alpha)(1+\sigma_1)(1+\theta)(1+\delta_1)(1+\delta_2)}\Big)^2 \\
&\frac{\alpha}{\sigma_1 \delta_1 \Psi(\alpha, \theta, \delta_2)} \Big(\frac{x_i}{\sigma_1}\Big)^{\frac{1}{\delta_1}-1} \\
&\times \Big[\Big(\big(1 + \frac{x_i}{\sigma_1}\big)^{\frac{1}{\delta_1}}\Big)^{-(\alpha+1)} - \Big(1 + \theta^{\frac{1}{\delta_2}} + \big(\frac{x_i}{\sigma_1}\big)^{\frac{1}{\delta_1}}\Big)^{-(\alpha+1)}\Big] \, d\alpha \, d\sigma_1 \, d\theta \, d\delta_1 \, d\delta_2,
\end{aligned}
$$

where

$$R_1 = (\alpha > 0, \sigma_1 > 0, \theta > 0, \delta_1 > 0, \delta_2 > 0).$$

So that the marginal posterior density of α is given by

$$f(\alpha|\underline{X} = \underline{x}) = A^{-1} \int_{R_{11}} \cdots \int f(\alpha, \sigma_1, \delta_1, \delta_2, \theta|\underline{X} = \underline{x}) d\sigma_1\, d\theta\, d\delta_1\, d\delta_2, \tag{4.27}$$

where

$$R_{11} = (\sigma_1 > 0, \theta > 0, \delta_1 > 0, \delta_2 > 0).$$

Again the marginal posterior density of σ_1 is given by

$$f(\sigma_1|\underline{X} = \underline{x}) = A^{-1} \int_{R_{12}} \cdots \int f(\alpha, \sigma_1, \delta_1, \delta_2, \theta|\underline{X} = \underline{x}) d\alpha\, d\theta\, d\delta_1\, d\delta_2, \tag{4.28}$$

where

$$R_{12} = (\alpha > 0, \theta > 0, \delta_1 > 0, \delta_2 > 0).$$

Also the marginal posterior density of δ_1 is given by

$$f(\delta_1|\underline{X} = \underline{x}) = A^{-1} \int_{R_{13}} \cdots \int f(\alpha, \sigma_1, \delta_1, \delta_2, \theta|\underline{X} = \underline{x}) d\alpha\, d\theta\, d\sigma_1\, d\delta_2, \tag{4.29}$$

where

$$R_{13} = (\alpha > 0, \theta > 0, \sigma_1 > 0, \delta_2 > 0).$$

The marginal posterior density of δ_2 is given by

$$f(\delta_2|\underline{X} = \underline{x}) = A^{-1} \int_{R_{14}} \cdots \int f(\alpha, \sigma_1, \delta_1, \delta_2, \theta|\underline{X} = \underline{x}) d\alpha \, d\theta d\delta_1 \, d\sigma_1, \tag{4.30}$$

where

$$R_{14} = (\alpha > 0, \theta > 0, \delta_1 > 0, \sigma_1 > 0).$$

Finally the marginal posterior density of δ_1 is given by

$$f(\theta|\underline{X} = \underline{x}) = A^{-1} \int_{R_{15}} \cdots \int f(\alpha, \sigma_1, \delta_1, \delta_2, \theta|\underline{X} = \underline{x}) d\alpha \, d\delta_1 d\sigma_1 \, d\delta_2, \tag{4.31}$$

where

$$R_{15} = (\alpha > 0, \delta_1 > 0, \sigma_1 > 0, \delta_2 > 0).$$

4.9.2 Posterior simulation study

With the following choice of priors we first draw samples of size $n = 100, 200$ from a hidden truncated density for a particular choice ($\alpha = 2, \sigma_1 = 6, \theta = 1, \delta_1 = 1, \delta_2 = 2$) of the parameters. However in our case we utilize a Metropolis-Hastings algorithm which is a general term for a family of Markov chain simulation methods that are useful for drawing samples from Bayesian posterior distributions. For the jumping distributions we consider gamma distributions but with different shape and scale parameters. The posterior analysis is based on the posterior modes and also the posterior means for each of the five parameters. Below we provide various choices for the starting distribution, the choices of the parameters of the jumping distribution along with posterior modes:

Initial choices of the parameters: $\alpha = 1.2, \sigma_1 = 0.65, \theta = 0.81, \delta_1 = 0.82, \delta_2 = 3.29$.

Jumping distribution for the parameters:

- $\alpha \sim \Gamma(5.5, 0.67)$.
- $\sigma_1 \sim \Gamma(2.6, 1.59)$.
- $\theta \sim \Gamma(1.9, 0.52)$.
- $\delta_1 \sim \Gamma(2.05, 0.72)$.
- $\delta_2 \sim \Gamma(1.91, 3.72)$.

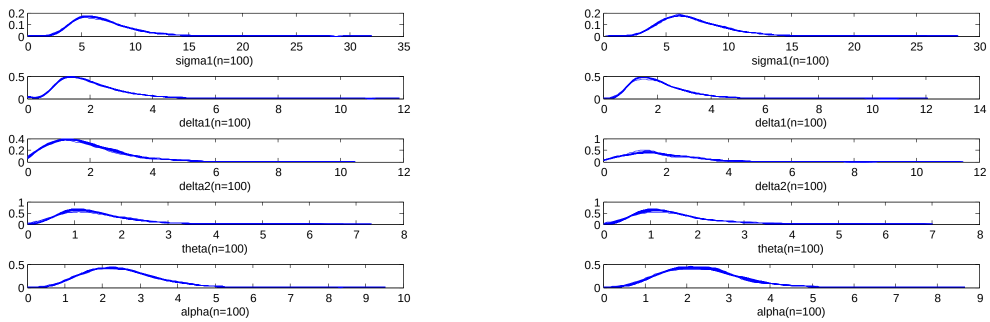
For the graphical reference we consider

n	Mode(α)	Mode(σ_1)	Mode(θ)	Mode(δ_1)	Mode(δ_2)
100	2.0868	5.9482	1.0049	1.1673	1.8941
200	2.0127	6.0709	1.1118	1.0231	1.9536

Table 4.2: Bayesian estimates of the parameters using the posterior mode.

n	Mean(α)	Mean(σ_1)	Mean(θ)	Mean(δ_1)	Mean(δ_2)
100	2.0723	5.9317	1.0049	1.0974	1.8829
200	2.0472	6.0245	1.1118	1.0187	1.9734

Table 4.3: Bayesian estimates of the parameters using the posterior mean.



(a) Posterior Density of the parameters for $n = 100$. (b) Posterior Density of the parameters for $n = 200$.

Figure 4.3: Posterior density for all the parameters for different choices of the sample size.

4.10 Bayesian analysis using dependent priors for the shape and scale parameter with an independent prior for the truncation parameter

In this case we at first consider $\tau = \frac{1}{\sigma_1}$, where τ is the precision parameter then our density is given by

$$
\begin{aligned}
& f^{HT}_{X|Y\leq\theta}(x) \\
&= \frac{\alpha\tau}{\delta_1\Psi(\alpha,\theta,\delta_2)}(x\tau)^{\frac{1}{\delta_1}-1}[((1+x\tau)^{\frac{1}{\delta_1}})^{-(\alpha+1)} \\
&- (1+\theta^{\frac{1}{\delta_2}}+(x\tau)^{\frac{1}{\delta_1}})^{-(\alpha+1)}]I(x\geq 0). \qquad (4.32)
\end{aligned}
$$

Next we set $\theta^{\frac{1}{\delta_2}} = \theta_1$, so that our density reduces to

$$\begin{aligned} & f^{HT}_{X|Y \le \theta_1}(x) \\ & = \frac{\alpha\tau}{\delta_1 \Psi(\alpha, \theta_1)} (x\tau)^{\frac{1}{\delta_1}-1} [((1 + x\tau)^{\frac{1}{\delta_1}})^{-(\alpha+1)} \\ & - (1 + \theta_1 + (x\tau)^{\frac{1}{\delta_1}})^{-(\alpha+1)}] I(x \ge 0), \end{aligned} \tag{4.33}$$

where $\Psi(\alpha, \theta_1) = 1 - (1 + \theta_1)^{-\alpha}$. So the likelihood function is given by

$$\begin{aligned} & L(\alpha, \tau, \delta_1, \theta_1) = [\frac{\alpha\tau}{\delta_1 \Psi(\alpha, \theta_1)}]^n \prod_{i=1}^{n} [((1 + x_i\tau)^{\frac{1}{\delta_1}})^{-(\alpha+1)} \\ & - (1 + \theta_1 + (x_i\tau)^{\frac{1}{\delta_1}})^{-(\alpha+1)}] \end{aligned} \tag{4.34}$$

Note that in this case our parameter space

$$\Omega = (\alpha > 0, \tau > 0, \delta_1 > 0, \theta_1 > 0).$$

Next we consider the following choice of priors:

- For each $(\tau, \delta_1) \in \mathbb{R}^+$, the conditional density of α given τ, δ_1 is exponential with shape paramter=$\xi_1(\tau, \delta_1)$,

  and intensity parameter=$\lambda_1(\tau, \delta_1)$, i.e.,

  $$f(\alpha|\tau, \delta_1) \propto \alpha^{\xi_1(\tau,\delta_1)-1} \exp(-\alpha\lambda_1(\tau, \delta_1)) I(\alpha > 0).$$

- For each $(\alpha, \delta_1) \in \mathbb{R}^+$, the conditional density of τ given α, δ_1 is exponential with shape paramter=$\xi_2(\alpha, \delta_1)$,

and intensity parameter=$\lambda_2(\alpha, \delta_1)$, i.e.,

$$f(\tau|\alpha, \delta_1) \propto \tau^{\xi_2(\alpha,\delta_1)-1} \exp(-\tau\lambda_2(\alpha, \delta_1))I(\alpha > 0).$$

- For each $(\alpha, \tau) \in \mathbb{R}^+$, the conditional density of δ_1 given α, τ is exponential with shape paramter=$\xi_3(\alpha, \tau)$,

  and intensity parameter=$\lambda_3(\alpha, \tau)$, i.e.,

$$f(\delta_1|\alpha, \tau) \propto \delta_1^{\xi_3(\alpha,\tau)-1} \exp(-\delta_1\lambda_3(\alpha, \tau))I(\alpha > 0).$$

- While for θ_1 we consider an non-informative prior (mildly informative) which is independent of (α, τ, δ_1) and it is of the form

$$f(\theta_1) = \frac{1}{(1+\theta_1)^2} I(\theta_1 \geq 0).$$

Since all the conditional distributions are members of the exponential family, so we can write their joint density as

$$
\begin{aligned}
f(\alpha, \tau, \delta_1) &\propto \exp[1 - a_{11}\alpha + a_{12}\log(\alpha) - \tau(a_{13} + (a_{14} - a_{15})\alpha) \\
&+ \log\tau(a_{16} + (a_{17} - a_{18})\log\alpha) - \delta_1(a_{19} - a_{11} + (a_{20} - a_{21})\alpha + a_{22}\log\alpha \\
&+ (a_{23} - a_{24}\alpha + a_{25}\log\alpha)\log\tau) + ((a_{26} - a_{27}\alpha + a_{28}\log\alpha) \\
&- \tau(a_{29} + a_{30}\log\alpha - a_{31}\alpha) \\
&+ (a_{32} + a_{33}\log\alpha - a_{34}\alpha)\log\tau)\log\delta_1]I(\alpha \geq 0, \tau \geq 0, \delta_1 \geq 0).
\end{aligned}
$$

So our joint posterior distribution will be given by

$$
\begin{aligned}
&\text{posterior} \propto L(\alpha, \tau, \delta_1, \theta_1) f(\alpha, \tau, \delta_1) f(\theta_1) \\
&\propto [\frac{\alpha\tau}{\delta_1 \Psi(\alpha, \theta_1)}]^n \prod_{i=1}^{n} [((1 + x_i\tau)^{\frac{1}{\delta_1}})^{-(\alpha+1)} - (1 + \theta_1 + (x_i\tau)^{\frac{1}{\delta_1}})^{-(\alpha+1)}] \\
&\times \exp[1 - a_{11}\alpha + a_{12}\log(\alpha)\alpha) \\
&- \tau(a_{13} + (a_{14} - a_{15}) + \log\tau(a_{16} + (a_{17} - a_{18})\log\alpha) \\
&- \delta_1(a_{19} - a_{11} + (a_{20} - a_{21})\alpha + a_{22}\log\alpha \\
&+ (a_{23} - a_{24}\alpha + a_{25}\log\alpha)\log\tau) + ((a_{26} - a_{27}\alpha + a_{28}\log\alpha) \\
&- \tau(a_{29} + a_{30}\log\alpha - a_{31}\alpha) + (a_{32} + a_{33}\log\alpha - a_{34}\alpha)\log\tau)\log\delta_1] \\
&\times \frac{1}{(1 + \theta_1)^2} I(\alpha \geq 0, \tau \geq 0, \delta_1 \geq 0, \theta_1 \geq 0).
\end{aligned}
$$

4.10.1 Posterior simulation study

We consider a random sample of size ($n = 100, 200$) from our density and we run a multiple sequence of chains for the MCMC algorithm. However for the proposal densities for the parameters we consider the following:

True value of the parameters: $\alpha = 2, \tau = 2, \theta = 0.50, \delta_1 = 0.50$.

Initial choices of the parameters: $\alpha = 1.27, \tau = 0.65, \theta = 0.81, \delta_1 = 0.62$.

Jumping distribution for the parameters:

- $\alpha \sim \Gamma(5.5, 0.67)$.
- $\tau \sim \Gamma(2.6, 1.59)$.
- $\theta \sim \Gamma(2.1, 0.55)$.
- $\delta_1 \sim \Gamma(2.01, 0.72)$.

n	Mean (α)	Mean (τ)	Mean (θ)	Mean (δ_1)
100	2.0681	1.8809	0.4699	0.5405
200	2.0258	1.9461	0.4817	0.5145

Table 4.4: Bayesian estimates of the parameters using the posterior mean.

n	Mode(α)	Mode (τ)	Mode (θ)	Mode (δ_1)
100	2.1029	1.9214	0.4567	0.5218
200	2.0456	1.9348	0.4785	0.5089

Table 4.5: Bayesian estimates of the parameters using the posterior mode.

4.11 Comment on the posterior simulation study

For the bivariate hidden truncation $P(IV)$ model, we observe that with a sample size of 200, the estimates of the parameters based on the Bayesian posterior modes and the posterior means are quite good. However when the sample size is 100, both the posterior modes and the posterior means for all the parameters are quite far away from the true value of the parameter. More informative priors (for example a proper prior, or a at prior for the index parameter α only) might result in a substantial amount of improvement in our posterior mean as well as posterior modal values. From our simulation study, we can not differentiate which one among the posterior means or the posterior modes are uniformly better than the other. From the output in Table(4.4) and in Table(4.5), it may be said that one can use either of them.

4.12 Concluding remarks

Precise inference under both the classical and Bayesian paradigm for the parameters of hidden truncation models is an area which still is in it's infancy stage. One obvious factor which played a significant role in developing the subject matter of this chapter is that since $P(II)$ model is nested in $P(IV)$ model, the majority of the results which we have obtained in chapter 3 could well be found using the more general setting by using the available results for the hidden truncated $P(IV)$ model and also $P(IV)$ results can be used to predict the behavior of other Pareto-like distributions under the hidden truncated paradigm since the $P(IV)$ model is the most general model. Since the hidden truncated $P(IV)$ model does not belong to the exponential families of densities, essentially no reduction in complexity of the data can be obtained by invoking sufficiency arguments. Moreover under the classical approach it has been observed that the performance of the maximum likelihood estimation procedure for the $P(IV)$ model is really poor to the extent that we do not report the output. The other two estimation procedures have been found to be effective in this case and at least to the best of author's knowledge, these are the only two procedures that can be recommended under the classical approach. The development of inference procedures for Pareto distributions and their close relatives has been predictably uneven. Also interval estimation for Pareto populations has not been extensively investigated. A lack of convenient pivotal quantities hampers efforts in the case of more general Pareto distributions. Moreover the same situation occurs in the setting of testing parametric hypotheses. However in this context, it is to be noted that testing for the truncation parameter under the hidden truncated bivariate $P(II)$ model has been addressed in the previous chapter. Inference under the Bayesian framework as mentioned earlier is severely restricted in such settings. A study under the Bayesian set up has been carried out and is reported here by both considering the independence assumption on the choice of prior distributions of the parameters and also by invoking a dependent prior set-up. The results based on a small simulation study are quite encouraging in the sense that the Bayesian estimates of the parameters are quite good under both the set-ups.

Chapter 5

Hidden truncation for multivariate Pareto data

5.1 Introduction

As in the case of univariate Pareto distributions, mathematical simplicity and tractability have generated a lot of interest in the theory and applications of multivariate Pareto distributions especially with the advent of super efficient algorithms for generating samples from multivariate populations. We will discuss k-dimensional distributions which qualify as being multivariate $P(II)$ and $P(IV)$ distributions by virtue of having $P(II)$ and $P(IV)$ marginals respectively. Sometimes conditional and other related distributions are Paretian, but this phenomenon is not very frequent. The first author to study systematically k-dimensional Pareto distribution was Mardia (1962). A detailed discussion of multivariate Pareto distributions can be found in the monograph by Arnold (1983). Numerous papers dealing with bivariate and multivariate Pareto distributions have subsequently appeared in the literature. In addition, a good reference in this context is the book on multivariate continuous distributions by Kotz, Balakrishnan and Johnson (2006).

However discussions on inferential aspects of various forms of multivariate Pareto distributions have been somewhat restricted. A scarcity of multivariate data, a lack of appropriate models which predict multivariate Paretian behavior, and the relative newness of the introduction of the majority of multivariate Pareto distributions, have all combined together to restrain the development of inferential techniques. Inferential techniques which capitalize on the multivariate structure of the data are not well developed.

The pioneering work of Mardia (1962) on multivariate Pareto distribution of the first kind was followed by Arnold's (1983) discussion on estimation procedures for multivariate Pareto distributions of the second, third and fourth kinds and also on the Bayesian estimation for multivariate Pareto distribution of the second kind. The contributions of Targhetta (1979), Tajvedi (1996) and Hanagal (1996) are also noteworthy to mention. In this chapter we will focus on the construction of a multivariate hidden truncated model for k-variate $P(II)$ and $P(IV)$ distributions. Specifically we will address the following two types of hidden truncation :(1) Single variable truncation from above and (2) k_1-variable (where $k_1 \leq k$, $k_1 \geq 2$) truncation from above. In particular, in the present chapter, we will focus on inferential aspects for a hidden truncated trivariate $P(II)$ distribution when one of the co-variables is truncated from above. However we will in the first few sections of this chapter consider the hidden truncation paradigm for a multivariate $P(II)$ distribution with attention directed towards single variable truncation from above. In other words we will be looking at the distribution of $X_1, X_2, \cdots, X_{k-1}$ given that X_k is less than some arbitrary positive quantity c, with the only assumption being that c will be bigger than the location parameter of the unobserved covariable X_k. Later on in subsequent sections we will implement the above set-up for a trivariate $P(II)$ model. This chapter is organized in the following way. In section 1, we will consider the hidden truncation paradigm (single variable truncation) for a k-variate $P(II)$ distribution and will derive the density function for such a model. In section 2, we will consider the hidden truncated density for a trivariate $P(II)$ density when one of the co-variables is truncated from above. In section 3, we will consider maximum likelihood estimation for the above model. In section 4, we will consider the method of moments estimation strategy. In section 5, we will consider estimation of all the parameters for a hidden truncated trivariate $P(II)$ model using a simulation study. In section 6, we will consider estimation of all the parameters under the Bayesian paradigm.

5.2 Single variable hidden truncation from above

Following Arnold (1983), we write $\underline{X} \sim MP^{(k)}(II)(\underline{\mu}, \underline{\sigma}, \alpha)$ if it has the following joint survival function:

$$\bar{F}_{\underline{X}}(\underline{x}) = [1 + \sum_{i=1}^{k} (\frac{x_i - \mu_i}{\sigma_i})^{-\alpha}], x_i > \mu_i, i = 1, 2, \ldots, k. \tag{5.1}$$

Note that in this case the corresponding marginals are, again multivariate $P(II)$; the univariate marginals being $P(II)$ with suitable parameters. Conditional distributions are, again, multivariate $P(II)$; in fact introducing the dot-double dot notation if we define $\underline{X} = (\underline{\dot{X}}^{(k_1)}, \underline{\ddot{X}}^{(k-k_1)})$, then

$$\underline{\dot{X}}^{(k_1)} \sim MP^{(k_1)}(II)(\underline{\dot{\mu}}, \underline{\dot{\sigma}}, \alpha),$$

and

$$\underline{\dot{X}} | \underline{\ddot{X}} = \underline{\ddot{x}} \sim MP^{(k_1)}(II)(\underline{\dot{\mu}}, \ddot{c}(\underline{x})\underline{\dot{\sigma}}, \alpha + k - k_1), \tag{5.2}$$

where

$$\ddot{c}(\underline{x}) = [1 + \sum_{j=k_1+1}^{k} (\frac{x_j - \mu_j}{\sigma_j})^{-\alpha}].$$

In (5.1), the σ_i 's are non-negative marginal scale parameters, and the μ_i´s are marginal location parameters. The non-negative parameter α is an inequality parameter (common to all marginals).

Let us consider the following (for $\underline{c} > \underline{\mu}_2$, where the elements of $\underline{c}$ are any set of $(k - k_1)$ real numbers)

$$\begin{aligned} P(\underline{X}_1 > \underline{x}_1 | \underline{X}_2 > \underline{c}) &= \frac{P(\underline{X}_1 > \underline{x}_1, \underline{X}_2 > \underline{c})}{P(\underline{X}_2 > \underline{c})} \\ &= \frac{[1 + \sum_{i=1}^{k_1} (\frac{x_{1i} - \mu_{1i}}{\sigma_{1i}}) + \sum_{i=k_1+1}^{k} (\frac{c_i - \mu_{2i}}{\sigma_{2i}})]^{-\alpha}}{[1 + \sum_{i=k_1+1}^{k} (\frac{c_i - \mu_{2i}}{\sigma_{2i}})]^{-\alpha}} \quad (5.3) \end{aligned}$$

Let us write for notational simplicity

$$[1 + \sum_{i=k_1+1}^{k} (\frac{c_i - \mu_{2i}}{\sigma_{2i}})]^{-\alpha} = A_{11}(\underline{c}, \underline{\mu_2}, \underline{\sigma_2}).$$

Then equation (5.3) reduces to

$$\begin{aligned} P(\underline{X}_1 > \underline{x}_1 | \underline{X}_2 > \underline{c}) &= \bar{F}(\underline{x}_1 | \underline{X}_2 > \underline{c}) \\ &= [1 + (A_{11}(\underline{c}, \underline{\mu}_2, \underline{\sigma}_2))^{-1} \sum_{i=1}^{k_1} (\frac{x_{1i} - \mu_{1i}}{\sigma_{1i}})]^{-\alpha} \quad (5.4) \end{aligned}$$

So it is evident that the above multivariate hidden truncation density, involving lower truncation (equivalently truncation from below), is again an $MP^{(k)}(II)$ density with a new scale parameter. So this form of hidden truncation does not result in any augmentation our original model for $\underline{X}$, and as a consequence, there is no way to determine whether or not such hidden truncation has occurred. However in contrast, we consider two cases for upper truncation separately which are listed as follows:

- Single variable truncation from above. In this case we consider the conditional distribution of (say) $\underline{\dot{X}}$ given that $X_k \leq c$ where $\underline{\dot{X}}$ is a $(k-1 \times 1)$ vector and the constant $c > \mu_k$.
- More than one variable truncation in which we will consider the conditional distribution of $\dot{X}$ given that $\underline{\ddot{X}} \leq \underline{\ddot{c}}$ where

$$\underline{X} = (\underline{\dot{X}}^{(k_1 \times 1)}, \underline{\ddot{X}}^{(k-k_1 \times 1)}).$$

We discuss the above two types of upper truncation separately in detail as follows:

(a) First we focus on single variable hidden truncation from above for our model. In this case we consider

$$\bar{F}(\underline{\dot{x}}|X_k \leq c) \;=\; \frac{P(\underline{\dot{X}} > \underline{\dot{x}}, X_k \leq c)}{P(X_k \leq c)}. \tag{5.5}$$

Next since the marginals are also $P(II)$ with suitable parameters so we can write

$$P(X_k \leq c) = 1 - [1 + (\frac{c - \mu_k}{\sigma_k})]^{-\alpha},$$

assuming that $c > \mu_k$. However for the numerator we can write

$$\begin{aligned}
&P(\underline{\dot{X}} > \underline{\dot{x}}, X_k \leq c) \\
&= P(\underline{\dot{X}} > \underline{\dot{x}}) - P(\underline{\dot{X}} > \underline{\dot{x}}, X_k > c) \\
&= [1 + \sum_{i=1}^{k-1} (\frac{x_i - \mu_i}{\sigma_i})]^{-\alpha} - [1 + \sum_{i=1}^{k-1} (\frac{x_i - \mu_i}{\sigma_i}) + (\frac{c - \mu_k}{\sigma_k})]^{-\alpha}.
\end{aligned}$$

For the sake of notational simplicity let us write

$$\psi(\alpha, c) = 1 - [1 + (\frac{c - \mu_k}{\sigma_k})]^{-\alpha}.$$

So that the conditional survival function is of the form:

$$\begin{aligned}
&\bar{F}(\underline{\dot{x}} | X_k \leq c) \\
&= \frac{1}{\psi(\alpha, c)} [[1 + \sum_{i=1}^{k-1} (\frac{x_i - \mu_i}{\sigma_i})]^{-\alpha} - [1 + \sum_{i=1}^{k-1} (\frac{x_i - \mu_i}{\sigma_i}) + (\frac{c - \mu_k}{\sigma_k})]^{-\alpha}].
\end{aligned} \tag{5.6}$$

If we define $\theta = \frac{c-\mu_k}{\sigma_k}$, then our survival function reduces to

$$\bar{F}(\underline{\dot{x}}|X_k \leq \theta)$$
$$= \frac{1}{\psi(\alpha,\theta)}[1+\sum_{i=1}^{k-1}(\frac{x_i-\mu_i}{\sigma_i})]^{-\alpha} - [1+\sum_{i=1}^{k-1}(\frac{x_i-\mu_i}{\sigma_i})+\theta]^{-\alpha}, x_i \geq \mu_i, \tag{5.7}$$

where

$$\psi(\alpha,\theta) = 1-[1+\theta]^{-\alpha}.$$

So the corresponding density will be

$$\begin{aligned} & f_{\underline{\dot{X}}|X_k\leq\theta}(\underline{\dot{x}}) \\ &= \frac{\partial^{k-1}}{\partial x_1 \partial x_2 \cdots \partial x_{k-1}} \bar{F}(\underline{\dot{x}}|X_k \leq \theta) \\ &= \frac{1}{\psi(\alpha,\theta)} \prod_{i=1}^{k-1}(\frac{\alpha-i+1}{\sigma_i})[1+\sum_{i=1}^{k-1}(\frac{x_i-\mu_i}{\sigma_i})]^{-(\alpha+k)} \\ & -[1+\sum_{i=1}^{k-1}(\frac{x_i-\mu_i}{\sigma_i})+\theta]^{-(\alpha+k)} I(x_i \geq \mu_i). \end{aligned} \tag{5.8}$$

(b) Next we consider k_1 variable (where $k_1 \leq k, k_1 \geq 2$) truncation from above. First of all we consider the following:

$$\bar{F}(\underline{\dot{x}}|\underline{\ddot{X}} \leq \underline{\ddot{c}}) = \frac{P(\underline{\dot{X}} > \underline{\dot{x}}, \underline{\ddot{X}} \leq \underline{\ddot{c}})}{P(\underline{\ddot{X}} \leq \underline{\ddot{c}})}, \tag{5.9}$$

where we assume that $\underline{\ddot{c}} > \underline{\ddot{\mu}}$. Next for our model (5.11), the denominator is given by

$$P(\underline{\ddot{X}} \leq \underline{\ddot{c}})$$

$$= 1 - P(\bigcup_{j_1=k_1+1}^{k} (X_{j_1} > c_{j_1}))$$

$$= 1 - [\sum_{j_1=k_1+1}^{k} P(X_{j_1} > c_{j_1}) - \sum\sum_{j_1<j_2} P(X_{j_1} > c_{j_1}, X_{j_2} > c_{j_2}) + \cdots$$

$$+ (-1)^{k-k_1} P(\bigcap_{j_1=k_1+1}^{k} [X_{j_1} > c_{j_1}])] \tag{5.10}$$

Now since in our case we have

- $\underline{\dot{X}}^{(k_1)} \sim MP^{(k_1)}(II)(\underline{\dot{\mu}}, \underline{\dot{\sigma}}, \alpha)$
- $\underline{\ddot{X}}^{(k-k_1)} \sim MP^{(k-k_1)}(II)(\underline{\ddot{\mu}}, \underline{\ddot{\sigma}}, \alpha)$.

So that we can rewrite (5.12), as follows:

$$P(\underline{\ddot{X}} \leq \underline{\ddot{c}})$$

$$= 1 - [\sum_{j_1=k_1+1}^{k} (1 + \frac{c_{j_1} - \mu_{j_1}}{\sigma_{j_1}})^{-\alpha} - \sum\sum_{j_1<j_2} (1 + \frac{c_{j_1} - \mu_{j_1}}{\sigma_{j_1}} + \frac{c_{j_2} - \mu_{j_2}}{\sigma_{j_2}})^{-\alpha}$$

$$+ \cdots + (-1)^{k-k_1} (1 + \sum_{j_1=k_1+1}^{k} \frac{c_{j_1} - \mu_{j_1}}{\sigma_{j_1}})^{-\alpha}]$$

$$= B(\underline{\ddot{c}}, \underline{\ddot{\mu}}, \underline{\ddot{\sigma}}, \alpha), \tag{5.11}$$

say, while the numerator will be

$$
\begin{aligned}
&P(\underline{\dot{X}} > \underline{\dot{x}}, \underline{\ddot{X}} \leq \underline{\ddot{c}}) \\
&= P(\underline{\dot{X}} > \underline{\dot{x}}, [\bigcup_{j_1=k_1+1}^{k} (X_{j_1} > c_{j_1})]^c) \\
&= P(\underline{\dot{X}} > \underline{\dot{x}}) - \sum_{j_1=k_1+1}^{k} P(\underline{\dot{X}} > \underline{\dot{X}}, X_{j_1} > c_{j_1}) \\
&+ \sum\sum_{j_1<j_2} P(\underline{\dot{X}} > \underline{\dot{x}}, X_{j_1} > c_{j_1}, X_{j_2} > c_{j_2}) \\
&+ \cdots + (-1)^{k-k_1} P(\underline{\dot{X}} > \underline{\dot{x}}, \bigcap_{j_1=k_1+1}^{k} [X_{j_1} > c_{j_1}]) \\
&= (1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i})^{-\alpha} - \sum_{j_1=k_1+1}^{k} (1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \frac{c_i - \mu_i}{\sigma_i})^{-\alpha} \\
&- \sum\sum_{j_1<j_2} (1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \frac{c_{j_1} - \mu_{j_1}}{\sigma_{j_1}} + \frac{c_{j_2} - \mu_{j_2}}{\sigma_{j_2}})^{-\alpha} \\
&+ \cdots + (-1)^{k-k_1} (1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \sum_{j_1=k_1+1}^{k} 1 + \frac{c_{j_1} - \mu_{j_1}}{\sigma_{j_1}})^{-\alpha}.
\end{aligned}
\tag{5.12}
$$

So the conditional survival function is given by

$$\begin{aligned}
&\bar{F}(\underline{\dot{x}}|\underline{\ddot{X}} \leq \underline{\ddot{c}}) \\
&= \frac{1}{B(\underline{\ddot{c}}, \underline{\ddot{\mu}}, \underline{\ddot{\sigma}}, \alpha)}[(1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i})^{-\alpha} \\
&- \sum_{j_1=k_1+1}^{k} (1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \frac{c_{j_1} - \mu_{j_1}}{\sigma_{j_1}})^{-\alpha} \\
&- \sum\sum_{j_1<j_2} (1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \frac{c_{j_1} - \mu_{j_1}}{\sigma_{j_1}} + \frac{c_{j_2} - \mu_{j_2}}{\sigma_{j_2}})^{-\alpha} \\
&+ \cdots + (-1)^{k-k_1}(1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \sum_{j_1=k_1+1}^{k} 1 + \frac{c_{j_1} - \mu_{j_1}}{\sigma_{j_1}})^{-\alpha}], \\
&x_i \geq \mu_i, i = 1, 2, \cdots, k_1.
\end{aligned} \tag{5.13}$$

Consequently the corresponding hidden truncated density will be given by

$$
\begin{aligned}
& f_{\underline{\dot{X}}|\underline{\ddot{X}}\le\underline{\ddot{c}}}(\underline{\dot{x}}) \\
&= \frac{\partial^{k-1}}{\partial x_1 \partial x_2 \cdots \partial x_{k_1}}[\bar{F}(\underline{\dot{x}}|\underline{\ddot{X}} \le \underline{\ddot{c}})] \\
&= \frac{1}{B(\underline{\ddot{c}}, \underline{\ddot{\mu}}, \underline{\ddot{\sigma}}, \alpha)} \prod_{i=1}^{k_1-1} (\frac{\alpha - i + 1}{\sigma_i})[(1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i})^{-(\alpha+k_1)} \\
&- \sum_{j_1=k_1+1}^{k} (1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \frac{c_{j_1} - \mu_{j_1}}{\sigma_{j_1}})^{-(\alpha+k_1)} \\
&- \sum \sum_{j_1<j_2} (1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \frac{c_{j_1} - \mu_{j_1}}{\sigma_{j_1}} + \frac{c_{j_2} - \mu_{j_2}}{\sigma_{j_2}})^{-(\alpha+k_1)} \\
&+ \cdots + (-1)^{k-k_1} (1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \sum_{j_1=k_1+1}^{k} 1 + \frac{c_{j_1} - \mu_{j_1}}{\sigma_{j_1}})^{-(\alpha+k_1)}] \\
&\times I(\underline{\dot{x}} \ge \underline{\dot{\mu}}). \qquad (5.14)
\end{aligned}
$$

For notational simplicity let us consider $\frac{c_l - \mu_l}{\sigma_l} = \theta_l$. Then our hidden truncated density reduces to

$$
\begin{aligned}
& f_{\underline{\dot{X}}|\underline{\ddot{X}}\leq\underline{\ddot{c}}}(\underline{\dot{x}}) \\
&= \frac{\partial^{k-1}}{\partial x_1 \partial x_2 \cdots \partial x_{k_1}}[\bar{F}(\underline{\dot{x}}|\underline{\ddot{X}} \leq \underline{\ddot{c}})] \\
&= \frac{1}{B(\underline{\ddot{c}}, \underline{\ddot{\mu}}, \underline{\ddot{\sigma}}, \alpha)} \prod_{i=1}^{k_1-1}(\frac{\alpha - i + 1}{\sigma_i})[(1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i})^{-(\alpha+k_1)} \\
&- \sum_{j_1=k_1+1}^{k}(1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \theta_{j_1})^{-(\alpha+k_1)} \\
&- \sum\sum_{j_1<j_2}(1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \theta_{j_1} + \theta_{j_2})^{-(\alpha+k_1)} \\
&+ \cdots + (-1)^{k-k_1}(1 + \sum_{i=1}^{k_1} \frac{x_i - \mu_i}{\sigma_i} + \sum_{j_1=k_1+1}^{k} 1 + \theta_{j_1})^{-(\alpha+k_1)}]I(\underline{\dot{x}} \geq \underline{\dot{\mu}}).
\end{aligned} \tag{5.15}
$$

However we focus on in this chapter a simpler case where we consider

$$\underline{X}^{(3\times1)} \sim MP^{(3)}(II)(\underline{\mu}, \underline{\sigma}, \alpha),$$

where $\underline{X}^{(3\times1)} = (X_1, X_2, X_3)^T$. where all the marginals as well as all the conditionals are again members of $P(II)$ family with suitable choice of the parameters. We are interested in the distribution of X_1 when both X_2 and X_3 are truncated from above. Let us consider a situation that we will observe X_1 and X_2 if and only if $X_3 \leq c$. So we want to find the joint density of X_1 and X_2 given $X_3 \leq c$. Then from (5.10), our density will be (by substituting $k = 3$, $k_1 = 2$)

$$f_{X_1,X_2|X_3\leq\theta}(x_1,x_2) = \frac{\alpha(\alpha+1)}{\sigma_1\sigma_2\psi(\alpha,\theta)}[(1+\frac{(x_1-\mu_1)}{\sigma_1}+\frac{(x_2-\mu_2)}{\sigma_2})^{-(\alpha+2)}$$
$$-(1+\frac{x_1}{\sigma_1}+\frac{x_2}{\sigma_2}+\theta)^{-(\alpha+2)}]I(x_1\geq\mu_1,x_2\geq\mu_1), \tag{5.16}$$

where

$$\psi(\alpha,\theta) = 1-(1+\theta)^{-\alpha}.$$

and $\theta = \frac{c-\mu_3}{\sigma_3}$. Note that the parameters are constrained to be such that: $\mu_1 \in (-\infty,\infty)$, $\mu_2 \in (-\infty,\infty)$, $\sigma_1 \in (0,\infty)$, $\sigma_2 \in (0,\infty)$, $\theta \geq 0$, $\alpha > 0$. In subsequent sections we will focus on the problem of estimating and making inferences about these parameters based on a random sample of size n drawn from the density in (5.18). Note that the situation where $\theta = 0$ reduces to the non-truncated case.

5.3 Parameter estimation for the trivariate hidden truncated $P(II)$ model

We will consider two types of estimation using a simulation study. In each case the available data will consist of n i.i.d observations with common hidden truncation $P(II)$ density (5.16).

5.4 Maximum likelihood estimation

We draw a random sample of size n from the density in (5.16). Denote the observations by $(X_{1j}, X_{2j}, j = 1(1)n)$.

So the likelihood function is given by

$$L = \prod_{j=1}^{n} \frac{\alpha(\alpha+1)}{\sigma_1 \sigma_l \psi(\alpha,\theta)}[(1+\frac{X_{1j}-\mu_1}{\sigma_1}+\frac{X_{2j}-\mu_2}{\sigma_2})^{-(\alpha+2)}$$
$$-(1+\frac{X_{1j}-\mu_1}{\sigma_1}+\frac{X_{2j}-\mu_2}{\sigma_2}+\theta)^{-(\alpha+2)}]. \tag{5.17}$$

Equivalently the log-likelihood function is given by

$$\begin{aligned}
&\log L \\
&= n[\log(\alpha(\alpha+1)) - \log\psi(\alpha,\theta) - \log(\sigma_1\sigma_2)] \\
&+ \sum_{j=1}^{n} \log[(1+\frac{X_{1j}-\mu_1}{\sigma_1}+\frac{X_{2j}-\mu_2}{\sigma_2})^{-(\alpha+2)} \\
&- (1+\frac{X_{1j}-\mu_1}{\sigma_1}+\frac{X_{2j}-\mu_2}{\sigma_2}+\theta)^{-(\alpha+2)}] \\
&= n[\log(\alpha(\alpha+1)) - \log\psi(\alpha,\theta) - \log(\sigma_1\sigma_2)] \\
&- (\alpha+2)\sum_{j=1}^{n}[\log(1+\frac{X_{1j}-\mu_1}{\sigma_1}+\frac{X_{2j}-\mu_2}{\sigma_2})] \\
&+ \sum_{j=1}^{n}[\log(1-(1+\frac{\theta}{1+\frac{X_{1j}-\mu_1}{\sigma_1}+\frac{X_{2j}-\mu_2}{\sigma_2}})^{-(\alpha+2)})].
\end{aligned}$$

First of all observe that keeping all other parameters fixed if we replace (μ_1, μ_2) by the minimum of the sample observations i.e.,

$$X_{1(1:n)} = \min_{1 \leq j \leq n} X_{1j},$$

$$X_{2(1:n)} = \min_{1 \leq j \leq n} X_{2j}$$

then the likelihood function (or equivalently the log-likelihood function) becomes a monotonically increasing function in $(X_{1(1:n)}, X_{2(1:n)})$.

So the ML estimates for (μ_1, μ_2) will be given by

$$X_{1(1:n)} = \min_{1 \leq j \leq n} X_{1j}$$

and

$$X_{2(1:n)} = \min_{1 \leq j \leq n} X_{2j}$$

respectively. For notational simplicity let us denote $X_{1(1:n)} = q1$ and $X_{2(1:n)} = q2$.

So the log-likelihood function now becomes

$$\begin{aligned} \log L = {} & n[\log(\alpha(\alpha+1)) - \log\psi(\alpha,\theta) - \log(\sigma_1\sigma_2)] \\ & - (\alpha+2)\sum_{j=1}^{n}[\log(1 + \frac{X_{1j}-q1}{\sigma_1} + \frac{X_{2j}-q2}{\sigma_2})] \\ & + \sum_{j=1}^{n}[\log(1 - (1 + \frac{\theta}{1 + \frac{X_{1j}-q1}{\sigma_1} + \frac{X_{2j}-q2}{\sigma_2}})^{-(\alpha+2)})]. \end{aligned} \tag{5.18}$$

The likelihood equation for the remaining parameters are obtained by differentiating the likelihood function. For α we have

$$\frac{\partial}{\partial \alpha}[\log L] = 0. \tag{5.19}$$

Equivalently we can write

$$\begin{aligned} & n(\frac{1}{\alpha}+\frac{1}{\alpha+1}) - n\frac{(1+\theta)^{-\alpha}\log(1+\theta)}{\psi(\alpha,\theta)} \\ & + \sum_{j=1}^{n} \frac{(1+\frac{\theta}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-\mu_l}{\sigma_2}})^{-(\alpha+2)}}{(1-(1+\frac{\theta}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-q2}{\sigma_2}})^{-(\alpha+2)})} \\ & \times \log((1+\frac{\theta}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-q2}{\sigma_2}})) = 0. \end{aligned}$$

Here $\psi(\alpha,\theta) = 1-(1+\theta)^{-\alpha}$. So that $\frac{\partial}{\partial\alpha}\psi(\alpha,\theta) = (1+\theta)^{-\alpha}\log(1+\theta)$. Again we have

$$\frac{\partial}{\partial \sigma_1}[\log L] = 0. \tag{5.20}$$

Equivalently we can write

$$\begin{aligned} & -\frac{n}{\sigma_1} + (\alpha+2)\sum_{j=1}^{n}[\frac{1}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-q2}{\sigma_2}}(\frac{X_{1j}-q1}{(\sigma_1)^2})] \\ & +(\alpha+2)\sum_{j=1}^{n}[\frac{(1+\frac{\theta}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-q2}{\sigma_2}})^{-(\alpha+3)}}{(1-(1+\frac{\theta}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-q2}{\sigma_2}})^{-(\alpha+2)})} \\ & \times\frac{\theta(\frac{X_{1j}-q1}{\sigma_1})}{(\sigma_1(1+\frac{X_{2j}-q2}{\sigma_2})+(X_{1j}-q1))^2}] = 0. \end{aligned}$$

Similarly

$$\frac{\partial}{\partial \sigma_2}[\log L] = 0.$$

Equivalently we can write

$$-\frac{n}{\sigma_2} + (\alpha+2)\sum_{j=1}^{n}[\frac{1}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-q2}{\sigma_2}}(\frac{X_{1j}-q1}{\sigma_1^2})]$$
$$+ (\alpha+2)\sum_{j=1}^{n}[\frac{(1+\frac{\theta}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-q2}{\sigma_2}})^{-(\alpha+3)}}{(1-(1+\frac{\theta}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-q2}{\sigma_2}})^{-(\alpha+2)})}$$
$$\times \frac{\theta(\frac{X_{2j}-q1}{\sigma_2})}{(\sigma_2(1+\frac{X_{1j}-q1}{\sigma_1})+(X_{2j}-q2))^2}] = 0.$$

Finally we have

$$\frac{\partial}{\partial \theta}[\log L] = 0. \tag{5.21}$$

Equivalently we can write

$$-n\frac{\psi'(\alpha,\theta)}{\psi(\alpha,\theta)} + (\alpha+2)\sum_{j=1}^{n}\frac{1}{(1-(1+\frac{\theta}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-q2}{\sigma_2}})^{-(\alpha+2)})}$$
$$\times ((1+\frac{\theta}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-q2}{\sigma_2}})^{-(\alpha+3)})(1+\frac{\theta}{1+\frac{X_{1j}-q1}{\sigma_1}+\frac{X_{2j}-q2}{\sigma_2}})^{-1} = 0.$$

Note that here

$$\frac{\partial}{\partial \theta}\psi(\alpha,\theta) = \alpha(1+\theta)^{-(\alpha+1)}.$$

5.5 Method of moments estimation

We have our density (from earlier)

$$\begin{aligned}
& f_{X_1,X_2|X_3\le\theta}(x_1,x_2) \\
&= \frac{\alpha(\alpha+1)}{\sigma_1\sigma_2\psi(\alpha,\theta)}[(1+\frac{x_1}{\sigma_1}+\frac{x_2}{\sigma_2})^{-(\alpha+2)} \\
&- (1+\frac{x_1}{\sigma_1}+\frac{x_2}{\sigma_2}+\theta)^{-(\alpha+2)}]I(x_1\ge 0, x_2\ge 0).
\end{aligned}$$

Let us consider for any $\delta_1 \ge 1$,$\delta_2 \ge 1$,

$$\begin{aligned}
E(X_1^{\delta_1}X_2^{\delta_2}|X_3\le\theta) &= \int_0^\infty\int_0^\infty x_1^{\delta_1}x_2^{\delta_2}\frac{\alpha(\alpha+1)}{\sigma_1\sigma_2\psi(\alpha,\theta)}[(1+\frac{x_1}{\sigma_1}+\frac{x_2}{\sigma_2})^{-(\alpha+2)} \\
&\quad -(1+\frac{x_1}{\sigma_1}+\frac{x_2}{\sigma_2}+\theta)^{-(\alpha+2)}]dx_1dx_2 \\
&= I,
\end{aligned}$$

say, where

$$I = \int_0^\infty \frac{\alpha(\alpha+1)}{\psi(\alpha,\theta)\sigma_1}(I_1 - I_2)dx_1$$

in which

$$\begin{aligned}
I_1 &= \int_0^\infty \frac{x_l^{\delta_2}}{\sigma_2}[(1+\frac{x_1}{\sigma_1}+\frac{x_2}{\sigma_2})^{-(\alpha+2)}]dx_2 \\
&= \frac{(\sigma_2)^{\delta_2}}{(\alpha+1)}(1+\frac{x_1}{\sigma_1})^{\delta_2+1-(\alpha+2)}B(\delta_2+1,\alpha+2-\delta_2),
\end{aligned}$$

and

$$\begin{aligned} I_2 &= \int_0^\infty \frac{x_l^{\delta_2}}{\sigma_2}[(1+\frac{x_1}{\sigma_1}+\frac{x_2}{\sigma_2}+\theta)^{-(\alpha+2)}]dx_2 \\ &= \frac{(\sigma_2)^{\delta_2}}{(\alpha+1)}(1+\frac{x_1}{\sigma_1}+\theta)^{\delta_2+1-(\alpha+2)}B(\delta_2+1,\alpha+2-\delta_2). \end{aligned}$$

So that

$$\begin{aligned} & I = \alpha B(\delta_2+1,\alpha+2-\delta_2)\int_0^\infty \frac{x_1^{\delta_1}}{\sigma_1}[(1+\frac{x_1}{\sigma_1})^{-(\alpha+2)} \\ & -(1+\frac{x_1}{\sigma_1}+\theta)^{-(\alpha+2)}]dx_1 \\ & = (\sigma_1)^{\delta_1}(\sigma_2)^{\delta_2}B(\delta_1+1,\alpha+2-\delta_1-\delta_2-1) \\ & \times B(\delta_2+1,\alpha+2-\delta_2)(1+\theta)^{\delta_1+1-(\alpha+2-\delta_2-1)}. \end{aligned}$$

Therefore

$$\begin{aligned} & E(X_1^{\delta_1}X_2^{\delta_2}|X_3\leq\theta) \\ & = \frac{\alpha(\alpha+1)}{\psi(\alpha,\theta)}\sigma_1^{\delta_1}\sigma_2^{\delta_2}B(\delta_1+1,\alpha+2-\delta_1-\delta_2-1)B(\delta_2+1,\alpha+2-\delta_2) \\ & \times(1-(1+\theta)^{\delta_1+1-(\alpha+2-\delta_2-1)}). \end{aligned} \tag{5.22}$$

Note:

- If $\delta_1=\delta_2=\delta$, then we have $E(X_1^{\delta}X_2^{\delta}|X_3\leq\theta)=\frac{\alpha(\alpha+1)}{\psi(\alpha,\theta)}B(\delta+1,\alpha+1-2\delta)B(\delta+1,\alpha+2-\delta)[1-(1+\theta)^{2\delta-\alpha}]$.

- If $\delta_1 = \delta_2 = 1$, then from equation (5.22) we have

$$\begin{aligned}
&E(X_1X_2|X_3 \leq \theta) \\
&= \frac{\alpha(\alpha+1)}{\psi(\alpha,\theta)}\sigma_1\sigma_2 B(2,\alpha-1)B(2,\alpha+1)(1-(1+\theta)^{2-\alpha}).
\end{aligned}$$

- If $\delta_1 = \delta_2 = \frac{1}{2}$, then from equation (5.22) we have

$$\begin{aligned}
&E(X_1^{\frac{1}{2}}X_2^{\frac{1}{2}}|X_3 \leq \theta) \\
&= \frac{\alpha(\alpha+1)}{\psi(\alpha,\theta)}B(2,\alpha)B(\frac{3}{2},\alpha+1)(\sigma_1\sigma_2)^{\frac{1}{2}}(1-(1+\theta)^{1-\alpha}).
\end{aligned}$$

Again

- If $\delta_1 = \delta_2 = \frac{1}{3}$, then from equation (5.22), we have

$$E(X_1^{\frac{1}{3}}X_2^{\frac{1}{3}}|X_3 \leq \theta) = \frac{\alpha(\alpha+1)}{\psi(\alpha,\theta)}B(\frac{4}{3},\alpha+\frac{1}{3})B(\frac{4}{3},\alpha+\frac{4}{3})(\sigma_1\sigma_2)^{\frac{2}{3}}.$$

Equivalently

$$\begin{aligned}
&E(X_1^{\frac{1}{3}}X_2^{\frac{1}{3}}|X_3 \leq \theta) \\
&= \frac{\alpha(\alpha+1)}{\psi(\alpha,\theta)}B(\frac{4}{3},\alpha+\frac{1}{3})B(\frac{4}{3},\alpha+\frac{4}{3}) \\
&\times (\sigma_1\sigma_2)^{\frac{1}{3}}(1-(1+\theta)^{\frac{2}{3}-\alpha}).
\end{aligned}$$

Again

$$
\begin{aligned}
&E(X_2^{\frac{1}{4}} X_2^{\frac{1}{4}} | X_3 \le \theta) \\
&= \frac{\alpha(\alpha+1)}{\psi(\alpha,\theta)} B(\frac{1}{4}+1, \alpha+\frac{1}{2}) B(\frac{1}{4}+1, \alpha+\frac{3}{2}) \\
&\times (\sigma_1\sigma_2)^{\frac{1}{4}} (1-(1+\theta)^{\frac{3}{2}-\alpha}).
\end{aligned}
$$

Next we define the following quantity $M_{a1,a2} = \frac{\sum_{i=1}^{n}(X_{1i}-q1)^{a1}(X_{2i}-q2)^{a2}}{n-2}$, which is the sample bivariate moment of order $(a1, a2)$, where both $(a1, a2)$ are positive numbers (where both $q1$ and $q2$ have been defined earlier).

Next equating the sample moments with the population moments we get the following:

$$
\begin{aligned}
\frac{M_{1,1}}{M_{\frac{1}{2},\frac{1}{2}}} &= \frac{\sigma_1\sigma_2 B(2,\alpha-1) B(2,\alpha+1)(1-(1+\theta)^{2-\alpha})}{B(2,\alpha) B(\frac{3}{2},\alpha+1)(\sigma_1\sigma_2)^{\frac{1}{2}}(1-(1+\theta)^{1-\alpha})} \\
&= \frac{2(\alpha+1)\Gamma(\alpha+\frac{5}{2})(\sigma_1\sigma_2)^{\frac{1}{2}}(1-(1+\theta)^{2-\alpha})}{\sqrt{\pi}(\alpha-1)\Gamma(\alpha+3)(1-(1+\theta)^{1-\alpha})}.
\end{aligned}
$$

Equivalently we can write

$$
\begin{aligned}
&M_{1,1}[\sqrt{\pi}(\alpha-1)\Gamma(\alpha+3)(1-(1+\theta)^{1-\alpha})] - M_{\frac{1}{2},\frac{1}{2}} \\
&[2(\alpha+1)\Gamma(\alpha+\frac{5}{2})(\sigma_1\sigma_2)^{\frac{1}{2}}(1-(1+\theta)^{2-\alpha})] = 0. \qquad (5.23)
\end{aligned}
$$

Again

$$\begin{aligned}\frac{M_{\frac{1}{2},\frac{1}{2}}}{M_{\frac{1}{3},\frac{1}{3}}} &= \frac{B(2,\alpha)B(\frac{3}{2},\alpha+1)(\sigma_1\sigma_2)^{\frac{1}{2}}(1-(1+\theta)^{1-\alpha})}{B(\frac{4}{3},\alpha+\frac{1}{3})B(\frac{4}{3},\alpha+\frac{4}{3})(\sigma_1\sigma_2)^{\frac{1}{3}}(1-(1+\theta)^{\frac{2}{3}-\alpha})} \\ &= \frac{\Gamma(\alpha)\sqrt{\pi}\Gamma(\alpha+\frac{5}{3})\Gamma(\alpha+\frac{8}{3})(\sigma_1\sigma_2)^{\frac{1}{6}}(1-(1+\theta)^{1-\alpha})}{2(\alpha+1)\Gamma(\alpha+\frac{5}{2})(\Gamma(\frac{4}{3}))^2\Gamma(\alpha+\frac{1}{3})\Gamma(\alpha+\frac{4}{3})(1-(1+\theta)^{\frac{2}{3}-\alpha})}.\end{aligned}$$

Equivalently we can write

$$\begin{aligned}&M_{\frac{1}{2},\frac{1}{2}}[2(\alpha+1)\Gamma(\alpha+\frac{5}{2})(\Gamma(\frac{4}{3}))^2\Gamma(\alpha+\frac{1}{3})\Gamma(\alpha+\frac{4}{3})(1-(1+\theta)^{\frac{2}{3}-\alpha})] \\ &-M_{\frac{1}{3},\frac{1}{3}}[\Gamma(\alpha)\sqrt{\pi}\Gamma(\alpha+\frac{5}{3})\Gamma(\alpha+\frac{8}{3})(\sigma_1\sigma_2)^{\frac{1}{6}}(1-(1+\theta)^{1-\alpha})]=0.\end{aligned} \tag{5.24}$$

Similarly we can have

$$\begin{aligned}\frac{M_{1,1}}{M_{\frac{1}{3},\frac{1}{3}}} &= \frac{\sigma_1\sigma_2 B(2,\alpha-1)B(2,\alpha+1)(1-(1+\theta)^{2-\alpha})}{B(\frac{4}{3},\alpha+\frac{1}{3})B(\frac{4}{3},\alpha+\frac{4}{3})(\sigma_1\sigma_2)^{\frac{1}{3}}(1-(1+\theta)^{\frac{2}{3}-\alpha})} \\ &= \frac{\Gamma(\alpha+\frac{5}{3})\Gamma(\alpha+\frac{8}{3})(\sigma_1\sigma_2)^{\frac{2}{3}}}{(\alpha+2)(\alpha+1)(\alpha-1)\alpha(\Gamma(\frac{4}{3}))^2\Gamma(\alpha+\frac{1}{3})\Gamma(\alpha+\frac{4}{3})}.\end{aligned}$$

Equivalently we can write

$$\begin{aligned}&M_{1,1}[(\alpha+2)(\alpha+1)(\alpha-1)\alpha(\Gamma(\frac{4}{3}))^2\Gamma(\alpha+\frac{1}{3})\Gamma(\alpha+\frac{4}{3})] \\ &-M_{\frac{1}{3},\frac{1}{3}}[\Gamma(\alpha+\frac{5}{3})\Gamma(\alpha+\frac{8}{3})(\sigma_1\sigma_2)^{\frac{2}{3}}]=0.\end{aligned} \tag{5.25}$$

Also we have

$$\begin{aligned}\frac{M_{1,1}}{M_{\frac{1}{4},\frac{1}{4}}} &= \frac{\sigma_1\sigma_2 B(2,\alpha-1)B(2,\alpha+1)(1-(1+\theta)^{2-\alpha})}{B(\frac{1}{4}+1,\alpha+\frac{1}{2})B(\frac{1}{4}+1,\alpha+\frac{3}{2})(\sigma_1\sigma_2)^{\frac{1}{4}}(1-(1+\theta)^{\frac{3}{2}-\alpha})} \\ &= \frac{\Gamma(\alpha+\frac{7}{4})\Gamma(\alpha+\frac{11}{4})(1-(1+\theta)^{2-\alpha})(\sigma_1\sigma_2)^{\frac{3}{4}}}{\Gamma(\alpha+\frac{1}{2})\Gamma(\alpha+\frac{3}{2})(\Gamma(\frac{5}{4}))^2(\alpha+2)(\alpha+1)(\alpha-1)\alpha(1-(1+\theta)^{\frac{1}{2}-\alpha})}.\end{aligned}$$

Equivalently we can write

$$\begin{aligned}&M_{1,1}[\Gamma(\alpha+\frac{1}{2})\Gamma(\alpha+\frac{3}{2})(\Gamma(\frac{5}{4}))^2(\alpha+2)(\alpha+1)(\alpha-1)\alpha(1-(1+\theta)^{\frac{1}{2}-\alpha})] \\ &-M_{\frac{1}{4},\frac{1}{4}}[\Gamma(\alpha+\frac{7}{4})\Gamma(\alpha+\frac{11}{4})(1-(1+\theta)^{2-\alpha})(\sigma_1\sigma_2)^{\frac{3}{4}}]=0. \qquad (5.26)\end{aligned}$$

5.6 Estimation of the parameters using a simulation study

By looking at the method of (fractional) moments equations and also the likelihood equations one can easily understand that it is not possible to find an analytic expression for the parameter estimates. So we at first consider one simple situation where we specify all the parameter values and then generate some data from our hidden truncated density. We then consider the estimation for all the parameters by maximum likelihood and also by the fractional moments method.

5.6.1 Sample generation from the truncated density

For our simulation study we consider (without loss of generality), $\alpha = 4$, $\mu_1 = 1$, $\mu_2 = 1$, $\sigma_1 = 1$, $\sigma_2 = 1$, $c = 2$, $\mu_3 = 0$, $\sigma_3 = 1$, so that

$\theta = \frac{c-\mu_3}{\sigma_3} = 2$, and $\psi(\alpha,\theta) = 1 - (1 + (\frac{c-\mu_3}{\sigma_3}))^{-\alpha} = \frac{80}{81}$.

But note that we usually estimate μ_1 and μ_2 by $X_{1(1:n)}$ and $X_{2(1:n)}$ respectively and so we can subtract and assume that $\mu_1 = 0$ and $\mu_2 = 0$. Our density reduces to

$$f_{X_1,X_2|X_3\le 2}(x_1,x_2) = \frac{81}{4}[(1+x_1+x_2)^{-6} - (1+x_1+x_2+2)^{-6}]I(x_1 \ge 0, x_2 \ge 0).$$

In Table 5.1, we illustrate for simulated samples of various sizes the estimated values of the parameters (using the ML method):

Table 5.1: Estimates of the parameters using the method of maximum likelihood

n	$\hat{\alpha}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\theta}$
50	3.8649117	0.9394654	0.9923756	1.9953699
100	3.8664039	0.9413608	0.9904680	1.9949965
200	3.8965841	0.9640310	0.9550397	1.9990409

In Table 5.2, parallel results obtained using the method of moments are displayed.

Table 5.2: Estimates of the parameters using the method of moments

n	$\hat{\alpha}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\theta}$
50	3.9816431	0.9625353	0.9573555	1.9745285
100	3.9931764	0.9867361	0.9527272	1.9945825
200	3.9954369	0.9895283	0.9812457	1.9939241

5.7 Bayesian inference for the hidden truncated $P(II)$ model

In this section we will consider Bayesian estimation of all the parameters under study. As before we have our density as

$$
\begin{aligned}
& f_{X_1,X_2|X_3\leq\theta}(x_1,x_2) \\
& = \frac{\alpha(\alpha+1)}{\sigma_1\sigma_2\psi(\alpha,\theta)}[(1+\frac{x_1}{\sigma_1}+\frac{x_2}{\sigma_2})^{-(\alpha+2)} \\
& - (1+\frac{x_1}{\sigma_1}+\frac{x_2}{\sigma_2}+\theta)^{-(\alpha+2)}]I(x_1\geq 0, x_2\geq 0).
\end{aligned}
$$

5.7.1 Sample and prior information

We consider a random sample of size n from above density (based on some particular choices of the parameter). Next we propose the following choice of independent priors (mildly informative) for the four parameters:

- $f(\alpha) \propto \frac{1}{\alpha}I(\alpha > 0)$.
- $f(\theta) \propto \frac{1}{\theta}I(\theta > 0)$.
- $f(\sigma_1) \propto \frac{1}{\sigma_1}I(\sigma_1 > 0)$.
- $f(\sigma_2) \propto \frac{1}{\sigma_2}I(\sigma_2 > 0)$.

5.7.2 Posterior distribution of the parameters

In this case our likelihood function is given by

$$
\begin{aligned}
&L(\alpha, \sigma_1, \sigma_2\theta | \underline{X}_1 = \underline{x}_1, \underline{X}_2 = \underline{x}_2, X_3 \leq \theta) \\
&= \prod_{i=1}^{n} [\frac{\alpha(\alpha+1)}{\sigma_1\sigma_2\psi(\alpha,\theta)}((1+\frac{x_{1i}}{\sigma_1}+\frac{x_{2i}}{\sigma_2})^{-(\alpha+2)} - (1+\frac{x_{1i}}{\sigma_1}+\frac{x_{2i}}{\sigma_2}+\theta)^{-(\alpha+2)})] \\
&= \frac{(\alpha(\alpha+1))^n}{(\sigma_1\sigma_2\psi(\alpha,\theta))^n} \prod_{i=1}^{n} [((1+\frac{x_{1i}}{\sigma_1}+\frac{x_{2i}}{\sigma_2})^{-(\alpha+2)} - (1+\frac{x_{1i}}{\sigma_1}+\frac{x_{2i}}{\sigma_2}+\theta)^{-(\alpha+2)})].
\end{aligned}
\tag{5.27}
$$

So the joint posterior of the four parameters is given by

$$
\begin{aligned}
&f(\alpha, \sigma_1, \sigma_2, \theta | \underline{X}_1 = \underline{x}_1, \underline{X}_2 = \underline{x}_2, X_3 \leq \theta) \\
&= \frac{L(\alpha, \sigma_1, \sigma_2\theta | \underline{X}_1 = \underline{x}_1, \underline{X}_2 = \underline{x}_2, X_3 \leq \theta) f(\alpha) f(\theta) f(\sigma_1) f(\sigma_2)}{B},
\end{aligned}
$$

where B is the normalizing constant which is given by

$$
\begin{aligned}
&B \\
&= \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty L(\alpha, \sigma_1, \sigma_2, \theta | \underline{X}_1 = \underline{x}_1, \underline{X}_2 = \underline{x}_2, X_3 \leq \theta) \\
&\times f(\alpha) f(\theta) f(\sigma_1) f(\sigma_2) d\sigma_1 d\sigma_2 d\theta d\alpha \\
&= \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \xi(\alpha, \sigma_1, \sigma_2, \theta) d\sigma_2 d\theta d\alpha,
\end{aligned}
$$

where $\xi(\alpha, \sigma_1, \sigma_2, \theta)$ is given by

$$\xi(\alpha,\sigma_1,\sigma_2,\theta)$$
$$= \frac{(\alpha(\alpha+1))^n}{(\sigma_1\sigma_2\psi(\alpha,\theta))^n}\prod_{i=1}^{n}[[((1+\frac{x_{1i}}{\sigma_1}+\frac{x_{2i}}{\sigma_2})^{-(\alpha+2)}-(1+\frac{x_{1i}}{\sigma_1}+\frac{x_{2i}}{\sigma_2}+\theta)^{-(\alpha+2)})]$$
$$\times (1+\theta)^{-2}(1+\alpha)^{-2}(1+\sigma_1)^{-2}(1+\sigma_2)^{-2}].$$

So the posterior density of α is given by

$$f(\alpha|\underline{X}_1=\underline{x}_1,\underline{X}_2=\underline{x}_2,X_3\leq\theta) \;=\; \frac{\int_0^\infty\int_0^\infty\int_0^\infty \xi(\alpha,\sigma_1,\sigma_2,\theta)d\theta d\sigma_1 d\sigma_2}{B}.$$

Also the posterior density of θ is given by

$$f(\theta|\underline{X}_1=\underline{x}_1,\underline{X}_2=\underline{x}_2,X_3\leq\theta) \;=\; \frac{\int_0^\infty\int_0^\infty\int_0^\infty \xi(\alpha,\sigma_1,\sigma_2,\theta)d\theta d\sigma_1 d\sigma_2}{B}.$$

Again the posterior density of σ_1 is given by

$$f(\sigma_1|\underline{X}_1=\underline{x}_1,\underline{X}_2=\underline{x}_2,X_3\leq\theta) \;=\; \frac{\int_0^\infty\int_0^\infty\int_0^\infty \xi(\alpha,\sigma_1,\sigma_2,\theta)d\theta d\sigma_1 d\sigma_2}{B}.$$

Finally the posterior density of σ_2 is given by

$$f(\sigma_2|\underline{X}_1=\underline{x}_1,\underline{X}_2=\underline{x}_2,X_3\leq\theta) \;=\; \frac{\int_0^\infty\int_0^\infty\int_0^\infty \xi(\alpha,\sigma_1,\sigma_2,\theta)d\theta d\sigma_1 d\sigma_2}{B}.$$

5.7.3 Posterior simulation study

In our case we consider Metropolis-Hastings algorithm which is a general term for a family of Markov chain simulation methods that are useful for drawing samples from Bayesian posterior distributions. First of all we draw random samples of size 100 and 200 from the hidden truncated density as described earlier for a particular choice ($\alpha = 3, \sigma_1 = 4, \sigma_2 = 4, \theta = 2$) of the parameters. For the jumping distribution we consider the same Gamma distribution but with different shape and scale parameters for the four parameters under study. The posterior analysis is based on the posterior modes for each of those four parameters. Below we provide various choices for the starting distribution, the choices of the parameters of the jumping distribution along with posterior modes:

Initial choices of the parameters : $\alpha = 2.8, \sigma_1 = 0.89, \sigma_2 = 0.98, \theta = 1.4$.

Jumping distribution for all the parameters:

- $\alpha \sim \Gamma(25.2/5, 6.9/5)$.
- $\sigma_1 \sim \Gamma(4.2/2, 1.8/2)$.
- $\sigma_2 \sim \Gamma(4.3/2, 1.8/2)$.
- $\theta \sim \Gamma(8/3, 1.9/3)$.

Note: The acceptance/rejection rule of the Metropolis-Hastings algorithm requires the ability to calculate the importance ratio r, for which we need to have jumping distribution for the parameters under study. For a detailed discussion, see [18].

The following table provides the posterior mode values for all the parameters based on the Bayesian posterior simulation study.

Table 5.3: Posterior modes of the parameters

n	Mode(α)	Mode(σ_1)	Mode(σ_2)	Mode(θ)
100	2.76741	3.8507	3.8520	1.85432
200	2.91723	3.9727	3.9686	1.96717

5.8 Hidden truncation for the trivariate $P(II)$ distribution

In this section we focus our attention to a situation where we consider

$$\underline{X}^{(3\times1)} \sim MP^{(3)}(II)((\underline{\mu}, \underline{\sigma}, \alpha),$$

where we consider

$$\underline{X}^{(3\times1)} = (X, X_2, X_3)^T,$$

and all the marginals as well as all the conditionals are again members of $P(II)$ family with suitable choice of the parameters. In particular we consider a situation where all the location parameters are zero and the scale parameters are one. So that in this case we have

$$\underline{X}^{(3\times1)} \sim MP^{(3)}(II)((\underline{0}, \underline{1}, \alpha).$$

We are interested in finding the distribution of X_1 when both X_2 and X_3 are truncated from above. Thus we consider a situation in which we will observe X if and only if $X_2 \leq \theta_1$ and $X_3 \leq \theta_2$. So we want to find the density of X given $X_2 \leq \theta_1$ and $X_3 \leq \theta_2$. First let us consider

$$P(X > x | X_2 \leq \theta_1, X_3 \leq \theta_2) = \frac{P(X > x, X_2 \leq \theta_1, X_3 \leq \theta_2)}{P(X_2 \leq \theta_1, X_3 \leq \theta_2)}, x \geq 0. \tag{5.28}$$

However the denominator

$$\begin{aligned}
&P(X_2 \leq \theta_1, X_3 \leq \theta_2) \\
&= P(X_2 \leq \theta_1) + P(X_3 \leq \theta_2) + P(X_2 > \theta_1, X_3 > \theta_2) - 1 \\
&= 1 - (1+\theta_1)^{-\alpha} + 1 - (1+\theta_2)^{-\alpha} + (1+\theta_1+\theta_2)^{-\alpha} - 1 \\
&= 1 - (1+\theta_1)^{-\alpha} - (1+\theta_2)^{-\alpha} + (1+\theta_1+\theta_2)^{-\alpha} \\
&= \Psi(\theta_1, \theta_2, \alpha).
\end{aligned}$$

While the numerator is given by

$$\begin{aligned}
&P(X > x, X_2 \leq \theta_1, X_3 \leq \theta_2) \\
&= P(X_1 > x_1) - [P(X_1 > x_1, X_2 > \theta_1) + P(X > x, X_3 > \theta_2) \\
&- P(X > x, X_2 > \theta_1, X_3 > \theta_2)] \\
&= (1+x)^{-(\alpha+1)} - [(1+x+\theta_1)^{-(\alpha+1)} + (1+x+\theta_2)^{-(\alpha+1)} \\
&- (1+x+\theta_1+\theta_2)^{-(\alpha+1)}].
\end{aligned}$$

Hence we have

$$
\begin{aligned}
&P(X > x | X_2 \leq \theta_1, X_3 \leq \theta_2) \\
&= \frac{1}{\Psi(\theta_1, \theta_2, \alpha)}(1+x)^{-(\alpha+1)} - [(1+x+\theta_1)^{-(\alpha+1)} + (1+x+\theta_2)^{-(\alpha+1)} \\
&- (1+x+\theta_1+\theta_2)^{-(\alpha+1)}]. \qquad (5.29)
\end{aligned}
$$

The conditional density of X given $X_2 \leq \theta_1$ and $X_3 \leq \theta_2$, (from (5.29)) will be

$$
\begin{aligned}
&f_{X|X_2 \leq \theta_1, X_3 \leq \theta_2}(x) \\
&= \frac{\alpha}{\Psi(\theta_1, \theta_2, \alpha)}[(1+x)^{-(\alpha+1)} - ((1+x+\theta_1)^{-(\alpha+1)} + (1+x+\theta_2)^{-(\alpha+1)} \\
&- (1+x+\theta_1+\theta_2)^{-(\alpha+1)})]I(x \geq 0). \qquad (5.30)
\end{aligned}
$$

5.8.1 Maximum likelihood estimation

We draw a random sample of size n from the density in (5.30). Denote the observations by $X_1, X_2, \cdots, X_n$. In this case the likelihood function is given by

$$
\begin{aligned}
L = \prod_{i=1}^{n} \frac{\alpha}{\Psi(\theta_1, \theta_2, \alpha)}&[(1+X_i)^{-(\alpha+1)} - ((1+X_i+\theta_1)^{-(\alpha+1)} \\
&+ (1+X_i+\theta_2)^{-(\alpha+1)} - (1+X_i+\theta_1+\theta_2)^{-(\alpha+1)})].
\end{aligned}
$$

Equivalently the log-likelihood function takes the form

$$\log L = n \log \alpha - n \log \Psi(\theta_1, \theta_2, \alpha) + \sum_{i=1}^{n} \log[(1 + X_i)^{-(\alpha+1)}$$

$$- ((1 + X_i + \theta_1)^{-(\alpha+1)} + (1 + X_i + \theta_2)^{-(\alpha+1)} - (1 + X_i + \theta_1 + \theta_2)^{-(\alpha+1)})].$$

So the likelihood equation for α is

$$\frac{\partial}{\partial \alpha} \log L = 0.$$

Equivalently

$$\frac{n}{\alpha} - \frac{\frac{\partial}{\partial \alpha}\Psi(\theta_1, \theta_2, \alpha)}{\Psi(\theta_1, \theta_2, \alpha)} + \sum_{i=1}^{n} [((1 + X_i)^{-(\alpha+1)} - ((1 + X_i + \theta_1)^{-(\alpha+1)}$$

$$+ (1 + X_i + \theta_2)^{-(\alpha+1)}$$

$$- (1 + X_i + \theta_1 + \theta_2)^{-(\alpha+1)}))^{-1}(-(1 + X_i)^{-(\alpha+1)} \log(1 + X_i)$$

$$+ (1 + X_i + \theta_1)^{-(\alpha+1)} \log(1 + X_i + \theta_1)$$

$$+ (1 + X_i + \theta_2)^{-(\alpha+1)} \log(1 + X_i + \theta_2)$$

$$- (1 + X_i + \theta_1 + \theta_2)^{-(\alpha+1)} \log(1 + X_i + \theta_1 + \theta_2))] = 0. \qquad (5.31)$$

where

$$\frac{\partial}{\partial \alpha}\Psi(\theta_1, \theta_2, \alpha)$$
$$= (1+\theta_1)^{-\alpha}\log(1+\theta_1) + (1+\theta_2)^{-\alpha}\log(1+\theta_2)$$
$$- (1+\theta_1+\theta_2)^{-\alpha}\log(1+\theta_1+\theta_2).$$

The second likelihood equation is

$$\frac{\partial}{\partial \theta_1}\log L = 0.$$

Equivalently

$$n\frac{\frac{\partial}{\partial \theta_1}\Psi(\theta_1, \theta_2, \alpha)}{\Psi(\theta_1, \theta_2, \alpha)} + \sum_{i=1}^{n}[((1+X_i)^{-(\alpha+1)}$$
$$- ((1+X_i+\theta_1)^{-(\alpha+1)} + (1+X_i+\theta_2)^{-(\alpha+1)}$$
$$- (1+X_i+\theta_1+\theta_2)^{-(\alpha+1)}))^{-1}((1+X_i+\theta_1)^{-(\alpha+1)}\log(1+X_i+\theta_1))$$
$$- (1+X_i+\theta_1+\theta_2)^{-(\alpha+1)}\log(1+X_i+\theta_1+\theta_2)] = 0. \tag{5.32}$$

where

$$\frac{\partial}{\partial \theta_1}\Psi(\theta_1, \theta_2, \alpha) = (1+\theta_1)^{-\alpha}\log(1+\theta_1) - (1+\theta_1+\theta_2)^{-\alpha}\log(1+\theta_1+\theta_2).$$

Finally, we have

$$\frac{\partial}{\partial \theta_2}\log L = 0.$$

Equivalently

$$n\frac{\frac{\partial}{\partial\theta_2}\Psi(\theta_1,\theta_2,\alpha)}{\Psi(\theta_1,\theta_2,\alpha)}+\sum_{i=1}^{n}[((1+X_i)^{-(\alpha+1)}-((1+X_i+\theta_1)^{-(\alpha+1)}$$
$$+(1+X_i+\theta_2)^{-(\alpha+1)}$$
$$-(1+X_i+\theta_1+\theta_2)^{-(\alpha+1)}))^{-1}((1+X_i+\theta_2)^{-(\alpha+1)}\log(1+X_i+\theta_2))$$
$$-(1+X_i+\theta_1+\theta_2)^{-(\alpha+1)}\log(1+X_i+\theta_1+\theta_2)]=0. \tag{5.33}$$

where

$$\frac{\partial}{\partial\theta_2}\Psi(\theta_1,\theta_2,\alpha)=(1+\theta_2)^{-\alpha}\log(1+\theta_2)-(1+\theta_1+\theta_2)^{-\alpha}\log(1+\sigma_1+\theta_2).$$

5.8.2 Estimation by the method of moments

Let us consider for any $(r\geq 1)$

$$\begin{aligned}
&E[X^r]\\
&=\int_0^\infty x^r f_{X|X_2\leq\theta_1,X_3\leq\theta_2}(x)dx\\
&=\frac{\alpha}{\Psi(\theta_1,\theta_2,\alpha)}\int_0^\infty[(1+x)^{-(\alpha+1)}-((1+x+\theta_1)^{-(\alpha+1)}\\
&+(1+x+\theta_2)^{-(\alpha+1)}-(1+x+\theta_1+\theta_2)^{-(\alpha+1)})]dx\\
&=\frac{\alpha}{\Psi(\theta_1,\theta_2,\alpha)}[B(r+1,\alpha-r)(1-(1+\theta_1)^{r-\alpha}-(1+\theta_2)^{r-\alpha}\\
&+(1+\theta_1+\theta_2)^{r-\alpha})].
\end{aligned}\tag{5.34}$$

Next substituting $r=1,\frac{1}{2},\frac{1}{3}$, we get

- $E[X] = \frac{\alpha}{\Psi(\theta_1,\theta_2,\alpha)}[B(2, \alpha - 1)(1 - (1 + \theta_1)^{1-\alpha} - (1 + \theta_2)^{1-\alpha} + (1 + \sigma_1 + \theta_2)^{1-\alpha})]$.
- $E[X^{\frac{1}{2}}] = \frac{\alpha}{\Psi(\theta_1,\theta_2,\alpha)}[B(\frac{3}{2}, \alpha - \frac{1}{2})(1 - (1 + \theta_1)^{\frac{1}{2}-\alpha} - (1 + \theta_2)^{\frac{1}{2}-\alpha} + (1 + \theta_1 + \theta_2)^{\frac{1}{2}-\alpha})]$.
- $E[X^{\frac{1}{3}}] = \frac{\alpha}{\Psi(\theta_1,\theta_2,\alpha)}[B(\frac{4}{3}, \alpha - \frac{1}{3})(1 - (1 + \theta_1)^{\frac{1}{3}-\alpha} - (1 + \theta_2)^{\frac{1}{3}-\alpha} + (1 + \theta_1 + \theta_2)^{\frac{1}{3}-\alpha})]$.

Next we define the following quantity based on a random sample of size n drawn from above density in (5.30),

$$M_a = \frac{1}{n}\sum_{i=1}^{n} X_i^a,$$

which can be viewed as the a-th order sample raw moment. Next we equate the sample moments with the corresponding population moments to find the estimates of the parameters. In this case the resulting equations are as follows:

$$\frac{M_1}{M_{\frac{1}{2}}} = \frac{[B(2, \alpha - 1)(1 - (1 + \theta_1)^{1-\alpha} - (1 + \theta_2)^{1-\alpha} + (1 + \theta_1 + \theta_2)^{1-\alpha})]}{[B(\frac{3}{2}, \alpha - \frac{1}{2})(1 - (1 + \theta_1)^{\frac{1}{2}-\alpha} - (1 + \theta_2)^{\frac{1}{2}-\alpha} + (1 + \theta_1 + \theta_2)^{\frac{1}{2}-\alpha})]}, \tag{5.35}$$

$$\frac{M_{\frac{1}{2}}}{M_{\frac{1}{3}}} = \frac{[B(\frac{3}{2}, \alpha - \frac{1}{2})(1 - (1 + \theta_1)^{\frac{1}{2}-\alpha} - (1 + \theta_2)^{\frac{1}{2}-\alpha} + (1 + \theta_1 + \theta_2)^{\frac{1}{2}-\alpha})]}{[B(\frac{4}{3}, \alpha - \frac{1}{3})(1 - (1 + \theta_1)^{\frac{1}{3}-\alpha} - (1 + \theta_2)^{\frac{1}{3}-\alpha} + (1 + \theta_1 + \theta_2)^{\frac{1}{3}-\alpha})]}, \tag{5.36}$$

and

$$\frac{M_1}{M_{\frac{1}{3}}} = \frac{[B(2,\alpha-1)(1-(1+\theta_1)^{1-\alpha}-(1+\theta_2)^{1-\alpha}+(1+\theta_1+\theta_2)^{1-\alpha})]}{[B(\frac{4}{3},\alpha-\frac{1}{3})(1-(1+\theta_1)^{\frac{1}{3}-\alpha}-(1+\theta_2)^{\frac{1}{3}-\alpha}+(1+\theta_1+\theta_2)^{\frac{1}{3}-\alpha})]}. \tag{5.37}$$

Solving the above sets of 3 non-linear equations we will find the estimates of the parameters.

5.8.3 Estimation of the parameters using a simulation study

We draw random samples of sizes n=50, 100 and 200 from our density in (5.30), for a particular choice of the parameters which are $\alpha = 2.5$, $\theta_1 = 0.5$, and $\theta_2 = 0.5$. So that in this case

$$\Psi(\theta_1,\theta_2,\alpha) = 1-(1+\theta_1)^{-\alpha}-(1+\theta_2)^{-\alpha}+(1+\theta_1+\theta_2)^{-\alpha} = 0.270877.$$

Hence our density reduces to

$$\begin{aligned} &f_{X|X_2\leq\theta_1,X_3\leq\theta_2}(x) \\ &= \frac{2.5}{0.270877}[(1+x)^{-(\alpha+1)}-2(1+x+0.5)^{-3.5}+(2+x)^{-3.5}]I(x\geq 0). \end{aligned} \tag{5.38}$$

In the following tables we provide the estimates of the parameters using the maximum likelihood method and using the method of moments.

n	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$
50	0.4784212	0.4847938	2.431728
100	0.4831572	0.4857404	2.471352
200	0.4955152	0.4863323	2.489523

Table 5.4: Estimates of the parameters using the method of moments

n	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\alpha}$
50	0.4923689	0.4733277	2.45321768
100	0.4926100	0.515632	2.48431713
200	0.4928690	0.5012083	2.5003729

Table 5.5: Estimates of the parameters using the method of maximum likelihood

5.8.4 Comment on the simulation study

For our simulation study, we observe that our estimates for all the parameters are quite good under both of the estimation procedures and they are consistent in the sense that as we increase the sample size, the estimated values are closer to the true value of the parameters. The result of this small simulation study is quite encouraging and one can extend this idea to the more general trivariate $P(IV)$ model.

5.8.5 Bayesian estimation of the parameters under a non informative prior

In this situation we will consider the following choice of independent (vague) priors for each of our parameters and will perform a Bayesian analysis based on that. Let us consider

- $f(\alpha) \propto \frac{1}{\alpha} I(\alpha > 0)$.
- $f(\theta_1) \propto \frac{1}{\theta_1} I(\theta_1 > 0)$.

- $f(\theta_2) \propto \frac{1}{\theta_2} I(\theta_2 > 0)$.

So the joint posterior will be given by (for a random sample of size n drawn from the density in (5.30))

$$f(\alpha, \theta_1, \theta_2 | \underline{X} = \underline{x}) = f(\alpha) f(\theta_1) f(\theta_2) A^{-1} \prod_{i=1}^{n} \frac{\alpha}{\Psi(\theta_1, \theta_2, \alpha)}$$
$$[(1 + X_i)^{-(\alpha+1)} - (1 + X_i + \theta_1)^{-(\alpha+1)}$$
$$- (1 + X_i + \theta_2)^{-(\alpha+1)} + (1 + X_i + \theta_1 + \theta_2)^{-(\alpha+1)}] \tag{5.39}$$

where A is the normalizing constant given by

$$A = \iiint_{(\mathbb{R}^+)^3} f(\alpha, \theta_1, \theta_2 | \underline{X} = \underline{x}) d\theta_1 d\theta_2 d\alpha.$$

So the marginal posterior density of α is given by

$$f(\alpha | \underline{X} = \underline{x}) = \frac{\int_0^\infty \int_0^\infty f(\alpha, \theta_1, \theta_2 | \underline{X} = \underline{x}) d\theta_1 d\theta_2}{A}.$$

Similarly the marginal posterior density of θ_1 is given by

$$f(\theta_1 | \underline{X} = \underline{x}) = \frac{\int_0^\infty \int_0^\infty f(\alpha, \theta_1, \theta_2 | \underline{X} = \underline{x}) d\alpha d\theta_2}{A}.$$

And the marginal posterior density of θ_2 is given by

$$f(\theta_2 | \underline{X} = \underline{x}) = \frac{\int_0^\infty \int_0^\infty \Pi(\alpha, \theta_1, \theta_2 | \underline{X} = \underline{x}) d\alpha d\theta_1}{A}.$$

5.8.6 Comment on the choice of Priors

Suppose that we have some specific information (in the form of prior belief) about any one of the parameters(say α) such as that it can take any values between (say) 0 and 2. Then a reasonable choice of prior distribution for α would be any kind of flat prior, (say) a uniform distribution with support (0,2). This will reduce the complexity in the posterior analysis. Although the use of non-informative and /or dependent priors will increase the complexity of the analysis, one may still want to consider use of such priors because of the fact that for analytically intractable models like ours, the corresponding posterior analysis can be efficiently performed by a Markov Chain Monte Carlo (MCMC) algorithm which is specifically designed for complicated models.

5.8.7 Posterior simulation study

First of all we draw random samples of size 50 and 100 from our density for a particular choice ($\alpha = 1.5, \sigma_1 = 0.3, \theta_2 = 0.3$) of the parameters. For the jumping distributions we consider gamma distributions but with different shape and scale parameters. The posterior analysis is based on the posterior modes for each of those three parameters. The corresponding posterior density plots are included in this chapter also. The following table shows various choices for the starting distribution, the choices of the parameters of the jumping distribution along with posterior modes:

Initial choices of the parameters:$\alpha = 1.07, \sigma_1 = 0.47, \theta_1 = 0.43, \theta_2 = 0.47$.

Jumping distribution of the parameters:

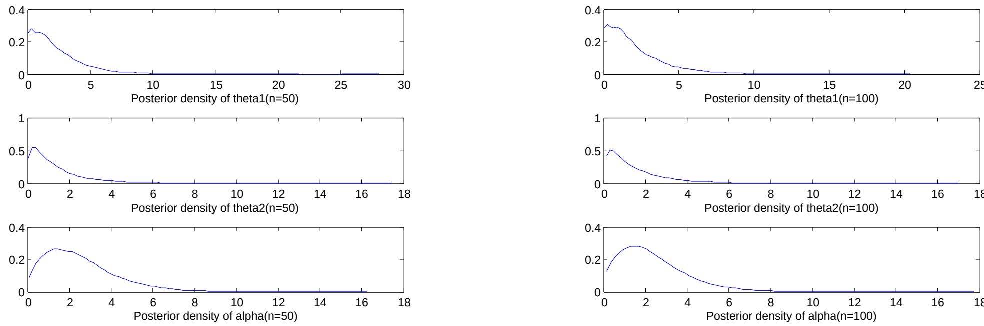
- $\alpha \sim \Gamma(8.1/5, 6.1/5)$
- $\theta_1 \sim \Gamma(1.8/3, 7.2/3)$
- $\theta_2 \sim \Gamma(3.2/3, 6.1/3)$

In the following table for each gamma density the first value corresponds to the shape parameter and the next value is the scale parameter. In the following table the posterior modes of all the parameters are displayed.

n	Mode(α)	Mode(θ_1)	Mode(θ_2)
50	1.4343	0.2813	0.3900
100	1.4744	0.2753	0.3063

Table 5.6: Bayesian estimates of the Parameters

The corresponding marginal posterior density plots for the three parameters based on samples of sizes 50 and 100 are displayed in Figures (5.1 a) and (5.1 b).



(a) Posterior Density of the parameters for $n = 50$ (b) Posterior Density of the parameters for $n = 100$

Figure 5.1: Posterior density for all the parameters for different choices of the sample size

5.8.8 Remarks on the posterior simulation study

For the trivariate hidden truncation model, we observe that with a sample size of 100, the estimates of the parameters based on the Bayesian posterior modes are quite good. However when the sample size is 50, the posterior mode for θ_2 is quite far away from the true value of the parameter. More informative priors (for example a proper prior, or a flat prior for the index parameter α only) might result in a substantial amount of improvement in our posterior modal values.

Appendix

Asymptotic distribution of the smallest order statistic

Suppose we draw a random sample of size n from the density (5.11). Denote the observations by $\underline{X}^1, \underline{X}^2, \ldots, \underline{X}^n$, where $\underline{X}^i$, for $i = 1, 2, \cdots, n$, are $(k-1)$ dimensional. Then we define

$$\underline{Y} = \min_{1 \le j \le n} \underline{X}^j.$$

Let us consider

$$\begin{aligned} P(\underline{Y} > \underline{y}) &= P(\underline{X}^1 > \underline{y}, \underline{X}^2 > \underline{y}, \ldots, \underline{X}^n > \underline{y}) \\ &= [P(\underline{X}^1 > \underline{y})]^n \\ &= [\frac{1}{\psi(\alpha, c)}[[1 + \sum_{i=1}^{k-1}(\frac{y_i - \mu_i}{\sigma_i})]^{-\alpha} \\ &- [1 + \sum_{i=1}^{k-1}(\frac{y_i - \mu_i}{\sigma_i}) + (\frac{c - \mu_k}{\sigma_k})]^{-\alpha}]]^n. \end{aligned}$$

However for the notational simplicity we consider $(\frac{c-\mu_k}{\sigma_k}) = \theta$. and also we assume that $\mu_i = 0, \forall i = 1, 2, \ldots, k$ so that $\psi(\alpha, c) = 1-(1+\theta)^{-\alpha} = \psi(\alpha, \theta)$.

Thus

$$P(\underline{Y} > \frac{1}{n}\underline{y}) = [\frac{1}{\delta_1(\alpha, \theta)}[[1+\sum_{i=1}^{k-1}(\frac{y_i}{n\sigma_i})]^{-\alpha} - [1+\sum_{i=1}^{k-1}(\frac{y_i}{n\sigma_i})+\theta]^{-\alpha}]]^n. \quad (5.40)$$

Next consider the following:

$$[1+\sum_{i=1}^{k-1}(\frac{y_i}{n\sigma_i})]^{-\alpha}-[1+\sum_{i=1}^{k-1}(\frac{y_i}{n\sigma_i})+\theta]^{-\alpha}$$

$$
\begin{aligned}
&= [1-\frac{\alpha}{n}(\sum_{i=1}^{k-1}(\frac{y_i}{\sigma_i}))+\frac{\alpha(\alpha+1)}{2}(\sum_{i=1}^{k-1}(\frac{y_i}{n\sigma_i}))^2 \\
&- \ldots]-[(1+\theta)^{-\alpha}-\frac{\alpha(1+\theta)^{-\alpha+1}}{n}(\sum_{i=1}^{k-1}(\frac{y_i}{\sigma_i})) \\
&+ \frac{\alpha(\alpha+1)(1+\theta)^{-\alpha+2}}{2}(\sum_{i=1}^{k-1}(\frac{y_i}{n\sigma_i}))^2-\ldots] \\
&= \delta_1(\alpha,\theta)-\frac{\alpha}{n}(1-(1+\theta)^{-\alpha+1})(\sum_{i=1}^{k-1}(\frac{y_i}{\sigma_i}))+o(n^{-2}). \quad (5.41)
\end{aligned}
$$

So that

$$P(\underline{Y}>\frac{1}{n}\underline{y})=[1-\frac{\alpha}{n}(1-(1+\theta)^{-\alpha+1})(\sum_{i=1}^{k-1}(\frac{y_i}{\sigma_i}))+o(n^{-2})]^n.$$

Hence

$$
\begin{aligned}
\lim_{n\to\infty}P(\underline{Y}>\frac{1}{n}\underline{y}) &= \lim_{n\to\infty}[1-\frac{\alpha}{n}(1-(1+\theta)^{-\alpha+1})(\sum_{i=1}^{k-1}(\frac{y_i}{\sigma_i}))+o(n^{-2})]^n \\
&= \exp[-B(\sum_{i=1}^{k-1}(\frac{y_i}{\sigma_i}))], \quad (5.42)
\end{aligned}
$$

for $\underline{y}>\underline{0}$ where $B=\alpha(1-(1+\theta)^{-\alpha+1})$. So the asymptotic distribution of the vector of smallest order statistics when the samples are drawn from a hidden truncated (from above) k-variate $P(II)$ distribution has independent exponential marginals.

Chapter 6

Conclusion

6.1 Introduction

In this chapter we will discuss in detail plausible reasons for considering hidden truncation paradigms for bivariate and multivariate Pareto data. We will also describe the challenges that have been faced during the course of study of such hidden truncation models, and what effective steps have been taken into consideration regarding inference for such models. A broad spectrum of flexible univariate and multivariate models can be constructed by a hidden truncation paradigm. Such models can be viewed as being characterized by a basic marginal density, a family of conditional densities and a specified hidden truncation point. Skewed multivariate distributions can arise in situations in which the observed variables represent a sample that has been truncated with respect to one or more hidden covariable. In chapters 3 through 5, a survey of such models arising from a bivariate and multivariate Pareto distributions has been provided with discussion on related inference questions. In particular we consider the fact that, in income modeling, quite often there are instances of unreported income. In such a situation it is reasonable to assume that the true income (say, X), the observable income (Y, say) and the unreported income (U, say) are related in the following way: $X = Y + U$.

Such a model was considered by Krishnaji (1970) and later on by Hartley and Revankar(1974) and Hinkley and Revankar (1977). Now, as mentioned earlier, we focus on in particular situations where the income data for an individual is available if and only the unreported income does not exceed a certain value which could be any value within the support set of the conditioning variable. Next we focus on inference for such models and in particular estimation of all the parameters of the model. First of all for hidden truncation models, it is not possible to construct unbiased estimates nor to consider a routine classical approach on inferential aspects because of the following:

- The models do not constitute one parameter exponential families of distributions.
- The moment generating function does not exist in these cases.

Because of the form of the density, as in (3.14) or in (4.5), we have observed that for certain choices of the truncation parameter (in particular for the bivariate case $\theta \geq 2$, and for the trivariate hidden truncation case ($\min(\theta_1, \theta_2) \geq 0.7$) and also for the index of inequality in the range ($\alpha \geq 5$), our densities are indistinguishable. Furthermore since such models are not members of one parameter exponential family, no reduction in the complexity of the data is possible by invoking sufficiency arguments. For the various estimation strategies that we have discussed in earlier chapters for hidden truncation Pareto families, it would be really difficult to claim whether one is superior to another although for both types of models (i.e; both bivariate and multivariate hidden truncation) for small samples the performance of the maximum likelihood estimation procedure is not that good. A possible reason could be that the likelihood functions associated with these type of models often do not have easily identified modes because of the unavailability of analytical expressions for the maximum likelihood estimates. Furthermore in all other estimation procedures, the estimates of the parameters have been obtained numerically. Also because of the lack of analytical tractability of expected moments for the likelihood estimates, only the observed Fisher Information Matrix is available. As a consequence the asymptotic distribution of the maximum likelihood estimates for hidden truncated Pareto models is known only approximately. But we have observed chapters 3 through 5, that the asymptotic distribution of the smallest order statistic when the samples are taken from those hidden truncated distributions, is exponential so in large samples it belongs to a well known family of density for which well known results are available in terms of estimation and testing for the location parameter.

6.2 Discussion on the performance of Bayesian analysis for hidden truncation Pareto models

We now focus on the performance of Bayesian Analysis for the hidden truncated Pareto models. The complicated form of the likelihood as in (3.29) and also in (5.19) is a warning that friendly conjugate priors will not be encountered. As a consequence, little attention has been devoted in this direction. In all our Bayesian analysis for such types of hidden truncation models we have proposed non-informative priors for all the parameters, which seems to be the only plausible choice in implementing a Bayesian approach. Such a method predictably yields results that are very similar to those obtained using the method of maximum likelihood. However the only difference will be in their interpretation. In our case it would be unrealistic to think of conditional priors for all the parameters α, σ and θ. However one might consider a situation where some prior information on say, the index of inequality parameter α is provided, perhaps that it can only take any value between 1 and 2. Then we might consider any kind of flat prior on α say for example $\alpha \in U(1, 2)$ and subsequently we can use this information for our posterior analysis for the remaining parameters. Although again this will be a rare situation. The Bayesian estimates of the parameters based on posterior modes are quite good for all the hidden truncated bivariate and multivariate Pareto models except for small sample sizes where the estimates for the truncation parameter is quite far away from the true value of the parameter. A more extensive simulation study will be required to check whether the behavior of this estimate is artifactual when a small sample (for example a random sample of size 50) is drawn from the density as in (3.14) or in (4.5).

6.3 Future work

As mentioned in chapter 3, the $P(II)$ model is a viable competitor of the log-normal distribution which is very useful for survival data analysis. Thus it would be natural to investigate the application of hidden truncation $P(II)$ or the more general $P(IV)$ model in survival models in addition to income data modeling. The role of multivariate hidden truncation models which we have discussed in chapter 5 can not be ignored either. Efforts will be given to derive hidden truncation models for the multivariate $P(II)$ and also for multivariate $P(IV)$ when all parameters are unknown. Also we will consider separately the two situations : (1) Single variable truncation from above and (2) k_1-variable (where $k_1 \leq k, k_1 \geq 2$) truncation from above, for a k variate $P(II)$ and $P(IV)$ distribution. In chapter 5 we have discussed both types of hidden truncation but we have considered a simple situation for the types of hidden truncation where each of the scale parameters were considered to be one while the location parameters were considered to be zero. Undoubtedly one can imagine that for a more general family of multivariate $P(II)$ or $P(IV)$ distributions, the estimation of all the parameters will have to be performed numerically as was done for the hidden truncated bivariate $P(II)$ and $P(IV)$ models earlier. Furthermore, so far we have considered one sided hidden truncation (from above). It would be natural to think of two sided hidden truncation and to investigate first of all whether or not the resulting model augments the original distribution. And if that happens then it would be interesting to see whether we can find an application of such models in real life situations. Also efforts will be made to find more applicability of such models beyond the economic sphere and in particular in conditional survival analysis and also in stress and strength analysis. Needless to say much work remains to be done before the more complicated hidden truncation models can become useful tools for the applied statisticians working in this area.

Bibliography

[1] Aczel, J. (1966). *Lectures on Functional Equations and their Applications.* Academic Press, New York.

[2] Arnold, B.C. (1983). *Pareto Distributions.* International Cooperative Publishing House, Fairland, MD.

[3] Arnold, B.C. (1997). Characterizations involving conditional specification. *Journal of Statistical Planning and Inference,* 63, 117-131.

[4] Arnold, B.C. and Beaver, R.J. (2000). Hidden truncation models. *Sankhya Series A*, 62, 23-35.

[5] Arnold, B.C., Castillo, E., and Sarabia, J.M. (1999). *Conditional Specification of Statistical Models.* Springer, New York.

[6] Arnold, B.C., Castillo, E., and Sarabia, J.M. (2002). Exact and near compatibility of discrete conditional distributions. *Journal of Computational Statistics and Data Analysis*, 40, 231-252.

[7] Arnold, B.C. and Ghosh, I. (2010). Inference for Pareto data subject to hidden truncation. *Journal Of The Indian Society for Probability And Statistics*, 12, 1-16.

[8] Arnold, B.C. and Ghosh, I. (2011). Hidden truncation in multivariate Pareto data. To appear in *Contemporary Mathematics with Applications,* Asian Books Private Ltd, Kolkata.

[9] Arnold, B.C. and Gokhale, D.V. (1994). On uniform marginal representations of contingency tables. *Statistics and Probability Letters*, 21, 311-316.

[10] Arnold, B.C. and Gokhale, D.V. (1998). Distributions most nearly compatible with given families of conditional distributions. The finite discrete case, *Test, 7*, 377-390.

[11] Arnold, B.C. and Laguna, L. (1977). *On Generalized Pareto Distributions with Applications to Income Data.* International Studies in Economics, Monograph 10, Department of Economics, Iowa State University, Ames, Iowa.

[12] Arnold, B.C. and Press, S.J. (1983). Bayesian inference for Pareto populations. *Journal of Econometrics*, 21, 287-306.

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

[14] Arnold, B.C. and Strauss, D. (1991). Bivariate distributions with conditionals in prescribed exponential families. *Journal of the Royal Statistical Society, Ser.B,* 53, 365-375.

[15] Cox, D.R. (1958). Some problems connected with statistical inference. *Annals of Mathematical Statistics*, 29, 357-372.

[16] Efron, B. and Hinkley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 65, 457-482.

[17] Feder, P.I. (1968). On the distribution of the log likelihood ratio test statistic when the true parameter is near the boundaries of the hypothesis regions. *Annals of Mathematical Statistics,* 39, 2044-2055.

[18] Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995). *Bayesian Data Analysis.* Chapman and Hall, New York.

[19] Hanagal, D.D. (1996). A multivariate Pareto distribution. *Communications in Statistics-Theory and Methods*, 25, 1471-1488.

[20] Kotz, S., Balakrishnan N., and Johnson N.L. (2006). *Continuous Multivariate Distributions, Vol.I.* Wiley, New York.

[21] Kullback, S. (1968). *Information Theory and Statistics.* Wiley, New York.

[22] Lehman, E.L., and Casella G. (2001). *Theory Of Point Estimation.* Springer, New York.

[23] Mardia, K.V. (1962). Multivariate Pareto distributions. *Annals of Mathematical Statistics*, 33, 1008-1015.

[24] Pareto, V. (1897). *Cours d'Economie Politique, Vol.II.* F.Rouge, Lausanne.

[25] Quandt, R.E. (1966). Old and new methods of estimation and the Pareto distribution. *Metrika,* 10, 55-82.

[26] Read, T.R.C. and Cressie, N.A.C. (1998). *Goodness-of-fit Statistics for Discrete Multivariate Data.* Springer, New York.

[27] Self, S.G. and Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *Journal of the American Statistical Association,* 82, 605-610.

[28] Singh, S.K. and Maddala, G.S. (1976). A function for size distribution of incomes. *Econometrika*, 44, 963-970.

[29] Tajvidi, N. (1996). Characterization and some statistical aspects of univariate and multivariate generalized Pareto distributions. *Technical report, Department of Mathematics, Chalmers University of Technology,* Gothenburg, Sweden.

[30] Targhetta, M.L. (1979). Confidence intervals for a one parameter family in a mixture of distributions. *Biometrika*, 65, 687-688.