Bayesian inference for the bivariate Pareto((II)) distribution under the hidden truncation paradigm (from above)

Indranil Ghosh [1]

Austin Peay State University, TN 37044, USA

ghoshi@apsu.edu

Abstract

Data from a hidden truncated Pareto(II) distribution are to be used to make inferences about the inequality, precision and truncation parameters. Two different types of dependent prior analyses are reviewed and compared with each other. It is argued that mathematical tractability should be, perhaps, a minor consideration in choosing the appropriate choices of the hyperparameters for the prior densities. Some illustrative examples are provided.

1 Introduction

Income and wealth data is typically modeled using some variant of the classical Pareto distribution. In practice, it is frequently likely that the observed data has been truncated with respect to some unobserved covariable. In this paper a hidden truncation formulation of this scenario is considered. A bivariate Pareto (II) distribution is assumed for the variable of interest and the unobserved covariable. Thus we are dealing with data from a hidden truncated Pareto(II) distribution with shape or inequality parameter α, precision parameter τ, and truncation parameter θ. In this case the natural parameter space is $\Omega = (\alpha > 0, \tau > 0, \theta > 0)$. Two different types of conditional prior densities have been taken into consideration. In one, dependent priors for the inequality and precision parameter are assumed together with an independent prior for the truncation parameter. While for the second approach, we consider a full conditional prior set-up for all the three parameters. This article is concerned with inference using both of these proposed priors. This paper draws heavily on the works published earlier in Arnold and Press (1983, 1986, 1989). In the Arnold and Press papers, discussion dealt with Bayesian inference for the classical Pareto family. In contrast we focus our attention on associated Bayesian inference for a Pareto(II) distribution under the hidden truncation paradigm. The approach advocated in the present paper involves use of conditionally specified prior families. This is attained as follows. Suppose we have k-parameters $\underline{\zeta} = (\zeta_1, \zeta_2, \cdots, \zeta_k)$. A convenient notation for the k-dimensional case dealing with $\underline{\zeta}$ involves use of the symbol $\underline{\zeta}^{(i)}$ to denote the vector $\underline{\zeta}$ with the i-th co-ordinate deleted, $i = 1, 2, \cdots, k$. First suppose that for a fixed i, $\underline{\zeta}^{(i)}$ is known. We then determine, under such circumstances, a natural conjugate prior family for ζ_i , say $f(\zeta_i|\underline{\tau}^{(i)})$ where $\underline{\tau}^{(i)}$ are hyperparameters. We then seek a joint prior in the family of joint distribution for $\underline{\zeta}$ in which, for each i, the conditional distribution of ζ_i given $\underline{\zeta}^{(i)}$ will be a member of the family $f(\zeta_i|\underline{\tau}^{(i)})$ with hyperparameters $\underline{\tau}^{(i)}$ which may depend on $\underline{\zeta}^{(i)}$. The number of hyperparameters may be large, but this is a price we have to pay to avoid facing a possibly un-natural dependence structure in the class of priors we use to match our informed expert's belief. In this paper suggestions will be made regarding efficient ways of eliciting values of the hyperparameters. This paper is organized in the following way. In section 2, we will discuss in detail the concept of hidden truncation. In section 3, we will consider the hidden truncation density for a bivariate Pareto(II) model. In section 4, we consider the choice of priors in which the inequality and

[1]**Key Words and Phrases: bivariate hidden truncated $P(II)$ distribution, conditionally specified priors, hyperparameters.**

precision parameters are dependent but independent of the truncation parameter. In section 5, we consider a reasonable way to choose hyperparameters for the prior distribution with an illustrative example. In section 6, we consider a real life application of our proposed procedure. In section 7, we consider the full conditional prior set-up in which we have conditionally specified priors for all the parameters along with an example. In section 8, we provide a short discussion on the use of locally uniform or diffuse priors in such situations.

2 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

$$f_{X|Y \in M}(x) = f_X(x) \frac{P(Y \in M|X = x)}{P(Y \in M)}. \tag{1}$$

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

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

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

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 (1) reduces to

$$f_{c-}(x) = f_X(x) \frac{P(Y \leq c|X = x)}{P(Y \leq c)}. \tag{2}$$

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. In the expression in (2) we are actually using $P(Y \leq c|X = x)$ for one value of c. Moreover, $P(Y \leq c|X = x)$ could be any function $g_c(x)$, satisfying $0 < g_c(x) < 1$ which is an increasing function in c so that for the denominator in (2) we can define

$$P(Y \leq c) = \int_{-\infty}^{\infty} g_c(x) f_X(x) dx. \tag{3}$$

In this case we can rewrite equation (2) as

$$f_{c-}(x) = f_X(x) \frac{g_c(x)}{\int_{-\infty}^{\infty} g_c(x) f_X(x) dx}.$$

3 Hidden truncated density for the bivariate Pareto model

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

$$P(X > x, Y > y) = [1 + (\frac{x - \mu_1}{\sigma_1}) + (\frac{y - \mu_2}{\sigma_2})]^{-\alpha}, x \geq \mu_1, y \geq \mu_2, \tag{4}$$

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. A convenient source for discussion of bivariate and multivariate Pareto models is Arnold(1983). Also see Johnson, Kotz and Balakrishnan(2000, Chapter 52).

The joint distribution function of X and Y is given by

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

and the joint density of (X,Y) is given by

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\partial}{\partial x}\frac{\partial}{\partial y}[F(x,y)] \\ &= \frac{\alpha(\alpha+1)}{\sigma_1\sigma_2}[1 + (\frac{x-\mu_1}{\sigma_1}) + (\frac{y-\mu_2}{\sigma_2})]^{-(\alpha+2)} I(x \geq \mu_1, y \geq \mu_2). \end{aligned} \quad (5)$$

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

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\ &= \frac{(\alpha+1)}{\sigma_2(1+(\frac{x-\mu_1}{\sigma_1}))}[1 + \frac{(\frac{y-\mu_2}{\sigma_2})}{1+(\frac{x-\mu_1}{\sigma_1})}]^{-(\alpha+2)} I(y \geq \mu_2). \end{aligned} \quad (6)$$

after some algebraic simplification. 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 truncated density of X given that $Y \geq b$ with the condition that $b > \mu_2$ is given by

$$f^{HT}_{X|Y\geq b}(x) = f_X(x)\frac{P(Y \geq b|X = x)}{P(Y \geq b)} I(x \geq b). \quad (7)$$

In the bivariate Pareto(II) 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 (7) we get the following

$$f^{HT}_{X|Y\geq b}(x) = \frac{\alpha}{\sigma_1(1+(\frac{b-\mu_2}{\sigma_2}))}(1 + \frac{\frac{x-\mu_1}{\sigma_1}}{1+(\frac{b-\mu_2}{\sigma_2})})^{-(\alpha+1)} I(x \geq \mu_1). \quad (8)$$

From the above it is quite clear that the conditional distribution of X given $Y \geq b$ is again a member of a Pareto(II) family i.e.,

$$X|Y \geq b \sim \text{Pareto(II)}(\mu_1, \sigma_1^* = \sigma_1(1 + (\frac{b-\mu_2}{\sigma_2})), \alpha).$$

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

$$f^{HT}_{X|Y\leq c}(x) = \frac{\alpha}{\sigma_1(1-(1+(\frac{c-\mu_2}{\sigma_2}))^{-\alpha})}[(1 + (\frac{x-\mu_1}{\sigma_1}))^{-(\alpha+1)} - (1 + (\frac{(x-\mu_1)}{\sigma_1}) + (\frac{c-\mu_2}{\sigma_2}))^{-(\alpha+1)}] I(x \geq \mu_1). \quad (9)$$

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

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

where $\theta = \frac{c-\mu_2}{\sigma_2} > 0$. With this notation, the density can be written as

$$f^{HT}_{X|Y\leq\theta}(x) = \frac{\alpha}{\sigma_1\psi(\alpha,\theta)}[(1+\frac{x}{\sigma_1})^{-(\alpha+1)} - (1+\frac{x}{\sigma_1}+\theta)^{-(\alpha+1)}]I(x\geq 0). \tag{10}$$

It is evident from the density in (10) that the hidden truncated density involving upper truncation really augments the original model for X. This is in contrast to the fact that lower truncation does not augment the original model for X. Hidden truncated densities involving lower truncation are simply Pareto(II) densities with a revised set of parameters. In succeeding sections we address the problem of estimating and making inferences under the Bayesian paradigm about all the parameters (involved in the model) based on a sample of n i.i.d. observations $X_1, X_2, \cdots, X_n$ with common density (10). Note that the case $\theta = 0$ corresponds to the situation in which no hidden truncation has occurred.

4 Choice of dependent priors for the inequality and shape parameters and an independent prior for the truncation parameter

Let us consider a random sample of size n from the density (10), denoted by $X_1, X_2, \cdots, X_n$. Then the corresponding likelihood will be

$$L(\alpha,\tau,\theta) = \prod_{i=1}^{n} \frac{\alpha\tau}{\psi(\alpha,\theta)}[(1+X_i\tau)^{-(\alpha+1)} - (1+X_i\tau+\theta)^{-(\alpha+1)})], \tag{11}$$

where $\tau = \frac{1}{\sigma_1}$. 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. 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

$$f(\alpha|\tau) \propto \alpha^{\xi_1(\tau)-1}\exp(-\alpha\lambda_1(\tau))I(\alpha>0). \tag{12}$$

- The conditional density of τ given α is a Gamma distribution with shape parameter $= \xi_2(\alpha)$ and intensity parameter $= \lambda_2(\alpha)$. So that the conditional density of τ given α will be

$$f(\tau|\alpha) \propto \tau^{\xi_2(\alpha)-1}\exp(-\tau\lambda_2(\alpha))I(\tau>0). \tag{13}$$

- 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 gamma family. So the joint density (alternatively the joint prior) will be of the form (Arnold, Castillo and Sarabia (1999))

$$\begin{aligned} f(\alpha,\tau) &\propto \exp[-\alpha c_{11} + c_{12} + c_{13}\log(1+\alpha) + \log\tau(-\alpha c_{21} + c_{22} + c_{23}\log\alpha) \\ &- \tau(-\alpha c_{31} + c_{32} + c_{33}\log\alpha)]I(\alpha>0)I(\tau>0), \end{aligned} \tag{14}$$

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

$$
\begin{aligned}
& f(\alpha,\tau,\theta|data) \\
& \propto L(\alpha,\tau,\theta) \times f(\alpha,\tau) \times f(\theta) \\
& \propto \prod_{i=1}^{n} [\frac{\alpha\tau}{\psi(\alpha,\theta)}((1+x_i\tau)^{-(\alpha+1)} - (1+x_i\tau+\theta)^{-(\alpha+1)})] \\
& \times \exp[-\alpha c_{11} + c_{12} + c_{13}\log(1+\alpha) + \log\tau(-\alpha c_{21} + c_{22} + c_{23}\log\alpha) \\
& - \tau(-\alpha c_{31} + c_{32} + c_{33}\log\alpha)] \times \frac{1}{\theta} I(\alpha>0)I(\tau>0)I(\theta>0).
\end{aligned} \tag{15}
$$

Our proposed conditionally specified priors include, as special cases, independent gamma marginals in (14). This corresponds to initially setting $c_{21} = c_{23} = c_{31} = c_{33} = 0$. This particular approach has been advocated by those who consider marginal assessment of prior beliefs to be most viable (see e.g., Press (2003)). However, the use of the full family of conditionally specified priors comes with a price that must be paid for the flexibility of this family of priors. We have in total 8 hyperparameters to assess. But this problem of assessment, as we shall discuss in the next section, is not really an insurmountable task to perform.

5 Choice of the hyperparameters

In our case the conditional density of α given τ , for each fixed value of τ (from (14))

$$
\begin{aligned}
f(\alpha|\tau) &= \frac{f(\alpha,\tau)}{f(\tau)} \\
&= \alpha^{-\tau c_{33}+c_{13}-c_{23}\log\tau} \exp[-\alpha(c_{11} - \tau c_{31} + c_{21}\log\tau)] \\
&\times (c_{11} - \tau c_{31} + c_{21}\log\tau)^{-\tau c_{33}+c_{13}-c_{23}\log\tau+1} \\
&\times [\Gamma(-\tau c_{33} + c_{13} - c_{23}\log\tau + 1)]^{-1} I(\alpha>0).
\end{aligned} \tag{16}
$$

Equivalently the conditional distribution is α given τ is Gamma with shape$=-\tau c_{33} + c_{13} + c_{23}\log\tau$ and intensity$=c_{11} - \tau c_{31} + c_{21}\log\tau$.

So that for any real $r \geq 1$ we will have,

$$
\begin{aligned}
& E(\alpha^r|\tau) \\
&= \frac{\Gamma(-\tau c_{33} + c_{13} - c_{23}\log\tau + 1 + r)}{(c_{11} - \tau c_{31} + c_{21}\log\tau)^{-\tau c_{33}+c_{13}+c_{23}\log\tau+1+r} \times \Gamma(-\tau c_{33} + c_{13} + c_{23}\log\tau + 1)}.
\end{aligned} \tag{17}
$$

Again the conditional density of τ given α , for each fixed value of α is given by (from (14))

$$
\begin{aligned}
f(\tau|\alpha) &= \frac{f(\alpha,\tau)}{f(\alpha)} \\
&= \tau^{c_{22}+c_{23}\log\alpha-\alpha c_{21}} \exp[-\tau(-\alpha c_{31} + c_{32} + c_{33}\log\alpha)] \\
&\times (-\alpha c_{31} + c_{32} + c_{33}\log\alpha)^{c_{22}+c_{23}\log\alpha-\alpha c_{21}+1} \\
&\times [\Gamma(c_{22} + c_{23}\log\alpha - \alpha c_{21} + 1)]^{-1} I(\tau>0).
\end{aligned} \tag{18}
$$

Equivalently the conditional distribution is τ given α is Gamma with shape$=c_{22} + c_{23}\log\alpha - \alpha c_{21}$ and intensity$=-\alpha c_{31} + c_{32} + c_{33}\log\alpha$.

So that for any real $r \geq 1$ we will have,

$$
\begin{aligned}
& E(\tau^r|\alpha) \\
&= \frac{\Gamma(c_{22} + c_{23}\log\alpha - \alpha c_{21} + 1 + r)}{(-\alpha c_{31} + c_{32} + c_{33}\log\alpha)^{c_{22}+c_{23}\log\alpha-\alpha c_{21}+r+1} \times \Gamma(c_{22} + c_{23}\log\alpha - \alpha c_{21} + 1)}.
\end{aligned} \tag{19}
$$

Note: Here the notation Γ would imply a Gamma distribution and from hereafter whenever it appears it would mean the same as mentioned above. If we insist on proper prior densities, then there are constraints which must be imposed on the parameters in (14) to ensure that certain parameters appearing in the conditional densities are always positive to yield proper conditional densities. Adequate constraints are: $c_{11} > 0, c_{21} > 0, c_{31} < 0, c_{32} > 0, c_{33} > 0, c_{22} + c_{23} - c_{21} > -1,$ and $c_{13} + c_{23} - c_{33} > -1$. From the above it is quite clear that we have the complete knowledge of conditional distribution of τ given α and α given τ for each fixed values of α and τ respectively. For a conditionally specified priors such as (16) and (18), it is quite natural to try to match conditional moments whose approximate values will be provided by our informed expert who has collected the data. In our case 8 such conditional moments will suffice to evaluate all the hyperparameters. We propose more generally to ask the experimenter to provide prior values for more than 8 conditional moments. We recognize the fact that it is highly unlikely that such choices of prior values will be consistent and what we propose is to select a prior of the form (14) that will have conditional moments that are minimally disparate from those provided apriori by the expert. Suppose that prior assessed values for the conditional mean and variances for several different choices of the precision τ and for different given choices of the index of inequality α are solicited. Thus the experimenter provides his/her best guesses for the quantities (the subscript AV stands for assessed value)

$$E_{AV}(\alpha|\tau = \tau_i) = \delta_i, i = 1, 2, \cdots, m, \tag{20}$$

$$Var_{AV}(\alpha|\tau = \tau_i) = \eta_i, i = 1, 2, \cdots, m, \tag{21}$$

$$E_{AV}(\tau|\alpha = \alpha_j) = \beta_j, j = 1, 2, \cdots, l \tag{22}$$

$$Var_{AV}(\tau|\alpha = \alpha_j) = \chi_j, j = 1, 2, \cdots, l, \tag{23}$$

where $2m + 2l \geq 8$. Note that $[\tau_i]_{i=1}^m$ and $[\alpha_j]_{j=1}^l$ are known quantities. If indeed a density of the form (14) approximates the joint distribution of (α, τ) then the values of the conditional moments (20)-(23) will be well approximated by expressions derived from equations(18), (20). One reasonable approach (since exact equality is unlikely to be possible for any choice of the parameters $\underline{c}$) is to set up as an (although seemingly arbitrary) objective function the sum of squared differences between the left and right hand sides of (20)-(23),($2m+2l$ terms in total) and using a convenient optimization program to minimize this objective function. The assessed prior would then be (14) with this choice of parameters.

Note: We know that if $X \sim Gamma(shape = \alpha, intensity = \lambda)$ then $E(X) = \frac{\alpha}{\lambda}$ and $Var(X) = \frac{\alpha}{\lambda^2}$, so we can write $\frac{[E(X)]^2}{Var(X)} = \alpha$ and $\frac{E(X)}{Var(X)} = \lambda$, and we will use this idea in defining the objective function to be minimized for choosing appropriate values of c_{ij}'s which will be minimally inconsistent with the given values of conditional moments. Specifically we set up our objective function as follows:

$$\begin{aligned}
&\Delta(\underline{c}) \\
&= \sum_{i=1}^{m}[\frac{(\delta_i)^2}{\eta_i} - (-\tau_i c_{33} + c_{13} - c_{23}\log\tau_i + 1)]^2 + \sum_{i=1}^{m}[\frac{\delta_i}{\eta_i} - (c_{11} - \eta_i c_{31} + c_{21}\log\eta_i)]^2 \\
&+ \sum_{j=1}^{l}[\frac{(\beta_j)^2}{\chi_j} - (c_{22} + c_{23}\log\alpha_j - \alpha_j c_{21} + 1)]^2 + \sum_{j=1}^{l}[\frac{\beta_j}{\chi_j} - (c_{32} + c_{33}\log\alpha_j - \alpha_j c_{31})]^2.
\end{aligned} \tag{24}$$

where $\underline{c} = (c_{11}, c_{13}, c_{21}, c_{22}, c_{23}, c_{31}, c_{32}, c_{33})$. Next we consider partial derivatives with respect to c_{ij}'s and set them equal to zero. So we will have (i.e., $\frac{\partial}{\partial c_{ij}}\Delta(\underline{c}) = 0$, $\forall(i,j)$) the following set of 8 equations:

$$mc_{11} - [\sum_{i=1}^{m}\tau_i]c_{31} + [\sum_{i=1}^{m}\log\tau_i]c_{21} = \sum_{i=1}^{m}\frac{\delta_i}{\eta_i}. \tag{25}$$

$$mc_{13} - [\sum_{i=1}^{m}\tau_i]c_{31} + [\sum_{i=1}^{m}\log\tau_i]c_{21} = \sum_{i=1}^{m}\frac{\delta_i^2}{\eta_i} - m. \tag{26}$$

$$
\begin{aligned}
&[\sum_{i=1}^{m}(\log\tau_i)^2+\sum_{j=1}^{l}\alpha_j^2]c_{21}+[\sum_{i=1}^{m}\log\tau_i]c_{11}\\
&-[\sum_{i=1}^{m}\tau_i\log\tau_i]c_{31}-(\sum_{j=1}^{l}\alpha_j)c_{22}-(\sum_{j=1}^{l}\alpha_j\log\alpha_j)c_{23}\\
&=\sum_{i=1}^{m}\frac{\delta_i}{\eta_i}\log\tau_i+\sum_{j=1}^{l}\alpha_j-\sum_{j=1}^{l}\frac{(\beta_j)^2}{\chi_j}\alpha_j.
\end{aligned} \tag{27}
$$

$$
lc_{22}+[\sum_{j=1}^{l}\log\alpha_j]c_{23}-[\sum_{j=1}^{l}\alpha_j]c_{21}=-l+\sum_{j=1}^{l}\frac{(\beta_j)^2}{\chi_j}. \tag{28}
$$

$$
\begin{aligned}
&[\sum_{i=1}^{m}(\log\tau_i)^2+\sum_{j=1}^{l}(\log\alpha_j)^2]c_{23}-[\sum_{i=1}^{m}\tau_i\log\tau_i]c_{33}\\
&-[\sum_{j=1}^{l}\alpha_j\log\alpha_j]c_{21}+[\sum_{j=1}^{l}\alpha_j]c_{22}\\
&=-(m+l)+\sum_{i=1}^{m}\frac{(\delta_i)^2}{\eta_i}\log\tau_i+\sum_{j=1}^{l}\frac{(\beta_j)^2}{\chi_j}\log\alpha_j
\end{aligned} \tag{29}
$$

$$
\begin{aligned}
&-[\sum_{i=1}^{m}\tau_i]c_{11}-[\sum_{i=1}^{m}\tau_i\log\tau_i]c_{21}+[\sum_{j=1}^{l}\alpha_j]c_{32}\\
&[\sum_{i=1}^{m}\tau_i^2+\sum_{j=1}^{l}\alpha_j^2]c_{31}-[\sum_{j=1}^{l}\alpha_j\log\alpha_j]c_{33}\\
&=-\sum_{i=1}^{m}\frac{\delta_i}{\eta_i}\tau_i-\sum_{j=1}^{l}\frac{\beta_j}{\chi_j}\alpha_j.
\end{aligned} \tag{30}
$$

$$
-[\sum_{j=1}^{l}\alpha_j]c_{31}-lc_{32}-[\sum_{j=1}^{l}\log\alpha_j]c_{33}=\sum_{j=1}^{l}\frac{\beta_j}{\chi_j}. \tag{31}
$$

$$
\begin{aligned}
&-[\sum_{i=1}^{m}\tau_i]c_{13}--[\sum_{i=1}^{m}\tau_i\log\tau_i]c_{23}-[\sum_{j=1}^{l}\alpha_j\log\alpha_j]c_{31}\\
&+[\sum_{i=1}^{m}\tau_i^2+\sum_{j=1}^{l}(\log\alpha_j)^2]c_{33}\\
&=\sum_{i=1}^{m}\tau_i-\sum_{i=1}^{m}\frac{(\delta_i)^2}{\eta_i}\tau_i-\sum_{j=1}^{l}\frac{(\beta_j)^2}{\chi_j}\log\alpha_j.
\end{aligned} \tag{32}
$$

An alternative approach would involve matching conditional percentiles. However since for the Gamma family the percentile functions are not available in closed form, this approach will be more difficult to implement.

In the next section we provide a real life situation where we apply our above mentioned procedure.

6 Real life example

We consider the North Dakota Counties data(2010) (data source: US Census Bureau 2010), consists of 53 observations. Each observation represents the total average income(in dollars) per week for all the 53 counties. Plausibly the data set will be described by a hidden truncated Pareto(II) model of the form (10). To this end, first we apply all 3 of our classical estimation procedures for the hidden truncated bivariate Pareto(II) model (estimation by fractional method of moments, quartile method of estimation and also estimation by maximum likelihood) and we get the following estimates for the parameters:

- Estimation by the fractional method of moments: $\alpha = 3.5884$, $\tau = 2.6543$ and $\theta = 1.1114$.
- Estimation by the sample quantile method : $\alpha = 3.5909$, $\tau = 2.6491$ and $\theta = 1.1068$.
- Estimation by the method of maximum likelihood: $\alpha = 3.5947$, $\tau = 2.6617$ and $\theta = 1.1288$.

So we can consider (approximately) the following numbers as the estimates for our parameters: $\alpha = 3.59$, $\theta = 1.11$ and $\tau = 2.65$. So our fitted hidden truncated density will be

$$f^{HT}_{X|Y\leq 1.11}(x) = \frac{3.59 \times 2.65}{0.9675254}[(1 + 2.65x)^{-(3.59+1)} - (1 + 2.65x + 1.11)^{-(3.59+1)}]I(x \geq 0). \tag{33}$$

First of all we consider a graph in which we plot both the density of our data and also the density of X to see how well our probability distribution of X fits the data.

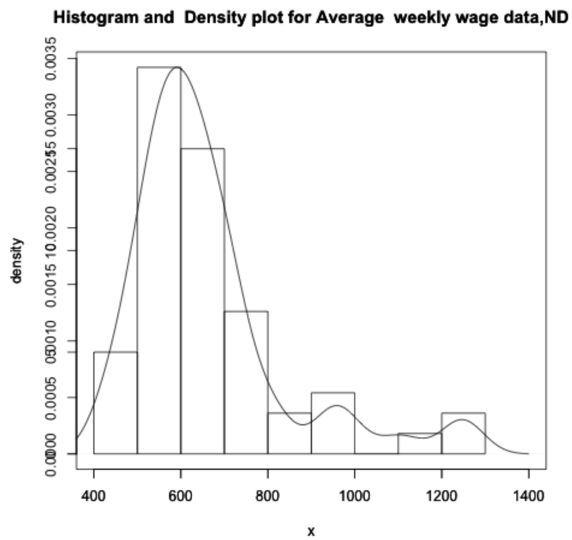

Figure 1: Histogram and fitted density plot of the North Dakota weekly average wage data(2010).

Next we consider the corresponding Bayesian estimation based on our conditional prior set up. For illustrative purpose we assume that our informed expert has supplied us with the following :

- $E(\alpha|\tau = 2) = 3$.
- $Var(\alpha|\tau = 2) = 2$.
- $E(\tau|\alpha = 2.5) = 1.2$.
- $Var(\tau|\alpha = 2.5) = 1.7$.
- $E(\alpha|\tau = 1) = 2$.
- $Var(\alpha|\tau = 1) = 1.5$.
- $E(\tau|\alpha = 3) = 2.1$.
- $Var(\tau|\alpha = 3) = 1.05$.

We need to choose c_{ij}'s that will minimize the sum of squares of the discrepancies between the elicited values and the corresponding functions of the c_{ij}'s.

Using the method as mentioned in section 5, we get the following estimated values of the c_{ij}'s which are consistent with the given information on conditional moments: $\underline{\hat{c}} = (c_{11} = 1.3625, c_{13} = 0.1206, c_{21} = 0.3967, c_{22} = 1.6026, c_{23} = 2.1194, c_{31} = -1.4574, c_{32} = 1.2676, c_{33} = 1.3812)$.

Note: It is quite easy to establish that the above choices of c_{ij}'s correspond to a proper (joint) posterior density for α, τ, θ.

We perform a Bayesian analysis with the above set of assessed values of the hyperparameters for the prior distributions assumed for our model.

6.1 Posterior simulation study

We consider an MCMC procedure for the Bayesian estimation of the parameters. For the jumping distribution we consider dependent gamma distributions for α and τ and independent gamma prior for θ. The posterior analysis is based on 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: $\alpha = 2.1$, $\tau = 1.81$, $\theta = 0.57$. Jumping distribution of the parameters :

- $\alpha|\tau \sim \Gamma$(shape=0.1206+2.1194 $\log \tau$ -1.3812 τ,intensity=1.3624+1.5748 τ+0.3967 $\log \tau$),
- $\tau|\alpha \sim \Gamma$(shape=1.6026+2.1194 $\log \alpha$-0.3967 α, intensity=1.4574 α+1.2676+1.3812log α).
- $\theta \sim \Gamma(2.6, 2.2)$,

after substituting the estimated values of the hyperparameters in the conditional densities for α and τ given earlier. Next we provide the results of our Bayesian analysis given below: Posterior mean(α)=3.1643, Posterior mean(τ)=2.6241, Posterior mean(θ)=0.7324. The corresponding (95%) prediction intervals are (1.1567,3.2239), (2.3121,4.4763), and (0.3287,1.9843) for α, τ and θ respectively.

7 An illustrative example with using simulated data

Suppose we have (from our informed expert) the following set of information:

- $E(\alpha|\tau = 2) = 1.5$.
- $Var(\alpha|\tau = 2) = 2$.
- $E(\tau|\alpha = 3) = 0.17$.
- $Var(\tau|\alpha = 3) = 1.16$.
- $E(\alpha|\tau = 1.5) = 2.4$.
- $Var(\alpha|\tau = 1.5) = 3.9$.
- $E(\tau|\alpha = 2) = 0.29$.
- $Var(\tau|\alpha = 2) = 1.05$.

We need to choose c_{ij}'s that will minimize the sum of squares of the discrepancies between the elicited values and the corresponding functions of the c_{ij}'s.

On solving the above set of linear equations (using the method as mentioned in section 5) we get the following optimum choices for the c_{ij}'s $\underline{c} = (1.9843, -0.0446, -0.4021, 0.9649, 1.0125, -1.3018, 0.7081, -0.1203)$. Next we perform a Bayesian analysis for our model based on the following choices of the c_{ij}'s as the hyperparameters of the prior distribution assumed for the α and τ.

7.1 Simulation study

We draw samples of size $n = 100$ and $n = 200$ from our density given earlier(for a particular choice of the parameter values $\alpha = 3$, $\tau = 1$ and $\theta = 1.5$). We consider dependent gamma priors for α and τ while for θ, we consider independent gamma prior as before. Specifically we consider the following:

- $\alpha|\tau \sim \Gamma$(shape=-0.0446-1.3018 $\log \tau$ +1.0125 τ,intensity=1.9843+1.3018 τ-0.4021 $\log \tau$),
- $\tau|\alpha \sim \Gamma$(shape=0.9649+1.0125 $\log \alpha$+0.4021 α, intensity=-1.3018 α+0.7081-0.1203log α).
- $\theta \sim \Gamma(2.9, 2.4)$,

after substituting the estimated values of the hyperparameters in the conditional densities for α and τ given earlier. Then for the Bayesian analysis we consider the MCMC technique to draw samples from the posterior distribution for each of the parameters and we provide the results in the table given below:

The values in the last column of the table are Bayesian confidence intervals(95%) for α, τ and θ respectively.

n	Posterior Mean(α)	Posterior Mean(τ)	Posterior Mean(θ)	Credible interval
100	2.9410	0.9811	1.5489	(0.1014,5.9793),(0.0163,1.9734),(0.0215,3.4642)
200	2.9732	0.9532	1.5129	(0.1314, 4.3523),(0.0252,1.7565),(0.1235, 3.2307)

Table 1: Bayesian estimates of the parameters

8 Full conditional prior set-up

Here we consider the following:

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

$$f(\alpha|\tau, \theta) \propto \alpha^{\xi_1(\tau,\theta)-1} \exp(-\alpha\lambda_1(\tau, \theta))I(\alpha > 0). \tag{34}$$

2. The conditional density of τ given α and θ is a Gamma distribution with shape parameter $= \xi_2(\alpha, \tau,)$ and intensity parameter $= \lambda_2(\alpha, \theta)$. So that the conditional density of τ given α and θ will be

$$f(\tau|\alpha, \theta) \propto \tau^{\xi_2(\alpha,\theta)-1} \exp(-\tau\lambda_2(\alpha, \theta))I(\tau > 0). \tag{35}$$

3. The conditional density of θ given τ and α is a Gamma distribution with shape parameter $= \xi_3(\tau, \alpha)$ and intensity parameter $= \lambda_3(\tau, \alpha)$. So that the conditional density of θ given α and τ will be

$$f(\theta|\tau, \alpha) \propto \theta^{\xi_3(\tau,\alpha)-1} \exp(-\theta\lambda_3(\tau, \alpha))I(\theta > 0). \tag{36}$$

So the joint density of α τ and θ will be (since all the conditionals are members of exponential families of densities) (according to Arnold, Castillo and Sarabia(1999))

$$\begin{aligned}
&f(\alpha, \tau, \theta)\\
&= \exp[m_{000} - m_{100}\alpha + m_{200} \log \alpha - m_{010}\tau + m_{110}\tau\alpha - m_{210}\tau \log \alpha\\
&+ m_{020} \log \tau - m_{120}\alpha \log \tau + m_{220} \log \alpha \log \tau - m_{001}\theta + m_{101}\theta\alpha - m_{201}\theta \log \alpha\\
&+ m_{011}\theta\tau - m_{111}\alpha\theta\tau - m_{021}\theta \log \tau + m_{121}\theta\alpha \log \tau + m_{211}\theta\tau \log \alpha\\
&- m_{221}\theta \log \tau \log \alpha + m_{002} \log \theta\\
&- m_{102}\alpha \log \theta + m_{202} \log \theta \log \alpha - m_{012}\tau \log \theta\\
&+ m_{112}\tau\alpha \log \theta - m_{212}\tau \log \theta \log \alpha - m_{022} \log \tau \log \theta\\
&- m_{122}\alpha \log \tau \log \theta + m_{222} \log \theta \log \alpha \log \tau]I(\alpha > 0) \times I(\tau > 0) \times I(\theta > 0).
\end{aligned} \tag{37}$$

Now if we consider $m_{ijk} = 0$ whenever $i + j + k > 2$, then equation(37) reduces to

$$\begin{aligned}
&f(\alpha, \tau, \theta)\\
&= \exp[m_{000} - m_{100}\alpha + m_{200} \log \alpha - m_{010}\tau + m_{110}\tau\alpha + m_{020} \log \tau\\
&- m_{001}\theta + m_{101}\theta\alpha + m_{011}\theta\tau + m_{002} \log \theta]I(\alpha > 0) \times I(\tau > 0) \times I(\theta > 0).
\end{aligned} \tag{38}$$

Note: The reason that we have imposed the restriction $m_{ijk} = 0$ whenever $i + j + k > 2$ is simply because we are looking for a simple but flexible family with less parameters than the 26 which appear in (37).

8.1 Appropriate choice of hyperparameters

We have from (38), the joint (marginal) density of τ and θ will be

$$\begin{aligned} f(\tau,\theta) &= \exp[m_{000} - m_{010}\tau + m_{020}\log\tau - m_{001}\theta + m_{011}\theta\tau + m_{002}\log\theta] \\ &\times \int_0^\infty \alpha^{m_{200}} \exp[-\alpha(m_{100} - m_{110}\tau - m_{101}\theta)]d\alpha \\ &= \exp[m_{000} - m_{010}\tau + m_{020}\log\tau - m_{001}\theta + m_{011}\theta\tau + m_{002}\log\theta] \\ &\times \frac{\Gamma(m_{200}+1)}{(m_{100} - m_{110}\tau - m_{101}\theta)^{m_{200}+1}} I(\tau > 0) \times I(\theta > 0). \end{aligned} \tag{39}$$

So the conditional density of α given τ and θ for each fixed choices of τ and θ will be

$$\begin{aligned} & f(\alpha|\tau,\theta) \\ &= \frac{f(\alpha,\tau,\theta)}{f(\tau,\theta)} \\ &= \exp[-\alpha(m_{100} - m_{110}\tau - m_{101}\theta)]\alpha^{m_{200}} \\ &\times \frac{(m_{100} - m_{110}\tau - m_{101}\theta)^{m_{200}+1}}{\Gamma(m_{200}+1)} I(\alpha > 0). \end{aligned} \tag{40}$$

Equivalently we can say that the conditional density of α given τ and θ will be a gamma with shape=$m_{200}+1$ and intensity=$m_{100} - m_{110}\tau - m_{101}\theta$.

- Similarly, the conditional density of τ given α and θ will be a gamma with shape=$m_{020}+1$ and intensity=$m_{010} - m_{110}\alpha - m_{011}\theta$.
- Also the conditional density of θ given α and τ will be a gamma with shape=m_{002}+1 and intensity=$m_{001} - m_{101}\alpha - m_{011}\tau$.
- So we have in total 9 parameters to estimate which are $m_{010}, m_{020}, m_{001}, m_{011}, m_{002}, m_{100}, m_{110}, m_{101}, m_{200}$.

Also note that for all the expressions for the conditional means and variances to be positive we need the following restriction on the choices of m_{ijk}'s: ($m_{100} > 0, m_{110} < 0, m_{101} < 0, m_{010} > 0, m_{011} < 0, m_{001} > 0, m_{101} < 0, m_{200} > -1, m_{020} > -1, m_{002} > -1$). We can use the corresponding conditional moments for all the parameters to evaluate those hyperparameters.

From (49)-(50), we can write that,

1. $E[\alpha|\tau,\theta] = \frac{m_{200}+1}{m_{100}-m_{110}\tau-m_{101}\theta}$, and $Var[\alpha|\tau,\theta] = \frac{m_{200}+1}{(m_{100}-m_{110}\tau-m_{101}\theta)^2}$.
2. $E[\tau|\alpha,\theta] = \frac{m_{020}+1}{m_{010}-m_{110}\alpha-m_{011}\theta}$, and $Var[\tau|\alpha,\theta] = \frac{m_{020}+1}{(m_{010}-m_{110}\alpha-m_{011}\theta)^2}$.
3. $E[\theta|\alpha,\tau] = \frac{m_{002}+1}{m_{001}-m_{101}\alpha-m_{011}\tau}$, and $Var[\theta|\alpha,\tau] = \frac{m_{002}+1}{(m_{001}-m_{101}\alpha-m_{011}\tau)^2}$.

Next we suppose that we have information (from our informed expert) on the conditional mean and variances for all the above. For example suppose that we have,

- $E[\alpha|\tau = 1, \theta = 0.5] = 2.31$
- $Var[\alpha|\tau = 1, \theta = 0.5] = 1.05$
- $E[\alpha|\tau = 1.5, \theta = 0.6] = 2.17$
- $Var[\alpha|\tau = 1.5, \theta = 0.6] = 2.91$

- $E[\tau|\alpha = 2, \theta = 0.5] = 1.16$
- $Var[\tau|\alpha = 2, \theta = 0.5] = 0.65$
- $E[\tau|\alpha = 3, \theta = 0.6] = 1.78$
- $Var[\tau|\alpha = 3, \theta = 0.6] = 0.87$
- $E[\theta|\alpha = 2, \tau = 1] = 3.09$
- $Var[\theta|\alpha = 2, \tau = 1] = 0.17$
- $E[\theta|\alpha = 3, \tau = 1.5] = 3.45$
- $E[\theta|\alpha = 3, \tau = 1.5] = 1.09.$

Again as before, by using the methods mentioned in section 5, we get the following optimum choices for the m_{ijk}'s $\underline{m} = (m_{010} = 2.0836, m_{110} = -0.1883, m_{011} = -0.1495, m_{101} = -0.0918, m_{100} = 0.9090, m_{001} = 8.5926, m_{020} = 9.3539, m_{200} = 2.3501, m_{002} = 32.5425)$.

As before, We perform a Bayesian analysis with the above set of assessed values of the hyperparameters for the prior distributions assumed for our model.

8.2 Simulation study

We draw samples of size $n = 100$ and $n = 200$ from our density given earlier(for a particular choice of the parameter values $\alpha = 3$, $\tau = 1$ and $\theta = 1.5$). We consider dependent gamma priors for all the parameters for our study. Specifically we consider the following for the jumping distribution:

- $\alpha|\tau, \theta \sim \Gamma$(shape=2.3501+1, intensity=0.9090+0.1883 τ+0.09182θ),
- $\tau, |\alpha, \theta \sim \Gamma$(shape=9.3539+1, intensity=2.0836+0.1883 α+0.1495 θ),
- $\theta|\alpha, \tau \sim \Gamma$(shape=32.5425+1, intensity=8.5926+0.0918 α+0.1495τ),

after substituting the estimated choices of the hyperparameters in the conditional densities of α, τ and θ given earlier. Then for the Bayesian analysis we consider the MCMC technique to draw samples from the posterior distribution for each of the parameters and we provide the results in the table given below:

n	Posterior Mean(α)	Posterior Mean(τ)	Posterior Mean(θ)	Credible interval
100	2.9415	0.9643	1.46893	(0.1354,4.2795),(0.0365,1.8872),(0.3278,3.8654)
200	2.96312	0.9704	1.4817	(0.2534, 4.1629),(0.1247,1.7449),(0.2267, 3.2345)

Table 2: Bayesian estimates of the parameters.

The values in the last column of the table are Bayesian confidence interval(95) for α, τ and θ respectively.

9 Use of diffuse priors (or equivalently locally uniform priors)

The use of conditionally specified priors as mentioned in earlier section involves assessment of several hyperparameters which at times can be quite challenging. If we envisage a situation where our informed expert does not have any knowledge (or in other words he or she is unable to provide plausible values) of conditional means, variances or percentile functions as described in the suggested prior assessment approach. In such a situation it is quite reasonable to use such values of hyperparameters which reflect ignorance or diffuseness of prior information about the parameters. If our informed expert is unable to provide any information on the conditional moments for his choice of prior then one might consider of using a locally uniform joint prior which will correspond to the situation in which all the hyperparameters in (15) are set equal to zero.

10 Concluding remarks

Precise inference under both classical and Bayesian paradigm for the parameters of hidden truncation models is an area which still is in it's infancy stage. Very recently Arnold and Ghosh (2011) have provided discussion on Bayesian inference for trivariate hidden truncated Pareto(II) model with a diffuse prior set-up. Inarguably the complicated form of the likelihood as in (12) is a warning that friendly conjugate priors will not be encountered. As a consequence, little attention has been devoted in this direction. The output of the associated Bayesian analysis for two different conditional set-up do provide a strong indication that although the number of hyperparameters is high, the assessment problem is manageable, since the methods either involve systems of linear equations in the hyperparameters or nonlinear equations in a single unknown. To this end our procedure can also be extended for a more general Pareto (IV) distribution under the hidden truncation paradigm. However needless to say, from the computational point of view this will be quite challenging as there will be three additional parameters and with that the associated prior distributions assumed for the model will have quite a large set of hyperparameters.

References

[1] Arnold, B.C. (1983). *Pareto Distributions.* International Cooperative Publishing House, Fairland, MD.

[2] Arnold, B.C. (1997). Characterizations involving conditional specification. *Journal of Statistical Planning and Inference,* 63, 117-131.

[3] Arnold, B.C. and Beaver, R.J. (2000). Hidden truncation models. *Sankhya Series A*, 62, 23-35.

[4] Arnold, B.C., Castillo, E., and Sarabia, J.M. (1999). *Conditional Specification of Statistical Models.* Springer, New York.

[5] 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.

[6] Arnold, B.C. and Ghosh, I. (2011). Hidden truncation in multivariate Pareto data. To appear *Contemporary Mathematics with Applications,* Asian Books Private Ltd, Kolkata.

[7] Arnold, B.C. and Press, S.J. (1983). Bayesian inference for Pareto populations. *Journal of Econometrics*, 21, 287-306.

[8] Arnold, B.C. and Press, S.J. (1983). Bayesian estimation and prediction for Pareto data. *Journal of the American Statistical Association*, 84, 1079-1084.

[9] Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995). *Bayesian Data Analysis.* Chapman and Hall, New York.

[10] Kotz, S., Balakrishnan, N., and Johnson, N.L. (2000). *Continuous Multivariate Distributions.* John Wiley & Sons, New York.

[11] Press, S.J. (2003). *Subjective and Objective Bayesian Statistics.* Wiley, New York.