International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 6 (Special Issue), 2007, pp 137-150

Bayesian Prediction and Estimation from Pareto Distribution of the First Kind

Hatem A. Howlader
Dept. of Mathematics and Statistics
University of Winnipeg
Manitoba, Canada R3B 2E9

Anwar M. Hossain and Oleg Makhnin
Dept. of Mathematics
New Mexico Tech
Socorro, NM 87801, U.S.A

[Received February 16, 2005; Revised August 7, 2007; Accepted August 29, 2007]

Abstract

This paper presents the Highest Posterior Density (HPD) interval for the Pareto parameter and the associated reliability function based on natural conjugate prior (NCP) and minimal information prior (MIP). The Bayes predictive estimator and the HPD prediction interval for a future observation are also presented. Bessel function of the third kind and its asymptotic expansion have been employed in order to overcome the intractability of the integrals under the minimal information prior. A numerical example is given to illustrate the results.

Keywords and Phrases: Bessel Function, Incomplete Gamma Function, Loss Function, Minimal Information Prior, Natural Conjugate Prior, Posterior Density, Prediction Interval, Reliability Function.

AMS Classification: 62N05.

1 Introduction

The Pareto distribution was introduced (Pareto, 1897) as a model for the distribution of income. In addition to economics, its models in several different forms are now being used in a wide range of fields such as insurance, business, engineering, survival analysis, reliability and life testing. See Nelson (1982), Lwin (1972), Lomax (1954), Cohen and Whitten (1988), Cox and Oakes (1984), Davis and Feldstein (1979), Lawless (1982). The probability density function (pdf) of the Pareto distribution is given by

$$f(\mathbf{x}|\theta_0, \beta) = \beta {\theta_0}^{\beta}\ \mathbf{x}^{-(\beta+1)}, \quad \beta > 0, \quad 0 < \theta_0 < \mathbf{x} < \infty \tag{1}$$

The quantity θ_0 is a threshold parameter above which the distribution of income follows the paretian law (1). The parameter β measures the degree of inequality of income. Cramer (1971, p.57) remarked that the values of β have increased from between 1.6 and 1.8 in the nineteenth century to between 1.9 and 2.1 in the developed countries in the present time. The pdf (1) is also known as the Pareto distribution of the first kind, most commonly used in economic modelling and risk analysis in insurance and business.

A two parameter Pareto distribution with pdf

$$f(\mathbf{x}|\lambda, \beta) = \frac{\lambda}{\beta}\left(1 + \frac{\mathbf{x}}{\beta}\right)^{-(\lambda+1)}, \mathbf{x} > 0;\ \lambda, \beta > 0$$

is known in the literature as the Pareto distribution of the second kind or the Lomax distribution or Pearson's Type VI distribution (Johnson, Kotz & Balakrishnan (1994)). It has been found to provide a good model in biomedical problems, such as survival time following a heart transplant (Bain & Englehardt (1992)). Lomax (1954) used this model in the analysis of business failure data. The length of wire between flaws has also been found to follow a Pareto distribution of the second kind (Bain & Englehardt (1992)). A general three parameter form was used, among others by Charek, Moore & Coleman (1988). Arnold (1983) studied an extended four parameter Pareto distribution that has all of the above types as special cases.

In this study, we consider only the Pareto distribution of the first kind. The objective of this paper is to obtain and compare the Highest Posterior Density (HPD) intervals for β and also for the associated reliability function R_t based on natural conjugate and minimal information priors. HPD-intervals for a future observation are also derived under the above priors.

2 Prior and Posterior Distributions

For a random sample $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n})$ of size n from (1), the likelihood function is given by

$$l(\mathbf{x}|\theta_0, \beta) = \left(\frac{\beta}{G}\right)^{n} exp(-nc_0\, \beta) \tag{2}$$

where $G = (\prod_i^n \mathbf{x_i})^{1/n}$ is the geometric mean of the $\mathbf{x}$'s and $c_0 = log(G/\theta_0)$.

The Pareto distribution belongs to the general exponential family and that $\sum log(\mathbf{x_i})$ is a sufficient statistic for β. The likelihood function (2) can be rewritten in order to express it in the form of the general exponential family as:

$$l(\mathbf{x}|\theta_0, \beta) = \beta^n \; [exp(-\beta\, c_0)/G]^n \tag{3}$$

Natural Conjugate Prior (NCP):

The NCP is generated by replacing all the quantities in the kernel of the likelihood function, after expressing it in the form of the general exponential family, dependent on the sample by parameters known as prior parameters.
Thus, the NCP for β from (3) is

$$g_1(\beta) \propto \beta^{c-1} exp(-p\beta), \quad p, c > 0; \beta > 0 \tag{4}$$

where c and p are prior parameters and the normalizing constant of (4) is $\frac{p^c}{\Gamma(c)}$. Hence, the NCP for the Pareto distribution is a gamma distribution, $G(c, p)$.
Combining (2) and (4), the posterior density of β is,

$$\Pi(\beta|\theta_0, \mathbf{x}) = \frac{(nc_0 + p)^{(c+n)}}{\Gamma(c+n)} \; \beta^{c+n-1} \; exp\{-\beta(nc_0 + p)\}, \; \beta > 0 \tag{5}$$

The posterior pdf (5) is also gamma, $G(c+n, nc_0+p)$, showing that the posterior has the same functional form as the prior, and hence, the gamma priors are closed under sampling. The Bayes estimator of β under the squared error loss function (SEL)from (5) is $\beta^*_N = (c+n)/(nc_0+p)$. The SEL is appropriate when large errors of estimation are considered to be more serious, compared to, for example, an absolute loss error function, for which large errors are not quite as serious.

Minimal Information Prior (MIP):

A class of non-informative priors called Minimal Information Priors (MIP) was proposed by Zellner (1971) by using information theoretic approach. These priors are dependent upon a particular parametrization used. In order to generate an MIP, let $f(\mathbf{x}|\theta)$ be the pdf of X, and $g(\theta)$ be a prior density of θ. Define,

$$I_{\mathbf{x}}(\theta) = \int f(\mathbf{x}|\theta) \log f(\mathbf{x}|\theta) \, d\mathbf{x},$$

as a measure of information in $f(\mathbf{x}|\theta)$, and

$$\int g(\theta) \log g(\theta) \, d\theta,$$

as a measure of information in the prior $g(\theta)$. Define,

$$\int I_{\mathbf{x}}(\theta)g(\theta)d\theta,$$

as measuring the prior average information in the data. Then,

$$G = \int I_{\mathbf{x}}(\theta)g(\theta)d\theta - \int g(\theta)\log g(\theta)\, d\theta \tag{6}$$

represents the gain in information associated with an observation $\mathbf{x}$ over the information in the prior $g(\theta)$.
The MIP is defined as the prior that maximizes G by varying $g(\theta)$ subject to

$$\int g(\theta)\, d(\theta) = 1.$$

Using the Lagrange multiplier, it can be easily shown that the MIP is

$$g(\theta) \propto exp\{I_{\mathbf{x}}(\theta)\}.$$

For the Pareto pdf (1), $I_{\mathbf{x}}(\beta) = \log\beta - \frac{1}{\beta} - 1 - \log\theta$, and hence, the MIP,

$$g_2(\beta) \propto \beta\, exp(-\frac{1}{\beta}),\ \ \beta > 0 \tag{7}$$

Combining (2) and (7), the posterior distribution of β under the MIP is given by

$$\Pi(\beta|\theta_0,\ \mathbf{x}) = \frac{\beta^{n+1}\, exp\left\{-(n\, c_0\, \beta + \frac{1}{\beta})\right\}}{\int_0^\infty \beta^{n+1}\, exp\left\{-(n\, c_0\, \beta + \frac{1}{\beta})\right\} d\beta},\quad \beta > 0 \tag{8}$$

We evaluate the denominator of (8) by using the modified Bessel function of the third kind of order ν (Erdelyi,et.al., 1953) given by

$$\frac{2}{a^\nu} K_\nu(az) = \int_0^\infty t^{-(\nu+1)}\, exp\left\{\frac{-z}{2}\left(t + \frac{a^2}{t}\right)\right\} dt \tag{9}$$

where

$$K_\nu(az) = K_{-\nu}(az).$$

Replacing z, a and ν in (9), respectively by $2nc_0$, $(nc_0)^{-\frac{1}{2}}$ and $-(n+2)$, we have from (8),

$$\Pi(\beta|\theta_0, \mathbf{x}) = \frac{(nc_0)^{\frac{n+2}{2}}}{2K_{n+2}(2\sqrt{nc_0})}\ \beta^{n+1}\ exp\{-(nc_0\beta + \frac{1}{\beta})\},\ \ \beta > 0 \tag{10}$$

The Bayes estimator of β from (10) under the SEL can be shown to be

$$\beta_M^* = E[\beta|\theta_0, \mathbf{x}] = \frac{K_{n+3}(2\sqrt{nc_0})}{K_{n+2}(2\sqrt{nc_0})\sqrt{nc_0}}.$$

In order to assess the performance of β_M^*, we perform a Monte Carlo simulation study for different true values of β and different n. In each case, $10,000$ random samples were generated from (1) using $\theta_0 = 100$. The estimates and their corresponding mean squared errors (MSE's) in parentheses are presented in Table 1. It is clear from Table 1 that the MIP somewhat overestimates β for all sample sizes, especially for small n.

Table 1: Posterior means and MSE's (in parentheses) of β under MIP.

β	n =20	n = 40	n = 60	n= 100
1.0	1.201	1.100	1.067	1.040
	(0.115)	(0.041)	(0.023)	(0.013)
1.5	1.785	1.638	1.594	1.555
	(0.250)	(0.085)	(0.052)	(0.027)
2.0	2.353	2.178	2.122	2.074
	(0.413)	(0.156)	(0.092)	(0.049)
2.5	2.954	2.705	2.645	2.585
	(0.675)	(0.225)	(0.140)	(0.075)
3.0	3.502	3.257	3.164	3.099
	(0.908)	(0.338)	(0.200)	(0.107)

3 Numerical Example

In order to illustrate our results, we use the Dyer (1981) annual wage data (in multiples of 100 U.S. dollars) of a random sample of 30 production line workers in a large industrial farm as follows:

112	154	119	198	112	156	123	103	115	107
125	119	128	132	106	151	103	104	116	140
108	105	158	104	119	111	101	157	112	115

As in Dyer (1981) we assume the minimum wage θ_0 for these workers as 100 U.S. dollars. The Pareto model was found to adequately fit the data with a p-value of 0.25.

First, we investigate the conditions under which the posteriors (5) and (10) under NCP and MIP, respectively, are close. The prior (7) is improper, as its mass increases with the increasing values of β. Intuitively, we should reach the best agreement with NCP in equation (4) when its density is also improper, that is $p \approx 0$. One can consider it a limiting case when the parameter p becomes small, and more and more mass is assigned to higher values of β. If we set $p = 0$, the Gamma prior (4) does not make

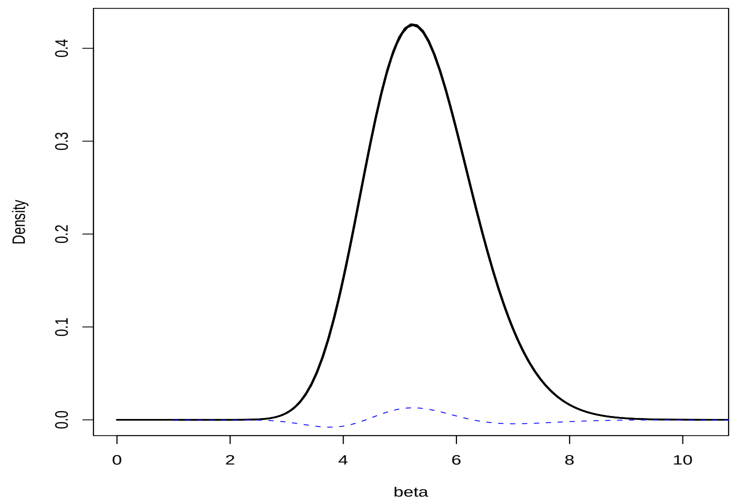

Figure 6: Posteriors under MIP and NCP (virtually indistinguishable) with $p = 0$, $c = 2.19$. Broken line shows the difference in posteriors magnified by a factor of 10.

sense, but setting $p = 0$ in (5) still gives us a valid posterior.

Next, we choose the value of c. This can be done simply by equating the posterior mean under MIP, β^*_M, with that under NCP, β^*_N, resulting in $c = \beta^*_M(nc_0 + p) - n$.

Empirically, we observed that the absolute difference between (5) with above choice of p and c, and (10) decreases when n gets large. Fig. 1 shows the two posterior densities, computed for the above annual wage data. They are extremely close, actually, indistinguishable on the graph.

4 HPD Intervals for β

A highest posterior density (HPD) interval is one in which the posterior density for every point inside the interval is greater than that for every point outside of it so that the interval includes the more probable values of the parameter and excludes the less probable ones. For a uni-modal posterior pdf, a $(1 - \alpha)$ HPD interval (h_1, h_2) for θ must satisfy the following two equations simultaneously:

$$\int_{h_1}^{h_2} \Pi(\theta|\mathbf{x})d\theta = 1 - \alpha$$

and

$$\Pi(h_1|\mathbf{x}) = \Pi(h_2|\mathbf{x}).$$

Since the pdf (5) is unimodal, the HPD interval for β under the NCP can be shown to be the simultaneous solution of the equations (11) and (12):

$$\Gamma(c+n, q_2) - \Gamma(c+n, q_1) = 1 - \alpha \tag{11}$$

$$\left(\frac{h_1}{h_2}\right)^{c+n-1} = \left(\frac{e^{h_1}}{e^{h_2}}\right)^{nc_0+p} \tag{12}$$

where $q_i = (nc_0 + p)h_i, \;\; i = 1, 2$ and

$$\Gamma(c+n, q_i) = \frac{1}{\Gamma(c+n)} \int_0^{q_i} e^{-y} \; y^{(c+n-1)} \; dy, \tag{13}$$

the incomplete gamma function ratio or the gamma cdf.
Similarly, the HPD interval (h_1', h_2') for β under the MIP from the unimodal pdf (10) is the simultaneous solution of the equations (14) and (15):

$$\left(\frac{h_1'}{h_2'}\right)^{n+1} = \frac{exp\left(nh_1'c_0 + \frac{1}{h_1'}\right)}{exp\left(nh_2'c_0 + \frac{1}{h_2'}\right)} \tag{14}$$

$$d_1[u_2 - u_1] = 1 - \alpha \tag{15}$$

where

$$d_1 = \frac{(nc_0)^{\frac{n+2}{2}}}{2K_{n+2}\{2(nc_0)^{\frac{1}{2}}\}}$$

$$u_1 = \int_0^{h_1'} \beta^{n+1} \, exp\{-(nc_0\beta + \frac{1}{\beta})\} d\beta$$

$$u_2 = \int_0^{h_2'} \beta^{n+1} \, exp\{-(nc_0\beta + \frac{1}{\beta})\} d\beta$$

Numerical estimates for wage data of Section 3 were obtained using R statistical package, via Newton-Raphson method. The resulting HPD intervals for MIP tend to be close to those for NCP with prior parameters chosen as discussed above (see Table 2).

Table 2: HPD intervals for β (95%) under NCP (matching the MIP mean $\beta^*_M = 5.3970$) and MIP.

p	c	HPD interval
2.0	12.9851	(3.8309, 7.0344)
1.0	7.5881	(3.7268, 7.1490)
0.5	4.8896	(3.6662, 7.2160)
0.1	2.7308	(3.6127, 7.2755)
0.0	2.1911	(3.5985, 7.2913)
MIP	–	(3.6040, 7.2857)

5 HPD Interval for the Reliability Function

The reliability function of (1) is given by

$$R_t = \left(\frac{\theta_0}{t}\right)^{\beta}, \quad t \geq \theta_0$$

Substituting $\beta = -\frac{\log R_t}{r}$ in (5), where $r = log(t/\theta_0)$, we obtain the posterior pdf of R_t as

$$\Pi(R_t|\theta_0, \mathbf{x}) = \frac{(nc_0+p)^{c+n}}{\Gamma(c+n)} r^{-(c+n)} (-log\, R_t)^{c+n-1} R_t^{\frac{nc_0+p}{r}-1}, \quad 0 < R_t < 1$$

The Bayes estimator of R_t under the squared error loss function and the NCP is

$$R^*_{t,N} = E(R_t|\theta_0, \mathbf{x}) = \left(1 + \frac{r}{nc_0+p}\right)^{-(c+n)} = \left(1 + \frac{log(t/\theta_0)}{nc_0+p}\right)^{-(c+n)} \tag{16}$$

The $(1-\alpha)$ HPD interval $(I_{1,N},\, I_{2,N})$ for R_t is given by the simultaneous solution of the equations (17) and (18):

$$\Gamma(c+n, b_2) - \Gamma(c+n, b_1) = 1 - \alpha \tag{17}$$

and

$$\left(\frac{log\, I_{1,N}}{log\, I_{2,N}}\right)^{c+n-1} = \left(\frac{I_{2,N}}{I_{1,N}}\right)^{\frac{nc_0+p}{r}-1} \tag{18}$$

where $b_j = -\left(\frac{nc_0+p}{r}\right) log\, I_{j,N},\ j = 1,2$ and $\Gamma(c+n, b_j),\ j = 1,2$ is the incomplete gamma function ratio as in (13).
Again, substituting $\beta = \frac{-log\, R_t}{r}$ and $r = log(t/\theta_0)$ in (10), we obtain the posterior pdf of R_t under the MIP as:

$$\Pi(R_t|\theta_0, \mathbf{x}) = d_2\, (-log\, R_t)^{n+1}\, exp\left\{ (\frac{-nc_0}{r} + 1)(-log\, R_t) + \frac{r}{logR_t} \right\}, 0 < R_t < 1 \tag{19}$$

The normalizing constant d_2 of (19) is evaluated by using (9) with $\nu = -(n+2)$, $z = \frac{2(nc_0+2r)}{r}$ and $a = r(nc_0+2r)^{-1/2}$ as $d_2 = \frac{d_1}{r^{n+2}}$. The Bayes estimator of R_t under the squared error loss function and the MIP is

$$R^*_{t,M} = E(R_t|\theta_0, \mathbf{x}) = \left(1 + \frac{r}{nc_0}\right)^{-(\frac{n+2}{2})} \frac{K_{n+2}\{2(nc_0+r)^{1/2}\}}{K_{n+2}\{2(nc_0)^{1/2}\}}$$

Using an asymptotic expansion (Copland(1965)), namely,

$$K_\nu(y) \cong \frac{2^{\nu-\frac{1}{2}}\,\nu^{\nu-\frac{1}{2}}\sqrt{\pi}\,exp(-\nu)}{y^\nu} \tag{20}$$

we obtain $R^*_{t,M}$ as

$$R^*_{t,M} = \left(\frac{nc_0}{nc_0+r}\right)^{n+2} \tag{21}$$

The $(1-\alpha)$ HPD interval $(I_{1,M}, I_{2,M})$ for R_t is given by the simultaneous solution of the equations (22) and (23):

$$d_2 \int_{I_{1,M}}^{I_{2,M}} (-logR_t)^{n+2}\, exp\left\{\left(-\frac{nc_0}{r}+1\right)(-logR_t) + \frac{r}{log\, R_t}\right\} dR_t = 1-\alpha \tag{22}$$

and

$$\left(\frac{logI_{1,M}}{logI_{2,M}}\right)^{n+1} \left\{e^{r\left(\frac{1}{I_{1,M}}-\frac{1}{I_{2,M}}\right)}\right\} = \left(\frac{I_{2,M}}{I_{1,M}}\right)^{\frac{nc_0}{r}+1} \tag{23}$$

Fig. 2 shows the bounds of HPD intervals for different values of t for our numerical example of Section 3. The nonparametric (empirical) estimate of R_t is also shown. The estimates for $t = 120$ and the HPD intervals are given in Table 3.

Table 3: Estimates of R_t, $t = 120$ and 95% HPD-intervals under NCP and MIP.

p	c	Posterior Mean and HPD interval
2.0	12.9851	0.3780 (0.2703, 0.4879)
1.0	7.5881	0.3786 (0.2636, 0.4960)
0.5	4.8896	0.3789 (0.2598, 0.5008)
0.1	2.7308	0.3793 (0.2564, 0.5050)
0.0	2.1911	0.3794 (0.2555, 0.5061)
MIP	–	0.3793 (0.2558, 0.5057)

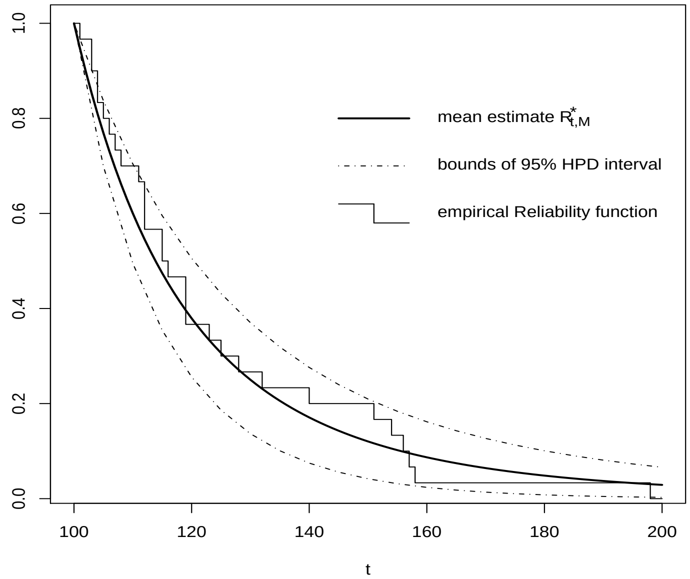

Figure 7: Estimates of $R^*_{t,M}$, bounds of 95% HPD interval and empirical reliability function.

6 Bayes Predictive Interval For a Future Observation

The prediction problems of the Pareto life time models are very important and have been studied, among others, by Arnold & Press (1989), Madi & Raqab (2004), Nigm & Hamdy (1987) and Soliman (2000). For other prediction problems, see Khan (2004, 2006), Khan & Chattopadhyay (2003) and Geisser (1984).
Let y be a future observation from (1). Given the data $\mathbf{x}$, the conditional joint pdf of y and β is

$$\begin{aligned} h(y, \beta|\theta_0, \mathbf{x}) &= f(y|\theta_0, \beta, \mathbf{x})\ \Pi(\beta|\theta_0, \mathbf{x}) \\ &= f(y|\theta_0,\ \beta)\ \Pi(\beta|\theta_0, \mathbf{x}) \end{aligned} \tag{24}$$

since y and $\mathbf{x}$ are independent.
Using (1) and the posterior pdf (5) under the NCP, the equation (24) takes the form:

$$h(y,\ \beta|\theta_0, \mathbf{x}) \propto \frac{1}{y}\ \beta^{c+n}\ exp\left\{-\beta(nc_0 + p + log(\frac{y}{\theta_0}))\right\} \tag{25}$$

Integrating out β from (25) and restoring the normalizing constant, the predictive density of y is

$$p(y|\theta_0, \mathbf{x}) = \frac{c+n}{nc_0+p}\,\frac{1}{y}\ \left(1 + \frac{log(y/\theta_0)}{nc_0+p}\right)^{-(c+n+1)},\ \theta_0 < y < \infty$$

Since $p(y|\theta_0, \mathbf{x})$ is strictly decreasing (see Figure 3), the HPD-prediction interval is $[\theta_0, y^*]$, where y^* is obtained from the given equation (26)

$$P(Y > y^*) = \left(1 + \frac{log(y/\theta_0)}{nc_0 + p}\right)^{-(c+n)} = \alpha \tag{26}$$

and hence the upper end of the $(1-\alpha)100\%$ HPD interval under NCP is

$$y_N^* = \theta_0 \exp\left[(nc_0 + p)(\alpha^{-\frac{1}{n+c}} - 1)\right] \tag{27}$$

Again using (1) and the posterior pdf (10) under the MIP, the equation (24) takes the form:

$$f(y, \beta|\theta_0, \mathbf{x}) \propto \frac{1}{y}\ \beta^{n+2}\ exp\left\{-(nc_0\beta + \beta\, log(y/\theta_0) + \frac{1}{\beta})\right\},$$

Thus, the corresponding predictive density of y is:

$$p(y|\theta_0, \mathbf{x}) \propto \int_0^\infty \frac{1}{y}\ \beta^{n+2}\ exp\left\{-(nc_0\beta + \beta\, log(y/\theta_0) + \frac{1}{\beta})\right\} d\beta$$

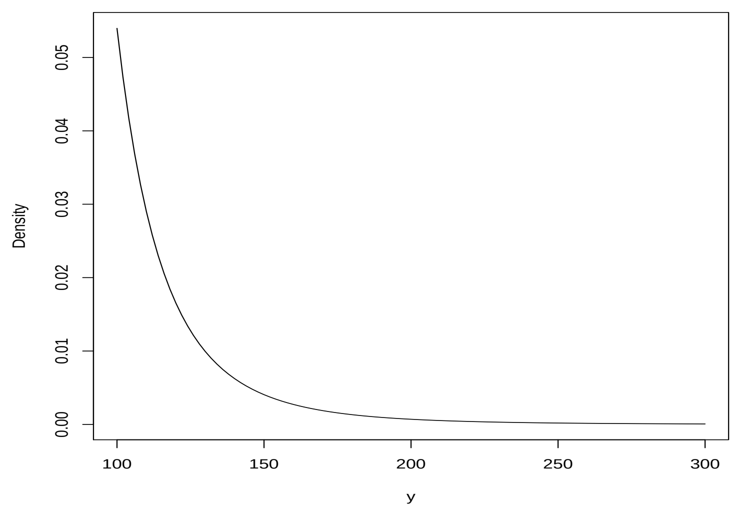

Figure 8: Predictive density for NCP ($p = 0$, $c = 2.19$).

Using the modified Bessel function of the third kind given in (9) with $\nu = -(n+3)$, $z = 2(nc_0 + log(y/\theta_0))$ and $a = (nc_0 + log(y/\theta_0))^{-1/2}$, the above integral takes the form:

$$p(y|\theta_0, \mathbf{x}) \propto \frac{1}{y}\ \left\{(nc_0 + log(y/\theta_0))^{-\frac{n+3}{2}}\right\}\ K_{n+3}\left\{2(nc_0 + log(y/\theta_0))^{1/2}\right\}$$

Using the asymptotic result (20) and restoring the normalizing constant the predictive density under the MIP is

$$p(y|\theta_0, \mathbf{x}) = \left(\frac{n+2}{nc_0}\right)\ \frac{1}{y}\ \left(1 + \frac{log\, y/\theta_0}{nc_0}\right)^{-(n+3)},\ \theta_0 < y < \infty$$

The upper end of the $(1-\alpha)100\%$ HPD interval under MIP is

$$y_M^* = \theta_0 \exp\left[nc_0(\alpha^{-\frac{1}{n+2}} - 1)\right] \tag{28}$$

Using the above wage data, the upper ends of a 95 percent HPD prediction interval for both NCP (for $p = 0$, $c = 2.19$) and MIP are computed as 178.91 and 179.56, respectively.

7 Conclusion

In this study we considered natural conjugate and minimal information priors for Bayesian estimation and prediction from Pareto distribution of the first kind. It is evident from the results that:

1. There were no computational difficulties or extreme behavior with either of the priors. There were also no gross over- or underestimation by either of these two sets of Bayes estimators.

2. The simulation study (Table 1) indicated some overestimation of the Pareto parameter, especially for small sample sizes, by the minimal information prior (MIP). The estimates under MIP are positively biased making them less appealing with respect to both bias and MSE.

3. For certain values of the prior parameters p and c ($p \approx 0$ and $c = 2.19$) the NCP is seen to produce HPD interval for β almost coinciding with that under MIP (Table 2). A similar pattern can be observed from Table 3 for HPD interval for the reliability function.

4. The length of the HPD interval for β (Table 2) decreases as the values of the prior parameters p and c under NCP get larger, therefore reducing uncertainties in estimation and prediction. The above is also true for HPD intervals for the reliability function (Table 3).

5. The Bayes estimators for the reliability function (Table 3) are extremely robust on the choice of the values of the prior parameters. It implies that the prior is dominated by the likelihood function, a situation known in literature (Box & Tiao, 1973, p.22) as the principle of "stable estimation".

6. The posterior density under NCP is gamma, which is easily amenable, and hence this prior might be preferred over MIP for Bayesian estimation and prediction from Pareto distribution of the first kind.

Acknowledgement

The authors express sincere thanks to the referee for valuable comments and suggestions.

References

1. Arnold, B. C. (1983). Pareto Distributions. International Cooperative Publishing House, Marcel Dekker.

2. Arnold, B.C. and Press, S. J. (1989). Bayesian estimation and prediction for Pareto data. Journal of the Americal Statistical Association, 84, 1079-1084.

3. Bain, L. J. and Englehardt, M. (1992). Introduction to probability and Mathematical Statistics. 2nd edition, Duxbury Press.

4. Box G.E.P. and Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis, Addison-Wesley.

5. Charek, D. J., Moore, A. H. and Coleman, J. W. (1988). A comparison of Estimation techniques for the three-parameter Pareto distribution. Comm. Statist-Simulation, 17(4), 1395-1407.

6. Cohen, A. C. and Whitten, B. J. (1988). Parameter estimation in reliability and life span models. Marcel Dekker, N.Y.

7. Copland, T. E. (1965). Asymptotic Expansions. Cambridge University Press.

8. Cox, D. R. and Oakes, D. (1984). Analysis of Survival Data. Chapman Hall, London.

9. Cramer, J. S. (1971). Empirical Economics. Amsterdam: North-Holland.

10. Davis, T. H. and Feldstein, L. M. (1979). The generalized Pareto law as a model for progressively censored survival data. Biometrika, 66(2), 299-306.

11. Dyer, D. (1981). Structural probability for the strong Pareto law, Canadian Journal of Statistics, 9, 71-77.

12. Erdelyi, A. (1953). Higher transcendental functions, vol 2. McGraw Hill, N. Y.

13. Geisser, S. (1984). Predicting Pareto and exponential observables. Canadian J. Statist., 12, 143-152.

14. Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). Continuous Univariate Distributions, vol. 1, 2nd ed. Wiley Interscience Publication.