*Aligarh Journal of Statistics*
*Vol.* 31 (2011), 17-29

# ESTIMATION OF PARAMETERS OF GAMMA DISTRIBUTION IN THE PRESENCE OF OUTLIERS IN RIGHT CENSORED SAMPLES

U. J. Dixit and M. Jabbari Nooghabi

For homogenous data, Wilks, Gnanadesikan and Huyett (1962) derived the maximum likelihood estimator based on order statistics $(X_{(1)}, X_{(2)}, ..., X_{(n)})$, $(r < n)$. Here we deal with the estimation of these parameters in the case of right censored samples for non-homogeneous data.

## 1. INTRODUCTION

Many authors have considered the problem of estimation of the parameters of gamma distribution. For gamma distribution considerable amount of literature is available on estimation of shape $p$ and scale $\sigma$. Bowman and Shenton (1987) give references to most of papers in this area. There are many situations in which it is reasonable to assume that the items may not be homogenous and hence the assumption of independent identically distributed random variable may be unrealistic and then the model may have to be modified suitably. Gather and Kale (1988) had considered the maximum likelihood estimator in the presence of $k$ outliers and Dixit (1989) considered the estimation of the parameters of gamma distribution in the presence of $k$ outliers. Dixit (1991) considered the estimation of power of the scale parameter of the gamma distribution in the presence of $k$ outliers Many authors have considered the problem of estimation of the mean of an exponential distribution in the presence of an outlying observation having higher expected values than the others. We refer Kale and Sinha (1971), Joshi (1972), Chikkagoudar and Kunchar (1980), Rauhut (1982), Dixit and Nasiri (2001) and Dixit, Ali and Woo (2003). Barnett and Lewis (1984, p. 146) remarked that outliers in gamma distribution with an arbitrary shape parameter arise with skew distributed data, for which a gamma distribution is often a useful pragmatic model. Outliers in gamma distribution arise in any context where Poisson processes are appropriate basic models, for example, in studying traffic flow, failures of electronic equipment, biological aggregation or even death from horse kicks.

Suppose experimental animals are subjected to massive dose of radiation and their survival times are recorded. During the administering of radiation dose, it is known that exactly $k$ animals received a dose of radiation far in excess to others. It may then be assumed that the expected survival times of the overdosed animals are different than the other $(n-k)$ animals. Gross, Hunt and Odeh (1986) have considered the above example when $k=1$. For definiteness we assume that $(X_1, X_2, ..., X_n)$ contains exactly $k$ (known), the number of outliers but the outliers themselves are not known whose expected life is large (or small) as compared to the expected life of rest of them. We assume that the random variables $(X_1, X_2, ..., X_n)$ are such that $k$ of them are distributed as:

$$f(x; p+b, \rho) = \frac{x^{p+b-1} e^{-\frac{x}{\sigma}}}{\sigma^{p+b}\Gamma(p+b)} \tag{1}$$

where $x>0$, $\sigma>0$, $p+b>0$, $b \neq 0$ and the remaining $(n-k)$ random variables are distributed as:

$$g(x; p, \sigma) = \frac{x^{p-1} e^{-\frac{x}{\sigma}}}{\sigma^{p}\Gamma(p)} \tag{2}$$

where $x>0$, $\sigma>0$ and $p>0$.

For homogenous case Wilk, Gnanadesiken and Huyett (1962) derived the maximum likelihood estimates of $p$ and $\sigma$ based on order statistics $(X_{(1)}, X_{(2)}, ..., X_{(n)})$, $(r<n)$ from (1).

## 2. JOINT DISTRIBUTION OF $(X_{(1)}, X_{(2)}, ..., X_{(r)})$ WITH $k$ OUTLIERS

Suppose a random sample of size $n$ (known) is taken from the outlier model are represented in (1) and (2). Suppose now that the observations are available only for the $r-$th smallest values of this random sample. Let these $r$ order statistics be denoted by $X_{(1)}, X_{(2)}, ..., X_{(r)}$, where $0 \leq X_{(1)} \leq X_{(2)} \leq ... \leq X_{(r)}$.

The problem considered is

Given the values of $X_{(1)}, X_{(2)}, ..., X_{(r)}$, $n$ and $k$ to find the maximum likelihood estimate of $p$, $\sigma$ and $b$. The joint distribution of the order statistics $X_{(1)}, X_{(2)}, ..., X_{(r)}$ in a sample of size $n$ is such that $k$ of them is distributed as $G(x)$ (cumulative distribution function $(CDF)$ of $g$) and remaining $(n-k)$ is distributed as $F(x)$ $(CDF$ of $f)$. Out of the first $r$ order statistics, let $A_j$ be the event that $j$ of them are $G(x)$ where $j$ is $Max(0, r-n+k) \leq j \leq Min(k, r)$. The contribution of $A_j$ to the joint probability density function (pdf) of $X_{(1)}, X_{(2)}, ..., X_{(r)}$ is

$$a_j \prod_{t=1}^{j} \frac{g(x_{(i_t)})}{f(x_{(i_t)})} \prod_{i=1}^{r} f(x_{(i)}) \left[1 - F(x_{(r)})\right]^{n-r-k+j} \left[1 - G(x_{(r)})\right]^{k-j}, \tag{3}$$

where $C(a,b) = \dfrac{a!}{b!(a-b)!}$, $a_j = C(n-k, r-j) C(k, j) j!(r-j)!$ and summing over $j$, the joint pdf of $(X_{(1)}, X_{(2)}, ..., X_{(r)})$ is given as:

$$f(x_{(1)}, x_{(2)}, ..., x_{(r)}) = \prod_{i=1}^{r} f(x_{(i)}) \left[1 - F(x_{(r)})\right]^{n-r-k} \left[1 - G(x_{(r)})\right]^{k}$$

$$\times \sum_j a_j \sum_{i_1, i_2, ..., i_j} \prod_{t=1}^{j} \frac{g(x_{(i_t)})}{f(x_{(i_t)})} \left[\frac{1 - F(x_{(r)})}{1 - G(x_{(r)})}\right]^{j} \tag{4}$$

Note that $\sum\limits_{i_1, i_2, ..., i_j} = \sum\limits_{i_i=1}^{r-k+1} \sum\limits_{i_2=i_1+1}^{r-k+2} ... \sum\limits_{i_j=i_{j-1}+1}^{r}$.

When $j = 0$, then $\sum\limits_{i_1, i_2, ..., i_j} \prod\limits_{t=1}^{j} \dfrac{g(x_{(i_t)})}{f(x_{(i_t)})} = 1$.

For brevity $x_{(1)}, x_{(2)}, ..., x_{(r)}$ is taken as $x_1, x_2, ..., x_r$.

In the Appendix it had been shown that

$$\int_{-\infty}^{\infty} \int_{x_1}^{\infty} ... \int_{x_{r-2}}^{\infty} \int_{x_{r-1}}^{\infty} f(x_1, x_2, ..., x_r) dx_1 dx_2 ... dx_r = 1.$$

Now, the likelihood function of $X_{(1)}, X_{(2)}, ..., X_{(r)}$ with (1) and (2) is given as:

$$L(x_1, x_2, ..., x_r; \sigma, p+b) = \exp\left\{-\sum_{i=1}^{r} \frac{x_i}{\sigma}\right\} \prod_{i=1}^{r} x_i^{p-1} \left[\int_{x_r}^{\infty} \frac{e^{-\frac{x_r}{\sigma}} x_r^{p-1}}{\sigma^p \Gamma(p)} dx_r\right]^{n-r-k}$$

$$\times \left[\int_{x_r}^{\infty} \frac{e^{-\frac{x_r}{\sigma}} x_r^{p+b-1}}{\sigma^{p+b} \Gamma(p+b)} dx_r\right]^{k} \sum_{j=0}^{k} \frac{(n-k)!}{(n-k-r+j)!} \frac{k!}{(k-j)!}$$

$$\times \sum_{(i_1, i_2, ..., i_j)} \prod_{t=1}^{j} x_{i_t}^{b} \left[\frac{\Gamma(p)}{\Gamma(p+b)\sigma^b}\right]^{j} \left[\int_{x_r}^{\infty} \frac{e^{-\frac{x_r}{\sigma}} x_r^{p-1}}{\sigma^p \Gamma(p)} dx_r\right]^{j}$$

$$\times\left[\int_{x_r}^{\infty}\frac{e^{-\frac{x_r}{\sigma}}x_r^{p+b-1}}{\sigma^{p+b}\Gamma(p+b)}dx_r\right]^{-j}$$

$$=\exp\{-r\rho S\}R^{r(p-1)}x_r^{-r}\rho^{np+bk}\left[J(p,\rho)\right]^{n-r-k}\left[J(p+b,\rho)\right]^{k}$$

$$\times\left[\Gamma(p)\right]^{k-n}\left[\Gamma(p+b)\right]^{-k}\sum_{j=0}^{k}a_j\sum_{(i_1,i_2,\ldots,i_j)}\prod_{t=1}^{j}x_{i_t}^{b}\frac{1}{x_r^{jb}}\left[\frac{J(p,\rho)}{J(p+b,\rho)}\right]^{j} \quad (5)$$

where $R=\frac{\left[\prod_{i=1}^{r}x_i\right]^{\frac{1}{r}}}{x_r}$, $\rho=\frac{x_r}{\sigma}$, $S=\frac{\prod_{i=1}^{r}x_i}{rx_r}$ and $J(p,\rho)=\int_1^{\infty}e^{-\rho t}t^{p-1}dt$.

As a check, consider $b=0$, i.e. $k=0$ in (4), then we get

$$L(x_1,x_2,\ldots,x_r;\sigma,p)=\frac{n!}{(n-r)!}\exp\{-r\rho S\}\frac{R^{r(p-1)}}{x_r^{r}}\frac{\rho^{np}}{\left[\Gamma(p)\right]^{n}}\left[J(p,\rho)\right]^{n-r},$$

which is the joint likelihood function of $(X_{(1)},X_{(2)},\ldots,X_{(r)})$ in the homogenous case.

## 3. MAXIMUM LIKELIHOOD EQUATIONS

The log likelihood function obtained from (5) is

$$Ln(L(p,b,\rho\mid x))=-r\rho S+r(p-1)Ln(R)-rLn(x_r)+(np+bk)Ln(\rho)$$

$$+(n-r-k)Ln(J(p,\rho))+k\left[J(p+b,\rho)\right]-(n-k)Ln(\Gamma(p))$$

$$-kLn(\Gamma(p+b))+Ln(U). \quad (6)$$

The maximum likelihood equations obtained from (6) are

$$Ln(R)=\frac{n-k}{r}\Psi(p)+\frac{k}{r}\Psi(p+b)-\frac{n}{r}Ln(\rho)-\frac{n-r-k}{r}\frac{J'_p(p,\rho)}{J(p,\rho)}$$

$$-\frac{k}{r}\frac{J'_p(p+b,\rho)}{J(p+b,\rho)}-\frac{U'_p}{U}, \quad (7)$$

$$S=\frac{p}{\rho}-\frac{e^{-\rho}}{J(p,\rho)}\left[\frac{n}{r}-\frac{k}{r}-1\right]-\frac{ke^{-\rho}}{r\rho J(p+b,\rho)}+\frac{U'_\rho}{rU}, \quad (8)$$

$$Ln(\rho)+\frac{J'_b(p+b)}{J(p+b,\rho)}-\Psi(p+b)+\frac{U'_b}{kU}=0 \quad (9)$$

where $\Psi(p)=\dfrac{dLn(\Gamma(p))}{dp}$ is digamma function of $p$. Calculations details are given in Abramowitz and Stegun (1972, pp. 810) and Dixit (1989).

Also

$$J'_p(p,\rho)=\int_1^{\infty} Ln(t)e^{-\rho t}t^{p-1}dt\,, \tag{10}$$

$$J'_p(p+b,\rho)=\int_1^{\infty} Ln(t)e^{-\rho t}t^{p+b-1}dt\,, \tag{11}$$

$$\frac{J'_p(p,\rho)}{J(p,\rho)}=-\frac{e^{-\rho}}{\rho J(p,\rho)}-\frac{p}{\rho}, \tag{12}$$

$$\frac{J'_p(p+b,\rho)}{J(p+b,\rho)}=-\frac{e^{-\rho}}{\rho J(p+b,\rho)}-\frac{p+b}{\rho}, \tag{13}$$

$$U=\sum_{j=0}^{k}\frac{(n-k)!}{(n-k-r+j)!}\frac{k!}{(k-j)!}\sum_{(i_1,i_2,\ldots,i_j)}\prod_{t=1}^{j}x_{i_t}^{b}\frac{1}{x_r^{jb}}\left[\frac{J(p,\rho)}{J(p+b,\rho)}\right]^{j}, \tag{14}$$

$$U'_p=\sum_{j=0}^{k}\frac{(n-k)!}{(n-k-r+j)!}\frac{k!}{(k-j)!}\sum_{(i_1,i_2,\ldots,i_j)}\prod_{t=1}^{j}x_{i_t}^{b}\frac{j}{x_r^{jb}}\left[\frac{J(p,\rho)}{J(p+b,\rho)}\right]^{j}$$

$$\times\left[\frac{J'_p(p,\rho)}{J(p,\rho)}-\frac{J'_p(p+b,\rho)}{J(p+b,\rho)}\right] \tag{15}$$

$$U'_\rho=\sum_{j=0}^{k}\frac{(n-k)!}{(n-k-r+j)!}\frac{k!}{(k-j)!}\sum_{(i_1,i_2,\ldots i_j)}\prod_{t=1}^{j}x_{i_t}^{b}\frac{j}{x_r^{jb}}\left[\frac{J(p,\rho)}{J(p+b,\rho)}\right]^{j}$$

$$-\left[\frac{e^{-\rho}}{J(p+b,\rho)}-\frac{e^{-\rho}}{J(p,\rho)}+\frac{b}{\rho}\right]\;, \tag{16}$$

$$U'_b=\sum_{j=0}^{k}\frac{(n-k)!}{(n-k-r+j)!}\frac{k!}{(k-j)!}\left\{\sum_{(i_1,i_2,\ldots i_j)}\prod_{t=1}^{j}x_{i_t}^{b}\sum_{t=1}^{j}\frac{Ln(x_{i_t})}{x_r^{jb}}\left[\frac{J(p,\rho)}{J(p+b,\rho)}\right]^{j}\right.$$

$$\left.-\sum_{(i_1,i_2,\ldots i_j)}\prod_{t=1}^{j}x_{i_t}^{b}\left[\frac{J(p,\rho)}{J(p+b,\rho)}\right]^{j}\frac{1}{x_r^{jb}}\left[\frac{bj}{x_r}+\frac{j\,J'_b(p+b,\rho)}{J(p+b,\rho)}\right]\right\} \tag{17}$$

$$J'_x(x,y)=\frac{\partial J(x,y)}{\partial x},\qquad J'_y(x,y)=\frac{\partial J(x,y)}{\partial y},\; J'_z(x+z,y)=\frac{\partial J(x+z,y)}{\partial z},$$

$$J'_y(x+z,y)=\frac{\partial J(x+z,y)}{\partial y},\ U'_p=\frac{\partial U}{\partial p},\ U'_b=\frac{\partial U}{\partial b}\ \text{and}\ U'_\rho=\frac{\partial U}{\partial \rho}.$$

Consider for $k=1$

$$\frac{U'_p}{U}=\frac{DUP1}{(n-r)+U1}, \tag{18}$$

$$\frac{U'_\rho}{U}=\frac{DU\rho 1}{(n-r)+U1}, \tag{19}$$

and

$$\frac{U'_b}{U}=\frac{DUB1}{(n-r)+U1}, \tag{20}$$

where

$$U=\frac{(n-1)!}{(n-r-1)!}+\frac{(n-1)!}{(n-r)!}U1,\ U1=\sum_{i=1}^{r}x_i^b x_r^{-b}H(p,b,\rho)\ \text{and}$$

$$H(p,b,\rho)=\frac{J(p,\rho)}{J(p+b,\rho)},$$

$$DUP1=\sum_{i=1}^{r}x_i^b x_r^{-b}H(p,b,\rho)M(p,b,\rho),$$

where

$$M(p,b,\rho)=\frac{J'(p,\rho)}{J(p,\rho)}-\frac{J'(p+b,\rho)}{J(p+b,\rho)},$$

$$DU\rho 1=\sum_{i=1}^{r}x_i^b x_r^{-b}H(p,b,\rho)Q(p,b,\rho),$$

where

$$Q(p,b,\rho)=\frac{e^{-\rho}}{\rho J(p+b,\rho)}-\frac{e^{-\rho}}{\rho J(p,\rho)}+\frac{b}{\rho},$$

and

$$DUB1=x_r^{-b}H(p,b,\rho)\left[\sum_{i=1}^{r}x_i^b Ln(x_i)-\sum_{i=1}^{r}x_i^b\left\{\frac{b}{x_r}+\frac{J'_b(p+b,\rho)}{J(p+b,\rho)}\right\}\right]$$

Now, $k=2$

$$\frac{U'_p}{U} = \frac{2(n-r)DUP1 + 4(DUP2)}{(n-r)(n-r-1) + 2(n-r)U1 + 2(U2)}, \tag{21}$$

$$\frac{U'_\rho}{U} = \frac{2(n-r)DU\rho 1 + 4(DU\rho 2)}{(n-r)(n-r-1) + 2(n-r)U1 + 2(U2)}, \tag{22}$$

$$\frac{U'_b}{U} = \frac{2(n-r)DUB1 + 2(DUB2)}{(n-r)(n-r-1) + 2(n-r)U1 + 2(U2)}, \tag{23}$$

where

$$U = \frac{(n-2)!}{(n-r-2)!} + \frac{(n-2)!}{(n-r-1)!}2(U1) + \frac{(n-2)!}{(n-r)!}2(U2), \tag{24}$$

$$U2 = x_r^{-2b} H^2(p,b,\rho) \sum_{i=1}^{r-1} \sum_{u=i+1}^{r} x_i^b x_u^b, \tag{25}$$

$$DUP2 = x_r^{-2b} H^2(p,b,\rho) M(p,b,\rho) \sum_{i=1}^{r-1} \sum_{u=i+1}^{r} x_i^b x_u^b, \tag{26}$$

$$DU\rho 2 = x_r^{-2b} H^2(p,b,\rho) Q(p,b,\rho) \sum_{i=1}^{r-1} \sum_{u=i+1}^{r} x_i^b x_u^b, \tag{27}$$

and

$$DUB2 = x_r^{-2b} H^2(p,b,\rho) \sum_{i=1}^{r-1} \sum_{u=i+1}^{r} x_i^b x_u^b [Ln(x_i) + Ln(x_u)] - \sum_{i=1}^{r-1} \sum_{u=i+1}^{r} x_i^b x_u^b \left[ \frac{2b}{x_r} + \frac{2J'_b(p+b,\rho)}{J(p+b,\rho)} \right], \tag{28}$$

Similarly, one can write the above expressions for any value of $k(k < n)$.

The maximum likelihood estimates of $p$, $b$ and $\rho$ are then defined by the solution of the simultaneous equations (7), (8) and (9).

## 4. THE SOLUTIONS OF THE MAXIMUM LIKELIHOOD EQUATION

An iterative procedure was used in obtaining the necessary roots. Let $R = R(p,b,\rho)$, $S = S(p,b,\rho)$ and $Z = Z(p,b,\rho)$ to represent the equations (6), (7) and (8) respectively.

The functions $R(p,b,\rho)$, $S(p,b,\rho)$ and $Z(p,b,\rho)$ involve $\Gamma(p)$, $\Psi(p)$, $\Psi(p+b)$, $J(p,\rho)$, $J'(p,\rho)$, $J(p+b,\rho)$, $J'(p+b,\rho)$.

The numerical approximations are same as given by Wilk, Gnanadesikan and Huyett (1962).

Let $(p,b,\rho)$ be sufficiently close to $(p_0,b_0,\rho_0)$. Then following will be the reasonable approximations.

Hence,

$$R(p,b,\rho)=R(p_0,b_0,\rho_0)+(p-\rho_0)\frac{\partial R}{\partial p}|_{p=p_0}+(b-b_0)\frac{\partial R}{\partial b}|_{b=b_0} +(\rho-\rho_0)\frac{\partial R}{\partial \rho}|_{\rho=\rho_0} \tag{29}$$

$$S(p,b,\rho)=S(p_0,b_0,\rho_0)+(p-p_0)\frac{\partial S}{\partial p}|_{p=p_0}+(b-b_0)\frac{\partial S}{\partial b}|_{b=b_0} +(\rho-\rho_0)\frac{\partial R}{\partial \rho}|_{\rho=\rho_0} \tag{30}$$

and

$$Z(p,b,\rho)=Z(p_0,b_0,\rho_0)+(p-p_0)\frac{\partial Z}{\partial p}|_{p=p_0}+(b-b_0)\frac{\partial Z}{\partial b}|_{b=b_0} +(\rho-\rho_0)\frac{\partial R}{\partial \rho}|_{\rho=\rho_0} \tag{31}$$

Thus, suppose that values $(x_1,x_2,...,x_{1r})$ are given and the values of $p$, $b$ and $\rho$ is to satisfy $R=R(p,b,\rho)$, $S=S(p,b,\rho)$ and $Z=Z(p,b,\rho)$ are wanted. Let the values $p_0$, $b_0$ and $\rho_0$ be known such that $R_{000}=R(p_0,b_0,\rho_0)$, $S_{000}=S(p_0,b_0,\rho_0)$ and $Z_{000}=Z(p_0,b_0,\rho_0)$ are close to $R$, $S$ and $Z$, respectively. Further $p_1$, $b_1$ and $\rho_1$ are close to $p_0$, $b_0$ and $\rho_0$ respectively are selected and the corresponding

$$R_{100}=R(p_1,b_0,\rho_0),\ R_{010}=R(p_0,b_1,\rho_0),\ R_{001}=R(p_0,b_0,\rho_1),$$

$$S_{100}=S(p_1,b_0,\rho_0),\ S_{010}=S(p_0,b_1,\rho_0),\ S_{001}=S(p_0,b_0,\rho_1),$$

$$Z_{100}=Z(p_1,b_0,\rho_0),\ Z_{010}=Z(p_0,b_1,\rho_0),\ Z_{001}=Z(p_0,b_0,\rho_1),$$

are computed. The partial derivatives here can themselves be approximated by suitable divided differences. Therefore

$$\frac{\nabla R}{\nabla p}=\frac{R_{100}-R_{000}}{p_1-p_0},\ \frac{\nabla R}{\nabla b}=\frac{R_{010}-R_{000}}{b_1-b_0},\ \frac{\nabla R}{\nabla \rho}=\frac{R_{001}-R_{000}}{\rho_1-\rho_0}, \tag{32a}$$

$$\frac{\nabla S}{\nabla p}=\frac{S_{100}-S_{000}}{p_1-p_0},\ \frac{\nabla S}{\nabla b}=\frac{S_{010}-S_{000}}{b_1-b_0},\ \frac{\nabla S}{\nabla \rho}=\frac{S_{001}-S_{000}}{\rho_1-\rho_0}, \tag{32b}$$

$$\frac{\nabla Z}{\nabla p}=\frac{Z_{100}-Z_{000}}{p_1-p_0},\ \frac{\nabla Z}{\nabla b}=\frac{Z_{010}-Z_{000}}{b_1-b_0},\ \frac{\nabla Z}{\nabla \rho}=\frac{Z_{001}-Z_{000}}{\rho_1-\rho_0}. \tag{32c}$$

Then the solution of three simultaneous linear equations

$$A\frac{\nabla R}{\nabla p}+B\frac{\nabla R}{\nabla b}+C\frac{\nabla R}{\nabla \rho}=R-R_{000}, \tag{33}$$

$$A\frac{\nabla S}{\nabla p}+B\frac{\nabla S}{\nabla b}+C\frac{\nabla S}{\nabla \rho}=S-S_{000}, \tag{34}$$

and

$$A\frac{\nabla Z}{\nabla p}+B\frac{\nabla Z}{\nabla b}+C\frac{\nabla Z}{\nabla \rho}=Z-Z_{000} \tag{35}$$

will yield correction $A$, $B$ and $C$ such that one might expect $p^*=p_0+A$, $b^*=b_0+B$ and $\rho^*=\rho_0+C$ to be improved to the required roots of $p$, $b$ and $\rho$. Not that

$$\hat{\sigma}=\frac{x_r}{\rho} \tag{36}$$

**Note:** If $k$ is unknown then $k$ can be selected by evaluating the likelihood function for different values of $k$ and choosing the one that maximizes the likelihood.

## 5. APPENDIX

The joint distribution of $(X_{(1)},X_{(2)},...,X_{(r)})$ is given as:

$$f(x_{(1)},x_{(2)},...,x_{(r)})=\prod_{i=1}^{r}f(x_{(i)})\left[1-F(x_{(r)})\right]^{n-r-k}\left[1-G(x_{(r)})\right]^{k}$$

$$\times\sum_{i_1,i_2,...,i_j}\prod_{t=1}^{j}\frac{g(x_{(i_t)})}{f(x_{(i_t)})}\left[\frac{1-F(x_{(r)})}{1-G(x_{(r)})}\right]^{j}, \tag{37}$$

where $-\infty\le x_{(1)}\le x_{(2)}\le...\le x_{(r)}<\infty$ and $a_j=C(n-k,r-j)C(k,j)j!(r-j)!$,

when $j=0$ then $\sum_{i_1,i_2,...,i_j}\prod_{t=1}^{j}\frac{g(x_{(i_t)})}{f(x_{(i_t)})}=1$.

To prove that

$$\int_{-\infty}^{\infty}\int_{x_1}^{\infty}\int_{x_{(2)}}^{\infty}\cdots\int_{x_{(r-2)}}^{\infty}\int_{x_{(r-1)}}^{\infty} f(x_1,x_2,\ldots,x_r)\prod_{i=1}^{r} dx_{(i)} = 1, \tag{38}$$

let $B_j^r$ be the event that there are $j$ out of $r$ order statistics which belong to $g(x)$ then $x_{(r)}$ may be on $g(x)$ or it may not be on $g(x)$. The contribution of these two events is as:

$$L_j^r = L_{g,j-1}^{(r)} + L_{f,j}^{(r)},$$

where $L_{g,j-1}^{(r)}$ is the sum corresponding to the $r-$th observation on $g(x)$ and $(j-1)$ of $g(x)$'s are less than $x_{(r)}$ and $L_{f,j}^{(r)}$ is the sum corresponding to the $r-$th observation on $f(x)$ and $j$ of $g(x)$ are less than $x_{(r)}$. Now

$$\sum_{j=0}^{k}\int_{x_{(r-1)}}^{\infty} L_j^{(r)} dx_{(r)} = \sum_{j=0}^{k}\int_{x_{(r-1)}}^{\infty} L_{g,j-1}^{(r)} dx_{(r)} + \sum_{j=0}^{k}\int_{x_{(r-1)}}^{\infty} L_{f,j}^{(r)} dx_{(r)}$$

and

$$L_{g,j-1}^{(r)} = g(x_{(r)})\left[1-F(x_{(r)})\right]^{m-r+j}\left[1-G(x_{(r)})\right]^{k-j}$$

where $m = (n-k)$.

Therefore

$$\begin{aligned}\int_{x_{(r-1)}}^{\infty} L_{g,j-1}^{(r)} dx_{(r)} &= \int_{x_{(r-1)}}^{\infty} g(x_{(r)})\left[1-F(x_{(r)})\right]^{m-r+j}\left[1-G(x_{(r)})\right]^{k-j} dx_{(r)}\\ &= \frac{\left[1-G(x_{(r-1)})\right]^{k-j+1}}{k-j+1}\left[1-F(x_{(r-1)})\right]^{m-r+j}\\ &\quad - \frac{(m-r+j)}{k-j+1}\int_{x_{(r-1)}}^{\infty} f(x_{(r)})\left[1-F(x_{(r)})\right]^{m-r+j-1}\left[1-G(x_{(r)})\right]^{k-j+1} dx_{(r)}\end{aligned}$$

Therefore

$$\begin{aligned}\sum_{j=1}^{k}\int_{x_{(r-1)}}^{\infty} L_{g,j-1}^{(r)} dx_{(r)} &= \sum_{j=1}^{k}\left[L_{j-1}^{(r-1)} - \int_{x_{(r-1)}}^{\infty} L_{f,j-1}^{(r)} dx_{(r)}\right]\\ &= \sum_{j=0}^{k-1} L_{j-1}^{(r-1)} - \sum_{j=0}^{k-1}\int_{x_{(r-1)}}^{\infty} L_{f,j}^{(r)} dx_{(r)}\end{aligned}$$

and

$$\sum_{j=0}^{k}\int_{x_{(r-1)}}^{r} L_j^{(r)} dx_{(r)} = \sum_{j=0}^{k-1}\left[ L_j^{(r-1)} - \sum_{j=0}^{k-1}\int_{x_{(r-1)}}^{\infty} L_{f,j}^{(r)} dx_{(r)} \right] + \sum_{j=0}^{k}\int_{x_{(r-1)}}^{\infty} L_{f,j}^{(r)} dx_{(r)}$$

$$= \sum_{j=0}^{k-1} L_j^{(r-1)} + \int_{x_{(r-1)}}^{\infty} L_{f,k}^{(r)} dx_{(r)} .$$

Now

$$\int_{x_{(r-1)}}^{\infty} f(x_{(r)})\left[1 - F(x_{(r)})\right]^{m-r+k} dx_{(r)} = \frac{\left[1 - F(x_{(r-1)})\right]^{m-r+k+1}}{m-r+k+1} = L_k^{(r-1)} .$$

Hence

$$\sum_{j=0}^{k}\int_{x_{(r-1)}}^{\infty} L_j^{(r)} dx_{(r)} = \sum_{j=1}^{k-1} L_j^{(r-1)} + L_k^{(r-1)} = \sum_{j=0}^{k} L_j^{(r-1)} .$$

Similarly

$$\sum_{j=0}^{k}\int_{x_{(r-2)}}^{\infty} L_j^{(r-1)} dx_{(r-1)} = \sum_{j=1}^{k-1} L_j^{(r-2)} .$$

Therefore

$$\sum_{j=0}^{k}\int_{x_{(1)}}^{\infty} L_j^{2} dx_{(1)} = \sum_{j=1}^{k} L_j^{(1)} .$$

Now

$$\sum_{j=0}^{k} L_j^{(1)} = kg(x_{(1)})\left[1 - F(x_{(1)})\right]^m \left[1 - G(x_{(1)})\right]^{k-1} + mf(x_{(1)})\left[1 - G(x_{(1)})\right]^k \left[1 - F(x_{(1)})\right]^{m-1}$$

and

$$\int_{-\infty}^{\infty} \sum_{j=0}^{k} L_j^{1} = \left[1 - F(x_{(1)})\right]^m \left[1 - G(x_{(1)})\right]^k \Big|_{-\infty}^{\infty}$$

$$- m\int_{-\infty}^{\infty} f(x_{(1)})\left[1 - F(x_{(1)})\right]^{m-1}\left[1 - G(x_{(1)})\right]^k dx_{(1)}$$

$$+ m\int_{-\infty}^{\infty} f(x_{(1)})\left[1 - F(x_{(1)})\right]^{m-1}\left[1 - G(x_{(1)})\right]^k dx_{(1)}$$

$$= 1$$

## 6. AN EXAMPLE

The following is the life distribution (in units of 100 hours) of 25 electronic tubes. But we could observe that only 20 electronic tubes due to insufficient power supply. Further, it is also observe that some tubes (say one or two) are of different quality. This data follows a gamma distribution.

| | | | | |
|---|---|---|---|---|
| 0.1415 | 0.5937 | 2.3467 | 3.1356 | 3.5681 |
| 0.3484 | 1.1045 | 2.4651 | 3.2259 | 3.7287 |
| 0.3994 | 1.7323 | 2.6155 | 3.4177 | 9.2817 |
| 0.4174 | 1.8348 | 2.7425 | 3.5551 | 9.3208 |

Then form the above data $R = 0.19$, $S = 539.35$ and $Z = 2.57$. Then for $k = 1$, $r = 20$, $p^*$, $b^*$ and $\rho^*$ from (33), (34) and (35) are $1.2164$, $0.2212$ and $4.0485$, respectively. Next for $k = 2$, $r = 20$, $p^*$, $b^*$ and $\rho^*$ are $1.2238$, $0.1869$ and $4.0747$, respectively.

## 7. Acknowledgment

The author wishes to thank Dr. Y.S. Sathe, former Professor and Head, Department of Statistics, University of Mumbai for his valuable comments and suggestions in preparing this paper. We are also thankful to Professor A.H. Khan and the referees for their valuable comments.

## REFERENCES

Barnett, V. and Lewis, T., (1984): Outliers in Statistical Data, John Wiley and Sons, New York.

Bowman, K.O. and Shenton, L.R. (1987): Properties of estimators for the Gamma distribution. *Marcel Dekker Inc.*, New York.

Chikkagouder, M.S. and Kunchur, S.H. (1980): Estimation of the mean of an exponential distribution in the presence of an outlier, *Canadian Journal of Statistics,* 8(1), 59-63.

Dixit, U.J. (1989): Estimation of the parameters of Gamma distribution in the presence of k outliers, *Commun. Statist. Theor. Meth.*, **18**, 8, 3071-3085.

Dixit, U.J. (1991): On the Estimation of the power of the scale parameter of the Gamma distribution in the Presence of Outliers, *Commun. Statist. Theor. Meth.*, **20,** 4, 1315-1328.

Dixit, U.J. and Nasiri, F.P. (2001): Estimation of parameters of the exponential distribution in the presence of outliers generated from uniform distribution, *Metron*,. **LIX**, 3-4, 187-198.

Dixit, U.J., Ali, M.M. and Woo Jungsoo. (2003): Efficient Estimation of parameters of a uniform distribution in the presence of outliers, *Soochow Journal of Mathematics*, **29(4)**, 363-369.

Gather, U. and Kale, B.K. (1988): Maximum likelihood estimation in the presence of outliers, *Commun. Statist. Theory. Meth.*, **17(11)**, 3767-3784.

Gross, A.J., Hunt, H.H. and Odeh, R.E. (1986): The correlation coefficient between the smallest and largest observations when (N-1) of the n observations are iid exponentially distributed, *Commun. Statist. Theor. Meth.*, **15(4)**, 1113-1123.

Joshi, P.C., (1972): Efficient estimation of the mean of an exponential distribution when an outlier is present, *Technometrics*, **41**, 13, 143.

Kale, B.K. and Sinha, S.K. (1971): Estimation of expected life in the presence of an outlier observation, *Technometrics*, **13**, A, 755-759.

Rauhut, B.O. (1982): Estiamtion of the mean of an exponentail distribution with an outlying observation, *Commun. Statist. Theor. Meth.*, **11**, 1439-1452.

Wilk, M.B., Gnanadesikan, R. and Huyett, M.J. (1962): Estimation of the parameters of Gamma distribution using order statistics, *Biometrika*, **49**, 525-545.

Received : 19-07-2010

Revised : 04-12-2010

U. J. Dixit and M. Jabbari Nooghabi

Department of Statistics, University of Mumbai,
Mumbai-India.

Email: ulhasdixit@yahoo.co.in

Email: jabbarinm@yahoo.com