ESTIMATION OF $P[X \leq Y]$FOR THE UNIFORM DISTRIBUTION IN THE PRESENCE OF OUTLIERS FROM UNIFORM DISTRIBUTION

K.D.PHAL

Abstract: The maximum likelihood and uniformly minimum variance unbiased estimator (U.M.V.U.E) of P[X ≤ Y] are derived ,where both X and Y have uniform distribution and outliers are generated from Uniform distribution .It is shown that U.M.V.U.E is better than M.L.E when one parameter of Uniform distribution is known. It is shown that estimator of P[X ≤ Y] is consistent.

Keywords: Consistent estimator, Maximum likelihood estimator (M.L.E), Outliers ,Uniformly Minimum Variance Unbiased estimator (U.M.V.U.E).

Introduction: The probability of estimating R=P[X ≤ Y] plays an important role in reliability analysis. Ooms and Moore [2] had shown that as a plant develops into the reproductive phase of growth ,a mat of smaller roots grows near the surface to a depth of $\frac{1}{6}^{\text{th}}$ of maximum depth achieved. The mass of roots has a uniform distribution.

Dixit etal[2] assumes that a set of random variables (X_1 ,X_2 ,....... X_n) represents the mass of roots. Some of these roots have different mass of roots .Therefore , it has different uniform distribution with some α . Dixit and Phal [3] , have estimated R=P[X ≤ Y] in the presence of outliers from Generalized uniform distribution. Here we assume that out of n random variables some say kare with different distribution .Further , we assume that these k random variables are distributed with p.d.f

g(x ,θ_1 ,α) where,

$$g(x ,\theta_1 ,\alpha) = \begin{cases} \frac{1}{\theta_1\alpha} & 0\leq x \leq \theta_1\alpha. \ .\theta_1 ,\alpha >0 \\ 0 & Otherwise \end{cases}$$

(1.1)

Then the remaining (n-k)random variables are distributed with p.d. f. f(x ,θ_1)

$$f(x ,\theta_1) = \begin{cases} \frac{1}{\theta_1} & 0\leq x \leq \theta_1. \ .\theta_1 >0 \\ 0 & Otherwise \end{cases} \quad (1.2)$$

The joint distribution of (x_1,.......x_n) in the presence of k outliers as in Dixit et al (2003) is given by using (1.1) and (1.2) ,

$$f((x_1,....... x_n)\ \theta_1 \alpha) = \{h_k \prod I(\theta_1 - x_{A_i}) G(x,\theta_1 ,\alpha)$$

(1.3)

Where $h_k=\left[\binom{n}{k}\theta^n\alpha^k\right]^{-1}$,G(x ,$\theta_1\alpha$)=

$$\sum_{A_1=1}^{n-k+1}\sum_{A_2=A_1+1}....\sum_{A_k}\prod_{i=1}^{k} I(\alpha\theta_1 - x_{A_i})$$

$$I(u)=\begin{cases} 1 & u<0 \\ 0 & u\geq 0 \end{cases} \quad (1.4)$$

Let the set of random variables represents the mass of roots .We assume that these random variables are distributed with p.d.f. h(y ,θ_2) , where

$$h(y ,\theta_2) = \begin{cases} \frac{1}{\theta_2} & 0\leq y \leq \theta_2. \ .\theta_2 >0 \\ 0 & Otherwise \end{cases} \quad (1.5)$$

The joint distribution of (Y_1,Y_2,....Y_n) is given by using (1.5)

h(y ,θ_2) =

$$\begin{cases} \frac{\prod_{i=1}^{n} I(\theta_2 - y_i)}{\theta_2^{\,n}} & 0\leq y_i \leq \theta_2. \ ;\theta_2 >0 \\ 0 & Otherwise \end{cases}$$

(1.6)

Now the marginal p.d.f. of X from (1.1)and (1.2) is given as

q(x ,θ_1,α)=

$$\begin{cases} \frac{\bar{b}}{\theta_1} + \frac{b}{\alpha\theta_1} & 0\leq x \leq \theta_1\alpha. \ .\theta_1 ,\alpha >0 \\ 0 & Otherwise \end{cases} \quad (1.7)$$

Where $b = \frac{k}{n}$, $\bar{b} = 1 - b$

$$R=P[X \leq Y] = \int_0^{\theta_2}\int_0^{y} (\frac{\bar{b}}{\theta_1} + \frac{b}{\alpha\theta_1})\frac{1}{\theta_2} dy \quad (1.8)$$

2. Estimation of R when α is known

Theorem 1 :

M.L.E of R is given by $\hat{R}_M = \left(\bar{b} + \frac{b}{\alpha}\right)\left(\frac{Y_{(n)}}{2X_{(n)}}\right)$ (2.1)

Proof: If α is known , then the maximum likelihood estimator of $\theta_1 = X_{(n)}/\alpha$

Where $X_{(n)}$= Max($X_1, X_2, \ldots X_n$); See Dixit etal [1]

From (1.6) the M.L.E of ,$\theta_2$is $Y_{(n)}$ =Max ($Y_1, Y_2, \ldots Y_n$).

Hence the M.L.E of R is given by

$$\hat{R}_M = \left(\bar{b} + \frac{b}{\alpha}\right)\left(\frac{Y_{(n)}}{2X_{(n)}}\right)$$

Theorem 2 :

U.M.V.U.E of R is given by

$$\hat{R}_U = \left(\bar{b} + \frac{b}{\alpha}\right)\left(\frac{\theta_2^*}{2\theta_1^*}\right) \qquad (2.2)$$

Proof: If α is known , then the maximum likelihood estimator of $\theta^*_1 = X_{(n)}/e_{-1}$

Where$X_{(n)}$= Max($X_1, X_2, \ldots X_n$);

$\theta^*_2 = \frac{n+1}{n} Y_{(n)}$ where$Y_{(n)}$ =Max ($Y_1, Y_2, \ldots Y_n$).

Where

$e_r = n_o(r)\alpha_k(r) + n_k(r)\bar{\alpha}_k(r)$

$n_k(r) = \frac{n-k}{n-k-r}$;$n_o(r) = \frac{n}{n-r}$;$\alpha_k(r) = \alpha^{n-k-r}$;$\bar{\alpha}_k(r) = 1 - \alpha_k(r)$

Let T= $\frac{Y_{(n)}}{X_{(n)}}$ and the distribution of T is given by $h_1(t)$.Where$h_1(t)$ is given as

$$h_1(t) = \begin{cases} [C_1(1-\alpha^{2n-k}) + \alpha^{2n-k}\frac{n}{2}]\left(\frac{\theta_1}{\theta_2}\right)^n t^{n-1}; & 0 \le t < \frac{\theta_2}{\theta_1} \\ C_1\left(\frac{\theta_2}{\theta_1}\right)^{n-k} t^{-n+k-1} - C_1\alpha^{2n-k}\left(\frac{\theta_1}{\theta_2}\right)^n t^{n-1} + \alpha^{2n-k}\frac{n}{2}\left(\frac{\theta_1}{\theta_2}\right)^n t^{n-1} & \frac{\theta_2}{\alpha\theta_1} \le t < \frac{\theta_2}{\alpha\theta_1} \\ \frac{n}{2}\alpha^{-k}\left(\frac{\theta_2}{\theta_1}\right)^n t^{-n-1} & \frac{\theta_2}{\alpha\theta_1} \le t < \infty \end{cases}$$

(2.3)

Where, $C_1 = \frac{n(n-k)}{2n-k}$

$$E(T^r) = \left(\frac{\theta_2}{\theta_1}\right)^r \left[\alpha_k(r)n_0(r)n_0(-r) + \bar{\alpha}_k(r)n_k(r)n_0(-r)\right]$$

(2.4)

From (1.5) and (1.7), one can see easily that $X_{(n)}$and $Y_{(n)}$ are sufficient and complete for known α .Hence , by using the Lehmann-Sheffe's theorem , we can get U.M.V.U.E of R in the presence of outliers if α is known . U.M.V.U.E of R is

$$\hat{R}_U = \left(\bar{b} + \frac{b}{\alpha}\right)\left(\frac{\theta_2^*}{2\theta_1^*}\right) \qquad (2.5)$$

$$= \left(\bar{b} + \frac{b}{\alpha}\right)\left(\frac{Y_{(n)}}{2X_{(n)}}\right)\left(\frac{n+1}{e_1 n}\right) \qquad (2.6)$$

is U.M.V.U.E of R.

Hence ,

$E(\hat{R}_U) = \left(\bar{b} + \frac{b}{\alpha}\right)(E(T))\left(\frac{n+1}{2e_1 n}\right)$ So,

$$E(T) = \left(\frac{\theta_2}{\theta_1}\right)\left[\alpha_k(1)n_0(1)n_0(-1) + \bar{\alpha}_k(1)n_k(1)n_0(-1)\right]$$

$$= \left(\frac{\theta_2}{\theta_1}\right)\left[\frac{e_1}{n_0(-1)}\right]$$

$$= \left(\frac{\theta_2}{\theta_1}\right)\left[\frac{ne_1}{n+1}\right] \qquad (2.7)$$

$$E(\hat{R}_U) = \left(\bar{b} + \frac{b}{\alpha}\right)\left(\frac{n+1}{2e_1 n}\right)\left(\frac{\theta_2}{\theta_1}\right)\left[\frac{ne_1}{n+1}\right]$$

$$= \left(\bar{b} + \frac{b}{\alpha}\right)\left(\frac{\theta_2}{2\theta_1}\right) = R$$

Thus $\hat{R}_U$ is U.M.V.U.E of R

Theorem 3: $\hat{R}_U$ is consistent estimator of R .

Proof:

$$V(\hat{R}_U) = E(\hat{R}_U - R)^2$$

$$= \left(\frac{(\bar{b} + \frac{b}{\alpha})^2}{4}\right)\left[\left(\frac{n+1}{e_1 n}\right)^2 E(T^2) - 2\left(\frac{n+1}{e_1 n}\right)E(T)\left(\frac{\theta_2}{\theta_1}\right) + \left(\frac{\theta_2}{\theta_1}\right)^2\right]$$

$$= \left(\frac{(\bar{b} + \frac{b}{\alpha})^2}{4}\right)\left(\frac{\theta_2}{\theta_1}\right)^2 \left[\left(\frac{n+1}{e_1 n}\right)^2 [\alpha_k(2)n_0(2)n_0(-2) + \bar{\alpha}_k(2)(n_k(2))(n_0(-2))]\right]$$

$$+ \left(\frac{(\bar{b} + \frac{b}{\alpha})^2}{4}\right)\left(\frac{\theta_2}{\theta_1}\right)^2 \left[[\alpha_k(1)n_0(1)n_0(-1) + \bar{\alpha}_k(1)(n_k(1))(n_0(-1))](-2\frac{n+1}{e_1 n}) + 1\right]$$

(2.8)

As n $\rightarrow\infty$; k $\rightarrow\infty$, b$\rightarrow\gamma$ and $\bar{b} \rightarrow \bar{\gamma}$ =1- γ Where ,0 <γ< 1
Therefore,

$$\lim_{k\to\infty\ n\to\infty} \alpha_k(r)n_0(r)n_0(-r) + \bar{\alpha}_k(r)(n_k(r))(n_0(-r)) = 1$$

Similarly , as As n $\rightarrow\infty$; k $\rightarrow\infty$ $n_0(-1)e_1 \rightarrow 1$
And $n_k(r) \rightarrow 1$

Hence, As n $\rightarrow\infty$; k $\rightarrow\infty$ V($\hat{R}_U$)$\rightarrow$0 $\hat{R}_U$ is consistent estimator of R

Theorem 3: $\hat{R}_M$ is consistent estimator of R
Proof:

$$E(\hat{R}_M) = \left(\bar{b} + \frac{b}{\alpha}\right) E\left(\frac{Y_{(n)}}{2X_{(n)}}\right)$$

$$= \left(\bar{b} + \frac{b}{\alpha}\right) E\left(\frac{T}{2}\right)$$

$$= \left(\frac{\theta_2}{2\theta_1}\right)(\bar{b} + \frac{b}{\alpha})[\alpha_k(1)n_0(1)n_0(-1) + \bar{\alpha}_k(1)(n_k(1)n_0(-1))]$$

(2.9)
As n $\rightarrow\infty$; k $\rightarrow\infty$

$$[\alpha_k(1)n_0(1)n_0(-1) + \bar{\alpha}_k(1)(n_k(1)n_0(-1))]$$

Tends to 1 Similarly , as
As n $\rightarrow\infty$; k $\rightarrow\infty$ $n_0(-1)e_1 \rightarrow 1$ and $n_k(r) \rightarrow 1$

E ($\hat{R}_M$)$\rightarrow$ R asn $\rightarrow\infty$; k $\rightarrow\infty$

M.S.E ($\hat{R}_M$) = E($\hat{R}_M$ -R)^2

=E($\hat{R}_M$)^2-2 R E($\hat{R}_M$) + R^2

$$= \left[\frac{(\bar{b} + \frac{b}{\alpha})^2}{4}\right]\left[E(T^2) - 2E(T)\left(\frac{\theta_2}{\theta_1}\right) + \left(\frac{\theta_2}{\theta_1}\right)^2\right]$$

$$= \left[\frac{(\bar{b} + \frac{b}{\alpha})^2}{4}\right]\left(\frac{\theta_2}{\theta_1}\right)^2 [[\alpha_k(2)n_0(2)n_0(-2) + \bar{\alpha}_k(2)(n_k(2))(n_0(-2)) + 1]$$

$$+ \left[\frac{(\bar{b} + \frac{b}{\alpha})^2}{4}\right]\left(\frac{\theta_2}{\theta_1}\right)^2 - 2\,[[\alpha_k(1)n_0(1)n_0(-1) + \bar{\alpha}_k(1)(n_k(1))(n_0(-1))]]$$

As

$$\lim_{k\to\infty\ n\to\infty} \alpha_k(r)n_0(r)n_0(-r) + \bar{\alpha}_k(r)(n_k(r))(n_0(-r)) = 1$$

Hence,M.S.E ($\hat{R}_M$)$\rightarrow$0 asn $\rightarrow\infty$; k $\rightarrow\infty$

Hence $\hat{R}_M$ is consistent estimator of R.

Hence both $\hat{R}_M$ and $\hat{R}_U$ are consistent estimators.Further ,

Corollary: $\hat{R}_U$ is more efficient than $\hat{R}_M$.

Proof: $M.S.E$($\hat{R}_M$)= E($\hat{R}_M$ -R)^2

=E($\hat{R}_M$ -R + $\hat{R}_U$ - $\hat{R}_U$)^2

=$E(\hat{R}_U - R)^2 + E(\hat{R}_M - \hat{R}_U)^2$

=$V(\hat{R}_U) + E(\hat{R}_M - \hat{R}_U)^2$

ThusM.S.E($\hat{R}_M$) ≥V($\hat{R}_U$).

Therefore U.M.V.U.EofRi.e., $\hat{R}_U$ is more efficient than M.L.EofRi.e., $\hat{R}_M$.

Conclusion: Here it is concluded that ifα is known then M.S.E of $\hat{R}_U$ is less than M.S.E $\hat{R}_M$ .Hence U.M.V.U.E of R, $\hat{R}_U$ should be selected .

References:

1. U. J. Dixit, M.Masoom Ali and Jungsoo Woo, “Efficient estimation of parameters of a uniform distribution in the presence of outliers”. Soochow journal of Mathematics volume 29,No. 4, pp363-369,2003.
2. 2 G. Ooms and K. L. Moore, “ A model assay for genetic and environmental changes in the architecture of intact roots systems of plants grown in vitro.” Plantcell , Tissue and Organ Culture ,27,pp 129-139,1991.
3. 3 U.J. Dixit,andK.D.Phal “ Estimation of P[X ≤Y] for the uniform distribution in the presence of outliers”.Journal of Probability and Statistical Sciences volume 7,No. 1,pp11-18,2009.

***

Associate Professor,Statistics
B.N. Bandodkar College,Thane.
Email: kdflgd@gmail.com