Youngnam Statistical Letters
June 1993, Vol. 4, 131-136

The Correlation Coefficient between the Smallest and Largest Observations in the Weibull Model in the Presence of an Unidentified Outlier

Jungsoo Woo and Changsoo Lee [1]

ABSTRACT

We shall consider the trends of correlation coefficient between the smallest and largest observations in the Weibull model in the presence of an unidentified outlier, and derive the density functions of order statistics by the permanent theory.

1. Introduction

Consider a life testing experiment consisting of n observations, (n-1) of which have the same expected life time, while one of them could have a different expected life time.

The correlation coefficient between the smallest and largest observations in the probability model could be used as a measure to direct if an arbitrary unidentified outlier is present in a simple random sample. A small value of the correlation coefficient between the samllest and largest observations would indicated the presence of an unidentified outlier, whereas a large value of the correlation coefficient would indicated the absence of an arbitrary unidentified outlier.

[1] Department of Statistics, Yeungnam University, Kyungsan 712-749, Korea.

Gross, Hunt and Odeh(1986) have studied the correlation coefficient between the smallest and largest observations when all but one of n observations are exponentially distributed with a mean life time θ and remaining obsrevation has a mean life time θ_1.

Here we shall consider the trends of the correlation coefficient between the smallest and largest observations in the Weibull model in the presence of an unidentified outlier.

The exact formulas of density function and joint density function of the order statistics in the Weibull model in the presence of an unidentified outlier will be derived by the method of permanent theory in Bapat and Beg(1989), and we shall evaluate the correlation coefficient between the smallest and largest observations in the Weibull model in the presence of an unidentified outlier, this will be an extension of results in Gross, Hunt and Odeh(1986).

2. The Correlation Coefficient

Consider a life testing experiment consisting of n observations, the life time of the items is often discribed as the Weibull random variable with the p.d.f.

$$f(x;\beta,\alpha) = \frac{\alpha}{\beta^\alpha}\cdot x^{\alpha-1}\cdot \exp\{-(\frac{x}{\beta})^\alpha\}, \qquad x>0,\ \ \alpha,\ \beta>0,$$

where α and β are referred as the shape and scale parameters, respectively, denoted by X $\sim$ WEI(β,α).

Suppose $X_1, X_2, \cdots, X_n$ are independent observations where all but one of them are from WEI(β,α), but one remaining observation is from WEI(β_1,α). Define $v \equiv \beta_1/\beta$. Before the start of the experiment we have no prior knowledge as to which one of these n is the outlier. Let $Y_1 \leq Y_2 \leq \cdots \leq Y_n$ be the order statistics of the sample. Then the marginal p.d.f's and joint p.d.f. of Y_1 and Y_n can be obtained by the parmanent theory (cf. Vaughn and Venables(1972)).

Vaughn and Venables have shown that the joint p.d.f of $Y_1, Y_2, \cdots Y_n$ at the point $(y_1, y_2, \cdots y_n)$ may be written as the permanent of the matrix of the marginal p.d.f's as follows;

$$f_{1,2,\cdots,n}(y_1, y_2, \cdots y_n) = \text{per}\left| F \right|,$$

where $0 \le y_1 \le y_2 \le \cdots \le y_n$ and F is the $n \times n$ matrix

$$F = \begin{pmatrix} f_1(y_1) & f_2(y_1) & \dots & f_n(y_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(y_n) & f_2(y_n) & \dots & f_n(y_n) \end{pmatrix}$$

The permanent of an $n \times n$ square matrix A, denoted as per| A |, has the same definition as its determinant except that all signs are positive. Thus per| A | is the sum of n!-terms. Each term in this sum is formed as the product of n terms obtained by choosing an element from each row and column.

Let f_o and F_o be the p.d.f. and c.d.f. of an unidentified outlier random variable, respectively.

Then the p.d.f.'s of Y_1 and Y_n are given by

$$\begin{aligned} f_{Y_1}(y_1) &= \frac{1}{(n-1)!} \cdot \text{per} \begin{vmatrix} f(y_1) & 1-F(y_1) & \dots & 1-F(y_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_o(y_1) & 1-F_o(y_1) & \dots & 1-F_o(y_1) \end{vmatrix} \\ &= \frac{\alpha(n-1+v^{-\alpha})}{\beta^\alpha} \cdot y_1^{\alpha-1} \exp\{-(n-1+v^{-\alpha})(\frac{y_1}{\beta})^\alpha\}, \quad 0 \le y_1 \end{aligned}$$

and

$$\begin{aligned} f_{Y_n}(y_n) &= \frac{1}{(n-1)!} \text{per} \begin{vmatrix} F(y_n) & \dots & F(y_n) & f(y_n) \\ \vdots & \ddots & \vdots & \vdots \\ F_o(y_n) & \dots & F_o(y_n) & f_o(y_n) \end{vmatrix} \\ &= \sum_{k=0}^{n-1} \frac{(-1)^k \cdot \alpha \cdot \binom{n-1}{k}}{v^\alpha \cdot \beta^\alpha} \cdot y_n^{\alpha-1} \exp\{-(k+v^{-\alpha})(\frac{y_n}{\beta})^\alpha\} \\ &\quad + \sum_{k=0}^{n-2} \frac{(-1)^k \alpha(n-1)\binom{n-2}{k}}{\beta^\alpha} \cdot y_n^{\alpha-1} \Big\{ \exp\{-(k+1)(\frac{y_n}{\beta})^\alpha\} \\ &\quad - \exp\{-(k+1+v^{-\alpha})(\frac{y_n}{\beta})^\alpha\} \Big\}, \qquad 0 \le y_n. \end{aligned}$$

Therefore, the m-th moments of Y_1 and Y_n are given by

$$E(Y_1^m) = \{\frac{\beta}{(n-1+v^{-\alpha})^{\frac{1}{\alpha}}}\}\Gamma(1+\frac{m}{\alpha}) \tag{2.1}$$

and

$$\begin{aligned} E(Y_n^m) =& \sum_{k=0}^{n-1} \frac{(-1)^k \beta^m \Gamma(1+\frac{m}{\alpha})\binom{n-1}{k}}{v^\alpha (k+v^\alpha)^{1+\frac{m}{\alpha}}} \\ &+\sum_{k=0}^{n-2}(-1)^k(n-1)\beta^m\Gamma(1+\frac{m}{\alpha})\binom{n-2}{k} \\ &\times \left\{(k+1)^{-(1+\frac{m}{\alpha})} - (k+1+v^{-\alpha})^{-(1+\frac{m}{\alpha})}\right\}, \end{aligned} \tag{2.2}$$

respectively.

In the same method of permanents, the joint p.d.f. of Y_1 and Y_n is given by

$$\begin{aligned} g_{Y_1,Y_n}(y_1,y_n) =& \frac{1}{(n-2)}\times \\ & \begin{vmatrix} f(y_1) & F(y_n)-F(y_1) & \dots & F(y_n)-F(y_1) & f(y_n) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ f_o(y_1) & F_o(y_n)-F_o(y_1) & \dots & F_o(y_n)-F_o(y_1) & f_o(y_n) \end{vmatrix} \\ =& \sum_{k=0}^{n-2} \frac{(-1)^k\alpha^2(n-1)\binom{n-2}{k}}{v^\alpha \cdot \beta^{2\alpha}} \cdot y_1^{\alpha-1}\cdot y_n^{\alpha-1} \\ & \Big\{\exp\{-(n-1-k)(\frac{y_1}{\beta})^\alpha - (k+v^{-\alpha})\frac{(y_n}{\beta)}^{\alpha}\} \\ & +\exp\{-(n-2-k+v^\alpha)(\frac{y_1}{\beta})^\alpha - (k+1)(\frac{y_n}{\beta})^\alpha\}\Big\} \\ & +\sum_{k=0}^{n-3} \frac{(-1)^k\alpha^2(n-1)(n-2)\binom{n-3}{k}}{\beta^{2\alpha}} \cdot y_1^{\alpha-1}\cdot y_n^{\alpha-1} \\ & \Big\{\exp\{-(n-2-k+v^{-\alpha})(\frac{y_1}{\beta})^\alpha - (k+1)(\frac{y_n}{\beta})^\alpha\} \\ & -\exp\{-(n-2-k)(\frac{y_1}{\beta})^\alpha - (k+1+v^{-\alpha})(\frac{y_n}{\beta})^\alpha\}\Big\} \end{aligned}$$

Therefore, the expectation of $Y_1 \cdot Y_n$ is given by

$$E(Y_1 \cdot Y_n) = \sum_{k=0}^{n-2} \frac{(-1)^k (n-1)\binom{n-2}{k}}{v^{\alpha}\beta^{2\alpha}}$$
$$\times \left\{ \frac{\Gamma^2(1+\frac{1}{\alpha})}{[(n-1-k)\beta^{-2\alpha}(k+v^{-\alpha})]^{1+\frac{1}{\alpha}}} IB_{(n-1-k)/c}(1+\frac{1}{\alpha}, 1+\frac{1}{\alpha}) \right.$$
$$\left. + \frac{\Gamma^2(1+\frac{1}{\alpha})}{[(n-2-k+v^{-\alpha})\beta^{-2\alpha}(k+1)]^{1+\frac{1}{\alpha}}} IB_{(n-2-k+v^{-\alpha})/c}(1+\frac{1}{\alpha}, 1+\frac{1}{\alpha}) \right\}$$
$$+ \sum_{k=0}^{n-3} \frac{(-1)^k (n-1)(n-2)\binom{n-3}{k}}{\beta^{2\alpha}}$$
$$\times \left\{ \frac{\Gamma^2(1+\frac{1}{\alpha})}{[(k+1)(n-2-k+v^{-\alpha})\beta^{-2\alpha}]^{1+\frac{1}{\alpha}}} IB_{(n-1-k+v^{-\alpha})/c}(1+\frac{1}{\alpha}, 1+\frac{1}{\alpha}) \right.$$
$$\left. - \frac{\Gamma^2(1+\frac{1}{\alpha})}{[(n-2-k)(k+1+v^{-\alpha})\beta^{-2\alpha}]^{1+\frac{1}{\alpha}}} IB_{(n-2-k)/c}(1+\frac{1}{\alpha}, 1+\frac{1}{\alpha}) \right\} \quad (2.3)$$

where $IB_a(d_1, d_2)$ is the Karl Pearson's incomplete beta function, $\Gamma(\cdot)$ is the gamma function and $c = n - 1 + v^{-\alpha}$.

Hence, by the results (2.1), (2.2), and (2.3), we can evaluate numerical correlation coefficient between the smallest and largest observations in the Weibull model in the presence of an unidentified outlier. The numerical values of the correlation coefficient between the smallest and largest observations in the Weibull model in the presence of an unidentified outlier are given in Table for v= 1/5, 1/2, 1, 2, 5 and the shape parameter α = 2 and 1/2. When the shape parameter is one, our result of correlation coefficient is the same as that of Odeh, Gross and Hunt(1986).

From the table, correlation coefficient between the smallest and largest observations in the Weibull model in the presence of an unidentified outlier is locally maximum at a neigborhood of $v = \beta/\beta_1 = 1$.

References

1. Balakrishnan, N. and Cohen, A.C.(1991), Order Statistics and Inference, AP, NY.

2. Bapat, R.B and Beg, M.I.(1989), Order Statistics for nonidentically Distributed Variables and Permanents, Sankhya, 51, 79-93

3. Gradshteyn, I.G. and Ryzhik, I.N.(1980), Table of Integrals, Series, and Products, AP, NY.

4. 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 I.I.D. Exponential Distribution, Commun. Statist.-Theor. Meth., 15(4), 1113-1124.

5. Lieblein, J.(1965), On Moments of Order Statistics from the Weibull Distribution, Ann. Math. Stat., 26, 330-333

6. Vaughn, R.J. and Venables, W.N.(1972), Permanent Expressions for Order Statistics Density, Journal of the Royal Statistical Society, Series B, 34, 308-310.