



The Correlation between Variate-Values and Ranks in Samples from Complete Fourth Power Exponential Distribution

Tahani A. Maturi (Corresponding author)

Department of Statistics, Faculty of Science

King Abdulaziz University, Jeddah, Saudi Arabia

E-mail: tahanistat@hotmail.com

Anga Elsayigh

Department of Statistics, Faculty of Science

King Abdulaziz University, Jeddah, Saudi Arabia

E-mail: angasoft@hotmail.com

Abstract

In this paper, we derive the correlation between variate-values and ranks in a sample from the Complete Fourth Power Exponential (CFPE) distribution. A sample from the CFPE distribution could be misclassified as if it is drawn from the normal distribution due to some similarities between the two distributions. In practice, ranks are used instead of real values (variate-values) when there is hardly any knowledge about the underlying distribution. This may lead to loss of some of the information contained in the actual values. In this paper we found that the amount of information loss, by using ranks instead of real data, is larger when the sample is from the CFPE distribution than if it is from the normal distribution. However, there is still a relatively high correlation between variate-values and the corresponding ranks. Comparisons between the correlation between variate-values and ranks for the CFPE distribution and some other distributions are provided.

Keywords: Complete Fourth Power Exponential Distribution, Variate-values, Normal distribution, Correlation coefficient, Ranks

1. Introduction

Statistical methods based on ranks have been heavily studied in the literature, specifically when there is hardly any knowledge of the underlying distribution of the data at hand. Such methods fall under the nonparametric techniques umbrella. However, ignoring the underlying distribution may lead to loss of some of the information contained in the data. And, in some circumstances, when there is a lack of information with regards to the underlying distribution, nonparametric techniques – including the methods based on ranks – could be useful and lead to robust inferences. For more details on these methods, we refer the reader to Lehmann and D'Abrera (1976).

It can be difficult to discriminate between the Complete Fourth Power Exponential (CFPE) distribution and the normal distribution due to some similarities between these distributions, e.g. their shapes and other properties. Some people, especially non-statisticians, may mistakenly assume their dataset comes from a normal distribution, when in fact the data may actually come from the CFPE distribution, since it has properties similar to a normal distribution.

Amira and Mazloun (1993) studied in detail the CFPE distribution and focused on the geometric and statistical properties of this distribution and compared it to the normal distribution. Unfortunately, this distribution

has limited application at the moment with possible future works. We briefly give an overview of the CFPE distribution and some of its properties in the next section.

The misclassification of the dataset from the CFPE distribution as a normal distribution was studied in an unpublished MSc thesis entitled “Study on the Error in Random Samples Classification between the Normal Distribution and the Complete Fourth Power Exponential” by Baeshen (2000), King Abdulaziz University, Jeddah, Saudi Arabia.

In practice, the ranks of data are used instead of the original data to make an inference. Additionally, to find how much information we may lose by this action, Stuart (1954) derived the formula to calculate the relationship (correlation) between variate-values and their ranks. It showed that for some distributions we did not lose much information when the original data is replaced by their corresponding ranks. Moreover, Stuart (1955) considered the situation when the variance of a specific distribution does not exist and it showed that for a continuous distribution with no moments which is eventually monotone, the correlation between variate-values and their ranks is zero.

Later, O’Brien (1982) estimated via simulation the average correlation between variate-values and their ranks for small size samples from different distributions. The term “the degree of distortion or error” is used to indicate the loss of information by replacing the variate-values by their ranks. O’Brien found that these correlations are generally high, and when the sample size increase they reach the limiting values presented by Stuart (1954, 1955). This finding supports the idea of using the ranks instead of variate-values with a small degree of distortion or error even for small samples.

In Section 2, we will overview the CFPE distribution and some of its properties. In Section 3, we use Stuart’s formula to derive the correlation between variate-values and ranks in samples from the CFPE Distribution. Finally, we provide the exact and the approximation correlations between variate-values and ranks from some distributions in Section 4.

2. The Complete Fourth Power Exponential (CFPE) Distribution

In this section we overview briefly the CFPE distribution and some of its summaries that are needed in this paper. For more details we refer to Amira and Mazloum (1993).

Let X be a random variable from the CFPE distribution with the density function given by

$$f(x; \alpha, \beta) = \frac{2}{\beta \Gamma(\frac{1}{4})} \exp\{-(x - \alpha)^4 / \beta^4\} \quad , -\infty < x < \infty, -\infty < \alpha < \infty, \beta > 0 \quad (1)$$

where α is the location parameter and β is the scale parameter. The characteristic function of this distribution is

$$\phi_X(t) = \sum_{m=0}^{\infty} \sum_{j=0}^{\lfloor \frac{m}{2} \rfloor} C_{2j}^m \alpha^{(m-2j)} \beta^{2j} \frac{\Gamma(\frac{2j+1}{4})}{\Gamma(\frac{1}{4})} \frac{(it)^m}{m!}$$

where $\lfloor m/2 \rfloor$ is the greatest integer number less than $m/2$. Then, from this characteristic function, we can obtain the central and non-central moments as follows:

$$\begin{aligned} \mu'_1 = \mu = E(X) &= \alpha, \quad \mu'_2 = E(X^2) = \alpha^2 + \beta^2 \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}, \quad \mu'_3 = E(X^3) = \alpha^3 + 3\beta^2 \alpha \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})}, \\ \mu'_4 = E(X^4) &= \alpha^4 + 6\beta^2 \alpha^2 \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})} + \frac{1}{4}\beta^4, \quad \mu_2 = \sigma^2 = \beta^2 \frac{\sqrt{2}\pi}{[\Gamma(\frac{1}{4})]^2}, \quad \mu_3 = 0 \quad \text{and} \quad \mu_4 = \frac{1}{4}\beta^4 \end{aligned}$$

For simplicity and easy notation, we work on the special case of the CFPE distribution when $\alpha = 0$, since it is easy to centralize the data around the mean. Therefore

$$f(x; \beta) = \frac{2}{\beta \Gamma(\frac{1}{4})} \exp\{-(x/\beta)^4\} \quad , -\infty < x < \infty, \beta > 0$$

and $\mu = \alpha = 0$, $V(X) = \sigma^2 = \beta^2 \frac{\sqrt{2}\pi}{[\Gamma(\frac{1}{4})]^2}$.

3. The correlation between variate-values and ranks in samples from the CFPE Distribution

In this section, we use Stuart's formula (1954) to derive the correlation between variate-values and the ranks in sample from the CFPE distribution. From Stuart (1954), the correlation between variate-values X_i and their ranks R_i , $\rho(X_i, R_i)$, is given by

$$\rho(X_i, R_i) = \left[\int_{-\infty}^{\infty} xF(x)dF(x) - \frac{1}{2}\mu \right] \left[\frac{12(n-1)}{(n+1)\sigma^2} \right]^{\frac{1}{2}} \quad (2)$$

where n is the sample size. The second factor on the right side in (2) tends to $2\sqrt{3}/\sigma$ as n tends to infinity. In the following theorem, we use (2) to derive the correlation between variate-values and ranks in a sample from the CFPE distribution.

Theorem:

The correlation between variate-values and ranks in a sample from the CFPE distribution when $\alpha = 0$ is given by

$$\rho = \rho(X_i, R_i) = \frac{\sqrt{3}}{2\sqrt[4]{2}} \sqrt{\frac{n-1}{n+1}} \quad \text{and as } n \rightarrow \infty, \quad \rho \rightarrow \frac{\sqrt{3}}{2\sqrt[4]{2}} = 0.728 \quad (3)$$

Proof:

For ease notation, we consider the centered CFPE distribution, i.e. when $\alpha = 0$, therefore

$$f(x; \beta) = \frac{2}{\beta \Gamma(\frac{1}{4})} \exp\{-(x/\beta)^4\} \quad , -\infty < x < \infty, \beta > 0$$

and $\mu = \alpha = 0$, $V(X) = \sigma^2 = \beta^2 \frac{\sqrt{2}\pi}{[\Gamma(\frac{1}{4})]^2}$ then

$$E[XF(X)] = \int_{-\infty}^{\infty} \{x F(x)\} f(x) dx = \frac{4}{\beta^2 [\Gamma(\frac{1}{4})]^2} \int_{-\infty}^{\infty} \int_{-\infty}^x x e^{-\frac{1}{\beta^4}(t^4+x^4)} dt dx$$

Let $\frac{x^2}{\beta^2} = r \cos \theta \Rightarrow x = \beta r^{\frac{1}{2}} \cos^{\frac{1}{2}} \theta$ and $\frac{t^2}{\beta^2} = r \sin \theta \Rightarrow t = \beta r^{\frac{1}{2}} \sin^{\frac{1}{2}} \theta$.

The Jacobian of the transformation is

$$\begin{aligned} J &= \begin{vmatrix} \frac{1}{2}\beta r^{-\frac{1}{2}} \cos^{\frac{1}{2}} \theta & -\frac{1}{2}\beta r^{\frac{1}{2}} \cos^{-\frac{1}{2}} \theta \sin \theta \\ \frac{1}{2}\beta r^{-\frac{1}{2}} \sin^{\frac{1}{2}} \theta & \frac{1}{2}\beta r^{\frac{1}{2}} \sin^{-\frac{1}{2}} \theta \cos \theta \end{vmatrix} = \frac{1}{4}\beta^2 \left\{ \sin^{-\frac{1}{2}} \theta \cos^{\frac{3}{2}} \theta + \cos^{-\frac{1}{2}} \theta \sin^{\frac{3}{2}} \theta \right\} \\ &= \frac{1}{4}\beta^2 \sin^{-\frac{1}{2}} \theta \cos^{-\frac{1}{2}} \theta \end{aligned}$$

then

$$\begin{aligned} E[XF(X)] &= \frac{8}{\beta^2 [\Gamma(\frac{1}{4})]^2} \int_0^{\infty} \int_0^{\frac{\pi}{4}} (\beta r^{\frac{1}{2}} \cos^{\frac{1}{2}} \theta) e^{-(r^2 \cos^2 \theta + r^2 \sin^2 \theta)} \left(\frac{1}{4}\beta^2 \sin^{-\frac{1}{2}} \theta \cos^{-\frac{1}{2}} \theta \right) dr d\theta \\ &= \frac{2\beta}{[\Gamma(\frac{1}{4})]^2} \int_0^{\infty} \int_0^{\frac{\pi}{4}} r^{\frac{1}{2}} e^{-r^2} \sin^{-\frac{1}{2}} \theta dr d\theta \end{aligned}$$

Let $z = r^2 \Rightarrow \frac{1}{2}z^{-\frac{1}{2}}dz = dr$, then

$$\begin{aligned} E[XF(X)] &= \frac{2\beta}{[\Gamma(\frac{1}{4})]^2} \left(\frac{1}{2} \int_0^{\infty} z^{-\frac{1}{4}} e^{-z} dz \right) \left(\int_0^{\frac{\pi}{4}} \sin^{-\frac{1}{2}} \theta d\theta \right) \\ &= \frac{2\beta}{[\Gamma(\frac{1}{4})]^2} \left\{ \frac{1}{2} \Gamma(\frac{3}{4}) \right\} \left\{ \frac{1}{4} B(\frac{1}{4}, \frac{1}{2}) \right\} = \frac{\beta \Gamma(\frac{1}{2})}{4\Gamma(\frac{1}{4})} \end{aligned}$$

where $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Now by using (2) we have

$$\rho(X_i, R_i) = \left[\frac{12(n-1)[\Gamma(\frac{1}{4})]^2}{(n+1)\beta^2 \sqrt{2} \pi} \right]^{\frac{1}{2}} \frac{\beta \Gamma(\frac{1}{2})}{4\Gamma(\frac{1}{4})} = \frac{\sqrt{3}}{2 \sqrt[4]{2}} \sqrt{\frac{n-1}{n+1}}$$

and as $n \rightarrow \infty$, $\rho \rightarrow \frac{\sqrt{3}}{2 \sqrt[4]{2}} = 0.728$.

4. Comparisons results

Stuart (1954) provided, using the formula in (2), the correlation between variate-values and the ranks in samples from Uniform (0, 1), Exponential ($\lambda = 1$), and Normal (0, 1) distributions. From (3), we have the correlation between variate-values and the ranks in the sample from the CFPE distribution (0, b), which is equal to 0.728 when n is significantly large, see Table 1.

Moreover, by using the simulation technique to estimate the quantity $E[XF(X)]$ in (2), we obtained the approximation results for the correlation between variate-values and the ranks in samples from the other distributions which did not have explicit formula. These approximation results are given in Table 2. We can point out that the corresponding correlation for the Log-normal distribution (0, 1) is closer to (less than) the CFPE distribution compared to other distributions.

5. Conclusion

In this paper, we derived the correlation between variate-values and ranks in a sample from the Complete Fourth Power Exponential (CFPE) distribution. We found that we lost more information when the data set came from the CFPE distribution than when the data came from normal distribution, when we used ranks instead of variate-values. This finding emphasizes the importance of distinguishing between these distributions. However, the correlation between variate-values and ranks in samples from the CFPE is still relatively high, which allowed us to use ranks instead of variate-values without losing a lot of information.

Acknowledgment

The authors are grateful to the referees for useful comments on the earlier version of this paper.

References

- Amira, T. and Mazloun, R. (1993). Complete Fourth Power Exponential Distribution Geometric and statistical study. *Econometric and Business Review*, Ain Shams University.
- Lehmann, E. L. and D'Abrera, H. J. M. (1976). *Nonparametric : Statistical Methods Based on Ranks*. San Francisco, Holden-Day ; New York ; London : McGraw-Hill.
- O'Brien R.M. (1982). The correlation between variate-values and ranks in small size samples. *British Journal of Mathematical and Statistical Psychology*, 35, 106-112.
- Rotar, V. (1997). *Probability Theory*. World Scientific .
- Stuart, A. (1954). The correlation between variate-values and ranks in samples from a continuous distribution. *British Journal of Statistical Psychology*, 7, 37-44.
- Stuart, A. (1955). The correlation between variate-values and ranks in samples from distributions having no variance. *British Journal of Statistical Psychology*, 8, 25-27.

Table 1. The Exact correlation between variate-values and ranks for some distributions

Distribution	μ	σ^2	$E [X F(x)]$	ρ	$n \rightarrow \infty$
Uniform (0,1)	1/2	1/12	1/3	$\sqrt{\frac{n-1}{n+1}}$	1
Exponential $\lambda = 1$	1	1	3/4	$\frac{\sqrt{3}}{2} \sqrt{\frac{n-1}{n+1}}$	$\frac{\sqrt{3}}{2} \approx .866$
Normal (0,1)	0	1	$\frac{1}{2\sqrt{\pi}}$	$\sqrt{\frac{3}{\pi}} \sqrt{\frac{n-1}{n+1}}$	$\sqrt{\frac{3}{\pi}} \approx .977$
Logistic (0,1)	0	$\frac{\pi^2}{3}$	1/2	$\frac{3}{\pi} \sqrt{\frac{n-1}{n+1}}$	$\frac{3}{\pi} \approx .955$
CFPE (0,b)	0	$\beta^2 \frac{\sqrt{2}\pi}{[\Gamma(\frac{1}{4})]^2}$	$\frac{\beta\Gamma(\frac{1}{2})}{4\Gamma(\frac{1}{4})}$	$\frac{\sqrt{3}}{2} \sqrt{\frac{n-1}{n+1}}$	$\frac{\sqrt{3}}{2} \approx .728$

The exact results of Uniform, Exponential and Normal are reported by Stuart (1954) while the logistic result presented in O'Brien (1982)

Table 2. The approximation correlation between variate-values and ranks for some distributions

Distribution	μ	σ^2	$E [X F(x)]$	ρ
Student's T (5 df)	0	1.667	0.346	0.928
Chi-square (5 df)	5	10	3.347	0.928
Gamma (5, 1)	5	5	3.117	0.956
F (4, 10)	1.250	1.563	0.909	0.787
Beta (1, 2)	0.333	0.056	0.233	0.973
Lognormal (0, 1)	1.649	4.671	1.253	0.687
Weibull (5, 1)	0.918	0.044	0.518	0.974

The approximation results obtained from 100 000 simulation samples of size 100 ($n = 100$)