On a Newly Developed Estimator for More Accurate Modeling with an Application to Civil Engineering

Saleha Naghmi Habibullah
Professor, Department of Statistics, Kinnaird College For Women, Lahore, Pakistan

Syeda Shan-E-Fatima
Lecturer, Department of Statistics, Government College University, Lahore, Pakistan

ABSTRACT: Maintenance of the construction equipment fleet being an indispensably important concern in megaprojects such as construction of bridges and dams, equipment reliability metrics such as failure-rates, availability of equipment, time between failures and time required for repair are of paramount interest for contractors and project managers. The availability of data on variables such as time to failure, repair-time and the like motivates the determination of appropriate probability models that fit the data with a high degree of accuracy and facilitate estimation of probabilities that may be valuable in project-planning. Estimation of parameters of the proposed model by efficient estimation procedures is one of the first steps in achieving a model that 'best' fits the data. Only very recently, a property of a particular class of continuous probability distributions that has been named 'self-inversion at unity' has begun to be utilized for obtaining modifications to well-known estimators so that the modified estimators are *more efficient* than their well-known counterparts. In this paper, we focus on the *more general case* that we call 'self-inversion at A', where A can be any arbitrary real number, and propose a modification to the formula of the sample mean on the basis of this property. By applying the newly proposed modified mean to a data-set pertaining to repair-times of construction equipment, we demonstrate the *usefulness* of this approach in achieving probability models that are likely to fit, with *a higher degree of accuracy* than that which is achievable through the utilization of the well-known estimators, reliability and maintenance-related data encountered in megaprojects undertaken by civil engineers as well as in a variety of other engineering endeavors.

For obvious reasons, maintenance of the construction equipment fleet forms an integral part of the overall project-strategy in megaprojects such as construction of sky-scrapers, bridges and dams. A variety of reliability metrics such as failure-rates, availability, time between failures and time required for repair enable contractors and project managers to track the performance of the equipment fleet during the course of a high-profile construction project.

Reputed construction companies all over the world routinely keep records of variables such as time to failure, time between failure and time to repair pertaining to various types of equipment, and a number of them utilize the data generated by the record-keeping process for the determination of appropriate probability models that fit the data with a high degree of accuracy. Such determinations are important as they facilitate the computation of probabilities that guide the contractors and project-managers in taking important decisions at various points in time during the course of a construction project such as the choice of the optimal course of action out of a number of alternatives.

Estimation of parameters of a proposed probability model is one of the first steps in achieving a model that is likely to fit a real-life data-set. More often than not, parameter-estimation is achieved by computing statistics that are 'sample-counterparts' of the population parameters to be estimated. For example, in order to determine the mean time to failure (MTTF) of a piece of equipment, one would compute the mean value of the data-set that is available on the failure-times of such pieces of equipment. Of the various desirable properties of point estimators well-known in the statistical literature, the property known as *efficiency* carries special significance in that a probability model based on an 'efficient' estimator is likely to fit the data *better* than models based on estimators that are inferior in terms of efficiency.

In the context of fitting probability models to reliability-related data, it is worth mentioning a particular class of continuous probability distributions that has been named 'Self-Inverse at Unity' (SIU). For any such distribution, the $(1-q)^{th}$ quantile is the reciprocal of the q^{th} quantile and, as a result, the median of the distribution lies at unity. The important point is that the reciprocity property provides a method of *modifying* well-known estimators of distribution parameters such that the modified estimators are *more efficient* than the well-known estimators. Only very recently, a number of papers have appeared containing modifications to well-known estimators of parameters of SIU probability models such that the sampling distributions of the modified estimators are *narrower* than those of the well-known estimators.

SIU distributions can be regarded as a special case of distributions that can be given the nomenclature 'Self-Inverse at A' ('SIA') where A can be any positive real number and represents the median of the distribution. The lognormal distribution and the Birnbaum Saunders distribution, two well-known probability models that are extensively used in various areas of engineering, belong to the class of 'SIA' distributions. In this paper, we propose a modification to the formula of the ordinary sample mean for estimating the mean of an SIA probability distribution. We prove that, similar to the ordinary sample mean, the SIA-based modified mean is an *unbiased* estimator of the distribution mean. By conducting a simulation study based on repeated sampling from the lognormal distribution, we show that the SIA-based modified mean is a *more efficient* estimator of the distribution mean than the ordinary sample mean. By fitting the lognormal distribution to a data-set pertaining to repair-times of construction equipment using (i) the ordinary sample mean and (ii) the SIA-based modified mean, we show that the utilization of the modified mean provides *a better fit*. It appears that adoption of the proposed modification may lead to *more accurate modelling* in situations where one has reasons to believe that an SIA distribution is an appropriate probability model for the data at hand.

1. INVERTED OR 'INVERSE' DISTRIBUTIONS

Inverted distributions have been of interest to researchers in the area of distribution theory and a number of inverted or 'inverse' distributions have been derived during the past few decades. For example, Lin et al. (1989) present the utilization of the inverted gamma distribution as a life distribution and provide an example based on a maintenance data set , Khan et al. (2008) discuss the flexibility of the Inverse Weibull distribution in that, for different values of the parameters, the Inverse Weibull approaches various distributions, and Soliman et al. (2010) discusses Bayesian and non-Bayesian estimation problem of the unknown parameter for the inverse Rayleigh distribution based on lower record values.

2. PROBABILITY DISTRIBUTIONS SELF-INVERSE AT UNITY

Distributions possessing the property of invariance under the reciprocal transformation can be regarded as an interesting sub-class of the class of inverted distributions. Snedecor (1934) developed the *F* distribution and utilized the fact that $1/F(m,n)$ is distributed as $F(n,m)$ to obtain lower percentage points of the distribution. Evidently, the case $m=n$ yields invariance under the reciprocal transformation. Finney (1938) derived the distribution of $\omega = s_1 / s_2$ and mentioned that because of the fact that the distributions of ω and ω^{-1} are identical. Seshadri (1965) discussed in detail distributions of non-negative continuous random variables for which the distribution of 1/X is the same as that of the original random variable X. He (1965) presented a necessary and sufficient condition for the fulfillment of this property, and provided a number of examples of such distributions.

Saunders (1974) generalizes Seshadri (1965)'s reciprocal property for the normal family of distributions and presents a number of important theoretical results. Also, he shows that if Y is any nonnegative random variable with density g, then, for $t > 0$, the random variable T with density $\left[g(t)/2\right] + \left[g(1/t)/\left(2t^2\right)\right]$ also possesses the reciprocal property. Habibullah et al. (2010) provide two types of *differential equations* for generating distributions possessing this property.

Habibullah et al. (2010) adopt the nomenclature "Strictly Closed Under Inversion" for distributions invariant under the reciprocal transformation whereas Habibullah and Saunders (2011) introduce the term Self-Inverse at Unity for such distributions. Habibullah (2012) adopts the abbreviation SIU for distributions self-inverse at unity.

3. MODIFIED ESTIMATORS BASED ON THE SIU PROPERTY

Whereas the phenomenon of self-inversion at unity caught the interest of researchers nearly three quarters of a century ago, only very recently has this property been begun to be utilized for obtaining estimators of distribution parameters that are more efficient than their well-known counterparts. Habibullah and Saunders (2011) use the SIU property to modify the formula of the well-known estimator of the *cumulative distribution function* and show that the sampling distribution of the modified formula is tighter than that of the original estimator when sampling from a distribution self-inverse at unity; Fatima et. al. (2013) achieve a similar result for the *cumulative hazard function*.

Fatima and Habibullah (2013a,b) propose self-inversion-based modifications of *L-estimators* of central; tendency and *dispersion* whereas Habibullah and Fatima (2014 a, b & c) propose SIU-based modifications to the formulae of the well-known *percentile coefficient of kurtosis*, *Crow & Siddiqui's Coefficient of Kurtosis* and *Kelley's Measure of Skewness*. Through simulation studies, they show that the newly proposed estimators are likely to be *more efficient* than the corresponding well-known formulae when sampling from SIU distributions.

4. PROBABILITY DISTRIBUTIONS SELF-INVERSE AT A

Habibullah and Saunders (2011) show that the self-inversion at unity property is a special case of the property that we will refer to as "Self-Inverse at A" ("SIA") and which can be stated as follows:

Definition 4.1: For any positive real number A, the probability distribution of a non-negative continuous random variable X can be regarded as being Self-Inverse at A ("SIA") if the distribution of X/A is identical to that of A/X.

Of the probability models that belong to the class of SIA distributions, the lognormal distribution and the Birnbaum Saunders distribution, are extensively used in engineering and reliability studies.

5. SIA-BASED MODIFIED MEAN

SIU distributions can be regarded a sub-class of the class of SIA distributions i.e. those for which A=1. For any SIA distribution, the *(1-q)th* quantile is connected to the *qth* quantile by the equation $X_{(1-q)} / A = A / X_q$ and the median of the distribution is A. In this paper, we propose the following modification to the formula of the ordinary sample mean for estimating the mean of an SIA probability distribution:

$$\bar{x}_{SIA} = \frac{\sum_{i=1}^{n} x_i + A^2 \sum_{j=1}^{n} x_j^{-1}}{2n} \qquad (6.1)$$

where n is the sample size and A is the median of the sample. (Here, the sample median is taken to be the same as the median of the distribution by a method similar to the well-known method of moments.)

It is to be noted that, letting A=1 on the right-hand side of (6.1), we obtain the SIU-based modified mean proposed by Fatima and Habibullah (2013) for the case when the sample has been drawn from an SIU distribution.

6. MATHEMATICAL PROOF OF UNBIASEDNESS OF THE SIA-BASED MODIFIED MEAN

It is well-known that the sample mean $\bar{X}$ is an unbiased estimator of the distribution mean μ which is one of its highly desirable properties. In this section, we present a mathematical proof of the fact that the newly proposed SIA-based modified mean is also an unbiased estimator of the distribution mean μ.

From (6.1) we have

$$E\left(\bar{X}_{SIA}\right) = \frac{1}{2n}\left\{\left(\sum_{i=1}^{n} E\left(X_i\right)\right) + A^2\left[\sum_{j=1}^{n} E\left(\frac{1}{X_j}\right)\right]\right\} \qquad (7.1)$$

Now, it is easy to show that, for a random variable X having an SIA distribution, where A is an arbitrary constant,

$$E\left(\frac{X}{A}\right) = E\left(\frac{A}{X}\right)$$

provided these expectations exist.

Hence eq. (7.1) can be written as

$$E\left(\bar{X}_{SIA}\right) = \frac{1}{2n}\left\{\left(A\sum_{i=1}^{n} E\left(\frac{X_i}{A}\right)\right) + A\sum_{i=1}^{n} E\left(\frac{X_i}{A}\right)\right\}$$
$$= \frac{1}{2n}\left\{2\sum_{i=1}^{n} E\left(X_i\right)\right\} = \frac{1}{n}\left(\sum_{i=1}^{n}\mu\right) = \frac{1}{n}(n\mu) = \mu$$

implying that $\bar{X}_{SIA}$ is an unbiased estimator of μ.

7. SIMULATION STUDY

In this section, we present the results of a simulation study that has been carried out in order to demonstrate that the sampling distribution of the modified estimator (6.1) is narrower than that of the ordinary sample mean. One thousand samples of size 50 each were drawn from the lognormal distribution with $\mu = 2$ and $\sigma = 1$ and the simple arithmetic mean was computed for each of the 1000 samples. Utilizing the fact that the median of the lognormal distribution is equal to e^{μ}, the SIA-based Modified Mean was computed for each of the 1000 samples by setting A=7.389 in (6.1).

7.1 Comparison of Sampling Distributions of the Ordinary Sample Mean and the SIA-Based Modified Mean

Table 1 presents a comparison of the sampling distributions of the ordinary sample mean and the SIA-based modified mean. From the table, it is obvious that:

- The range of the sampling distribution of the SIA-based modified mean is *less* than that of the ordinary sample mean.
- The coefficient of range, a relative measure of dispersion given by $\left(X_m - X_0\right)/\left(X_m + X_0\right)$ where X_m stands for the maximum value and X_0 for the minimum value is *much smaller* in the case of the SIA-based modified mean than in the case of the ordinary sample mean.

Table 1: Comparison of the sampling distributions of the Ordinary Sample Mean and the SIA-Based Modified Mean when drawing 1000 samples of size 50 each from the Lognormal Distribution with $\mu = 2$ and $\sigma = 1$ through coefficient of range and coefficient of variation

	Sampling distribution of Ordinary Sample Mean	*Sampling distribution of SIA-Based Modified Mean*
Minimum	*06.3538*	*09.4225*
Maximum	*21.7630*	*19.3176*
Range	*15.4092*	*09.8951*
Half-Range	*07.7046*	*04.9476*
Mid-Range	*14.0584*	*14.3700*
Coefficient of Range	*00.5480*	*00.3443*
Mean	*12.2390*	*12.1876*
Variance	*5.1648*	*1.6346*
Coefficient of Variation	*0.1857*	*0.1049*

- The simulation is *verifying* the fact that, similar to the ordinary sample mean, the SIA-based modified mean is an unbiased estimator of the mean of the lognormal distribution. (This is so due to the fact that the mean of the lognormal distribution with $\mu = 2$ and $\sigma = 1$ is $e^{\mu+\sigma^2/2} = e^{2.5} = 12.1825$.)
- The variance of the modified mean is *much smaller* than that of the ordinary sample mean.
- The coefficient of variation of the modified mean is *a little more than half* of the coefficient of variation of the ordinary sample mean.

8. APPLICATION

In this section, we apply the newly proposed estimator to a data-set given in Fan and Fan (2015). According to the authors, the data used in this research-paper comes from a contractor's equipment fleet which works on 3-shift schedule around the clock in Canada and consists of full working records of downtime, uptime, failure events, and repair details on each unit. Of the 30 values of the time to repair (TTR) data of one piece of construction equipment given in the paper, we pick up all but one, the one that reads 0.00. (This value is discarded in order to facilitate the computation of the SIA-based modified mean the formula of which involves the reciprocals of the observations.)

The 29 values of the TTR data (other than 0.00) are as follows:

3.20, 11.58, 38.37, 1.83, 6.52, 27.43, 15.93, 1.02, 0.50, 7.25, 1.17, 27.40, 92.90, 62.77, 0.33, 0.33, 0.95, 8.58, 1.00, 0.50, 13.83, 1.72, 10.25, 5.25, 40.02, 31.93, 88.27, 0.50, 4.35.

The histogram of this data-set is given in Fig. 1. The shape of the histogram being positively skewed, we decide to fit the lognormal distribution to this data-set.

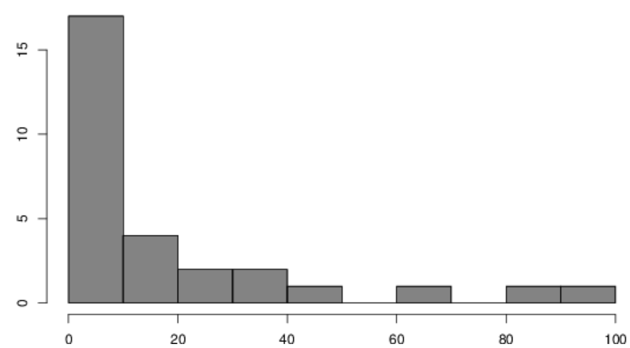

Fig. 1: Histogram of the TTR data drawn using the free online software available at http://www.wessa.net/rwasp_histogram.wasp

8.1 Model-fitting Using Ordinary Sample Mean

For these 29 values, we find that the median is equal to 6.52 and the ordinary arithmetic mean is 17.4372 whereas, for the lognormal distribution with location parameter μ and scale parameter σ, we have

$$median = e^{\mu} \Rightarrow \mu = \ln(median)$$

and

$$mean = e^{\mu+\frac{\sigma^2}{2}} = median.\left(e^{\frac{\sigma^2}{2}}\right)$$

implying that

$$\sigma = \sqrt{2\left[\ln\left(mean\right) - \mu\right]}$$

Applying a method similar to the method of moments, we obtain

$$\hat{\mu} = \ln\left(median\right) = \ln\left(6.52\right) = 1.8749$$

and

$$\hat{\sigma} = \sqrt{2\left[\ln\left(17.4372\right) - 1.8749\right]} = 1.4026$$

In order to compare our data with the lognormal distribution with location parameter 1.8749 and scale parameter 1.4026, we apply the Kolmogorov-Smirnov test, a well-known non-parametric procedure for testing the goodness of fit. The value of the K-S statistic D comes out to be 0.200 which is less than the critical value 0.246 at 5% level of significance. Hence, the lognormal distribution with $\mu = 1.8749$ and $\sigma = 1.4026$ seems to fit the TTR data adequately.

8.2 Model-fitting Using SIA-Based Modified Mean

Substituting $A = 6.52$ in (6.1), the SIA-based modified mean comes out to be 22.446. As such, we have $\hat{\mu} = 1.8749$ (as before) and

$$\hat{\sigma} = \sqrt{2\left[\ln\left(mean_{SIA}\right) - \ln\left(median\right)\right]} = 1.5724.$$

Using *these* as the parameter-values of the lognormal distribution, the value of the K-S statistic D comes out to be 0.173 which is ***less than*** 0.200, the value that we obtained when using the simple mean for estimating σ. Hence we conclude that the lognormal distribution obtained through the use of the SIA-based modified mean fits the TTR data *better* than the lognormal distribution obtained through the use of the ordinary sample mean.

9. CONCLUDING REMARKS

Maintenance of the construction equipment fleet being crucially important for timely completion of megaprojects undertaken by civil engineers, the advantageousness of utilizing available data on failure-times, repair-times, etc. cannot be over-emphasized. Probability models that fit the data with a high degree of accuracy facilitate determination of probabilities that may be very valuable in project-planning and in taking informed decisions on allocations of equipment and maintenance resources. In this paper, we have proposed a modification to the formula of the ordinary sample mean on the basis of a property of continuous distributions of non-negative random variables for which we have introduced the terminology "Self-Inversion at Unity". We have shown that the newly proposed SIA-based estimator matches the sample mean in that it is an *unbiased* estimator of the distribution mean, and *surpasses* the sample mean in that its sampling distribution is 'tighter' than that of the sample mean. The newly proposed estimator can be profitably used when probability density functions such as the lognormal distribution, the Birnbaum Saunders distribution or some other distributions belonging to the class of SIA distributions seem to be viable models for modeling the data at hand. The SIA-based modified mean's potential for *more accurate modelling* has important implications for better decision-making during the course of megaprojects undertaken by civil engineers as well as in other branches of engineering.

10. REFERENCES

Fan, Q. and Fan, H. (2015), Reliability Analysis and Failure Prediction of Construction Equipment with Time Series Models, *Journal of Advanced Management Science Vol. 3, No. 3, September 2015.*

Fatima, S.S. and Habibullah, S.N. (2013a), "Self-Inversion-Based Modifications of L-Estimators of Central Tendency for Probability Distributions in the Field of Reliability and Safety", Oral Presentation at the *International Conference on Safety, Construction Engineering and Project Management (ICSCEPM)* (Islamabad, Pakistan, August 19-21, 2013).

Fatima, S.S. and Habibullah, S.N. (2013b), "On Modifications of L-Estimators of Dispersion in the Case of Self-Inverse Distributions", Proceedings of 11th International Conference on Statistical Sciences: Social Accountability, Global Economics and Human Resource Development with Special Reference to Pakistan (Indus International Institute (NCBA&E Sub-Campus), Dera Ghazi Khan, Pakistan, October 21-23, 2013).

Fatima, S.S., Habibullah, S.N. and Saunders, S.C. (2013), "Some Results Pertaining to the Hazard Functions of Self-Inverse Life-Distributions"; S. N. Habibullah invited to render Oral Presentation of this paper as Keynote Speaker at the *Third International Conference on Aerospace Science and Engineering (ICASE 2013)* (Islamabad, Pakistan, August 21-23, 2013).

Finney, D.J. (1938). The Distribution of the Ratio of Estimates of the Two Variances in a Sample from a Normal Bi-variate Population. *Biometrika*, 30, 190-192.

Habibullah, S.N. (2012), "A Generalization of the Standard Half–Cauchy Distribution", Proceedings of the *Twelfth Islamic Countries Conference on Statistical Sciences (ICCS-12)*, Doha, Qatar (Dec 19 to 22, 2012).

Habibullah, S.N. and Fatima, S.S. (2014a), "SIU-Based Modification in Kelley's Measure of Skewness to Achieve Gains in Efficiency" presented at the Second ISM International Statistical Conference 2014 with Applications in Sciences and Engineering (ISM II) organized by Faculty of Industrial Sciences and Technology, Universiti Malaysia Pahang, Kuantan, Malaysia on Aug 12-14, 2014; *Sponsoring Agency: Higher Education Commission, Pakistan.*

Habibullah, S.N. and Fatima, S.S. (2014b), "SIU-Based Modification in Crow and Siddiqui's Coefficient of Kurtosis to Achieve Gains in Efficiency" accepted for Oral Presentation at the Thirteenth Islamic Countries Conference on Statistical Sciences (ICCS 13) organized by ISOSS and Department of Statistics, Bogor Agricultural University, Indonesia on Dec 18-21, 2014.

Habibullah, S.N. and Saunders, S.C. (2011), "A Role for Self-Inversion", *Proceedings of International Conference on Advanced Modeling and Simulation (ICAMS, Nov 28-30, 2011)* published by Department of Mechanical Engineering, College of Electrical and Mechanical Engineering, National University of Science and Technology (NUST), Islamabad, Pakistan, Copyright 2011, ISBN 978-869-8535-11-7.

Habibullah, S.N., Memon, A.Z. and Ahmad, M. (2010).*On a Class of Distributions Closed Under Inversion*, Lambert Academic Publishing (LAP), ISBN 978-3-8383-4868-1

Khan, M.S., Pasha, G.R. and Pasha, A.H. (2008), Theoretical Analysis of Inverse Weibull Distribution, WSEAS Transactions on Mathematics, ISSN: 1109-2769 33 Issue 2, Volume 7, Page 30-38.

Lin, C.T., Duran, B.S. and Lewis, T.O. (1989) Inverted Gamma as a Life Distribution, Microelectronics Reliability, Volume 29, Issue 4, Pages 619–626.

Saunders, S.C. (1974). A Family of Random Variables Closed Under Reciprocation. *J. Amer. Statist. Assoc.*, 69(346), 533-539.

Seshadri, V. (1965). On Random Variables which have the Same Distribution as their Reciprocals. *Can. Math. Bull.*, 8(6), 819-824.

Snedecor, G.W. (1934). *Calculation and Interpretation of Analysis of Variance and Covariance.* Collegiate Press Inc., Ames, IA.

Soliman, A., Amin, E.A. and Abd-El Aziz, A.A. (2010) Applied Mathematical Sciences, Vol. 4, No. 62, 3057 - 3066 Estimation and Prediction from Inverse Rayleigh Distribution Based on Lower Record Values