Contents lists available http://www.kinnaird.edu.pk/

Journal of Research & Reviews in Social Sciences Pakistan

Journal homepage: http://journal.kinnaird.edu.pk

MODELING BLADDER CANCER PATIENTS DATA THROUGH AN SIA LOG-SYMMETRIC DISTRIBUTION FOR ACHIEVING BETTER FIT

Saleha Naghmi Habibullah[1*], Kessica Xavier[2]
[1] Department of Statistics Kinnaird College For Women Lahore, Pakistan.
[2] Department of Statistics Kinnaird College For Women Lahore, Pakistan.

Article Info

*Corresponding Author
Tel: +92 333 4249674
Email Id: saleha.habibullah@kinnaird.edu.pk

Keywords
Remission times data, SIA Log- Symmetric distributions, better-fitting model.

Abstract

Demography is the study of facts and figures regarding human populations such as births, deaths, incidence of disease, etc. which illustrate the changing structure of populations of various countries of the world. In this paper, we consider a data-set pertaining to remission times of bladder cancer patients which has already been modeled by a few researchers using various probability distributions. With a view to obtaining an even better fit than those achieved so far, we apply one of the SIA log-symmetric distributions to this data-set utilizing the self-inversion property for parameter-estimation. Evidently, the better the fit, the more accurate will be the estimation of probabilities of early remission, an information necessary for oncologists and cancer-care providers.

1. Introduction

Demography is a branch of social sciences concerned with the study of human populations and, in this regard, about half a century ago, it was realized that a study of phenomena such as fertility, mortality and migration needs to be interdisciplinary and must incorporate several other disciplines (Mayer 1962).
These disciplines include economics, climate change, environmental studies, health sciences Health of human beings is an indispensably important concern, and cancer is a severe and epidemiology. (See, for example, Preston 1975 and McMichael et al. 2005). According to Napalkov, (2004), the phenomenon of epidemiological transition based on medical innovation in disease or sickness therapy relates to the frequency, distribution, and control of disease in a population and this, in turn, relates to the changes which are occurring in the incidence of malignant tumors, as well as to mortality from them and to the opportunities for cancer control.

Health of human beings is an indispensably important concern, and cancer is a sever disease that attracts the attention of medical researchers all over the world. Obviously, the shorter the remission time of a cancer patient, the greater the need for patient care and, in this regard, collection of *data* on remission times of various categories of cancer patients is important as this can provide *information* conducive to proper planning for treatment and care of returning patients. The information contained in the data facilitates estimation of *probabilities of early remission* subsequent to attainment of *a well-fitting probability model*. The better a probability model fits the data, the more *accurate* will be the estimation of probabilities leading to informed decision-making.

The objective of this paper is to demonstrate the usefulness of *SI*A *log-symmetric distributions* (Habibullah, 2017) in obtaining models that are likely to fit data-sets such as those of remission times of cancer patients *better* than some other probability models. This is achieved through the development and utilization of self-inversion-based estimators of distribution parameters that are *more efficient* than the well-known method of moments estimators. The term *'SI*A-*estimator'* was adopted by Ali and Habibullah (2016) as well as by Xavier and Habibullah (2016) to represent any such estimator. The terminology *'SI*A-*estimation'* can also be used for this methodology.

In this section, we present a set of real data on remission times of bladder cancer patients. The data-set has been extracted from Lee and Wang (2003) and consists of remission times *in months* of a random sample of 128 patients. The data-values are as shown below:

0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 3.80, 25.74, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2..62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 3.88, 5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.2, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93, 1.46, 18.10, 11.79, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 13.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 12.07, 6.76, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63, 22.69

The data-set given above has already been modeled by a few researchers using various probability distributions. We will be applying an SIA log-symmetric distribution to this data-set using the method of SIA-estimation, the objective being to obtain a fit which is even better than that achieved by previous researchers.

2. SIA LOG-SYMMETRIC DISTRIBUTIONS (SOME HISTORY)

Seshadri (1965) focused on the distributions of non-negative continuous random variables for which the reciprocal of the distribution is the same as that of the original random variable. He (1965) derived a necessary and sufficient condition for the fulfillment of this property and provided quite a few examples of such distributions.

Saunders (1974) generalized Seshadri (1965)'s reciprocal property for the normal family of distributions and presented a number of important theoretical results. Also, he showed that, for any nonnegative random variable with density g, the random variable T with density

$$\left[g(t)/2\right]+\left[g(1/t)/\left(2t^2\right)\right], t>0$$

also possesses the reciprocal property.

Habibullah et al. (2010) formulated two types of *differential equations* for generating distributions possessing this property and adopted the nomenclature "Strictly Closed Under Inversion" for such distributions, Habibullah and Saunders (2011) introduced the term 'self-inverse at unity' for these distributions and Habibullah (2012) adopted the abbreviation SIU to represent the term 'Self-Inverse at Unity'. Habibullah and Fatima (2015) introduced the terminology 'Self-Inverse at A (SIA)' for distributions of the non-negative continuous random variable X for each of which, X/A is distributed exactly as A/X where A represents the median of the distribution. It was obvious that 'SIU' is a special case of 'SIA' and that the class of 'SIA-distributions' is *much wider* than the class of SIU distributions.

In almost the same period, Jones (2008) discussed the class of distributions of non-negative continuous random variables X for which the distribution of X/median is identical to the distribution of median/X and adopted the nomenclature 'log-symmetric' for these distributions due to the fact that application of the transformation Y = log X to any such distribution yields a symmetric distribution extending over the entire real line.

Habibullah (2017) proposed that the two nomenclatures be *merged* and such distributions be called 'SIA log-symmetric distributions'. Some of the well-known distributions belonging to this class are the lognormal, log-logistic, log-Laplace and log-Cauchy distributions.

3. 'SIU' and 'SIA'-ESTIMATORS'

Just as the term 'SIA-estimator' represents estimators of distribution parameters based on the self-inversion property of distributions that are self-inverse at A, the term 'SIU-estimator' can be used to represent estimators based on the self-inversion property of distributions that are self-inverse at unity. As the substitution A=1 converts an SIA distribution to an SIU-distribution and an 'SIA-estimator' to an 'SIU-estimator', obviously, the set of SIU-estimators is a subset of the set of SIA-estimators.

Habibullah and Saunders (2011) proposed an SIU-estimator of the cumulative distribution function (CDF) that is applicable when the sample has been drawn from an SIU distribution and is more efficient than the well-known empirical cumulative distribution function (ecdf). Fatima et. al. (2013) achieved an SIU-estimator of the cumulative hazard function.

Fatima and Habibullah (2013a, b) proposed 'SIU-estimators' of central tendency and dispersion whereas Habibullah and Fatima (2014 a, b & c) proposed SIU-estimators of the well-known Kelley's Measure of Skewness, Percentile Coefficient of Kurtosis and Crow & Siddiqui's Coefficient of Kurtosis. Through simulation studies, it was shown that the SIU-estimators are more efficient than the corresponding method of moments estimators when sampling from SIU distributions.

Habibullah and Fatima (2015) focused on the general case and proposed an SIA-estimator of the mean of a probability distribution self-inverse at A. The SIA-estimator is given by

$$\bar{x}_{\mathrm{SIA}} = \frac{\sum_{i=1}^{n} x_i + \hat{\mathrm{A}}^2 \sum_{j=1}^{n} x_j^{-1}}{2n}$$

where n is the number of observations and is the median of the sample.

Subsequently, Ali and Habibullah (2016) proposed a decile-based SIA-estimator of the scale parameter of the Log-Cauchy distribution whereas Xavier and Habibullah (2016) proposed a decile-based SIA-estimator of the shape parameter of the Log-Logistic distribution.

4. MODEL-FITTING THROUGH LOG-LOGISTIC DISTRIBUTION

The SIA log-symmetric distribution that has been chosen for this paper is the well-known Log-Logistic distribution the probability density function (PDF) of which can be written as

$$f(x) = \frac{(\beta/\mathrm{A})(x/\mathrm{A})^{\beta-1}}{\left(1+(x/\mathrm{A})^{\beta}\right)^2}, 0 < x < \infty,\ \mathrm{A} > 0,\ \beta > 0 \quad (2)$$

where A is the median as well as the scale parameter, and β is the shape parameter.

The mean of the distribution is given by

$$E(X) = \frac{(\mathrm{A}\pi/\beta)}{\sin(\pi/\beta)}, \beta > 1 \quad (3)$$

The distribution finds a variety of applications; among other things, it has been used for modeling lifetime of an object or of an organism, modeling service-time and modeling water demand.

In the following two sub-sections, we present the results obtained by fitting the Log-Logistic distribution to the data pertaining to bladder cancer patients' remission times given in Section 2 using (i) Non-SIA estimator and (ii)

SIA-estimator of the distribution's shape parameter. We compare the fit based on the SIA-estimator with the one based on Non-SIA estimator.

4.1 MODEL-FITTING USING NON-SIA ESTIMATOR OF THE SHAPE PARAMETER

For these 128 values of bladder cancer patients' remission times, we find that the median is equal to 6.395 and the ordinary arithmetic mean is 9.4455. Substituting these values in equation (3), the value of the shape parameter β turns out to be 2.138. We fit the distribution taking A= 6.395 and β = 2.138 and, in order to test for the goodness of fit, we apply the Kolmogorov-Smirnov test, a well-known non-parametric procedure. The value of the K-S statistic D comes out to be 0.074 which is less than the critical value 0.120 at 5% level of significance. Hence, we conclude that the Log-Logistic distribution with the Non-SIA estimator of β fits the bladder cancer patients' remission times data fairly satisfactorily.

4.2 MODEL-FITTING USING SIA-ESTIMATOR OF THE SHAPE PARAMETER

In order to obtain the SIA-estimator of the shape parameter β, we employ Habibullah and Fatima (2015)'s SIA-estimator of the distribution mean. The value of the sample median being 6.395, the SIA-estimator given by equation (1) comes out to be 12.918. Substituting these values in equation (3), the value of the shape parameter β turns out to be 1.647. We fit the distribution taking A= 6.395 and β = 1.647 and again apply the Kolmogorov-Smirnov test in order to test for the goodness of fit. The value of the K-S statistic D comes out to be 0.044 which is not only less than the critical value 0.120 at 5% level of significance but also less than the value of the K-S statistic in the case of the Non-SIA estimator of β. Hence, we conclude that the Log-Logistic distribution with the SIA-estimator of β fits the bladder cancer patients' remission times data better than the Log-Logistic distribution with the Non-SIA-estimator of β.

4.3 COMPARISON WITH OTHER DISTRIBUTIONS (MERGING SECTION 5 & 6)

In this section, we compare the two models fitted in Section 5 above with five different probability models that have already been fitted to the bladder cancer patients' remission times data given in Section 2. We begin with a review of the literature in this regard:

Khan et al. (2014) introduced the three-parameter Transmuted Inverse Weibull (TIW) distribution. Having derived its fundamental properties, they fitted the distribution to the bladder cancer patients' remission times data by the method of maximum likelihood, they fitted three sub-models of the TIW model i.e. the Transmuted Inverse Rayleigh (TIR), the Transmuted Inverted Exponential Distribution (TIE) and the Inverse Weibull Distribution (IW) models to the same data-set and utilized the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the Mean Square Error (MSE) and the Kolmogorov–Smirnov (K-S) Test to compare the goodness of fit of the four models.
The TIW distribution having the lowest AIC, BIC, MSE and K-S values, they concluded that the TIW distribution proposed by them provides a better fit to the bladder cancer patients' remission times data than each of the other three lifetime distributions.

Kumar et al. (2015) proposed the use of sine function for obtaining a new life time distribution in terms of some life time distribution used as baseline distribution. They adopted the nomenclature 'SS transformation' for their method and applied it to the exponential distribution with inverse scale parameter □ in order to obtain the 'SSE (□)-distribution'. Having computed the MLE of the parameter □ of SSE(□)-distribution for the bladder cancer patients' remission time data given in Section 2, they fitted the distribution to the data and, comparing the values of AIC, BIC, log-likelihood and KS statistic obtained for the SSE(□)-distribution with those for TIW, TIR, TIE and IW distributions considered by Khan et al. (2014), concluded that the SSE(□)-distribution provides a better fit to this data-set than each of the other four lifetime distributions.

In Table 1, we present the values of the K-S statistic D for each of the four models fitted by Khan et al. (2014) and the SSE (□)-model fitted by Kumar et al. (2015) as well as the two models fitted in Sub-sections 5.1 and 5.2 (i.e. the Log-Logistic distribution using Non-SIA estimator and the Log-Logistic distribution using SIA-estimator).

TABLE 1: Computed Values of The K-S Statistic D For $Ss_e(\Theta)$, Tiw, Tie, Iw, Tir And Log-Logistic Distribution Using Sia Estimator And Non-Sia Estimator

PROBABILITY DISTRIBUTION	VALUE OF K-S STATISTIC
Log-Logistic distribution using SIA-estimator of shape parameter	0.044
$SS_E(\theta)$ distribution	0.067
Log-Logistic distribution using Non-SIA estimator of shape parameter	0.074
TIW distribution	0.119
IW distribution	0.131
TIE distribution	0.155
TIR distribution	0.676

As indicated in Sub-section 5.1, for sample size 128, the critical value of the K-S statistic at 5% level of significance is 0.120. As such, we see that four of the seven fitted models yield K-S values that fall in the acceptance region whereas three do not. Of the four that can be regarded as good fits to the bladder cancer patients' data, the Log-Logistic distribution using SIA estimator has the lowest K-S value in comparison with the others. As such, we conclude that the Log-Logistic distribution using SIA estimator of the shape parameter provides the best fit to the bladder cancer patients' remission times data among the seven models contained in Table 1.

6. CONCLUDING REMARKS

Demography deals with the changing structure of human populations and phenomena such as incidence of a life-threatening disease have a direct relationship with potential demographic change. Human health being of paramount importance, it is important to collect *data* pertaining to severe diseases such as cancer. Collection of data on remission times of various categories of cancer patients, for example, can provide *information* conducive to proper planning for treatment and care of returning patients. The data facilitates estimation of *probabilities of interest* through well-fitting probability models; the *better* a probability model fits the data, the more *accurate* is the estimation of probabilities leading to informed decision-making, In this paper, we have focused on a data-set pertaining to *remission times of bladder cancer patients* which has already been modeled by a few researchers using five different probability distributions. The shape of the data being positively skewed, we have chosen the two-parameter Log-Logistic distribution for modeling this data-set and, utilizing the method of SIA-estimation for estimating the shape parameter of the distribution, we have obtained a fit that has turned out to be *better* than all five distributions used by previous researchers. As such, we have demonstrated the usefulness of *SI*A *log-symmetric distributions* in conjunction with the method of *SI*A-*estimation* in obtaining models that are likely to fit data on remission-times, survival times and the like *better* than some of the Non-SIA probability models.

REFERENCES

Ali, A. and Habibullah, S. N. (2016). On a Decile-Based SIA-Estimator of the Scale Parameter of the Log-Cauchy Distribution. Paper presented at the 14th International

Conference on Statistical Sciences, Karachi, Pakistan.

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.

Fatima, S.S. and Habibullah, S.N. (2013b). On Modifications of L-Estimators of Dispersion in the Case of Self-Inverse Distributions. Paper presented at the 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.

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

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.

Habibullah, S.N. (2017). On the Power Generalization of Log-Symmetric SIA Distributions. *Journal of ISOSS,* 3(1), 45-54.

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, Kuantan, Malaysia.

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

Habibullah, S.N. and Fatima, S.S. (2014c). Self-Inversion-Based Modification of Crow and Siddiqui's Coefficient of Kurtosis to Achieve Gains in Efficiency. Presented at the Thirteenth Islamic Countries Conference on Statistical Sciences (ICCS 13) organized by Bogor Agricultural University, Bogor, Indonesia.

Habibullah, S.N. and Fatima, S.S. (2015). On a Newly Developed Estimator for More Accurate Modeling with an Application to Civil Engineering. Proceedings of the 1^{2th} International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP12) organized by CERRA (Vancouver, BC, Canada, July 12-15, 2015*). Sponsoring Agency: Higher Education Commission, Pakistan.*

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, 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

Jones, M. C. (2008), On Reciprocal Symmetry, *Journal of Statistical Planning and Inference*, 138(10) pp. 3039–3043.

Khan, M. S., King, R. and Hudson, I. (2014): *Characterisations of the Transmuted Inverse Weibull Distribution*. ANZIAM J. 55, (EMAC2013), C197–C217.

Kumar, D., Singh, U. and Singh, S. K. (2015), A New Distribution Using Sine Function- Its Application To Bladder Cancer Patients Data, *Journal of Statistics Applications & Probability,* 3(4**)**, 417-427.

Lee, E. T. and Wang, J. W. (2003). *Statistical Methods for Survival Data Analysis*. Wiley, New York, DOI: 10.1002/0471458546.

Mayer, K. (1962). "Developments in the Study of Population". *Social Research* 3(29), 293-320.

McMichael, A. J. and Woodruff, R. E. (2005). Climate change and human health, *Encyclopedia of World Climatology*, Springer.

Napalkov, N. P. (2004), Cancer and demographic transition, *Vopr Onkol*. 2004; 50(2):127-44. Review. Russian.

Preston, S. H. (1975). The Changing Relation between Mortality and Level of Economic Development. *Population Studies*, 2(29), 231-248.

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.

Xavier, K. and Habibullah, S.N. (2016). On a Decile-Based SIA-Estimator of the Shape Parameter of the Log-Logistic Distribution. Presented at *the 14th International Conference on Statistical Sciences*, Karachi, Pakistan.