SEMI-PARAMETRIC ESTIMATION OF $P_{X,Y}(\{(X,Y)/X>Y\})$ FOR THE POWER FUNCTION DISTRIBUTION

Dhanya M*

E S Jeevanand**

Abstract:

The stress-strength model describes the life of a component which has a random strength X and is subjected to random stress Y, in the context of reliability. The component will function satisfactorily whenever X>Y and it fails at the instant the stress applied to it exceeds the strength. R=P(Y<X) is a measure of component reliability .In this paper, we obtain semi parametric estimators of the reliability under stress- strength model for the Power function distribution under complete and censored samples. We illustrate the performance of the estimators using a simulation study.

Key words: *Stress strength model, Power function distribution, Censored sample, Semi parametric estimates.*

* Department of Management Studies, Federal Institute of science and Technology, Angamaly , Kochi-683577, India

** Department of Mathematics, Union Christian College, Aluva-2 , Kerala, India

1. Introduction

The stress-strength model describes the life of a component which has a random strength X and is subjected to random stress Y, in the context of reliability. The component will function satisfactorily whenever X>Y and it fails at the instant the stress applied to it exceeds the strength. R=P(Y<X) is a measure of component reliability .Another interpretation of R is that it measures the effect of the treatment when X is the response for a control group and Y refers to the treatment group. It has many applications in engineering concepts such as structures ,deterioration of rocket motors, static fatigue of ceramic components ,the ageing of concrete pressure vessels. For more real life application in engineering ,reliability, quality control, medicine and psychology we refer Jeevanand (1997,1998) ,Kotz et .al. (2003), Nadarajah (2005), Nadarajah and Kotz (2006).

Most of the studies relating to estimation of the reliability measures are restricted to the parametric framework, either in the classical or Bayesian methods. It is often felt that statistical models based on conventional parametric distributions are not flexible enough to provide a reliable description of survival data. This has led to a wide spread use of non-parametric estimators, notably the Kaplan- Meier estimator for the survival function and the Nelson – Aalen estimator for the cumulated hazard. Semi parametric estimators are suggested as a compromise to the conventional parametric and non-parametric estimators.

Power laws appear widely in physics, biology, earth and planetary sciences, economics and finance, computer science, demography and the social sciences. The city populations size, the sizes of earthquakes, moon craters, solar flares, computer files and wars, the frequency of use of words in any human language, the frequency of occurrence of personal names in most cultures, the numbers of papers scientists write the number of citations received for the papers, the number of hits on web pages, the sales of books, music recordings sales of branded commodity, the numbers of species in biological taxa, people’s annual incomes and a host of other variables all follow power-law distributions. Power law implies that the small occurrence are extremely common, where as large instance are extremely rare. The power function distribution is given by pdf

$$f(x,\alpha,\beta) = \frac{\alpha}{\beta^{\alpha}}(x)^{\alpha-1}, 0 < x < \beta, \alpha, \beta > 0 \quad (1.1)$$

In this paper, we present semi parametric estimators for reliability of Power function distribution using complete and censored sampling scheme. In the next section least square estimation based on survival function are obtained. In section 3 , we consider the estimation under complete sampling scheme. In section 4, estimation under type I censoring scheme are discussed. In section 5,we illustrate the performance of the estimators using a simulation study.

2. Least square estimator based on Survival function

Let X and Y are independent power function distribution with pdf

$$f(x,\alpha_1,\beta) = \frac{\alpha_1}{\beta^{\alpha_1}} x^{\alpha_1 - 1}, 0 < x < \beta, \alpha_1, \beta > 0 \quad (2.1)$$

and

$$f(y,\alpha_2,\beta) = \frac{\alpha_2}{\beta^{\alpha_2}} y^{\alpha_2 - 1}, 0 < y < \beta, \alpha_2, \beta > 0 \quad (2.2)$$

$$P(X>Y) = \frac{\alpha_1}{\alpha_1+\alpha_2} \quad (2.3)$$

Afify (2003) derived the least square estimator using the regression of survival function on the observations, for the Pareto distribution. We used this approach to derive another estimator for the reliability function R. The survival function at the sample points $x_1, x_2, ..., x_n$,`from a power function distribution is

$$R(x_i) = P(X_i \geq x_i) = 1 - (\frac{x}{\beta})^{\alpha_1}, i = 1,2,3,... \quad (2.4)$$

Taking logarithms on both sides of (2.4) we get

$$\ln R(x_i) = \alpha_1 \ln \beta - \alpha_1 \ln x_i, i = 1,2, ... n \quad (2.5)$$

$$E = \sum_{i=1}^{n} [\ln R(x_i) + \alpha_1 \ln x_i - \alpha_1 \ln \beta]^2 \quad (2.6)$$

Differentiating partially with respect to the unknown parameters α_1 and β, then equating the results equal to zero we get normal equations as

$$\sum_{i=1}^{n} \ln R(x_i)\ln(x_i) + \alpha_1 \sum_{i=1}^{n} (\ln x_i)^2 - \alpha_1 \sum_{i=1}^{n} \ln\beta \ln x_i = 0 \tag{2.7}$$

(((

$$\frac{\alpha_1}{\beta}\sum_{i=1}^{n} \ln R(x_i) + \frac{\alpha_1^{\,2}}{\beta}\sum_{i=1}^{n} \ln x_i - \frac{n\alpha_1^{\,2}}{\beta}\ln\beta = 0 \tag{2.8}$$

These normal equations can be solved numerically to obtain the values of $\hat{\alpha}_{1LS}$. In a similar manner we will obtain the value of $\hat{\alpha}_{2LS}$. Thus

$$\hat{R}_{LS} = \frac{\widehat{\alpha_{1LS}}}{\widehat{\alpha_{1LS}} + \widehat{\alpha_{2LS}}} \tag{2.9}$$

3. Estimation of R using complete sampling

In this section we obtained the estimate based on sample $\underline{x} = (x_1, x_2, \dots, x_n)$ from the power function density (2.1). An estimate of the survival function $S(x_i)$ is 1-$\hat{F}(x_{i:n}:\theta)$ where $x_{i:n}$ is the i^{th} order statistic and $\hat{F}(x_{i:n}:\alpha_1,\beta) = \frac{i}{n}$,the empirical distribution function. In order to avoid log(0), D'Agostino and Stephans (1986) suggested that $\hat{F}(x_{i:n}:\theta)$ can be approximated by,$\frac{i-c}{n-2c+1}, i = 1,2,\dots n$ where 0≤c≤1. In this paper we take the three popular values for c considered in Wu (2001), Faucher and Tyson (1988) ,viz c=0,0.3 and 0.5.Then (2.7) and (2.8) becomes

$$\sum_{i=1}^{n} \ln\left(\frac{n+1-c-i}{n+1-2c}\right)\ln(x_i) + \alpha_1 \sum_{i=1}^{n} (\ln(x_i))^2 - \alpha_1 \sum_{i=1}^{n} \ln(\beta)\ln(x_i) = 0 \tag{3.1}$$

$$\frac{\alpha_1}{\beta}\sum_{i=1}^{n} \ln\left(\frac{n+1-c-i}{n+1-2c}\right) + \frac{\alpha_1^{\,2}}{\beta}\sum_{i=1}^{n} \ln x_i - \alpha_1^{\,2}\frac{n}{\beta}\ln\beta = 0 \tag{3.2}$$

These normal equations can be solved numerically to obtain $\widehat{\alpha_{1cs}}$ *and* $\hat{\beta}$. Proceeding similarly with $\hat{G}(y_{j:m}:\alpha_2,\beta) = \frac{j-c}{m-2c+1}, j = 1,2,\dots m$ and 0≤c≤1 one can obtain the estimate of$\widehat{\alpha_{2cs}}$ *and* $\hat{\beta}$. So the estimate of R is obtained by

$$\hat{R}_C = \frac{\hat{\alpha}_{1_{cs}}}{\hat{\alpha}_{1_{cs}} + \hat{\alpha}_{2_{cs}}} \tag{3.3}$$

4.Estimation of R using type I censored sample

Finaly here we propose a semi parametric estimator for the reliability function of the power function distribution (2.1) when the strength is censored at a pre determined time T. Suppose n components with power function life times are put on test and we observe the number of components failed at each time point x, x+k, $x+2k$,….up to the time T. Define, n_j as the number of components still functioning at the time $x_j = x+jk$, j=0,1,2,…,$x_j \leq$ T and d-j as the number of components whose failure occurs in the time interval (x_j-1, x_j). Then the Kaplan –Meir estimator of the survival function S(t) (see Jan et al.(2005)) for a given t is

$$S^*(t) = \prod_{j:x_{(j)}\leq x}\left(\frac{n_j - d_j}{n_j}\right) \tag{4.1}$$

For t<t_1,S*(t)=1,where m_j as the number of components still functioning at the time y, y_j= y+jk, and d_j as the number of components whose failure occurs in the time interval (y_j-1, y_j).

$$\sum_{i=1}^{n}\left(\sum_{j:x_{(j)}\leq x}\ln(1-\frac{n_j}{d_j})\right)\ln x_i + \alpha_1\sum_{i=1}^{n}(\ln x_i)^2 - \alpha_1\sum_{i=1}^{n}\ln\beta\ln(x_i) = 0 \tag{4.2}$$

$$\frac{\alpha_1}{\beta}\sum_{i=1}^{n}\left(\sum_{j:x_{(j)}\leq x}\ln(1-\frac{n_j}{d_j})\right) + \frac{\alpha_1^2}{\beta}\sum_{i=1}^{n}\ln x_i - \frac{n\alpha_1^2}{\beta}\ln\beta = 0 \tag{4.3}$$

These normal equations can be solved numerically to obtain $\hat{\alpha}_{1TCS} and\ \hat{\beta}$.. Similarly we will obtain $\hat{\alpha}_{2TCS} and\ \hat{\beta}$

This leads to the estimate of R as

$$\hat{R}_{TCS} = \frac{\hat{\alpha}_{1TCS}}{\hat{\alpha}_{1TCS} + \hat{\alpha}_{2TCS}} \tag{4.4}$$

5. Simulation results

In this section, we present the results of a simulation study in order to compare the performance of these estimators. We perform a simulation study of 2000 samples of sizes n = 25; 50; 100 and 200 generated from (2.1) and (2.2). We have presented the simulation results concerning the bias and mean square errors of all these estimators. In all the simulation results presented here, the bias of an estimator can be determined as the (average value of the estimate report in the table - True value). The variance of an estimator was determined as the sample variance obtained from all the simulations carried out. Finally, the mean square error of estimator is (variance of the estimator + $(\text{Bias})^2$). The bias and mean squared errors of the estimators under complete sampling and censored sampling are presented in Table 1.

Table 1 : Means and MSEs (in parentheses) of the estimates of *R* under complete and censored sampling

(α_1, α_2,β)		(2,2.5,15)	(3,3.2,15)	(2.6,1.8,15)	(1.9,2.6,15)
True R		**0.55555**	**0.51612**	**0.40909**	**0.57778**
n = 25	$\hat{R}_{LS}$	0.39032 (0.0223)	0.43207 (0.0324)	0.56284 (0.0239)	0.37476 (0.0221)
	$\hat{R}_{C}$	0.41263 (0.0056)	0.46160 (0.0046)	0.57057 (0.0049)	0.40989 (0.0045)
	$\hat{R}_{TCS}$	0.40263 (0.0031)	0.4560 (0.0029)	0.56356 (0.0031)	0.39389 (0.0027)
n = 50	$\hat{R}_{LS}$	0.40621 (0.0123)	0.44608 (0.0214)	0.55602 (0.0117)	0.40603 (0.0113)
	$\hat{R}_{C}$	0.40111 (0.0037)	0.44363 (0.0024)	0.55233 (0.0027)	0.40522 (0.0033)

	$\hat{R}_{TCS}$	0.41661 (0.0017)	0.46663 (0.0014)	0.57663 (0.0017)	0.38670 (0.0013)
n = 100	$\hat{R}_{LS}$	0.41227 (0.0011)	0.45725 (0.0017)	0.56228 (0.0034)	0.41326 (0.0017)
	$\hat{R}_{C}$	0.42828 (0.0005)	0.46777 (0.0002)	0.57426 (0.0008)	0.41345 (0.0003)
	$\hat{R}_{TCS}$	0.43827 (0.00012)	0.44604 (0.0040)	0.57816 (0.0020)	0.41592 (0.00012)
n=200	$\hat{R}_{LS}$	0.42521 (0.0009)	0.4609 (0.0001)	0.56912 (0.0005)	0.40231 (0.0004)
	$\hat{R}_{C}$	0.44333 (0.0000)	0.48011 (0.0000)	0.59512 (0.0000)	0.42012 (0.0000)
	$\hat{R}_{TCS}$	0.44194 (0.0010)	0.48911 (0.00002)	0.59191 (0.0009)	0.42371 (0.0009)

The performance of the estimators are evaluated. It is revealed that the estimators does not seem very sensitive with variation of the parameters α_1, α_2 and β. It seems that the MSE and bias of the estimators become smaller as the sample size increases.

References

[1] Ali.M.Masoom and Woo J (2005) On the ratio X/(X+Y) for the power function distribution, *Pakistan Journal of Statistics*, 21(2), 131-138.

[2] Enis, P. and Geisser, S. (1971) Estimation of probability that $Y > X$, *Journal of American Statistical Association*, 66, 162-168.

[3] Faucher, B. and Tyson, W. R. (1988) On the determination of Weibull parameters. *Journal of Material Science Letters*, 7,1199-1203.

[4] Jan, B., Shah, S. W. A., Shah, S. and Qadir, M. F. (2005) Weighted Kaplan Meier estimation of survival function in heavy censoring, *Pakistan Journalof Statistics*, 21 (1), 55-83.

[5] Jeevanand, E.S. and Nair, N.U. (1994) Estimating $P[X > Y]$ from exponential samples containing spurious observations, *Communications in Statistics - Theory and Methods*, Vol. 23(9), 2629-2642.

[6] Jeevanand, E.S. (1997) Bayes Estimation of $P(X2 < X1)$ for a bivariate Pareto distribution, *The Statistician*, A Journal of the Royal Statistical Society, Vol. 46(1) , 93-99.

[7] Jeevanand, E.S. (1998) Bayes estimate of the reliability under stress-strength model for the Marshal- Olkin bivariate exponential distribution, *IAPQR Transactions*,1998, 23(2), 133-136.

[8] Kaplan, E.L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* 53:457-481.

[9] Kelley, G.D., Kelly, J.A. and Schucany, W.R. (1976) Efficient Estimation of $P(Y < X)$ in the exponential case, *Technometrics*, 18, 359-360.

[10]Kim, C., Bae, W., Cho, H., and Park, B. U. (2005) Non-parametric hazard function estimation using the Kaplan - Meier estimator, *Non-parametricStatistics* 17 (8), 937 - 948.

[11] Kotz, S., Lumelskii, Y and Pensky, M (2003) The Stress-Strength Model and its Generalizations, World Scientific, London.