Journal of Statistical
Theory & Methods
1997, Vol. 8, No. 2, pp.231 ~ 237

A Comparative Study on Nonparametric Reliability Estimation for Koziol-Green Model with Random Censorship [1]

Young Joon Cha and Jae Man Lee [2]

Abstract

The Koziol-Green(KG) model has become an important topic in industrial life testing. In this paper we suggest MLE of the reliability function for the Weibull distribution under the KG model. Futhermore, we compare Kaplan-Meier estimator, Nelson estimator, Cheng & Chang estimator, and Ebrahimi estimator with proposed estimator for the reliability function under the KG model.

Key Words and Phrases: Reliability estimation, Koziol-Green model, Random censorship

1. Introduction

Let $X_1, X_2, \cdots, X_n$ be a random sample from a continuous reliability function $\overline{F}(x)$ on $[0, \infty)$. Let $Y_1, Y_2, \cdots, Y_n$ be a random sample from a continuous reliability function $\overline{G}(y)$. Suppose that X_i is assumed to be independent of Y_i for each i. In random censoring model, the true lifetimes $X_i's$ are censored on the right by the censoring times $Y_i's$, so that we can only observe (Z_i, δ_i), where

$$Z_i = min(X_i, Y_i)$$

and

$$\delta_i = \begin{cases} 1, & \text{if } X_i \le Y_i \\ 0, & \text{if } X_i > Y_i, \end{cases}$$

for $i = 1, ..., n$. Let $Z_{(1)} < \cdots < Z_{(n)}$ denote the ordered observed lifetimes, and let $\delta_{(1)}, \cdots, \delta_{(n)}$ be their corresponding unordered indicator values. For this random

[1]This paper was supported by Andong National University Research Fund, 1996.
[2]Department of Statistics, Andong National University

censoring scheme, we consider the KG model $\overline{G}(t) = \overline{F}(t)^{\beta}$ with for each t and an unknown positive constant β. In recent years, KG model has become an important topic in industrial life testing and has received special attention since the paper of Koziol and Green(1976).

Under this KG model, in section 2 we consider the presentation of several nonparametric estimators for the reliability function. In section 3, we suggest the MLE of the reliability function for the Weibull distribution under the KG model. Finally, in secton 4, we compare several nonparametric estimators with proposed estimator based on the mean squared error criterion.

2. Nonparametric Reliability Estimation under KG Model

In reliability and survival analysis, the Kaplan-Meier(1957) estimator(KME) plays an important role and has wide range of applications. The KME of the reliability function $\overline{F}(x)$ is defined by

$$\widehat{\overline{F}}_{KM}(t) = \prod_{i:Z_{(i)} \le t} \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}} \tag{1}$$

Chen, Hollander, and Langberg(1982) compared exact variance with asymptotic variance of KME under KG model. There are some comparable estimators of the KME. Nelson(1972) proposed the estimator $\widehat{\Lambda}(t)$ for the cumulative hazard function $\Lambda(t)$ as follows

$$\widehat{\Lambda}(t) = \sum_{i:Z_{(i)} \le t} \frac{\delta_{(i)}}{n-i+1}. \tag{2}$$

From the relation $\Lambda(t) = -\log\overline{F}(t)$, the Nelson-type reliability estimator $\widehat{\overline{F}}_N(t)$ by substituting $\widehat{\Lambda}(t)$ for $\Lambda(t)$ can be considered as

$$\widehat{\overline{F}}_N(t) = e^{-\widehat{\Lambda}(t)}. \tag{3}$$

On the other hand, Ebrahimi(1985) suggested a nonparametric estimator $\widehat{\overline{F}}_E(t)$ for the reliability function $\overline{F}(t)$ as follows :

$$\begin{aligned} \widehat{\overline{F}}_E(t) &= \alpha_n \exp\{\alpha_n \ln\frac{1}{\alpha_n} + \alpha_n \ln\widehat{\overline{H}}(t)\} \\ &+ (1-\alpha_n)\exp\{\alpha_n \ln\frac{1}{1-\alpha_n} + \alpha_n \ln\widehat{\overline{S}}(t)\}, \end{aligned} \tag{4}$$

where

$$\alpha_n = \frac{1}{n}\sum_{i=1}^{n}\delta_i,\ \ \widehat{\overline{H}}(t) = \frac{1}{n}\sum_{i=1}^{n} I(Z_i > t, \delta_i = 1),\ \text{and}\ \ \widehat{\overline{S}}(t) = \frac{1}{n}\sum_{i=1}^{n} I(Z_i > t, \delta_i = 0).$$

It is known that the $\widehat{\overline{F}}_E(t)$ is strongly consistent, and when properly normalized that this estimator converges weakly to the Gaussian process. That is, the asymptotic distribution of $\widehat{\overline{F}}_E(t)$ is

$$N\Big(\overline{F}(t)\ ,\ \mathrm{Var}_{asy}(\widehat{\overline{F}}_E(t)\ )\ \Big), \tag{5}$$

where

$$\begin{aligned}\mathrm{Var}_{asy}\Big(\widehat{\overline{F}}_E(t)\ \Big) \ &=\ \frac{1}{n}\Big[\frac{\beta}{(\beta+1)^2}\overline{F}^2(t)\Big(\theta\ \ln\frac{1}{\theta} + \theta + (1-\theta)\ln\frac{1}{1-\theta} - 1 \\ &+\theta\ln\overline{H}(t) + (1-\theta)\ln\overline{S}(t)\Big)^2\Big] + \Big\{\overline{H}(t)(1-\overline{H}(t))\theta^4\frac{\overline{F}^2(t)}{\overline{H}^2(t)}\Big\} \\ &+\Big\{\theta^2(1-\theta)^2\frac{\overline{F}^2(t)}{\overline{S}(t)}(1-\overline{S}(t))\Big\},\end{aligned}$$

and $\theta = 1/(1+\beta)$.

Cheng & Chang(1985) proposed the nonparametric estimator $\widehat{\overline{F}}_C(t)$ for the reliability function under the KG model as follows :

$$\widehat{\overline{F}}_C(t) = \Big(\overline{F}_n(t)\Big)^{\alpha_n}. \tag{6}$$

where $\overline{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(Z_i > t)$ and $\alpha_n = \frac{1}{n}\sum_{i=1}^{n}\delta_i$.

They computed the small sample mean squared errors and the large sample simultaneous confidence bands. Under the KG model, the nonparametric estimation of the reliability function has been considered by Cheng and Lin(1987), Hollander and Pena(1989).

3. MLE of Reliability Function under KG model

In order to compare the nonparametric reliability estimators with the MLE of the reliability function, we consider the Weibull model as follows :

$$\overline{F}(x) = e^{-\gamma x^\alpha} \tag{7}$$

and

$$\overline{G}(y) = e^{-\gamma\beta x^\alpha} \tag{8}$$

Then the likelihood function of the sample $(z_i, \delta_i),\ i = 1, \cdots, n$ is

$$L(\gamma, \alpha, \beta) = \prod_u \gamma\alpha z_i^{\alpha-1} e^{-\gamma(1+\beta)z_i^\alpha} \prod_c \gamma\alpha\beta z_i^{\alpha-1} e^{-\gamma(1+\beta)z_i^\alpha} \tag{9}$$

where $\prod_u(\prod_c)$ denotes a product over the uncensored(censored) observations. Finding MLE $(\widehat{\gamma}, \widehat{\alpha}, \widehat{\beta})$ is equivalent to finding the solution to the likelihood equations

$$\begin{aligned}
\frac{\partial}{\partial\gamma}\log L(\gamma, \alpha, \beta) &= \frac{n}{\gamma} - (1+\beta)\sum_{i=1}^n z_i^\alpha = 0 \\
\frac{\partial}{\partial\alpha}\log L(\gamma, \alpha, \beta) &= \frac{n}{\alpha} + \sum_{i=1}^n \log z_i - \gamma(1+\beta)\sum_{i=1}^n z_i^\alpha \log z_i = 0 \\
\frac{\partial}{\partial\beta}\log L(\gamma, \alpha, \beta) &= \frac{n_c}{\beta} - \gamma\sum_{i=1}^n z_i^\alpha = 0
\end{aligned} \tag{10}$$

where n_c is the number of censored observations. The solution is obtained by the Newton-Raphson method. The method requires starting values $\widehat{\gamma}_0, \widehat{\alpha}_0$, and $\widehat{\beta}_0$ and the sample information matrix. To get reasonable starting values $\widehat{\gamma}_0, \widehat{\alpha}_0$, observe that

$$\log(-\log\overline{F}(z_i)) = \log\gamma + \alpha\log z_i,$$

so if we use estimates $\widehat{\overline{F}}(z_i)$, we could regress $\log(-\log\widehat{\overline{F}}(z_i))$ against $\log z_i$,and then let the regression coefficient be $\widehat{\alpha}_0$ and the constant be $\log\widehat{\gamma}_0$. Also, to get starting value $\widehat{\beta}_0$, we use the relationship

$$\text{censoring rate} = \frac{\beta}{1+\beta}.$$

Natually, we estimate censoring rate with n_c/n, then let $n_c/(n-n_c)$ be $\widehat{\beta}_0$. The sample information matrix at $(\widehat{\gamma}_0, \widehat{\alpha}_0, \widehat{\beta}_0)$ is

$$\begin{pmatrix}
\frac{n}{\widehat{\gamma}_0^2} & (1+\widehat{\beta}_0)\sum_{i=1}^n z_i^{\widehat{\alpha}_0}\log z_i & \sum_{i=1}^n z_i^{\widehat{\alpha}_0} \\
(1+\widehat{\beta}_0)\sum_{i=1}^n z_i^{\widehat{\alpha}_0}\log z_i & \frac{n}{\widehat{\alpha}_0} + \widehat{\gamma}_0(1+\widehat{\beta}_0)\sum_{i=1}^n z_i^{\widehat{\alpha}_0}(\log z_i)^2 & \widehat{\gamma}_0\sum_{i=1}^n z_i^{\widehat{\alpha}_0}\log z_i \\
\sum_{i=1}^n z_i^{\widehat{\alpha}_0} & \widehat{\gamma}_0\sum_{i=1}^n z_i^{\widehat{\alpha}_0}\log z_i & \frac{n_c}{\widehat{\beta}_0^2}
\end{pmatrix} \tag{11}$$

Then we have the MLE $\widehat{\overline{F}}_{MLE}(t)$ of the reliability function as follows :

$$\widehat{\overline{F}}_{MLE}(t) = \exp(-\widehat{\gamma}t^{\widehat{\alpha}}). \tag{12}$$

Also $\widehat{\overline{F}}_{MLE}(t)$ is a function of $(\widehat{\gamma}, \widehat{\alpha})$, so we can get an approximate distribution of $\widehat{\overline{F}}_{MLE}(t)$ by using the delta method.

4. Numerical Comparison and Conclusions

Monte Carlo simulation is performed for the following reliability distributions: (i) Weibull with parameters $\gamma = 1$ and $\alpha = 1$ in equation(7) (ii) Weibull with parameters $\lambda = 1$ and $\alpha = 0.5$ (iii) Weibull with parameters $\lambda = 1$ and $\alpha = 2$. There were chosen to represent hazard rates that are constant, increasing, and decreasing, respectively. Furthermore, we investigate the effects of varing the censoring rates(30sample sizes ($n = 30, 50, 100$). The simulation procedure is repeated 10000 times in order to get the mean squared error of the $\widehat{\overline{F}}_{KM}(t), \widehat{\overline{F}}_N(t), \widehat{\overline{F}}_E(t), \widehat{\overline{F}}_C(t)$, and $\widehat{\overline{F}}_{MLE}(t)$ evaluated at $t \ : F(t) = 0.30, 0.50, 0.70, 0.90, 0.95$. Since simulation results of constant, increasing, and decreasing hazard rates are similar, we report in the table for constant hazard rate.

From the simulation study, we have the following results :

(i) $\widehat{\overline{F}}_E(t)$ and $\widehat{\overline{F}}_C(t)$ are nearly always better than$\widehat{\overline{F}}_{KM}(t)$ and $\widehat{\overline{F}}_N(t)$ regardless of β in the sense of mean squared error criterion.

(ii) $\widehat{\overline{F}}_N(t)$ is slightly better than $\widehat{\overline{F}}_{KM}(t)$ under the KM model .

(iii) As expected, $\widehat{\overline{F}}_{MLE}(t)$ is better than all estimators regardless of censoring rates. For instance, when the reliability is 0.9 and the sample size is moderate, MSE of $\widehat{\overline{F}}_{KM}(t)$ and $\widehat{\overline{F}}_N(t)$ is about three times MSE of $\widehat{\overline{F}}_{MLE}(t)$. But MSE of $\widehat{\overline{F}}_C(t)$ and $\widehat{\overline{F}}_E(t)$ is nearly two times MSE of $\widehat{\overline{F}}_{MLE}(t)$. Therefore we prefer $\widehat{\overline{F}}_C(t)$ and $\widehat{\overline{F}}_E(t)$ to $\widehat{\overline{F}}_{KM}(t)$ and $\widehat{\overline{F}}_N(t)$ under the KG model.

(iv) $\widehat{\overline{F}}_C(t)$ is slightly better than $\widehat{\overline{F}}_E(t)$ in small sample size at the right tail. But $\widehat{\overline{F}}_C(t)$ and $\widehat{\overline{F}}_E(t)$ are similar at large sample size.

Table 1. MSE of the $\widehat{\widehat{F}}_{KM}(t), \widehat{\widehat{F}}_N(t), \widehat{\widehat{F}}_E(t), \widehat{\widehat{F}}_C(t)$ and $\widehat{\widehat{F}}_{MLE}(t)$ for Weibull model with $\gamma = 1 \& \alpha = 1$

β	n	$R(t)$	$\widehat{\widehat{F}}_{KM}(t)$	$\widehat{\widehat{F}}_N(t)$	$\widehat{\widehat{F}}_E(t)$	$\widehat{\widehat{F}}_C(t)$	$\widehat{\widehat{F}}_{MLE}(t)$
0.4286	30	0.30	.0080	.0072	.0075	.0067	.0001
		0.50	.0087	.0081	.0078	.0075	.0004
		0.70	.0074	.0070	.0060	.0059	.0015
		0.90	.0030	.0029	.0025	.0024	.0010
		0.95	.0018	.0017	.0015	.0015	.0004
	50	0.30	.0054	.0050	.0049	.0047	.0001
		0.50	.0052	.0049	.0045	.0044	.0003
		0.70	.0046	.0045	.0035	.0035	.0012
		0.90	.0018	.0018	.0016	.0016	.0008
		0.95	.0010	.0010	.0009	.0009	.0003
	100	0.30	.0026	.0026	.0024	.0023	.0001
		0.50	.0032	.0031	.0027	.0027	.0002
		0.70	.0025	.0024	.0022	.0022	.0008
		0.90	.0010	.0010	.0010	.0010	.0005
		0.95	.0006	.0006	.0005	.0005	.0002
1.0000	30	0.30	.0178	.0140	.0176	.0146	.0001
		0.50	.0119	.0113	.0105	.0097	.0005
		0.70	.0083	.0080	.0067	.0065	.0017
		0.90	.0032	.0030	.0019	.0019	.0010
		0.95	.0017	.0017	.0010	.0010	.0004
	50	0.30	.0108	.0095	.0106	.0087	.0001
		0.50	.0079	.0076	.0071	.0069	.0003
		0.70	.0053	.0052	.0041	.0040	.0012
		0.90	.0018	.0018	.0011	.0011	.0007
		0.95	.0010	.0010	.0006	.0006	.0003
	100	0.30	.0037	.0036	.0032	.0030	.0001
		0.50	.0039	.0039	.0035	.0035	.0002
		0.70	.0028	.0028	.0024	.0024	.0008
		0.90	.0010	.0010	.0007	.0007	.0005
		0.95	.0005	.0005	.0004	.0004	.0002

References

1. Chem, Y., Hollander. M., and Langberg, N. (1982). Small- sample results for the Kaplan-Meier estimator. *Journal of the American Statistical Association.* 77, 141-144.

2. Cheng, P. and Lin, G. (1987). Maximum likelihood estimation of survival

function under the Kozior-Green proportional hazards model. *Statistics and Probability Letters* 5, 75-80.

3. Cheng, P. E. and Chang, Y. C. (1985). PLE versus MLE of Survival Functions under a simple Proportional Hazards Model. *Journal of the Chinese Statistical Association 23, Special Issue*, 57-73.

4. Ebrahimi, N. (1985). Nonparametric Estimation of Survival Functions for Incomplete Observations when the Life Time Distribution is proportionally related to the Censoring Time Distribution, *Communication in Statistics: Theory and Methods* 14, 2887-2898.

5. Hollander, M. and Pena. E. (1989). Families of Confidence bands for the survival function under the general random censorship model and the Koziol-Green model, *Canadian Journal of Statistics* 17, 59-74.

6. Kaplan, E. L. and Meier, P. (1957). Nonparametric Estimation From Incomplete Sample, *Journal of the American Statistical Association.* 53, 457-481.

7. Koziol, J. and Green, S. (1976). A Cramer-von Mises statistic for randomly censored data, *Biometrika* 63, 465-481.

8. Nelson, W. B. (1972). Thory and Applications of Hazard Plotting for Censored Failure Data, *Technometrics* 14, 945-996.