The Korean Communications
in Statistics Vol. 3, No. 3, 1996
pp. 187-194

Nonparametric Estimation of Reliability in Strength-Stress Model

H. S. Jeong[1],
J. J. Kim, B. U. Park[2], H. W. Lee[3]

Abstract

We treat the problem of estimating reliability $R = P[Y < X]$ in the stress-strength model in which a unit of strength X is subjected to environmental stress Y. In this paper several nonparametric approaches to estimation of R are analyzed and compared by simulations.

1. Introduction

Let X and Y be independent random variables with cumulative distribution functions (c.d.f.'s) $F(\cdot)$ and $G(\cdot)$, respectively. We are interested in estimating functionals of the form $R = P[Y < X]$, given samples $X_1, X_2, \cdots, X_m$ of size m from X and $Y_1, Y_2, \cdots, Y_n$ of size n from Y.

Previous work in this field is mostly occupied by the parametric approaches for continuous data. The most common parametric model has both X and Y normally distributed. Church and Harris (1970), Downton (1973) and Reiser and Guttman (1986) discuss various estimation strategies in this case. Other parametric models have also been investigated; see, e.q., Sathe and Shah (1981),Tse and Karson (1986) and Awad and Charraf (1986).

Little work has been done on nonparametric estimation of R since Birnbaum (1956). He proposed the point estimate

$$\begin{aligned} \hat{R} &= \int G_n(x)\,dF_m(x) \\ &= \frac{1}{m}\sum_{i=1}^{m} G_n(X_i) \qquad (1.1) \\ &= \frac{1}{mn}\{\text{number of } (i,j) \text{ pairs such that } Y_j \le X_i\} \end{aligned}$$

1) Professor, Applied Statistics, Seowon University, Chongju, 360-742, Korea.
2) Professor, Department of Computer Science and Statistics, Seoul National University, Seoul, 151-742, Korea.
3) Department of Computer Science and Statistics, Seoul National University, Seoul, 151-742, Korea.

$F_m(\cdot)$ and $G_n(\cdot)$ are the empirical c.d.f.'s of the X's and Y's, respectively. Note that the numerator in the last expression is the Wilcoxon-Mann-Whitney statistic. Under the assumption that the underlying c.d.f.'s $F(\cdot)$ and $G(\cdot)$ are continuous, $\hat{R}_{WMW}$ is the uniform minimum variance unbiased estimator of R.

In Section 2 three new estimators are proposed. In Section 3 consistency of the estimators are discussed. In section 4 the three estimators and $\hat{R}_{WMW}$ are compared by a simulation study.

2. Estimators of R

Replacing G_n in (1.1) by other estimators of G, we propose the following three estimators of R.

2.1 Case of known stress distribution G

In some applications, the stress distribution may be known to the investigator. The following is an example of situations in which the stress distribution may be known. Consider the case of telephone poles or support towers for power-transmission cables. In this case, usually the wind loadings are so well known that the stress distribution can be derived almost exactly.

Assume the parametric model $Y \sim G(\cdot \mid \theta)$ for stress. Instead of G_n in (1.1), we use $G(\cdot \mid \hat{\theta})$ by calculating the maximum likelihood estimator (M.L.E.) $\hat{\theta}$ of θ. For example, when $Y \sim N(\mu, \sigma^2)$, let $G(\cdot \mid \hat{\theta}) = \Phi((\cdot - \hat{\mu})/\hat{\sigma})$ where $\hat{\mu}$ and $\hat{\sigma}$ are M.L.E. of μ and σ, respectively. Then

$$\hat{R}_{MLE} = \frac{1}{m} \sum_{i=1}^{m} G(X_i \mid \hat{\theta})$$

2.2 Kernel type estimator

An approach that does not depend at all on a parametric assumption would be to estimate the density $g(y)$ (probability density function of Y) using a nonparametric density estimation procedure. In this view, we replace $G_n(x)$ in (1.1) by $\int_{-\infty}^{x} \hat{g}(u)\,du$ where $\hat{g}(u)$ is kernel density estimator. The bandwidth is selected by the biased cross-validation method which is proposed by Scott and Terrell (1987). Then

$$\hat{R}_{KER} = \frac{1}{m} \sum_{i=1}^{m} \int_{-\infty}^{X_i} \hat{g}(u)\,du .$$

2.3 Estimator using median rank

Let $Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(n)}$ denote the order statistics from a random sample of G. Let $G_n^*(y) = (\sum_{i=1}^{n} I(Y_j \le y) - 0.3)/(n+0.4)$. Replacing G_n in (1.1) by G_n^*, we can obtain another estimator of R defined by

$$\hat{R}_{MED} = \frac{1}{m} \sum_{i=1}^{m} G_n^*(X_i) .$$

3. Consistency of the proposed estimators.

In this section, we discuss the consistency of the estimators proposed in section 2.

3.1 The maximum likelihood estimator

Under some regularity conditions on G (see Lehmann (1983), P409, for example), the MLE $\hat{\theta}$ is a consistent estimator of θ. Suppose, in addition, that $\sup_{|\theta-\theta'|\le c} |\frac{\partial}{\partial\theta} G(y|\theta')| \le M(y)$ for all y in the support of $G(\cdot|\theta)$ with $E(M(Y)) < \infty$ and $c > 0$. By Taylor expansion, we have

$$G(y \mid \hat{\theta}) = G(y \mid \theta) + (\hat{\theta} - \theta)^T \cdot \frac{\partial}{\partial \theta} G(y \mid \theta^*)$$

where θ^* is a point on the line segment joining $\hat{\theta}$ and θ.

Taking the sum of above equations for $i = 1, 2, \cdots, m$ and dividing by m, we see that

$$\frac{1}{m} \sum_{i=1}^{m} G(X_i \mid \hat{\theta}) = \frac{1}{m} \sum_{i=1}^{m} G(X_i \mid \theta) + (\hat{\theta} - \theta)^T \cdot \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta} G(X_i \mid \theta^*)$$

Thus

$$\begin{aligned} \left| \frac{1}{m} \sum_{i=1}^{m} G(X_i \mid \hat{\theta}) - \frac{1}{m} \sum_{i=1}^{m} G(X_i \mid \theta) \right| &\le \left| (\hat{\theta} - \theta)^T \cdot \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta} G(X_i \mid \theta^*) \right| \\ &\le |\hat{\theta} - \theta| \frac{1}{m} \sum_{i=1}^{m} M(X_i) \\ &\to 0 , \quad \text{in probability.} \end{aligned}$$

Hence

$$\frac{1}{m} \sum_{i=1}^{m} G(X_i \mid \hat{\theta}) \to E[G(X_1 \mid \theta)] = P[Y < X]$$

in probability. i.e. $\hat{R}_{MLE}$ is a consistent estimator for R.

3.2 The kernel estimation

We can rewrite $\hat{R}_{KER}$ as

$$\hat{R}_{KER} = \frac{1}{mnh} \sum_{i=1}^{m} \sum_{j=1}^{n} \int_{-\infty}^{X_i} k(\frac{z-Y_j}{h})\ dz\ ,$$

where $k(\cdot)$ is a kernel function and h is a bandwidth. Then

$$\begin{aligned} E[\ \hat{R}_{KER}\ ] &= \frac{1}{h} E \int_{-\infty}^{X_1} k(\frac{z-Y_1}{h})\ dz \\ &= \frac{1}{h} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{x} k(\frac{z-y}{h})\ dz\ f(x) g(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(\frac{x-y}{h})\ f(x) g(y)\, dx\, dy \\ &\rightarrow \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x-y)\ f(x) g(y)\, dx\, dy \quad \text{as} \quad h \rightarrow 0 \\ &= P[\,Y < X\,] \end{aligned}$$

where $I(x) = 1$ if $x > 0$ and otherwise 0, and $K(t) = \int_{-\infty}^{t} k(u)\, du$ i.e. the cumulative distribution function of $k(\cdot)$. So $\hat{R}_{KER}$ is asymptotically unbiased.

Let $\eta_h(X_i\ , Y_j) = \frac{1}{h} \int_{-\infty}^{X_i} k(\frac{z-Y_j}{h})\ dz$, then

$$\begin{aligned} Var(\ \hat{R}_{KER}) &= \frac{1}{m^2 n^2} \Big[\sum_{i=1}^{m} \sum_{j=1}^{n} Var[\eta_h(X_i, Y_j)] \\ &+ \sum_{i=1}^{m} \sum_{k \neq i}^{m} \sum_{j=1}^{n} Cov[\ \eta_h(X_i, Y_j)\ ,\ \eta_h(X_k, Y_j)] \\ &+ \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{l \neq j}^{n} Cov[\ \eta_h(X_i, Y_j), \eta_h(X_i, Y_l)] \Big] \\ &= \frac{1}{m^2 n^2} [\ mn \cdot Var[\eta_h(X_1\ , Y_1)] \\ &+ mn(m-1) \cdot Cov[\ \eta_h(X_1\ , Y_1)\ ,\ \eta_h(X_2\ , Y_1)] \\ &+ mn(n-1) \cdot Cov[\ \eta_h(X_1\ , Y_1)\ ,\ \eta_h(X_1\ , Y_2)]\ ] \end{aligned}$$

Now by the similar argument leading to the asymptotic unbisedness, we have

$$Var\,[\eta_h(X_1, Y_1)] \rightarrow P[Y<X]\{1-P[Y<X]\},$$

$$Cov[\eta_h(X_1, Y_1), \eta_h(X_2, Y_1)] \rightarrow P[Y<\min(X_1, X_2)] - \{P[Y<X]\}^2$$

and

$$Cov\,[\eta_h(X_1, Y_1), \eta_h(X_1, Y_2)] \rightarrow P[\max(Y_1, Y_2)<X] - \{P[Y<X]\}^2$$

as $h \rightarrow 0$.

This show that $\hat{R}_{KER}$ is a consistent estimator for R.

3.3 A estimator using median rank.

Considering the consitency of $\hat{R}_{WMW}$, we can easily prove that $\hat{R}_{MED}$ is consistent for R. Rewrite $\hat{R}_{MED}$ as

$$\begin{aligned}
\hat{R}_{MED} &= \frac{1}{m}\sum_{i=1}^{m}\frac{\sum_{j=1}^{n} I(Y_j \le X_i) - 0.3}{n+0.4} \\
&= \frac{1}{m(n+0.4)}\sum_{i=1}^{m}\sum_{j=1}^{n} I(Y_j \le X_i) - \frac{0.3}{n+0.4} \\
&= \frac{n}{(n+0.4)}\cdot\frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n} I(Y_j \le X_i) - \frac{0.3}{n+0.4} \\
&= \frac{n}{(n+0.4)}\cdot\hat{R}_{WMW} - \frac{0.3}{n+0.4} \\
&\rightarrow P[Y<X]
\end{aligned}$$

in probability. i.e. $\hat{R}_{MED}$ is a consistent estimator for R.

4. Simulation results.

To compare the accuracy of the estimators of R, computer simulations are performed. A FORTRAN program run on Pentium-100 computer system is used to evaluate the Mean Sqare Error (M.S.E.) of $\hat{R}_{WMW}$, $\hat{R}_{MLE}$, $\hat{R}_{KER}$ and $\hat{R}_{MED}$. The standard package IMSL " International Mathematical and Statistical Library" is used. Stress random variables $Y's$ are generated from the Weibull distribution with a fixed scale parameter $\alpha=1$ and shape parameter β=0.5, 1.0, 3.0 and 3.5, respectively. Simulations are done with 250 replications for selected values of sample size $(m, n)=(10,20),(20,10),(20,20),(20,30)$ and (30,20).

In order to vary the value of $R(0.5 \leq R < 1.0)$ strength random variable X's are with a scale parameter α varing from 1.0 to 2.5 and the same shape parameter with stress random variable Y. The MSE's of the four estimators with the fixed sample size (m, n)=(20,30) when β=0.5, 1.0, 3.0 and 3.5 are given Figure 1,2,3 and 4, respectively.

From these figures, five important points emerge. First, the four graphs are similar to the trends. Second, the M.S.E. decreases as the exact R increases except $\hat{R}_{KER}$ in β =0.5 case. Third, $\hat{R}_{MED}$ is the worst in all cases. Fourthly, for the interval where $0.5 \leq R < 0.8$, $\hat{R}_{KER}$ is always the best. Finally, M.S.E.'s of $\hat{R}_{WMW}$, $\hat{R}_{MLE}$ and $\hat{R}_{KER}$ are nearly equal for the large value of R.

Figure 5, 6, 7 and 8 show the results of the simulation with the various sample size (m, n)=(10,20), (20,10), (20,20) and (30,20) when fixed β=3.5. The four graph are similar to the trends but the MLE's decreses as the sample sizes increse.

We perform the simulation with the same sample size as figure 5 ~8 when β=0.5, 1.0 and 3.0. These simulation results, which are omitted, are simular to the above ones.

References

[1] Awad, M. A. and Charraf, K. C. (1986), Estimation of Pr(Y < X) in the Burr Case: A Comparative Study, *Communications in Statistics-Simulation and Computation* B15(2), 389-403.

[2] Birnbaum, Z. W. (1956), On a Use of the Mann-Whitney Statistic, *in the Proceedings of the Third Berkeley Symposium on Mathematical Stasistics and Probability* (Vol. 1), Berkeley: University of California Press, 13-17.

[3] Church, J. D. and Harris, B. (1970), The Estimation of Reliability from Stress-Strength Relationships, *Technometrics* 12, 49-54.

[4] Downton, F. (1973), The Estimation of the Pr(Y < X) in the Normal Case, *Technometrics* 15, 551-558.

[5] Reiser, B. and Guttman, I. (1986), Statistical Inference for Pr(Y < X) , *Technometrics* 28, 253-257.

[6] Lehmann E. L. (1983), *Theory of Point Estimation*, John Wiley & Suns, New York.

[7] Sathe, Y. S. and Shah, S. P. (1981), On Estimating Pr(Y < X) for the Exponential Distribution, *Communications in Statistics-Theory and Methods* A10, 39-47.

[8] Scott, D. W. and Terrell, G. R. (1987), Biased and Unbiased Cross-Validation in Density Estimation, *Journal of the American Statistical Association* 82, 1131-1146.

[9] Tse, S. K. and Karson, M. (1986), Estimation of Pr(Y < X) in the Gamma Case, *Communications in Statistics-Simulation and Computation* B15(2), 365-388.

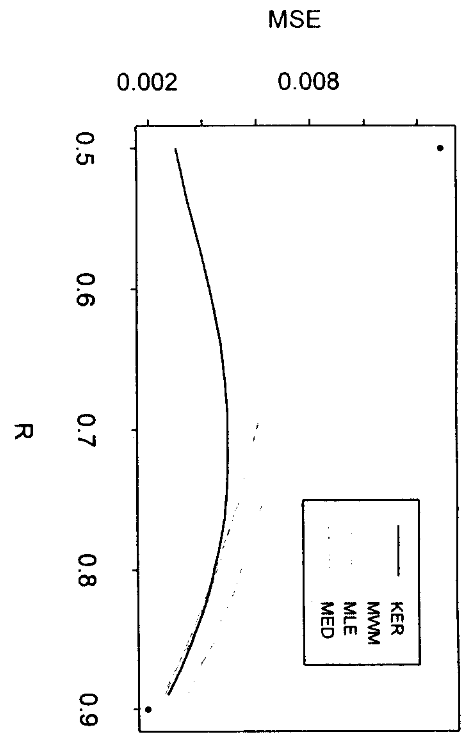

Figure 1. Weibull Case (Beta=0.5,m=20,n=30)

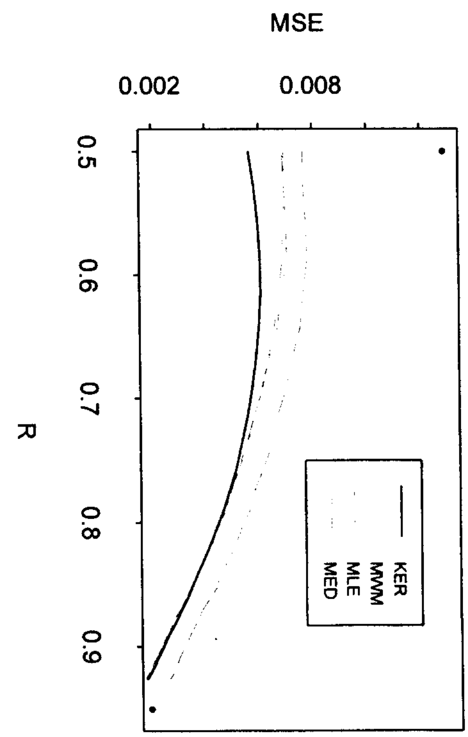

Figure 2. Weibull Case (Beta=1.0,m=20,n=30)

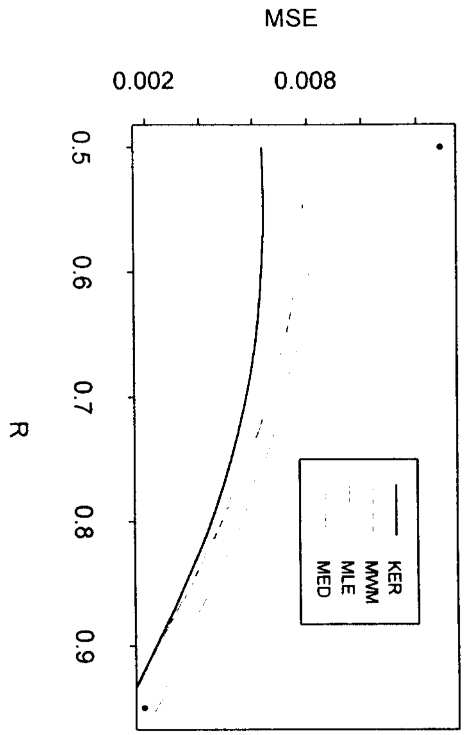

Figure 3. Weibull Case (Beta=3.0,m=20,n=30)

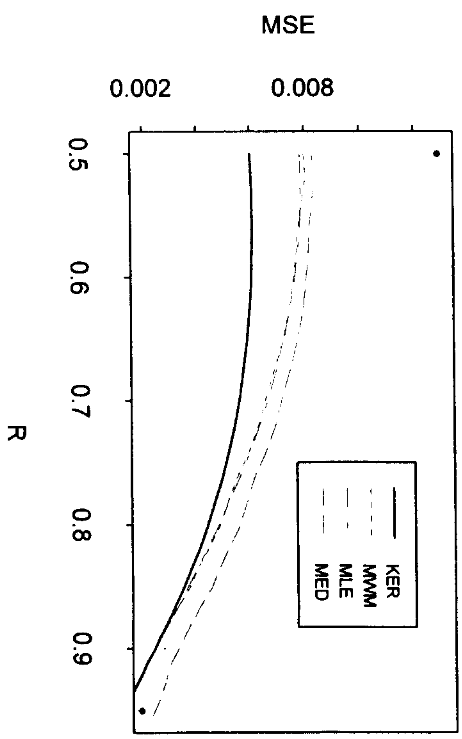

Figure 4. Weibull Case (Beta=3.5,m=20,n=30)

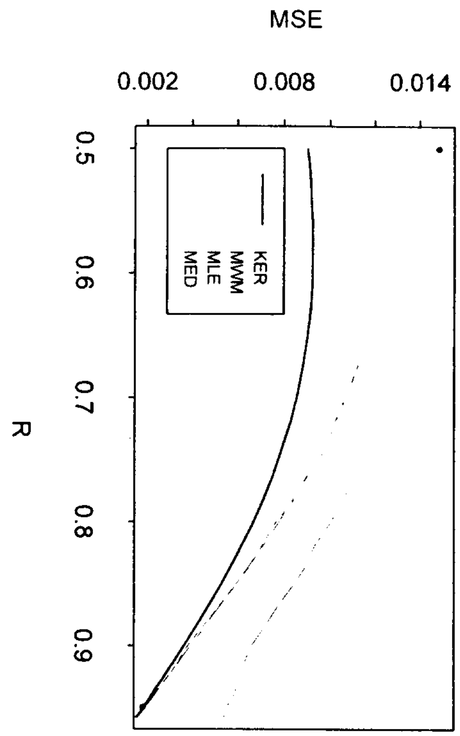

Figure 5. Weibull Case (Beta=3.5,m=10,n=20)

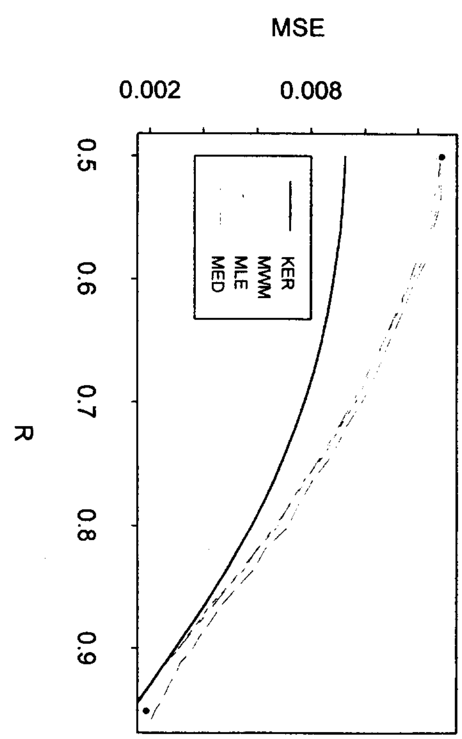

Figure 6. Weibull Case (Beta=3.5,m=20,n=10)

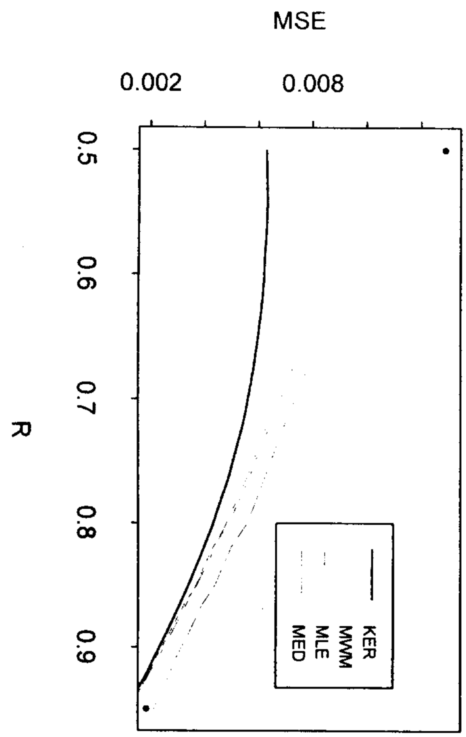

Figure 7. Weibull Case (Beta=3.5,m=20,n=20)

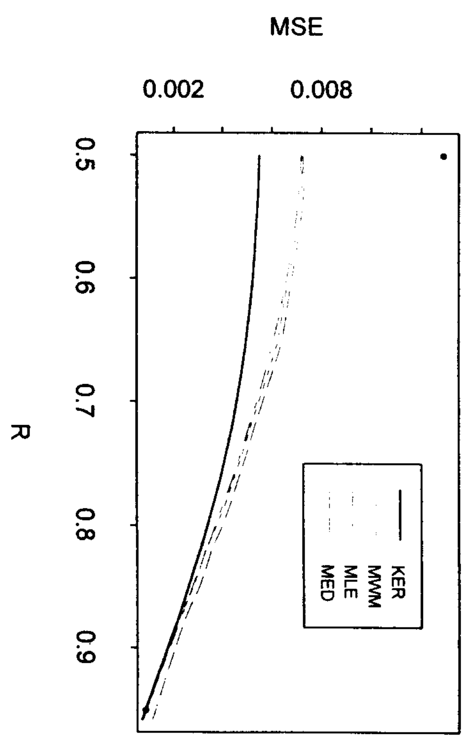

Figure 8. Weibull Case (Beta=3.5,m=30,n=20)