<**Research Paper**>

Journal of the Korean Statistical Society Vol. 19, No. 1, 1990

A Nonparametric Small Sample Estimator of Mean Residual Life

Farrokh Choobineh* and Dong-Ho Park**

ABSTRACT

In reliability and life testing the mean residual life(MRL) of an item plays a significant role. While there has been a great deal of discussion on the theoretical aspects of the MRL, good estimators of MRL have been difficult to obtain. In this paper we propose a new estimator of the MRL of items at a given age, which is especially good for a small sample. The new estimator compares favorably with the empirical MRL estimator for small samples.

1. Introduction

The mean residual life(MRL) at age $t(t \geq 0)$ is the expected remaining life of an item, given that the item is of age t. Let F be a life distribution function(i.e., $F(x)=0$ for $x \leq 0$) with a finite first moment, and let X be a continuous random variable with distribution F. The MRL at age t is defined as

$$m(t)=\begin{cases} E(X-t \mid X>t) & \text{if } \bar{F}(t)>0 \\ 0 & \text{if } \bar{F}(t)=0 \end{cases} \tag{1.1}$$

for $t \geq 0$, where $\bar{F}(t)=1-F(t)$. For $t=0$, m(t) is the usual mean. Note that we can express

* Industrial and Management Systems Engineering Department, University of Nebraska-Lincoln, Lincoln, Nebraska 68588, USA

** Department of Mathematics and Statistics, University of Nebraska-Lincoln, Linclon, Nebraska 68588, USA

$m(t)=\int_t^\infty \bar{F}(x)\,dx/\bar{F}(t)$ for $\bar{F}(t)>0$.

The MRL has been discussed extensively in the literature in the last two decades. Not only does the MRL have a wide range of applications among reliabilitists and engineers but also the knowledge of the MRL completely determines the life distribution via an inversion formula. The MRL is also used to define various nonparametric classes of life distributions such as decreasing mean residual life, increasing mean residual life, new is better than used in expectation, and new is worse than used in expectation. A number of nonparametric tests have been proposed to test the hypotheses concerning these classes of life distributions. Guess and Proschan(1988) provides an excellent discussion and references on theory and applications of the MRL.

While there has been a great deal of discussion on the theoretical aspects of the MRL, a good estimator of the MRL has been difficult to obtain except the empirical MRL estimator. The empirical MRL estimator, $\hat{m}_e(t)$, is obtained by replacing $\bar{F}$ of equation (1.1) with the empirical distribution as shown by equation (1.2).

$$\hat{m}_e(t)=\begin{cases} \dfrac{\sum_{i=k+1}^{n}(X_{(i)}-t)}{n-k} & \text{if } X_{(k)}\le t<X_{(k+1)} \\ 0 & \text{if } t\ge X_{(n)} \end{cases} \tag{1.2}$$

for $k=0, 1,\cdots, n\text{-}1$; $X_{(0)}=0$, where $X_{(1)}<\cdots<X_{(n)}$ is the corresponding order statistics of a random sample $X_1,\cdots,X_n$ from F.

Yang(1978) proved that the empirical MRL estimator is strongly uniformly consistent and weakly convergent to a Gaussian process on a finite interval, and Hall and Wellner(1979) generalized Yang's results and proposed nonparametric confidence bounds for the MRL at age t based on the large sample theory. Lawrence(1966) derived a lower bound of the MRL at age t, assuming that the distribution has decreasing failure rate and the mean and a percentile are known.

The empirical MRL estimator only benefits from the information contained in the observations exceeding t. The only contribution by the observations before or at t to the empirical MRL estimator is through the number of observations before or at t. This poses a very serious problem in accuracy of the estimate when the sample size is small, and consequently the number of observations exceeding t is very small.

In practice, it is common to select few items for testing since, in general, the total testing cost is an increasing function of the sample size, and practitioners are inclined to minimize cost. Thus, in the presence of a pressing need to estimate the MRL using a small sample size, it is prudent to use the total information contained in a small sample. More discussions on such situations are given in Section 4.

This paper proposes a new estimator of the MRL at age t that benefits from the information contained in the entire sample. Our approach to estimate the MRL is to represent the MRL in terms of a partial moment, and then use an approximation of the partial moment.

The organization of the paper is as follows. In Section 2 partial moments and their approximations are presented, and the new estimator of the MRL is proposed. In Section 3 the new estimator is compared numerically with the empirical MRL estimator. Concluding remarks and some situations in which our estimator is useful are presented in Section 4.

2. Estimation of Mean Residual Life

The kth partial moment of a random variable X about the origin is the partial expectation of X^k over $(-\infty, t)$ for fixed t, that is,

$$E_{-\infty}^{t}(X^k)=\int_{-\infty}^{t} x^k \, dF(x)$$

if the integral exists. An approximation for $E_{-\infty}^{t}(X^k)$ can be obtained by solving the following optimization model: $\max_y Py^k$ such that $\mu_q=Py^q+(1-P)z^q$, $q=1,\cdots,Q$, $y<z$, and $P=\Pr\{X\leq t\}$, where μ_q is the qth moment of X with respect to the origin, and masses P and (1-P) are placed at points y and z, respectively. The larger the value of Q, the closer the optimum value is to $E_{-\infty}^{t}(X^k)$. Also the approximation performs better when $k\leq Q$. Choobineh and Branting(1986) considered the similar method to approximate the semivariance.

Let the approximate value of $E_{-\infty}^{t}(X^k)$ be denoted by $APL_t^k=\max_y Py^k$. For our purpose of estimating the MRL, we consider the case when $k=1$ and $Q=2$(i.e. matching only the first two moments). Then the problem is reduced to determine y so that

$$\begin{aligned}\mu&=Py+(1-P)z,\\ \mu^2+\sigma^2&=Py^2+(1-P)z^2\end{aligned} \tag{2.1}$$

and

$$y<z.$$

Solving (2.1), a unique solution $y=\mu-\left(\frac{1-P}{P}\right)^{1/2}\sigma$ is obtained and thus we have:

$$APL_t^1 = P\left(\mu-\left(\frac{1-P}{P}\right)^{1/2}\sigma\right) \tag{2.2}$$

where μ and σ^2 are the mean and variance of X, respectively. We utilize APL_t^1 to derive the estimator of the MRL at age t which relies on all observations $X_1,\cdots,X_n$.

The MRL at age t, defined in (1.1), can be rewritten as

$$m(t)=\frac{\int_t^\infty x\, dF(x)}{\bar{F}(t)}-t=\frac{\mu-\int_0^t x\, dF(x)}{\bar{F}(t)}-t=\frac{\mu-E_0^t(X)}{1-P}-t$$

Replacing $E_0^t(X)$ by APL_t^1 of (2.2), we obtain

$$m_p(t)=\mu+\left(\frac{P}{1-P}\right)^{1/2}\sigma-t \tag{2.3}$$

which is an approximation to m(t).

Let $\bar{X}=\frac{1}{n}\Sigma X_i$ and $S^2=\frac{1}{n}\Sigma(X_i-\bar{X})^2$ be the sample mean and variance, respectively and let $\hat{P}=\frac{1}{n}\Sigma I(X_i\le t)$ be the proportion of observations that are less than or equal to t, where $I(\cdot)$ is an indicator function. A new nonparametric estimator of m(t) is obtained by substituting $\bar{X}$, S and $\hat{P}$, respectively, for μ, σ and P in the expression of $m_p(t)$ in (2.3) as follows:

$$\hat{m}_p(t)=\begin{cases}\bar{X}+\left(\frac{\hat{P}}{1-\hat{P}}\right)^{1/2}S-t & \text{if } t<X_{(n)}\\ 0 & \text{if } t\ge X_{(n)}\end{cases} \tag{2.4}$$

where $X_{(n)}=\max_{1\le j\le n} X_j$. In the rest of this paper $\hat{m}_p(t)$ is referred to as the partial moment (PM) estimator. Although the PM estimator is not a consistent estimator of m(t) unless m(t) is equal to $m_p(t)$, it fully utilizes the entire sample through the statistics $\bar{X}$ and S^2, whereas the empirical MRL estimator, given in (1.2), utilizes only the observations exceeding t. Thus, the PM estimator is expected to perform better than the empirical MRL estimator when the sample size is small.

3. A Numerical Comparison

A Monte Carlo experiment was conducted to investigate and compare the behavior of the proposed PM estimator and the empirical MRL estimator. The underlying life distribution was assumed to follow a gamma distribution with the scale parameter $\alpha=1$ and the shape parameter λ, i.e. $G(\alpha, \lambda)$. Five different values of the shape parameter λ were used namely, $\lambda=1$, 3, 5, 7, and 10. Since the PM estimator is more suitable for small samples, the sample size n was set to be 3, 4, 5, 6, and 10. Five values for t were used. These values are the mean, μ, and plus and minus 10% and 20% of μ(i.e. $t=(1\pm\beta)\mu$, $\beta=0$, 0.1 and 0.2).

Tables 1 and 2 respectively show the results of the Monte Carlo experiment for the empirical MRL estimator and the PM estimator along with the true values of the MRL. The values of estimators in the tables are the average of 1,000 replications for a given combination of λ, t and n.

Comparing the values of the estimators with the true values the following general observations can be made. First, the PM estimators are closer to the true value of MRL than the empirical MRL estimator for $n\leq5$. Second, the PM estimator overestimates the true value of MRL for i) $n>5$, ii) $t<\mu$, and iii) increasing values of λ. Third, the PM estimators are greater than the empirical MRL estimators in all cases. Fourth, although the biases of both estimators increase in general as t increases, the corresponding increase in the bias is smaller for the PM estimator than for the empirical MRL estimator. However, since the empirical MRL estimator is a consistent estimator, it is better than the PM estimator for moderate or large sample sizes.

4. Concluding Remarks

As Tables 1 and 2 indicate, the bias of the PM estimator is consistently smaller than that of the empirical MRL estimator for small samples such as $n=3$ and $n=4$. Although the empirical MRL estimator is better for large samples, the PM estimator provides more accurate approximation to the MRL at a fixed age in such situations where only small samples are available.

Table 1. Numerical Values of the Empirical MRL Estimators, $\hat{m}_e(t)$, for Various n and t when the Underlying Distribution is G(α, λ). m(t) Denotes the True Value.

Dist	t	m(t)	n=3	n=4	n=5	n=6	n=10
	0.80	1.0000	0.8270	0.8902	0.9531	0.9652	0.9949
	0.90	1.0000	0.7884	0.8646	0.9344	0.9572	0.9968
G(1,1)	1.00	1.0000	0.7484	0.8291	0.9081	0.9358	0.9868
	1.10	1.0000	0.7063	0.7913	0.8769	0.9093	0.9813
	1.20	1.0000	0.6728	0.7602	0.8512	0.8888	0.9708
	2.40	1.7006	1.5180	1.6114	1.6794	1.6754	1.7150
	2.70	1.6398	1.3532	1.4689	1.5616	1.5758	1.6326
G(1,3)	3.00	1.5882	1.2218	1.3663	1.4717	1.4955	1.5945
	3.30	1.5438	1.0706	1.2331	1.3618	1.4122	1.5347
	3.60	1.5054	0.9336	1.0980	1.2354	1.3089	1.4700
	4.00	2.2427	2.1545	2.1924	2.2334	2.2298	2.2414
	4.50	2.1052	1.9274	1.9886	2.0414	2.0557	2.0955
G(1,5)	5.00	1.9917	1.6788	1.7743	1.8619	1.8877	1.9703
	5.50	1.8970	1.4195	1.5400	1.6622	1.7070	1.8422
	6.00	1.8173	1.1773	1.3044	1.4342	1.5033	1.7053
	5.60	2.7234	2.5701	2.6509	2.6707	2.6839	2.7368
	6.30	2.4996	2.2186	2.3588	2.4151	2.4461	2.5076
G(1,7)	7.00	2.3193	1.8495	2.0508	2.1464	2.1870	2.3375
	7.70	2.1725	1.5039	1.7389	1.8821	1.9574	2.1596
	8.40	2.0518	1.1814	1.4437	1.6238	1.7145	1.9892
	9.00	3.0187	2.8515	2.9799	2.9984	3.0361	2.9985
G(1,10)	10.00	2.7320	2.3136	2.4885	2.6162	2.6750	2.7056
	11.00	2.5058	1.8527	2.0788	2.2409	2.3510	2.4783
	12.00	2.3251	1.4067	1.6424	1.8487	1.9935	2.2647

Table 2 also shows that in most cases the PM estimator provides an optimistic estimate of MRL and may possibly be viewed as a tight upper bound for the MRL at a fixed age.

The need to restrict the sample size to a small number (e.g. 4) is common in practice. The impetus to require small samples is economically motivated since decision makers want to minimize experimentation costs. The experimentation costs generally consist of the cost of unit(or system) being tested and the testing costs. The cost of a unit becomes a significant part of the experimentation cost when the unit has a high production cost and

Table 2. Numerical Values of the PM Estimator, $\hat{m}_p(t)$, for Various n and t when the Underlying Distribution is G(α, λ). m(t) Denotes the True Value.

Dist	t	m(t)	n=3	n=4	n=5	n=6	n=10
	0.80	1.0000	0.8723	0.9754	1.0820	1.1093	1.2003
	0.90	1.0000	0.8320	0.9438	1.0500	1.0973	1.2004
G(1,1)	1.00	1.0000	0.7900	0.9080	1.0205	1.0743	1.1910
	1.10	1.0000	0.7440	0.8630	0.9826	1.0429	1.1891
	1.20	1.0000	0.7086	0.8293	0.9533	1.0227	1.1832
	2.40	1.7006	1.5934	1.7571	1.8749	1.9056	2.0182
	2.70	1.6398	1.4289	1.6151	1.7665	1.8245	1.9500
G(1,3)	3.00	1.5882	1.2984	1.5169	1.6853	1.7521	1.9325
	3.30	1.5438	1.1444	1.3898	1.5847	1.6822	1.9087
	3.60	1.5054	1.0050	1.2554	1.4658	1.5935	1.9002
	4.00	2.2427	2.2585	2.3799	2.4739	2.5064	2.5981
	4.50	2.1052	2.0354	2.1776	2.2896	2.3468	2.4682
G(1,5)	5.00	1.9917	1.7843	1.9712	2.1288	2.1978	2.3782
	5.50	1.8970	1.5241	1.7509	1.9563	2.0643	2.3286
	6.00	1.8173	1.2846	1.5276	1.7615	1.9092	2.3139
	5.60	2.7234	2.6942	2.8654	2.9505	3.0018	3.1513
	6.30	2.4996	2.3405	2.5885	2.7194	2.7981	2.9429
G(1,7)	7.00	2.3193	1.9855	2.3108	2.4845	2.5793	2.8362
	7.70	2.1725	1.6367	2.0212	2.2787	2.4350	2.8067
	8.40	2.0518	1.3002	1.7216	2.0370	2.2825	2.8635
	8.00	3.3851	3.4477	3.6145	3.6822	3.7600	3.8442
	9.00	3.0187	2.9998	3.2545	3.3527	3.4466	3.5105
G(1,10)	10.00	2.7320	2.4928	2.8025	3.0374	3.1691	3.3250
	11.00	2.5058	2.0315	2.4428	2.7523	2.9655	3.3074
	12.00	2.3251	1.5720	2.0090	2.4103	2.7074	3.4273

after the experiment either it is destroyed or it is considerably depreciated. The testing cost becomes a significant part of the experimentation cost when the testing procedures are complex and/or the testing duration is long.

For example, in testing the overall performance of tractors, where the unit cost is moderately high and the test duration is long, manager would not be willing to commit more than a few units for testing. Another example is testing a new treatment procedure or drug on, for example, AIDS patients. The unit cost(human life) is very high and consequently the number of patients treated should be small.

References

1. Choobineh, F. and Branting, D.(1986). A Simple Approximation for Semivariance. *European Journal of Operations Research,* Vol. 27, 364-370.
2. Guess, F. and Proschan, F.(1988). Mean Residual Life. Theory and Applications. Handbook of Statistics, Vol. 7, Reliability and Quality Control. P.R. Krishnaiah and C.R. Rao(eds.), 215-224.
3. Hall, W.J. and Wellner, J.A.(1979). Estimation of Mean Residual Life. University of Rochester Department of Statistics Technical Report.
4. Lawrence, M.J.(1966). An Investigation of the Burn-In Problem. *Technometrics,* Vol. 8, 61-71.
5. Yang, G.L.(1978). Estimation of a Biometric Function. *The Annals of Statistics,* Vol. 6, 112-116.