A New Method for Estimating Inverse Data from Destructive Regular Storage Life Test

H. Wang, X.B. Ma

School of Reliability and Systems Engineering
Beihang University
Beijing, China

*Abstract*一**In the destructive regular storage life test of products, there is "inverse" data sometimes, resulting in inaccurate estimates of its reliability index. Based on the theory of isotonic regression and minimum chi-square estimation, we proposed a new method for processing "inverse" data. First, alter the possible "inverse" of original frequency into the frequency meeting sequence constraint by using PAVA algorithm. Then, estimate reliability parameters by means of minimum chi-square estimation. Comparing with traditional methods, we increased the distribution-test of overall failure probability function, the point estimates and confidence intervals of reliability parameters during storage period of products. Finally, example shows that the coefficient of variation obtained by this method are smaller than MLE and the coefficient of variation changes little when sample size changes, reflecting its superiority for estimating small sample data.**

Keywords- inverse data; isotonic regression; minimum chi-square estimation; coefficient of variation

I. INTRODUCTION

Under normal circumstances, the products after produced in the factory, for some reasons need to be stored in the warehouse for some time. Due to the effect of storage conditions, products' reliability indices will decline. To identify the change of the quality of products at different stages of the storage period, we need to test products during storage period and analyze test data. Usually destructive regular storage life test is carried out as follows:

Mark $t = (t_1, t_2, \cdots, t_k)$ as the test time, after a storage time t_i, extract n_i products and we can find X_i products which are failed. Then we obtain the data:

$$(t_i, n_i, X_i), \quad i = 1, 2, \cdots, k \qquad (1)$$

in which, $0 < t_1 < t_2 < \cdots < t_k$. All tests are independent.

In practice, by the time and cost constraints, etc. The sample size n_i is usually small, so the test results will be seriously affected by the randomness of the samples. And sometimes it will come as $X_{i+1} / n_{i+1} \; \square \; X_i / n_i$, which does not meet the characteristic that the probability of failure reduces with time increasing. Such data is always called "inverse" data [8] in engineering, and the estimate result calculated by this would seriously have a great bias.

To solve the problem above and regulate "inverse" data, scholars do a lot of related work, like the "inverse" data estimating method proposed in [9, 10] based on Bayes theory. But Bayes method requires the prior distribution information of products' life, which has limited applications. In this paper the idea of isotonic regression [1, 3, 6] is applied to analyzing data from the storage life test, regulating the inversed frequency of failure due to the randomness. And then, combined with Pearson goodness of fit theory, we give a complete method for estimating test data.

II. ISOTONIC REGRESSION BASED ON FAILURE FREQUENCY

In the destructive regular storage life test, the failure frequency at time t_i is:

$$f_i = \frac{X_i}{n_i}, \quad i = 1, 2, \cdots, k \qquad (2)$$

We assume $T = \{t_1, t_2, \cdots, t_k\}$ is a finite set, and $f = (f_1, f_2, \cdots, f_k)'$ is a bounded function defined on T. Also we define a semi order " $\prec$ " on T, when:

$$t_i \in T, \quad t_j \in T, \quad t_i \prec t_j \qquad (3)$$

the inequality $f_i^* = f^*(t_i) \le f_j^* = f^*(t_j)$ is always workable, then we call $f^* = (f_1^*, f_2^*, \cdots, f_k^*)'$ is an order-preserving function relative to " $\prec$ " defined on T.

Mark G as the whole set for the order-preserving functions. If there is $f^* \in G$, which satisfied with:

$$\sum_{i=1}^{k} (f_i - f_i^*)^2 \omega_i = \min_{\forall g \in G} \sum_{i=1}^{k} (f_i - g_i)^2 \omega_i \qquad (4)$$

then we call $f^* = (f_1^*, f_2^*, \cdots, f_k^*)'$ as an isotonic regression of f, when $\omega = (\omega_1, \omega_2, \cdots, \omega_k)'$, $\omega_i > 0$ is a weight function. By using the calculation method of isotonic regression below, we can alter the possible "inverse" of original frequency into the frequency meeting sequence constraint.

Algorithm PAVA:

a. If $f \in G$, then $f^* = f$;

b. If there is a j making $f_j > f_{j+1}$, we make:

$$B = \{j, j+1\}, \quad f_B = A_V(B) = \frac{\sum_{i \in B} f_i \omega_i}{\sum_{i \in B} \omega_i}, \quad \omega_B = \omega_j + \omega_{j+1} \quad (5)$$

Moreover:

$$\tilde{f} = (f_1, \cdots, f_{j-1}, f_B, f_{j+2}, \cdots, f_k)' \quad (6)$$

$$\tilde{\omega} = (\omega_1, \cdots, \omega_{j-1}, \omega_B, \omega_{j+2}, \cdots, \omega_k)' \quad (7)$$

c. Repeat step b until subscript set K is divided into l pieces: $B_1, B_2, \cdots, B_l$, which is satisfied with:

$$A_V(B_1) < A_V(B_2) < \cdots < A_V(B_l) \quad (8)$$

Then we can get:

$$f_i^* = A_V(B_t), \quad i \in B_t, \quad i = 1, 2, \cdots, l \quad (9)$$

III. PEARSON CHI-SQUARE GOODNESS OF FIT TEST WITH VARIABLE SAMPLES AND MINIMUM CHI-SQUARED ESTIMATION

In the theory of goodness of fit test [2,4], Pearson proposed χ^2 statistics used to test and verify the joint distribution of a group of independent samples whether belonging to the family of distributions with specific properties. Among these, the form of composite variable sample Pearson χ^2 statistic is:

$$\chi^2(\vec{\lambda}) = \sum_{i=1}^{k} \frac{(X_i - n_i p_i(\vec{\lambda}))^2}{n_i p_i(\vec{\lambda})} \quad (10)$$

in which, $\vec{\lambda} = (\lambda_1, \lambda_2, \cdots)$ represents parameter vector.

Pearson χ^2 statistic describes the difference between desired frequencies and observed frequencies. When $n_i \to \infty$, the limit distribution of $\chi^2(\vec{\lambda})$ is χ^2 distribution whose degree of freedom is $k-1$, that is $\chi^2(\vec{\lambda}) \sim \chi_{k-1}^2$. When conducting the test, without knowing the true value of $\vec{\lambda}$, a natural thought is to replace $\vec{\lambda}$ with the estimated value of $\vec{\lambda}$, which is $\hat{\vec{\lambda}}$. To calculate $\chi^2(\vec{\lambda})$, we use:

$$\chi^2(\hat{\vec{\lambda}}) = \sum_{i=1}^{k} \frac{(X_i - n_i p_i(\hat{\vec{\lambda}}))^2}{n_i p_i(\hat{\vec{\lambda}})} \quad (11)$$

as the test statistic, Fisher proves that when $\hat{\vec{\lambda}}$ is the consistency estimation of $\vec{\lambda}$, the limit distribution of $\chi^2(\vec{\lambda})$ is χ_{k-s-1}^2, when s is the dimension of parameter $\vec{\lambda}$. So if significance level is α, we have:

$$P(\chi^2(\hat{\vec{\lambda}}) \geq \chi_{k-s-1}^2(1-\alpha) \mid H_0) \leq \alpha \quad (12)$$

When $\chi^2(\hat{\vec{\lambda}}) \geq \chi_{k-s-1}^2(1-\alpha)$, small probability events will reject the null hypothesis.

The meaning of minimum chi-square estimation is that we consider the value $\tilde{\lambda}$ which minimizes Pearson χ^2 statistic as the best estimated value of $\vec{\lambda}$, which is:

$$\chi^2(\tilde{\lambda}) = \min\{\chi^2(\vec{\lambda}) : \vec{\lambda} \in \Lambda\} \quad (13)$$

Fisher also points out that $\tilde{\lambda}$ and $\hat{\vec{\lambda}}$ are equivalent, in other words, $\chi^2(\tilde{\lambda})$ and $\chi^2(\hat{\vec{\lambda}})$ share the same limit distribution.

Solve the equation $\dfrac{\partial \chi^2(\vec{\lambda})}{\partial \lambda_j} = 0, \quad j = 1, 2, \cdots, s$, we can get the minimum chi-square estimation of $\vec{\lambda}$.

$$\begin{aligned}
\frac{\partial \chi^2(\vec{\lambda})}{\partial \lambda_j} &= \sum_{i=1}^{k} \{-\frac{2n_i(X_i - n_i p_i(\vec{\lambda}))}{n_i p_i(\vec{\lambda})} - \frac{(X_i - n_i p_i(\vec{\lambda}))^2}{n_i p_i^2(\vec{\lambda})}\} \cdot \frac{\partial p_i(\vec{\lambda})}{\partial \lambda_j} \\
&= \sum_{i=1}^{k} \{\frac{-X_i^2 + n_i^2 p_i^2(\vec{\lambda})}{n_i p_i^2(\vec{\lambda})}\} \cdot \frac{\partial p_i(\vec{\lambda})}{\partial \lambda_j} \\
&= \sum_{i=1}^{k} (1 - (\frac{f_i}{p_i(\vec{\lambda})})^2) \cdot \frac{n_i \cdot \partial p_i(\vec{\lambda})}{\partial \lambda_j}
\end{aligned}$$

So the minimum chi-square estimation $\tilde{\lambda}$ of $\vec{\lambda}$ is the solution to:

$$\sum_{i=1}^{k} \{1 - [\frac{f_i}{p_i(\vec{\lambda})}]^2\} \cdot \frac{n_i \cdot \partial p_i(\vec{\lambda})}{\partial \lambda_j} = 0, \quad j = 1, 2, \cdots, s \quad (14)$$

IV. INVERSE DATA ESTIMATION METHOD BASED ON ISOTONIC REGRESSION AND MINIMUM CHI-SQUARED ESTIMATION

a. We use PAVA algorithm to regulate the inverse in our data: select $\omega = (\omega_1, \cdots, \omega_k)'$, $\omega_i > 0$, then get the isotonic regression of $\{f_i = \dfrac{X_i}{n_i}\}$, which is f_i^*, then mark:

$$X_i^* = f_i^* \cdot n_i \quad (15)$$

b. Assuming that the overall theoretical failure rate $p_i(\vec{\lambda})$, which is at the time t_i, obeys specific distribution function $F(t, \vec{\lambda})$, in other words:

$$p_i(\vec{\lambda}) = F(t_i, \vec{\lambda}) \quad (16)$$

c. Calculate the chi-square statistic:

$$\chi^2(\vec{\lambda}) = \sum_{i=1}^{k} \frac{(X_i^* - n_i p_i(\vec{\lambda}))^2}{n_i p_i(\vec{\lambda})} = \sum_{i=1}^{k} \frac{n_i(f_i^* - p_i(\vec{\lambda}))^2}{p_i(\vec{\lambda})} \quad (17)$$

d. Solve the equation:

$$\sum_{i=1}^{k} \{1 - [\frac{f_i^*}{p_i(\vec{\lambda})}]^2\} \cdot \frac{n_i \cdot \partial p_i(\vec{\lambda})}{\partial \lambda_j} = 0, \quad j = 1, 2, \cdots, s \quad (18)$$

and we can get $\tilde{\lambda}$, which is the minimum chi-square estimation of $\vec{\lambda}$.

e. Do Pearson chi-square goodness of fit test. If the result meet with the requirement, end the calculation; if not, go back to step b.

V. $(\tilde{\lambda}_L, \tilde{\lambda}_U)$, THE CONFIDENCE INTERVAL OF $\tilde{\lambda}$

In practical engineering applications, we should be more concerned about the range of a reliability parameter in a certain probability, which is $(\tilde{\lambda}_{jL}, \tilde{\lambda}_{jU})$, the confidence interval of $\tilde{\lambda}_j$. Under the given confidence level α, make $P(\tilde{\lambda}_{jL} \le \tilde{\lambda}_j \le \tilde{\lambda}_{jU}) = 1 - \alpha$. After regulating the inverse data, we calculate:

$$\frac{\partial \chi^2(\vec{\lambda})}{\partial \lambda_j} = \sum_{i=1}^{k} n_i \frac{\partial p_i(\vec{\lambda})}{\partial \lambda_j}[1 - (\frac{f_i^*}{p_i(\vec{\lambda})})^2] \qquad (19)$$

And function $\chi^2(\vec{\lambda})$ can get its min when $\vec{\lambda} = \tilde{\lambda}$. So when $j = 1, 2, \cdots$, we can always get:

$$\chi^2(\tilde{\lambda}_j) \le \chi^2(\tilde{\lambda}_{jL}) \qquad (20)$$

$$\chi^2(\tilde{\lambda}_j) \le \chi^2(\tilde{\lambda}_{jU}) \qquad (21)$$

Assuming $\chi^2(\tilde{\lambda}_{jL})$ equals to $\chi^2(\tilde{\lambda}_{jU})$, the necessary and sufficient condition of $P(\tilde{\lambda}_{jL} \le \tilde{\lambda}_j \le \tilde{\lambda}_{jU}) = 1 - \alpha$ is:

$$P(\chi^2(\tilde{\lambda}_j) \le \chi^2(\tilde{\lambda}_{jL}) = \chi^2(\tilde{\lambda}_{jU})) = 1 - \alpha \qquad (22)$$

From $\chi^2(\tilde{\lambda}) = \sum_{i=1}^{k} \frac{(X_i^* - n_i p_i(\tilde{\lambda}))^2}{n_i p_i(\tilde{\lambda})} \sim \chi_{k-s-1}^2$, we know that:

$$\chi^2(\tilde{\lambda}_{jL}) = \chi^2(\tilde{\lambda}_{jU})$$

is the $(1 - \alpha)$ quantile of χ_{k-s-1}^2, which is:

$$\chi^2(\tilde{\lambda}_{jU}) = \chi^2(\tilde{\lambda}_{jL}) = \chi_{k-s-1}^2(1 - \alpha) \qquad (23)$$

Then we can get $(\tilde{\lambda}_{jL}, \tilde{\lambda}_{jU})$, which is the confidence intervals of $\tilde{\lambda}_j$, by solving the equation.

VI. SIMULATION RESULTS AND ANALYSIS

In order to investigate the effect of the parameter estimation method in this paper for estimating data, we conducted a simulation contract. Assuming the product storage life to obey two-parameter Weibull distribution:

$$F(t) = 1 - \exp(-\left(\frac{t}{\eta}\right)^m) \qquad (24)$$

Compare the maximum likelihood estimation and the estimation method proposed in this paper which is based on isotonic regression and minimum chi-square theory. Conduct the destructive regular storage life test by the method below:

According to practical engineering experience, we mark k as the group number of each test and choose $k = 4$, $k = 6$, $k = 8$ respectively, the values of $\{(t_i, n_i)\}$ can be referred in Table 1 below:

TABLE I. THE VALUE OF $\{(t_i, n_i)\}$.

k		(t_i, n_i)
$k = 4$		$(8,12),(11,15),(13,10),(15,8)$
$k = 6$	①	$(5,6),(8,7),(11,8),(13,6),(15,5),(18,8)$
	②	$(5,10),(8,12),(11,15),(13,10),(15,8),(18,15)$
	③	$(5,30),(8,35),(11,40),(13,30),(15,25),(18,40)$
$k = 8$		$(3,8),(5,10),(8,12),(11,15),(13,10),(15,8),(18,15),(20,12)$

When $\eta = 10$, $m = 3$, simulating 100 times, according to the coefficient of variation of the estimated parameters, we assess the superiority of an estimation method, here is the coefficient of variation:

$$C \cdot V(\zeta) = \frac{\sigma_\zeta}{\mu_\zeta} = \frac{\sqrt{D(\zeta)}}{E(\zeta)} = \frac{\sqrt{\frac{1}{l-1}\sum_{i=1}^{l}(\tilde{\zeta}_i - \zeta)^2}}{\frac{1}{l}\sum_{i=1}^{l}\tilde{\zeta}_i} \qquad (25)$$

in which, ζ can be referred as a parameter, which represents η or m here.

To make our conclusions more representative, we compare the coefficient of variation while k and (t_i, n_i) are changing. We can see the simulation results in Table 2 and 3.

TABLE II. COEFFICIENT OF VARIATION FOR DIMENSION PARAMETER η

Simulation Results			MLE	Proposed Method	
				$\omega_i = n_i$	$\omega_i = X_i(1 - \frac{X_i}{n_i})$
$C \cdot V(\eta)$	$k = 4$		0.341467	0.254468	0.250832
	$k = 6$	①	0.325576	0.198375	0.192727
		②	0.281259	0.224421	0.200703
		③	0.236022	0.197084	0.184062
	$k = 8$		0.286446	0.203605	0.201239

TABLE III. COEFFICIENT OF VARIATION FOR SHAPE PARAMETER m

Simulation Results			MLE	Proposed Method	
				$\omega_i = n_i$	$\omega_i = X_i(1 - \frac{X_i}{n_i})$
$C \cdot V(m)$	$k = 4$		0.150583	0.060764	0.060607
	$k = 6$	①	0.137321	0.073290	0.068794
		②	0.134984	0.065273	0.065867
		③	0.129926	0.044917	0.040352
	$k = 8$		0.143542	0.053715	0.053082

We can see that the traditional method of maximum likelihood estimation has a strict requirement for the size of the sample, to obtain a smaller coefficient of variation; we can only

increase the sample size, which is difficult to carry out in practical engineering applications. The method proposed in this paper has a smaller coefficient of variation comparing to MLE while doing parameter estimation and the coefficient of variation changes little when the sample size changes. This is mainly because isotonic regression reduces the adverse effect of "inverse" data for parameter estimation, while minimum chi-square estimation can be more stable when dealing with the small sample data compared to maximum likelihood estimation theory. When conducting the storage life test, by the method proposed in this paper, we can solve the problem that the sample size is too small by increasing the value of k properly.

When conducting isotonic regression, we can set different weights ω_i. In this paper we select $\omega_i = n_i$ and $\omega_i = X_i(1 - X_i/n_i)$ to contrast. We can see that the coefficient of variation obtained by the second weight is smaller, because the second weight considers both the randomness of number of failures and the selection of samples. When we choose $\omega_i = n_i$, we can deal with the data more conveniently.

REFERENCES

[1] T. Robertson, F. T. Wright, and R. L. Dykstra. Order Restricted Statistical Inference. New York: John Wiley & Sons, 1988.

[2] J. Berkson. Minimum Chi-squared, not Maximum Likelihood! The Annals of Statistics, 1980, 8(2): 457-487.

[3] N. Balakrishnan, E. Beutner, M. Kateri. Order Restricted Inference for Exponential Step-Stress Models. IEEE Transactions on Reliability, 2009, 58(1): 132-142.

[4] Quang H. Vuong. Minimum chi-square estimation and tests for model selection. Journal of Econometrics, 1993, 56(1-2): 141-168.

[5] H. M. Fu. Method for Reliability Assessment with Very Small Sample. Journal of Mechanical Strength, 2005, 27 (3): 335-338.

[6] N. Z. Shi. Isotonic Regression and Maximum Likelihood Estimation. Applied Probability and Statistics, 1993, 9(2): 203-214.

[7] N. Z. Shi. Maximum likelihood estimation of means and variances from normal populations under simultaneous order restrictions. Journal of Multivariate Analysis, 1994, 50(2): 282-293.

[8] R. T. Zhang, H. X. Liu. A Method for Estimating "Inverse" Data in Statistical Analysis of Reliability, 1991, 4(4): 22-28.

[9] B. J. Song, Y. X. Qin. A Reliability Assessment Method for small sample of products based on Bayes. Missiles and Guidance, 1999, 20(2): 52-57.

[10] T. A. Mazzuchi, R. Soyer. A Bayes method for assessing product-reliability during development testing. IEEE Trans. Reliability, 42(18): 503-510.