Reliability Engineering and System Safety 160 (2017) 114–121

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Bootstrap analysis of designed experiments for reliability improvement with a non-constant scale parameter

Guodong Wang[a,*], Zhen He[b], Li Xue[a], Qingan Cui[c], Shanshan Lv[b], Panpan Zhou[b]

[a] Department of Management Engineering, Zhengzhou University of Aeronautics, Zhengzhou 450015, China
[b] College of Management and Economics, Tianjin University, Tianjin 300072, China
[c] School of Management Engineering, Zhengzhou University, Zhengzhou 450001, China

ARTICLE INFO

Keywords:
Design of experiments
Censored data
Percentile
Weibull distribution

ABSTRACT

Factors which significantly affect product reliability are of great interest to reliability practitioners. This paper proposes a bootstrap-based methodology for identifying significant factors when both location and scale parameters of the smallest extreme value distribution vary over experimental factors. An industrial thermostat experiment is presented, analyzed, and discussed as an illustrative example. The analysis results show that 1) the misspecification of a constant scale parameter may lead to misidentify spurious effects; 2) the important factors identified by different bootstrap methods (i.e., percentile bootstrapping, bias-corrected percentile bootstrapping, and bias-corrected and accelerated percentile bootstrapping) are different; 3) the number of factors affecting 10th percentile lifetime significantly is less than the number of important factors identified at 63.21th percentile.

1. Introduction

1.1. Background

With markets increasingly competitive, leading practitioners have to find efficient and economical methods for improving product reliability [1]. Many factors may affect product reliability. Practitioners need to identify important factors and choose levels of these factors that lead to improved reliability. Design of experiments (DOE) has received a great deal of attention as a quality improvement tool [2–4]. However, statistical techniques of DOE cannot be directly applied to lifetime data analysis, because the distribution of lifetime data is usually non-normal. The usual *F*-tests from analysis of variance are not valid. Generalized linear models (GLM) can deal with non-normal distribution [5]. However, Weibull distribution which is often used in reliability analysis does not belong to the exponential family. Additionally, the lifetime data may be censored. When censoring exists, DOE that based on least square method cannot be used to improve product reliability [6].

Maximum likelihood method, easily accounts for censored data and non-normal data, is generally recommended for calculating parameter estimates for lifetime models [7–9]. Some statistical software packages are available for analysis of lifetime data, such as SAS, R, MINITAB, and JMP. To test the significance of the estimates, they need to be compared with their standard errors. For example, in SAS software [10], the LIFEREG procedure estimates the parameters by maximum likelihood with a Newton-Raphson algorithm. The standard errors of the parameter estimates are computed from large sample normal approximations by using the observed information matrix. If convergence of the maximum likelihood estimates (MLEs) is attained, the Chi-square test statistic is computed for each effect, testing whether there is any contribution from any of the levels of the effect. Condra [11], Hamada [12], and Wu and Hamada [4] gave several examples to illustrate how to improve product reliability using design of experiments. Rigdon et al. [13] discussed a number of examples and presented results using these statistical software packages.

1.2. Motivating example

The aforementioned methods can deal with problems by assuming scale parameter is a constant. In fact, sometimes, the scale parameters under different treatment combinations are different. Lawless [7] suggested using likelihood ratio tests to decide whether the scale parameter is constant.

Bullington et al. [14] used a Plackett-Burman design for improving reliability of an industrial thermostat and analyzed lifetime data using LIFEREG procedure in SAS software. The results showed that all factors are significant at the 0.0025 level. They pointed out that

* Corresponding author.
E-mail address: gdwang@tju.edu.cn (G. Wang).

http://dx.doi.org/10.1016/j.ress.2016.12.006
Received 27 October 2015; Received in revised form 8 December 2016; Accepted 15 December 2016
Available online 21 December 2016
0951-8320/

Nomenclature

Acronyms

ALTs Accelerated Life Tests
BCa Bias-corrected and Accelerated Percentile Bootstrapping
BCIs Bootstrapping Confidence Intervals
BCPB Bias-corrected Percentile Bootstrapping
CDF Cumulative Distribution Function
DOE Design of Experiments
GLM Generalized Linear Models
MLEs Maximum Likelihood Estimates
MTTF Mean Time to Failure
PB Percentile Bootstrapping
SEV Smallest Extreme Value

Notation

m number of treatment combinations
n number of items within each treatment combination
K repeat times of resampling
$\boldsymbol{X}$ matrix of treatment combinations
C_{pp} process incapability index
t_{ij} failure time of jth item in the ith treatment combination
y_{ij} $\log(t_{ij})$
y_{ip} 100pth percentile of log lifetime in the ith treatment combination
$\boldsymbol{a}$ vector of parameters of model with y_{ij}
$\boldsymbol{b}$ vector of parameters of model with y_{ip}
$\hat{\boldsymbol{b}}$ the estimate of $\boldsymbol{b}$ obtained from original samples
$\hat{\boldsymbol{b}}^*$ the estimate of $\boldsymbol{b}$ obtained from resampling
$\boldsymbol{P}_0$ $\Pr[\hat{\boldsymbol{b}}^* \le \hat{\boldsymbol{b}}]$
$\boldsymbol{P}_L$ the probability of lower quantile
$\boldsymbol{P}_U$ the probability of upper quantile
z_α the lower α th quantile of standard normal distribution
z_0 $\Phi^{-1}(P_0)$
ε_{ij} the measurement errors of $\log(t_{ij})$
μ_i the location parameter of SEV distribution in the ith treatment combination
σ_i the scale parameter of SEV distribution in the ith treatment combination
z_{ij} $z_{ij} = (y_{ij} - \mu_i)/\sigma_i$
$F(\cdot)$ CDF of log(lifetime)
Φ_{sev} the standard SEV distribution
δ_{ij} censoring indicator of jth item in the ith treatment combination
$\mathcal{L}(\cdot)$ likelihood function
$\ell(\cdot)$ log-likelihood function
$\hat{\Sigma}$ the estimate of covariance matrix

"scarcity of effects" principle is the case in most screening experiments, but a number of the effects included in this study had already been demonstrated to have an effect. We test the null hypothesis that the scale parameter is constant. The result shows that the null hypothesis is rejected with a very small p-value.

1.3. Literature review

Examining the literature, we find few publications that use DOE method to improve product reliability considering the scale parameter is non-constant. Besseris [15] analyzed lifetime data from designed experiments using a rank sum test method. Mee and Lu [16] noted that there is a fundamental flaw to using rank sum tests for analyzing factorial designs. That is, rank sum statistics for the factorial effects are not independently distributed. Piña-Monarrez and Ortiz-Yañez [17] analyzed lifetime data from Taguchi experiments using multiple linear regression. However, the proposed method cannot be used to identify important factors which affect 100pth percentile significantly.

In accelerated life tests (ALTs), the scale parameter may be non-constant, which is widely studied in the literature [18–24]. All of these works assume that the scale parameter is a function of stress factors. The coefficients of the models can be estimated by maximum likelihood method. The variance estimates of coefficients can be obtained by computing the inverse of Fisher information matrix. However, this procedure is not available for computing the asymptotic variance-covariance matrix when we assume the scale parameter is non-constant, because the number of factors in DOE is much larger than the number of stress factors in ALT. Furthermore, the aforementioned literature on ALTs assume that the scale parameter and the location parameter are functions of stress factors, respectively. Thus, we cannot decide which factors affect the product lifetime significantly even if we have known which factors are important for scale parameter and which factors are important for location parameter of location-scale models.

In design of experiments for quality improvement, interest is focused on the mean or average response. In reliability improvement experiments, the focus is often on percentile lifetime. These low percentiles provide engineers with an evaluation of the product's early failures as well as information of specification limits, warranty, and cost analysis. Additionally, for smallest extreme value distribution, mean is a special case of percentiles, i.e., 42.96th percentile; location parameter is also a special case of percentiles, i.e., 63.21th percentile.

In order to identify the important factors that affect percentiles of a product, we must test the significance of the estimates which needs to be compared with standard errors. It is difficult to assess significance via DOE when the data is censored, because it is very hard to obtain the degrees of freedom. Maximum likelihood method can deal with censored data, but it assumes that the ratio follows normal distribution. Bootstrapping is a direct method to identify important experimental factors. It does not have to consider the distribution of estimator and does not need to calculate the degrees of freedom. Efron and Tibshirani [25] pointed out that bootstrapping in statistics is a computer-intensive resampling method used to estimate properties of a statistic that are difficult to calculate analytically. DiCiccio and Efron [26] presented several types of bootstrap confidence intervals including standard, percentile, and bootstrap-t. Edwards et al. [27] compared two wood plastic composite extruders using bootstrapping confidence interval of percentiles. Wang et al. [28] analyzed lifetime data with subsampling via two-stage bootstrapping. Kenett et al. [29] proposed the application of the bootstrapping for the analysis of designed experiments as an alternative to a standard regression approach. Kenett et al. treated the right-censored times as actual failure times and used bootstrap method in the analysis of screening design. Ignoring the censoring information, however, can lead to wrong decisions because the unobserved failure times and right-censored times may differ greatly, depending on the particular factor level combination. A simulation study in Hamada and Wu [30] showed that this method can perform poorly by missing some important factors and misidentifying spurious factors. Efron [31] presented bias-corrected percentile method to reduce the median-bias of bootstrap distribution. In order to improve the coverage properties of the percentile confidence intervals, Efron [32] suggested achieving the normality of bootstrap distribution by some transformation. Chou et al. [33] computed process incapability index C_{pp} by using bootstrap confidence intervals.