ARTICLE IN PRESS
Chinese Journal of Aeronautics, (2016), xxx(xx): xxx–xxx

Chinese Society of Aeronautics and Astronautics & Beihang University
Chinese Journal of Aeronautics
cja@buaa.edu.cn
www.sciencedirect.com

Scatter factor confidence interval estimate of least square maximum entropy quantile function for small samples

Wu Fuxian [a,b], **Wen Weidong** [a,b,*]

[a] Jiangsu Province Key Laboratory of Aerospace Power System, College of Energy and Power Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[b] State Key Laboratory of Mechanics and Control of Mechanical Structures, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Received 22 December 2015; revised 28 March 2016; accepted 19 May 2016

KEYWORDS
Confidence intervals;
Maximum entropy;
Quantile function;
Reliability;
Scatter factor;
Small samples

Abstract Classic maximum entropy quantile function method (CMEQFM) based on the probability weighted moments (PWMs) can accurately estimate the quantile function of random variable on small samples, but inaccurately on the very small samples. To overcome this weakness, least square maximum entropy quantile function method (LSMEQFM) and that with constraint condition (LSMEQFMCC) are proposed. To improve the confidence level of quantile function estimation, scatter factor method is combined with maximum entropy method to estimate the confidence interval of quantile function. From the comparisons of these methods about two common probability distributions and one engineering application, it is showed that CMEQFM can estimate the quantile function accurately on the small samples but inaccurately on the very small samples (10 samples); LSMEQFM and LSMEQFMCC can be successfully applied to the very small samples; with consideration of the constraint condition on quantile function, LSMEQFMCC is more stable and computationally accurate than LSMEQFM; scatter factor confidence interval estimation method based on LSMEQFM or LSMEQFMCC has good estimation accuracy on the confidence interval of quantile function, and that based on LSMEQFMCC is the most stable and accurate method on the very small samples (10 samples).

1. Introduction

The estimation of quantile function to test samples is commonly encountered in the reliability analysis of engineering system (such as aviation product properties). In traditional methods of statistical inference, the steps for quantile function estimation involve fitting an analytical probability distribution

* Corresponding author. Tel.: +86 25 84892251.
E-mail addresses: wfxfjxy@163.com (F. Wu), gswwd@nuaa.edu.cn (W. Wen).

Peer review under responsibility of Editorial Committee of CJA.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.cja.2016.08.015
1000-9361

that is judged empirically from the available sample data, and then distribution parameters are suitably estimated using methods such as maximum likelihood method or least square method. However the bias and efficiency of quantile function estimate remain sensitive to the type of assumed distribution.

An alternative approach for the distribution fitting comes from the modern information theory in which a robust measure of probabilistic information, the entropy, has been developed. The maximum entropy principle is presented as a rational approach for choosing the least biased probability distribution among all possible distributions which are consistent with available sample data and contain the minimum amount of spurious information. In Refs.[1,2], the maximum entropy is used to characterize a large number of discrete and continuous distributions under certain constraints. In Refs.[3–6], the maximum entropy is applied in the engineering problems, and compared with the common probability distributions. In Refs.[7,8], the maximum entropy is also applied in the climate science to estimate the probability distribution of rainfall amount, temperature and emissivity. There are also some researches to improve the performance of maximum entropy. Maximum entropy is combined with adaptive importance sampling for modeling different kinds of loss distributions.[9] The maximum entropy method is incorporated into Bayesian probability theory to estimate the distribution of intensity of weighted image.[10] The minimum Kullback entropy rather than the maximum entropy is used in the derivation of exponential distribution with given mean and variance.[11] The maximum entropy method is combined with bootstrap method to resolve the problem about parameter estimation under the condition that the number of test times is small while at every test time, the number of test data is large.[12] The maximum entropy method is combined with the Monte Carlo simulation to estimate the dam overtopping probability.[13]

However, the estimates of higher order moments (order >2) from small samples (size less than 30) tend to have large sampling errors. The maximum entropy distribution derived from poor moment estimates would lead to inaccurate quantile values.[14–16] This difficulty can be circumvented by using the probability weighted moments (PWMs) in place of ordinary moments. The PWMs are firstly introduced by Greenwood and Landwehr[17] The PWMs can be viewed as the expectations of order statistics and moments of the quantile function of any nonnegative random variable. In contrast with ordinary moments, PWMs are less sensitive to the effects of sampling variability. Also higher order PWMs can be accurately estimated from small samples, because they are linear combinations of the observed sample values.[18] PWMs shows the potential to provide a robust solution to characterizing the statistical nature of the underlying distributions especially for small samples. For this reason, PWMs have been extended to a variety of field of science and engineering.[19–21] But there will be some sampling errors on the very small samples (size less than 10), which can be verified in Section 4.

Because of expensiveness of aviation product and high test cost, the test is conducted only on the small samples. There will be some errors between the estimated quantile function and theoretical quantile function on the very small samples, so scatter factor method is introduced to estimate the confidence interval of quantile function, where there is assigned probability for including the theoretical quantile function.[22] The scatter factor is widely used to estimate the fatigue life of air plane and air engine components.[23,24] In Ref.[25], life scatter factor method based on kth order experiment life with logarithm normal distribution is deduced. In Refs.[26–28], the scatter factor formula is deduced for common probability distributions and has a good application in the aviation product. But there is a weakness that probability distribution type and distribution parameters of the random variable must be determined before it is used.

In this paper, to overcome the low accuracy of classic maximum entropy quantile function method (CMEQFM) on the very small samples, least square maximum entropy quantile function method (LSMEQFM) and that with constraint condition (LSMEQFMCC) are proposed. To improve the confidence level of quantile function estimate, scatter factor method is introduced. The results show that LSMEQFMCC is the most stable and accurate method for quantile function estimate on the very small samples.

2. Maximum entropy quantile function methods

2.1. CMEQFM

For a continuous random variable X, with the quantile function $x(u)$ where $u(x) = P(X \leqslant x)$ is the cumulative distribution function (CDF) and $0 \leqslant u(x) \leqslant 1$, the classic maximum entropy of quantile function $x(u)$ is defined as[12,14,16,20,21,29]

$$\begin{cases} \max\ S = -\int_0^1 x(u)\ln x(u)\mathrm{d}u \\ \text{s.t. } \int_0^1 u^j x(u)\mathrm{d}u = b_j \quad j = 0, 1, \ldots, m \end{cases} \tag{1}$$

where S is the entropy of quantile function $x(u)$; b_j is the one form of PWMs; m is the highest order of PWMs considered in the analysis.

From an ordered random sample of X with size n $(x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n)$, b_j can be obtained as

$$b_j = \frac{1}{n}\sum_{i=1}^{n}\left[\binom{i-1}{j}\Big/\binom{n-1}{j}x_i\right] \tag{2}$$

To account for constraints Eq. (2), the entropy function is augmented as

$$\bar{S} = -\int_0^1 x(u)\ln x(u)\mathrm{d}u - (\lambda_0 - 1)\left(\int_0^1 x(u)\mathrm{d}u - b_0\right) - \sum_{j=1}^{m}\lambda_j\left(\int_0^1 u^j x(u)\mathrm{d}u - b_j\right) \tag{3}$$

where λ_j denotes an unknown Lagrangian multiplier. To derive the quantile function, the entropy is maximized using the following condition

$$\frac{\partial \bar{S}}{\partial x(u)} = 0 \tag{4}$$

Substituting Eq. (3) into Eq. (4) with subsequent simplification leads to quantile function as

$$x(u) = \exp\left(-\sum_{j=0}^{m}\lambda_j u^j\right) \tag{5}$$

Substituting Eq. (5) into $\int_0^1 x(u)\mathrm{d}u = b_0$, we can write λ_0 as

$$\lambda_0 = \ln\left[\int_0^1 \exp\left(-\sum_{j=1}^{m}\lambda_j u^j\right)\mathrm{d}u/b_0\right] \tag{6}$$

Please cite this article in press as: Wu F, Wen W Scatter factor confidence interval estimate of least square maximum entropy quantile function for small samples, *Chin J Aeronaut* (2016), http://dx.doi.org/10.1016/j.cja.2016.08.015