Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation 97 (2014) 1–13
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Original article

Prediction intervals of future observations for a sample of random size from any continuous distribution

H.M. Barakat [a], Magdy E. El-Adll [b,*], Amany E. Aly [b]

[a] Department of Mathematics, Faculty of Science, Zagazig University, Zagazig, Egypt
[b] Department of Mathematics, Faculty of Science, Helwan University, Ain Helwan, Cairo, Egypt

Received 27 April 2012; received in revised form 26 January 2013; accepted 12 June 2013
Available online 19 August 2013

Abstract

In this paper, a general method for predicting future observations from any arbitrary continuous distribution is proposed. Two pivotal statistics are developed to construct prediction intervals of future observations in two cases. In the first case, we assume fixed sample size, while in the second case, the sample size is assumed to be positive integer-valued random variable independent of the observations. Explicit forms for the distribution functions of the pivotal statistics are derived. Some special cases for the random sample size are considered. An algorithm is constructed to demonstrate the practical importance of the theoretical results. Moreover, simulation study is applied on some important distributions to investigate the efficiency of the suggested method. Finally, an example for real lifetime data is analyzed, where it is assumed that the distribution of the data is unknown.

Keywords: Order statistics; Random sample size; Predicative interval; Probability coverage; Monte Carlo simulation

1. Introduction

In many applications of extreme-value theory, predictive inference of future or censored observations is the main interest. The prediction interval is a basic tool for predicting future observations. It is widely used in reliability theory, lifetime problems and can be applied in industrial applications to predict the number of defective units that will be produced during future production of a product. Censoring occurs when exact survival times are known only for a portion of the individuals or items under study. The complete survival times may not have been observed by the experimenter either intentionally or unintentionally.

It is well known that the subject of order statistics is a basic tool for lifetime prediction methods because if m items are put simultaneously in a life test, the weakest component will fail first, followed by the second weakest, and so on until all have failed. For example, in industry we are interested in the time to failure after n units are put in a life test. In such cases, the lifetimes are already arranged in ascending order of magnitude and do not have to be ordered. The practical importance of such experiments is evident. Moreover, the possibility is now open of terminating the experiment before its conclusion, by stopping after a given time (Type I censoring) or after a given number of failures

* Corresponding author. Tel.: +201099344955.
E-mail addresses: hbarakat2@hotmail.com (H.M. Barakat), meladll2@yahoo.com (M.E. El-Adll), a_abusabh@yahoo.com (A.E. Aly).

0378-4754/$36.00
http://dx.doi.org/10.1016/j.matcom.2013.06.007

(Type II censoring). It may be of interest, to predict the time at which future observations will have failed or to predict the mean failure time of the unobserved lifetimes. In these cases the interval or point predict are of interest.

Prediction problems have been studied by many authors. Lawless [11] examined predicting the smallest value for a future sample of k observations based on an observed sample and presented conditional confidence intervals. Hsieh [7] considered the case where n items are put on test and the first r failures are observed (type II censoring). He derived prediction intervals for the lifetime of the remaining components. Nelson [13] derives simple prediction limits for the number of failures that will be observed in a future inspection of a sample of units, where the past data consist of the cumulative number of failures in a previous inspection of the same sample of units. Recently, Raqab et al. [15], considered the prediction problem, for Pareto distribution based on progressively Type-II censored samples.

The prediction intervals for future order statistics from the exponential distribution have been studied by many authors, among them are Lawless [10], Geisser [6] and Balakrishnan and Lin [1]. A comprehensive survey of developments on prediction problems has been prepared by Kaminsky and Nelson [8]. Estimation of the parameters of the distribution and prediction of future observations are main problems in statistical inference.

In many applications, when it is impossible to predict the size of samples with absolute accuracy beforehand, it is necessary to consider that the size of a sample is a random variable itself. Perhaps, one of the major reasons for this phenomenon is that in many biological, agricultural and some quality control problems it is almost impossible to have a fixed sample size, because some observations always get lost for various reasons. Recently, Barakat et al. [2] obtained prediction intervals for future exponential lifetimes based on random generalized order statistics. Moreover, El-Adll [5] studied the prediction intervals for future lifetimes based on random number of three-parameter Weibull distributions.

In this paper, two pivotal statistics are modified to construct prediction intervals of future observations from any continuous distribution, when the sample size is assumed to be fixed and when the sample size is assumed to be a positive integer-valued random variable. The paper is organized as follows. In Section 2 the distribution functions (df's) of the two proposed pivotal quantities are derived. In Section 3 a simulation study is applied on two important lifetime distributions (namely, the normal and gamma distributions) for some special cases of the random sample size (namely, binomial and negative binomial random sample size) to investigate the efficiency of the suggested method. In Section 4, we study the case when the distribution of the underlying random variable for lifetime is unknown. Moreover, an example for real lifetime data is presented. Finally, Section 5 is devoted to some concluding remarks.

2. The main results

In this section we present the theoretical results of the paper. For simplicity we restrict ourselves with absolutely continuous distributions.

Lemma 2.1. *Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from an absolutely continuous distribution F and let $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ be the corresponding order statistics. Then the normalized spacings*

$$Z_i = (n-i+1)(-\ln \overline{F}(X_{i:n}) + \ln \overline{F}(X_{i-1:n})), \quad i = 1, 2, \ldots, n, \text{with} X_{0:n} \equiv F^{-1}(0+), \tag{2.1}$$

are independent and identically distributed random variables each of them has the standard exponential distribution (Exp(1)), where $\overline{F}(t) = 1 - F(t)$ and $F^{-1}(u) = \inf\{x : F(x) \geq u\}$.

Proof. Consider the following transformation

$$X_i^* = -\ln \overline{F}(X_i), \quad i = 1, 2, \ldots, n. \tag{2.2}$$

It is clear that, whatever the distribution F, X_i^* have $Exp(1)$, for all $i = 1, 2, \ldots, n$. Therefore, the random variables $Z_1 = nX^*_{1:n}$, $Z_2 = (n-1)(X^*_{2:n} - X^*_{1:n}), \ldots, Z_i = (n-i+1)(X^*_{i:n} - X^*_{i-1:n})$, $Z_n = (X^*_{n:n} - X^*_{n-1:n})$ are all independent standard exponential random variables (cf. David and Nagaraja [4], p. 18), which was to be proved. □

In order to construct a prediction interval for the value of a future order statistic $X_{s:n}$, which are based on the observed values of the order statistics $X_{1:n} < X_{2:n} < \cdots < X_{r:n}$, $r < s$, we apply the pivotal method. The pivotal method uses a pivotal quantity Q, which is an explicit function of both $X_{1:n}, X_{2:n}, \cdots, X_{r:n}$ and $X_{s:n}$, i.e., $Q = Q(X_{1:n}, X_{2:n}, \cdots, X_{r:n}, X_{s:n})$. Moreover, Q has two characteristics.