Proposal Report

Research Practice I

Use of metaheuristic methods in the estimation of indices of non-normal processes

***R**esponsible:*

Andrea Molina-Alonso
amolin13@eafit.edu.co

***T**utor:*

Myladis Rocio Cogollo-Flórez, PhD(c)
mcogollo@eafit.edu.co

Mathematical Engineering
Department of Mathematical Sciences
EAFIT University

September 8, 2015

Problem Formulation

The two parameter Burr distribution, called Burr type XII distribution, was introduced by Burr [1]. The specific forms of the cumulative distribution function and the probability density function are

$$F(x) = 1 - \frac{1}{(1+x^c)^k}; \quad x > 0,\ c > 0,\ k > 0$$

$$f(x) = \frac{k\ c\ x^{c-1}}{(1+x^c)^{k+1}}; \quad x > 0,\ c > 0,\ k > 0$$

where k is shape parameter and c is scale parameter. Changing the values of these parameters, it can be obtained several distributions commonly used such as gamma, lognormal, loglogistic and beta.

The Burr type XII distribution have been used in a variety of applied mathematics contexts [2]. Within them are the process capability analysis [3], quality control [4], reliability analysis [5] and failure time modeling [6]. Liu and Chen [4] demonstrated and justified that the use of Burr XII distribution improves the accuracy of the estimation of process capability indices or quality control charts obtained in processes that do not follow a normal distribution, as this distribution can be used to describe the behavior of real-world data.

The formulation of non-normal indices capability of processes using Burr type XII distribution requires the estimation of the parameters c and k from sample data. However, this procedure is not done directly, but the parameter estimation is done using tables. Specifically, the procedure used is as follows: (i) calculate the coefficients of kurtosis and skewness sample, (ii) search proposed values c and k in the table for these coefficients. By making the estimations with tables, there may be a loss of accuracy when using the distribution, especially when it is put into practice in real-world processes, in this case the calculation of process capability indices and therefore when is necessary to show that a process is reliable and is under control.

Moreover, to estimate the parameters c and k from the sample data, conventional methods of estimation: Maximum Likelihood (MLE) and Least Squares (LS) have been considered [7, 8, 9, 2]. Under these methods, the best estimate is given by

$$\boldsymbol{\theta}_{opt} = \arg \min_{\boldsymbol{\theta} \in \mathbf{R}^+ \times \mathbf{R}^+} L_T(\boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = (c,\ k)$ and

$$L_T(\boldsymbol{\theta}) = n(\ln(c) + \ln(k)) + (c-1)\sum_{i=1}^{n} \ln(x_i) - (k+1)\left(\sum_{i=1}^{n} \ln(1+x_i^c)\right)$$

is the logarithm of the likelihood function for the Burr XII distribution. In the process of optimization the following equations are obtained:

$$\frac{n}{c} + \sum_{i=1}^{n} \ln(x_i) - (k+1) \sum_{i=1}^{n} \frac{x_i^c \ln(x_i)}{1 + x_i^c} = 0 \tag{1}$$

$$\frac{n}{k} - \left(\sum_{i=1}^{n} \ln(1 + x_i^c) \right) = 0 \tag{2}$$

The set of equations, Equation 1 and Equation 2, are intractable analytically, and numerical methods must be used to resolve them. The difficulty of mathematical manipulation, negatively impacts the development of statistics asymptotic properties of the estimators [8, 2]. Furthermore, depending on the numerical method used the process of estimating c and k can be very costly in terms of computational time, which limits the use of Burr type XII distribution to build process capability indices of non-normal data in real time.

Objectives

General Objective

Propose an accurate parameter estimation method for the Burr type XII distribution.

Specific Objectives

- Identify which heuristics methods are appropriate to find a general way to estimate the parameters of the Burr XII distribution.
- Develop the found estimation method in a programming language.
- Compare the results obtained with the estimation method with other proposed in the literature, using experimental data.

Background

The estimation of process capability indices using the Burr type XII distribution is proposed to evaluate the outputs of processes that are not normally distributed. Liu and Chen [4] modify the Clements method and uses the Burr XII distribution for estimating indices capability of non-normal processes. Ahmad, Abdollahian, and Zeephongseku [10] show that the method of Burr is better in accuracy than the method of Clements. However in these articles, the estimation of the parameters of the Burr type XII distribution is not directly but through tabulated values.

In the literature there are several proposals for estimating the parameters of the Burr type XII distribution. Abbasi [7] uses a neural network type Multilayer perceptron (MLP) to estimate the distribution parameters, however is not presented a systematic procedure for the construction of model. Malinowska et al. [8] presents the theoretical development of obtaining the minimum variance linear unbiased estimators (MVLUE), the best linear invariant estimators (BLIE) and the maximum likelihood estimators (MLE) using generalized order statistics for the parameters of the Burr XII distribution. Wang and Lee [9] use the least squares (LS) method and the M-estimator to estimate the parameters based on the quantile function for complete data with outliers. Watkins [2] uses the MLE, exploiting the link between the Burr XII and the two parameter Weibull distribution. The MLE of the distribution must be found numerically, and therefore losing accuracy in the estimation.

Justification

Process capability indices measure how much variation a process experiences according to its specification limits [11] and they are generally defined based on three basic assumptions [4]: (i) the system determining which data are collected is under control, (ii) the collected process data are independent and identically distributed, (iii) the collected process data are normally distributed. But in practice, industrial production involves many non-normal processes, implying the need of estimation of process capability indices where the process output does not follow a normal distribution.

As discussed above, the Burr XII has been useful in recent researches in which the information does not follow a normal distribution, but the estimations of the parameters are not accurate. For this reason, establishing a method for estimating parameters of said distribution to obtain results without accessing tables would impact positively, not only statistical areas, but also process analysis areas, specifically in process capability indices.

The importance and originality of this research is:

- In the systematic literature review on non-normal process capability indices was found that there is not a proposed method for estimation of parameters that does not include tabulated values or the use of numerical methods.
- A suitable monitoring on production processes requires the use of methods that consider all data generating process (DGP), and therefore represents very well the reality of the process.
- In this project, the idea is to propose another estimation method more accurate where use of values tables is not required, allows to work in real-time with non-normal data, and thus be more faithful with the behavior of processes in reality

Scope

In this research project the intention is to propose an appropriate estimation method for the Burr XII distribution not associated with tabulated values, but its use in the analysis of process capability would not be implemented. In addition, validation would be done with experimental (not real) data.

Methodology

In order to ensure a proper development of the project, there will be one hour per week available to be focused in the current research with the tutor plus the time invested on individual study. Initially there will be a search of the state of art databases. Then a study of the necessary contents explored in order to achieve the objectives of the project and then the identification of the method will be started followed by the implementation of it in the appropriate programming language. Finally the experimental data for validation of the proposed method is generated with the respective analysis. Reports and presentations will be made to document the results obtained throughout the investigation.

Schedule of Activities

It is presented in Table 1 an estimated schedule of the different phases of the project with the respective deadlines. Different documents and presentations must be elaborated according to the Research Practice terms.

Activity	**Start**	**End**
Review of literature	July 20	September 29
Proposal report	July 25	August 9
Oral presentation of the proposal report	August 9	August 14
Identification of the method	August 10	September 15
Oral progress report	September 15	September 25
Method implementation	September 15	September 29
Validation with experimental data	September 29	October 15
Project report	August 15	November 6
Project presentation	October 24	November

Table 1: Schedule of Activities.

Intellectual Property

According to [12], Andrea Molina-Alonso and Myladis Rocio Cogollo-Flórez share intellectual property in this research equally.

The development of this research practice is associate with the research project "Estimation of processes quality control indexes using inaccurate data with deviations from normality" (internal code P14221) which is carried out jointly between EAFIT University and the Instituto Tecnológico Metropolitano (ITM).

References

[1] I. W. Burr, "Cumulative frequency functions," *The Annals of Mathematical Statistics*, vol. 13, no. 2, pp. 215–232, 1942.

[2] A. J. Watkins, "An algorithm for maximum likelihood estimation in the three parameter Burr XII distribution," *Computational statistics & data analysis*, vol. 32, no. 1, pp. 19–27, 1999.

[3] S. Ahmad, M. Abdollahian, P. Zeephongsekul, and B. Abbasi, "Multivariate nonnormal process capability analysis," *The International Journal of Advanced Manufacturing Technology*, vol. 44, no. 7-8, pp. 757–765, 2009.

[4] P.-H. Liu and F.-L. Chen, "Process capability analysis of non-normal process data using the Burr XII distribution," *The International Journal of Advanced Manufacturing Technology*, vol. 27, no. 9-10, pp. 975–984, 2006.

[5] M. D. Pant and T. C. Headrick, "A method for simulating Burr type III and type XII distributions through-moments and-correlations," *ISRN Applied Mathematics*, vol. 2013, 2013.

[6] P. F. Paranaíba, E. M. Ortega, G. M. Cordeiro, and R. R. Pescim, "The Beta Burr XII distribution with application to lifetime data," *Computational Statistics & Data Analysis*, vol. 55, no. 2, pp. 1118–1136, 2011.

[7] B. Abbasi, S. Z. Hosseinifard, and D. W. Coit, "A neural network applied to estimate Burr XII distribution parameters," *Reliability Engineering & System Safety*, vol. 95, no. 6, pp. 647–654, 2010.

[8] I. Malinowska, P. Pawlas, and D. Szynal, "Estimation of location and scale parameters for the Burr XII distribution using generalized order statistics," *Linear algebra and its applications*, vol. 417, no. 1, pp. 150–162, 2006.

[9] F.-K. Wang and C.-W. Lee, "M-estimator for estimating the Burr type III parameters with outliers," *Mathematics and Computers in Simulation*, vol. 105, pp. 144–159, 2014.

[10] S. Ahmad, M. Abdollahian, and P. Zeephongsekul, "Process capability estimation for non-normal quality charactersitics using Clement, Burr and Box–Cox methods," *ANZIAM Journal*, vol. 49, pp. 642–665, 2008.

[11] R. Miao, X. Zhang, D. Yang, Y. Zhao, and Z. Jiang, "A conjugate Bayesian approach for calculating process capability indices," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8099–8104, 2011.

[12] Universidad-EAFIT, “Reglamento de propiedad intelectual,” 2009.