

THE BOOTSTRAP METHOD AS A TOOL TO IMPROVE THE DESIGN OF EXPERIMENTS

doi: 10.2478/czoto-2019-0092

Date of submission of the article to the Editor: 30/11/2018

Date of acceptance of the article by the Editor: 05/02/2019

Renata Dwornicka¹ – *orcid id: 0000-0001-6761-9623*

Norbert Radek² – *orcid id: 0000-0002-1587-1074*

Jacek Pietraszek¹ – *orcid id: 0000-0003-2851-1606*

¹Cracow University of Technology, **Poland**, *renata.dwornicka@mech.pk.edu.pl*

²Kielce University of Technology, **Poland**

Abstract: The paper considers the use of the bootstrap method to improve the determination of confidence intervals identified by the DOE (design of experiment) procedure. Two different approaches have been used: one that is appropriate for factorial designs and the other one relevant to the methodology of the response surface. Both approaches were tested on the real experiment datasets and compared with the results obtained from the classical statistical expressions based on well-known asymptotic formulas derived from the t distribution.

Keywords: bootstrap, statistics, design of experiments, numerical simulation

1. INTRODUCTION

Conducting real experiments is usually expensive, while the implementation of industrial experiments is extremely expensive. For this reason, since the earliest beginnings of modern industry, there has been a natural desire to minimize the costs of such experiments. The development of modern science methods in the 19th century, and in particular the emergence of the applied statistics, gave industry the basic tools necessary to analyze industrial processes; however, they were still very simplistic analyzes based on the old One-Factor-At-Time scheme (OFAT), known since Galileo, while the interpretations were not very sophisticated and often naive.

The effective solution to this problem was DOE (the design of experiments) methodology, originally developed by Fisher in the 20s and 30s of the 20th century as factorial approach (Fisher, 1925; Fisher 1935). In the following years, it was extended by Yates (Yates, 1935) to two-level experimental designs, much easier to use and analyze than the Fisher's Latin squares. In the 50s of the 20th century, Box and Wilson proposed a new methodology (Box and Wilson, 1951), more appropriate to continuous factors than factorial methodology, and named it the response surface methodology (RSM). In 1958, Scheffé extended RSM into a special variant appropriate for mixtures (Scheffé, 1958), with the constant sum condition i.e. the assumption that all factors settings (percentages of mixture components) have to add

up to a constant value (usually 100%). In 1959, Kiefer and Wolfowitz, based on Kolmogorov's axiomatization, developed a complete generalized theory of DOE (Kiefer and Wolfowitz, 1959). DOE has been very successful in industrial applications, of which the most spectacular in the American Apollo lunar program.

However, it should be noted and remembered that the DOE methodology uses numerous simplifying assumptions that enable, among others, model parameters identification by the least squares method (LSQ) (Wolfberg, 2006). One of the most important assumptions is normality, non-bias and additivity of random errors. If this assumption is not met, the estimation with LSQ is not correct. Further, the normality assumption projects – through the LSQ – onto the resultant estimates and their confidence intervals: they are estimated by the asymptotic expressions based on t distribution. So, the question arises: what happens when the probability distribution does not meet the normality assumption? Three solutions are possible:

- a) a use of traditional methods with a belief that inevitable errors are small,
- b) a use of a preprocessing transformation of data e.g. Box-Cox (Box and Cox, 1964),
- c) a use of the Monte-Carlo re-sampling scheme based on raw data i.e. the bootstrap method (Shao, 1995).

The first approach is typical for those simple industrial analyzes in which a significant margin of error is acceptable or deviations from the normal distribution are of little importance. The second approach uses a specially tailored function to transform the original data set into one whose distribution is close to the normal distribution. The third approach uses the idea proposed by Efron (Efron, 1979), which allows to estimate the complete distribution of selected statistics from the data set, although the set allows to calculate directly only its single value. The details of this method are described in the next chapter. It should be also noted, that an alternative idea, derived from other mathematical assumptions, was proposed by Owen (Owen, 2001): non-parametric maximum likelihood (NPML).

2. METHODOLOGY OF THE BOOTSTRAP

The main goal is to estimate distribution of the S statistics for the random X variable which has the unknown F distribution and may be sampled. The main idea of the bootstrap method, which is oriented to achieve this goal, focuses on the following steps:

- a) the random X variable is sampled n times and creates the original D dataset,
- b) the D dataset is randomly sampled (with repetitions) n times to create the bootstrapped DB dataset,
- c) the S statistics is evaluated for the DB dataset and collected.

The sequence b-c is iterated as many times as needed until the empirical distribution of the collected statistics S reaches the desired smoothness.

The main assumption of the method is related to sampled values of X taken from the F distribution: they have to be *independent and identically distributed* (i.i.d). In the practical applications, the key problem is proper identification of the i.i.d. variable inside the mathematical model, because this variable will be the basis of the bootstrap procedure i.e. it will be the random variable X from the above description.

DOE uses two completely different models, a fixed effects model and RSM, although in both cases it uses LSQ to determine parameter. Additionally, RSM models may be identified over the fixed (nonrandom) or random factor settings. The fixed effects

model and the nonrandom RSM may be bootstrapped in the same manner by the *residual based bootstrap* (RB), while the random RSM requires different approach: the *paired bootstrap* (PB). The details are described below.

2.1. Bootstrap in fixed effects model and nonrandom RSM

The simplest fixed effects model is defined by a following formula (Montgomery, 2008):

$$y_{ijk} = \mu + \sum_{i,j} \alpha_{ij} + \varepsilon_{ijk}, \quad (1)$$

Where:

y – measurement of the response, μ - mean response, α_{ij} – effect of i -th factor at j -th level (treatment), ε_{ijk} – random error component at k -th replication. This model contains only linear effects i.e. does not include interactions between factors. More complicated models include interactions, usually two-way (second order) or three-way (third order), while higher order ones are rare.

The RSM model is defined by a following formula:

$$y = f(x_1, \mathbf{K}, x_i) + \varepsilon, \quad (2)$$

Where:

f is virtually any function of quantitative factors x_1, \dots, x_i , while ε represents the noise observed in the response y (Montgomery, 2008). Usually, due to the possibility of applying LSQ, the model includes a function being a linear expression of parameters β :

$$y = \beta_0 + \sum_j \beta_j f_j(x_1, \mathbf{K}, x_i) + \varepsilon. \quad (3)$$

In most problems, the exact form of the relationship between the response and the independent variables is unknown. There is also a significant risk that the set of independent variables is incomplete and some important factor may be omitted. Therefore, the selection of an appropriate function and set of factors is a researcher's risk. The random error component ε is the core of a bootstrap procedure i.e. i.i.d. element. The bootstrap scheme starts from typical identification of the fixed effect model usually by the least squares method. This model and its predictions will be used in the subsequent iterations of the bootstrap. Next, residuals are determined as deviations between predictions and measurements, and the dataset of residual deviations from the average residual is calculated. At this point, the initial operations end and the iteration of the bootstrap begins.

Now, a dataset of bootstrapped measurements is calculated as a sum of the model predictions and residuals randomly taken (drawn with repetitions) from the dataset of residuals deviations. The bootstrapped measurements are a base for the identification of the bootstrapped model. This model is used to determine all required bootstrapped

statistics and these statistics are collected. At this point, a single iteration of the bootstrap ends, and then the next iteration may begin.

2.2. Bootstrap in random RSM model

In contrast to the previous models, the i.i.d. component is created by a pair: a point in the design space (a combination of all factor settings) and a related measurement. It means that the original raw data are directly the source for the bootstrap draw.

The bootstrap iteration starts from the creation of the bootstrapped data randomly taken (drawn with repetitions) from the original raw data. Practically, it means that original records are invariant but their replications counters may vary, even reaching zero if the record has not been drawn. The bootstrapped model, the general form (Eq.2) or the linearized (Eq.3), and its statistics are identified from the bootstrapped data, and then collected.

3. RESULTS

3.1. Fixed effects model – ceramic shell mould of the airfoil blade casting

The raw data were obtained during investigation of the ceramic shell mould of the airfoil blade casting (Szczotok et al., 2015; Pietraszek et al., 2016). The purpose of the study was to check if the microstructure is homogeneous. The cross-section of a casting were analyzed by SEM to obtain a microstructure data. The images were analyzed by an image analysis program and quantitative data were gathered. The dataset was created by counting a number and a size of eutectic islands detected at six different traces. The one-way ANOVA was selected as a comparing method but non-gaussian distribution questioned the desirability of using the classical method i.e. assessment of the p value taken from F distribution. Due to the fact that the fixed effect model is an integral part of ANOVA, the RB bootstrap of this model was chosen. The bootstrap was used to gather full empirical distributions of the six effects. Next, this distributions were mutually compared to detect possible significant differences from zero i.e. statistical significance of considered effect. If at least one effect would be significantly different from zero, it would mean that the cross-section is not homogeneous. Fortunately, all effects turned out to be statistically insignificant.

3.2. Random RSM – strength of human vertebrae

The raw data were obtained during investigation on the development and the optimization of diagnosis methods for the estimation risk of fractures in osteoporosis based on a three-dimensional images of trabecular bones obtained *in vivo* (Czerwinski et al., 2012). The 23 vertebrae were taken from corpses and tested for compression on a universal testing machine. The stress related to 30% compression was the measured outcome. The selection of 30% as a threshold value was caused by typical compression strain observed during vertebral fracture.

Due to the natural origin, all tested vertebrae were different. It caused that all measurements, data acquired from a computer tomography (CT) before test and the stress observed during compression test, should be treated with the PB procedure. The procedure was conducted (Pietraszek and Wojnar, 2016) to check the stability of results obtained from the classic analysis. The analyzed linear model was dependent from three variables related to relative density of a trabecular bone, an average number of branches in trabecular bone and an average number of junctions on the branches in trabecular bone. The PB bootstrap included determination of distributions

for the mean response coefficient and coefficients associated with independent variables. The obtained results showed that the empirical distributions of coefficients are slightly asymmetrical, and thus their confidence intervals are asymmetrical also.

4. CONCLUSION

After the carried out investigations, the following conclusions can be drawn:

- the bootstrap is convenient method to support DOE analysis when non-gaussian data are met i.e. sample is small or a normality test failed,
- if controlled factors may be set at fixed values, then the RB bootstrap should be used – fixed effect model or RSM with fixed settings,
- if independent factors are only observed, not controlled, then the PB bootstrap should be used – RSM with random settings,
- the bootstrap approach is very convenient to automatize in computational workflow and further statistical postprocessing.

Similar methods of data-driven analysis i.e. a statistical analysis without additional assumption about distribution, may be useful in other areas like e.g. industrial management (Maszke et al., 2018), materials science (Weglowski and Osocha, 2009; Ulewicz and Novy, 2016), especially supported by an image analysis (Gadek-Moszczak, 2017; Gadek-Moszczak and Matusiewicz, 2017), even in biomaterials (Gadek-Moszczak et al., 2015), hydraulic machines design (Guzowski and Sobczyk, 2014; Walczak and Sobczyk, 2014) and rolling industry (Sygut et al., 2016). It may be also very useful in pharmaceutical and biotechnology industry (Skrzypczak-Pietraszek, 2016; Skrzypczak-Pietraszek et al., 2018), where phytochemistry investigation (Skrzypczak-Pietraszek and Pietraszek, 2009; Skrzypczak-Pietraszek et al., 2018) are conducted with a huge random noise from individual differences in plant reactions.

REFERENCES

- Box, G.E.P., Cox, D., 1964. *An analysis of transformations*. J. R. Stat. Soc. Ser. B-Stat. Methodol. B, 26, 211-246.
- Box, G.E.P., Wilson, K.B., 1951. *On the Experimental Attainment of Optimum Conditions*. J. Roy. Stat. Soc. B, 13, 1-45.
- Czerwiński, E., Gadek-Moszczak, A., Konopka, T., Latała, Z., Petryniak, R., Tabor, Z., Wojnar, L., Pietraszek, J., 2012. *Development and optimization of diagnosis methods for estimation risk of fractures in osteoporosis based on a three-dimensional images of trabecular bones obtained in vivo*. [in Polish]. Cracow University of Technology, Kraków, Poland.
- Efron, B., 1979. *Bootstrap Methods: Another Look at the Jackknife*. Ann. Statist., 7(1), 1-26.
- Fisher, R.A., 1925. *Statistical Methods for Research Workers*. Oliver and Boyd Press, Edinburgh, UK.
- Fisher, R.A., 1935. *The Design of Experiments*. Oliver and Boyd Press, Edinburgh.
- Gadek-Moszczak, A., 2017. *History of stereology*. Image Anal. Stereol., 36, 151-152.
- Gadek-Moszczak, A., Pietraszek, J., Jasiewicz, B., Sikorska, S., Wojnar, L., 2015. *The bootstrap approach to the comparison of two methods applied to the evaluation of the growth index in the analysis of the digital x-ray image of a bone regenerate*. New Trends in Comput. Collective Intelligence, 572, 127-136.

- Gadek-Moszczak, A., Matusiewicz, P., 2017. *Polish Stereology – a Historical Review*. *Image Anal. Stereol.*, 36, 207-221.
- Guzowski A, Sobczyk A., 2014. *Reconstruction of Hydrostatic Drive and Control System Dedicated for Small Mobile Platform*. ASME – Fluid Power Systems Technology, 8th FPNI Ph.D Symposium on Fluid Power, art.V001T05A012.
- Kiefer, J., Wolfowitz, J., 1959. *Optimum Designs in Regression Problems*. *Ann. Math. Stat.*, 30, 271-294.
- Maszke, A., Dwornicka, R., Ulewicz, R., 2018. *Problems in the implementation of the lean concept at a steel works – Case study*. MATEC Web Conf., 183, art. 01014.
- Montgomery, D.C., 2008. *Design and analysis of experiments*. Wiley, Hoboken.
- Owen, A.B., 2001. *Empirical Likelihood*. Chapman & Hall/CRC, Boca Raton.
- Pietraszek, J., Dwornicka, R., Szczotok, A., 2016. *The Bootstrap Approach to the Statistical Significance of Parameters in the Fixed Effects Model*. ECCOMAS Congress 2016 – Proc. 7th European Congress on Computational Methods in Applied Sciences and Engineering, Vol.3, 6061-6068.
- Pietraszek, J., Wojnar, L., 2016. *The Bootstrap Approach to the Statistical Significance of Parameters in RSM Model*. ECCOMAS Congress 2016 – Proc. 7th European Congress on Computational Methods in Applied Sciences and Engineering, Vol.1, 2003-2009.
- Scheffé, H., 1958. *Experiments with Mixtures*. *J. Roy. Stat. Soc. B*, 20, 344-360.
- Shao, J., Tu, D., 1995. *The Jackknife and Bootstrap*. Springer, New York.
- Skrzypczak-Pietraszek, E., 2016. *High production of flavonoids and phenolic acids for pharmaceutical purposes in Vitex agnus castus L. shoot culture*. *New Biotechnol.*, 33, S155-S155.
- Skrzypczak-Pietraszek, E; Hensel, A., 2000. *Polysaccharides from Melittis melissophyllum L. herb and callus*. *Pharmazie*, 55, 768-771.
- Skrzypczak-Pietraszek, E., Pietraszek, J., 2009. *Phenolic acids in in vitro cultures of Exacum affine Balf. f.* *Acta Biol. Cracov. Bot.*, 51, 62-62.
- Skrzypczak-Pietraszek, E., Reiss, K., Zmudzki, P., Pietraszek, J., 2018. *Enhanced accumulation of harpagide and 8-O-acetyl-harpagide in Melittis melissophyllum L. agitated shoot cultures analyzed by UPLC-MS/MS*. *PLoS ONE* 13, e0202556.
- Sygut, P., Klimecka-Tatar, D., Borkowski, S., 2016. *Theoretical analysis of the influence of longitudinal stress changes on band dimensions during continuous rolling process*. *Arch. Metall. Mater.*, 61, 183-188.
- Szczotok, A., Nawrocki, J., Gadek-Moszczak, A., Kołomycki, M., 2015. *The Bootstrap Analysis of One-Way ANOVA Stability in the Case of the Ceramic Shell Mould of Airfoil Blade Casting*. *Solid State Phenomena*, 235, 24-30.
- Ulewicz, R.; Novy, F. R., 2016. *The influence of the surface condition on the fatigue properties of structural steel*. *Journal of the Balkan Tribological Association*, 22, 1147-1155.
- Walczak, P., Sobczyk A., 2014. *Simulation of water hydraulic control system of francis turbine*. ASME – Fluid Power Systems Technology, 8th FPNI Ph.D Symposium on Fluid Power, art. V001T04A001.
- Weglowski, B., Osocha, P., 2009. *Modelling of creep for Y pipe from ferritic-martensitic P91 steel*. *Rynek Energii*, 6, 140-145.
- Wolfberg, J., 2006. *Data Analysis Using the Method of Least Squares. Extracting the Most Information from Experiments*. Springer, Berlin-Heidelberg.
- Yates, F., 1935. *Complex experiments*. *J. Roy. Statist. Soc., Suppl.*2, 181-247.