

STOCHASTIC OPTIMIZATION OF LABORATORY TEST WORKFLOW AT METALLURGICAL TESTING CENTERS

Received – Prispjelo: 2015-11-06

Accepted – Prihvaćeno: 2016-04-30

Preliminary Note – Prethodno priopćenje

The objective of the paper is to present a way to shorten the time required to perform laboratory tests of materials in metallurgy. The paper finds a relation between the time to perform a test of materials and the number of technicians carrying out the test. The relation can be used to optimize the number of technicians. The approach is based on probability theory, as the amount of material to be tested is unknown in advance, and uses powerful modelling techniques involving the generalized estimating equations.

Key words: metallurgy, laboratory test, testing time, dependent data, generalized linear model

INTRODUCTION

In metallurgy, quality of materials and final products is heavily tested [1]. The tests involve control of mechanical and other properties of the products, and are carried out in laboratories by certified personnel. The laboratories are often owned by metallurgical companies, which is one of the reasons why the companies urge the laboratories to test quality fast so that product distribution is not delayed. Any delays mean additional costs because the firms fund their daily activities with short-term bank credits, and they also rent freight trains to transport the output [2]. The same companies, however, do not pay much attention to whether it is possible at all to run the tests swiftly without sacrificing their precision. The precision can be ensured, but sometimes only at the cost of hiring more technicians. In that case, a proper number of technicians must be determined, which is not a straightforward task because it depends on the amount of work to be done, and that amount fluctuates randomly. Due to the randomness, one can only hope of estimating the proper number of technicians in such a way that there will be a low probability of not getting the work done in time. To achieve this objective, it is convenient to explore the mechanisms of testing embedded in the historical data of a laboratory, and use the mechanisms found to set up a mathematical model which could describe the testing time as a function of the amount of products tested. Such a model may then imply how many technicians are needed. The aim of the paper is to formulate such a model, based on experience from a specific testing environment in a North Moravian metallographic laboratory. One of the features of this

environment was the way the laboratory stored its data on testing because the storage created dependences in the data. The paper is divided into several sections. In the first section, the process of testing in the laboratory is described. In the following sections, models for testing times are presented, discussed and compared through simulation.

TESTING IN THE LABORATORY

The observed laboratory received in the course of time doses of production to be tested on quality. Each dose contained product samples. When a dose reached the storage room of the laboratory, it was divided into two batches of the same size so that cluttering was avoided during the testing. After the samples were pre-processed in line with a technical norm, the first batch was transferred to the laboratory and underwent an examination. The time during which the samples resided in the laboratory until their testing was over was recorded in a computer. Next, the second batch arrived at the laboratory, but wasn't tested immediately, because the staff finished tasks on the first batch, as well – it recorded the testing time also on paper, it removed the tested samples, and so on. During this pause, however, the time of testing the second batch had begun. This pause was proportional to the size of the first batch. One might say that a necessary delay occurred when the second batch was about to be processed by the laboratory. On the other hand, the second batch was not encumbered with the preparatory work that accompanied the first batch, which involved a more complex initiation of microscopes, labour division, etc. Figure 1 shows the timeline of how the two batches were processed. The handling itself of the first batch lasted from a time point A to a time point B, and this was the time recorded for the batch. As for the second batch, its recorded time

F. Tošenovský, D. Vykydal, P. Klaput, P. Halfarová, VŠB – Technical University of Ostrava, Faculty of Metallurgy and Materials Engineering, Ostrava, Czech Republic

spanned the period from the point B to a point D, although the batch was handled from a point C to the point D. The point C was not recorded. The times B – A and D – C could be regarded as independent.

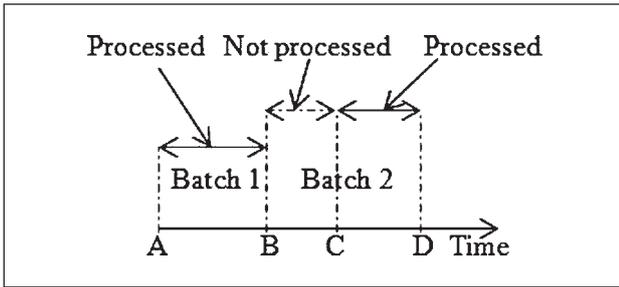


Figure 1 Timeline of processing two batches making up a single production dose

The laboratory was required to test each batch so that the test wouldn't last longer than specified. Figure 2 shows the character of a sequence of batches to be processed. Each rectangular represents a batch, the rectangulars of the same type describe the batches of the same size.

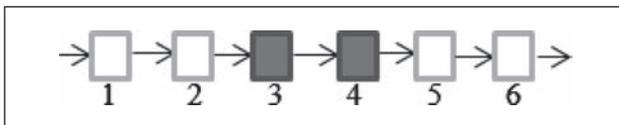


Figure 2 Sequence: each rectangular is a batch, two neighbouring batches form a dose

Each batch contained samples the technicians tested simultaneously. The technicians were also certified and experienced similarly, and so they differed from one another negligibly in terms of how much time each of them needed to test a single sample. These facts point to the principles that determine the rate of work acceleration when the number of technicians is altered. To illustrate the principles, let us imagine a batch of eleven samples. If there are three technicians in the laboratory, the first three samples will be processed in t_1 hours. It can also be said that each technician processes a sample in t_1 hours because of the simultaneity in their work. A similar time can be expected for the second triplet, the third triplet and the final pair of samples. Thus, the batch can be expected to be processed in about $4 \cdot t_1$ hours. If the number of technicians is raised to four, for instance, four samples will be processed at virtually the same time, and the batch will be processed in about $3 \cdot t_1$ hours. This way, one can imagine batches of different sizes and various changes in the number of technicians, and set up a function $f(k, x)$ of two variables, where k is a new number of technicians and x is a batch size. The function would calculate the work acceleration for given values of the two variables. We shall refer to it as the *conversion function*.

There is also another aspect related to the observed process. Denoting the *recorded* time of processing the

i -th batch by Z_i , and using Y_i to denote the time of *actually* processing it, we may write for the batches from the same dose: $Z_1 = Y_1, Z_2 = Y_2 + \varphi \cdot Y_1$, where $\varphi \cdot Y_1$ reflects the fact that processing the second batch depends on how long the first batch is processed because this affects the necessary delay time C - B. If the Y_i 's, $i = 1, 2$, are independent random variables, the covariance of Z_1 and Z_2 is $cov(Z_1, Z_2) = \varphi \cdot var(Y_1)$. This implies the recorded times are not independent for $\varphi \neq 0$, a fact that can be exploited when building a model that would describe the testing time as a function of the batch size. It will be discussed in the following text.

MODEL FOR INDEPENDENT DATA

In this section, we shall describe a model that determines how many technicians are needed to process batches on time with a high probability. The recorded times needed to process the batches are assumed to be independent random variables first, as this serves well as a starting point for a more complex situation. Since the processing times of batches from the same dose were similar, we also assume their distributions are the same. All times are modelled with gamma distributions which are typically used for these and similar purposes. The relation for the required number of technicians can now be built within the framework of generalized linear models [3]. Let $Z(x_i) \sim \Gamma(\mu(x_i), \phi)$, $i = 1, 2, \dots, m$, where $Z(x_i)$ is the recorded time of processing a batch of size x_i , and $\mu(x_i) = E[Z(x_i)]$ and ϕ are the two parameters of the distribution. The effect the batch size has on its recorded processing time is described by a link function which, for gamma distributions, may be logarithmic:

$$\ln \mu(x_i) = \beta_0 + \beta_1 \cdot x_i, i = 1, 2, \dots, m \tag{1}$$

Knowing the historically recorded times z_{ij} , the symbol denoting the recorded processing time for the j -th batch of size x_i , the unknown parameters β_0, β_1, ϕ of the model can be estimated by the maximum likelihood method, i.e. by maximizing the parameter function

$$g(\beta_0, \beta_1, \phi) = \sum_{i,j} \ln f(z_{ij}, \mu(x_i), \phi), \tag{2}$$

where the function $f(z, \mu(x_i), \phi)$ is the density of the gamma distribution. This problem can be solved in Stata software with the command [4]:

$$glm \ z \ x, \text{family}(\text{gamma}) \ \text{link}(\text{log})$$

We note that this calculation is based on the *existing* number of technicians k . Having the estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\phi}$, it is possible to estimate the probability that the recorded time Z exceeds an upper time limit L . If the batch size X is considered a random variable, taking on values x_1, x_2, \dots, x_m , the probability $P(Z > L)$ equals

$$P\left\{\bigcup_{i=1}^m ((Z > L) \cap (X = x_i))\right\} \tag{3}$$

$$= \sum_{i=1}^m P\{(Z > L) \cap (X = x_i)\} \tag{4}$$

$$= \sum_{i=1}^m P\{(Z > L | X = x_i)\} \cdot P\{X = x_i\} \tag{5}$$

$P \{X = x_i\}$ can be estimated by the relative frequency of occurrence of the batch size x_i in the past, whereas $P \{Z > L | X = x_i\}$ can be calculated with $f(z, \hat{\mu}(x_i), \hat{\phi})$.

All that remains is to determine a new number of technicians k^* , so that $P \{Z > L\}$ is low. For a given k^* , the time $Z(x_i)$ shifts to time $f(k^*, x_i) \cdot Z(x_i)$, where the term $f(k^*, x_i)$ is the conversion function, and so we obtain $f(k^*, x_i) \cdot Z(x_i) \sim \Gamma(f(k^*, x_i) \cdot \hat{\mu}(x_i), \hat{\phi})$, given the properties of any gamma distribution. The new density can now be used again to calculate $P \{Z > L\}$ for the new number of workers.

MODEL FOR DEPENDENT DATA

The model just described can be improved. We work with m data panels, the i -th of which is defined by the batch size x_i , and contains processing times $z_{ij}, j = 1, 2, \dots, n_i$. In the i -th panel, the times z_{i1}, \dots, z_{in_i} are realizations of random variables $Z_1(x_i), \dots, Z_{n_i}(x_i)$ from the distribution $\Gamma(\mu(x_i), \phi)$. We may form a random vector $\mathbf{v}(x_i) = (Z_{k_1}(x_i), \dots, Z_{k_{n_i}}(x_i))$ from these variables, so that the first and second component of the vector relate to the same dose of size $2 \cdot x_i$, and the same is true for its third and fourth component, fifth and sixth, etc. To see this, refer to Figure 2. In this case, the first and second batch times would be the first two components of the vector, and the fifth and sixth batch times would be its third and fourth component. This representation of a panel is convenient because the vector $\mathbf{v}(x_i)$ has the same correlation matrix, regardless of the panel:

$$\text{corr } \mathbf{v}(x_i) = \begin{pmatrix} 1 & \phi & 0 & \cdot & \cdot & 0 \\ \phi & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \phi & \cdot & 0 \\ \cdot & \cdot & \phi & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \phi \\ 0 & 0 & 0 & 0 & \phi & 1 \end{pmatrix} \quad (6)$$

Due to the constant form (6), the parameters β_0, β_1, ϕ can be estimated by the generalized estimating equations (GEE) [5], which utilize the information embedded in the correlation. The method may give more accurate estimates of β_0, β_1, ϕ [6,7]. Once the estimates are obtained, the resulting model can be used the same way as before to determine the necessary number of technicians. In Stata, this version of GEE is run by typing the command:

```
xtgee z x, family(gamma) link(log)
corr(nonstationary 1).
```

SIMULATION

It is convenient to compare the two approaches just described, therefore a standard simulation procedure was also run [8] in this case. A thousand sequences of batches, each the length of four hundred batches, was generated. Also, different batch sizes were randomly se-

lected. The size, as suggested by the historical data from the analysed laboratory, varied from six to thirty - two samples. The expected value of the recorded processing time related to the smallest batch was set to two hours, for other batch sizes, the value was shifted using the conversion function and a normally distributed noise with a small variance. The parameter was set to values 0,1, 0,3 and 0,5. The simulation showed that when at least 6 panels were present, and was at least 0,3, the model based on the dependent data led to parameter estimates with smaller variances. This suggests that a fairly low correlation can already justify the use of the GEE - based model here because it leads to more precise estimates of the model, the precision being measured by variance. The models based on independent and dependent data may as well be compared in practice, using the QIC criterion, a generalized version of the Akaike's Information Criterion [9].

CONCLUSION

In order to guarantee smooth quality testing of materials in metallurgy, testing laboratories must have, aside from good management practices [10] and proper measuring equipment [11], enough certified personnel at hand, which depends on the amount of testing to be performed. Since that amount is unknown in advance because of its random fluctuation, a stochastic model should be established to estimate how many workers are needed at a laboratory. Several models of this sort were presented in the paper, drawing on findings from a specific metallographic laboratory, and performance of the models was compared through standard simulation techniques. It turns out that reflecting dependence in the data will often be a better approach to building the model for the working environment similar to the one analysed.

Acknowledgement

This paper was elaborated within the framework of the specific research project SP2016/91 which has been solved at the Faculty of Metallurgy and Materials Engineering, VŠB – TU Ostrava with a support of the Ministry of Education, Youth and Sports, the Czech Republic.

REFERENCES

- [1] American Society for Testing and Materials, 2010 Annual Book of ASTM Standards – Volume 03.01 Metals – Mechanical Testing; Elevated and Low-Temperature Tests. A.S.T.M., West Conshohocken, Pennsylvania, 2010.
- [2] P. Lyons, Corporate Lending. Institute of Financial Services, Canterbury, Great Britain, 2004.
- [3] J. W. Hardin, J. M. Hilbe, Generalized Linear Models and Extensions. Stata Press, College Station, Texas, 2012.
- [4] U. Kohler, F. Kreuter, Data Analysis Using Stata. Stata Press, College Station, Texas, 2012.

- [5] J. W. Hardin, J. M. Hilbe, *Generalized Estimating Equations*. CRC Press, Boca Raton, Florida, 2013.
- [6] K. Y. Liang, S. L. Zeger, *Longitudinal Data Analysis Using Generalized Linear Models*, *Biometrika* 73 (1986) 1, 13-22.
- [7] A. Ziegler, C. Kastner, U. Gromping, M. Blettner, *The Generalized Estimating Equations in the Past Ten Years: An Overview and Biomedical Application*, Collaborative Research Center 386, Discussion Paper 24 (1996), 1-15.
- [8] R. Y. Rubinstein, D. P. Kroese, *Simulation and the Monte Carlo Method*. A John Wiley and Sons, Hoboken, New Jersey, 2008.
- [9] W. Pan, *Akaike's Information Criterion in Generalized Estimating Equations*, *Biometrika* 57 (2001) 1, 120-125.
- [10] K. Zgodavová, P. Bober, *An Innovative Approach to the Integrated Management System Development, Quality, Innovation, Prosperity* 16 (2012) 2, 59-70.
- [11] K. Zgodavová, *Proceedings, 21st International DAAAM Symposium, 2010*, B. Katalinic (ed.), vol. 21, DAAAM International, Vienna, 2010, pp. 0183.
- Note:** The responsible translator for English language is Dr. Karolina Slamová, Department of Languages, VŠB – Technical University of Ostrava, Czech Republic