Journal of Business; ISSN 2233-369X; Volume 3, Issue 1, 2014

The Generalized Technique for Evaluation of Quality Level of Manufacturing, Business and Educational Processes

Ahmet DEMIR*
Irakli RODONAIA**

Abstract

The requirements for production and learning process quality are different in various manufacturing, business and educational organizations. A new approach to fit these requirements and evaluate the closeness of realistic (actual) quality of production or learning processes (based on quality indicators of output or scores of examination tests) is proposed in the paper. The technique uses the strictly defined approximation procedures and allows users automatically evaluate closeness of actual quality level when quality requirements change.

Keywords: business processes quality, generalized lambda distribution, learning process quality, manufacturing processes quality, non-parametric approximation, percentile function
JEL: C46

Introduction

Let us assume that we are given the next requirement for the learning process quality: "weak" (failed) students can be thought those ones whose grades are less than 60 and the percentage of them should be 30%; "ordinary" (of acceptable level) students are those ones whose grades are between 61 and 95 grades, the percentage of them should be 65%; in latter range so called "middle" level students are those whose percentage is no more than 50% of total number of students (including failed ones) and 20% of "ordinary" students; say, the grade of these "middle" students turns out to be 80 (or any other value), so the grade 80 can be considered as a median of grades distribution; "excellent" students are those ones whose grades are above 95% and the percentage of them is 5%. The corresponding cumulative distribution function (CDF) is shown in Figure 1. Obviously, the pattern distribution cannot be approximated by the normal distribution.

However, in all known for us papers such distributions were approximated by either the normal distribution or by some another well-known distributions (beta distribution, gamma distribution, Weibull distribution, etc.) (Rohwer, 2012; Schwarz, 2011; Milnikova, 2012). But in case of applying normal distribution the adequacy and precision of results strongly depends on the degree of "skewness" and often may not be acceptable. In case of applying other distributions (beta distribution, gamma distribution, Weibull distribution, etc.) the problem of estimating adequate distribution parameters arises. In many cases analytical expression cannot be obtained in close form. Besides, when requirements for quality changes, the corresponding shapes of PDF and CDF functions also change. As a result, it is necessary to use frequently complicated procedures of distribution parameters estimation.

The similar task is commonly met in the area of product quality control (Manzini, Regattieri, Pham, & Ferrari, 2010). Suppose that the quality requirement to the product quality is as follows. The percentage of deviation from required level of some quality parameter must be no more than ±5% in 95 % of the output of the product; in this case the quality of the product is regarded as "excellent". To be regarded as "acceptable" the product quality must be as follows: deviation from required level of the quality parameter is ±6%-20% in 3% of the output of the product. The product quality is regarded as "unacceptable" (or defective) if there is the de-

* MA, Ishik University, Erbil, Iraq
E-mail: ahmet.demir@ishik.edu.iq

** Prof. Dr., Faculty of Computer Technologies and Engineering, International Black Sea University, Tbilisi, Georgia
E-mail: irakli.rodonaia@ibsu.edu.ge

Journal of Business; ISSN 2233-369X; Volume 3, Issue 1, 2014

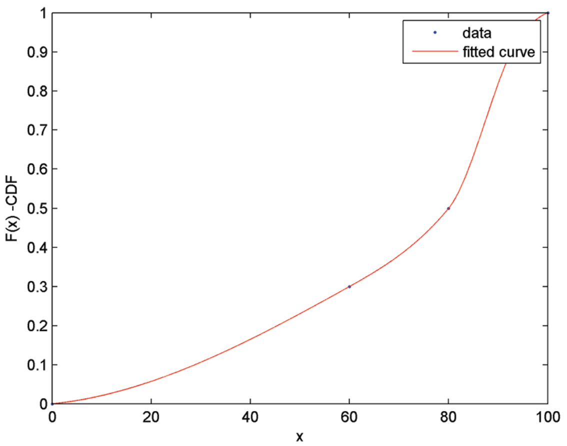

Figure 1. *CDF of pattern distribution*

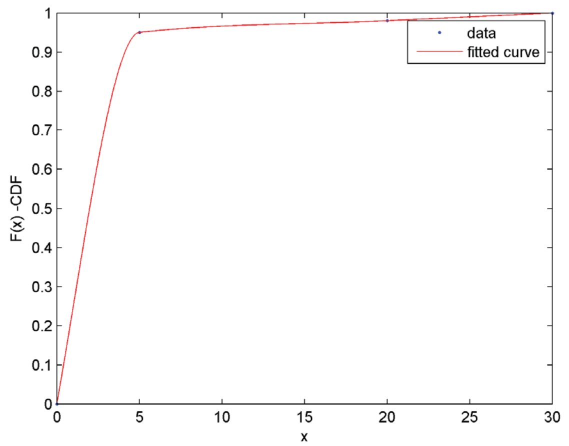

Figure 2. *CDF of pattern distribution in product quality control*

viation of more than 20% (so the percentage of defective production must be no more than 2%).The CDF is shown in Figure 2. As one can see, the distribution is also skewed and the problem of choosing the right type of distribution occurs here.

It is not clear in advance which type of distribution should be used in this case. The above given distributions (reflecting quality requirements) are called hereinafter "pattern" distributions (functions). It is desirable that distribution of grades of actual exams is as close to the pattern distribution as possible. The question of closeness degree is a problem (and is considered further in the paper). Moreover, the pattern distribution presents quality requirement for total learning process (which must take into account results of all relevant tests).

Grades of many subjects (obtained by a group of students in tests held during one of morecourses) must match the pattern distribution in order to get group regarded as successful and meeting the requirements of learning quality. Of course, it is possible to compare grades of each actual test with the pattern distribution and then summarize the results. But this approach is associated with a large amount of additional and repeated calculations.

Taking into account all the above-mentioned, a new general method of using a unified non-parametric (Erceg-Hurn & Mirosevich, 2008) estimation of relevant grades distributions and further application of its results to the evaluation process of learning quality is developed in this paper. It is important to point out that the method does not require the execution of rather complicated procedures of estimating distribution parameters (mean, standard deviation, third and fourth moments)). The method can be applied to fit grades of various multiple tests and compare them with pattern distribution by using the same unified techniques and algorithms. The approach provides forming of overall quality criterion for all test scores and method of comparing it with pattern quality requirement.

General part

To provide fitting the wide variety of distribution shapes and to describe data by using a single functional form the approach used in the paper implements the Generalized Lambda Distribution (GLD)[6].The method specifies four parameter values for each case, instead of giving the basic data (which is what the empirical distribution essentially does) for each case. The one functional form allows us to group cases that are similar, as opposed to being overburdened with a mass of numbers or graphs.

The generalized lambda distribution family with parameters λ1, λ2, λ3, λ4, GLD (λ1, λ2, λ3, λ4), is most easily specified in terms of its quantile or percentile function.

$$Q(y) = Q(y;\lambda_1,\lambda_2,\lambda_3,\lambda_4) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}$$

where $0 \le y \le 1$. The parameters λ1 and λ2 are, respectively, location and scale parameters, while λ3 and λ4 determine the skewness and kurtosis of the GLD (λ1, λ2, λ3, λ4). Recall that percentile function (PF) of the stochastic variable X is the function Q(y) which, for each y between 0 and 1, tells us the value of x such that F (x) = y: Q (y) = (The value of x such that F (x) = y), $0 \le y \le 1$

Here F(x) is the cumulative distribution function (CDF) of the variable X:

$$F(x) = P(X \le x), -\infty < x < +\infty.$$

The restrictions on λ1, λ2, λ3, λ4that yield a valid GLD(λ1, λ2, λ3, λ4) distribution and the impact ofλ3 andλ4 on the shape of the GLD(λ1, λ2, λ3, λ4) PDF (Probability Density Function) will be considered later.

It is relatively easy to find the probability density function

Journal of Business; ISSN 2233-369X; Volume 3, Issue 1, 2014

from the percentile function of the GLD (Karian, 2011). For the GLD (λ1, λ2, λ3, λ4), the probability density function is:

$$f(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1-y)^{\lambda_4 - 1}}$$

at x=Q(y)

As we have seen above, very often the quality requirements are given in the form of required percentiles (percent of failed, ordinary, middle and excellent students, percent of deviation of some product's quality parameters from their nominal values and so on). The percentile-based approach [1] fits a GLD(λ1, λ2, λ3, λ4) distribution to a given dataset by specifying four percentile-based sample statistics and equating them to their corresponding GLD (λ1, λ2, λ3, λ4) statistics. The resulting equations are then solved for λ1, λ2, λ3, λ4, with the constraint that the resulting GLD would be a valid distribution.

The method, described above, requires usage of the complex tables of various values of parameters λ3 and λ4. To automate the fitting process the **algorithm P-KS** (Fournier, Rupin, Bigerelle, Najjar, & Iost, 2011) is used in the paper. The strategy is to find the set of parameters (λ1, λ2, λ3, λ4) that give the lowest value of the Kolmogorov-Smirnov estimator E_{KS}:

$$E_{ks} = \max \left| \widehat{F_n} - F(x) \right|$$

where $\widehat{F_n}$ is the empirical cumulative distribution function (ECDF).

As it was stated above, the pattern distribution is given in the form of some percent. For the example of the section we have the following data (expressed in the form of Matlab statements):

x= [0, 60, 80, 95,100];

y= [0, 0.30, 0.50, 0.95, 1];

In order to form the pattern distribution (with which the actual tests grades should be compared) we need to fit a curve to the given data. The fitted curve will be used to generate data values in intermediate points (other than the original data points) -interpolation points. To provide the smoothness and maximum accuracy of generated data in interpolation points the technique of the shape-preserving cubic splines is used. The plot of the ECDF for pattern distribution looks like (Figure 3). The corresponding PDF function can be obtained similarly and is shown in Figure 4.

As one can see, the shape of the PDF is non-standard and it is difficult to guess which theoretical distribution can successfully fit it.

Now we can estimate (using relevant Matlab statements) values of the pattern distribution in interpolation

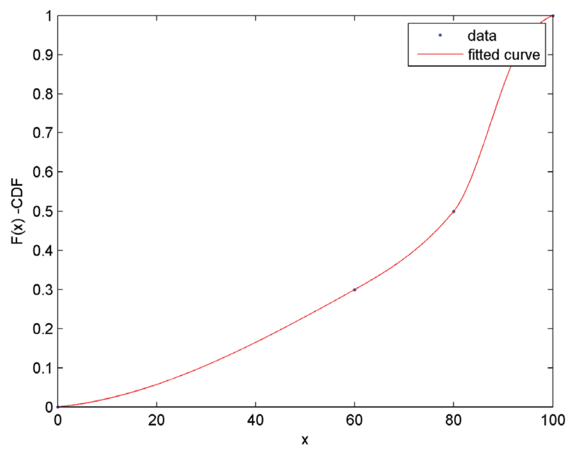

Figure 3. *ECDF for pattern distribution*

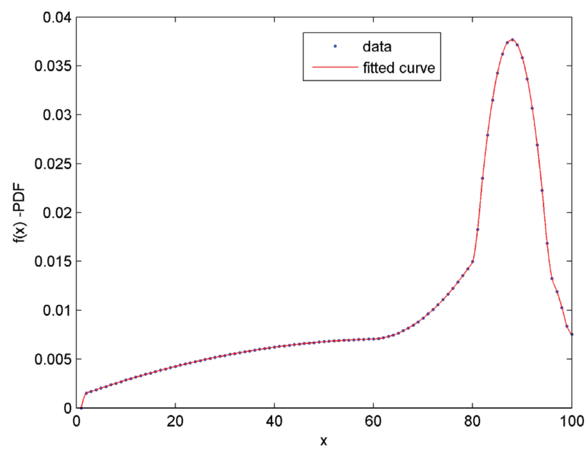

Figure 4. *PDF function for pattern distribution*

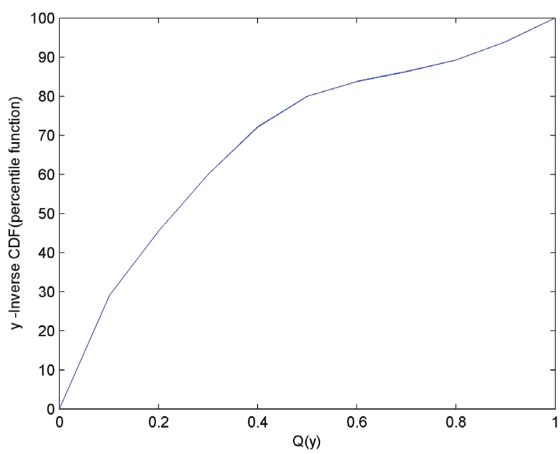

Figure 5. *PCHIP to estimate values of ICDF*

points, that is, we can estimate the values of various percentiles (namely, 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th percentiles) of the pattern distribution to be compared with actual tests grades' percentiles. As we stated above, the GLD Percentile-Based Approach to Fitting Distributions intensively uses operations with percentile functions PF (inverse cumulative distribution functions ICDF). We can compute a nonparametric estimate of the inverse CDF. In fact, the inverse CDF estimate is just the CDF estimate with the axes swapped. Here we again use the Piecewise Cubic Hermite Interpolant Polinomial (PCHIP) to estimate values of ICDF (Figure 5).

Having values of PF we can compute now the values of $\widehat{\rho_1}, \widehat{\rho_2}, \widehat{\rho_3}$ and $\widehat{\rho_4}$. Having computed these values, we now run the procedure P-KS. The solution with the best KS criteria for all possible combinations of pairs (λ3, λ4) and associated with them pairs of (λ1, λ2) are selected. As it was explained above, knowing λ1, λ2, λ3, λ4 and using formulas (1) and (2) , we can build the PDF curve: we take a grid of y values (such as .01, .02, .03, . . ., .99, that give us the 1%, 2%, 3%, . . ., 99% points), find x at each of those points from (1), and find f(x) at that x from (2). Then, we plot the pairs (x, f(x)) and link them with a smooth curve.

This pattern distribution will be compared with actual test grades in the following manner: for samples of each test grades the GLD fitting technique described above will be applied, as a result we will obtain data, corresponding lambdas and curves for each test. Using the special desirability (Trautmann, 2009) function, we will create single integrated PDF curve (which represent PDF curves of all actual tests). The integrated PDF curve should be compared with the pattern PDF curve obtained above. To determine the closeness (or distinction) of distribution functions (and, thereby, determine the quality of learning process) we will use Kullback–Leibler Divergence (Perez-Cruz & Kullback-Leibler, 2008) (this is the subject of the future research).

Conclusion

• The problem of evaluation of manufacturing, business and learning processes is defined.

• The needs to use non-parametrical approximation methods are shown.

• The new approach to the above problems is formed and described.

• This approach might be used in manufacturing, Business and Educational fields.

References

Erceg-Hurn, D. M., & Mirosevich, V. (2008). Modern robust statistical methods. University of Western Australia. *Amercan Psychologist*, 63(7), 591–601.

Fournier, B., Rupin, N., Bigerelle, M., Najjar,D., & Iost, A. (2011). Fitting a generalized lambda distribution using a percentile-KS (P-KS) Adequacy Criterion.

Karian, Z. A. & Dudewicz, E.J. (2011). *Handbook of fitting statistical distributions with* R. CRC Press.

Manzini, R. A., Regattieri, H., Pham, E., & Ferrari (2010). Maintenance for industrial systems. Springer.

Milnikova, I. (2012). *Elaboration of statistical quality control models in education process*. (PhD thesis, Georgian technical University, Georgia).

Perez-Cruz, F. & Kullback-Leibler. (2008). *Divergence estimation of continuous distributions, department of electrical engineering*. Princeton University, New Jersey.

Rohwer, G. (2012). *Statistical methods in sociological research of education*. Ruhr University Publish, Germany.

Schwarz, J. (2011). *Sampling, regression, experimental design and analysis for environmental scientists, biologists, and resource managers*. Simon Fraser University, Canada