Journal of Quality Technology
A Quarterly Journal of Methods, Applications and Related Topics
ISSN: 0022-4065 (Print) 2575-6230 (Online) Journal homepage: http://www.tandfonline.com/loi/ujqt20

Discussion

Rob Goedhart, Marit Schoonhoven & Ronald J. M. M. Does

To cite this article: Rob Goedhart, Marit Schoonhoven & Ronald J. M. M. Does (2018) Discussion, Journal of Quality Technology, 50:1, 17-19, DOI: 10.1080/00224065.2018.1404886

To link to this article: https://doi.org/10.1080/00224065.2018.1404886

Published online: 01 Feb 2018.

JOURNAL OF QUALITY TECHNOLOGY
2018, VOL. 50, NO. 1, 17–19
https://doi.org/10.1080/00224065.2018.1404886

Discussion

Rob Goedhart, Marit Schoonhoven, and Ronald J. M. M. Does

Department of Operations Management, University of Amsterdam, Amsterdam, The Netherlands

1. Introduction

As mentioned in Jensen et al. (2018), JQT has always had a focus on real-world problems and solutions that can be applied by practitioners. One of the fields of interest in JQT where large developments have been observed is statistical process monitoring (SPM). To this end, we elaborate further on the trends and developments that are encountered in JQT in the field of SPM in this discussion.

2. Trends and developments in statistical process monitoring

At the time the first edition of *JQT* was published, researchers were already aware that parameter estimation has an effect on control chart performance. The original Shewhart control chart based on three-sigma limits lies at the basis of SPM. In the classical setting, data from the process under consideration are assumed to be i.i.d. normally distributed variables with mean μ and standard deviation σ. Based on these assumptions, an in-control process is expected to produce about 2.7 false alarms in every 1,000 measurements. However, even though the assumptions can be summarized in a single sentence, they are often violated in practice, which leads to undesirable control chart performance. Jensen et al. (2018) argue that it is extremely rare to see an SPM article do any meaningful assessment of the assumptions involving the real data from an actual process. Although we do agree that such assessments are generally lacking, we observe a development of the published articles where methods become less dependent on these assumptions.

In practice, μ and σ are generally unknown and need to be estimated. The effects of estimating the parameters on the control chart performance depend on the amount of available Phase I data as well as on possible contaminations (outliers) in the data. In addition, variables often contain serial correlation and are thus not independent. Furthermore, the assumption of normality, or a known distribution in general, is an extremely strong one. Violations of any of these assumptions lead to a deterioration in control chart performance. We discuss the developments and publications in JQT concerning these issues and focus mainly on Shewhart $\bar{X}$-charts throughout this article. However, similar developments are observed for other charts.

2.1. Estimated parameters

In the first edition of *JQT*, Hillier (1969) shows that the overall probability of false alarms for $\bar{X}$ and R charts differs from the anticipated nominal value (i.e., 0.0027 for three-sigma limits) when only a small number of subgroups is available. In a later article, Yang and Hillier (1970) build on this work and propose to replace the average range $\bar{R}$ by more efficient estimators of dispersion, such as the pooled standard deviation. Subsequently, more articles were written on the considerations when choosing between different estimators (e.g., Schoonhoven et al. 2011).

Although robust and efficient estimation is definitely important in practice, this alone does not solve all of the problems associated with parameter estimation. An influential article in the field of SPM to this end is Quesenberry (1993), who pointed out the dependency between consecutive false alarm probabilities for individual points when parameters are estimated. This means that, in contrast to the known parameters case, the run length distribution is not geometric. He suggests the use of at least $400/(n-1)$ Phase I samples in order for the control charts to behave like the known parameters situation. In Jensen et al. (2006), these and other findings are discussed in an overview article on the effects of parameter estimation in SPM.

The need for review and overview articles as mentioned in Jensen et al. (2018) also became clear here, as the authors introduced large changes to how control charts were evaluated. Originally, performance measures

CONTACT Ronald J. M. M. Does r.j.m.m.does@uva.nl Department of Operations Management, University of Amsterdam, Plantage Muidergracht 12, 1018 TV Amsterdam, The Netherlands.

consisted of unconditional performance measures such as the unconditional false alarm rate (FAR) or the unconditional average run length (ARL). Recently, more emphasis has been placed on conditional performance. Even when sampling from the same distribution, different samples will lead to different parameter estimates and, consequently, different control limits and control chart performance. This variability is sometimes also referred to as practitioner-to-practitioner variability (see e.g., Saleh et al. 2015a). With this thought in mind, measures such as the unconditional ARL or unconditional FAR make much less sense because they only show how a control chart performs in expectation, but say nothing about the individual performances. Saleh et al. (2015a) illustrate the conditional performance of $\bar{X}$ control charts and show that the earlier sample size requirements of $400/(n-1)$ samples from Quesenberry (1993) are not sufficient.

On occasions where the required sample sizes for a sufficient conditional control chart performance are not available, another solution is to adjust the control limits accordingly. The idea is to guarantee a minimum conditional in-control control chart performance (in terms of conditional FAR or conditional ARL) to practitioners with a prespecified probability. Saleh et al. (2015b) provide a bootstrap approach to determine the adjusted control limits for given sample sizes. In the case of normally distributed data, Goedhart Schoonhoven, and Does (2017) provide analytical expressions for the adjusted limits.

2.2. Autocorrelated data

Another topic where the interest of *JQT* in real-world problems becomes clear is methods on how to deal with autocorrelated data. In many practical applications, data are not independently distributed but instead display some degree of autocorrelation. Vasilopoulos and Stamboulis (1978) is one of the first articles in *JQT* that considers control charts for process data that contain serial correlation. They use an autoregressive (AR) model to determine adjustments required to the control limits. Later works on autocorrelated data also incorporate AR models of general order p (AR(p)). Similarly to the estimation of the process location and dispersion, the actual order p is generally unknown and has to be estimated. For other models under consideration, such as ARMA or ARIMA, the number of required time-series parameters that need to be estimated is even greater. Adams and Tseng (1998) and Lu and Reynolds (1999, 2001) investigate the consequences of estimating the time-series parameters in several residual control charts. They find that estimating the time-series parameters severely decreases the control chart performance. As already mentioned by Jensen et al. (2006), the effect of model misspecification in combination with the estimation of process parameters has not yet been studied and still requires further investigation. Again, this illustrates the importance of discussion articles for an overview of work that has been done as well as work that has to be done.

2.3. Non-normally distributed data

Jensen et al. (2018) argue that the applicability of approaches across many different situations becomes more crucial. One of the most commonly made assumptions in SPM is that the data under consideration follow a normal distribution. Deviations from normality often lead to undesirable and unpredictable control chart performance. This means that other control chart designs are required for the cases where such deviations are present. These methods are thus not as broadly applicable as desired.

A first alternative is to transform the original data so that the transformed data follow a normal distribution. Such a transformation is considered in Chou, Polansky, and Mason (1998), who use the Johnson system of distributions in combination with a best-fit estimation procedure to transform non-normal data to normality. After the transformation, control charts designed for normally distributed data can be applied. Another option is the use of nonparametric control chart designs. The increasing interest in nonparametric methods was already anticipated by Woodall and Montgomery (1999) and Chakraborti, Van der Laan, and Bakir (2001). Common approaches are the use of order statistics and/or change-point methods. A disadvantage of nonparametric methods in general is that they require large amounts of data to provide a proper control chart performance. However, this large sample size requirement is becoming less of an issue in recent years. Data supply is abundant in many processes, providing a great opportunity for the development of improved nonparametric methods in SPM. However, as can be seen in Table 1 of Jensen et al. (2018), nonparametric methods have not yet received much attention in *JQT*. With the increase of data supply as also mentioned there, we expect to observe an increase in the number of articles in *JQT* regarding nonparametric methods.

3. Future considerations and research ideas

As discussed, the case of perfect normal i.i.d. variables is quite unrealistic in practice, and new works on this setting would provide only marginal improvements to SPM in practice. However, there are several problems that are encountered during the application of SPM that still deserve attention, as well as new issues to be tackled given the development of modern-day data availability.

This availability goes up to the point where it is unclear what to actually do with the data and how to summarize it. Therefore, we agree with Jensen et al. (2018) that new methods that deal with large data sets are required. Multiple issues arise when the size of data sets increases. Some examples are multicollinearity when multiple variables measure more or less the same thing, a high degree of autocorrelation due to the high frequency of sampling, and the problem that p-values will decrease with the sample size.

One of the specific areas where methods for large data sets are extremely important is the area of (social) network monitoring. Many networks collect immense data sets on their users regarding activities, contacts, and much more. How to monitor these types of data and how to define an out-of-control situation formally are two examples of questions that are more often asked than answered. With an increasing interest in this topic, giving (directions to get to) answers to these questions might very well be one of the future challenges for *JQT*.

Since a trend of increasing data availability has been observed in recent years, the future will most likely bring forth a new shift of interest toward methods for large data sets, with social network monitoring as one of the specific applications. The future of *JQT* provides room for approaches and guidelines on how to obtain the right information and detect out-of-control situations from large data sets.

About the authors

Rob Goedhart is a Ph.D. student at the Department of Operations Management, Faculty of Economics and Business, and a consultant at the Institute for Business and Industrial Statistics (IBIS UvA). His email address is r.goedhart2@uva.nl.

Marit Schoonhoven is an Associate Professor at the Department of Operations Management, Faculty of Economics and Business, and a senior consultant at IBIS UvA. Her email address is m.schoonhoven@uva.nl.

Ronald J. M. M. Does is a Professor at and the Head of the Department of Operations Management, and the Director of IBIS UvA. He is a Fellow of the ASQ. His email address is r.j.m.m.does@uva.nl.

Acknowledgments

The authors are grateful to the editor and the authors of "50 years of the *Journal of Quality Technology*" for their work and the opportunity to write this discussion.

References

Adams, B. M., and I. T. Tseng. 1998. Robustness of forecast-based monitoring schemes. *Journal of Quality Technology* 30 (4):328–39.

Chakraborti, S., P. Van der Laan, and S. T. Bakir. 2001. Nonparametric control charts: An overview and some results. *Journal of Quality Technology* 33 (3):304–15.

Chou, Y. M., A. M. Polansky, and R. L. Mason. 1998. Transforming non-normal data to normality in statistical process control. *Journal of Quality Technology* 30 (2): 133–41.

Goedhart, R., M. Schoonhoven, and R. J. M. M. Does. 2017. Guaranteed in-control performance for the Shewhart X and $\bar{X}$ control charts. *Journal of Quality Technology* 49 (2): 155–71.

Hillier, F. S. 1969. $\bar{X}$- and R-chart control limits based on a small number of subgroups. *Journal of Quality Technology* 1 (1):17–26.

Jensen, W. A., L. A. Jones-Farmer, C. W. Champ, and W. H. Woodall. 2006. Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology* 38 (4):349–64.

Jensen, W. A., D. C. Montgomery, F. Tsung, and G. G. Vining. 2018. 50 years of the *Journal of Quality Technology*. *Journal of Quality Technology* 50(1).

Lu, C. W., and M. R. Reynolds Jr. 1999. EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality Technology* 31 (2):166–88.

Lu, C. W., and M. R. Reynolds Jr. 2001. CUSUM charts for monitoring an autocorrelated process. *Journal of Quality Technology* 33 (3):316–34.

Quesenberry, C. P. 1993. The effect of sample size on estimated limits for $\bar{X}$ and X control charts. *Journal of Quality Technology* 25 (4):237–47.

Saleh, N. A., M. A. Mahmoud, M. J. Keefe, and W. H. Woodall. 2015a. The difficulty in designing Shewhart $\bar{X}$ and X control charts with estimated parameters. *Journal of Quality Technology* 47 (2):127–38.

Saleh, N. A., M. A. Mahmoud, L. A. Jones-Farmer, I. M. Zwetsloot, and W. H. Woodall. 2015b. Another look at the EWMA control chart with estimated parameters. *Journal of Quality Technology* 47 (4):363–82.

Schoonhoven, M., H. Z. Nazir, M. Riaz, and R. J. M. M. Does. 2011. Robust location estimators for the X-bar control chart. *Journal of Quality Technology* 43 (4):363–79.

Vasilopoulos, A. V., and A. P. Stamboulis. 1978. Modification of control chart limits in the presence of data correlation. *Journal of Quality Technology* 10 (1):20–30.

Woodall, W. H., and D. C. Montgomery. 1999. Research issues and ideas in statistical process control. *Journal of Quality Technology* 31 (4):376–86.

Yang, C. H., and F. S. Hillier. 1970. Mean and variance control chart limits based on a small number of subgroups. *Journal of Quality Technology* 2 (1):9–16.