Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Statistical Planning and Inference 136 (2006) 909–941

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Data driven choice of control charts☆

Willem Albers, Wilbert C.M. Kallenberg*, Sri Nurdiati

Department of Applied Mathematics, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Received 30 January 2004; received in revised form 8 June 2004; accepted 18 July 2004
Available online 15 September 2004

Abstract

Standard control charts are often seriously in error when the distributional form of the observations differs from normality. Recently, control charts have been developed for larger parametric families. A third possibility is to apply a suitable (modified version of a) nonparametric control chart. This paper deals with the question when to switch from the control chart based on normality to a parametric control chart, or even to a nonparametric one. This model selection problem is solved by using the estimated model error as yardstick. It is shown that the new combined control chart asymptotically behaves as each of the specific control charts in their own domain. Simulations exhibit that the combined control chart performs very well under a great variety of distributions and hence it is recommended as an omnibus control chart, nicely adapted to the distribution at hand. The combined control chart is illustrated by an application on real data. The new modified nonparametric control chart is an attractive alternative and can be recommended as well.

MSC: 62G32; 62P30; 65C05

Keywords: Statistical process control; Phase II control limits; Second order unbiasedness; Normal power family; Model error; Nonparametric; Large deviations; Model selection

☆ This research was supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs.

* Corresponding author. Tel.: +31-53-4893374; fax: +31-53-7893069.
E-mail address: w.c.m.kallenberg@math.utwente.nl (W.C.M. Kallenberg).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2004.07.003

1. Introduction

Classical control charts for monitoring the mean of a production process are based on the assumption that the observations are normally distributed. This assumption is in practice not always fulfilled and the control chart often is seriously in error when the distributional form of the observations differs from normality, see e.g. Chan et al. (1988), Pappanastos and Adams (1996), Albers et al. (2002, 2004). An obvious solution to the problem is to assume a larger parametric model, containing normality as a submodel, and to produce a control chart in this new setting. This step has been made in Albers et al. (2002, 2004), where the merits of such an approach are comprehensively discussed.

It is clear that as long as we are (very) close to normality, the control chart based on normality is preferable. The first reason is that in that case the performance of this control chart is better, since it is matched to normality of the observations. The second reason is that the normal control chart is easier and more familiar and hence people prefer this chart as long as possible.

However, for distributions farther away from normality, but still close to the larger parametric family, the best choice is the parametric control chart. In that case we should no longer stick to the normal control chart, but move towards the parametric chart.

Of course, there also are distributions so far outside the larger parametric family that the parametric control chart is not satisfactory either and a nonparametric approach should be applied. One may ask why not always apply a nonparametric control chart or a parametric control chart in a very large parametric family. The point is that there are two types of error: the model error (due to the distance between the true distribution and the most suitable distribution in the supposed model) and the stochastic error (caused by estimating parameters or, in the nonparametric case, an extreme quantile). The larger the parametric model, the smaller the model error (with a vanishing model error in the nonparametric control chart), but the larger the stochastic error. For instance, estimating the 0.999-quantile with 100 observations makes no sense in a nonparametric setting.

The theme of this paper is how to choose between the three control charts: the normal control chart, the parametric control chart as developed for the normal power family, cf. Albers et al. (2004) and the nonparametric control chart, cf. Albers and Kallenberg (2004c). The idea is to let the data tell what control chart to use.

A first idea might be to execute a (standard) goodness of fit test to investigate normality. If normality is not rejected, use the normal control chart. If we do reject, apply a goodness of fit test for the normal power family. Again, when not rejecting, apply the parametric control chart and otherwise use the nonparametric control chart (if this makes sense).

Although this way of thinking looks attractive, it has a serious drawback. Standard goodness of fit tests are looking at the majority of the data, and as such concentrate on the middle of the distribution, while here we are not interested in this middle part, but in the (extreme) tail. Therefore, standard goodness of fit tests are not appropriate for the situation at hand. For the same reason, less formal methods like “a good look at the data” or “an inspection of a histogram” are completely insufficient to judge the possible normality in the far tail.

The choice between the three control charts can be seen as a model selection problem. Again, we cannot use standard selection rules, since the common selection rules are intended for the bulk of the data and not for the extreme tail. The motivation to switch from the