Journal of Statistical Planning and Inference 142 (2012) 1924–1932

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Control charts for high-quality processes: *MAX* or *CUMAX*?

Willem Albers

Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

A R T I C L E I N F O

Article history:
Received 15 August 2011
Received in revised form
17 January 2012
Accepted 17 January 2012
Available online 30 January 2012

Keywords:
Statistical process control
Health care monitoring
Geometric charts
Average run length
Estimated parameters
Order statistics
Sets method

A B S T R A C T

For attribute data with (very) small failure rates control charts were introduced which are based on subsequent groups of r failure times, for some $r \geq 1$. Within this family, it was shown to be attractive to stop once the maximum of such a group is sufficiently small, because this choice allows a very satisfactory nonparametric adaptation. The question we address here is whether a cumulative approach offers even further improvement. Thus instead of fixed groups, we shall use the first sequence of r consecutive sufficiently small failure times to produce a signal. A further reason for considering this type of chart is the fact that it forms the nonparametric counterpart of the well-known sets method.

1. Introduction and motivation

By virtue of ever increasing standards, high-quality processes are more and more common in industrial settings. Moreover, for health care monitoring they even form the standard, as in this area failure probabilities are tiny by nature: events such as malfunctioning equipment, surgical errors, recurrence of cancer and birth defects, should be avoided as much as possible. In both fields, application of control charts to improve and maintain quality is strongly advocated (see e.g. Sonesson and Bock, 2003; Thor et al., 2007; Shaha, 1995 for some health care monitoring review papers). In view of the really small failure probabilities p involved it is quite common to use charts based on the geometrically distributed waiting times from one failure till the next. A group of size $r(r \geq 1)$ of such waiting times is inspected and an alarm is raised if their sum (or alternatively, their maximum) falls below some boundary value. Of course, a two-sided version can also be used, but in practice the focus with respect to going out of control (*OoC*) typically is on increases of p, and thus on shorter waiting times than during in control (*IC*). The boundary value is determined such that the alarm rate during *IC* (the so-called false alarm rate (*FAR*)), remains sufficiently small (e.g. 0.001). Guidelines are available on how to optimally choose r in relation to the underlying parameters. See Albers (2011) for a description and further references.

Nevertheless, as discussed in Albers (2011) as well, a major remaining problem to be dealt with concerns the estimation issues involved. Usually such problems are conveniently ignored in practice. The typically unknown p is estimated using a so-called Phase I sample and the result $\hat{p}$ is simply plugged in. However, as the *FAR* is (very) small, the relative errors due to this estimation step are far from negligible for practical sample sizes and further analysis and corrections are needed. Moreover, the problem may not stop at estimating a single parameter p. Health care applications often involve patients showing considerable heterogeneity, leading to overdispersion of the formerly geometric waiting times. A rigorous way to solve the resulting distributional estimation problems of course is to adopt a nonparametric approach. However, for sample sizes common in

E-mail address: w.albers@utwente.nl

0378-3758/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2012.01.014

practice, the aforementioned relative estimation errors may now become huge: if e.g. the *FAR* should equal 0.001, a customary sample size like 100 is not very useful towards estimating the corresponding 0.001-quantile.

Consequently, further adaptation is required. The usual form of the chart is based on the sum of the r waiting times for each group, as in the homogeneous case this statistic is negative binomially distributed and moreover clearly optimal. However, as already mentioned above, an alternative is to adopt the maximum, rather than the sum, of the waiting times as our statistic. A simple example (roughly) illustrates the advantage: for $r=3$, the probability that all waiting times in a group fall below their 0.1-quantile is $(0.1)^3=0.001$. But, unlike a 0.001-quantile, estimating a 0.1-quantile based on a sample of moderate size is quite feasible. Of course, solving the estimation problems in this way only makes sense if the step from sum to maximum merely causes a small loss of detection power when the assumption of a homogeneous case happens to hold after all. Fortunately, in Albers (2011) it is demonstrated that this is indeed true. Hence such a loss can be viewed as a small insurance premium to be paid for safeguarding against the risk of making substantial errors with the basic negative binomial chart once the ideal of homogeneity fails. Having settled this issue, the estimation aspects of the proposed nonparametric *MAX*-chart are subsequently dealt with in Albers (2011), resulting in a straightforward empirical version which is easy to understand and to apply.

The topic of the present paper now is the question whether there is still room for further improvement. The motivation for raising it is twofold. In the first place, essentially the same question received a positive answer in the corresponding continuous (and typically normal) case of controlling the mean of a process (see Albers and Kallenberg, 2009). Here the focus usually is on detecting increases in such a mean during *OoC*, implying that a signal typically should result if the minimum of a group of size r is too large. But obviously the analysis for a *MIN*-chart is essentially identical to that for its *MAX*-counterpart, so we can ignore this difference between the two situations. In Albers and Kallenberg (2009) it is suggested to replace the fixed-group approach by a sequential or cumulative version: produce a signal as soon as r consecutive observations all exceed a suitable boundary value. The resulting *CUMIN*-chart definitely looks better than the more rigid *MIN*-chart. It is an accelerated version, starting anew as soon as an observation falls below the boundary value and it thus makes no sense to complete the full group of size r.

However, note that matters are in fact a bit more subtle. For the *MIN*-chart, the average run length (*ARL*) during *IC* equals r/FAR. Making a fair comparison now requires that the *CUMIN*-chart matches this *ARL*-value during *IC*. But, as the runs of failed attempts in this latter chart are mostly shorter than r, a lower value of *FAR* is needed here to actually achieve this. In other words, the *CUMIN*-chart has to be more strict in the sense that it employs a higher boundary value to be exceeded. In view of this, it is not at all trivial anymore that it should be the winner. Nevertheless, in Albers and Kallenberg (2009) it is demonstrated that for the basic situation of normal underlying distributions, this typically indeed is the case, both empirically (Fig. 1 and Table 1) and theoretically (asymptotic results in Lemmas 3.1 and 3.2). Consequently, it seems worthwhile to investigate whether this state of affairs also holds true for attribute rather than continuous data, i.e. when dealing with geometric rather than normal distributions. In other words, is it true that *CUMAX* beats *MAX*?

The second reason for posing the question is quite straightforward. As already remarked in Albers (2011), in the present attribute data setting the cumulative version of the *MAX*-chart is nothing but the well-known sets method, introduced by Chen (1978): a signal results once all of r successive waiting times are too small. But, just as in Albers (2011), the focus here will be on showing how this type of approach can form the basis for a satisfactory nonparametric procedure, thus adequately solving the aforementioned serious underlying estimation issues. Even if these problems are not ignored in practice, usually at best the effect of estimating a single parameter is studied. Typically, the latter already turns out to be substantial; to give an example in the present context, Chen et al. (1997) mention a 30–90% increase in *FAR* for a 10% bias in $\hat{p}$. Nevertheless, as the *CUMAX*-chart actually is the more prominent of the two proposals, it certainly makes sense to figure out whether it is an adequate competitor to (or even an improvement over) the less well-known *MAX*-chart from Albers (2011). In that case, its empirical nonparametric version, introduced here along the lines of Albers and Kallenberg (2009) and Albers (2011), offers an attractive robust alternative to many existing methods which rely with often unfounded optimism on negligibility of both estimation and model errors.

In Section 2 the charts are introduced and compared for the basic homogeneous case, i.e. where the underlying distributions are simply geometric. The comparisons of detection power are based on the commonly used form of *ARL*, which conveniently assumes that the process goes *OoC* prior to the onset of monitoring, or just as monitoring begins. Several authors (e.g. see Sego et al., 2008) have pointed out that this is somewhat artificial. Indeed, in Albers and Kallenberg (2009) it was observed in this connection that the shift will rarely coincide precisely with the start of a new group and thus that the impact of going *OoC* will probably be delayed till the next group starts. For the *MIN*-chart with its groups of fixed size r, this effect will be more pronounced than for the more quickly reacting *CUMIN*-chart. In the present context this consequently translates into an (additional?) advantage of *CUMAX* over *MAX*. This effect, and more generally the impact of using a different form of *ARL*, will be studied in Section 3. Having settled this issue, the empirical nonparametric version of the *CUMAX*-chart is the topic of Section 4. For convenience, the conclusions reached, as well as a summary of the resulting procedure, are presented in Section 5.

2. The homogeneous case

As explained in Introduction, the nonparametric chart, which is our ultimate goal, has to satisfy two requirements. First of all, the estimation effects should be manageable (i.e. not too large and, if desired, allowing suitable corrections). On the