

# Improved binomial charts for high-quality processes

Willem Albers\*

w.albers@utwente.nl, University of Twente, Netherlands

## Abstract

For processes concerning attribute data with (very) small failure rate  $p$ , often negative binomial control charts are used. The decision whether to stop or continue is made each time  $r$  failures have occurred, for some  $r \geq 1$ . Finding the optimal  $r$  for detecting a given increase of  $p$  first requires alignment of the charts in terms of in-control behavior. In the present paper binomial charts are subjected to this same requirement. Subsequent study reveals that the resulting charts are quite attractive in several aspects, such as detection power. For the case of unknown  $p$ , an estimated version of the chart is derived and studied.

## Keywords

Statistical process control. Health care monitoring. Geometric charts. Average run length. Estimated parameters.

## 1. Introduction and motivation

Due to continuing efforts to raise production standards, the occurrence of high-quality processes in industrial production becomes more and more common. Another area where (very) small fractions of defectives are typical is that of health care monitoring. Failures like malfunctioning equipment, surgical errors or recurrence of cancer, should by their very nature be avoided as much as possible and thus occur only very rarely. Some review papers in this latter field are Sonesson and Bock (2003), Thor et al.(2007), Shaha (1995) and Woodall (2006). In such references, the use of control charts to improve and maintain quality, is strongly advocated.

Traditionally the way to monitor attribute data is to apply a  $p$ -chart: consider a fixed number of incoming items or patients and give a signal if the number of defectives is too high. However, quite a few authors have argued that for really small failure probability  $p$  it is preferable to use so-called time-between-events charts. Essentially these are based on waiting times till  $r$  ( $r \geq 1$ ) failures have occurred. A signal then follows if the corresponding negative binomial random variable (r.v.)  $X$  attains a value which is judged to be too small. See Albers (2010) for extensive references on such geometric ( $r = 1$ ) and negative binomial charts. In this latter paper, a more detailed analysis of these charts is presented,

allowing in particular determining which choice of  $r$  is best for a given configuration of underlying parameters. Moreover, the important problem is tackled how to deal with the fact that  $p$  is typically unknown and thus has to be estimated. Simple corrections are derived which control the estimation effects.

While performing this analysis, yet another interesting question arises, and this we will address in the present paper. The above mentioned too small value of the negative binomial  $X$  simply means that  $X \leq n$ , for some suitably chosen lower limit  $n$ . But one could argue that it is in fact a waste of (waiting) time to continue after such a point  $n$ , all the way till the  $n^{\text{th}}$  failure has occurred, i.e. to obtain the actual realization  $x$  of  $X$ . It seems sufficient to just check at this time  $n$  whether at least  $r$  failures have occurred. If so, give a signal; if not, do not continue till  $x$ , but start again right away. Note that in this way we are in fact back at the binomial  $p$ -chart: consider a given number  $n$  of items or patients and give a signal if this group contains  $r$  or more defectives. Otherwise, look at the next batch of size  $n$ .

It might seem that this argument makes the time-between-events charts superfluous after all. Fortunately, matters are less straightforward and several reasons can be given to keep using such

charts as well. First of all, convenience is an aspect to take into account. In monitoring, it can be quite natural to use the occurrence of defectives as alerts. As soon as  $r$  of these have been registered, the corresponding  $x$  is obtained and compared to the lower limit  $n$ . Another advantage is that this  $r$  does not depend on the, usually unknown,  $p$ : one just picks some attractive value like 3 or 5 (ALBERS, 2010) for guidance, e.g. a simple rule of thumb. On the other hand, it is intuitively clear (details follow in later sections) that the lower limit  $n$  does depend on  $p$ . This makes the binomial chart already less straightforward: the batch size here may be a fixed number, but only after the relevant  $p$  has been chosen (and usually estimated).

However, the main complication is of a more technical nature. Restarting each time at point  $n$ , rather than waiting for the realization  $x$ , means that typically much smaller groups of items are used before the next instant occurs at which we decide whether or not to stop. But making a fair comparison requires that still the same Average Run Length (ARL) during in-control has to be realized. To achieve this for such smaller groups thus implies that the False Alarm Rate (FAR) has to be lowered considerably as well. Hence comparison of the two types of charts is more complicated than it might seem at first glance and thus certainly not simply a matter of the negative binomial charts typically 'wasting'  $(x-n)$  observations at each step.

In passing, note that several of the remarks above also apply when comparing to charts of CUSUM-type (cf. Woodall (2006)). These accumulate the information over time and thus will typically be more efficient. However, such charts definitely are more complicated to apply. In particular, the important issue of the impact of estimation and how to deal with that, remains largely unsolved. Hence the (negative) binomial charts can be viewed as robust alternatives.

Consequently, we shall study the properties of the binomial chart in the present paper. Fortunately, the techniques involved will be similar to those from the negative binomial case in Albers (2010), so we will be quite brief here. In section 2 we study the situation during in-control. The out-of-control behavior is the subject of section 3. In section 4, the estimated version of the chart is treated. The conclusions are presented in section 5. Here the implementation of the chart is summarized as well. Finally, to facilitate independent reading, the proofs of two lemmas from Albers (2010) are reproduced in an Appendix.

## 2. The binomial chart

Consider a sequence  $D_1, D_2, \dots$  of independent identically distributed (i.i.d.) random variables (r.v.'s) with  $P(D_i=1) = 1 - P(D_i=0) = p$ , where  $p$  is small (e.g.  $p \leq 0.01$ ). This models the monitoring situation during in-control. At some unknown point, however, matters may change and the process goes out-of-control. This we model by subsequently replacing  $p$  by  $\theta p$ , for some  $\theta > 1$ . The idea of course is to detect such a change as quickly as possible. The time-between-events approach tries to achieve this by considering a new sequence  $X_1, X_2, \dots$ , based on  $D_1, D_2, \dots$ . Here  $X_1$  is the number of  $D_i$  observed when the  $r^{\text{th}}$  nonconforming item occurs, for some given  $r \geq 1$ . Likewise,  $X_2$  is the next number of  $D_i$  required to obtain  $r$  failures, etc. Clearly, these  $X_i$  are i.i.d. copies of a negative binomial r.v.  $X_{r,p}$  such that

$$P(X_{r,p} = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \quad (2.1)$$

where  $k = r, r+1, \dots$ . If no confusion is likely, we simply write  $X$  instead of  $X_{r,p}$ .

A signal is in order whenever an  $X_i$  is too small, i.e. falls below a suitably chosen lower limit  $n$ . To achieve fairness in comparing the charts for various  $r$ , it seems reasonable to choose this  $n = n_r$  such that during in-control  $\text{FAR} = P(X \leq n) = r\alpha$ , for some suitably chosen small  $\alpha > 0$  (e.g. between 0.001 and 0.01). In this way the ARL (measured in terms of numbers of failures) will equal  $n/\text{FAR} = 1/\alpha$  for all  $r$ . In other words, increasing  $r$  means using longer stretches between the subsequent opportunities for stopping. This is then simply balanced by proportionally increasing the corresponding FAR. Hence for the negative binomial chart we arrive at

$$n = n_{r,p} = F_{r,p}^{-1}(r\alpha) \quad (2.2)$$

where  $F_{r,p}$  denotes the negative binomial distribution function (d.f.) and  $F_{r,p}^{-1}$  its inverse. (Either use standard interpolation in (2.2) or let  $n$  be the largest integer such that  $F_{r,p}(n) \leq r\alpha$ ; in practice the differences involved will be negligible.)

Instead of measuring ARL in terms of numbers of failures, we can of course equivalently base it on the numbers of items inspected. As  $EX = 1/p$ , it is immediate that for the negative binomial case  $\text{ARL} = \text{ARL}_X = 1/\alpha$  will be replaced by  $\text{ARL}_D = 1/(\alpha p)$ . Note that this result readily allows us to make the step to the binomial charts mentioned in the introduction. As discussed, these charts do not continue till the realization  $x$ , but decide whether to stop straight away at the lower limit  $n$  itself.

Nevertheless, for fairness' sake, such charts should also satisfy the requirement that  $ARL_D = 1/(\alpha p)$  for all  $r$ . Only in this way it makes sense to compare these new charts, both among themselves for varying  $r$ , as well as to the negative binomial charts. In the new situation, the number of items between two inspection moments has changed from an expected value  $n/p$  to a fixed value  $n$ . Consequently, the value of  $n$  to be used for the binomial chart has to satisfy.

$$P(X_{r,p} \leq n) = n p \alpha \tag{2.3}$$

in order to produce the required  $ARL_D = n/(n p \alpha) = 1/(p \alpha)$ .

Next observe that the old  $n$  from (2.2) solved  $P(X_{r,p} \leq n) = \alpha$ , which already led to an outcome considerably smaller than  $EX_{r,p} = n/p$ . Hence if this  $\alpha$  is replaced by the smaller  $n p \alpha$ , as happens in (2.3), the resulting new  $n$  will even be smaller than before. Consequently, the decision to look at shorter intervals (not the full  $x$ , but just  $n$ ) leads to an even further shortening, because of the need to align the ARL's for the purpose of fair comparison.

Another observation is that the result from (2.3) is less explicit than the one from (2.2): now we need to solve  $n$  from  $F_{r,p}(n) = n p \alpha$ . Of course, even the  $n$  from (2.2) is less explicit than it may seem: merely a numerical answer follows, e.g. by using Maple. In Albers (2010) it was already argued that such numerical outcomes are not very enlightening for the purpose of understanding how  $n$  varies as a function of  $r$ ,  $\alpha$  and  $p$ . For that purpose, the derivation of approximations which are both accurate and transparent, is much more useful and this task was performed in Albers (2010). Fortunately, after minor modifications, these results can be used here as well. The following well-known relations are used:

$$F_{r,p}(n) = P(X_{r,p} \leq n) = P(Y_{n,p} \geq r) \approx P(Z_{np} \geq r) \tag{2.4}$$

where  $Y_{n,p}$  is a binomial r.v. with parameters  $n$  and  $p$ , while  $Z_{np}$  is a Poisson r.v. with parameter  $\lambda = np$  and the latter step in (2.4) assumes  $n$  to be large. Incidentally, observe that from comparison of (2.3) and (2.4) it is immediate that the present binomial chart can also be simply characterized by the requirement that  $P(Y_{n,p} \geq r) = n p \alpha$ . Hence the formulation using a negative binomial r.v. in (2.3) could be avoided. The obvious reason to nevertheless use it here is the convenient link to the already covered negative binomial charts.

Together (2.3) and (2.4) show that for large  $n$  we have as a first approximation step

$$n = n_{r,p} \approx \lambda / p \tag{2.5}$$

where  $\lambda$  is such that  $P(Z_\lambda \geq r) = \lambda \alpha$ . The second step consists of finding an appropriate approximation  $\tilde{\lambda}$  for  $\lambda$ . Since in the negative binomial case this task was already performed for  $P(Z_\lambda \geq r) = \alpha$ , it is hardly surprising that virtually the same method can be used here. Hence to avoid repetition, we just present the result, after a brief explanatory remark on the underlying steps. The first of these entails replacing  $P(Z_\lambda \geq r)$  by  $\sum_{j=r}^{r+2} P(Z_\lambda = j)$ , invoking a result from Klar (2000), which shows that the error involved is sufficiently small. A third order Taylor expansion w.r.t.  $\lambda$ , followed by a suitable inversion step then readily produces:

**Lemma 2.1.** *Let  $\alpha_r = (r!)^{1/(r-1)}$ , then  $\lambda$  such that  $P(Z_\lambda \geq r) = \lambda \alpha$  can be approximated for  $p \leq 0.01$ ,  $3 \leq r \leq 6$  and  $\alpha \leq 0.01$  by*

$$\tilde{\lambda} = \alpha_r (1 + \zeta_r), \text{ with } \zeta_r = r \alpha_r / (r^2 - 1) + \frac{1}{2} \alpha_r^2 r (3r^2 + 5r + 1) / \{(r^2 - 1)^2 (r + 2)\} \tag{2.6}$$

**Proof.** See the Appendix: Lemma 2.1 from Albers (2010).

Hence in addition to the exact result for  $n$  from (2.3) we now have, in view of (2.5) and Lemma 2.1, the approximation

$$\tilde{n} = \tilde{\lambda} / p \tag{2.7}$$

with  $\tilde{\lambda}$  as given in (2.6).

**Remark.** The choice for the region  $3 \leq r \leq 6$  in Lemma 2.1 is explained as follows. Clearly,  $\alpha_r$  increases sharply in  $r$  for given  $\alpha$ . Consequently,  $n$  will be large for  $r \geq 3$ , which means that the error due to the Poisson step will indeed be small for all  $p$  involved. As concerns the values below 3, for  $r=1$  an exact solution of (2.3) can easily be found by looking at  $1 - (1-p)^n = n p \alpha$  directly. Unfortunately, this only produces the useless root  $n \approx 1/(p \alpha)$ , and thus no counterpart of the geometric chart exists. (For  $r > 1$ , the equation  $P(Z_\lambda \geq r) = \lambda \alpha$  has two roots, from which we obviously need the smaller one, and not the second, very large one, which indeed  $\approx 1/\alpha$ ). For  $r=2$ , a refinement of (2.6) can be derived, using binomial, rather than Poisson probabilities. However, as larger  $r$  are more interesting anyhow, this does not seem worth bothering. As concerns the upper end, values of  $r > 6$  could be considered as well, but on practical grounds there seems to be little need to go beyond this value (also see optimality considerations later on.)

While using the Poisson step, the actual value of  $p$  (as long as it is at most 0.01), plays almost no role as far as the approximation quality is concerned and

in studying the behavior of the binomial chart we can focus on comparing  $\tilde{\lambda}$  from (2.6) for various  $(\alpha, r)$  to the 'exact'  $\lambda^* = np$ , with  $n$  as in (2.3). In Table 1 below some illustrative values are collected.

The conclusions from Table 1 completely parallel those from table 1 from Albers (2010): the approximation is fine, with decreasing accuracy as  $r\alpha$  increases (and thus for small  $\alpha$  like 0.001, values of  $r > 6$  can be considered as well). This is hardly surprising, as a result for  $(r, \alpha)$  from the present Table produces a  $\lambda$  such that  $P(Z_k \geq r) = r\tilde{\alpha}$ , with  $\tilde{\alpha} = (\lambda/r)\alpha$ . E.g., the choice  $(r, \alpha) = (5, 0.005)$  gives  $\lambda = 1.11$ , which corresponds to  $(r, \tilde{\alpha}) = (5, 0.00111)$ . Indeed from Albers (2008) we have for  $(5, 0.001)$  that  $\lambda = 1.08$ .

More interesting is the observation that the present  $\lambda$  for given  $(r, \alpha)$  are close to the negative binomial ones for  $(r-1, \alpha)$ . Indeed, solving  $P(Z_k \geq r) = \lambda\alpha$  means finding  $\lambda$  such that  $exp(-\lambda)\lambda^{r-1}/(r-1)! \{1 + \sum_{j=1}^{\infty} \lambda^j / (\prod_{k=1}^j (r+k))\} = r\alpha$ ,

while solving  $P(Z_k \geq r-1) = (r-1)\alpha$  requires  $\lambda$  such that  $exp(-\lambda)\lambda^{r-1}/(r-1)! \{1 + \sum_{j=1}^{\infty} \lambda^j / (\prod_{k=1}^j (r+k-1))\} = (r-1)\alpha$ .

The latter value is slightly smaller, while the relative difference between the two decreases in  $r$ . Yet another way to see this is by noting from Lemma 2.1 that  $\alpha_{r+1} = (r!/(r+1)\alpha)^{1/r}$ , which is slightly larger than  $(r!\alpha)^{1/r}$ , the  $\alpha_r$  for the negative binomial case. Likewise, the coefficient of the leading term of  $\zeta_{r+1}$  (cf. (2.6)) is  $(r+1)/\{r(r+2)\}$ , which slightly exceeds  $1/(r+1)$ , the corresponding coefficient of the negative binomial  $\zeta_r$  from Albers (2010). Hence, roughly speaking, we can use one and the same value of  $n$  in two ways: either check at  $n$  whether at least  $r$  failures have occurred and stop or restart right away, or wait with deciding between stopping and restarting till  $r-1$  failures have occurred and check whether this has happened within  $n$  steps. To conclude the section we provide an explicit example.

**Example 2.1.** Suppose we choose  $\alpha = 0.005$ . If for the binomial chart we want to check after  $n$  observations whether at least  $r = 5$  failures have occurred, Table 1 gives  $\lambda = 1.11$  (or  $\tilde{\lambda} = 1.08$ ). Hence for e.g.  $p = 0.001$  we have that  $n = 1110$

(or  $\tilde{n} = 1080$ ), which indeed is substantially smaller than the expected value 5000 for the time of the fifth failure. From Table 1 in Albers (2010) we find for  $\alpha = 0.005$  and  $r = 4$  that  $\lambda = 1.02$  (or  $\tilde{\lambda} = 1.00$ ), and thus that the somewhat smaller  $n = 1020$  (or  $\tilde{n} = 1000$ ) can be used as the lower limit for the negative binomial chart that pauses after each 4<sup>th</sup> failure.

### 3. The out-of-control situation

At some unknown point in the sequence  $D_1, D_2, \dots$ , the process may go out-of-control and  $p = P(D_i = 1)$  is replaced by  $\theta p$  for some  $\theta > 1$ . In Albers (2010) it was argued that a region like  $3/2 \leq \theta \leq 4$  is of interest, and here we will typically stick to this choice. During out-of-control the probability of a signal is given by  $F_{r, \theta p}(n)$ , and therefore

$$ARL = ARL_{X=np/F_{r, \theta p}}(n) \tag{3.1}$$

Note that the scale we are using here is again the number of failures, which will make it easier to compare the results to the previously obtained ones for the negative binomial case. If desired, the transition from this  $ARL_X$  to  $ARL_D = n/F_{r, \theta p}(n)$  is of course immediate. For all charts, ARL decreases from  $1/\alpha$  (cf. (2.3)) for  $\theta = 1$  to  $np$  for  $\theta = 1/p$ . As  $\lambda = np$  increases in  $r$ , for very large  $\theta$  it is clearly better to take small  $r$ . Just as in section 2, results in (3.1) can be obtained numerically, but again it is much more illuminating to apply a suitable approximation. Following Albers (2010) we obtain

**Lemma 3.1.** The exact  $ARL$  from (3.1) can be approximated for  $p \leq 0.01$ ,  $3 \leq r \leq 6$ ,  $\alpha \leq 0.01$  and  $3/2 \leq \theta \leq 4$  by

$$A \tilde{R}L = A \tilde{R}L_{r, \theta} = \tilde{\lambda} / \{1 - exp(-\theta\alpha_r)\} [1 + \theta\alpha_r + \dots + (\theta\alpha_r)^{r-1} (1 - \theta\alpha_r \zeta_r) / (r-1)!] \tag{3.2}$$

with  $\tilde{\lambda}$ ,  $\alpha_r$  and  $\zeta_r$  as in (2.6).

**Proof.** See the Appendix 1: Lemma 3.1 from Albers (2010). The only changes are again the region of values  $r$  considered (cf. Lemma 2.1) and the different  $\tilde{\lambda}$ ,  $\alpha_r$  and  $\zeta_r$ .

Table 1. Comparison of the approximation  $\tilde{\lambda}$  from (2.6) to  $\lambda^* = np$  with  $n$  as in (2.3) for various  $\alpha$  and  $r$ . The first value is  $\lambda^*$ ; the second one is  $\tilde{\lambda}$ .

$\alpha/r$	3	4	5	6
0.001	0.081	0.080	0.315	0.313
0.005	0.187	0.186	0.576	0.570
0.01	0.272	0.270	0.760	0.749
			0.679	0.674
			1.11	1.08
			1.39	1.35
			1.14	1.12
			1.73	1.67
			2.12	2.00

Just as in Albers (2010), the approximation is again quite satisfactory in the area considered in Lemma 3.1, with decreasing quality as  $r\alpha$  increases. Hence, to avoid repetition, we just list in Table 2 some illustrative values of the exact ARL's for this region, without bothering to accompany these for  $3 \leq r \leq 6$  by the (close) approximate values from (3.2).

From Table 2 it is clear that also for the binomial charts increasing  $r$  leads to large improvements. More importantly, by comparing the values obtained here to those from table 2 in Albers (2010), we see that the present results are consistently better than the corresponding negative binomial ones. Especially for larger  $r$  and  $\alpha$  the relative difference is substantial. This reflects the fact that in the binomial case ARL eventually decreases to  $\lambda$ , whereas the negative binomial chart has  $r$  as a lower boundary for its ARL. Of course, the improvement is not accidental: we have that

**Lemma 3.2.** For  $\theta > 1$ , the ratio  $k = k(\theta, \lambda) = P(Z_\lambda \geq r) / P(Z_{\theta\lambda} \geq r)$  increases in  $\lambda$ .

**Proof.** Since  $\partial \{P(Z_{\theta\lambda} \geq r)\} / \partial \lambda = rP(Z_{\theta\lambda} = r) / \lambda$ , it follows that  $\partial k / \partial \lambda = r \{P(Z_\lambda = r)P(Z_{\theta\lambda} \geq r) - P(Z_\lambda \geq r)P(Z_{\theta\lambda} = r)\} / \{\lambda P(Z_\lambda \geq r)P(Z_{\theta\lambda} \geq r)\}$ . This is indeed positive, as  $\{P(Z_\lambda = r)P(Z_{\theta\lambda} \geq r) - P(Z_\lambda \geq r)P(Z_{\theta\lambda} = r)\} = P(Z_\lambda = r)P(Z_{\theta\lambda} = r) \{ [1 + \sum_{j=1}^{\infty} (\theta\lambda)^j / (\prod_{k=1}^j (r+k))] - [1 + \sum_{j=1}^{\infty} \lambda^j / (\prod_{k=1}^j (r+k))] \}$  and  $\theta > 1$ .

From (2.4) together with (3.1) it is clear that  $ARL \approx \lambda / P(Z_{\theta\lambda} \geq r)$ . Hence as a function of  $\theta$ , the ratio  $k$  from Lemma 3.1 is just the factor by which ARL is reduced when going from 1 to some  $\theta > 1$ . Since  $k$  increases in  $\lambda$ , the reduction for given  $\theta$  is maximized by choosing  $\lambda$  as small as possible. As mentioned before, the binomial charts employ  $\lambda$  such that  $P(Z_\lambda \geq r) = \lambda\alpha$ . This value is indeed smaller, and thus better, than the negative binomial  $\lambda$ , which solves  $P(Z_\lambda \geq r) = r\alpha$ . In fact, the two charts can be seen as the opposite ends of an interval: in general let  $l$ , such that  $n = n_l \leq l \leq EX_{r,p} = r/p$ , be the number of  $D_i$  between two consecutive inspection

moments, then we need  $P(X_{r,p} \leq n_l) = l/p\alpha$  in order to keep satisfying the requirement that  $ARL = 1/\alpha$ . In analogy to (2.4) and (2.5), this leads to  $n_l = \lambda_l/p$ , where  $\lambda_l$  solves  $P(Z_{\lambda_l} \geq r) = l/p\alpha$  and thus  $\lambda_{Bin} \leq \lambda_l \leq \lambda_{NB}$ .

To illustrate matters, we consider the following continuation of Example 2.1:

**Example 3.1.** Again  $\alpha = 0.005$ , and thus the in-control  $ARL = 200$  for all charts involved. Suppose we focus on detecting a possible doubling of the value of  $p$ , i.e. on the case  $\theta = 2$ . For  $r = 5$  we then have from Table 2 that  $ARL = 15.0$ . As  $\lambda = 1.11$  for this case, this means on average 13.5 inspection steps before stopping. In comparison, the negative binomial chart for  $r = 5$  has  $ARL = 21.9$  (see table 2 from Albers (2010)), which is indeed larger than the 15.0 obtained here. On the other hand, there on average only 4 to 5 inspection steps are needed, as the step size is  $r = 5$ , rather than  $\lambda = 1.11$ . To put matters in perspective, do note that the simple geometric chart has  $ARL \approx 1/(\theta\alpha)$ , which still is as high as 100 here.

After showing that increasing  $r$  is very worthwhile, it remains to provide further guidance on how to actually choose  $r$ . This issue has been studied in some detail in Albers (2010) for the negative binomial case. Fortunately it turns out that the conclusions obtained there continue to hold for the present situation. Hence we just quote the result here: a simple rule of thumb for finding  $r^{opt}$ , the value of  $r$  for which ARL is approximately minimal in the region of interest. For given  $\alpha$  and  $\theta$ , let

$$r^{opt} = 1 / \{ \alpha(2.6\theta + 2) + 0.01(4\theta - 3) \} \tag{3.3}$$

Hence e.g.  $\alpha = 0.01$  gives  $r^{opt} = 4$  for  $\theta = 4$  and  $r^{opt} = 5$  for  $\theta = 3$ , while for  $\alpha = 0.005$  we have  $r^{opt} = 5$  for  $\theta = 4$  (cf. Table 2). For practical application, it seems sensible to actually use a truncated version like  $\min(r^{opt}, 6)$ . One reason is that it may feel

Table 2. The exact ARL from (3.1) for various  $\alpha$ ,  $r$  and  $\theta$ .

$\alpha/r$	$\theta = 3/2$					$\theta = 2$				
	2	3	4	5	6	2	3	4	5	6
0.001	445	305	223	173	140	250	133	79.9	54.0	39.9
0.005	89.2	63.4	49.4	41.0	35.5	50.3	28.6	19.5	15.0	12.6
0.01	44.7	32.7	26.4	22.8	20.6	25.3	15.2	11.2	9.28	8.38
$\alpha/r$	$\theta = 3$					$\theta = 4$				
	2	3	4	5	6	2	3	4	5	6
0.001	111	41.6	20.1	12.2	8.70	62.6	18.6	8.09	4.89	3.72
0.005	22.4	9.72	5.94	4.60	4.14	12.7	4.68	2.87	2.44	2.51
0.01	11.4	5.49	3.87	3.42	3.47	6.50	2.81	2.10	2.13	2.50

awkward in practice to apply too large  $r$ , since this excludes the possibility to stop really quickly if  $\theta$  is evidently very large after all. However, note that this effect is less pronounced here than in the negative binomial case. In the latter, the minimum is  $r$ , while for the binomial case it is  $\lambda$ . The second reason is that most of the gain in  $ARL$  reduction compared to the geometric chart, has already been realized for an  $r$  like 6. To avoid repetition, we refer to Albers (2010), Example 3.3 for illustration of this point.

#### 4. The estimated chart

The estimation step for the binomial case closely resembles the one for the negative binomial situation. Hence we shall be very brief here; for more details, again consult Albers (2010). Often  $p$  will be unknown in practice and a Phase I sample has to precede the actual monitoring. To achieve fairness of comparison with respect to estimation as well, no  $r$  should be involved and we simply have  $m$  geometric r.v.'s  $X_{1,p}$ , or equivalently a single negative binomial r.v.  $X_{m,p}$  (cf.(2.1)). Let  $\bar{X} = m^{-1} \sum_{i=1}^m X_i$ , then  $E\bar{X} = 1/p$ ,  $\text{var}(\bar{X}) = (1-p)/(mp^2)$  and the unknown  $p$  can be estimated by  $\hat{p} = 1/\bar{X}$ . Through (2.3) and (2.5), this immediately produces for  $n$  the estimate  $\hat{n} = n_r$ ,  $\hat{p} \approx \lambda / \hat{p} = \lambda \bar{X}$ , where still  $\lambda$  is such that  $P(Z_\lambda \geq r) = \lambda\alpha$ . Likewise, through (2.7) we obtain  $\hat{n} = \tilde{\lambda} / \hat{p} = \tilde{\lambda} \bar{X}$ . Now the chart can be applied as before: following Phase I, after each batch of  $\hat{n}$  (or  $\hat{n}$ )  $D_i$ 's, we stop if at least  $r$  failures have occurred; otherwise the next batch is considered.

Just as in Albers (2010), it remains to study the impact of the estimation step. Performance characteristics like  $FAR$  and  $ARL$  have now become random. We e.g. have

$$F\hat{A}R = FAR(\bar{X}) = P(X_{r,p} \leq \hat{n} | \bar{X}) \quad (4.1)$$

and likewise  $A\hat{R}L = ARL(\bar{X})$ . Consequently, no unique criterion exists to appraise relative errors such as  $W = \{F\hat{A}R - \lambda\alpha\}/(\lambda\alpha)$ . The main candidates are the bias  $EW$  and the exceedance probability  $P(F\hat{A}R > \lambda\alpha(1+\varepsilon)) = P(W > \varepsilon)$ , for some small  $\varepsilon > 0$ . Incidentally, note that this latter criterion essentially also covers the exceedance probability for  $A\hat{R}L$ , as  $P(A\hat{R}L < (1-\varepsilon)/\alpha) = P((A\hat{R}L - 1/\alpha)/(1/\alpha) < -\varepsilon) = P(W > \tilde{\varepsilon})$  with  $\tilde{\varepsilon} = \varepsilon/(1-\varepsilon) \approx \varepsilon$ .

The idea is as follows: from (2.4) and (4.1) we observe that  $F\hat{A}R \approx P(Z_{\hat{n}p} \geq r | \bar{X})$ . Now  $\hat{n}p = \lambda(1+U)$ , with  $U = p/\hat{p} - 1$ . As  $U$  has  $E(U) = 0$  and  $\text{var}(U) = (1-p)/m \approx 1/m$ , expansion in powers of  $U$  will give the desired results on  $W$ . To be more precise, following Albers (2010), we obtain

(cf. Lemma 4.1) that to first approximation the relative bias of  $F\hat{A}R$  equals

$$EW = \gamma r(r-1-\lambda)/(2m) \quad (4.2)$$

where  $\lambda$  is such  $P(Z_\lambda \geq r) = \lambda\alpha$  and  $\gamma = P(Z_\lambda = r)/P(Z_\lambda \geq r)$  satisfies  $1-\lambda/(r+1) < \gamma < 1$  (cf. Klar (2000)). If desired, this bias can be removed by using the slightly more strict  $\hat{n}_c = \hat{n}(1-c) = \lambda\bar{X}(1-c)$ , with (cf. Lemma 4.2)

$$c = (r-1-\lambda)/(2m) \quad (4.3)$$

As concerns the exceedance probability, it can be shown (cf. Lemma 4.3) that, again to first approximation,

$$P(W > \varepsilon) = 1 - \Phi(m^{1/2}\varepsilon/(\gamma r)) \quad (4.4)$$

where  $\Phi$  is the standard normal d.f.. If desired, correction is possible here as well. Specifically, the probability in (4.4) can be reduced to any small value  $\beta > 0$  by choosing  $c$  in  $\hat{n}_c$  this time as

$$c = m^{-1/2}u_\beta - \varepsilon/(\gamma r) \quad (4.5)$$

where  $u_\beta$  satisfies  $1 - \Phi(u_\beta) = \beta$ . Evidently, both  $c$  from (4.3) and from (4.5) tend to 0 as the size  $m$  of the Phase I sample increases. However, the exceedance probability correction from (4.5), being of order  $m^{-1/2}$ , will typically be larger than the order  $m^{-1}$  correction for bias, as is intuitively clear. To avoid repetition, we once more refer to Albers (2010) for further comments and examples.

#### 5. Conclusions and summary

As indicated in the Introduction, the choice between negative binomial charts and binomial ones is less straightforward than it may seem at first sight. Nevertheless, in section 2 it was demonstrated that the technicalities involved are quite similar. This enabled us to be quite brief for the binomial case, as the results for the negative binomial charts were readily available from Albers (2010). In particular, it was noted that for given  $\alpha$  a particular  $n$  could be used either in a binomial chart by looking whether at least  $r$  failures had resulted at this point, or alternatively in a negative binomial chart by waiting for  $r-1$  failures and seeing whether these occurred within  $n$  steps. Hence it was hardly surprising that binomial charts, just like negative binomial ones, benefit from using a value of  $r$  larger than 1. In fact, the same rule of thumb as in Albers (2010) can be applied to find the optimal  $r$  for given  $\alpha$  and  $\theta$ . The main conclusion, however,

was that the detection power of the binomial charts is consistently better than that of the negative binomial ones. Especially for larger  $\alpha$  and  $r$  the relative difference is considerable. A proof for this phenomenon is provided in Lemma 3.2. Hence binomial charts form an attractive alternative to the slightly more simple negative binomial choice.

To conclude the paper, for convenience we summarize the application of the binomial chart as discussed in the previous sections:

- Select a desired in-control ARL =  $1/\alpha$  and a degree of change  $\theta$  during OoC that should be optimally protected against.
- Apply rule of thumb (3.3) to obtain the best  $r$  (typically truncate at 6 in practice).
- Find  $\lambda$  such that  $P(Z_\lambda \geq r) = \lambda\alpha$ , where  $Z_\lambda$  is Poisson, or simply use its approximation  $\tilde{\lambda}$  from (2.6).
- If desired, check whether the out-of-control behavior is satisfactory through ARL from (3.1) or its approximation ARL from (3.2).
- For known  $p$ , either use  $n = \lambda/p$  (cf.(2.5)), or simply  $\tilde{n} = \tilde{\lambda}/p$  (cf. (2.7)).
- If  $p$  is unknown, first wait till  $m$  failures have occurred. Take e.g.  $m = 100$ , or use section 4 (e.g. see (4.2) or (4.4)) to make a more elaborate choice.

- From this Phase I sample, obtain  $\hat{p} = 1/\bar{X}$ , where  $\bar{X} = m^{-1} \sum_{i=1}^m X_i$ , and use  $\hat{n} = \lambda/\hat{p}$ , or simply  $\hat{n} = \tilde{\lambda}/\hat{p}$ .
- Now monitoring starts: a series of batches is inspected, each consisting of  $n$  (or  $\tilde{n}$ ,  $\hat{n}$ ,  $\hat{\hat{n}}$ ) items.
- Give a signal as soon as a batch is encountered which contains at least  $r$  defectives.

## References

- ALBERS, W. The optimal choice of Negative Binomial control charts for monitoring high-quality processes. *Journal of Statistical Planning and Inference*, v. 140, p. 214-225, 2010.
- KLAR, B. Bounds on tail probabilities of discrete distributions. *Probability in the Engineering and Informational Sciences*, v. 14, p. 161-171, 2000.
- SHAHA, S. H. Acuity systems and control charting. *Quality Management in Healthcare*, v. 3, p. 22-30, 1995.
- SONESSON, C.; BOCK, D. A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society: Series A*, v. 166, p. 5-21, 2003.
- THOR, J. et al. Application of statistical process control in healthcare improvement: systematic review *Quality Management in Healthcare*, v. 16, p. 387-399, 2007.
- WOODALL, W. H. The use of control charts in health care and public health surveillance. *Journal of Quality Technology*, v. 38, p. 89-104, 2006.

# Melhorias em gráficos de controle por atributos para processos de alta qualidade

## Resumo

Geralmente gráficos de controle binomial negativa são utilizados em controle de processo por atributos cuja probabilidade de falha  $p$  é relativamente baixa. A decisão de parar ou continuar a produção é feita cada vez que ocorrer  $r > 1$  falhas. Determinar o valor ótimo de  $r$  para detectar um determinado aumento de  $p$ , requer um alinhamento do gráfico em termos de comportamento do processo sob controle. Neste artigo, gráficos de controle binomial estão sujeitos às mesmas exigências. Estudo subsequente revela que este gráfico apresenta resultados um tanto atraentes sob vários aspectos, como por exemplo, seu poder de detecção. Uma versão do mesmo gráfico empregando estimadores de  $p$  também é apresentada no caso de  $p$  ser desconhecido.

## Palavras-chave

Controle estatístico de processo. Monitoramento de indicadores de saúde. Gráfico controle geométrico. Estimação de parâmetros.

Appendix 1. To facilitate independent reading, below we quote Lemmas 2.1 and 3.1 for the negative binomial chart from Albers (2010).

**Lemma 2.1.** Let  $\alpha_r = (r!r\alpha)^{1/r}$ , then  $\lambda$  such that  $P(Z_\lambda \geq r) = r\alpha$  can be approximated for  $p \leq 0.01$ ,  $r \leq 5$  and  $\alpha \leq 0.01$  by

$$\tilde{\lambda} = \alpha_r (1 + \zeta_r), \text{ with } \zeta_r = \alpha_r / (r+1) + \frac{1}{2} \alpha_r^2 (3r+5) / \left\{ (r+1)^2 (r+2) \right\} \quad (\text{A.1})$$

**Proof.** From Klar (2000) we have that for  $k \geq 1$  and  $r > \lambda - 1$

$$\left\{ 1 - \lambda^k / \Pi_{j=1}^k (r+j) \right\} < \left\{ \sum_{j=r}^{r+k-1} P(Z_\lambda = j) \right\} / P(Z_\lambda \geq r) < 1 \quad (\text{A.2})$$

Hence for  $k = 3$ , this ratio lies between  $\left\{ 1 - \lambda^3 / \Pi_{j=1}^3 (r+j) \right\}$  and 1. Since we aim at situations  $\lambda = \eta r$  with  $\eta$  small, this typically means that the ratio from (A.2) is sufficiently close to 1 to allow us to solve

$$e^{-\lambda} \lambda^r \left[ 1 + \lambda / (r+1) + \lambda^2 / \left\{ (r+1)(r+2) \right\} \right] / r! = r\alpha \quad (\text{A.3})$$

rather than  $P(Z_\lambda \geq r) = r\alpha$ . In addition note that, as  $P(Z_\lambda \geq r)$  is increasing in  $\lambda$ , the solution from (A.3) will provide an upper bound for the true  $\lambda$ . The second step involves expanding  $\exp(-\lambda)$ : as  $|\exp(-\lambda) - (1 - \lambda + \frac{1}{2}\lambda^2)| \leq \lambda^3/6$  for  $\lambda > 0$ , the error involved here will also be acceptable for small  $\lambda$ . Hence (A.3) leads to e.g. the result that to first order  $\lambda/r = r\alpha$  and thus to  $\tilde{\lambda} = \alpha_r$  to first order; using expansion to third order w.r.t  $\lambda$  and inverting the result w.r.t  $\alpha_r$  produces (A.1) in a straightforward manner.

**Lemma 3.1.** The exact  $ARL = ARL_{r,\theta} = r! F_{r,\theta,p}(n_{r,p})$  can be approximated for  $p \leq 0.01$ ,  $r \leq 5$ ,  $\alpha \leq 0.01$  and  $3/2 \leq \theta \leq 4$  by

$$A \tilde{R} L = A \tilde{R} L_{r,\theta} = r / \left\{ 1 - \exp(-\theta\alpha_r) \left[ 1 + \theta\alpha_r + \dots + (\theta\alpha_r)^{r-1} (1 - \theta\alpha_r\zeta) / (r-1)! \right] \right\} \quad (\text{A.4})$$

with  $\alpha_r$  and  $\zeta$  as in (A.1).

**Proof.** From Lemma 2.1 it follows that in  $ARL = r! P(Z_{\theta\alpha_r} \geq r)$  we can replace  $\lambda$  by  $\tilde{\lambda} = \alpha_r (1 + \zeta)$  from (A.1). Since  $dP(Z_\mu \geq r) / d\mu = P(Z_\mu = r-1)$ , we have that  $P(Z_{\mu(1+\zeta)} \geq r)$  to first order equals  $P(Z_\mu \geq r) + \zeta\mu P(Z_\mu = r-1) = 1 - \exp(-\mu) [1 + \mu + \dots + \mu^{r-1} (1 - \zeta\mu) / (r-1)!]$ . Application of this result with  $\mu = \theta\alpha_r$  immediately produces (A.4).