The Rate of False Signals for Control Charts with Limits Estimated from Small Samples

Dan Trietsch
Diane Bischak
Department of Management Science and Information Systems
University of Auckland
New Zealand
d.trietsch@auckland.ac.nz
d.bischak@auckland.ac.nz

Abstract

Due to more frequent use of short production runs and the need for shorter setup times in order to make these runs cost-effective, there has lately been a great deal of interest in the statistical properties of $\bar{X}$ control charts with control limits that are based on unknown process parameters. Typically the true process distribution can be assumed to be normal, but the process mean μ and variance σ^2 are estimated from some number m of initial subgroups of size n each. For short production runs it would be desirable to use only a small number of subgroups to estimate μ and σ^2 in order to get on with charting the run as soon as possible. However, as a number of authors have discussed, basing the chart on estimates of μ and σ^2 made from a small number of subgroups may give rise to some unexpected and undesirable effects. In particular, a chart with limits estimated from only a few subgroup means will tend on average to produce a greater number of false signals, which will add to the cost of production.

The statistical properties of charts with limits estimated from small samples have usually been studied from the perspective of the effects of estimation on the average (expected) run length for a chart, known as its ARL. We examine the ARL and find that it is easily misunderstood and that, even when it is unambiguously defined, it has only moderate value as a focal point for the study of control charts. We argue that the rate at which false signals occur in a chart is both a more intuitive concept and a more useful one for determining a reasonable number of subgroups to sample in order to construct control limits.

---

1 Introduction

Since process parameters are seldom, if ever, known with certainty, the statistical properties of $\bar{X}$ control charts that are based on estimates of process parameters should be of particular interest to practitioners. Typically the true process distribution can be assumed to be normal, but the process mean μ and variance σ^2 are estimated from some number m of initial subgroups of size n each. The upper and lower control limits UCL and LCL are then a function of these estimates, for example $\bar{\bar{X}} \pm t\bar{S}/(c_4\sqrt{n})$, where $\bar{\bar{X}}$ is the average of the m subgroup averages $\bar{X}_i$, i=1,…, m, $\bar{S}$ is the average of the m subgroup standard deviations S_i, i=1,…, m, c_4 is the value such that $E[S_i/c_4] = \sigma$ for the given subgroup size n, and t is the desired number of standard deviations to use for process control.

From the practitioner's viewpoint, it would be desirable to use only a small number of subgroups to estimate μ and σ^2 in order to get on with charting the run as soon as possible. However, as a number of authors have discussed, basing the chart on estimates of μ and σ^2 made from a small number of subgroups may give rise to some unexpected and undesirable effects. In particular, Hillier [3] computes the probability that a subgroup mean will fall outside the control limits when the process is in control and shows that for small m this can be much larger than 0.0027, the probability of a false signal for a chart with 3σ limits constructed using a known mean and variance. Thus a chart with limits estimated from only a few subgroup means will tend on average to produce a greater number of false signals, which will add to the cost of production. Various authors have attempted to circumvent this problem; see, for example, Hillier's computation [3] of an adjusted constant for the number of process standard deviations enclosed by the control limits, or the Q-charts of Quesenberry [4], which allow charting to be done from the beginning of a run. Such schemes invariably trade power for safety: the price paid for a reduction in the number of false signals from an in-control process is a decrease in the number of signals sent when the process is truly out of control.

In the quality control literature, the statistical properties of charts with limits estimated from small samples have usually been studied from the perspective of the effects of estimation on the average (expected) run length for a chart, known as its ARL. In this paper, however, we examine the ARL and find that it is easily misunderstood and that, even when it is unambiguously defined, it has only moderate value as a focal point for the study of control charts. We argue that the rate at which false signals occur in a chart is both a more intuitive concept and a more useful one for determining a reasonable number of subgroups to sample in order to construct control limits. In the remainder of the paper we enlarge upon this argument, discussing the properties of the distribution of the rate of false signals in detail.

2 The average run length for charts with estimated limits

We will assume the following two-stage scenario for establishing an $\bar{X}$ control chart. (In our narrative we ignore the dispersion chart, but implicitly it is also created, at least for use in the calculation of control limits.) At Stage 1 we sample m subgroups of n items each and create a trial control chart. Assuming that the m points used to create the chart are strictly between the control limits, we declare the control chart as ready for Stage 2, i.e., it is no longer a "trial" chart (otherwise, we take action to obtain a "good" trial chart). At Stage 2 we start collecting data. We may then ask how long we can expect to sample data until we see the first out-of-control point. We may do this under two assumptions: one, that the process is in control, and therefore the out-of-control signal in question is false; two, that the process is out of control in some specific way (e.g., the adjustment is off center by a given amount). The discussion here will concentrate on the case that the first assumption is true. Under the first assumption, the quantity we are interested in is the ARL.

The ARL is the expected value of the random variable that represents the sample number on which the first (false) out-of-control point appears for a process that is operating in control. That is, for a Stage 2 charted process $\{\overline{X}_t, t=1, 2, \ldots\}$, ARL = E[RL], where the run length RL = min $\{t\text{: } \overline{X}_t \notin [\text{LCL, UCL}]\}$ (see [1]). The ARL can then be thought of as the average across a sample of charts of the number of samples plotted on a chart until the first false signal occurs, where every chart in the sample for given values of *m*, *n*, and *t* is counted exactly once in that average. Note that for the case that the control limits are known with certainty and thus are not estimated at Stage 1, a process producing independent samples will produce a series of false signals that can be considered to be independent of one another, and therefore the ARL has a geometric distribution with parameter $p = 0.0027$ for a 3σ chart.

Most previous authors have measured a chart's tendency to produce false signals by its ARL. Historically, this interest in the ARL has often come about because of the usefulness of the quantity in the economic design of control chart procedures: for example, Ghosh et al. [2] present an economic model of long-run average cost per unit time in which the ARL figures as one of the most important parameters. However useful the ARL may be to researchers, to a practitioner the ARL is a peculiar and somewhat irrelevant measure. As the definition of ARL above shows, the ARL is the average over a large number of charts of a single false signal per chart, the first one that the chart produces. It is defined as though each chart were to be used only until the first out-of-control point occurred, then thrown out and replaced by a new chart, even though investigation reveals that the process is still in control. Unless a chart is indeed thrown away after the first out-of-control point, which is unlikely in practice, a better measure of a chart's performance over time is not the expected run length but the expected run length on average (RLOA), which is the average of the run lengths between out-of-control points on a given chart when the process is in control. This is especially true since, as Quesenberry noted [4], there can be many very short runs between out-of-control points on a chart with estimated limits.

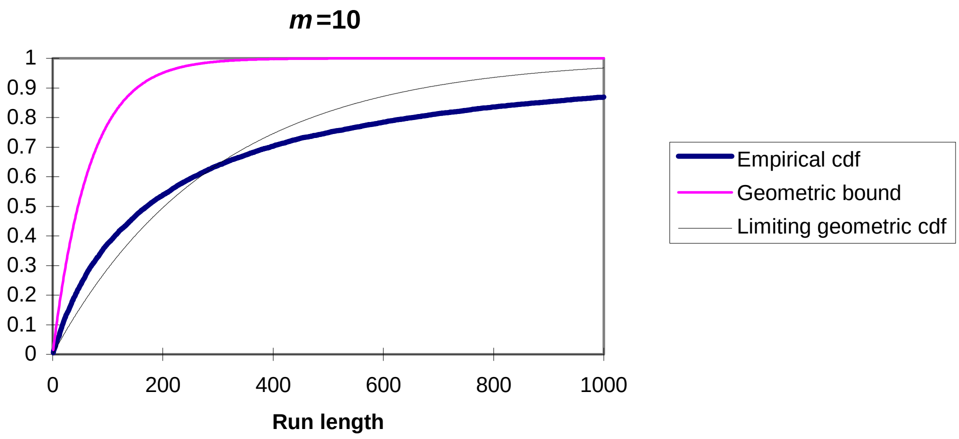
For a given chart, the RLOA has a negative binomial distribution with parameters *r*, equal to the number of runs averaged together, and *p*, equal to the probability of a point being out of control. Although the across-chart expected values and variances of the RL and the RLOA are the same, their distributions are not. Figure 1 (at end of paper) shows the estimated distribution of RLOA for m=50 and various numbers of run lengths averaged within a given chart, based on 10,000 (or, in the top graphs, 100,000) simulated charts with n=5 and t=3. If only one run length is gathered per chart, we have the ARL distribution. (The top graphs' curves are smoother than the bottom two because of the additional charts simulated.) Averaging ten or more run lengths together on a given chart produces very similar distributions, but the distribution of a single run length is different: the probability of seeing an RLOA of less than, say, 50 is very high if only one run length is gathered per chart but very low if ten or more run lengths per chart are averaged together. These distributions are all positively skewed with a heavy tail, and the single run length distribution is similar to a heavy-tailed geometric distribution. The RLOA distributions are clearly not normal, which would be the approximate distribution for a chart constructed with known limits. The Central Limit Theorem does not apply here, because each plotted point is a sample from a different distribution, that is, from a given chart's distribution of average run lengths. Thus the RLOA distribution should itself be of interest to both researchers and practitioners.

Previous authors have also used the ARL to explain why the properties of charts with control limits estimated from small *m* turn out to be inferior to those of charts with known limits. Ghosh et al. [2] point out that any procedure that uses the same estimate of σ^2 to judge whether each sample mean is in control, as will be true in the two-stage scenario we describe above, will result in a dependent, rather than independent, series of comparisons of subgroup means against the control limits. That is, as Quesenberry shows [4], there will be a dependence between the events "false signal at subgroup *i*" and "false signal at subgroup *j*". This dependence, which is negligible for large *m* but sizeable for very small *m*, has implications for the run length distribution and the ARL. Ghosh et al. derive an expression for the distribution of the run length for an estimated chart and show that small *m* increases both the ARL and the variance of the run length distribution.

As noted above, for a chart with known limits, false signals are independent of each other, and therefore the ARL for a single run length per chart has a geometric distribution. But the dependence found by Ghosh et al. [2] and Quesenberry [4] implies that the run length distribution (across charts) is not geometric when the control limits are estimated. Figure 2 shows, for m=10 with n=5 and t=3, the empirical cumulative distribution function (cdf) of the run length for charts with estimated limits, a geometric distribution which forms a bound for this cdf [2], and the limiting cdf for a chart with known limits. The bounding geometric distribution has parameter $p = \Pr(-t \le T_{(m-1)} \le t)$, where $T_{(m-1)}$ follows a Student's *t*-distribution with $m-1$ degrees of freedom. The limiting cdf is that of a geometric random variable with $p = 0.0027 = \Pr(-t \le Z \le t)$, where Z is a standard normal random variable. The bounding geometric is stochastically smaller than the estimated chart run length distribution, so although in this case it is a very loose bound it is a useful one, as for any given run length the probability of seeing a run length less than that length is always less for the chart with estimated limits than for the bounding distribution. The limiting cdf is of interest for comparison purposes, because it shows just how different the estimated chart's distribution is from the known limits case. As Quesenberry points out [4], estimated charts produced from small samples will have many more very short runs and more (although not as many) extremely long runs between false signals, which leads Quesenberry to recommend using larger *m* to produce charts in general. Seeing more short runs in a chart means chasing a false signal more frequently; seeing more extremely long runs means that the power of the chart to detect a real change in the process is reduced.

3 The rate of false signals for charts with estimated limits

Consider now the false signals that a chart with estimated limits will produce. It is true that, from a standpoint before the production of the chart, there is dependence between these false signals, so that an unusually low mean estimate, for example, will push the whole chart down relative to the data and will tend to result in a larger number of false signals. However, after a particular chart is actually produced from estimates of $\bar{\bar{X}}$ and σ^2, it will have independent runs between false signals and a geometric run length distribution, because the limits are fixed values. This claim is true even if the control



2. Empirical, bounding geometric, and limiting geometric cdf's of the run length for *n*=5 and *t*=3.

limits are determined by, say, reading tarot cards: once control limits are obtained, one way or another, the rate of false signals, *p*, is a constant, but it is a different constant for each control chart.

Unfortunately, the users of a given chart with estimated limits do not know their chart's *p*: all they know is that they sampled *m* subgroups in Stage 1 and subsequently the first, say, *k* subgroups of Stage 2 were not out of control. For these users the best estimate of the probability of a false signal on subgroup *k*+1 is the average rate of false signals of all charts based on *m* subgroups with no signal on the first *k* subgroups. This rate is highly unlikely to be exactly equal to *p*, the true rate of false signals that their own chart possesses. However, as users gain experience with a particular chart and obtain many signals on it that cannot be explained by investigating the process, they may have a much better idea of their chart's true rate of false signals. Thus treating the chart as if nothing is known about it, as though it is a random draw from the universe of possible charts, is invalid in the long run.

We refer to the rate of false signals across all possible charts as the RFS. Assuming that the underlying process is normal, the rate of false signals for a chart with fixed limits UCL = u and LCL = l is a constant equal to

$$1 - \Phi\left(\frac{u - \mu}{\sigma/\sqrt{n}}\right) + \Phi\left(\frac{l - \mu}{\sigma/\sqrt{n}}\right),$$

where μ and σ are the mean and standard deviation of the process and $\Phi(x)$ is the cumulative normal distribution at x. (This expression evaluates to 0.0027 for 3σ limits.) However, for estimated charts RFS is not a constant but is instead a random variable: it is the (random) probability that a subgroup mean will fall outside the (random) control limits UCL and LCL, given that the process is in control. The random variable RFS has the expected value

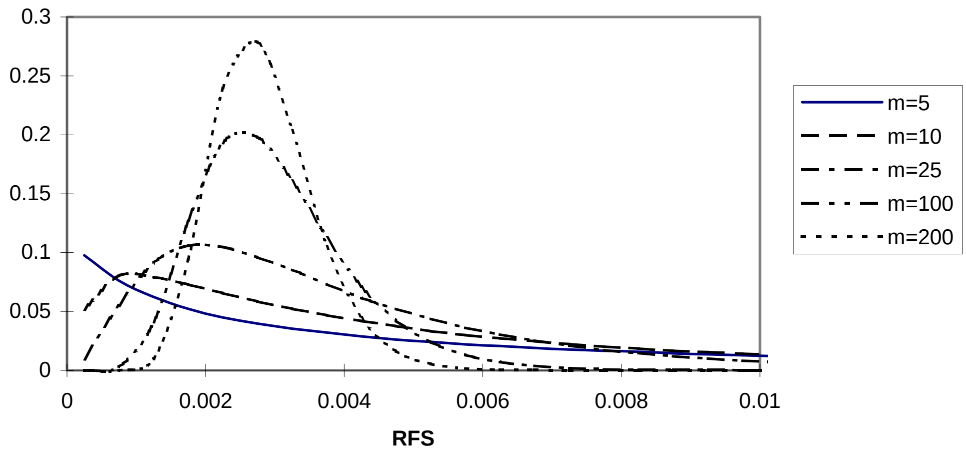
$$E[RFS] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [1 - \Phi\left(\frac{u-\mu}{\sigma/\sqrt{n}}\right) + \Phi\left(\frac{l-\mu}{\sigma/\sqrt{n}}\right)]\, f_{UCL,LCL}(u,l)\, du\, dl,$$

where $f_{UCL,LCL}(u,l)$ is the joint distribution of UCL and LCL. The probability that is given in Table 1 of [3] is the expected value of RFS for a 3σ chart with limits estimated from m subgroups of n=5 each, for various values of m; these probabilities approach 0.0027 as m becomes large.

To understand more clearly the difference between the ARL and the RFS as measures of chart performance, consider the following scenario. Suppose that a group of charts are all constructed by estimating limits by the same procedure in Stage 1, and suppose further that all the charts will be used for the same length of time. Then some charts will generate more false signals than others. A chart that has small ARL will tend to contribute many false signals to the pool of false signals during this time period, and a chart that has large ARL will tend to contribute fewer of them. An estimate of the RFS for this charting procedure based on these charts would average the charts by placing equal weight on each chart's contributions of run lengths (out-of-control points) that occur in the given period of time. An estimate of the ARL, on the other hand, would average only the run lengths until the first out-of-control point on each chart, thus counting each chart exactly once. Note that for estimated charts the average overall rate of false signals across all charts (as computed by Hillier [3]) is not simply the reciprocal of the average run length until the first false signal (as studied by, for example, Quesenberry [4]). If each constructed chart will be used over a period of time, it is likely to generate more than one false signal. The rate of false signals is then a natural measure of the chart's performance.

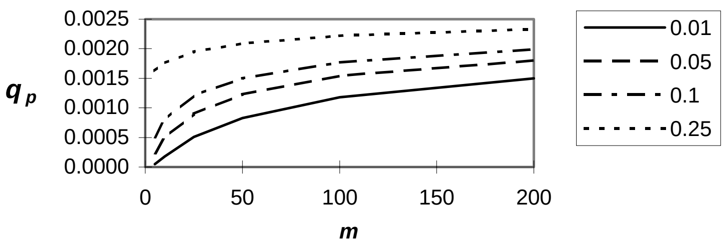
4 Quantiles of the rate of false signals

Because charts' limits will vary, it is useful to look not just at the expected rate of false signals but also at the probability distribution of that rate across charts. This will show, for example, just how likely it is to get a bad chart if a small number of subgroups is used. Figure 3 shows the probability density function of RFS for a 3σ chart using various values of m. These graphs were created from smoothed histograms of data from simulations, hence the graphs do not begin at the origin. The graphs are similar to Weibull density functions for various parameter values. At small values of m the density functions display an extremely large spread, indicating the wide range of possible false signal rates achievable across charts. As m increases the density gradually tightens up; if m were to go to infinity, which would in effect mean that the limits are not estimated but are known with certainty, the density function would become a single spike at 0.0027.



3. Probability density function of RFS for various values of *m*.

Figure 4 shows various quantiles of the rate of false signals as a function of *m*, based on simulation results. The values plotted in Figure 4 are (m, q_p), where pr (RFS ≤ q_p) = p, p = 0.01, 0.05, 0.10, and 0.25. For example, 1% of charts created from 25 subgroups, a number recommended by many authors, have an RFS which is less than 0.0005, and 10% have an RFS less than 0.0012. Even at *m*=100, recommended by [4], 10% of charts will have an RFS of 0.0018 or less, that is, at most two-thirds of the "planned" 0.0027 value.



4. Quantiles of RFS as a function of *m*.

We can look at these quantiles q_p as functions of t, the number of standard deviations the chart is based upon; z, where $\Phi(z) = p$; and m. Call these values $\lambda(t, z, m)$. Tabled values of $\lambda(t, z, m)$ are given in Table 1, calculated using the Mathematica software package. For large m the estimate of the true standard deviation, whether based on $\overline{S}$ or $\overline{R}$, will be approximately normally distributed. We can use this fact and approximations based on the normal distribution to interpolate in Table 1 to obtain quantiles for any p and m>25. Interpolation between tabled values for m>25 can be effectively performed with a two-step procedure, details of which are given in Trietsch and Bischak [5].

5 Conclusions

In this paper we have pointed out ambiguities in the definition of the ARL and have argued that the rate of false signals is a more intuitive and more useful measure of a chart's performance. Graphs and tables of RFS quantiles can be used by practitioners to determine a reasonable number of subgroups to sample for constructing control limits. The distribution of the rate of false signals directly provides information to the practitioner concerning the probability that a chart will have control limits which either have an excessive rate of false signals or a rate of false signals that is too low which consequently provides little power to determine that a process change has actually occurred.

References

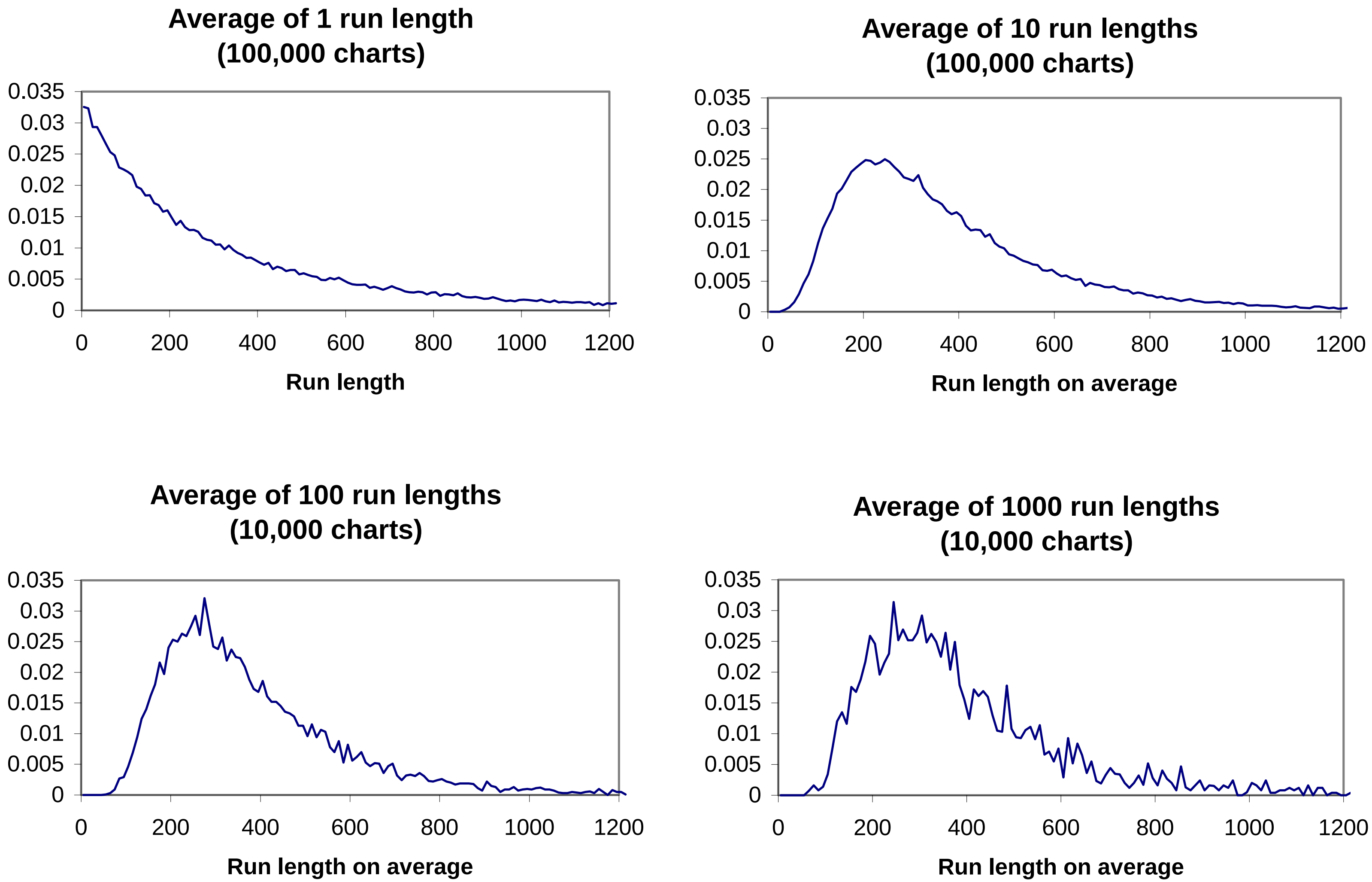
[1] E. Del Castillo, *Run Length Distributions and Economic Design of $\overline{X}$ Charts with Unknown Process Variance*, Metrika, 43 (1996), pp. 189-201.

[2] B.K. Ghosh, M.R. Reynolds, Y. Van Hui, *Shewhart $\overline{X}$-Charts with Estimated Process Variance*, Communications in Statistics: Theory and Methods, 18 (1981), pp. 1797-1822.

[3] F.S. Hillier, *$\overline{X}$ Chart Control Limits Based on a Small Number of Subgroups*, Industrial Quality Control, 20 (1964), pp. 24-29.

[4] C.P. Quesenberry, *The Effect of Sample Size on Estimated Limits for $\overline{X}$ and X Control Charts*, Journal of Quality Technology, 25 (1993), pp. 237-247.

[5] D. Trietsch, D.P. Bischak, *The Probability of False Signals in $\overline{X}$ Charts*, working paper, Department of Management Science and Information Systems, University of Auckland.



1. Estimated distribution of run length on average, m=50, n=5, and t=3.

	z	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5
	p	0.5	0.69146	0.84134	0.93319	0.97725	0.99379	0.99865	0.99977
t	*m*								
1	25	0.32680	0.34509	0.36402	0.38359	0.40378	0.42460	0.44604	0.46808
	100	0.31972	0.32879	0.33802	0.34743	0.35699	0.36672	0.37661	0.38667
	400	0.31792	0.32243	0.32699	0.33158	0.33622	0.34091	0.34563	0.35039
	1600	0.31746	0.31971	0.32198	0.32425	0.32653	0.32883	0.33113	0.33345
	∞	0.31731	0.31731	0.31731	0.31731	0.31731	0.31731	0.31731	0.31731
2	25	0.04986	0.05898	0.06945	0.08140	0.09497	0.11028	0.12749	0.14672
	100	0.04658	0.05081	0.05535	0.06022	0.06544	0.07103	0.07701	0.08339
	400	0.04577	0.04782	0.04995	0.05215	0.05444	0.05681	0.05927	0.06181
	1600	0.04557	0.04658	0.04761	0.04866	0.04973	0.05082	0.05193	0.05307
	∞	0.04550	0.04550	0.04550	0.04550	0.04550	0.04550	0.04550	0.04550
2.5	25	0.01423	0.01826	0.02325	0.02939	0.03687	0.04592	0.05677	0.46808
	100	0.01286	0.01463	0.01661	0.01882	0.02129	0.02403	0.02708	0.38667
	400	0.01253	0.01337	0.01427	0.01521	0.01621	0.01727	0.01839	0.35039
	1600	0.01245	0.01286	0.01329	0.01372	0.01417	0.01464	0.01512	0.33345
	∞	0.01242	0.01242	0.01242	0.01242	0.01242	0.01242	0.01242	0.01242
3	25	0.00326	0.00462	0.00647	0.00895	0.01226	0.01660	0.02225	0.02949
	100	0.00283	0.00339	0.00405	0.00482	0.00572	0.00677	0.00799	0.00940
	400	0.00273	0.00299	0.00328	0.00358	0.00391	0.00427	0.00466	0.00509
	1600	0.00271	0.00283	0.00297	0.00310	0.00325	0.00340	0.00355	0.00371
	∞	0.00270	0.00270	0.00270	0.00270	0.00270	0.00270	0.00270	0.00270

Table 1. $\lambda(t, z, m)$ for R-based $\overline{X}$ charts with n=5.