Center for Quality and Productivity Improvement
University of Wisconsin-Madison
610 Walnut Street
Madison, WI 53705

Report No. 33

A CONTOUR NOMOGRAM FOR DESIGNING CUSUM CHARTS FOR VARIANCE

José Ramírez[1] and Jesús Juan[2]

February 1989

1 Center for Quality and Productivity Improvement, University of Wisconsin-Madison.
2 Departamento de Estadística, Universidad Politécnica de Madrid, Madrid, SPAIN.

This research was sponsored by the National Science Foundation Grant DMS-8420968 and the United States Army under Contract No. DAAL03-87-0050; and was aided by access to the research computer of the University of Wisconsin-Madison Statistics Department and the Center for Mathematical Sciences. The research of Professor Juan was also supported by the Comunidad Autónoma de Madrid.

CENTER FOR QUALITY AND PRODUCTIVITY IMPROVEMENT
UNIVERSITY OF WISCONSIN-MADISON

Report No. 33

A CONTOUR NOMOGRAM FOR DESIGNING CUSUM CHARTS FOR VARIANCE

José Ramírez[1] and Jesús Juan[2]

February 1989

PRACTICAL SIGNIFICANCE

The objective of this report is to provide a method for the design of cumulative sum charts for the control of the variability of a process which produces observations that are normally distributed. Fredholm integral equations of the second kind are used to obtain average run lengths (A.R.L), for different charts. We discuss the construction and use of a contour nomogram based on the values of A.R.L at an acceptable quality level L_a and an rejectable level L_r. The robustness of the A.R.L when the normality hypothesis does not hold and the observations are distributed according to the exponential power family of distributions is discussed.

KEYWORDS: A.R.L; contour nomogram; cusum; exponential power distribution; Fredholm integral equations; variance.

1 Center for Quality and Productivity Improvement, University of Wisconsin-Madison.
2 Departamento de Estadística, Universidad Politécnica de Madrid, Madrid, SPAIN.

This research was sponsored by the National Science Foundation Grant DMS-8420968 and the United States Army under Contract No. DAAL03-87-0050; and was aided by access to the research computer of the University of Wisconsin-Madison Statistics Department and the Center for Mathematical Sciences. The research of Professor Juan was also supported by the Comunidad Autónoma de Madrid;.

1. INTRODUCTION

Cusum charts have been extensively used for the control of the mean level of industrial processes, see, for example, Woodward and Goldsmith (1964). In some processes, however, we are more interested in controlling the variabilty of the process rather than the mean level, or controlling simultaneously the mean level and the variability. In general, it is not only desiderable that the mean be on target but also that the variability around the target value be as small as possible.

If x_k is the observed value of the response at time k, distributed with known mean as $N(\mu, \sigma)$, cumulative sums charts can be constructed to monitor changes in the standard deviation of the process from an acceptable value σ_a to a rejectable value σ_r, by plotting the cumulative sum of $(x_k - \mu)^2$ versus the number of observations k.

A more practical cusum chart can be obtained by choosing a "reference" value s, as a value between σ_a and σ_r, and plotting the cumulative sum of

$$S_k = \sum_{i=1}^{k} [(x_i - \mu)^2 - s] \qquad (1.1)$$

versus the number of observations k. If the population variance is equal to this reference value, the resulting figure will be horizontal; otherwise, the figure will slope downwards if the population variance is less than s and upwards if it is greater than s.

Since we are concerned with departures from σ_a in one direction, we could plot the cumulative sums in (1.1) only when they are relevant towards taking a decision that the process variance has increased. This means that starting from zero, values of S_k less than zero need not be plotted and the process can be considered as satisfactory. However, as soon as a result exceeds zero, a cumulative sum is started. If the cusum subsequently reverts to zero, it is concluded that the process is satisfactory, but whenever S_k reaches or crosses a stated amount h, called the "decision limit", the process variance is said to have increased. See figure 1 below.

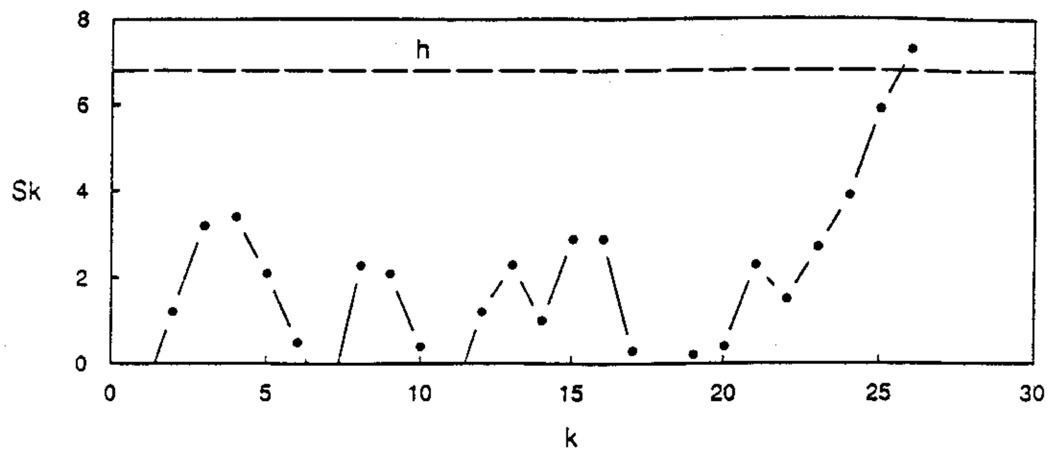

Fig.1 Cusum Chart

The control scheme just described corresponds to a sequence of Wald sequential tests, whose efficiency will depend on the choice of the parameters h and s. This choice, as in the case of cusum charts for the mean (Goel and Wu (1971)), will be based on some previously chosen Average Run Lengths.

2. AVERAGE RUN LENGTH

Following Page (1954), when the standard deviation of the process remains constant the Average Run Length (A.R.L) is defined as the expected number of items sampled before concluding that an increase in standard deviation has occurred.

The Average Run Length depends on the values of h, s, the true standard deviation of the process and the distribution of the observations. The values of h and s are chosen to yield large values of A.R.L when the process variability is at an acceptable level, and low values when is at a rejectable level.

The A.R.L of a cusum chart is defined in terms of $P(0)$ and $N(0)$,

$$ARL = \frac{N(0)}{1-P(0)}, \tag{2.1}$$

where $P(z)$ is the probability that a Wald test with boundaries $(0,h)$ and initial value z ends in the lower boundary and $N(z)$ the average sampling number that the test ends either in the lower boundary 0 or in the upper boundary h.

The functions $P(z)$ and $N(z)$ are the solutions to Fredholm integral equations of the second kind

$$P(z) = \int_{-s}^{-z} f(y)dy + \int_{0}^{h} P(x)f(x-z)dx, \quad 0 \le z \le h \tag{2.2}$$

$$N(z) = 1 + \int_{0}^{h} N(x)f(x-z)dx, \qquad 0 \le z \le h. \tag{2.3}$$

Where $f(y)$ is the density function of the increments of the cumulative sum; i.e., $f(y)$ is the density of $Y = [(X-\mu)^2 - s]$; where X is $N(\mu,\sigma^2)$

$$f(y) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi}}(y+s)^{-\frac{1}{2}} \exp\left[-\dfrac{(y+s)}{2\sigma^2}\right] & \text{if } y \ge -s, \\ 0 & \text{if } y < -s. \end{cases} \tag{2.4}$$

The solution of equations (2.2) and (2.3) can be obtained by replacing them by a system of linear equations as suggested by Kantorovich and Krylov (1964) (see appendix).

The figures 2a and 2b show the graphs of $P(0)$ and $N(0)$ for different values of h. Note how $P(0)$, the probability of a sequential test with boundaries $(0,h)$ ends on or below the lower boun-

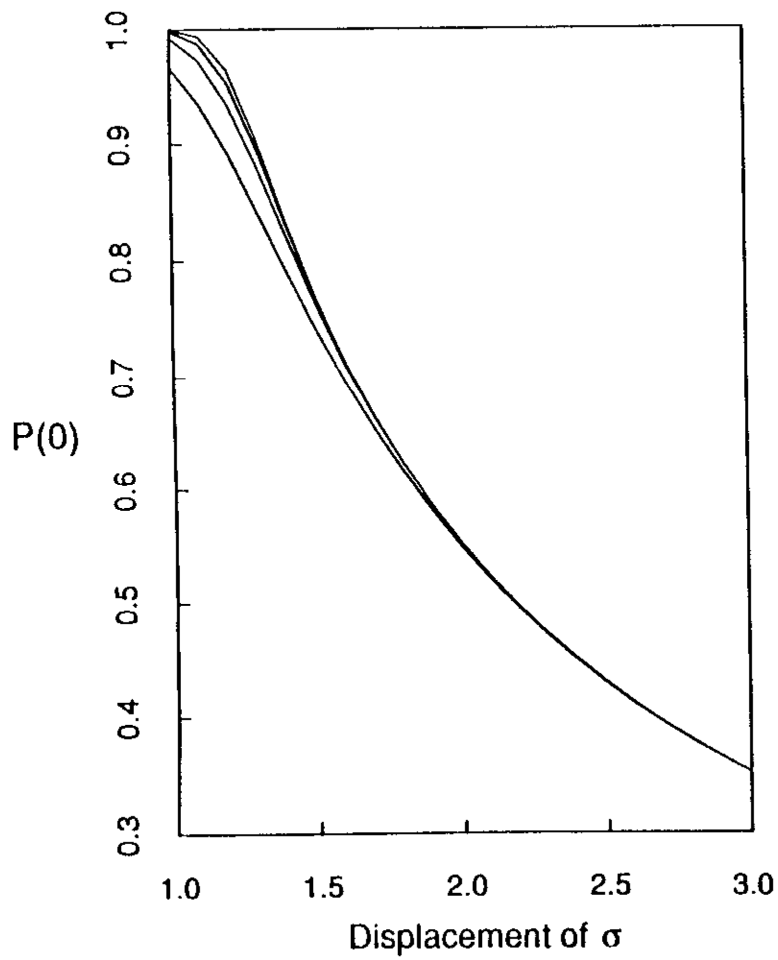

Fig. 2a Value of P(0) for h = 5,10,15,20

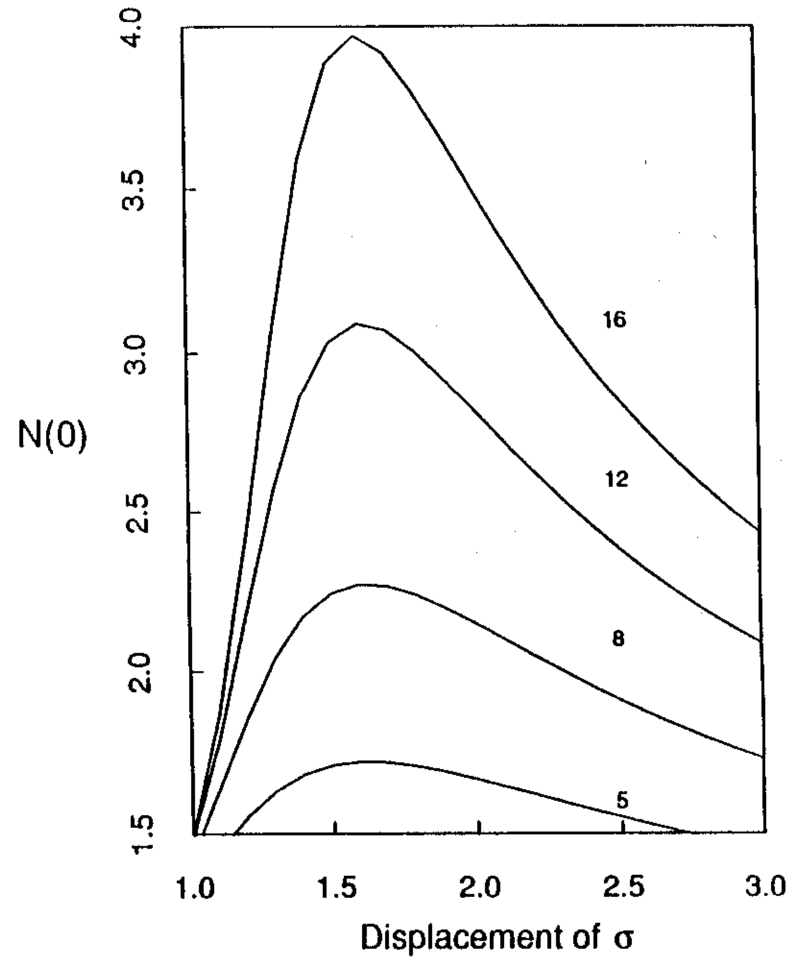

Fig. 2b Average Sample Number for h = 5,8,12,16

dary, is a decreasing function of σ and increases with h, for fixed values of σ. The average sample number, $N(0)$, is a unimodal function of σ and increases for fixed values of h.

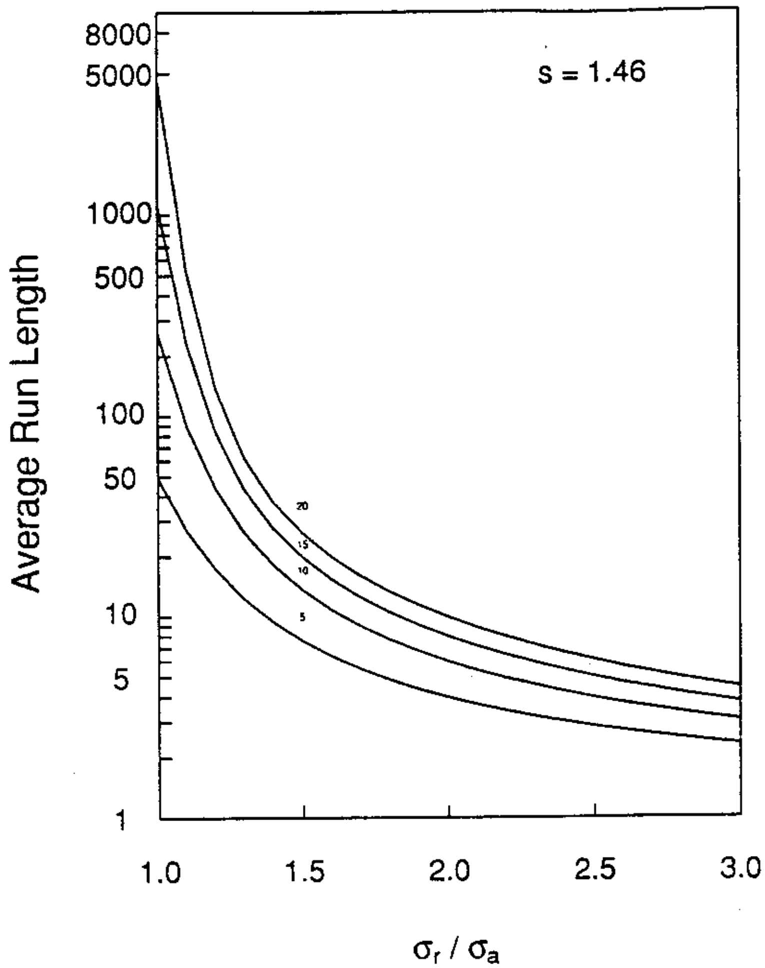

Fig. 3a Average Run Length for h = 5,10,15,20

Figure 3a shows the values of A.R.L, in logarithmic scale, for four different values of $h = 5,\ 10,\ 15,\ 20$ and $s = 1.46$. The figure was constructed to give values of the A.R.L for changes in the standard deviation from 1 to 3. Subsequently, we assume that the variable under study has been scaled to give $\sigma_a = 1$.

Note that the A.R.L is a decreasing function of σ, and that large values of A.R.L correspond to large values of h. Other graphs for values of h in the interval (5,20) can be obtained by logarithmic interpolation.

An example will help to clarify how this graphs can be used to obtain desired A.R.L values corresponding to some predefined values of h and s. Let us suppose that a cusum chart, like figure 1, has been constructed with $h = 10$ and $s = 1.46$, to detect changes in the standard deviation of a continuous process, and that the variable representing this process has been scaled to give $\sigma_a = 1$. Figure 3a shows that even if the process variability is maintained at an acceptable level, i.e. $\sigma_a = 1$, in the average we are going to have a false alarm every 250 observations.

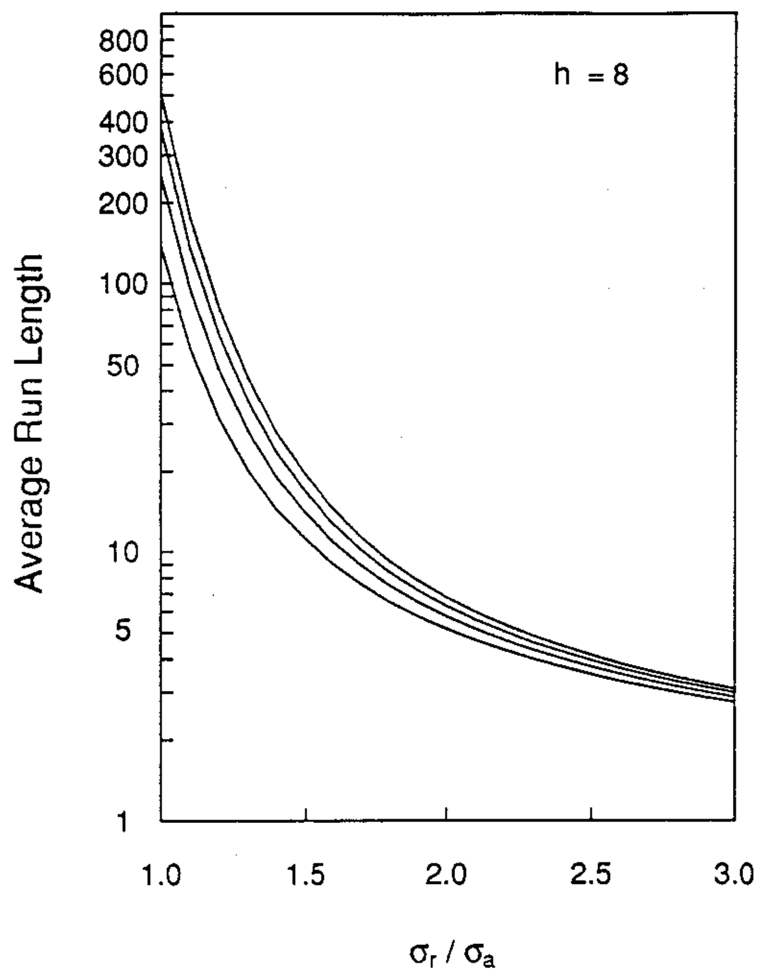

Fig. 3b Average Run Length for s = 1.46,1.85,2.18,2.47

On the other hand, if the variability of the process increases to a rejectable level $\sigma_r = 2$, on the average we expect the cusum chart to detect the change after 6 observations. If we choose the decision interval to be $h = 20$, we would have "false alarms" once in every 5000 observations, and we would be able to detect a change of a 100% ($\sigma_r = 2$), in approximately 10 observations. Figure 3b was constructed in a similar way with h fixed and equal to 8, and different values of s.

The figures 3a and 3b summarize the behaviour of cusum charts designed to detect changes in the standard deviation up to 200%, and can be used to select the optimum values of h and s. However, this procedure is impractical since it requires a lot of calculation. In the next section a contour nomogram, similar to the one given by Goel and Wu (1971) for means, is introduced, that simplifies the design of the cusum charts for variances.

3. CONTOUR NOMOGRAM

As we have seen, The A.R.L depends on the values of s, h and the process variability σ. At the two commonly specified values of σ, σ_a and σ_r, the A.R.L is denoted L_a and L_r respectively. It is desiderable to have large values of L_a (as few false alarms as possible), and small values of L_r (quick detection of changes).

The contour nomogram with contour lines L_a and L_r is a very useful aid in the design of cumulative sums charts, since it allows us to obtain A.R.L values without having to compute the integral equations (2.2) and (2.3).

The nomogram was constructed using a two dimensional grid with values of σ_r/σ_a ranging from 1.2 to 3, and values of h/σ_a^2 ranging from 5 to 20. The values of L_a and L_r were obtained by using the method and programs described in the appendix. Using these A.R.L values the contours of L_a from 40 to 75000, and the contours of L_r from 2.5 to 65, were computed using bivariate interpolation (Akima (1978)), as implemented in the statistical software S.

Figure 4 shows the contours of L_a and L_r superimposed and the values of σ_r/σ_a, h/σ_a^2 and s/σ_a^2. Since the s scale is nonlinear, is more accurate to use formula (3.1) to obtain s than trying to

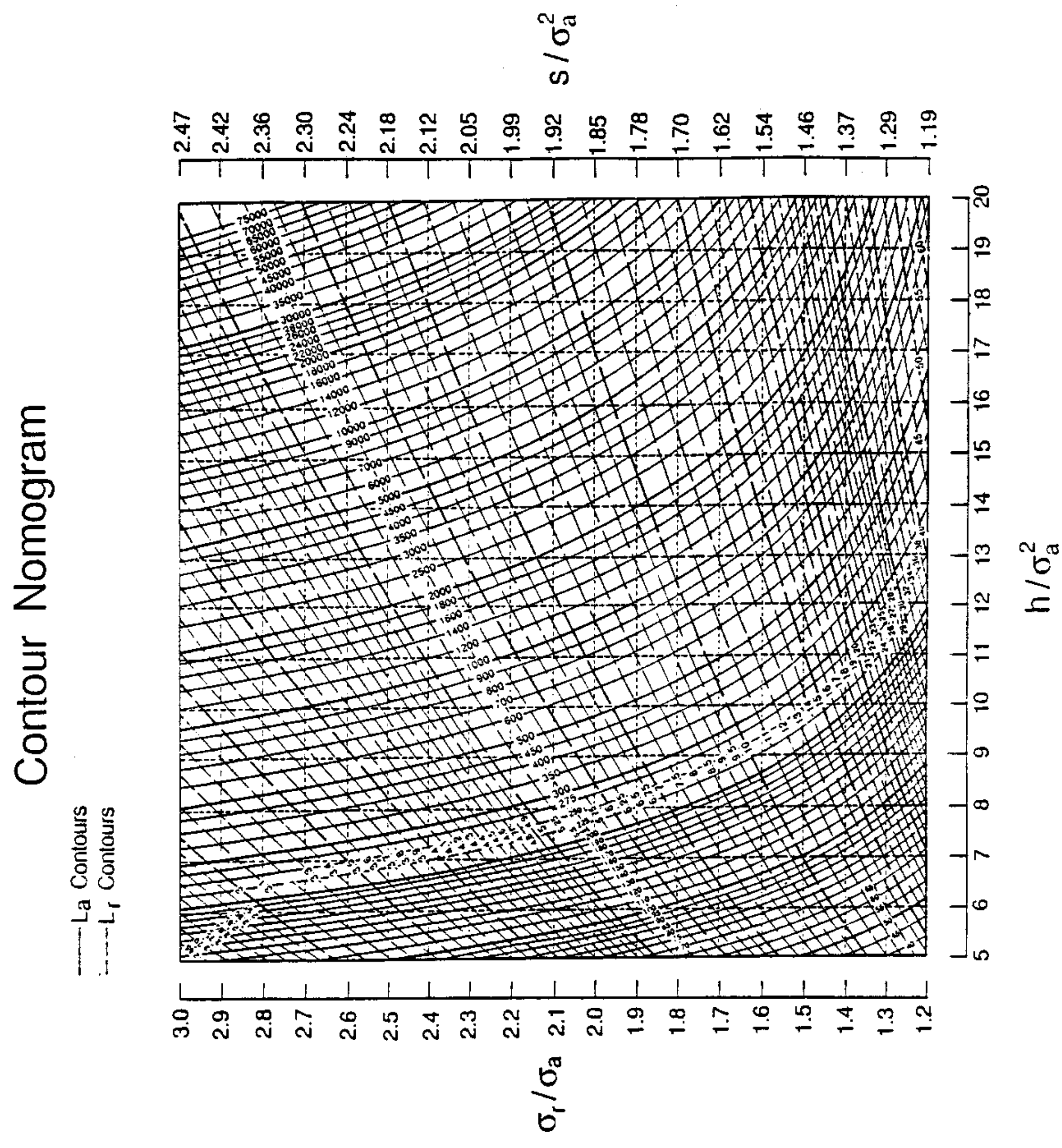

Fig. 4 Contour Nomogram showing values of L_a and L_r

interpolate in that scale.

Design of Cusum Charts

Let x_i be the results of a process whose variability we wish to monitor; and let σ_a and σ_r denote the acceptable and rejectable variability levels respectively. A cusum chart for monitoring changes in the variability of the process can be constructed as follows:

1) Determine the acceptable and rejectable quality levels of the process variability, σ_a and σ_r.

2) Once the values of σ_a and σ_r have been defined, the reference value s is obtained by means of the equation

$$s = \frac{log\,(\frac{\sigma_r^2}{\sigma_a^2})}{\frac{1}{\sigma_a^2} - \frac{1}{\sigma_r^2}}, \tag{3.1}$$

This is the value of the slope of the decision lines in a sequential test Wald (1947); and is a "mean" value between the acceptable and rejectable values of the process variability, that provides maximum discrimination between σ_a and σ_r.

3) The value of h is chosen with the help of the nomogram to yield the desired values of L_a and L_r. A horizontal line is drawn in the nomogram, corresponding to the value of σ_r/σ_a, giving various possible combinations of h, L_a and L_r.

4) For each observation x_k compute the cumulative sum

$$S_k = \sum_{i=1}^{k} [(x_i - \mu)^2 - s]$$

and declare that an increase in the variabilty of the process has ocurred whenever $S_k \geq h$. Reset to zero whenever $S_k < 0$.

Example 1: Suppose we wish to design a cusum chart to control changes in the standard deviation from $\sigma_a = 2$ to $\sigma_r = 4$. With these values a reference value s is calculated using the formula given above, giving $s = 7.39$. A horizontal line can now be drawn on the nomogram at the point $\sigma_r / \sigma_a = 2$, giving various possible combinations for L_a and L_r as shown in the table below.

$s = 7.39$		
h	L_a	L_r
20	73	4.7
24	112	4.8
28	169	5.3
32	252	5.8
36	374	6.2
40	552	6.7
44	812	7.2
48	1190	7.7
52	1742	8.1
56	2543	8.6
60	3710	9.1
64	5385	9.5
68	7844	10.0
72	11398	10.5
76	16531	10.9
80	23884	11.4

Table 1. Values of h, L_a and L_r

From this table a number of possible designs are available, the choice depending on the requirements of the process.

In some instances it is possible to specify the values of L_a and L_r that that are to be used in order to detect an increase in variability from some known acceptable level σ_a. In this case the nomogram can be used not only for designing the cusum chart but to describe the magnitude of the change that we can detect.

Example 2: Suppose we want to design a cusum chart with approximate Average Run Lengths, $L_a = 1200$ and $L_r = 7$ for a process with mean centered at target and equal to 13 and $\sigma_a = 2$. From the contour nomogram the point of intersection of the contours $L_a = 1200$ and $L_r = 7$ gives $\sigma_r / \sigma_a \approx 2.08$ and $h/\sigma_a^2 \approx 11.7$ as illustrated in figure 5. Hence, $\sigma_r \approx 4.16$, and $h \approx 46.8$. Using formula (3.1) with $\sigma_r = 4.16$, and $\sigma_a = 2$, we obtain $s = 7.62$.

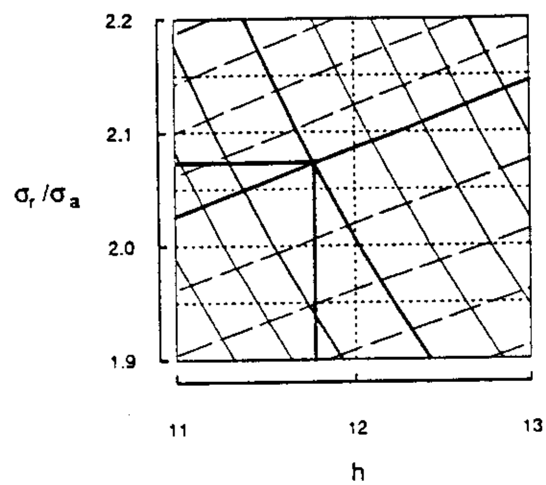

Fig. 5 Section of nomogram

The procedure consists of calculating or plotting, at each point k the cumulative sum

$$S_k = \sum_{i=1}^{k} [(x_i - 13)^2 - 7.62]$$

and declaring that an increased in standard deviation has occurred whenever $S_k \geq 46.8$; resetting to zero whenever $S_k < 0$.

4. A.R.L IN THE CASE OF NON-NORMALITY

Throughout this work we have been concerned with the determination of the Average Run Length in the case in which the observations coming from the process are normally distributed. In this section we relax that assumption, to check the robustness of the A.R.L values, by assuming that the observations are distributed according to the family of exponential power distributions. This is a class of symmetric distributions that includes the Normal, the Double Exponential, and the Rectangular distributions. The probability density function for this family can be written (Box and Tiao (1973))

$$f(y) = \frac{\omega(\gamma)}{\sigma} \exp\left[-c(\gamma) \left| \frac{y-\mu}{\sigma} \right|^{\frac{2}{1+\gamma}} \right] \qquad -\infty < y < \infty \tag{4.1}$$

where

$$\omega(\gamma) = \frac{\Gamma[(\tfrac{3}{2}(1+\gamma)]^{1/2}}{(1+\gamma)\,\Gamma[(\tfrac{1}{2}(1+\gamma)]^{3/2}}, \quad \text{and} \quad c(\gamma) = \left\{ \frac{\Gamma[(\tfrac{3}{2}(1+\gamma)]}{\Gamma[(\tfrac{1}{2}(1+\gamma)]} \right\}^{\frac{1}{1+\gamma}}$$

The parameters μ and σ are the mean and standard deviation of the population; and the parameter $-1 < \gamma \leq 1$ can be regarded as a measure of kurtosis indicating the extent of non-normality of the population. In particular, when $\gamma = 0$, the distribution is normal, when $\gamma = 1$ is the double exponential.

Under this assumption, the cumulative sums can be deduced from the sequential probability ratio test see Wald (1947), and are given by

$$S_k = \sum_{i=1}^{k} [(x_i - \mu)^{\frac{2}{1+\gamma}} - s\,] \tag{4.2}$$

and the distribution of the increments $[(x_k - \mu)^{\frac{1}{2+\gamma}} - s]$ is given by

$$f(y) = (1+\gamma)\frac{\omega(\gamma)}{\sigma^2}(y+s)^{\frac{\gamma-1}{2}} \exp\left[\frac{-c(\gamma)}{\sigma^{\frac{2}{1+\gamma}}}(y+s) \right] \qquad y \geq -s \tag{4.3}$$

and 0 otherwise.

The Average Run Length was calculated using the methods described in section 2 for a combination of values of s and h. For negative values of γ the convergence of the A.R.L values is slow and requires a large number of points in the quadrature to obtain desired accuracy.

Figure 6 shows A.R.L for $h = 8$ and $s = 2$. Note that assuming normality can lead to considerable inaccuracies, especially for γ's greater than 0.5 or −0.5. For example, $L_a = 250$ if we assume normality but $L_a \approx 1200$ for $\gamma = 0.5$. Similarly, $L_r = 4$ in the normal case, while $L_r \approx 2$ for $\gamma = -0.5$. Hence assuming normality will produce too many false positives in the first case, and will be slow in detecting changes in the second case. For values of $-0.1 \le \gamma \le 0.1$ the A.R.L values are close to those of the normal distribution; i.e. $\gamma = 0$.

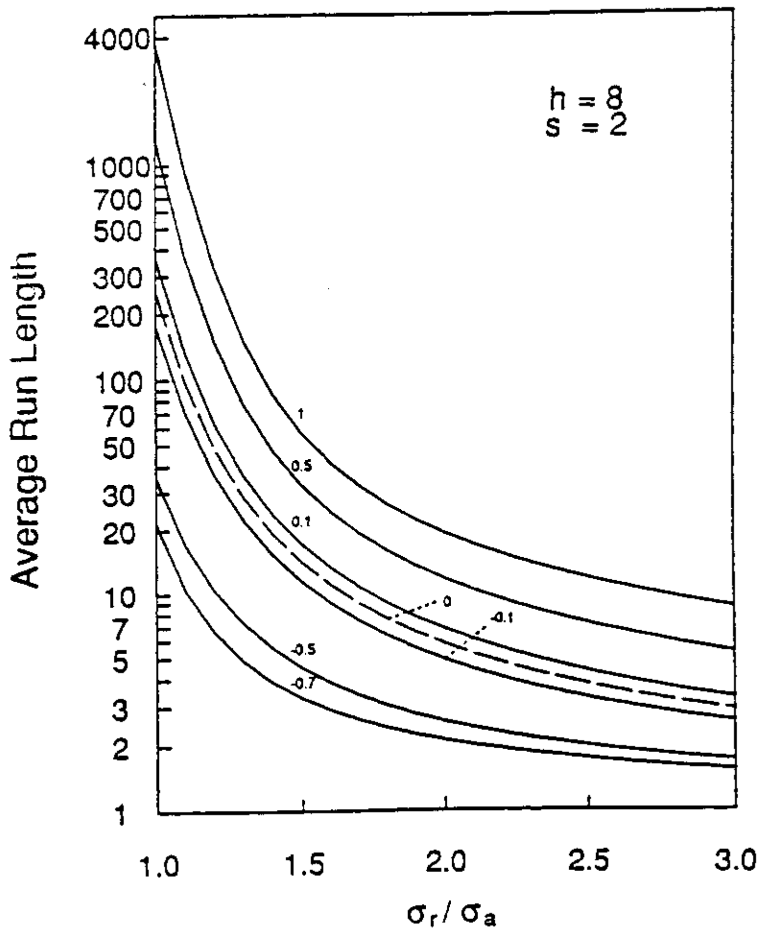

Fig. 6 A.R.L for γ = -0.7, -0.5, -0.1, 0, 0.1, 0.5, 1

5. CONCLUSIONS

Once the mean level of a process has been centered at the desired target value, the objective is to minimize or at least to maintain the variability of the process at acceptable levels of performance. In this report we have given a simple method to design a cusum chart when the objective is to control the variability of a continuous process.

The advantages of this procedure is that it allows one to monitor changes in variance by using only one observation at a given time k. It is also important to note that plotting the cusum chart is not necessary, since we can calculate the value of the cumulative sums and check when the sum is greater or equal to some decision interval h.

When the sample size at each time k is greater than 1, m say, we can compute the mean of the m observations at time k, $\bar{X}_{i\cdot} = \frac{1}{m}\sum_{j=1}^{m} x_{ij}$, and plot the cumulative sums

$$S_k = \sum_{i=1}^{k} [(\bar{X}_{i\cdot} - \mu)^2 - s]. \tag{5.1}$$

In this case the $\bar{X}_{i\cdot}$ are distributed as $N(\mu, \sigma^2/m)$; and this amounts to a change of scale from σ to $\sigma/\sqrt{m}$. Therefore, the contour nomogram given in section 3, can be used to obtain A.R.L values and design the cusum chart.

APPENDIX

The integral equations (2.2) and (2.3) can be replaced by a system of linear equations and solved for the unknown variables. The system employed is similar to the one given by Kantorovich and Krylov (1964), in which a partition of the interval of integration is constructed and the integrals are approximated in each of the resulting subintervals.

Let $\{a_0, a_1, ..., a_n\}$ be a partition of the interval $[0,h]$ where $a_0 = 0$ and $a_n = h$. Different partitions are obtained by using different quadratures. When h and n are small, Gaussian quadrature is recommended; however, when the value of h gets large it is necesary to use a larger partition of the interval and it is better, for computational purposes, to use a more flexible quadrature, for example a tangential one in which the distance between consecutive a_i's is constant and equal to h/n. The equation (2.2) can then be written as

$$P(z) = \int_{-s}^{-z} f(y)dy + \sum_{j=1}^{n} \int_{a_{j-1}}^{a_j} P(x) f(x-z)dx \tag{A.1}$$

Since the function $P(z)$ is smooth and decreasing in the interval $(0,h)$, and assuming that the subintervals (a_{j-1}, a_j) are sufficiently small so that $P(x)$ is constant and equal to $P(z_j)$, where $z_j = \dfrac{a_j + a_{j-1}}{2}$, equation (A.1) can be approximated by

$$\tilde{P}(z) = \int_{-s}^{-z} f(y)dy + \sum_{j=1}^{n} \tilde{P}(z_j) \int_{a_{j-1}}^{a_j} f(x-z)dx \tag{A.2}$$

Substituting z for the values z_i, $i = 1, .., n$, and letting

$$\mathbf{b} = (b_1, b_2, ..., b_n) \ , \quad \text{where} \quad b_i = \int_{-s}^{-z_i} f(y)dy \tag{A.3}$$

$$\mathbf{C} = (c_{ij}), \quad \text{where} \quad c_{ij} = \int_{a_{j-1}}^{a_j} f(x - z_i)dx \tag{A.4}$$

The vector $\tilde{\mathbf{P}} = (\tilde{P}(z_1), ..., \tilde{P}(z_n))$ can be computed as

$$\tilde{\mathbf{P}} = (\mathbf{I} - \mathbf{C})^{-1}\, \mathbf{b} \tag{A.5}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix. Similarly a numerical approximation for the function $N(z)$ is given by

$$\tilde{\mathbf{N}} = (\mathbf{I} - \mathbf{C})^{-1}\, \mathbf{e}, \qquad \text{where} \quad \mathbf{e} = (1, 1, ..., 1). \tag{A.6}$$

The value of $\tilde{P}(0)$ is obtained by setting $z = 0$ in equation (A.2)

$$\tilde{P}(0) = \int_{-s}^{0} f(y)dy + \sum_{j=1}^{n} \tilde{P}(z_j) \int_{a_{j-1}}^{a_j} f(x)dx \tag{A.7}$$

Similarly for $\tilde{\mathrm{N}}(0)$ we have

$$\tilde{N}(0) = 1 + \sum_{j=1}^{n} \tilde{N}(z_j) \int_{a_{j-1}}^{a_j} f(x)dx \tag{A.8}$$

A computer program using FORTRAN 77, was written to compute $N(0)$, $P(0)$, and A.R.L. The program was run on a VAX 11/780 computer running VMS. The IMSL subroutines, *LEQIF* and *MDCH*, were used to solve the system of linear equations and to compute the values of the Chi-squared distribution (2.4) respectively.

The accuracy of the approximation was evaluated by obtaining values of A.R.L from equation (2.1), and checking the convergence of the solutions for an increasing number of points in the quadrature. For the ranges of s and h under study, the computed A.R.L converged to a fixed value when the number of points in the quadrature was less than 500.

REFERENCES

Akima, H. (1978), "A Method of Bivariate Interpolation and Smooth Surface Fitting for Irregularly Distributed Data Points," *ACM Transactions on Mathematical Software*, 4, 148-164.

Box, G. E. P., and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, Reading, Massachusets: Addison-Wesley.

Goel, A. L., and Wu, S. M. (1971), "Determination of A.R.L. and a Contour Nomogram for Cusum Charts to Control Normal Mean," *Technometrics*, 13, 221-230.

Kantorovich, L. V., and Krylov, V. I. (1964), *Approximate Methods of Higher Analysis*, New York: John Wiley & Sons.

Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, 41, 100-114.

Wald, A. (1947), *Sequential Analysis*, New York: John Wiley & Sons.

Woodward, R. H., and Goldsmith, P. L. (1964), *Cumulative Sum Techniques*, London: Oliver & Boyd. (I.C.I. Monograph No. 3)