UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

CONTROLLING THE STANDARD DEVIATION BY CUSUMS AND WARNING LINES

by

E. S. Page

March, 1963

This research was supported by the Office of Naval Research under contract No. Nonr-855(09) for research in probability and statistics at the University of North Carolina, Chapel Hill, N. C.

Institute of Statistics
Mimeo Series No. 356

CONTROLLING THE STANDARD DEVIATION BY CUSUMS AND WARNING LINES

by

E. S. Page*

University Computing Laboratory, Newcastle upon Tyne

and

University of North Carolina, Chapel Hill

= =

INTRODUCTION

The usual process inspection scheme for controlling the standard deviation (S.D.) of a normal distribution is to plot a Shewhart chart on which are recorded the ranges of small samples; a common size is 5 items. Action is taken if any sample range exceeds a previously determined amount. The rapidity with which the scheme detects a departure from the accepted value of σ can be increased by considering the results of several of the latest samples together. A fixed number of samples can be taken into account by Warning Lines on the Shewhart-type chart (4); a more sensitive scheme uses cumulative sums (cusums) of the sample ranges together with a rule for deciding what change of direction of the cusum path should require action. Both these approaches can be applied to gauged observations. In this paper we compare these four schemes and provide a short table of cusum schemes using the range.

WARNING LINES SCHEME

The chart for a warning line scheme looks like Figure 1. We are interested in detecting an increase in variance so that corrective action can be taken. The chart has plotted on it the ranges of samples of size N; one rule that can be

---

*This research was supported in part by the Office of Naval Research under contract No. Nonr-855(09) for research in probability and statistics at the University of North Carolina, Chapel Hill, N. C. Reproduction in whole or in part is permitted for any purpose of the United States Government.

adopted is: "Take action if any sample point falls above the Action Line at $B_1\sigma$ or if n consecutive points fall above the Warning Line at $B_2\sigma$." The positions of the lines, defined by B_1, B_2, and the number, n, of consecutive points for action can be selected to give the scheme desirable properties while the sample size, N, is in practice usually dictated by convenience and the target standard deviation, σ, is subject to specification or is based on the history of satisfactory operation of the process.

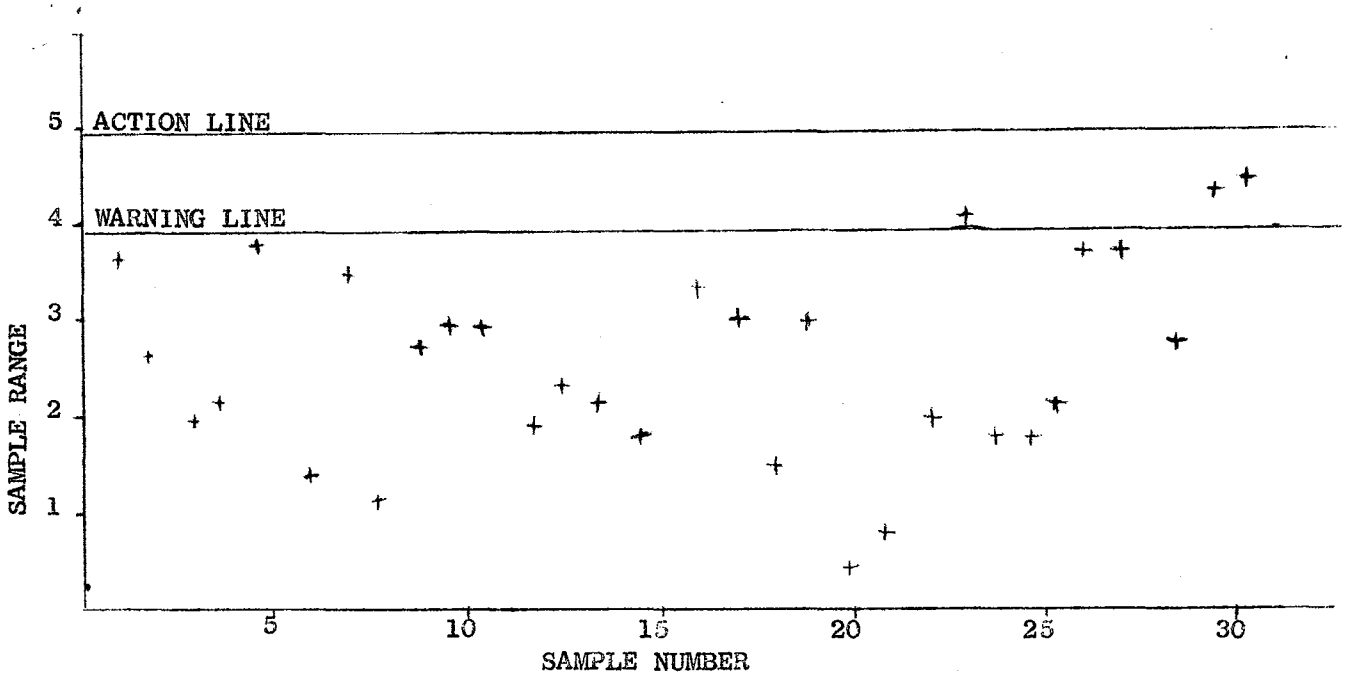

Figure 1. Warning line chart for ranges.

If the probabilities of a sample point falling into the several regions of the chart remain constant and samples are independent, the Average Article Run Length*

*The Average Run Length (A.R.L.) has become accepted in the literature as the average number of samples taken before a signal showing the need for action is given. Unfortunately in some early work the term was used in two senses. Accordingly we define the Average Article Run Length as the average number of items or articles sampled before action is taken and the Average Sample Run Length (A.S.R.L.) as the average number of samples before action is taken. Thus the A.S.R.L. is what is now the accepted usage of the A.R.L. but only when there is no risk of confusion will we drop the word "Sample".

(A.A.R.L.) is (4)

$$L = \frac{(1-p_2^n)N}{1-p_2-p_1(1-p_2^n)} \qquad (1)$$

where p_1 = prob. (point falls in the good region)

p_2 = prob. (point falls in the warning region).

The distribution of the run length may be obtained easily in terms of p_1, p_2, n (6). For the ranges of samples from a normal population $N(\mu,\sigma')$

$$\begin{aligned} p_1 &= F_N(B_2\sigma/\sigma') \\ p_2 &= F_N(B_1\sigma/\sigma') - F_N(B_2\sigma/\sigma') \end{aligned} \qquad (2)$$

where $F_N(w)$ is the distribution function of the range in samples of N from an N(0,1) population.

CUMULATIVE SUM SCHEME

The cumulative sum scheme to detect an increase in the mean, m, of a distribution with frequency function $g(x)$ can be represented by plotting the cusum $S_r = \sum_{i=1}^{r} (x_i - k)$ on a chart and taking action when S_r rises an amount h above its minimum. (Fig. 2) The quantity k which is subtracted from each of the x_i before entering the cusum is called the reference value and the critical rise, h, in the path of the cusum which requires action to be taken is the decision interval. Both k and h can be chosen to obtain satisfactory properties for the scheme. In order to control the variance the ranges of samples may be

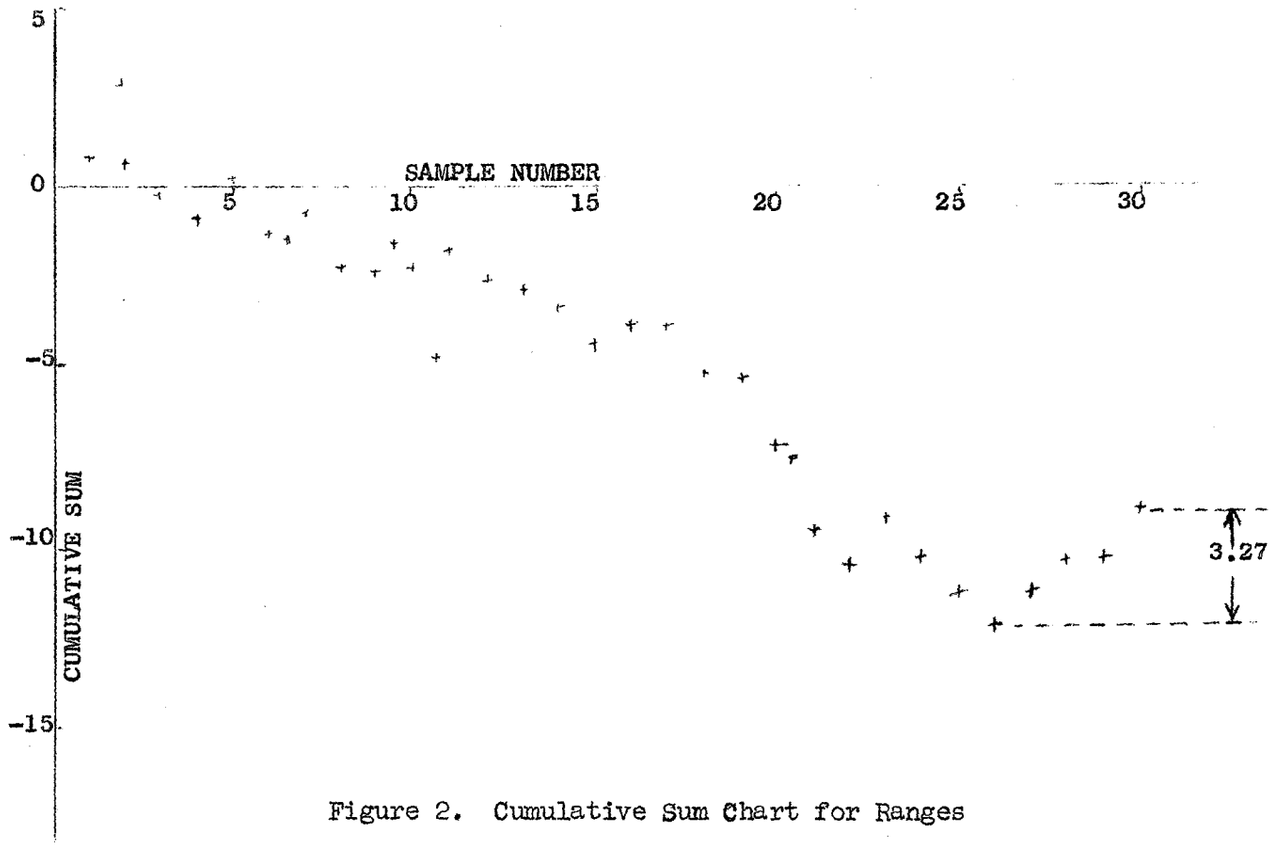

Figure 2. Cumulative Sum Chart for Ranges

plotted so that the x_i are ranges of normal samples of size N; in this case g(x) is the frequency function of the range in normal samples of size N and its mean is thus the expected range. If the expected range at the target variance is m the reference value will usually exceed m so that long uninterrupted runs of production will occur when the standard deviation is satisfactory.

The A.A.R.L. is NL(0) where L(Z) satisfies the integral equation (3)

$$L(Z) = 1 + \Psi(-Z)L(0) + \int_0^h L(y)\,\psi(y-Z)dy \qquad (3)$$

where $\psi(y)$ is the frequency function of the increments $x_i - k$ in the cusum and $\Psi(-Z) = \int_{-\infty}^{-Z} \psi(y)dy$ (of course $\Psi(y) = 0$ for $y \leq -k$). The A.A.R.L. may alternatively be derived in terms of the characteristics of a Wald sequential test; we

have

$$L(0) = \frac{N(0)}{1-P(0)} \qquad (4)$$

where $N(Z)$, $P(Z)$ are the Average Sample Number and Operating Characteristic of the Wald test with increments x_i-k, starting score Z and boundaries at 0, h. The solution of (3) and the calculation of $N(Z)$, $P(Z)$ are discussed in the Appendix and a table of schemes is given.

GAUGING

Stevens (8) described the operation of the conventional Shewhart chart using gauged observations. For controlling a normal population mean μ, S.D. σ, the numbers a, b, c of a sample of size N which fell in the regions $(-\infty, \mu-\alpha\sigma)$ $(\mu-\alpha\sigma, \mu+\alpha\sigma)$, $(\mu+\alpha\sigma, \infty)$ respectively were counted; charts of $(c-a)$ and $(c+a)$ were kept for indicating changes in mean and standard deviation. Some efficiency in the detection of changes is certainly lost but a suitable choice of the width, $2\alpha\sigma$, of the gauge can give efficiencies well worth considering if, as is usual, gauging is quicker, cheaper and more convenient than measuring. This type of gauge has been used in a cusum scheme for controlling the variance (5). The procedure, which can be operated automatically in a simple way, forms the sum $S'_r = \sum_{i=1}^{r} (x'_i - k')$, where $x'_i = 0,1$ according as the i-th observation is within or outside the gauges and takes action where S'_r rises a height h' above its previous minimum. In this case k', h' are the reference value and decision interval suitable for controlling the zero-one binomial variate. If k', h' are chosen to be rational a simple scoring scheme is obtained.

A warning line scheme could be applied using the Stevens technique of scoring or the warning line scheme could be applied to the single observations; for example the rule could be: Take action if any $|x_i-\mu| \geq B_1\sigma$ or if any sequence of n

consecutive observations have $B_2\sigma \leq |x_i-\mu| < B_1\sigma$. This rule has the disadvantage that a change in mean will call for speedy action and, thus, in isolation would give a false indication of the type of change. The A.A.R.L. is given by (1) with $N = 1$ and p_1, p_2, expressed by the normal integral. The scheme is a special case of one to control mean and variance simultaneously.

COMPARISONS

The first comparison that needs to be made is between the usual Shewhart range chart, a warning line scheme and a cumulative sum scheme using ranges from the same sizes of samples. In order to give an A.S.R.L. = 200 samples of size 5 observations when the process is at target S.D. the Shewhart chart needs an Action Line at $B_1\sigma$ where $B_1 = 4.886$. For the Shewhart Scheme, the A.S.R.L. = $1/p$ where p = prob (a given sample range exceeds $B_1\sigma$); accordingly here we take B_1 so that there is probability 0.005 in the upper tail of the distribution i.e. the 99.5% point. Many different combinations of the positions of Action and Warning Lines i.e. $B_1\sigma$ and $B_2\sigma$, and number of samples, n, used in the rule, give schemes with an A.S.R.L. = 200; similarly different combinations of decision interval and reference value, h and k, are possible for the cusum rules. This choice of A.S.R.L. to be 200 at the target value is for illustration only. The A.S.R.L. at the target value shows how frequently we shall interfere with a satisfactory process by following the rules of our process inspection scheme. It corresponds to the choice of the risk, α, of making errors of the first kind in significance testing. The performance of one rule of each kind is shown in Table 1. An example of their

TABLE I

Comparison of Control Chart and Cumulative Sum Schemes

	Average Sample Run Lengths						
σ	1.0	1.1	1.2	1.3	1.4	1.5	2.0
Shewhart	200	69	31	17	10	7.0	2.4
Warning Line	200	63	27	14	8.9	6.2	2.3
Cusum	200	46	19	10	7.1	5.3	2.5

Shewhart scheme : $B_1 = 4.886$

Warning Line scheme : $B_1 = 5.0$, $B_2 = 3.97$, $n = 2$

Cusum scheme : $h = 3.201$, $k = 2.8$

operation to control the S.D. at unity is given by Figs. 1 and 2. The Warning Line scheme calls for action after sample 31, since samples 30 and 31 both give ranges above the Warning level. The Shewhart scheme has not yet called for action. The cusum scheme which records $\Sigma(w_i - 2.8)$ where the w_i are the ranges of samples of five observations in this case calls for action after sample 30 since the rise of 3.27 in the path exceeds $h = 3.201$. The position of the minimum in Fig. 2 suggests that an increase in S. D. has taken place at about sample 26; a crude estimate (usually an over-estimate) of the new expected range is k + (height of the rise in path)/(number of observations since the minimum) = $2.8 + 3.27/5 = 3.45$ so that the new S. D. is thought to be about $3.45/2.33 \sim 1.5$ (since $E(w|\sigma = 1) = 2.33$). In fact the observations were derived from the first page of Wold's table of Random Normal Deviates and the S. D. was increased to 1.4 after sample 25. The recording can be done numerically instead of graphically if preferred.

The point to be made is that both the Warning Line scheme and the cusum scheme detect small and moderate changes in σ more quickly than the Shewhart scheme. The increase in speed of detection from the Warning Line scheme is quite small - 10 - 20$^{o}/o$ - but it is obtained with hardly any extra work or change in practice. All that is required is an extra line on the chart and the need to look at the last two results instead of just the last one. The cusum scheme has rather greater sensitivity and needs a different method of recording rather than greater effort; it has additional benefits by suggesting the position and amount of any departure from target value. The next comparison concerns the choice of decision interval, h, and reference value, k, for the cusum schemes. In their work on controlling the mean of a normal distribution Ewan and Kemp (2) suggested that k should be taken approximately halfway between μ_o or μ_1 where μ_o is the target value and μ_1 is the value of the mean that needs to be detected. Of all the cusum schemes with the same A.R.L. at the target value μ_o this choice of reference value, k, gives approximately the scheme with least A.R.L. at mean μ_1. For a mean of a normal distribution this suggestion has some theoretical backing which is absent for the range, but the rule seems to be a good guide in this case also. For example, if w is the range in samples of 5 from a normal population with S. D. equal to σ, $E(w|\sigma) = 2.326\sigma$, $E(w|3\sigma/2) = 3.489\sigma$ and $E(w|2\sigma) = 4.652\sigma$. Thus the rule suggests taking $k \doteq 2.91\sigma$, for detection of a change to 1.5σ, and $k \doteq 3.49\sigma$ for 2σ. Table 2 compares six schemes.

TABLE 2: Comparison of cusum schemes with different reference values

k	h_σ	Average Sample Run Lengths						
		1	1.1	1.2	1.3	1.4	1.5	2.0
2.8	3.201	200	46	19	10	7.1	5.3	2.5
2.9	2.791	200	48	19	11	7.1	5.2	2.4
3.0	2.473	200	51	20	11	7.1	5.2	2.3
3.25	1.908	200	57	23	12	7.6	5.4	2.2
3.5	1.513	200	62	26	13	8.3	5.7	2.2
3.75	1.196	200	65	28	15	8.9	6.1	2.3

The application of a cusum scheme to gauged observations is described in (5) and tables were given. In this case we use the Average Article Run Length for the comparison so that the numbers of observations correspond. The scheme taken from Appendix Table C of (5) which has A.A.R.L. nearest to 1000 to match the range scheme $N = 5$, A.S.R.L. = 200 is shown in Table 3. The gauges are placed at $\mu_o \pm G\sigma$, where, in this example, $G = 1.5$.

TABLE 3: Comparison of cusum schemes using gauges and ranges

Average Article Run Lengths

σ	1	1.1	1.5	2.0
Gauging ($G = 1.5$, $k = 0.2$, $h = 4.2$)	956	242	32	16.5
Ranges ($k = 2.8$, $h = 3.201$)	1000	230	27	12

The scheme based on ranges is certainly more sensitive to departures from target S. D. but the performance of the gauging scheme is sufficiently good for it to be worth consideration, especially as it is so easy to implement. However the gauging scheme on the data of Figs. 1 and 2 requires action at the same time as the other

Finally we consider a Warning Line scheme based on gauging; the scheme is: Take action if $|x-\mu_o| \geq B_1\sigma$ or if n consecutive points $|x_i-\mu_o|$ be in the interval $B_2\sigma$, $B_1\sigma$. Different positions of the Warning and Action lines giving the same A.A.R.L. at the target values of mean and S. D. make an appreciable change in the characteristics of the scheme. In addition, a change in process mean will lead to a reduced A.R.L. even though the S. D. remains constant. However the simplicity of the scheme makes it just worth considering. A good scheme with A.A.R.L. = 1000 at the target value is obtained with $B_1 = 3.50$, $B_2 = 2.259$ and $n = 2$. (Table 4). It is much less sensitive than the cusum scheme to

TABLE 4: Comparison of cusum scheme on ranges and Warning Line scheme on $|x_o-\mu_o|$

	Average Article Run Lengths			
σ	1.0	1.1	1.5	2.0
Warning lines on $\|x-\mu_o\|$	1000	346	32	9.3
Cusum on ranges	1000	230	27	12

small changes in σ but notices large changes well. The crossing of the A.A.R.L. curves at a large departure from the target value occurs in this case just as it does when controlling means. The difficulty is that changes in mean and S. D. often give the same indications in this rule.

THE TABLE OF CUSUM SCHEMES

Table 5 gives the reference value, decision interval and A.S.R.L. for six schemes. The three schemes with reference value $k = 2.9$ are appropriate for noticing quickly changes from the target S. D.σ to about 1.5σ, and those with

$k = 3.5$ changes to 2σ. For each reference value the table shows the corresponding decision interval to give a specified A.S.R.L. at the target value of the S.D. The three A.S.R.L. at the target have been chosen to be 100, 200 and 500 so that a process operating satisfactorily at the target S.D. would be interrupted only once in every 100 (200, 500) samples on the average. In all cases samples of five observations are taken and the values in the tables assume normality in the parent population.

TABLE 5: Table of cusum schemes for controlling the standard deviation of a normal population using ranges of samples of five.

Reference Value k	Decision Interval h	Average Sample Run Lengths						
		σ	1.1σ	1.2σ	1.3σ	1.4σ	1.5σ	2σ
2.9	2.268	100	31	14	8.4	5.8	4.4	2.1
2.9	2.791	200	48	19	11	7.1	5.2	2.4
2.9	3.529	500	86	29	14	8.9	6.5	2.8
3.5	1.184	100	36	17	9.8	6.5	4.7	2.0
3.5	1.513	200	61	26	13	8.3	5.7	2.2
3.5	1.946	500	120	43	20	11	7.4	2.6

EXAMPLE

A scheme is required to control the S. D. of a normal population at a target value σ so that any increase to S. D. 1.5σ is noticed as quickly as possible when interruption to a process with satisfactory S. D. should not occur more frequently on the average than once in a hundred samples.

In this case the A.S.R.L. on target is 100. For noticing changes to 1.5σ we need $k = 2.9$; from the table, the decision interval $h = 2.268$. Accordingly we

take samples of five observations and calculate their ranges x_i. The cumulative sum $S_r = \sum_{i=1}^{r} (x_i - 2.9\sigma)$ is recorded and action taken whenever S_r rises a height 2.268 above its previous minimum. The recording may be performed on a graph, or numerically as the sequence of figures $S_r - \min_{1 \leq i \leq r} S_i$.

ACKNOWLEDGEMENT

I wish to thank Miss E. D. Barraclough for programming and performing all the calculations described in this paper.

REFERENCES

1. Cox, D. R. (1949) The use of the range in sequential analysis. J. Roy. Statist. Soc. B 11, 101-114.

2. Ewan, W. D. and Kemp, K. W. (1960) Sampling inspection of continuous processes with no autocorrelation between successive results. Biometrika 47, 363-380.

3. Page, E. S. (1954) Continuous Inspection Schemes. Biometrika, 41, 100-115.

4. Page, E. S. (1955) Control charts with warning lines. Biometrika, 42 , 241-257.

5. Page, E. S. (1961) Cumulative sum schemes using gauging. Technometrics, 4, 97-109.

6. Page, E. S. (1962) The comparison of process inspection schemes. To appear.

7. Patnaik, P. B. (1950) The non-central χ^2 and F distributions and their application. Biometrika, 37, 202-232.

8. Stevens, W. L. (1948) Control by gauging. J. Roy. Statist. Soc. B 10, 54-108.

APPENDIX

The A. A. R. L. of the cusum scheme using ranges of N is $NL(0)$ where $L(Z)$ satisfies

$$L(Z) = 1 + \Psi(-Z)L(0) + \int_0^h L(y)\,\psi\,(y-Z)\,dy \quad , \qquad (A.1)$$

$\psi\,(y)$ is the frequency function of the increments $x_i - k$ and $\Psi(y)$ is the distribution function. Since x_i is a range of a sample of N from a normal population $\psi\,(y)$ involves the integration of products of normal ordinates and integrals. We would therefore wish to avoid its direct numerical evaluation if possible. An alternative approach is to use

$$L(0) = \frac{N(0)}{1-P(0)} \qquad (A.2)$$

where

$$N(Z) = 1 + \int_0^h N(y)\,\psi\,(y-Z)dy$$

$$P(Z) = \Psi\,(-Z) + \int_0^h P(y)\psi\,(y-Z)dy \quad , \qquad (A.4)$$

but similar objections arise. The simplest attack, and one which uses a previously prepared programme for the solution of the Fredholm equations, is to replace $\psi\,(y)$ by an approximation. Two approximations have been suggested in the literature: a simple one (χ^2) by Cox (1) and a more accurate one (χ) by Patnaik (7). On the first calculation performed these gave values 122 and 226 for the same $L(0)$ at the target value of σ. Such disagreement is almost entirely due to the smallness of $1 - P(0)$ at the target S. D. The approximations cannot provide sufficient correct significant figures in $P(0)$ to compensate for the loss in subtraction ($1 - P(0) = 0.009,\ 0.005$).

Another alternative is to proceed rather crudely and write (A.1) as

$$L(Z) = 1 + \Psi(-Z)L(0) + \sum_{i=1}^{n} L(y_i) \left\{ \Psi(y_i + \frac{h}{2n} - Z) - \Psi(y_i - \frac{h}{2n} - Z) \right\}$$

where $y_i = \frac{(2i+1)h}{2n}$; thus we are dividing the range of integration $(0,h)$ into n equal parts and replacing

$$\int_{\frac{ih}{n}}^{(i+1)h/n} L(y)\, \psi(y-Z)dy \quad \text{by} \quad L\left\{\frac{(2i+1)h}{2n}\right\} \int_{\frac{ih}{n}}^{(i+1)h/n} \psi(y-Z)\, dy \,. \qquad (A.5)$$

Since $L(y)$ is monotonic the approximation is likely to be fair. The values of $\psi(y)$ in (A.5) are obtained by interpolation within the computer in the values in the Biometrika table. The system of $(n+1)$ simultaneous linear equations derived from (A.5) by inserting $Z = 0$, $Z = \frac{2i+1}{2n} h$, $i = 1, \ldots n$ are then solved to give $L(0)$. The result of the first calculation using the four-decimal table was 199. This quantity was then computed using the correct analytical form for $\psi(y)$ with a six-decimal approximation to the error functions entering and the result obtained was 195. This agreement shown in Table 6 gives much more confidence than calculations using the approximations to $\psi(y)$.

TABLE 6: Comparison of methods of calculation of A.S.R.L.

Method	$\sigma/\sigma_o = 1$			$\sigma/\sigma_o = 1.5$		
	P(0)	N(0)	L(0)	P(0)	N(0)	L(0)
Quadrature	0.99427	1.1157	194	0.72253	1.5636	5.6
Interpolation in Table.	0.99440	1.1159	199	0.72463	1.5709	5.7
X-approximation	0.99506	1.1165	226	0.72150	1.5742	5.7
X^2-approximation	0.99091	1.1121	122	0.74097	1.5478	6.0

The "exact" approach by quadrature for $\psi(y)$ takes about 19 minutes computation on the Pegasus and the table method 1 minute for each $L(0)$. As σ increases the difference between the two approaches shrinks; e.g. even for $\sigma/\sigma_o = 1.05$, the two values for the same h, k as before give 103.4, 104.5, while there is close agreement for $\sigma/\sigma_o = 1.5$. Accordingly we use the "exact" method for finding $L(0)$ for $\sigma = \sigma_o$ and the table method for other values of σ.

The table is clearly of low accuracy; certainly errors of one unit in the second significant figure can be expected. However, a search for higher accuracy is hardly justified when the dependence of $L(0)$ on the distribution form is so marked.