

Biometrika Trust

Continuous Inspection Schemes

Author(s): E. S. Page

Source: *Biometrika*, Vol. 41, No. 1/2 (Jun., 1954), pp. 100-115

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: <http://www.jstor.org/stable/2333009>

Accessed: 29-01-2017 10:11 UTC

REFERENCES

Linked references are available on JSTOR for this article:

http://www.jstor.org/stable/2333009?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>



Biometrika Trust, Oxford University Press are collaborating with JSTOR to digitize, preserve and extend access to *Biometrika*

CONTINUOUS INSPECTION SCHEMES

By E. S. PAGE

Statistical Laboratory, University of Cambridge

1. INTRODUCTION

1.1. *Preliminary remarks*

Whenever observations are taken in order it can happen that the whole set of observations can be divided into subsets, each of which can be regarded as a random sample from a common distribution, each subset corresponding to a different parameter value of the distribution. The problems to be considered in this paper are concerned with the identification of the subsamples and the detection of the changes in the parameter value. Such problems can arise in a number of fields of application. For example, in an experiment in extra-sensory perception the proportion of correct answers given by a subject in response to a series of questions may change during the course of the experiment, and it may be desired to estimate the position of the change or to stop the experiment when a change is noticed. Again, in a psychological experiment in which a subject is required to guess the colour of the next ball to be drawn at random with replacement from a bag containing balls of two colours the subject's proportion of correct guesses in a series of trials with the same bag of balls may change as he gains some knowledge of the constitution of the bag from the results of earlier guesses; it may be of interest to estimate the point at which the change took place. More widely known are the occurrences in industry of problems of detecting changes in the quality of the output from a continuous production process. Some such processes maintain an approximately constant quality of output for considerable periods; occasionally, probably because of a fault at some point of the process, the quality worsens and a large proportion of the output becomes unacceptable. The quality of the output may be assessed by some measurable characteristic (e.g. when the length of articles is normally distributed with constant variance the mean length may be used as an indication of quality), or by the fraction of the output that fails to meet given specifications. In general, it will be possible to assign a quality number, θ , to the output which may be taken as a parameter of the distribution. We are interested in the changes in θ .

One of the simplest criteria for detecting a change in the mean, θ , of the distribution is a weighted sum of the last few, k say, observations, i.e. a moving average. If k is small, large changes in θ will be detected rapidly but small changes only slowly; on the other hand, a larger value of k will be required for the best detection of small changes in θ but then large changes will be noticed later owing to the moving average damping the effect of a single extreme observation. In general, the consequences of rules based on moving averages are difficult to evaluate. The theory of one such rule for identifying the subsamples in observations from a binomial population of changing mean has been given by Anscombe, Godwin & Plackett (1947).

It will be convenient in what follows to use the terminology of the industrial applications in view of its greater familiarity.

1.2. *Detection of a change in the parameter*

The first problem to be considered here is that of detecting a change in the parameter θ . It is this problem that process inspection schemes are designed to solve; it is required to detect a deterioration in the quality of the output from a continuous production process. When such a deterioration is suspected some action is taken; for example, the production may be suspended and a machine reset. A widely used scheme consists of examining samples of a fixed size at regular intervals of time; a statistic of the sample (e.g. mean, range, or number of defectives) is plotted on a control chart and corrective action is taken if the point falls outside control limits drawn on the chart (e.g. Dudding & Jennett, 1942; Duncan, 1952; Shewhart, 1931). In most cases of practical interest there is a probability of one that some point will eventually fall outside the limits and action will then be taken even when there is no change in the quality number, θ . If the fraction of the output that is sampled remains constant the amount of output produced before action is taken is proportional to the number of articles inspected. We are led, therefore, to consider what we shall call the *average run length* function for a process inspection scheme.

DEFINITION. When the quality remains constant the *average run length* (A.R.L.) of a process inspection scheme is the expected number of articles sampled before action is taken.

The A.R.L. is clearly a function of θ , the quality number.

When the quality of the output is satisfactory the A.R.L. is a measure of the expense incurred by the scheme when it gives false alarms, i.e. Type I errors (Neyman & Pearson, 1936). On the other hand, for constant poor quality the A.R.L. measures the delay and thus the amount of scrap produced before the rectifying action is taken, i.e. Type II errors. This measure is one of several suggested by Aroian & Levene (1950); another is the probability that action is taken within n observations.

For the control chart scheme described, on constant quality, the number, r , of samples each of N articles examined before action is taken is a geometric variable such that

$$\Pr(r=k) = P^{k-1}(1-P), \quad \text{where } P = P(\theta),$$

the probability that a given sample point falls between the control limits. $\mathcal{E}(r) = (1-P)^{-1}$ and the A.R.L. function is

$$L(\theta) = N/(1-P). \quad (1)$$

Although the results of previous samples are recorded on the chart none is used by the above process inspection rule; in order to avoid some of this loss, warning lines within the control limits are often drawn on the chart and the rule amended to read: 'Take action if any point falls outside the control limits or if l out of a sequence of m points fall outside the warning lines.' The A.R.L. for rules of this type can be calculated by enumerating the possible combinations of the positions of the last $m-1$ points and treating them as the states of a discrete Markov process; the methods described, for example, in Bartlett's paper (1953), will then give the expected number of samples before a combination demanding action is observed.† Such schemes will not be considered further here.

† When a chart is used for controlling the mean of a normal population, statements are sometimes made in the literature such as: '4 out of 16 points falling outside warning limits at $\mu \pm 1.96\sigma/\sqrt{N}$ are equivalent to one point outside control limits at $\mu \pm 3.09\sigma/\sqrt{N}$.' If the process mean is μ , these two events have the same probability, but rules using one or both of these criteria for action will have different A.R.L. functions and therefore a different effect in practice.

The process inspection rules described so far are based either on a single point recorded on the control chart or on a fixed number of the most recently recorded points. Such rules will therefore fail to make use of all the information that is available on the chart. Again, with these rules the best sample size and the position of the control limits vary with the magnitude of the change in the quality number that it is important to detect (Page, 1954). Consequently a rule that is optimum for a certain magnitude of change will not be optimum for other magnitudes of change, and in some cases the loss in efficiency may be serious. For example, if the best sample size for a single sample scheme to detect a change from θ_0 to θ_1 is N , the A.R.L. of this scheme is not less than N for any value of θ ; it may be possible by some other method to detect a large change in fewer than N observations. In the following sections we develop rules that use all the observations since action was last taken and that are suitable for the detection of any magnitude of change in the parameter.

2. ONE-SIDED PROCESS INSPECTION SCHEMES

2.1. A transition scheme

Process inspection schemes that are designed to detect deviations in θ in only one direction will be called *one-sided* schemes. We consider first a simple example of a rule to control the fraction of defective articles produced by an industrial process.

Suppose samples of twenty articles are taken every hour; let d_n be the number of defectives in the n th sample after action was last taken. Consider the rule:

$$\left. \begin{array}{l} \text{Take action if } d_n \geq 3, \\ \text{or } d_n + d_{n-1} \geq 4, \\ \text{or } d_n + d_{n-1} + d_{n-2} \geq 5, \\ \text{etc.} \\ \text{i.e. } \sum_{i=0}^r d_{n-i} \geq 3+r \text{ for at least one } r \quad (0 \leq r \leq n-1). \end{array} \right\} \quad (2)$$

After each sample is taken, the total number of defectives in the last one, two, three, etc., samples is examined and action is taken if, over any sequence of samples, the average number of defectives per sample is 'much' more than one. If 5% defective is the critical quality level for the process, so that any worse quality requires action to be taken, the above rule would be suitable.

With this rule the decision whether or not to take action is made after each sample and all the previous samples are used in making the decision. This rule may therefore be regarded as transitional between those rules for which the decisions are made after each sample on the results of a fixed number of samples and those for which a decision is made after each observation considering all previous observations (the latter type of rule will be called a *sequential process inspection scheme*).

The action criteria (2) may be specified in a different way. Let a score +19 be assigned to each defective occurring in a sample, and -1 for each non-defective. Let $S_n = \sum_{i=1}^n x_i$, where x_n is the total score in the n th sample after action was last taken. Then (2) is equivalent to:

$$\text{Take action if } S_n - S_{n-r} \geq 40 \text{ for any } r \quad (1 \leq r \leq n), \quad (3)$$

or, what is again equivalent,

$$\text{Take action if } S_n - \min_{0 \leq i < n} S_i \geq 40. \quad (4)$$

When the cumulative score S_n is plotted on a chart the mean path when the fraction defective, θ , is constant and less than 5% is below the horizontal. If $\theta > 0.05$, the mean path is above the horizontal so that the action criterion (4) can be expected to be satisfied speedily. Accordingly this rule would be appropriate for controlling a process for which more than 5% defectives could not be tolerated.

In the above scheme when the quality is satisfactory the mean path on the cumulative sum diagram is downwards, and a deterioration in quality is detected by an upward change of direction in the mean path. A similar criterion for controlling the parameter of other distributions when deviations in only one direction are important is provided by the following rule.

Rule 1. Take samples of fixed size N at regular intervals; assign a score x_k to the k th sample and plot the cumulative score $S_n = \sum_{k=1}^n x_k$ on a chart:

$$\text{Take action if } S_n - \min_{0 \leq i < n} S_i \geq h. \quad (5)$$

The A.R.L. for this rule will be obtained as a special case of that for the rule of the next section.

2.2. A sequential scheme

In what follows we shall suppose that observations are recorded singly at regular intervals, and that a decision whether or not to take action because of a suspected deterioration in the output is made after each observation. A rule fulfilling these requirements is obtained by modifying Rule 1.

Rule 2. Take observations at regular intervals; assign a score x_k to the k th observation and plot the cumulative score $S_n = \sum_{k=1}^n x_k$ on a chart. Take action if (5) is satisfied.

The system of scoring is chosen so that the mean sample path on the chart when quality is satisfactory is downwards, i.e. of negative gradient, and is upwards when quality is unsatisfactory (Fig. 1).

In order to evaluate the A.R.L., define

$$\left. \begin{aligned} S'_n &= \max(S'_{n-1} + x_n, 0) \quad (n \geq 1) \\ S'_0 &= 0, \end{aligned} \right\} \quad (6)$$

so that $S'_n = 0$ whenever $S_n < \min_{0 \leq i < n} S_i$. The condition (5) is then equivalent to:

$$\text{Take action after the } n\text{th observation if } S'_n \geq h. \quad (7)$$

It can now be seen that this rule breaks up into a sequence of Wald sequential tests with boundaries at $(0, h)$ and initial score zero.[†] The test is reapplied when the previous test ends on the lower boundary and action is taken when a test ends on the upper boundary.

Notation. Let the probability that a Wald test with boundaries $(0, h)$ and initial score Z end on the lower boundary be $P(Z)$ and the average sampling numbers, unconditional, conditional upon the test ending on the lower boundary, and conditional upon the test ending on the upper boundary, be $N(Z)$, $N_1(Z)$ and $N_2(Z)$ respectively. A test that ends upon the lower boundary will be called an 'acceptance test'; conversely for a 'rejection test'.

[†] The procedure in which observations x_1, x_2, \dots are taken so long as $a < Z_i < b$, where $Z_i = Z_{i-1} + x_i$, $Z_0 = x_0$ will be termed a Wald sequential test with boundaries (a, b) and initial score x_0 .

The probability that there are r acceptance tests before a rejection test occurs is $\{P(0)\}^r \{1 - P(0)\}$, and the expected number of such tests is therefore

$$\sum_{r=1}^{\infty} r \{P(0)\}^r \{1 - P(0)\} = P(0)/\{1 - P(0)\}. \quad (8)$$

It follows that the A.R.L. for this rule is

$$\begin{aligned} L &= \frac{P(0)}{1 - P(0)} N_1(0) + N_2(0) \\ &= \frac{N(0)}{1 - P(0)}. \end{aligned} \quad (9)$$

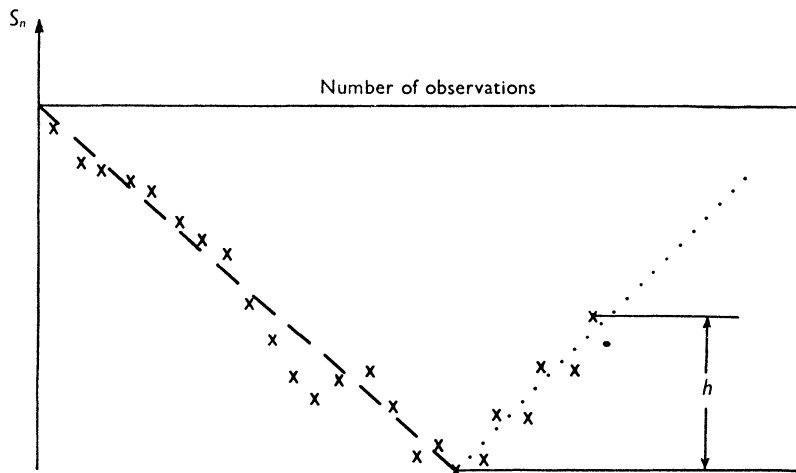


Fig. 1. Control diagram. $\times \times$ sample points. — mean path on satisfactory quality. ... mean path on unsatisfactory quality.

An approximation to the distribution of the run length, l , may be found when $P(0) \doteq 1$, i.e. when the number of acceptance tests before action is taken is large. If l_1 is the total number of observations in the acceptance tests its characteristic function (c.f.) is

$$\mathcal{E}(e^{itl_1}) = \frac{1 - P(0)}{1 - P(0)\phi_1(t)}, \quad (10)$$

where $\phi_1(t)$ is the c.f. of the number of observations in a single acceptance test. By considering the dominant terms in the repeated differentiation of (10) it follows that, for $P(0)$ near 1, the moments of $l_1/N_1(0)$, and hence those of $l/N_1(0)$, are approximately those of a geometric distribution with mean $P(0)/\{1 - P(0)\}$. Hence

$$\Pr(l \leq n) \doteq 1 - \{P(0)\}^{v+1}, \quad (11)$$

where

$$v = n/N_1(0).$$

If two independent rules of this type are operated simultaneously (as, for example, when it is desired to control both mean and variance of a normal population for one-sided deviations) the run length of the combination is distributed as the minimum of two independent run lengths, l and l^* ; using the approximation (11) the A.R.L. of the combined rules will be λ , where

$$\lambda^{-1} = L^{-1} + L^{*-1}. \quad (12)$$

This relation can be easily seen by considering the stops due to the two rules as occurring at random at rates L and L^* .

In order to calculate the A.R.L. for the transition scheme, the score assigned to each sample may be regarded as a single observation; the average number of samples is therefore given in terms of the characteristics of the appropriate Wald test by equation (9).

2.3. An integral equation for the A.R.L.

In general the operating characteristic (o.c.) and average sample number functions of a Wald test are the solutions of integral equations of Fredholm's type. For example, the integral equation for the average sample number $N(Z)$ of a Wald test with boundaries (a, b) and initial score Z is

$$N(Z) = 1 + \int_a^b N(y) f(y - Z) dy.$$

Using (9) it is therefore possible to calculate the A.R.L. of a general scheme as the ratio of the solutions of two integral equations. We can, however, derive a single integral equation satisfied by the A.R.L., and to do this we introduce an extension to Rule 2.

Rule 3. Take observations and assign scores as for Rule 2. Take action if either

$$(a) \quad S_n \geq h \quad \text{and} \quad S_i > 0 \quad \text{for } i = 1, 2, \dots, n-1,$$

or

$$(b) \quad S_n - \min_{0 \leq i < n} S_i \geq h,$$

where

$$S_0 = Z \quad (0 \leq Z < h) \quad \text{and} \quad S_n = S_{n-1} + x_n. \quad (13)$$

The rule modifies Rule 2 only near its start. A formulation similar to (6) shows that a Wald test with boundaries $(0, h)$ and initial score Z is first applied; if it ends in acceptance, Rule 1 then operates.

Let the distribution function of a single score, x , be $F(x)$, and let $L(Z)$ be the A.R.L. of Rule 3. The A.R.L. of Rule 2 is clearly $L(0)$. By considering the consequences of the first observation and taking expectations we obtain the equation, holding for $0 \leq Z < h$,

$$L(Z) = 1 + L(0) F(-Z) + \int_0^h L(x) dF(x - Z). \quad (14)$$

On substituting for $L(0)$, this equation reduces to the Fredholm form, although the convolution kernel is lost.

2.4. General remarks

In the last two sections we have been studying the expected values of a random variable, $l(S, m)$, the number of observations required to satisfy (5) when S is the cumulative score and m the previous minimum. It is plain that $l(S, m)$ is a Markov process in two dimensions where, in general, the states to which the process may move at the next observation are (S_1, m) and (m_1, m_1) , where $S_1 \geq m$ and where $m_1 < m$. When part of the information is suppressed, e.g. the value of m , the resulting one-dimensional process is no longer Markovian, but this property is regained if the scoring system is that of (6). The states are now defined by the S' score. It follows that every state S' is a regeneration point (Kendall, 1951; Lindley, 1951) of the process. It may also be regarded as a random walk between an absorbing and a 'holding' barrier; when $S'_{n-1} + x_n < 0$, the particle is held at zero until the next observation is taken. The A.R.L. is the mean absorption time, i.e. the mean first passage time to the states $S' \geq h$.

3. TWO-SIDED SCHEMES

3.1. *Simultaneous application of two one-sided schemes*

The schemes of § 2 are adequate for the detection of changes in the quality number in one direction only. It is often necessary, e.g. in controlling the mean of a normal population, to detect changes in either direction. A procedure immediately suggesting itself from the previous work is to use two one-sided schemes, one to detect a decrease, and the other an increase, in the quality number. The rule is then:

Rule 4. Take action after the n th article sampled if either

$$S_n - \min_{0 \leq i < n} S_i \geq h \quad \text{or} \quad \max_{0 \leq i < n} S_i - S_n \geq k. \quad (15)$$

When this rule is specified by the two co-ordinates $r_n = S_n - \min S_i$, $s_n = \max S_i - S_n$ the run length $l(r, s)$ is a two-dimensional Markov process and the A.R.L. is the mean first passage time to the states (r, s) when $r \geq h$ or $s \geq k$. An integral equation for the A.R.L. may be derived by introducing a scheme similar to Rule 3, but the equation is more complicated than (14). Alternatively, simultaneous integral equations for $L(r, 0)$ and $L(0, s)$ may be written down in terms of the characteristics of certain Wald tests and the distributions of the overlap of the boundaries at termination. In view of these complications we consider an alternative scheme.

3.2. *A two-sided Wald test*

Barnard (1947) has suggested a sequential test for discriminating between more than two hypotheses, and special cases have been considered by Armitage (1947, 1950) and Sobel & Wald (1949). In this section we specify the particular case needed for the process inspection scheme of § 3.3.

Consider the simultaneous application of two Wald tests, one with initial score on the acceptance boundary and the other with initial score on the rejection boundary. This procedure may be represented graphically as in Fig. 2. The test continues until both the simple Wald tests are ended.

Let the probability that the test ends in the region H_i be $P(H_i)$. Then Sobel & Wald have shown that $P(H_1) = \text{o.c.}$ for the first Wald test $= P'$, say. Similarly $P(H_3) = 1 - P''$, so that $P(H_2) = P'' - P'$. In the general case the average sample number (A.S.N.) of the two-sided test is difficult to obtain, but when two boundaries pass through the origin it may be expressed simply in terms of the A.S.N.'s of the composing Wald tests.

Let n , n' and n'' be the number of observations required for a decision in the two-sided test and the simple Wald tests respectively in a realization and let their expectations be denoted by the capital letter with the corresponding dashes.

$$\begin{aligned} \text{Then} \quad n &= \max(n', n'') \\ &= \max(n', n'') + \min(n', n'') - 1, \end{aligned}$$

since the first observation causes at least one of the simple tests to terminate.

$$\text{Hence} \quad n = n' + n'' - 1,$$

$$\text{and therefore} \quad N = N' + N'' - 1. \quad (16)$$

This equation may also be derived from the integral equations by considering the result of the first observation. The simplification leading to (16) arises because the composing tests are independent after the first observation.

3.3. A scheme based on the two-sided test

A two-sided process inspection scheme can now be simply specified in terms of the test T , say, described in § 3.2.

Rule 5. Apply T : if it ends in region H_2 , reapply T ; otherwise take action.

As for (9) we obtain the A.R.L. $L = N/[1 - P'' + P']$. (17)

This scheme is a modification of that based on Rule 4 in which suspected fluctuations in quality are investigated in one direction at a time. In the special case where all the boundaries are horizontal the rule can be stated in terms of maximum and minimum cumulative scores as for Rule 1, making application especially easy; even in the more general case acceptance and rejection scores can be recorded beforehand (cf. Wald, 1947, p. 93), so that only little calculation is necessary when carrying out the rule. The probability that the scheme ends in the H_3 region is

$$(1 - P'')/(1 - P'' + P'), \quad (18)$$

so that curves showing the probability that the scheme will indicate the correct direction of the change may be computed from the o.c.'s of the composing Wald tests.

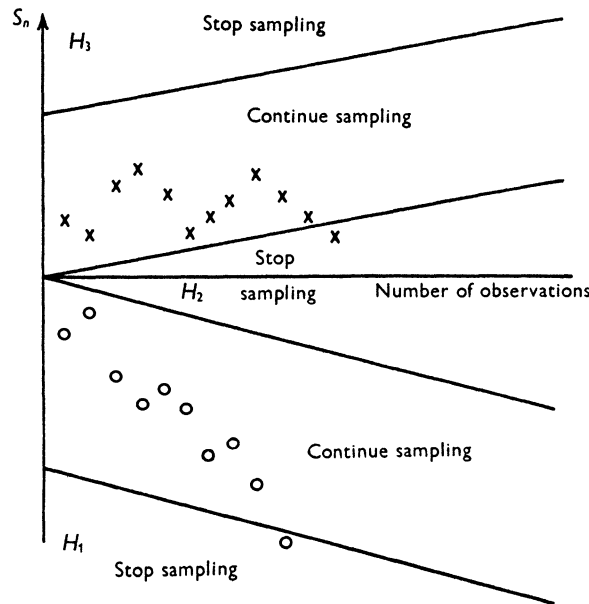


Fig. 2. A two-sided sequential test. $\circ \circ$ sample path ending in H_1 . $\times \times$ sample path ending in H_2 .

4. MISCELLANEOUS REMARKS ON PROCESS INSPECTION

4.1. Modifications of the rules

When one of the above rules has indicated a change in the quality of the output it may be desired to confirm the suspicions by an increased rate of testing. A possible extension to Rule 2 is as follows:

Inspect a fraction f of the output and apply Rule 2 until condition (5) is satisfied; then inspect a fraction f^* until the cumulative score S_n rises a further h_1 or falls h_2 . In the former case, take rectifying action, and in the latter, resume inspecting a fraction f and reapply Rule 2.

Another requirement may be to inspect the run of output that leads to the action being taken; this would entail holding production in 'bond' until it became clear that it was satisfactory. For example, with Rule 2, the output would be in bond until the current S_n was the minimum; the average amount held would be $N_1(0)/f$ while production continued and $N_2(0)/f$ when the action was taken.

In practice it might be inconvenient to examine single articles at frequent intervals so that sequential inspection schemes (in the sense of § 2.1) will not be suitable. In such a case a scheme of the transition type may be applied and the size of sample chosen as small as conveniently possible. An alternative procedure is to inspect a fixed number of articles on each occasion but to apply a sequential rule as each article is examined; a decision to take action will be delayed by an average time equivalent to about half the sample size. Another method of grouping is to inspect articles on each occasion until a specified number of the composing Wald tests have terminated or until action is required, whichever is the sooner. The fraction of output sampled would then be a function of the quality number and, for a given θ , could be adjusted by the choice of scheme.

4.2. Repeated tests

It has been shown how Rule 2 is equivalent to the repeated application of Wald tests with initial score on the acceptance boundary, and the A.R.L. has been evaluated using that fact. Process inspection schemes can be obtained by repeatedly applying general Wald tests and their A.R.L.'s found in the same manner. When the quality is uniform the behaviour of such schemes is comparable with that of Rules 2 and 3. There is a difference, however, when a deterioration in quality occurs; it is not clear what is the effect on the A.R.L. should an abrupt change in θ take place in the middle of one of the tests. When the tests have initial score on the acceptance boundary we have that $L(Z) \leq L(0)$ for Rules 2 and 3 so that the average number of observations until the action is taken is certainly no greater than the A.R.L. of Rule 2 on the new quality. No such statement can be made for more general tests.

Other process inspection schemes may be obtained by the repeated application of different tests or sequences of tests; for example, the control chart method is that of repeated fixed sample size tests.

4.3. Estimation problems

If it is required to estimate θ when a process inspection scheme has indicated a change, in practice it may be most convenient to take a further sample for this purpose. Alternatively, an estimator of sorts can be obtained by equating the number of observations in the rejection test to the appropriate conditional A.S.N. function, $N'(\theta)$, say. If there is no root of this equation the estimate can be θ^* , where $N'(\theta^*)$ is the maximum of $N'(\theta)$; if there are several roots, which will be the case in general, some rule such as 'take the largest root' can be used.

Since a process inspection scheme eventually suggests a change in θ whether one has occurred or not a rule for estimating the position of the change may be misleading. If it is known that a change has occurred somewhere within a set of n observations Lindley's Method of Minimum Unlikelihood (1952) will provide an estimator. If it is further known that the change is from θ_0 to θ_1 a simple unbiased estimator, not necessarily integral or within the range $(0, n)$, may be obtained from the meet of the mean paths from zero and

$$\sum_1^n x_i, \text{ i.e. we take } n = \left\{ n\theta_1 - \sum_1^n x_i \right\} / \{\theta_1 - \theta_0\}. \quad (19)$$

In this case the maximum-likelihood estimator is the minimum (maximum if $\theta_0 > \theta_1$) of the sample path for Rule 1 when the scoring is that of the sequential-likelihood ratio test,

$$z_i = \log \frac{f(x, \theta_1)}{f(x, \theta_0)}. \quad (20)$$

5. PROCESS INSPECTION FOR FRACTIONS DEFECTIVE

5.1. Average run length

In the case where the articles produced are classified as defective or non-defective explicit formulae for the o.c. and A.S.N. of the Wald test have been given by Burman (1946) and Walker (1950). Let the score for a non-defective article be $-a$ and that for a defective b , where a and b are integers.† The process inspection scheme is given by Rule 2.

The difference equations for the A.R.L. to which the integral equation (14) reduces in this case may be solved by a series transform method as in Walker's paper. We shall, however, use relation (9) and Burman's formulae in which $a = 1$ and it will be convenient to take the formulae in the limit as the fraction defective, $p \rightarrow 0$, $b \rightarrow \infty$ and $pb = X$. It is necessary to modify these formulae, which do not hold on the boundary, to obtain $P(0)$, $N(0)$. We have, by taking expectations conditional upon the first observation,

$$P(0) = q + pP(b) \quad (21)$$

$$\text{and} \quad N(0) = 1 + pN(b). \quad (22)$$

In the limit the A.R.L. is given by

$$\frac{L(0)}{b} = \frac{1 + XN(b)/b}{X[1 - P(b)]}. \quad (23)$$

5.2. Choice of scheme

If the sampling fraction is constant, the A.R.L. is proportional to the average number of articles produced before the inspection scheme causes action to be taken. In some applications it will be more important that stoppages when the quality is good be very infrequent while a fairly lengthy run of poor quality can be tolerated; in other cases a rapid action for bad quality is of prime importance. Sequential schemes to meet these requirements can be selected; e.g. for the first situation we choose the sequential scheme with specified A.R.L. on θ_0 (good quality) and minimum A.R.L. on θ_1 (bad quality). From (23), $pL(0)$ is a function of $X = pb$; consequently, for each h we can find X , and hence b , such that the A.R.L. is L_0 when the fraction defective is p_0 . The A.R.L. function for a few values of h may be calculated using the tables of Wald schemes given by Anscombe (1949); for the selection of schemes to the above requirements more extensive tables are needed; these are shown in the Appendix.‡ $pL(0)$ is tabulated as a function of X for $h/b = 1.25$ (0.25) 5.00. For $h/b \leq 1$, $pL(0) \equiv 1$. As an example, the scheme with $h/b = 2.75$ and A.R.L. 4000 at $p_0 = 0.02$ has $p_0L(0) = 80$ at $X \doteq 0.35$. Hence the required scoring is $b = 17.5$. The A.R.L. at $p = 0.05$ is given by $p_1L(0) = 11.6$, i.e. $L(0) = 232$ from the value $X_1 = 0.875$. In order to select schemes

† With this scoring a rise in the mean sample path corresponds to a rise in the fraction defective. Those who prefer to think in terms of a fall in the standard of production will wish to reverse the signs of the scores.

‡ I am indebted to the Director of the Mathematical Laboratory, University of Cambridge, for permission to use EDSAC for the preparation of these tables.

with given A.R.L. at p_0 and maximum or minimum A.R.L. at p_1 the schemes for different h/b can be compared and the appropriate value chosen.

It has been remarked (see § 2.2) that the process inspection scheme may be carried out as tests (Wald, 1947); it will, however, probably be found more convenient to record the cumulative score S_n on a chart and take action whenever $S_n - \min S_i \geq h$. As only the difference between S_n and the minimum is of importance only this information need be entered on a new chart when the old one is completed. If the old chart is withdrawn when the current cumulative score is the minimum the new chart may be started afresh.

5.3. Comparison of sequential and simple sampling schemes

The single sampling scheme in which a fixed number, N , of articles is inspected on each occasion and action is taken if c or more defectives are observed has A.R.L. L^* , where

$$L^* = N \sum_{i=c}^N \binom{N}{i} p^i q^{N-i}. \quad (24)$$

For p small and N large the Poisson approximation may be used; Molina's tables (1947) will be found useful for selecting the best scheme of this type (Page, 1954). In Table 1 the A.R.L. function is given for two values of the argument, p , for five single sampling schemes and in Table 2 the corresponding quantities for five sequential schemes taken from the Appendix table; the first three schemes have specified A.R.L. at $p = 0.01$ and minimum A.R.L. at $p = 0.03$ and the last two specified A.R.L. at $p = 0.03$ and maximum A.R.L. at $p = 0.01$.

Table 1. *Single sampling schemes*

N	c	$L^*(0.01)$	$L^*(0.03)$
63	3	2,500	214
103	4	5,000	275
150	5	8,000	321
70	3	2,050	200
165	5	6,300	300

Table 2. *Sequential schemes*

h/b	b	$L(0.01)$	$L(0.03)$
2.75	60	2,500	180
3.75	65	5,000	220
4.25	70	8,000	263
3.00	55	3,600	200
3.75	50	13,000	300

For the first three schemes $L(p=0.03) \simeq 0.8L^*(p=0.03)$, while the difference in the A.R.L.'s at $p = 0.01$ for the last two is proportionately rather more.

It will, however, be noticed that if a deterioration in quality much worse than $p = 0.03$ occurs the A.R.L. of the single sampling schemes is at least equal to the sample size while the lower limit of the A.R.L. of the sequential schemes is the largest integer less than $R + 1$. Accordingly the sequential schemes have the advantages of single sampling schemes using small samples for the detection of large changes in quality while retaining a desirable A.R.L. for small changes. These two requirements are in conflict for the choice of the best sample size for single sampling schemes; the best sample size becomes large as the magnitude of the change decreases.

The binomial process inspection scheme may be applied to continuous experiments designed to increase the rate of occurrence of a desirable but rare property. An analogous problem has been considered by Fisher (1952).

6. APPLICATIONS TO STORAGE PROBLEMS

6.1. *General remarks*

Storage and queuing problems have received attention in a number of recent papers (e.g. Kendall, 1951; Lindley, 1952; Smith, 1953). Many of the problems considered in these papers have placed no restriction on the capacity or content of the store except that the system should be in the equilibrium state where the distribution of input and output are such that the content of the store does not increase without limit. In this section some problems concerned with stores of finite capacity will be discussed briefly.

6.2. *A storage problem*

Consider a store of total capacity h ; at certain times ('movement epochs') goods of amount x are sent out and, at the same time, an amount y is received into the store. If the previous contents of the store are insufficient to meet the demand x , the deficiency is made up as far as possible from the new input. If the store would be left overfilled after a movement epoch some of the input is rejected so that the store is left just filled. Let S_n be the store content after the n th movement epoch.

$$\text{Then} \quad \left. \begin{aligned} S_n &= \min(h, S_{n-1} + y - x), \\ S_0 &= Z, \end{aligned} \right\} \quad (25)$$

where Z is the initial content of the store.

We are interested in the occasions when the store is empty and the demand cannot be satisfied; that is to say, when S_n becomes negative. The situation is that of the process inspection Rule 3 to detect a decrease in the quality number when the initial score is Z . The A.R.L. is the average number of movement epochs before the store becomes empty.

We have

$$L(Z) = N(Z) + \{1 - P(Z)\} L(h). \quad (26)$$

The probability that the store becomes empty before becoming full is $P(Z)$, and the expected number of movement epochs between successive epochs of shortage (i.e. when the store is empty) is $N_1(0)$. The number of consecutive epochs of shortage is a geometric variable with parameter $\varpi = \int_{-\infty}^0 f(z) dz$, where $f(z)$ is the frequency function of $z = y - x$.

If any unsatisfied demand is allowed to be supplied out of succeeding inputs, so that, for $S_n < 0$, $|S_n|$ is the total demand outstanding, the average number of movement epochs until the store becomes full is the A.S.N. of the Wald test with a single boundary at h and initial score Z .

6.3. *Stores with a steady drain*

Many problems of practical interest involve stores from which there is a steady output, e.g. storage of fuel for blast furnaces (this has been the subject of some at present unpublished work by H. Herne and D. G. Nikolls). If the store is allowed to accept input whenever the content of the store is $< h$ (even if the input causes the new total content to exceed h) but must reject otherwise, the expected time to an overflow can be written down as before. To find the expected time to the exhaustion of the store we need, in general, to create a regeneration point after the store has overflowed; for example, we can insert the condition that the next input after an overflow shall arrive when the content is Z_0 . The previous methods may then be applied.

When the upper limit, h , of the store may never be exceeded the problem is more complicated. The net increment between inputs is no longer sufficient to describe the behaviour of the store content. In the Wald test it is necessary to take two separate steps, first the amount of input and then the drain. The simple theory can be applied in some special cases. If the amount of input on each occasion is fixed (and equal to unity, say) and the drain between input epochs is a random variable, y , a Wald test with boundaries $(0, h-1)$ and increments $1-y$ may be used to construct the theory as before. Similarly, a Wald test with boundaries $(1, h)$ and increments $x-1$ may be used when there are inputs of amount x at unit time intervals. When the random variable has the exponential distribution an explicit solution for the characteristics of the Wald test is available (Anscombe & Page, 1954).

7. DEFERRED SENTENCING SCHEMES

7.1. General remarks

Process inspection schemes are designed to detect variations in the quality of output as it is being produced so that faults in the process may be traced promptly. In some cases it is impracticable to apply such a scheme, but after production it is necessary that runs of good and bad quality be separated, i.e. the output must be sentenced. Different action will then be taken with the two qualities. In the original formulation the problem is that of identifying the subsamples with the different parameter values. Some deferred sentencing schemes for articles classed as defective or not have been given by Anscombe *et al.* (1947); in these schemes articles are sampled and tested one by one and rules for the rejection of output are based on the occurrence of d defectives in any sequence of N articles examined. Roughly speaking, these moving average rules accept the output until a deterioration in quality is noticed and then reject until an improvement appears. Similar schemes can be constructed from process inspection schemes.

7.2. Repeated process inspection schemes

When a deterioration in quality is shown by a deviation in one direction (an increase, say) of the quality number, a deferred sentencing scheme may be obtained by applying two process inspection schemes consecutively. First, apply Rule 2 to detect an increase in θ ; let l be the position of the minimum S_i when the rule operates. Then apply Rule 2 to detect a decrease in θ ; let m be the position of the maximum S_i . Reject articles $l+1$ and m and all output within those limits. Because there is probability one that the process inspection scheme operates even if the quality, θ , remains constant some of the output will be rejected by this deferred sentencing scheme whatever the value of θ , satisfactory or not. In the notation of § 2.2 with $P^*(Z)$, etc., written for the characteristics of the Wald test with boundaries $(-k, 0)$, an estimate of the proportion of output rejected is

$$\frac{N_2(0) + \frac{\{1 - P^*(0)\}}{P^*(0)} N_2^*(0)}{L(0) + L^*(0)}. \quad (27)$$

Two-sided deferred sentencing schemes may be obtained in a similar manner. For example, a scheme could be:

Apply Rule 5 until it operates; suppose that it does so because of a hit on the $H_1(H_3)$ boundary (thus suggesting a decrease (increase) in θ). Then apply Rule 1 to detect an increase (decrease) in θ ; when it operates repeat the cycle.

The schemes suggested above may be regarded as a succession of Wald tests with the rule that all output leading to a rejection test is rejected. Other schemes may be constructed using different sequences of tests.

8. RECTIFYING SCHEMES

8.1. *General remarks*

The main purpose of rectifying inspection is to check that the output maintains a given standard and to improve the quality by replacing defective articles if necessary. Often it is required to ensure that the average quality of the product after inspection (the average outgoing quality: A.O.Q.) is better than some specified standard whatever be the quality before inspection. We suppose that articles are either defective or non-defective. A fraction (possibly large) of the output is inspected and any defective articles are rectified or replaced by non-defectives. Several schemes for the rectifying inspection of continuous output have been suggested, notably those of Dodge (1943) and Wald & Wolfowitz (1945). In these schemes a fraction, $f < 1$, of the output is inspected until there are signs that the quality is not as good as that required; 100 % inspection is then started and continued until an improvement is noticed, when the cycle is repeated. We consider schemes in which the signs of improvement or deterioration are based on the results of tests repeatedly applied.

8.2. *Repeated tests*

Suppose that we are given two statistical tests, T and T^* , for the quality of the production and that each has two possible results, acceptance and rejection. Then we consider the rule:

‘Inspect a fraction f of the output and apply test T : if it accepts, reapply T ; if it rejects, start 100 % inspection and apply T^* . If T^* rejects, continue 100 % inspection and reapply T^* ; if it accepts, inspect a fraction f and restart the cycle of tests.’

When the production process is under control so that the proportion of defectives, p , produced remains constant we can express the A.O.Q. in terms of the A.R.L.’s L and L^* , of the process inspection schemes obtained by repeatedly applying T and T^* respectively. We have

$$\text{A.O.Q.} = \frac{pL(1-f)}{L+fL^*}. \quad (28)$$

The fraction of production inspected, $I(p)$, is given by

$$I(p) = \frac{f(L+L^*)}{L+fL^*}. \quad (29)$$

More complex rules can be formulated in which more than two tests are used or in which succeeding tests depend upon characteristics of preceding ones. Tests allowing more than two decisions may be used to cater for multi-dimensional inspection.

8.3. *Special cases*

The scheme suggested by Dodge is:

‘Inspect a fraction f of the output until a defective occurs; then do 100 % inspection until a run of i non-defectives is observed.’ In this case the test T may be taken as that based on a fixed sample size of one article with acceptance for a non-defective and rejection for a defective. T^* is a binomial sequential test with boundaries $(0, i)$, starting score i and penalty for a defective $\geq i$.

This scheme bases its decision to start 100 % inspection on the appearance of one defective only; if this treatment of a first offender seems harsh it may be preferred to use for T a sequential test with initial score zero and boundaries $(0, h)$, where h is greater than the penalty for a defective.

Schemes may be obtained using Wald tests with general starting scores, but disadvantages similar to those given for process inspection schemes are present when the quality is not constant; these same disadvantages will apply in greater or less degree to the use of all tests with an A.S.N. much greater than one.

For the sampling plan A of Wald and Wolfowitz T is a sequential test with only a rejection boundary and T^* a fixed size test with only an acceptance criterion. The position of the boundary in the next test T and the number of observations in T^* depend on the overlap on the boundary in the preceding test.

I wish to express my thanks to Mr F. J. Anscombe, who suggested this subject of research and supervised it, and to Dr D. R. Cox, Mr D. V. Lindley and Dr J. Wishart for reading the draft of the paper. I also wish to thank the Department of Scientific and Industrial Research for the award of a maintenance grant.

REFERENCES

- ANSCOMBE, F. J. (1949). *J. R. Statist. Soc. A*, **62**, 180.
 ANSCOMBE, F. J., GODWIN, H. J. & PLACKETT, R. L. (1947). *J. R. Statist. Soc. Suppl.* **9**, 198.
 ANSCOMBE, F. J. & PAGE, E. S. (1954). *Biometrika*, **41**, 252.
 ARMITAGE, P. (1947). *J. R. Statist. Soc. Suppl.* **9**, 250.
 ARMITAGE, P. (1950). *J. R. Statist. Soc. B*, **12**, 145.
 AROIAN, L. A. & LEVENE, H. (1950). *J. Amer. Statist. Ass.* **45**, 520.
 BARNARD, G. A. (1947). *J. Amer. Statist. Ass.* **42**, 658.
 BARTLETT, M. S. (1953). *Proc. Camb. Phil. Soc.* **49**, 263.
 BURMAN, J. P. (1946). *J. R. Statist. Soc. Suppl.* **8**, 98.
 DODGE, H. F. (1943). *Ann. Math. Statist.* **14**, 264.
 DUDDING, B. P. & JENNET, W. J. (1942). British Standards Institution, B.S. 600 R.
 DUNCAN, A. J. (1952). *Quality Control and Industrial Statistics*. Chicago: Irwin.
 FISHER, R. A. (1952). *Biometrics*, **8**, 183.
 HERNE, H. & NICKOLLS, G. D. Unpublished.
 KENDALL, D. G. (1951). *J. R. Statist. Soc. B*, **13**, 184.
 LINDLEY, D. V. (1951). *J. R. Statist. Soc. B*, **13**, 173.
 LINDLEY, D. V. (1952). *Proc. Camb. Phil. Soc.* **48**, 277.
 LINDLEY, D. V. (1953). *J. R. Statist. Soc. B*, **15**, 30.
 MOLINA, E. C. (1947). *Poisson's Exponential Binomial Limit*. New York: Von Nostrand.
 NEYMAN, J. & PEARSON, E. S. (1936). *Statist. Res. Mem.* **1**, 113.
 PAGE, E. S. (1954). *J. R. Statist. Soc. B*, **16**, (at Press).
 SHEWHART, W. A. (1931). *Economic Control of Quality of Manufactured Product*. New York: Macmillan.
 SMITH, W. L. (1953). *Proc. Camb. Phil. Soc.* **49**, 449.
 SOBEL, M. & WALD, A. (1949). *Ann. Math. Statist.* **20**, 502.
 WALD, A. (1947). *Sequential Analysis*. New York: Wiley.
 WALD, A. & WOLFOWITZ, J. (1945). *Ann. Math. Statist.* **16**, 30.
 WALKER, A. M. (1950). *J. R. Statist. Soc. B*, **12**, 301.

APPENDIX. Table of $pL(0)$ (see § 5.2)

h/b $X = pb$	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00	4.25	4.50	4.75	5.00
0.05	28.1	41.0	77.4	829	1.41×10^3	2.94×10^3	9.14×10^3	4.64×10^4	1.01×10^5	2.77×10^5	1.03×10^6	3.20×10^6	8.15×10^6	—	—	—
0.10	14.8	21.0	37.7	216	354	688	1.79×10^3	5.65×10^3	1.14×10^4	2.71×10^4	7.47×10^4	1.81×10^5	4.19×10^5	1.07×10^6	2.71×10^6	6.28×10^6
0.15	10.4	14.4	31.1	100	158	290	660	1.65×10^3	3.02×10^3	6.58×10^3	1.52×10^4	3.20×10^4	6.65×10^4	1.46×10^5	3.19×10^5	6.67×10^5
0.20	8.13	11.1	18.3	59.1	90.3	156	321	686	1.21×10^3	2.36×10^3	4.80×10^3	9.18×10^3	1.75×10^4	3.46×10^4	6.78×10^4	1.30×10^5
0.25	6.80	9.13	14.5	39.7	58.9	97.0	183	351	588	1.06×10^3	1.96×10^3	3.46×10^3	6.14×10^3	1.11×10^4	2.00×10^4	3.56×10^4
0.30	5.92	7.81	12.0	29.1	41.8	66.0	116	204	326	548	941	1.56×10^3	2.60×10^3	4.39×10^3	7.37×10^3	1.23×10^4
0.35	5.29	6.88	10.3	22.5	31.6	47.9	79.1	130	199	316	509	799	1.26×10^3	2.00×10^3	3.16×10^3	4.97×10^3
0.40	4.82	6.18	9.04	18.2	25.0	36.5	57.3	89.3	138	198	301	450	674	1.02×10^3	1.53×10^3	2.28×10^3
0.45	4.45	5.65	8.07	15.3	20.4	29.0	43.5	64.6	91.2	132	192	274	393	565	810	1.16×10^3
0.50	4.16	5.22	7.31	13.1	17.2	23.7	34.2	48.8	66.7	92.9	130	178	245	388	465	638
0.55	3.93	4.87	6.71	11.5	14.8	19.9	27.8	38.3	50.8	68.4	92.0	122	162	216	286	377
0.60	3.73	4.59	6.21	10.2	12.9	17.0	23.1	30.9	40.0	52.3	68.1	87.8	113	145	186	238
0.65	3.56	4.34	5.80	9.21	11.5	14.9	19.7	25.7	32.5	41.3	52.4	65.7	82.2	103	128	158
0.70	3.42	4.14	5.46	8.42	10.4	13.2	17.1	21.8	27.0	33.6	41.6	50.9	62.2	75.7	91.7	111
0.75	3.30	3.96	5.17	7.77	9.47	11.8	15.1	18.8	22.9	28.0	33.9	40.7	48.7	58.0	68.7	81.1
0.80	3.19	3.81	4.91	7.23	8.72	10.8	13.5	16.6	19.9	23.8	28.4	33.5	39.3	45.9	53.4	61.8
0.85	3.10	3.67	4.69	6.78	8.10	9.88	12.2	14.8	17.5	20.7	24.2	28.2	32.6	37.5	42.8	48.8
0.90	3.01	3.55	4.50	6.40	7.57	9.15	11.1	13.3	15.6	18.2	21.1	24.2	27.6	31.4	35.4	39.7
0.95	2.94	3.45	4.33	6.07	7.13	8.53	10.3	12.2	14.1	16.3	18.7	21.2	23.9	26.8	30.0	33.3
1.00	2.87	3.35	4.19	5.78	6.75	8.01	9.56	11.2	12.9	14.7	16.7	18.8	21.1	23.4	25.9	28.5
1.05	2.81	3.27	4.05	5.54	6.42	7.56	8.96	10.4	11.9	13.5	15.2	17.0	18.8	20.8	22.8	24.9
1.10	2.76	3.19	3.93	5.32	6.13	7.18	8.44	9.75	11.0	12.5	14.0	15.5	17.1	18.7	20.4	22.1
1.15	2.71	3.12	3.83	5.13	5.88	6.85	8.00	9.18	10.3	11.6	12.9	14.3	15.6	17.0	18.5	20.0
1.20	2.67	3.06	3.73	4.96	5.66	6.56	7.62	8.70	9.75	10.9	12.1	13.2	14.5	15.7	17.0	18.3
1.25	2.63	3.00	3.64	4.81	5.48	6.30	7.28	8.28	9.24	10.3	11.3	12.4	13.5	14.6	15.7	16.9
1.50	2.47	2.78	3.31	4.26	4.75	5.38	6.12	6.85	7.53	8.25	8.98	9.72	10.4	11.2	11.9	12.7
1.75	2.36	2.62	3.08	3.91	4.31	4.89	5.44	6.03	6.58	7.16	7.75	8.31	8.91	9.48	10.1	10.6
2.00	2.28	2.50	2.92	3.68	4.02	4.46	5.00	5.52	5.99	6.48	7.00	7.50	7.98	8.50	9.00	9.48
2.25	2.22	2.42	2.79	3.52	3.80	4.21	4.70	5.18	5.58	6.03	6.50	6.95	7.38	7.85	8.30	8.75
2.50	2.18	2.35	2.70	3.40	3.65	4.02	4.48	4.93	5.30	5.72	6.15	6.55	6.97	7.40	7.82	8.22
2.75	2.14	2.30	2.63	3.32	3.54	3.83	4.32	4.74	5.09	5.47	5.91	6.27	6.63	7.04	7.45	7.84
3.00	2.12	2.26	2.56	3.25	3.44	3.76	4.20	4.61	4.92	5.31	5.70	6.06	6.42	6.81	7.20	7.56