Grouped Data-Sequential Probability Ratio Tests and Cumulative Sum Control Charts

Stefan H. STEINER

Department of Statistics and Actuarial Sciences
University of Waterloo
Waterloo, Ontario   N2L 3G1
Canada

P. Lee GEYER and George O. WESOLOWSKY

Faculty of Business
McMaster University
Hamilton, Ontario   L8S 4M4
Canada

Methodology is proposed for the design of sequential methods when data are obtained by gauging articles into groups. Exact expressions are obtained for the operating characteristics and average sampling number of Wald sequential probability ratio tests and for the average run length of cumulative sum (CUSUM) schemes based on grouped data. Step-by-step design algorithms are provided to assist the practitioner. The methodology is illustrated assuming a normal process with known standard deviation in which we wish to detect shifts in the mean. An example from a progressive die operation is presented. The methods proposed are simple to implement and are an economical alternative to variables-data-based sequential sampling plans and CUSUM control charts.

KEY WORDS: CUSUM; Fast initial response (FIR); Grouped data; Parametric multinomial; SPRT.

It is not always possible or practical to use variables or precise measurement data in quality control. The widespread occurrence of binomial pass/fail attribute data in industry attests to the economic advantages of collecting go/no-go data over exact measurements. Variables data provide more information, but gauging, or classifying, observations into one of several groups based on a critical dimension is often preferred because it takes less skill, is faster and less costly, and is a tradition in certain industries (Ladany 1976; Schilling 1982). Gauging observations results in grouped data, with binomial attribute data representing the special case of two groups. For more information on grouped data, see Haitovsky (1982).

Stevens (1948), Mace (1952), and Ott and Mundel (1954) attempted to bridge the gap in efficiency between variables and binomial attributes procedures by using go/no-go gauges set at artificial levels. The classification of units as conforming or nonconforming is inefficient when the proportion of nonconforming units is small. The sample size required for an attributes plan to achieve any given operating characteristics is inversely related to the size of the proportion nonconforming that it is required to detect. As a result, a gauge limit that classifies a higher proportion of items as nonconforming (pseudononconforming) will be statistically more efficient and offer more information about the characteristic of interest. The focus of much of this research has been the testing or control of the mean of a normal distribution.

Others have striven for greater efficiency by using three groups instead of two. Beja and Ladany (1974) proposed using three attributes to test for one-sided shifts in the mean of a normal distribution when the process dispersion is known. Ladany and Sinuary-Stern (1985) discussed the curtailment of artificial attribute-sampling plans with two or three groups. The first to consider the general k-group case were Steiner, Geyer, and Wesolowsky (1994, in press),

who developed methodology for one-sided and two-sided acceptance sampling plans, acceptance control charts, and Shewhart-type control charts.

In the realm of sequential quality-control methods, less work has been done. Schneider and O'Cinneide (1987) proposed a cumulative sum (CUSUM) scheme for monitoring the mean of a normal distribution with a single compressed limit gauge. They determined solutions based on the normal approximation to the binomial. Geyer, Steiner, and Wesolowsky (in press) extended this CUSUM to the use of two compressed limit gauges placed symmetrically about the midpoint between the target mean and the mean that the chart is intended to detect. The Geyer et al. (in press) solutions are exact and are derived through the theory of the random walk.

In this article we derive sequential probability ratio tests (SPRT's) and CUSUM procedures based on group data with any number of groups. Section 1 introduces the proposed integer scoring procedure based on the likelihood ratio. Section 2 considers the design and implementation of grouped-data SPRT's for testing simple hypotheses about a parameter of interest when data are grouped and the probability distribution of the quality characteristic is known. Using the theory of sequential analysis (Wald 1947), we derive exact expressions for the operating characteristics (OC's) and the average sampling number (ASN).

The design and implementation of grouped-data CUSUM quality-control schemes are discussed in Section 3. Following Page (1954), we consider the proposed grouped-data CUSUM as a sequence of grouped-data SPRT's and derive the average run length (ARL) using the properties of the

individual SPRT's. We also give results applicable when using the fast initial response (FIR) feature recommended by Lucas and Crosier (1982).

Sections 4 and 5 turn to practical considerations that arise when applying this methodology. Section 4 considers monitoring the mean of a normal distribution with known standard deviation and discusses the choice of gauge limits and the performance of grouped-data CUSUM's relative to traditional variables-based CUSUM's. Section 5 presents a step-by-step design procedure and an example concerning the production of metal fasteners in a progressive die environment. For simplicity, the analysis in Sections 1–5 assumes a unit-sequential implementation of the procedures. Section 6 shows that adapting the procedure to samples of size n is relatively straightforward.

Grouped-data SPRT and CUSUM procedures bridge the gap between the efficiency of binomial-attribute and compressed-limit sequential procedures and that of variables-based sequential methods.

1. A SEQUENTIAL SCORING PROCEDURE FOR GROUPED DATA

Whenever data are grouped, the need arises to assign the grouped observations a numerical value based on their grouping. For go/no-go gauges, observations are usually treated singly as Bernoulli random variables, being either conforming or nonconforming. When observations are grouped into multiple intervals, the likelihood ratio suggests a scoring system. The likelihood ratio is used because it has great prominence as a measure of statistical evidence in traditional hypothesis testing, sequential sampling, and the development of CUSUM control charts. The Neyman–Pearson lemma implies that the likelihood ratio test is the most powerful test for comparing simple hypotheses. Wald (1947) showed that a similar optimality property applies to the use of the likelihood ratio in sequential sampling: The SPRT minimizes the ASN under H_0 and H_1 among all sequential tests for given error probabilities. More recently, Moustakides (1986) proved that the CUSUM procedure based on the likelihood ratio minimizes the ARL under H_1 for a given ARL under H_0.

For the simple hypothesis test H_0: $\theta = \theta_0$ versus H_1: $\theta = \theta_1$, the likelihood ratio is given by the ratio of the likelihood of the data under H_1 to the likelihood of the data under H_0. When observations are grouped into k intervals, the likelihood ratio is a ratio of multinomial likelihoods in which the group probabilities depend on the parameter specifications in the underlying probability distributions under H_0 and H_1. Specifically, let the random variable X have probability distribution $f(x; \theta)$ and cumulative distribution function $F(x; \theta)$. Let $t_1 < t_2 < \cdots < t_{k-1}$ denote the $k - 1$ endpoints or gauge limits of the k grouping intervals. We assume for the moment that the $k - 1$ gauge limits are given. In many applications the grouping criteria are predetermined because they are based on some standard classification device or procedure. In Section 4 this assumption is relaxed and the optimal placement of group limits for detecting shifts in a normal mean is discussed. Defining $t_0 = -\infty$ and $t_k = \infty$, the probability that an observation

falls into the jth interval is denoted by

$$\pi_j = F(t_j; \theta) - F(t_{j-1}; \theta), \qquad j = 1, 2, \ldots, k, \quad (1.1)$$

the dependence of π_j on θ being understood. The contribution to the log-likelihood ratio of an observation that falls into the jth interval is thus given by the weight

$$l_j = \ln\{\pi_j(\theta_1)/\pi_j(\theta_0)\}, \qquad j = 1, 2, \ldots, k. \quad (1.2)$$

Using the analysis presented in Sections 2 and 3, the properties of sequential procedures based on integer scores can be found. Thus, for implementation, group scores are obtained by first scaling and then rounding off the likelihood ratio weights. Let

$$w_j = \text{round}(ql_j), \qquad j = 1, 2, \ldots, k, \quad (1.3)$$

denote the group score applied to any observation in the jth group, where q is the chosen scaling factor. Define $\mathbf{w} = (w_1, w_2, \ldots, w_k)$. We assume that all w_j scores are unique; if two or more groups lead to the same score, either the scaling factor should be increased or groups should be combined.

Due to the rounding of log-likelihood ratio weights, the resulting schemes are only approximately based on the optimal sequential probability ratio. The properties of the resulting random walk, however, can be made arbitrarily close to optimal by increasing the scaling factor. In subsequent sections we use the fact that, so long as the number of groups is greater than or equal to 2 and $\theta_0 \neq \theta_1$, at least one individual score is positive and at least one is negative. This implies $\max(\mathbf{w}) > 0$ and $\min(\mathbf{w}) < 0$ and ensures that the SPRT's and CUSUM schemes are capable of concluding either in favor of the null or the alternative hypothesis.

A trade-off is involved in the appropriate choice of the scaling factor q. The solution approach, presented in Sections 2 and 3, requires integer scores and is less computationally intensive to design (and easier to implement) when the scores are as close to 0 as possible. We wish, however, to stay as close as possible to the optimal relative weights suggested by the likelihood ratio. We have found that, in most cases, choosing a scaling factor so that the spread in the sample scores $[\max(\mathbf{w}) - \min(\mathbf{w})]$ is approximately 50; that is, setting $q = 50/[\max(l_1, \ldots, l_k) - \min(l_1, \ldots, l_k)]$ yields results that are indistinguishable from simulation results using the log-likelihood ratio weights. Naturally, smaller scaling factors are also feasible but may yield slightly inferior results. The relative size of the group scores drives the solution. Thus, if the group scores derived through (1.3) have a common factor, all the scores can be divided by this factor without affecting the efficiency of the solution.

2. SEQUENTIAL PROBABILITY RATIO TESTS WITH GROUPED DATA

Consider a sequential test of H_0: $\theta = \theta_0$ versus H_1: $\theta = \theta_1$, where each unit is assigned a sample score s and where $s_i = w_j$, as given by (1.3), if the ith unit is classified into group j. Choosing absorbing barriers at $\ln B$ and $\ln A$, the sampling terminates on the Nth trial, where N is the smallest integer for which either $\mathbf{S} = s_1 + s_2 + \cdots + s_N \geq \ln A$

or $S = s_1 + s_2 + \cdots + s_N \leq \ln B$, where $0 < B < 1 < A < \infty$. S is the value of the SPRT at termination. If $S \geq \ln A$, we conclude that the parameter has shifted to θ_1, whereas if $S \leq \ln B$, we decide in favor of θ_0. Because the observations are all independent and identically distributed, the sequence $S = s_1 + \cdots + s_N$ can be viewed as a random walk with step sizes w between absorbing barriers $\ln B$ and $\ln A$. Because the step sizes can take on only a finite number of integer values, we may use the theory of sequential analysis (Wald 1947) to derive the OC's and ASN.

To determine the OC's and ASN of this SPRT, we first derive the probability distribution for all possible terminating values of the SPRT. See the Appendix for a derivation of $\xi_j = \Pr(S = c_j)$, where the vector $c = (c_1, c_2, \ldots, c_d)$ denotes all the possible terminating values of the SPRT including overshoots of the absorbing barriers. Because all w_j's are integers, d, the number of different possible terminating values is finite and depends on the range in the group scores and the scale of the absorbing barriers. Let $[a]$ be the smallest integer greater than or equal to $\ln A$ and $[b]$ be the largest integer smaller than or equal to $\ln B$. Then the probability that the random walk terminates with $S \leq [b]$, and thus accepts the null hypothesis, is given by

$$P_{\text{accept}}(\theta; [a], [b]) = \sum_{j \in c^-} \xi_j, \qquad (2.1)$$

where $c^- = \{c_j : c_j \leq [b]\}$. This expression allows the determination of the OC curve of the sequential test.

Using the probability distribution of S and Wald's equations, we may derive the ASN of the sequential test, denoted $E(N)$. By Wald's first equation (Wald 1947, A: 69), if $E(N) < \infty$ and $E(s) \neq 0$, then

$$E(N; [a], [b]) = E(S)/E(s) = \frac{\sum_{j=1}^{d} \xi_j c_j}{\sum_{j=1}^{k} \pi_j w_j}. \qquad (2.2)$$

Wald (1947) also showed that if $E(s) = 0$ and $E(s^2) < \infty$, then

$$E(N; [a], [b]) = E(S^2)/E(s^2) = \frac{\sum_{j=1}^{d} \xi_j c_j^2}{\sum_{j=1}^{k} \pi_j w_j^2}. \qquad (2.3)$$

Equations (2.1)–(2.3) are valid for an SPRT with initial value 0 and absorbing barriers at $[b] < 0$ and $[a] > 0$. Through a translation, this is equivalent to an SPRT with initial value $\nu, 0 < \nu < h$, and absorbing barriers at 0 and h. This translation is of interest because the traditional $(0, h)$ CUSUM chart can be modeled as a geometric series of $(0, h)$ SPRT's. Define $E(N; \nu)$ and $P_a(\nu)$ as the ASN and probability of concluding in favor of the H_0, respectively, for a $(0, h)$ SPRT with initial score ν. Then $E(N; \nu) = E(N; [a] = h - \nu, [b] = -\nu)$ and $P_a(\nu) = P_{\text{accept}}(\theta; [a] = h - \nu, [b] = -\nu)$ as given by Equations (2.1) and (2.2) or (2.3).

3. CUSUM CONTROL CHARTS WITH GROUPED DATA

CUSUM charts consist of plotting $Y_i = \max(0, Y_{i-1} + \ln(f(x; \theta_1)/f(x; \theta_0)))$, where $Y_0 = 0$. The process is assumed to be in state H_0 as long as $Y_i < h$ and is deemed to shift to state H_1 if $Y_i \geq h$. The CUSUM represents a sequence of Wald tests with initial score 0 and absorbing barriers at 0 and h (Page 1954). It is easy to show that the ARL of a CUSUM chart is given by

$$\text{ARL} = \frac{E(N; \nu = 0)}{1 - P_a(\nu = 0)}, \qquad (3.1)$$

where $E(N; \nu = 0)$ and $P_a(\nu = 0)$ are the ASN and probability of acceptance of the component $(0, h)$ Wald tests with starting value 0 (Page 1954). Unfortunately, $E(N; \nu = 0)$ and $P_a(\nu = 0)$ are not directly obtainable from (2.1)–(2.3) because those expressions are derived assuming that the SPRT starting value is not equal to lower barrier values; that is, $\nu > 0$. Expressions for $E(N; \nu = 0)$ and $P_a(\nu = 0)$ can be derived by conditioning on the value of the first sample score, however. Notice also that the ARL of a CUSUM given by (3.1) is implicitly dependent on the true parameter value θ because changes in θ will change the group probabilities as given by (1.1).

Define w^+ and w^- as the set of all the possible sample scores that are positive and nonpositive, respectively. Then, remembering from (1.1) that $\pi_j = \Pr(s = w_j)$, we get

$$E(N; \nu = 0) = 1 + \sum_{j \in w^+} \pi_j E(N; \nu = w_j) \qquad (3.2)$$

and

$$P_a(\nu = 0) = \sum_{j \in w^-} \pi_j + \sum_{j \in w^+} \pi_j P_a(\nu = w_j), \qquad (3.3)$$

where $E(N; \nu = x) = 0$ and $P_a(\nu = x) = 0$ if $x \geq h$. Thus, the ARL of the grouped data CUSUM is given by

$$\text{ARL} = \frac{1 + \sum_{j \in w^+} \pi_j E(N; \nu = w_j)}{\sum_{j \in w^+} \pi_j (1 - P_a(\nu = w_j))}. \qquad (3.4)$$

Using an FIR (Lucas and Crosier 1982), the starting value of the CUSUM is $\omega > 0$. The FIR ARL, denoted ARL(ω), is determined by conditioning on the outcome of the first Wald test in the sequence. Only the initial Wald test is unique; if the initial test does not signal, then all subsequent Wald tests start at 0. With ARL given by (3.4) and $E(N)$ from Section 2,

$$\text{ARL}(\omega) = E(N; \nu = \omega) + P_a(\nu = \omega)\text{ARL}. \qquad (3.5)$$

4. OPTIMAL GAUGE–LIMIT PLACEMENT

In practice, the placement of group or gauge limits is often predetermined through the use of standard gauges. In some circumstances, however, design of the step gauge is possible, and we may wish to determine the optimal gauge limits. In any event, it is of interest to compare the efficiency of using grouped data relative to the traditional variables-based approaches. Clearly, grouped data will be less efficient because some information is lost due to the grouping. As will be shown, however, this loss of information is small for well-chosen group limits and, as a result, may be more than compensated for by lower data-collection costs.

Table 1.   Optimal SPRT Gauge Limits: Optimal Gauge Limits $t = (\mu_0 + \mu_1)/2 + \sigma\Delta t$

# groups	$\mu_1 - \mu_0 = .5\sigma$ Δt	$\mu_1 - \mu_0 = 1\sigma$ Δt	$\mu_1 - \mu_0 = 1.5\sigma$ Δt	$\mu_1 - \mu_0 = 2\sigma$ Δt
2	0	0	0	0
3	±.6209	±.6487	±.6976	±.7712
4	0, ±.9972	0, ±1.0439	0, ±1.1203	0, ±1.2224
5	±.3893, ±1.2652	±.4104, ±1.3259	±.4458, ±1.4201	±.4963, ±1.5385
6	0, ±.6716, ±1.4721	0, ±.7091, ±1.5437	0, ±.7696, ±1.6503	0, ±.8508, ±1.7790
7	±.2861, ±.8918, ±1.6397	±.3032, ±.9423, ±1.7201	±.3313, ±1.0211, ±1.836	±.3703, ±1.1223, ±1.972

The methodology presented in Sections 1–3 is applicable for hypothesis tests involving any parameters from any distribution so long as H_0 and H_1 completely specify the distribution. The efficiency of the method and optimal group limits depend, however, on the underlying distribution of the quality characteristic of interest. We derive optimal gauge limits for SPRT and CUSUM procedures for detecting mean shifts of a normal distribution with known standard deviation. We assume, without the loss of generality, that in control the process has mean $\mu_0 = 0$ and standard deviation $\sigma = 1$.

The goal of our SPRT is to distinguish between μ_0 and μ_1. As a result, we may maximize the SPRT's ability to differentiate between the two parameter values by determining the gauge limits that maximize the difference between the expected log-likelihood ratio weight (1.2) under μ_0 and under μ_1. With this goal in mind, we solve the following maximization problem:

$$\text{maximize} \quad E(l|\mu = \mu_1) - E(l|\mu = \mu_0), \quad (4.1)$$

where l is a random variable equal to l_j with probability π_j for $j = 1, 2, \ldots, k$. We use l rather than s, as defined in Section 2, to ensure that the optimal gauge limits do not depend on the scaling factor used. Strictly speaking the preceding optimization problem is appropriate only if we are equally interested in the parameter values μ_0 and μ_1. If not, we should consider a weighted difference of the expected log-likelihood ratio. The solution of (4.1), however, will provide guidance as to the best gauge limits in any event.

This maximization problem is solved using the Nelder–Mead simplex algorithm (Press, Flannery, Teukolosky, and Vetterling 1988). Results for various parameter values are given in Table 1. To save space, the optimal solutions in Table 1 are given in terms of Δt, where given μ_0, μ_1, and σ the optimal gauge limits are $t = (\mu_0 + \mu_1)/2 + \sigma\Delta t$. For example, when $\mu_0 = 12, \mu_1 = 15$, and $\sigma = 2$ and the number of groups equals four, we get optimal gauge limits (11.26,

Table 2.   Optimal CUSUM Gauge Limits: Assume $\mu_0 = 0$,
$\mu_1 = 1, \sigma = 1$, and $ARL_0 = 1,000$

# groups	t_1	t_2	t_3	t_4	t_5
2	.8861				
3	.3958	1.5637			
4	.0252	.9947	1.9090		
5	−.2945	.5720	1.3013	2.1194	
6	−.5591	.1787	.8415	1.5017	2.2019

13.5, 15.74). The effect of $\mu_1 - \mu_0$ on the optimal gauge limits written in terms of Δt is small. Thus, from Table 1, close to optimal gauge limits can be determined for most situations.

CUSUM's are typically evaluated in terms of their ARL at the alternate mean value, ARL(μ_1), when their ARL at the null, ARL(μ_0), equals some constant, ARL_0 say. For a grouped-data CUSUM, this comparison is difficult to make due to the discrete nature of the problem. For any given gauge limit design t, it is usually impossible to set the (integer) decision barrier h so that ARL(μ_0) equals ARL_0 precisely. To improve the comparison, we use interpolation. As h varies, a plot of $\ln(\text{ARL}(\mu_0))$ versus ARL(μ_1) forms approximately a straight line. Let h^- and h^+ equal the largest and smallest absorbing barriers, respectively, such that $\text{ARL}(\mu_0|h^-)$ is less than ARL_0 and $\text{ARL}(\mu_0|h^+)$ is greater than ARL_0. Then, the theoretical ARL at the alternate mean when the in-control ARL is ARL_0 is given by

$$\text{ARL}(\mu_1|\text{ARL}(\mu_0) = ARL_0)$$

$$= \text{ARL}(\mu_1|h^-) + \phi \ln\left(\frac{ARL_0}{\text{ARL}(\mu_0|h^-)}\right), \quad (4.2)$$

where

$$\phi = \frac{\text{ARL}(\mu_1|h^+) - \text{ARL}(\mu_1|h^-)}{\ln(\text{ARL}(\mu_0|h^+)) - \ln(\text{ARL}(\mu_0|h^-))}.$$

Optimal gauge limits that minimize the ARL given by (4.2) were found using the Nelder–Mead simplex algorithm (Press et al. 1988). Table 2 gives results for the case $\mu_1 = 1, ARL_0 = 1,000$. For different ARL_0 values, the results are very similar. For different values of μ_1, scaling the results in Table 2 by μ_1 yields near optimal gauge limits. Not surprisingly, the optimal gauge limits for CUSUM charts are different from the optimal limits for SPRT's.

It is of interest to evaluate the loss in efficiency that must be expected when articles are gauged into groups rather than measured precisely. A direct comparison of the grouped-data CUSUM and the traditional variables-based approach is difficult due to the discrete nature of any scheme that uses categorical data. Using interpolation, as given by (4.2), however, a comparison can be made. Using the solution approach suggested by Brook and Evans (1972) for a variables CUSUM when H_0: $\mu_0 = 0, H_1$: $\mu_1 = 1$, with $\sigma = 1$ and $h = 5$, we obtain ARL's of 904.81 and 10.39 at the null and alternate mean values. We consider grouped-data CUSUM's with two to six groups. The log-likelihood

Table 3.   Optimal Log-likelihood Ratio and Scores for Grouped-Data CUSUM

# groups	Log-likelihood ratio (1.2)	Assigned group scores w (1.3)
2	−.5802, 1.0661	−18, 32
3	−.8740, .4282, 1.5811	−18, 9, 32
4	−1.1296, .0092, .8883, 1.8653	−19, 0, 15, 31
5	−1.3688, −.3393, .4177, 1.1455, 2.0440	−20, −5, 6, 17, 30
6	−1.5774, −.6596, .0097, .6477, 1.4983, 2.1150	−21, −9, 0, 9, 18, 29

ratio weights presented in Table 3 are derived using the optimal gauge limits suggested in Table 2. The group scores are given by (1.3) with $q = 50/(l_k - l_1)$, where k equals the number of groups.

To conduct the comparison, we set $ARL_0 = 904.81$ and find h^- and h^+. Equation (4.2) then gives the theoretical matching ARL, $ARL(\mu_1|ARL(\mu_0) = 904.81)$. Using formulas analogous to (4.2), we also determine $ARL(\mu_0|ARL(\mu_1) = 10.39)$. Table 4 shows that the ARL properties improve significantly as the number of groups increases.

5.   DESIGN OF GROUPED–DATA SPRT'S AND CUSUM PROCEDURES

To assist the practitioner to implement grouped-data SPRT's or CUSUM's to test H_0: $\theta = \theta_0$ versus H_1: $\theta = \theta_1$, we present two iterative design procedures. A grouped-data SPRT with starting value 0 and absorbing barriers at $[a] > 0$ and $[b] < 0$ can be designed following steps S1–S9. Notice that an SPRT with starting value ν and absorbing barriers at 0 and h can be determined using these steps by applying the transformations $h = [a] - [b]$ and $\nu = -[b]$. The key problem in the design is determining appropriate values for $[a]$ and $[b]$.

Grouped-Data SPRT Design Algorithm

S1. Determine, based on application, the null and alternate parameter values θ_0 and θ_1 and the maximum desired Type I and Type II error rates α and β, respectively.

S2. Set group limits (t_i's) either at optimal values as discussed in Section 4 or at predetermined values based on the application.

S3. Use Equation (1.2) to calculate the log-likelihood ratio weight for each group.

S4. Derive integer group scores using (1.3). If the application allows, use $q = 50/[\max(l_1, \ldots, l_k) - \min(l_1, \ldots, l_k)]$. Common factors in the resultant scores can be removed without affecting the efficiency of the procedure.

Table 4.   Comparison of Variables-Based CUSUM and Grouped-Data CUSUM

# groups	ARL(μ₀) = 904.81; ARL(μ₁) estimate	ARL(μ₁) = 10.39; ARL(μ₀) estimate
2	14.37	240.0
3	12.04	458.0
4	11.28	592.6
5	10.96	694.0
6	10.78	756.0
Variables	10.39	904.81

S5. Choose initial values for the absorbing barriers $[a] > 0$ and $[b] < 0$.

S6. Using the current $[a]$ and $[b]$ and the methodology in the Appendix, derive the probability function $\xi_j = \Pr(S = c_j)$ for S, the terminating value of the SPRT.

S7. Using ξ_j and Equation (2.1), calculate P_{accept} at $\theta = \theta_0$ and $\theta = \theta_1$. Denote the actual error rates obtained by $\alpha_a = 1 - P_{accept}(\theta_0)$ and $\beta_a = P_{accept}(\theta_1)$.

S8. If $\alpha_a < \alpha$ and $\beta_a < \beta$ and $[a]$ and $[b]$ are the values closest to 0 that satisfy both these constraints, then stop.

S9. Otherwise,

- if $\alpha_a > \alpha$ and $\beta_a > \beta$, decrement $[b]$ and increment $[a]$ by one unit;
- if $\alpha_a < \alpha$ and $\beta_a > \beta$, decrement $[b]$ by one unit;
- if $\alpha_a > \alpha$ and $\beta_a < \beta$, increment $[a]$ by one unit;
- if $\alpha_a < \alpha$ and $\beta_a < \beta$, increment $[b]$ and decrement $[a]$ by one unit and go to S6.

For the continuous-variable problem, Wald (1947) suggested choosing $A = (1 - \beta)/\alpha$ and $B = \beta/(1 - \alpha)$. This choice is derived by ignoring the possibility of overshooting the barriers but is useful as a guide. We have found that good initial values for the absorbing barriers are the Wald approximations appropriately scaled; that is, let $[b]$ equal the largest integer smaller than $q \ln[\beta/(1 - \alpha)]$, and let $[a]$ equal the smallest integer larger than $q \ln[(1 - \beta)/\alpha]$, where q is the scaling factor discussed in Section 1.

CUSUM procedures are usually designed by specifying the minimum acceptable ARL at the null ARL_0 and the maximum acceptable ARL at the alternate ARL_1. We desire both $ARL(\theta_0) > ARL_0$ and $ARL(\theta_1) < ARL_1$. For our CUSUM procedure, however, h, the absorbing barrier, is the only design parameter. This implies that we cannot necessarily satisfy both ARL criteria simultaneously. Typically, the ARL at the null (i.e., the in-control ARL) is of greater concern. As a result, the following design algorithm ensures that the constraint $ARL(\theta_0) > ARL_0$ is satisfied while minimizing $ARL(\theta_1)$. The algorithm C1–C6 is easily adapted to design a CUSUM that maximizes $ARL(\theta_0)$ while $ARL(\theta_1) < ARL_1$.

Grouped-Data CUSUM Design Algorithm

C1. Specify the minimum acceptable ARL at the null, ARL_0.

C2. Perform Steps S2–S4.

C3. Choose an initial value for the absorbing barrier h. For typical CUSUM applications, the initial value of h should be a multiple of $\max(w)$.

C4. Use Equation (3.4) to determine $ARL(\theta_0|h)$.

Table 5. Design Iterations for SPRT Example

Iteration	[b]	[a]	α_a	β_a
1	−21	21	.0710	.0561
2	−20	20	.0743	.0624
3	−19	19	.0787	.0662
4	−18	18	.0876	.0758
5	−17	17	.1088	.0792
6	−17	18	.0869	.0811
7	−16	17	.1082	.0835
8*	−16	18	.0864	.0856
9	−15	17	.1050	.1112

C5. If we have found the smallest h, where $\text{ARL}(\theta_0|h) \geq \text{ARL}_0$, then stop.

C6. Adjust h as follows:

- If $\text{ARL}(\theta_0|h) < \text{ARL}_0$, increase h. Increment h by one unit if $\text{ARL}(\theta_0|h)$ is close to ARL_0 and a larger value if not.

- Otherwise, decrease h by one unit.

Go to step C4.

Consider the following example to illustrate the solution procedure. We first design an appropriate SPRT. In control, the process of interest is normal with mean $\mu_0 = 74.3$ and standard deviation $\sigma = 1.3$, and we wish to detect mean shifts to $\mu_1 = 75.6$ (i.e., mean shifts of one standard deviation unit). Using a standardized gauge, the group limits are $t = (74, 75, 76)$. This implies group probabilities of (.4087, .2961, .1997, .0955) and (.1092, .2130, .2986, .3792) under H_0 and H_1, respectively. The log-likelihood ratios (1.2) are thus $(-1.3200, -.3294, .4096, 1.3791)$. Based on the application, the error rates α and β should both be less than .1. This completes steps S1–S3. Scaling the weights as recommended in step S4 yields scores $(-24, -6, 8, 26)$. These scores have a common factor of 2; thus without any loss in efficiency, we use $\mathbf{w} = (-12, -3, 4, 13)$. Because the scaling factor needed to yield $\mathbf{w}$ is $q = 9.26$, the recommended initial estimates for the absorbing barriers using the scaled Wald approximations are -21 and 21. Table 5 shows the results of the iteration S6–S9. The best solution occurred at iteration 8. Based on $[a] = 18$ and $[b] = -16$, Equation (2.2) gives ASN's of 4.7767 and 4.7616 when $\mu = \mu_0$ and $\mu = \mu_1$, respectively.

To illustrate the design of a CUSUM, consider the example given by Steiner et al. (1994). That example concerned the manufacture of metal fasteners in a progressive die environment. The in-control process mean is 74 thousandths of an inch with a standard deviation of 1.3. Using a six-step gauge with group intervals defined by 73, 73.75, 74.35, 74.94, 75.55, and 76.3, a fixed-sample-size control chart to detect a mean shift to 75.3 thousandths of an inch with Type I and Type II error rates of .005 required a sample size of 27 units. Thus, the ARL's of this fixed-sample-size chart, in terms of the number of units examined before a signal, are $5,400 = 27/.005$ and 27.1 units when the process is "in-control" and "out-of-control," respectively. Because this example involves detection of a fairly small mean shift, a CUSUM chart is expected to outperform a Shewhart-type control chart.

Assuming that we wish to do at least as well as the fixed-sample-size approach, set $\text{ARL}_0 = 5,400$. Step C2 yields group weights $(-1.7492, -.9553, -.4503, 0, .4503, .9553, 1.7492)$ and corresponding group scores $\mathbf{w} = (-25, -14, -6, 0, 6, 14, 25)$. From previous experience, $h = 76$ was taken as an initial guess. The results of iterating through steps C3–C6 is shown in Table 6. Because the first iterations yielded $\text{ARL}(\mu_0)$ values significantly less than 5,400, h was incremented by five units at a time for the first four iterations. The best solution is $h = 98$ with $\text{ARL}(\mu_0) = 5,646$ and $\text{ARL}(\mu_1) = 14.6$. This grouped-data CUSUM has approximately the same performance as the fixed-sample-size approach at the null but is dramatically better at the alternate mean.

6. EXTENSION TO SAMPLES OF SIZE n

In this article we have focused on the unit sequential implementation of grouped-data SPRT's and CUSUM's. The same methodology is appropriate, however, when using larger samples. For a sample of size n, where n_j observations are classified into the jth group, the sample weight, defined as the sum of the individual log-likelihood ratios, is $\sum_{j=1}^{k} n_j \log\{\pi_j(\theta_1)/\pi_j(\theta_0)\}$. Scaling the weights to get scores as in Section 1, we may define our sample score as $y = \sum_{j=1}^{k} n_j w_j$. Thus, increasing the sample size increases the number and spread of the possible sample scores.

By the multinomial distribution, $\Pr(y = \sum_{j=1}^{k} n_j w_j) = (n!/\prod_{j=1}^{k} n_j!) \sum_{j=1}^{k} \pi_j^{n_j}$. Different combinations of n_j values may lead to the same sum, however. Let $\mathbf{z} = (z_1, z_2, \ldots, z_m)$ be the m distinct values of the sample score y. Define $p_i = \Pr(y = z_i)$, where the probability $\Pr(y = z_i)$ is the sum of all preceding multinomial probabilities in which the n_j's yield the sample score z_i. The number of possible sample scores m grows exponentially with the number of groups and polynomially with the sample size. Fortunately, we need only consider moderate sample sizes and numbers of groups. If the sample size is large, a normal approximation solution is appropriate, and if the number of groups is large, the problem can be accurately approximated by assuming variables data.

To derive solutions for samples of size n, make the following substitutions in the analysis of the previous sections: $k = m$, $w_i = z_i$, and $\pi_i = p_i$. Note that, if the desired sample size is large, then the scaling factor q may need to be reduced because the spread in the sample scores is now $n[\max(\mathbf{w}) - \min(\mathbf{w})]$ and that N refers to the number of samples until absorption, whereas n is the sample size.

Table 6. Design Iterations for CUSUM Example

Iteration	h	$\text{ARL}(\mu_0)$	$\text{ARL}(\mu_1)$
1	76	1200.4	11.45
2	81	1692.1	12.17
3	86	2470.3	12.97
4	91	3576.4	13.76
5	96	5026.3	14.49
6	97	5336.9	14.49
7*	98	5646.5	14.62

7. SUMMARY

We propose an integer scoring procedure for sequential tests with grouped data. Group scores are integer approximations of the parametric multinomial likelihood ratio. We show how to derive false-alarm rates, power, ASN, and ARL of grouped-data SPRT's and CUSUM procedures. Step-by-step algorithms to determine design parameters are presented. The resulting SPRT's and CUSUM procedures are easily implemented on the shop floor. Optimal grouping criterion for the normal mean case are also discussed. Grouped data are a natural compromise between the low data-collection and implementation costs of binomial (two-group) data and the high information content of variables (∞-group) data.

ACKNOWLEDGMENTS

APPENDIX: DERIVATION OF THE PROBABILITY DISTRIBUTION OF S

We derive the probability distribution of the cumulative sum S at the termination of the grouped-data SPRT. Based on (1.3), the moment-generating function of the group score s is given by $E(e^{st}) = \sum_{j=1}^{k} \pi_j u^{w_j} = \phi_s(t)$, say, where $u = e^t$. Thus, the moment-generating function of $S = \sum_{j=1}^{N} s_j$ is given by $\phi_S(t) = [\phi_s(t)]^N$. Consider

$$\phi_s(t) = \sum_{j=1}^{k} \pi_j u^{w_j} = 1. \qquad (A.1)$$

This is a polynomial in u, which has degree $d = \max(\mathbf{w}) - \min(\mathbf{w})$. Let $u_1, \ldots, u_d$ denote the d roots of (A.1), and assume that $u_i \neq u_j$ for $i \neq j$. The roots are unique so long as $E(s) \neq 0$. When the underlying process parameter θ is such that $E(s) = 0$, then $u = 1$ is a double root.

Now consider Wald's fundamental identity (Wald 1947, A: 16) for a random walk between two absorbing barriers, $E\{e^{St}[\phi_s(t)]^{-N}\} = 1$, which holds for any t in the complex plane such that the moment-generating function of s exists. Because by definition $\phi_s(u_i) = 1$, substituting u_i for e^t in the fundamental identity gives the d equations

$$E(u_i^S) = 1 \quad \text{for} \quad i = 1, \ldots, d. \qquad (A.2)$$

We may obtain the probability distribution of S by conditioning each of the d left sides of the fundamental identity on the terminating value of the process. Let $[a]$ be the smallest integer $\geq \ln A$ and $[b]$ the largest integer $\leq \ln B$. Then, because all w_j's are integer, the only possible terminating values of the random walk are $([b] + \min(\mathbf{w}) + 1), ([b] - \min(\mathbf{w}) + 2), \ldots, [b]$ for acceptance of $\theta = \theta_0$ and $[a], [a+1], \ldots, ([a] + \max(\mathbf{w}) - 1)$ for acceptance of $\theta = \theta_1$. Let $c_1, \ldots, c_d$ denote these d terminating values of the random walk. Then, the d equations of (A.2) may be

written as

$$\sum_{j=1}^{d} \xi_j u_i^{c_j} = 1 \quad \text{for} \quad i = 1, \ldots, d, \qquad (A.3)$$

where

$$\xi_j = \Pr(S = c_j) \quad \text{for} \quad j = 1, \ldots, d. \qquad (A.4)$$

The d equations in (A.3) are linear in ξ_j. Solving this system of linear equations, we can determine all the ξ_j's that give the probability distribution of S.

To illustrate all the calculations required to determine the probability distribution of S, consider a simpler version of the SPRT example introduced in Section 5. Rescaling the ratios through multiplication by 1.75 followed by rounding to the nearest integer results in the most compact possible distinct group scores—namely, $\mathbf{w} = (-2, -1, 1, 2)$. Solving Equation (A.1) for the four roots yields $\mathbf{u} = (-4.516, -.494, 1, 1.9188)$ when $\mu = \mu_0$, and $\mathbf{u} = (-2.1239, -.2367, .5729, 1)$ when $\mu = \mu_1$. Choosing absorbing barriers $[a] = 4$ and $[b] = -4$ the only possible terminating values of this SPRT—that is, possible values for S—are $\mathbf{c} = (-5, -4, 4, 5)$. Subsequently solving the system of four equations in four unknowns given by Equation (A.3) gives $\xi = (.2912, .6489, .0494, .0104)$ when $\mu = \mu_0$ and $\xi = (.016, .0706, .6367, .2767)$ when $\mu = \mu_1$. Based on Equations (2.1) and (2.2), this SPRT thus has a probability of acceptance and ASN, respectively, of .9402 and 5.26 when $\mu = \mu_0$ and .0866 and 5.70 when $\mu = \mu_1$.

REFERENCES

Beja, A., and Ladany, S. P. (1974), "Efficient Sampling by Artificial Attributes," *Technometrics*, 16, 601–611.

Brook, D., and Evans, D. A. (1972), "An Approach to the Probability Distribution of Cusum Run Length," *Biometrika*, 59, 539–549.

Geyer, P. L., Steiner, S. H., and Wesolowsky, G. O. (in press), "Optimal SPRT and CUSUM Procedures Using Compressed Limit Gauges," *IIE Transactions*, 1996.

Haitovsky, Y. (1982), "Grouped Data," in *The Encyclopedia of Statistical Sciences* (Vol. 3), eds. S. Kotz, N. L. Johnson, and C. B. Read, New York: John Wiley, pp. 527–536.

Ladany, S. P. (1976), "Determination of Optimal Compressed Limit Gaging Sampling Plans," *Journal of Quality Technology*, 8, 225–231.

Ladany, S. P., and Sinuary-Stern, Z. (1985), "Curtailment of Artificial Attribute Sampling Plans," *International Journal of Production Research*, 23, 847–856.

Lucas, J. M., and Crosier, R. B. (1982), "Fast Initial Response for CUSUM Quality Control Schemes," *Technometrics*, 24, 199–205.

Mace, A. E. (1952), "The Use of Limit Gages in Process Control," *Industrial Quality Control*, January, 24–31.

Moustakides, G. V. (1986), "Optimal Stopping Times for Detecting Changes in Distributions," *The Annals of Statistics*, 14, 1379–1387.

Ott, E. R., and Mundel, A. B. (1954), "Narrow-Limit Gaging," *Industrial Quality Control*, March, 221–228.

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

Press, W. H., Flannery, B. P., Teukolosky, S. A., and Vetterling, W. T. (1988), *Numerical Recipes*, Cambridge, U.K.: Cambridge University Press.

Schilling, E. G. (1982), *Acceptance Sampling in Quality Control*, New York: Marcel Dekker.

Schneider, H., and O'Cinneide, C. (1987), "Design of CUSUM Control Charts Using Narrow Limit Gauges," *Journal of Quality Technology*, 19, 63–68.

Steiner, S. H., Geyer, P. L., and Wesolowsky, G. O. (1994), "Control Charts Based on Grouped Data," *International Journal of Production Research*, 32, 75–91.

———(in press), "Shewhart Control Charts to Detect Mean and Standard Deviation Shifts Based on Grouped Data," *Journal of Quality and Reliability Engineering International*, 1996.

Stevens, W. L. (1948), "Control by Gauging," *Journal of the Royal Statistical Society*, Ser. B, 1, 54–108.

Wald, A. (1947), *Sequential Analysis*, New York: Dover.