Some aspects of the theory of statistical control schemes

by Emmanuel Yashchin

Control schemes (charts) are widely used in industrial quality control as means of monitoring the quality of manufactured products. These schemes provide a set of criteria for testing whether a given sequence of observations corresponds to an "on-target" state of the production process. In the present work we consider some graphical, computational, and statistical aspects of control charting—criteria of performance, methods of derivation, analysis, design, etc. We introduce the class of "Markov-type" control schemes and discuss some of its properties.

1. Introduction: Control schemes and characterization of their performance

Let $x_1, x_2, \cdots$ be a sequence of observations related to a certain process. The observation x_i may represent, for example,

- Sample percentage of defective chips in the *i*th produced lot;
- Total number of defects found in the *i*th produced wafer;
- Sample mean of four diameters of ball bearings chosen at random during the *i*th production period;
- Sample standard deviation of ten simultaneous measurements (corresponding to various locations) of polyethylene film thickness taken during the *i*th sampling period;
- Waiting time of the *i*th customer in the queue;
- Discrepancy between the actual amount of product shipped in the *i*th month and that predicted by a given model;

and so on; for purposes of our discussion the nature of incoming observations is immaterial. In most practical situations we would like our observations to behave in a certain way, for example to fall as close as possible to some target value, to stay below some prescribed limit, etc. Failure of the observations to comply with this desired behavior is considered as an out-of-control situation; we would like to detect such behavior as early as possible.

In order to monitor sequences of observations, we use control schemes. A control scheme is a set of criteria by which to test, at any given moment of time, whether the process generating the observations is under control. Clearly, many different control schemes can be associated with the same sequence of observations; some of the better known include Shewhart schemes, moving average schemes of various types, etc. In order to compare different types of schemes we need to introduce some criterion of performance of a control scheme. The most important one is represented by the run length (RL) of a scheme. If the input observations correspond to an on-target situation, we would like the RL to be as long as possible; otherwise, it should be as short as possible. Since the RL is a random variable, the actual comparison between control schemes is usually based on some of its characteristics, such as average run length (ARL), median or some other quantile of the run length, etc.

For example, let us assume that the observations are independent, identically distributed (iid), and normal, with mean μ and s.d. $\sigma = 1$. The target level of the process is $\mu = 0$. Then for every control scheme we can draw an ARL curve as a function of the process level μ. In particular, for a two-sided Shewhart scheme [a signal is triggered if a single observation falls outside the interval $(-3, 3)$] the on-target ARL is 370 and the ARL corresponding to $\mu = 1$ (i.e., a shift in mean by one standard deviation) is 45. One can find an alternative control scheme (Cusum) with the same degree of protection against false alarms, but much better sensitivity with respect to an assumed shift in mean: For this scheme ARL $(\mu = 1) = 10$. Thus, one way to characterize the performance of a control scheme with respect to a certain change in the process level would be in terms of its "resolution," which in the case of a fixed sampling interval can be defined as a ratio between the ARLs corresponding to acceptable ("good") and unacceptable ("bad") levels of the process (these levels are determined on the basis of practical and/or economic considerations).

In situations where the sampling interval is not a fixed number, it is natural to characterize the performance of a scheme in terms of the time to signal (TS) instead of the run length; in such cases one can define the resolution as a ratio between the ATSs (average time to signal) corresponding to "good" and "bad" levels of the process (several examples of this approach can be found in [1]). Such situations may occur either because of the inherent properties of the underlying process or because of the sampling policy, which calls, for example, for tightened, normal, or reduced sampling intensity depending on the current assessment of the process behavior.

The sampling intensity needed to achieve a certain resolution may by itself serve as an alternative criterion for comparison between control schemes. To illustrate this point, let us ask the following question: If (in the example above) we took n observations at a time and applied a Shewhart scheme to their sample means, how large should n be to ensure the same sensitivity at $\mu = 1$ as our Cusum scheme? One can show that to achieve that we need to take $n = 3$; a direct conclusion is that in some situations by using a Cusum scheme instead of a Shewhart one can reduce the sampling intensity by a factor of 3 and still keep the same "resolution" between "good" and "bad" levels of the process.

In most practical situations one can identify a set of several control schemes which are likely to have a "reasonable" resolution; once one limits the choice to representatives of this set, other features may become important, depending on the particular application. For example, in quality control applications one may be especially interested in schemes that enable easy visual (graphical) interpretation, including intuitively appealing procedures for estimating the current level of the controlled process, that are robust with respect to slight departures from the assumed features of the model (e.g., normality), that are based on few signal criteria, etc. In some other applications (e.g., robot control) one may be primarily interested in "parsimonious" schemes that are not associated with excessive computational effort or memory requirements.

However, one of the most universally desired features of a good control scheme is its "designability." In other words, once the "good" and "bad" levels of the process as well as corresponding sensitivity requirements are specified, one would like to be able to design, in a relatively straightforward way, a scheme of a given type and determine the sampling intensity needed to meet these requirements. Clearly, a "designable" type of a control scheme must also be "analyzable"; i.e., one must be able to examine (preferably by analytic means) the RL behavior of a scheme with respect to relevant stochastic patterns of incoming observations. As we shall see, these requirements give strong support to "Markov-type" schemes, especially when the monitored sequences of observations typically correspond to an iid process. In quality control applications, Cusum schemes are especially attractive because of their simplicity, easy graphical interpretation, availability of approximations for some more general (non-iid) inputs, etc. (see [2–5]).

Analysis of the RL and careful design of control schemes are especially important in situations where measurements are taken and processed automatically and/or where several parameters are controlled simultaneously. In such situations frequent out-of-control signals associated with *practically* nonimportant changes in process parameters may cause frequent unjustified corrective actions and eventually ruin the discipline of the operator; on the other hand, failure to detect a truly out-of-control situation rapidly may result in a substantial amount of poor-quality product.

For example, consider the following situation related to the production of surface-mounted printed circuit boards. Assume that a board has 400 pads, each containing a certain amount of solder paste deposited by squeezing it through a mask. Before the components are mounted on the board and the solder is reflowed, the volumes of solder paste on each pad are measured by an optical scanner. If the measurements corresponding to some pad show an erratic behavior (which may be caused, for example, by a partially clogged slot in the mask), an out-of-control signal is triggered. It is clear that use of a 3-sigma Shewhart scheme to control the subsequent volumes on a pad would result, on the average, in one false out-of-control signal per board! (Indeed, as we mentioned earlier, its on-target ARL is 370.) So, if we wanted the probability of a false alarm within an eight-hour shift not to exceed 5%, we should have undergone the appropriate design and analysis procedure. The final control scheme would probably represent some kind of a compromise between the desired sensitivity and degree of protection against false alarms.

This example makes it clear that one cannot blindly apply standard control schemes considered in some Quality Control textbooks to situations involving simultaneous control of several parameters. Yet, such situations are rather common in modern industry, and it is not unusual to see thousands of sequences monitored simultaneously. To summarize, any control scheme associated with automatic data processing and/or simultaneous control of several parameters should be thoroughly analyzed before it can be recommended for use. The analysis should involve identification of various possible patterns of incoming observations and assessment of the corresponding run length distributions. Its ultimate aim is to ensure that the run length of the scheme under consideration is sufficiently long if the changes in process parameters are not *practically* important, and sufficiently short if they are.

In the context of modern process control, another property of a control scheme becomes crucial, namely, its capability to incorporate new information immediately upon its arrival, and update itself accordingly. This criterion corresponds to one of the weakest points of Shewhart control schemes, which are typically associated with first subgrouping observations into samples and only then updating the scheme. Clearly, in situations where observations (measurements) are not "naturally" grouped, but rather arrive one at a time, such artificial subgrouping leads to waste of time and loss of resolution power of the scheme; it is not inherently tied to the problem of control itself. One of the main reasons for creating artificial samples when running Shewhart schemes is related to concern that individual observations may have other than normal distribution; by using sample averages one could bring the scheme characteristics closer to those predicted by the normal model.

Some other types of schemes, however, are free of such drawbacks. For example, in the case of Cusum-Shewhart schemes the process of cumulative summation itself brings us (by virtue of the Central Limit Theory) into the normal domain, eliminating any necessity for artificial grouping. In general, every scheme considered in the present work is based on the principle of *immediate utilization of incoming information* introduced in [1]; in this work one can also find discussion on some additional reasons of an economic or statistical nature for not updating the schemes immediately.

The area of statistical control is by no means limited to analysis of control schemes and their run length characteristics. Other important aspects include the problems of estimating the current process mean and detecting the points of change [6–11], forecasting and adaptive control [12–14], cost analysis and economic design of control procedures [15–17], and many others. The scope of the present work also does not enable us to discuss the questions related to actions following an out-of-control signal [various possibilities include an immediate stopping of the production process until the situation is clarified and the problems (if any) dealt with; increasing the sampling intensity and/or switching to a tighter mode of operation which, in turn, could lead to either more drastic actions or return to the normal operating mode—depending on subsequent behavior of the process; introducing an appropriate feedback correction, etc.]. The last topic is discussed in several books, e.g., [18, 19]; for results related to Cusum schemes see also the monographs by van Dobben de Bruyn [20], Woodward and Goldsmith [21], and Bissell [22], and the guide by the British Standards Institution [23].

2. How to derive a "good" control scheme

In this section we give several intuitively appealing ways of obtaining control schemes with good resolution properties. One can probably approach this problem from four different directions, which in many cases lead to the same result. We assume that the controlled parameter is the level (mean) of the sequence $\{x_i\}$.

The first approach calls for trying to estimate the current process level, analytically or by graphical means (e.g., see [6]), and trigger an out-of-control signal if the resulting estimate falls too far from the target region. This approach immediately leads to a Shewhart scheme (where estimation is based on the last observed point) and to a weighted moving average scheme based on the sequence of linear unbiased estimators

$$M_i = w_0 x_i + w_1 x_{i-1} + \cdots + w_{k-1} x_{i-k+1}, \qquad i = 1, 2, \cdots, \tag{1}$$

where $\{w_i\}$ is some (finite or infinite) sequence of weights satisfying $\Sigma_i w_i = 1$. Graphically, it uses a pair of horizontal lines as signal criteria; the signal is triggered at the first time the trajectory $\{M_i\}$ falls outside the strip defined by these lines. Some special cases include the simple (nonweighted) moving average scheme and the geometric moving average scheme [24, 25]:

$$M_i = (1 - w)M_{i-1} + wx_i, \qquad i = 1, 2, \cdots, \tag{2}$$

$(0 < w < 1)$; clearly, it corresponds to an infinite sequence of geometrically decreasing weights.

In general, moving average schemes have a strong "inertia"; i.e., they are typically slow in detecting large changes in the process level. Assigning larger weights to most recent observations may substantially reduce the "inertia" but leaves us with a control scheme based on too many parameters, which complicates the problem of designing a scheme with specified properties. Other unpleasant features of the moving average trajectory are related to its graphical representation: strong serial correlation within this sequence frequently produces an illusion of cycles [26], a single outlier enters into several averages and may create an illusion of a shift in the process level, etc. We must note, however, that the geometric moving average process is widely used in problems of forecasting and feedback control, and as a two-sided control scheme, primarily because of its simplicity, reasonable resolution, and ability to provide a simple estimate of the current process level [27, 28].

The second approach can be based on the likelihood ratio considerations. Let us suppose that $f_0(x)$ and $f_1(x)$ represent the density of observations under on-target and off-target conditions, respectively (for simplicity, we assume that observations are realizations of a continuous random variable). Now we can suggest a control scheme which calls for an out-of-control signal at time n if for some r the last r observations $(x_{n-r+1}, x_{n-r+2}, \cdots, x_n)$ are "significant" in the likelihood ratio sense, i.e., if for some h

$$\sum_{n-r+1}^{n} \log \frac{f_1(x_i)}{f_0(x_i)} > h, \tag{3}$$

the value of h serves primarily to determine the trade-off between the desired degree of protection against false alarms and sensitivity requirements. Each term of the above sum represents a *score* contributed by an appropriate observation. The scheme (3) can be alternatively formulated as follows (Page's scheme): Define the process $\{S_i, i = 0, 1, \cdots\}$ by means of

$$S_0 = 0, \qquad S_i = \left(S_{i-1} + \log \frac{f_1(x_i)}{f_0(x_i)}\right)^+, \qquad i = 1, 2, \cdots, \tag{4}$$

and trigger an out-of-control signal if $S_i > h$. One can see that this method is essentially based on a sequence of SPRTs (Sequential Probability Ratio Tests)—we start accumulating information (scores) in an attempt to reject the hypothesis that the observations come from the on-target population. If, however, the test leads to acceptance (i.e., $S_0 = 0$ for some i)—we immediately re-initiate the test. This interpretation enables one to derive formulas which relate the RL characteristics of the Page's scheme to quantities typically considered in the context of sequential hypothesis testing [29, 30].

As an example, consider the normal case, where $f_0(x)$ and $f_1(x)$ are both normal with a common standard deviation σ and means μ_0 and $\mu_1 > \mu_0$, respectively. Then the "score" is

$$\log \frac{f_1(x_i)}{f_0(x_i)} = \frac{1}{\Delta}(x_i - k); \qquad k = (\mu_0 + \mu_1)/2,$$

$$\Delta = \frac{\sigma^2}{\mu_1 - \mu_0}; \tag{5}$$

i.e., this type of control scheme suggests accumulating differences between observations and "reference value" k, and triggering a signal as soon as the process

$$S_0 = 0, \qquad S_i = (S_{i-1} + x_i - k)^+ \tag{6}$$

exceeds some signal level $h \geq 0$. What we have obtained is the classical *upper* Page's scheme for detecting changes of the process level upwards.

The scheme (4) was introduced by Page [29] and has been extensively studied since then. One of the reasons for its popularity is related to the fact that it has a very good resolution; in fact, Lorden [31] proved its asymptotic "worst-case" optimality as the signal level h tends to infinity. The meaning of worst-case is that the off-target ARL is computed under the assumption that at the moment the change occurs, the scheme is in the worst possible state, i.e., 0. [As opposed to that, we can consider the situation where at the moment of change the value of a control scheme has some "steady-state" distribution (see [32]), which may lead one to the notion of a "steady-state" optimality.] Another reason is related to the availability of an alternative graphical representation; for example, in the normal case one can just sequentially accumulate the successive deviations of the observations from the target, and then apply a V-mask [6] to the resulting cumulative sum trajectory; for that reason, the Page's schemes are also known as Cumulative Sum (Cusum) control schemes.

The third approach calls for testing, at each moment of time n, the hypothesis H_0: No change in distribution occurred before time n—against the alternative H_1: A change occurred before time n. In the case of iid observations with known on-target and off-target densities, this approach leads to the following likelihood ratio test: Reject H_0 at stage n (i.e., trigger an out-of-control signal at time n) if

$$\frac{\max_{1\leq r\leq n} \prod_{i=1}^{n-r} f_0(x_i) \prod_{i=n-r+1}^{n} f_1(x_i)}{\prod_{i=1}^{n} f_0(x_i)} > h, \tag{7}$$

in other words, if for some $1 \leq r \leq n$ the last r observations are "significant." Thus, in this case the control scheme merely reduces to Page's scheme considered earlier. It is clear that our third approach involves estimation of the change point (in the process of maximizing the likelihood in the numerator); the post-signal estimate of the change point is simply the last point at which the Page's scheme (4) had a value 0.

However, (7) represents a much more general approach, as it can be easily generalized to the case in which either $f_0(x)$ or $f_1(x)$ or both are unknown. So, for example, one can develop a control scheme to detect an increase of some given δ in the population mean without actually knowing the on-target mean. To the best of our knowledge, run length characteristics of control schemes obtained by using this approach have never been discussed in the literature, though the "static" problem of inference about the change point has been considered by several authors (e.g., see Hinkley [8, 9]). It seems that such an approach might be especially useful for the purpose of controlling the process variability: It enables one to detect the presence of a new "assignable" cause of variability by just tracking the data, i.e., without specifying the on-target region.

Finally, our fourth approach could be based on Bayesian considerations, assuming some *a priori* knowledge about how likely each point of time is to become a change point. One of the most robust results is obtained under the assumption that the underlying distribution of observations may switch from f_0 to f_1 only as a result of some "shock" the time to which is exponentially distributed (i.e., the on-target lifetime has a constant hazard rate). Thus, we can assume that for some small probability p, the process develops as follows: If the current density is f_0, at the time the next observation is taken it will stay the same with probability $(1 - p)$ or switch to f_1 with probability p. After the change the observations will be generated by f_1 until the change is detected (see [25, 33]).

In this "random shock"-type model, it is reasonable to adopt a control scheme which calls for a signal at time n once the posterior probability that the change occurred before n (given the data) exceeds some prescribed limit Π; i.e., the signal is triggered at time n if

$$\frac{C_n}{C_n + (1 - p)^n \prod_{j=1}^{n} f_0(x_j)} > \Pi, \tag{8}$$

where

$$C_n = \sum_{r=0}^{n-1} (1 - p)^r p \prod_{j=1}^{r} f_0(x_j) \prod_{j=r+1}^{n} f_1(x_j). \tag{9}$$

In other words, the scheme calls for an out-of-control signal at time n if

$$S_n = \frac{C_n}{p(1 - p)^n \prod_{j=1}^{n} f_0(x_j)} = (1 + S_{n-1}) \frac{f_1(x_n)}{(1 - p)f_0(x_n)} > \Pi/p(1 - \Pi). \tag{10}$$

This approach leads to an interesting (upper) control scheme in the normal case; by denoting $\tilde{S}_n = \Delta \log S_n$ [see (5)] we obtain the scheme with $\tilde{S}_n = -\infty$ and

$$\tilde{S}_n = \Delta \log(1 + e^{\tilde{S}_{n-1}/\Delta}) + [x_n - k - \Delta \log(1 - p)], \tag{11}$$

the structure of which looks very similar to that of the Page's scheme (6); in fact, when Δ tends to zero (i.e., the difference between on-target and off-target means becomes larger and larger with respect to σ) the scheme (11) turns into an upper Page's scheme.

3. Markov-type control schemes

As we have seen in the previous section, one can find some "good" (even optimal) control schemes of a Markov type; i.e., the value of a control scheme at any given moment of time depends only upon the new information which has arrived at this moment of time and the previous value of the scheme. To put this more formally, let us suppose that we are interested in detecting change of a certain process parameter *upward.* The Markov-type scheme to achieve this goal can be, in general, viewed as an operator transforming the original sequence of observations $\{x_i\}$ into the sequence of values of the scheme $\{S_i \geq 0\}$ of the type

$$S_i = g(S_{i-1}, x_i, \tilde{\alpha}), \tag{12}$$

where g is a monotonically increasing function in its first and (in many cases, but not necessarily) second argument, and $\tilde{\alpha}$ is the vector of scheme parameters. The signal level $h > 0$ is chosen so as to achieve the desired trade-off between the desired degree of protection against false alarms and sensitivity; the signal is triggered at time n if $S_n > h$. The initial value of the scheme, $0 \leq S_0 \leq h$ (headstart), is usually zero; however, in some cases a nonzero value is used to implement the so-called FIR (fast initial response) feature (see [34, 35]). It is clear that at each point of time i, the value S_i summarizes our evidence in support of the hypothesis that the process is presently out of control. The signal level, therefore, reflects the degree of accumulation of information in support of this hypothesis that we are able to tolerate; one can see that selecting $h = 0$ prevents us from any accumulation of evidence, i.e., turns our control scheme into a pure Shewhart scheme in which a signal is triggered on the basis of the last observation only.

The function g is chosen in such a way that the control scheme is a supermartingale when the process is in control (i.e., the sequence possesses a sort of "anchor" which presses it to zero and thus keeps it from drifting away toward the signal level) and a submartingale when it is out of control; i.e., in this case the evidence supporting the out-of-control hypothesis is able to outweigh the "anchor" and cause the scheme to "float up" and signal. This property explains the rationale for using a positive headstart to implement the FIR feature, which provides an instrument for detecting *initially present* out-of-control conditions earlier than similar conditions occurring later. Indeed, when the process is on target, the scheme will be (most likely) brought to the vicinity of zero by the anchoring mechanism, so that in this case the expected effect of the headstart is minimal; otherwise, however, the out-of-control signal will be triggered much sooner.

Examples of the upper Markov-type scheme include the "reflected" version of the geometric moving average scheme,

$$S_0 = \text{fixed}, \qquad S_i = \{\alpha S_{i-1} + (1 - \alpha)x_i\}^+, \tag{13}$$

the Girshick-Rubin scheme (11) and the Page's scheme [(4), (6)]. In the latter cases one can see that the role of the "anchor" is played by the reference value k, which is usually chosen to be midway between the "acceptable" and "unacceptable" levels of the controlled process.

Because of the process of accumulation of evidence associated with every non-Shewhart control scheme, the latter always has some degree of "inertia" which may delay its reaction with respect to sharp changes in the process level. To remove some of this inertia, one can supplement the scheme with a Shewhart control limit c, i.e., introduce a supplementary signal criterion which calls for an immediate signal at time i if $x_i > c$. Clearly, this action does not affect the Markovian property of the scheme.

In a similar way, we can define a *lower* Markov-type scheme in order to detect changes downward in the level of the controlled parameter. For example, if the parameter of interest is the mean of the sequence of observations $\{x_i\}$, it is natural to select some *upper* scheme and apply it to the sequence of reflected observations $\{-x_i\}$. For example, the lower Page's scheme for controlling the mean can be defined in terms of a signal level $h^- \geq 0$, reference value k^- [clearly, $(-k^-)$ should be chosen close to midway between the acceptable and (lower) unacceptable process level], and a headstart $0 \leq S_0^- \leq h^-$,

$$S_i^- = [S_{i-1}^- + (-x_i - k^-)]^+, \qquad i = 1, 2, \cdots, \tag{14}$$

signal if $S_i^- > h^-$.

One of the most attractive properties of the class of Markov-type schemes is that they are relatively easily "analyzable." Among the various approaches to the problem of analysis we can mention the method of integral equations originally suggested by Page [29], the method of systems of linear algebraic equations [36], the direct approach method [37], Brownian motion approximations [2, 3], and the method of Markov chains [38]. From the author's experience, the latter approach seems to be the simplest and the most efficient [5]. It is based on discretizing the interval $(0, h)$ into d parts and then, at each step, rounding the value of a control scheme to the center of the appropriate discretization interval. By doing this, we essentially replace our control scheme trajectory with its discretized version, and therefore turn the scheme into a simple Markov chain with $(d + 1)$ states, the last one being the absorbing state corresponding to a signal; the run length behavior can then be obtained in a relatively straightforward way from the associated transition matrix. The efficiency of this approach is primarily related to the fact that there is typically no need for high levels of discretization to obtain good results; the levels of magnitude $d = 30$ are usually sufficient for most practical purposes [1, 5]. Indeed, when the process is out of control, the run length characteristics depend primarily on the magnitude of drift toward the signal level rather than on the value of d. On the other hand, when the process is on-target, one can expect the approximation to be good because of the compensation of roundoff errors.

One can also expect the Markov-type schemes to be relatively easily "designable." For a class of Cusum-Shewhart schemes, the problem of design was considered in [4] and [5]. Solving this problem usually requires performing a repetitive analysis of a sequence of schemes with fixed parameters. Since such analysis is typically associated with an extensive computational effort, the number of steps needed to complete the design could be substantially reduced by the availability of procedures for efficient sensitivity analysis both by scheme and distribution parameters. For the class of Cusum-Shewhart schemes such procedures were developed in [39]; some of the results can be generalized for more general Markov-type schemes.

Now consider the situation in which one is interested in detecting rapidly both types of shift of the process from its on-target level. Such two-sided control can be achieved in one of two ways. The first type of two-sided scheme calls for separate design of an upper and a lower one-sided scheme, and then for running them in parallel. The second type defines a single scheme which, if the process is in control, is supposed to stay within certain prescribed control bounds. The usual geometric moving average scheme (2) is probably the most frequently used representative of this class of two-sided schemes (the "Cusum" analog of such a type of scheme was recently considered by Crossier [40]). The main benefit of using a two-sided scheme based on a single sequence is related to the fact that the latter can usually be defined in such a way that its value at any given moment of time provides an estimate of the current process level. (We must mention, however, that some schemes of the first type may be as good in this respect; for example, the V-mask version of a Cusum-Shewhart control scheme provides a very simple and efficient way of estimating the current process mean [6].)

On the other hand, however, two-sided schemes of the first type have a much better resolution with respect to the "worst case" scenario; i.e., they are better tuned to respond to most recent events. Indeed, consider the case in which the process level is at the lower bound of the target region, and then shifts upward. The two-sided scheme based on a single sequence is then likely to be near its lower signal level at the moment the shift occurs. In such a case it must waste some time in first coming back to the center line; only then can it proceed further toward the upper signal level. In comparison, schemes of the first type have an "always ready" upper scheme which never drops below zero and, therefore, in situations as described above, is able to signal much earlier. Moreover, schemes of the first type are less likely to produce a signal indicating change of the process level in one direction when the actual change is in the opposite direction.

Except for this worst-case comparison, the two-sided schemes of both types typically show a roughly similar performance. From the point of view of analysis and design, the schemes of the first type usually allow more flexibility and in some cases (e.g., for Cusum-Shewhart schemes, see [32]) require a smaller computational effort, since the analysis of such two-sided schemes can be decomposed into a separate analysis of upper and lower schemes. It is the opinion of the author that in most cases one will be more interested in schemes with better worst-case sensitivity, i.e., two-sided schemes which represent a combination of an upper and a lower scheme run in parallel.

References

1. E. Yashchin, "Package CONTRD for Design, Analysis and Running of Cusum-Shewhart Control Schemes," *Research Report RC-11880*, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 1986.
2. R. A. Johnson and M. Bagshaw, "The Effect of Serial

Correlation on the Performance of CUSUM Tests," *Technometrics* **16,** 103–122 (1974).
3. M. Bagshaw and R. A. Johnson, "The Effect of Serial Correlation on the Performance of CUSUM TESTS II," *Technometrics* **17,** 73–80 (1975).
4. W. H. Woodall, "The Design of Cusum Quality Control Charts," *J. Qual. Technol.* **18,** 99–103 (1986).
5. E. Yashchin, "On the Analysis and Design of Cusum-Shewhart Control Schemes," *IBM J. Res. Develop.* **29,** 377–391 (1985).
6. G. Barnard, "Control Charts and Stochastic Processes," *J. Roy. Statist. Soc. (B)* **21,** 239–257 (1959).
7. H. Chernoff and S. Zacks, "Estimating the Current Mean of the Normal Distribution Which Is Subject to Changes in Time," *Ann. Math. Statist.* **35,** 999–1018 (1964).
8. D. Hinkley, "Inference About the Change-Point in a Sequence of Random Variables," *Biometrika* **57,** 1–17 (1970).
9. D. Hinkley, "Inference About the Change-Point from Cumulative Sum Tests," *Biometrika* **58,** 509–523 (1971).
10. D. Siegmund, "Estimation Following Sequential Tests," *Biometrika* **65,** 341–349 (1978).
11. B. James, K. James, and D. Siegmund, "Tests for a Change-Point," *Technical Report No. 35,* Stanford University, Stanford, CA, 1985.
12. G. Box and G. Jenkins, "Some Statistical Aspects of Adaptive Optimization and Control," *J. Roy. Statist. Soc. (B)* **24,** 297–331 (1962).
13. G. Box and G. Jenkins, "Further Contributions to Adaptive Quality Control: Simultaneous Estimation of Dynamics I: Non-Zero Costs," *Bull. Int. Statist. Inst.* **34,** 943–974 (1963).
14. S. Crowder, "Kalman Filtering and Statistical Process Control," Ph.D. Thesis, Iowa State University, Ames, Iowa, 1986.
15. A. J. Duncan, "The Economic Design of $\bar{X}$-Charts Used to Maintain Current Control of the Process," *J. Amer. Statist. Assoc.* **51,** 228–242 (1956).
16. H. Taylor, "The Economic Design of Cumulative Sum Control Charts," *Technometrics* **10,** 479–488 (1968).
17. T. Lorenzen and L. Vance, "The Economic Design of Control Charts: A Unified Approach," *Technometrics* **28,** 3–10 (1986).
18. A. J. Duncan, *Quality Control and Industrial Statistics,* R. D. Irwin, Inc., Homewood, IL, 1974.
19. G. Box and G. Jenkins, *Time Series Analysis—Forecasting and Control,* Holden-Day, San Francisco, 1970.
20. D. S. van Dobben de Bruyn, *Cumulative Sum Tests: Theory and Practice,* Hafner Publishing Co., New York, 1968.
21. R. Woodward and P. L. Goldsmith, *Cumulative Sum Techniques,* ICI Monograph No. 3, Oliver and Boyd, London, 1964.
22. A. Bissell, "Cusum Techniques for Quality Control," *Appl. Statist.* **18,** 1–30 (1969).
23. British Standards Institution, *Guide to Data Analysis and Quality Control Using Cusum Techniques. Parts 1–4* (1980–1983).
24. S. Roberts, "Control Chart Tests Based on Geometric Moving Averages," *Technometrics* **1,** 239–250 (1959).
25. Roberts, S. "A Comparison of Some Control Chart Procedures," *Technometrics* **8,** 411–430 (1966).
26. L. Nelson, "The Deceptiveness of Moving Averages," *J. Qual. Technol.* **15,** 99–100 (1983).
27. J. Bather, "Control Charts and Minimization of Costs," *J. Roy. Statist. Soc. (B)* **25,** 49–80 (1963).
28. J. S. Hunter, "The Exponentially Weighted Moving Average," *J. Qual. Technol.* **18,** 203–210 (1986).
29. E. Page, "Continuous Inspection Schemes," *Biometrika* **41,** 100–115 (1954).
30. R. Khan, "On Cumulative Sum Procedure and the SPRT with Applications," *J. Roy. Statist. Soc. (B)* **46,** 79–85 (1984).
31. G. Lorden, "Procedures for Reacting to a Change in Distribution," *Ann. Math. Statist.* **42,** 1897–1908 (1971).
32. E. Yashchin, "On a Unified Approach to the Analysis of Two-Sided Cumulative Sum Schemes with Headstarts," *Adv. Appl. Prob.* **17,** 562–593 (1985).
33. M. Girshick and H. Rubin, "A Bayes Approach to a Quality Control Model," *Ann. Math. Statist.* **23,** 114–125 (1952).
34. J. M. Lucas and R. B. Crosier, "Fast Initial Response for Cusum Quality Control Schemes: Give Your Cusum a Head Start," *Technometrics* **24,** 199–205 (1982).
35. J. M. Lucas, "Combined Shewhart-Cusum Quality Control Schemes," *J. Qual. Technol.* **14,** No. 2, 51–59 (1982).
36. A. L. Goel and S. M. Wu, "Determination of ARL and a Contour Nomogram for Cusum Charts to Control Normal Mean," *Technometrics* **13,** 221–230 (1971).
37. W. H. Woodall, "The Distribution of the Run Length of One-Sided Cusum Procedures for Continuous Random Variables," *Technometrics* **25,** 295–300 (1983).
38. D. Brook and D. A. Evans, "An Approach to the Probability Distribution of Cusum Run Length," *Biometrika* **59,** 539–549 (1972).
39. E. Yashchin, "Sensitivity Analysis of Cusum-Shewhart Control Schemes," *Research Report RC-12078,* IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 1986.
40. R. Crossier, "A New Two-sided Cumulative Sum Quality Control Scheme," *Technometrics* **28,** 183–194 (1986).

Received December 18, 1986; accepted for publication January 16, 1987

Emmanuel Yashchin *IBM Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598.* Dr. Yashchin joined IBM at the Thomas J. Watson Research Center in 1983. His research interests include quality control, applied statistics, extreme value theory, and operations research. He received a Diploma in applied mathematics from Vilnius State University (U.S.S.R.) in 1974 and an M.Sc. in operations research and a D.Sc. in statistics from the Technion–Israel Institute of Technology in 1977 and 1981, respectively. In 1982 he was a Visiting Assistant Professor at Iowa State University, Ames. He is a member of the American Society for Quality Control, the American Statistical Association, and the Institute of Mathematical Statistics.