A RULE OF THUMB: RUN LENGTHS TO FALSE ALARM OF MANY TYPES OF CONTROL CHARTS RUN IN PARALLEL ON DEPENDENT STREAMS ARE ASYMPOTICALLY INDEPENDENT

BY MOSHE POLLAK

The Hebrew University of Jerusalem
and
Israel Central Bureau of Statistics

Consider a process that produces a series of independent identically distributed vectors. A change in an underlying state may become manifest in a modification of one or more of the marginal distributions. Often, the dependence structure between coordinates is unknown, impeding surveillance based on the joint distribution. A popular approach is to construct control charts for each coordinate separately and raise an alarm the first time any (or some) of the control charts signals. The difficulty is obtaining an expression for the overall average run length to false alarm (ARL2FA).

We argue that despite the dependence structure, when the process is in control, for large ARL’s to false alarm, run lengths of many types of control charts run in parallel are asymptotically independent. Furthermore, often, in-control run lengths are asymptotically exponentially distributed, enabling uncomplicated asymptotic expressions for the ARL2FA.

We prove this assertion for certain Cusum- and Shiryaev-Roberts-type control charts and illustrate it by simulations.

1. Introduction. In many applications, observations are multivariate, with the marginal behavior of each of \( p \) coordinates governed by its particular distribution. For example, vital statistics of a sequence of newborn infants may be monitored for public health purposes, where the distributions of the various measurements may be a mixed bag; the daily change in the price of a stock or a portfolio may be monitored for a change of volatility, where observations may be normally distributed; weekly traffic accidents on an assortment of roads may be monitored for an increase in mean, where observations may be Poisson-distributed; a behavioral change of measure-

*This work was supported by Grant No. 1450/13 of the Israel Science Foundation and a grant of the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem.

MSC 2010 subject classifications: Primary 62L10; secondary 94A13

Keywords and phrases: average run length, Cusum, Shiryaev-Roberts, exponential distribution, p-value
ments taken on a succession of articles emanating from a production line may be indicative of a deterioration of a machine, where the distributions of the various measurements may not even belong to a parametric family. A significant change in the state of affairs may express itself by a change of the stochastic behavior of any/some/all of the coordinates.

In many applications the vector observations are independent but the coordinates are not. Whereas the marginal distributions of the coordinates may belong to a known parametric family, typically the dependence structure is nebulous, impeding the construction of an efficient surveillance scheme, based on the joint distribution of the coordinates.

Many approaches to this problem have been proposed. Most have suggested combining the coordinates in one form or another into one or two statistics and applying a univariate control chart to each of the resulting sequences of statistics. For a review see Epprecht (2015). Shewhart-type control charts seem to be the choice of most of these.

An intuitively attractive approach is to construct control charts separately for each stream and stop the first time any of them (or some of them) calls for raising an alarm. In particular, this approach facilitates straightforward identification of the stream(s) where a change has taken place. A number of authors have considered this approach (cf. Meneces et al., 2008; Mei, 2010) and compared it to alternative methods. It is this approach that is the focus of the present article.

The inherent methodological problem that behooves all approaches to address is the ramifications of dependence between streams. Since this dependence could be thought to bring about dependence between the control statistics, the initial technical problem that needs to be addressed is the evaluation of the overall average run length to false alarm (ARL2FA). Some of the proposals described in Epprecht (2015) neglect this need; others (such as Meneces et al., 2008) try to account for the dependence. Methods that ignore the dependence may have an ARL2FA that is quite different from the nominal one.

We find that despite the dependence between streams, the in-control run lengths of various control charts applied individually to each stream behave asymptotically (as ARL2FA→∞) as if they are independent, a result that enables an asymptotic approximation to the overall ARL2FA. The intuition behind this is the following. Heuristically, a false alarm arises when “recent” observations aggregate in a way that gives credence to the impression that the process is out of control. When such a “spurt” takes place in one control sequence, it is plausible that corresponding observations in a parallel sequence may exhibit a “blip”. However, if the dependence between coor-
dinates is not too strong, the “blip” will most likely be weaker than the “spurt” (recall regression towards the mean), and will not be strong enough to signal a false alarm in the parallel sequence, too. A false alarm in the parallel sequence will take (or shall have taken) place at a distant point in time, rendering the stopping times approximately independent. Since the in-control run length of a Cusum or a Shiryaev-Roberts control chart is asymptotically exponentially distributed (cf. Pollak and Tartakovsky, 2008; Yakir, 1995, 1998), the overall average run length to false alarm of a policy of stopping after \( k \) of \( p \) control charts have signalled an alarm can be readily approximated. Hence, the practical aspect of our results is a contribution to the arsenal of methods of monitoring dependent streams.

Here we spell things out explicitly for (generalized) Cusum and Shiryaev-Roberts control charts, although it is easy to conjecture that our results hold for other control charts as well. We do not make comparisons to other approaches, although the application of Cusum or Shiryaev-Roberts procedures promises faster detection than Shewhart charts. Comparisons are of course begged for, but prior to making them one needs a handle on the ARL2FA, which is the heart of this paper. Suffice it to say that in a narrower context, application of separate Cusums has been shown to have merit: Mei (2010) proposed a surveillance method (based on the sum of Cusum statistics) in the case that all of the coordinates are independent, and compared his method with stopping after \( k \) of \( p \) Cusums defined separately on each stream have signaled. Mei’s results indicate that when many streams are affected by a change, his method may be superior; when few streams are affected then \( \{ k \text{ of } p \} \) may be a preferred method. The implication of our results is that since (when streams are dependent, asymptotically) run lengths to false alarms are independent, Mei’s insight may be extendable to the dependent streams case, too.

The paper is organized as follows. In Section 2 we present our main results. We illustrate them in Section 3 with a simulation study. In Section 4 we present a number of remarks. In Section 5 we give a short sketch of the idea behind the proofs. We relegate proofs to Supplementary Material; a formal proof is provided for the case where the sequential vector observations are independent and in-control observations are identically distributed with known marginal distributions. For ease of exposition, we deal formally with the case \( p = 2 \); the extension to \( p > 2 \) is straightforward (see Remark 1).

2. Main results. Let \( \{X_i, Y_i\} \) be a sequence of independent vectors. We assume that the in-control and out-of control marginal distributions of \( \{X_i\} \) and \( \{Y_i\} \) belong to exponential families. Specifically, the marginal den-
sity of $X$ (with respect to a sigma-finite measure $\mu_X$) is $f_0^X(x) = \exp(\theta x - \Psi_X(\theta))$; $-\infty < \theta < \infty$, where the difference between in-control and out-of-control manifests itself in a change in the parameter $\theta$ and the marginal density of $Y$ (with respect to a sigma-finite measure $\mu_Y$) is $f_0^Y(y) = \exp(\lambda y - \Psi_Y(\lambda))$; $-\infty < \lambda < \infty$, where the difference between in-control and out-of-control manifests itself in a change in the parameter $\lambda$. (By abuse of notation, we use $f_0, f_1, f_\theta, f_\lambda$ to denote both univariate and multivariate (joint) densities.) Without loss of generality, the in-control parameters of $X_i$ and $Y_i$ are $\theta = 0$ and $\lambda = 0$ respectively and $0 = \Psi_X(0) = \Psi'_X(0)$, $0 = \Psi_Y(0) = \Psi'_Y(0)$. We assume that $\Psi'_X(\theta) \neq 0, \Psi'_Y(\lambda) \neq 0$ for all $\theta, \lambda$. Thus, denoting $H^X_0 : \theta = 0$ and $H^Y_0 : \lambda = 0$ obtains log-likelihood ratios $Z^X(\theta) = \theta X - \Psi_X(\theta)$ and $Z^Y(\lambda) = \lambda Y - \Psi_Y(\lambda)$. Let $G_X, G_Y$ be prior distributions on $\{\theta \neq 0\}, \{\lambda \neq 0\}$ respectively. To avoid cumbersome proofs, we assume that $\theta$ and $\lambda$ are such that $Z^X(\theta)$ and $Z^Y(\lambda)$ are non-lattice. For technical reasons, if $G_X, G_Y$ are not concentrated on a finite set of atoms, we assume that the exponential families are strongly nonlattice (in the sense of Stone, 1965).

We denote: $H^X_0 : \theta = 0$, $H^Y_0 : \lambda = 0$, $H^X_1 : \theta \sim G_X, H^Y_1 : \lambda \sim G_Y$.

We assume that the $\{H^X_0, H^X_1\}$-distribution of $Z^X(\theta)$ conditional on $Z^Y(\lambda)$ is not degenerate and neither is the $\{H^Y_0, H^Y_1\}$-distribution of $Z^Y(\lambda)$ conditional on $Z^X(\theta)$, for all $\theta, \lambda$. (If correlation$_{H^X_0, H^Y_0}(Z^X, Z^Y) = -1$, this may be somewhat relaxed, though care must be taken. See Remarks 2 and 3 in Section 4.) We also assume that whatever the joint density of $X, Y$ be, $E(Z^X(\theta)|Z^Y(\lambda)), E(Z^Y(\lambda)|Z^X(\theta)), Var(Z^X(\theta)|Z^Y(\lambda)), Var(Z^Y(\lambda)|Z^X(\theta))$ are continuous in $\lambda \in support(G_Y), \theta \in support(G_X)$.

Denote the (separate) Cusum sequences by

$$W_n^X = \max_{k=1, \ldots, n} \int e^{\sum_{i=k}^n Z^X_i(\theta)} dG_X(\theta)$$

and

$$W_n^Y = \max_{k=1, \ldots, n} \int e^{\sum_{i=k}^n Z^Y_i(\lambda)} dG_Y(\lambda)$$

and the Cusum stopping times by

$$N_X = \min\{n|W_n^X > e^{by}\} \quad \text{and} \quad N_Y = \min\{n|W_n^Y > e^{by}\}.$$ 

Denote the (separate) Shiryaev-Roberts sequences by

$$R_n^X = \int \sum_{k=1}^n e^{\sum_{i=k}^n Z^X_i(\theta)} dG_X(\theta) \quad \text{and} \quad R_n^Y = \int \sum_{k=1}^n e^{\sum_{i=k}^n Z^Y_i(\lambda)} dG_Y(\lambda)$$
CONTROL CHARTS ON DEPENDENT STREAMS

and the Shiryaev-Roberts stopping times by

\[ M_X = \min\{n | R_n^X > e^{b_X}\} \quad \text{and} \quad M_Y = \min\{n | R_n^Y > e^{b_Y}\} \]

When both \( G_X, G_Y \) are concentrated at an atom, \( N_X, N_Y, M_X, M_Y \) are classical (simple) Cusum and Shiryaev-Roberts methods respectively, where either the post-change parameter is known or a (single) representative (i.e. the atom) is taken for the post-change parameter. When \( G_X, G_Y \) are continuous, the Cusum and Shiryaev-Roberts procedures are geared to the more complex case where the post-change parameter is unknown and a prior is taken over possible (or reasonable) post-change parameter values. We assume (in the continuous case) that on their supports \( G_X, G_Y \) have continuous positive densities \( g_X, g_Y \) respectively.

We assume that there exists a constant \( 0 < \zeta < \infty \) so that \( |b_X - b_Y| < \zeta \) and that for all \( \theta, \lambda \) neither the \( \{H_0^X, H_0^Y\} \)-distribution of \( Z^X(\theta) \) conditional on \( Z^Y(\lambda) \) nor that of the \( \{H_0^X, H_0^Y\} \)-distribution of \( Z^Y(\lambda) \) conditional on \( Z^X(\theta) \) is degenerate.

**Theorem 1** As \( b_X, b_Y \to \infty \), the pair \( (N_X/E_{H_0^X}(N_X), N_Y/E_{H_0^Y}(N_Y)) \) converges in distribution to \( (E_X, E_Y) \) where \( E_X \) and \( E_Y \) are independent Exponential(1)-distributed random variables.

**Theorem 2** As \( b_X, b_Y \to \infty \), the pair \( (M_X/E_{H_0^X}(M_X), M_Y/E_{H_0^Y}(M_Y)) \) converges in distribution to \( (E_X, E_Y) \) where \( E_X \) and \( E_Y \) are independent Exponential(1)-distributed random variables.

Consequently, the ARL2FA of a rule that raises an alarm the first time one of the two stopping times signals is

\[
\frac{1}{E_{H_0^X}^X(N_X)} \times (1 + o(1))
\]

and

\[
\frac{1}{E_{H_0^X}^X(M_X)} \times (1 + o(1))
\]

for \( N_X \) and \( M_X \) respectively, and the ARL2FA of a rule that raises an alarm after both stopping times signal is respectively

\[
\frac{[E_{H_0^X}^X(N_X)]^2 + E_{H_0^X}^X(N_X)E_{H_0^X}^Y(N_Y) + [E_{H_0^Y}^Y(N_Y)]^2}{E_{H_0^X}^X(N_X) + E_{H_0^Y}^Y(N_Y)} \times (1 + o(1))
\]

and

\[
\frac{[E_{H_0^X}^X(M_X)]^2 + E_{H_0^X}^X(M_X)E_{H_0^Y}^Y(M_Y) + [E_{H_0^Y}^Y(M_Y)]^2}{E_{H_0^X}^X(M_X) + E_{H_0^Y}^Y(M_Y)} \times (1 + o(1)).
\]

Two other results that are used as lemmas in the proof of the theorems may be of interest in their own right and are presented here. Lemma 3 is a probabilistic statement. Lemma 4 is a blueprint of the anatomy of a false
Lemma 3. Suppose \( G_X \) and \( G_Y \) are concentrated on \( \theta \) and on \( \lambda \) respectively. Let \( H^X_1 : \{ X_i \sim P^X_\theta \} \) and \( H^Y_1 : \{ Y_i \sim P^Y_\lambda \} \). Then
\[
E_{H^X_1} (Z^X_i (\theta)) < E_{H^Y_1} (Z^Y_i (\lambda)) \quad \text{and} \quad E_{H^X_1} (Z^Y_i (\lambda)) < E_{H^X_1} (Z^X_i (\theta)).
\]

For the second result, define and denote:
\[
\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \hat{\theta}_n \quad \tau^X_b = \min \{ n | \int \exp(\sum_{i=1}^n Z^X_i (\theta)) dG_X (\theta) \geq \exp(b) \}
\]
\[
\gamma_X (\theta) = \lim_{b \to \infty} E\theta \exp \{ - (\sum_{i=1}^n Z^X_i (\theta) - b) \} \quad \text{for} \quad G_X \quad \text{concentrated on} \quad \{ \theta \}
\]
\[
\gamma_{G,X} = \int \gamma_X (\theta) dG_X (\theta) \quad \text{for general} \quad G_X
\]

Regard the equation \( b_X = \int e^{h(t) - \theta \Psi (\theta)} dG_X (t) \). If the support of \( G_X \) contains only nonnegative (or only nonpositive) values then \( h(t) \) is unique; if the support contains both positive and negative values then there are two solutions to \( h(t) \), one positive, one negative. In any case, \( h(t) \) is a boundary that \( \tau^X_b \) must cross for \( \tau^X_b \) to be finite. Denote the excess of \( \sum_{i=1}^n X_i \) over/under the boundary by \( \varsigma \). Since \( O_{\tau^X_b} \) is stochastically bounded, so is \( \varsigma \).

Lemma 5. Conditional on \( \{ \tau^X_b < \infty \} \):

a) The \( P_\theta \)-stochastic behavior of the trajectory \( X_1, X_2, \ldots, X_{\tau^X_b} \) can be obtained by first randomly procuring \( \theta \) and then obtaining observations \( X_1, X_2, \ldots, X_{\tau^X_b-1}, (X_{\tau^X_b} - \varsigma), \varsigma \). The stochastic behavior of the observations \( X_1, X_2, \ldots, X_{\tau^X_b-1}, (X_{\tau^X_b} - \varsigma) \) is \( P_\theta \).

b) Given \( \theta \), \( \theta_{\tau^X_b} \to \theta \) in \( P_\theta \)-probability as \( b \to \infty \).

c) As \( b \to \infty \), the asymptotic distribution of \( \theta \) has density \( \gamma_X (\theta) dG (\theta) / \gamma_{G,X} \) and the convergence is uniform on bounded intervals of \( \theta \).

3. Simulations. The following is a report of simulations of Shiryaev-Roberts statistics and stopping times. When assessing the practical value of our results by the simulations, it should be borne in mind that if \( ARL_{2 \text{FA}} = B \) then the average speed of detection (of a true change) is proportional to \( \log B \) (asymptotically, as \( B \to \infty \)). Thus, when \( B \) is large, even a moderate discrepancy between the nominal \( ARL_{2 \text{FA}} \) and the true value may not have a marked effect on the speed of detection.

The simulation results reported in Table 1 are based on 1000 repetitions of Shiryaev-Roberts stopping times \( M_X, M_Y \) when the process is in control,
where $X$ and $Y$ are standard normal variables with correlation $\rho$ when the process is in control (IC). The control schemes are designed with out of control (OOC) parameters $\theta = \lambda = 1$. The simulations are reported for three cutoff levels $A = \exp(b_X) = \exp(b_Y) = 100, 500, 1000$ and six correlation values $\rho = .8, .6, .4, .2, -.4, -1$. In each case $E_{H_0^X H_0^Y}(M_X) = E_{H_0^X H_0^Y}(M_Y) (= 1.7845 \times A$ nominally, by Pollak’s (1987) renewal-theoretic approximation) and Theorem 2 implies that asymptotically the mean of $\min\{M_X, M_Y\}$ should be approximately half of the average of the means of $M_X$ and $M_Y$. Table 1 exhibits the dependence on $\rho$ and $A$ of the rate that the asymptotics kicks in: for a given value of $\rho$, the larger the cutoff level the better the approximation and for a given cutoff level, the larger the value of $\rho$ the worse the approximation. The approximation indicated by Theorem 1 seems to work reasonably well for standard ARL’s to false alarm if $\rho < .5$. The near equality of the means and the standard deviations is consistent with asymptotic exponentiality of the stopping times. The low correlations between $M_{\max}$ and $M_{\max} - M_{\min}$ and between $M_X, M_Y$ (when $\rho \leq .4$) are consistent with independence.

Note that when the correlations are negative, things behave as expected by Theorem 2 even if the cutoff level is low. Intuitively this makes sense; if the correlations are negative, high values of $X$ will go together with low values of $Y$, so if the values of the $X$’s are high (enough to signal an alarm) the $Y$’s will most likely not signal one. (See Remark 2 in the sequel for an extreme example of this.)

In Table 2 we describe what happens if each observation is a vector of five components, each of which has a standard normal distribution when the process is in control and each pair has correlation $\rho$. Suppose one is on the alert for an increase in the mean of the components and sets up a Shiryaev-Roberts control chart for each of them separately for a putative increase of one standard deviation. Suppose further that one is hesitant to raise an alarm caused by one component only, and prefers to raise an alarm only after two charts have signalled. In addition, suppose one wants an (overall) ARL2FA $\sim 741$ (as in the Shewhart control chart for a one-sided alternative). Theorem 2 (in a version extended to $p > 2$) implies that the five control charts are approximately independent and exponentially distributed. (In fact, none of the correlations between run lengths of different streams exceeded 0.1 as long as $\rho \leq 0.6$.) If a single control chart has ARL2FA=$\gamma$, then the average run length until the first of the five signals (when the process is in control, streams are independent and run lengths to false alarm are exponentially distributed) is $\gamma/5$ and the additional average run length until the next one signals is $\gamma/4$, so the overall ARL2FA is $0.45\gamma$. Applying Pollak’s
(1987) renewal-theoretic approximation to the ARL2FA, one gets (in this case) that for a single chart ARL2FA=1.7845A, where A is the Shiryaev-Roberts crossing boundary. Hence, if one desires an overall ARL2FA \( \sim 741 \), one should choose \( A = 741/(0.45 \times 1.7845) = 923 \).

The results described in Table 2 are based on 10,000 repetitions of the Shiryaev-Roberts stopping rule applied to each of the five streams. It is clear that the difference between the true and the nominal ARL2FA is insignificant when there exists a light positive correlation and even a moderate correlation does little harm, especially (as would be intuitively anticipated) since the ARL2FA is conservative.

Table 3 presents simulation results for a Poisson example, based on 1000 repetitions. The picture that is conveyed (in terms of the validity of the asymptotic formulae) is similar to the normal case.

The results described in Table 4 are based on 1000 repetitions of the Shiryaev-Roberts stopping rule applied to each of 100 streams, where for a given vector of observations the correlation between each pair of coordinates is \( \rho \) and where one stops after 10 streams have signalled. Again, each rule is designed (separately) to monitor a standard normal sequence for an increase of one standard deviation. The cutoff value \( A \) for a given nominal ARL2FA is again calculated via

\[
A = ARL2FA/(1.7845 \sum_{i=1}^{10}[1/(101 - i)]
\]

and the nominal standard deviation via

\[
SD = 1.7845A \sqrt{\sum_{i=1}^{10}[1/(101 - i)]^2}
\]

Evidently, things improve discernably as the nominal ARL2FA increases. It is interesting to note that the asymptotics for the SD kicks in somewhat more slowly than for the ARL2FA.

Intuitively, the case of equal correlations \( \rho \) between each pair of coordinates could be viewed as a “worst case scenario”; if some (or all) of the coordinate pairs have lesser correlation, the approximations would be expected to be better (as indicated by Tables 1, 3 and 4).

4. Remarks. 1. The multivariate case follows from the bivariate case. Suppose \( N_1, N_2, \ldots, N_p \) stopping times were run separately on each of \( p \) streams \( \{X_i^{(1)}\}, \{X_i^{(2)}\}, \ldots, \{X_i^{(p)}\} \). Setting up blocks of size \( \eta_{bX} \) and defining \( N_i^*, K_{X_i^*} \) as in Lemmas 8 and 9 in the Supplementary Material, the results of the paper imply that when everything is in control \( P(N_i \neq N_i^*) = o(1) \) and \( P(K_{X_i^*} = K_{X^*}) = o(1) \). The independence between blocks accounts for the asymptotic independence of the stopping times, which are exponentially distributed.

2. The two-sided (simple) Cusum control chart can be viewed as a special case of the minimum of two control charts (equivalent to \( X \sim f_0 \) pre-change with \( X \sim f_\theta \) or \( X \sim f_\lambda \) post-change, where \( \theta < 0 < \lambda \); here...
correlation\((Z'_X, Z'_Y) = -1\). Under a certain technical condition the equality in the ARL2FA following Theorems 1 and 2 is exact, without the \(o(1)\) piece. See Siegmund, 1985, page 28.

3. In Remark 2, the assumption that the distribution of \(Z^X\) conditional on \(Z^Y\) not be degenerate is obviously violated; nevertheless the result of Theorem 1 is valid. However, in general, if \(\text{correlation}(Z^X, Z^Y) = -1\), care must be taken, as the result may not be valid. For example, if \(X \sim N(0, 1)\) under \(H^X_0\) and \(Y = -X\), if \(G^X = G^Y\) are standard normal then \(\int e^{\sum_{i=k}^n Z^X_i(\theta)dG^X(\theta)}\) and \(\int e^{\sum_{i=k}^n Z^Y_i(\lambda)dG^Y(\lambda)}\) are identical.

4. In the problem of detection of a change in a normal mean, run lengths of parallel Shewhart charts (with similar ARL2FA) are obviously asymptotically independent. Although also in many other cases this will be true, this will not be true in general. For example, suppose \(X_i\) and \(Y_i\) are distributed \(\text{Cauchy}(0,1)\) when in control and \(\text{Cauchy}(1,1)\) when out of control, where \(Y_i = X_i\) if \(-2 < X_i < 2\) and otherwise \(Y_i \in \{(-\infty, -2] \cup [2, \infty)\}\) is independent of \(X_i\). Separate Shewhart charts based on the log-likelihood ratios \(Z^X_i = \log\left(\frac{1 + X_i^2}{1 + (X_i - 1)^2}\right)\) , \(Z^Y_i = \log\left(\frac{1 + Y_i^2}{1 + (Y_i - 1)^2}\right)\) that have the same ARL2FA will stop together if the ARL2FA is large enough.

5. In more complicated cases – such as when initial baseline parameters are unknown, and an invariance structure is exploited (cf. Yakir, 1998) – analogous results may be obtained. Intuitively, the reason for this is that asymptotically, it will take a long while for a false alarm to occur, and by then the unknown parameters are almost perfectly estimated. For example, consider the case where observations are distributed \(N(\mu, 1)\) when in control and \(N(\mu + \delta, 1)\) when out of control, where \(\delta\) is known (considered to be a representative of a post-change increase in mean) but \(\mu\) is unknown. An example of a procedure based on invariance calls for defining \(Y_i = X_i - X_1\) and monitoring the sequence \(Y_1, Y_2, \ldots\) by Cusum or Shiryaev-Roberts (cf. Pollak and Siegmund, 1991). A standard calculation obtains that the log-likelihood ratio of \(Y_2, Y_3, \ldots, Y_n\) for \(\nu = k\) vs. \(\nu = \infty\) (when translated back into the \(X's\), \(\nu\) is the first out-of-control observation and \(\nu = \infty\) means that the process is in control) is

\[
\delta \sum_{i=k}^n (X_i - \bar{X}_n) - \frac{1}{2} \delta^2 (n - k + 1) \frac{k - 1}{n}.
\]

Average run lengths do not depend on \(\mu\), so without loss of generality assume that \(\mu = 0\). In this case, when it is known that \(\mu = 0\), the log-likelihood
ratio of $X_1, X_2, \ldots, X_n$ for $\nu = k$ vs. $\nu = \infty$ is

$$\delta \sum_{i=k}^{n} X_i - \frac{1}{2} \delta^2 (n - k + 1).$$

Now argue:

a) If the cutoff boundary $A = e^b$ is very large, the probability that a false alarm will take place within the first $o(\sqrt{A})$ observations is negligible.

b) By virtue of Lemma 4, the “action” preceding a false alarm takes place in the last $O(\log A)$ observations.

c) Even after these $O(\log A)$ observations, $\bar{X}_n$ will be of order of magnitude $$(\log A) / n + O(1/n).$$

Hence, for $n > o(\sqrt{A})$ and $k > n - O(\log A)$, the difference $[(n - k + 1)[\delta^2 (n-k+1) - \delta \bar{X}_n]]$ between the two loglikelihood ratios is negligible, so with high probability they will raise a false alarm together.

Finally, to see why a) is true, recall that the Shiryaev-Roberts statistic $R_j$ has the property that $R_j - j$ is a zero-expectation martingale when the process is in control (IC), so $E_{IC} R_j = j$. Therefore

$$P_{IC} \{ \text{Shiryaev-Roberts stops before } m \} = P_{IC} \{ R_j > A \text{ for some } j < m \} \leq \sum_{j=1}^{m} P_{IC} \{ R_j > A \} \leq \sum_{j=1}^{m} E_{IC} (R_j) / A = \sum_{j=1}^{m} j / A = O(m^2 / A).$$


7. One of the deficiencies of changepoint detection methods is the lack of a p-value for an alarm being true. At least in principle, Lemma 4 provides a possible approach. Suppose $G_X$ is concentrated at $\theta$. At $N_X$, the last $b_X / I(\theta)$ observations should have mean $= \Psi'(\theta) \pm O_P(1/\sqrt{b_X})$ if the alarm is false. Therefore, if these observations have a significantly different mean (which is to be expected if a true change occurred, as one cannot predict exactly the value of a post-change parameter), it would be an indication that the alarm is not false.

8. The results have been formulated under the assumption that the vectors observed are iid when the process is in control. In fact, it suffices that the distributions of the marginals don’t change; if the correlations change the proofs remain valid (when the correlations are bounded away from 1).

9. The results hold also if $b_X, b_Y \to \infty$ and $|b_X - b_Y| \to \infty$. In that case, with probability $\to 1$, the stopping time with the lower threshold will stop before the other.
5. Sketch of proof. The idea behind the proof is to split the line into blocks of certain size, such that erasing the past at the start of each block has a negligible effect on the stopping times. What happens in one block thus becomes independent of that which takes place in other blocks. Thus the number of blocks until stopping is geometrically distributed, so that the stopping time is approximately exponential if the ARL2FA is large.

As intimated by the heuristics in the Introduction, false alarms occur because of a “spurt” in “recent” observations. (This is suggested by Lemma 4.) It was also surmised that a “spurt” in one control chart may cause at most a smaller “blip” in another chart. (This is indicated by Lemma 3.) The aforementioned blocks are set to be much larger than a “spurt”, but small enough that the likelihood of a “spurt” in a given block is small. Therefore, if a “spurt” in one stream takes place in a given block and the “blip” in another stream is not strong enough, a false alarm in this stream will take place at a different, independent, block.

The full proofs are relegated to the Supplementary Material.

References.


Table 1. Simulations: ARL2FA for monitoring \( \begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \)

when IC and \( E(X) = E(Y) = 1 \) when OOC by separate Shiryaev-Roberts control charts

<table>
<thead>
<tr>
<th>Threshold</th>
<th>( \rho = .8 )</th>
<th>( \rho = .6 )</th>
<th>( \rho = .4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_X \text{ mean} )</td>
<td>183 863 1704 172 890 1806 179 862 1793</td>
<td>177 855 1812 168 872 1749 170 864 1824</td>
<td>173 974 1777 162 920 1789 174 850 1809</td>
</tr>
<tr>
<td>( M_X \text{ sd} )</td>
<td>180 948 1697 183 892 1789 167 892 1785</td>
<td>177 921 1814 174 922 1745 185 843 1697</td>
<td>175 879 1807 161 906 1743 178 854 1748</td>
</tr>
<tr>
<td>( \min{M_X, M_Y} \text{ mean} )</td>
<td>119 567 1007 98 495 1002 100 457 922</td>
<td>114 596 1015 90 484 1028 96 440 957</td>
<td>.307 .266 .200 .115 .099 .136 .100 .020 .044</td>
</tr>
<tr>
<td>( \min{M_X, M_Y} \text{ sd} )</td>
<td>177 885 1812 168 872 1749 170 864 1824</td>
<td>173 974 1777 162 920 1789 174 850 1809</td>
<td>.8</td>
</tr>
</tbody>
</table>

Table 2. Simulations: monitoring \( \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{pmatrix} \right) \)

when IC and \( E(X_i) = 1, i = 1, \ldots, 5 \), when OOC by separate Shiryaev-Roberts control charts, stopping after two streams have signalled; nominal ARL2FA=741

<table>
<thead>
<tr>
<th>( \rho = \text{cor}(X_i, X_j) )</th>
<th>0</th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
<th>.6</th>
<th>.7</th>
<th>.8</th>
<th>.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARL2FA</td>
<td>745 744 760 772 787 802 833 876 964 1077 1661</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Simulations: ARL2FA of SR for $X \sim \text{Poisson}(9), Y \sim \text{Poisson}(4)$ when in control and $X \sim \text{Poisson}(12), Y \sim \text{Poisson}(6)$ when out of control

| Threshold($X$) | $A_X =$ | 100 | 500 | 1000 | 100 | 500 | 1000 | 100 | 500 | 1000 |
| Threshold($Y$) | $A_Y =$ | 70 | 350 | 700 | 70 | 350 | 700 | 70 | 350 | 700 |
| correlation($X,Y$) | $\rho = .2$ | $\rho = .4$ | $\rho = .6$ |
| $M_X$ mean | 178 | 945 | 1843 | 184 | 907 | 1879 | 178 | 883 | 1876 |
| $M_X$ sd | 169 | 903 | 1887 | 177 | 895 | 1809 | 170 | 870 | 1754 |
| $M_Y$ mean | 127 | 575 | 1230 | 131 | 606 | 1265 | 128 | 615 | 1309 |
| $M_Y$ sd | 115 | 572 | 1162 | 120 | 563 | 1240 | 123 | 606 | 1275 |
| $\min\{M_X, M_Y\}$ mean | 83 | 374 | 743 | 88 | 385 | 799 | 89 | 407 | 860 |
| $\min\{M_X, M_Y\}$ sd | 78 | 386 | 727 | 81 | 386 | 727 | 82 | 397 | 848 |
| $M_X, M_Y$ corr | .095 | .037 | .002 | .141 | .032 | .038 | .192 | .100 | .076 |
| $M_{max}, M_{max} - M_{min}$ corr | -.009 | -.050 | .002 | .045 | -.011 | .056 | -.016 | -.067 |

| $\frac{1}{\text{mean}(M_X)} + \frac{1}{\text{mean}(M_Y)}$ | 74 | 357 | 738 | 77 | 363 | 756 | 74 | 362 | 771 |

Table 4. Simulations: monitoring $E_{IC}(X_i) = 0$ vs. $E_{OOC}(X_i) = 1$, $i = 1, \ldots, 100$, where $\rho = \rho( (X_i)_j, (X_i)_k) = \text{correlation}( (X_i)_j, (X_i)_k)$, by separate Shiryaev-Roberts control charts, stopping after 10 of 100 streams have signaled, for various nominal values of ARL2FA

<table>
<thead>
<tr>
<th>nominal</th>
<th>simulated</th>
<th>$\rho = 0$</th>
<th>$\rho = .1$</th>
<th>$\rho = .2$</th>
<th>$\rho = .3$</th>
<th>$\rho = .4$</th>
<th>$\rho = .5$</th>
<th>$\rho = .6$</th>
<th>$\rho = .7$</th>
<th>$\rho = .8$</th>
<th>$\rho = .9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>741</td>
<td>ARL2FA</td>
<td>763</td>
<td>769</td>
<td>836</td>
<td>871</td>
<td>938</td>
<td>1063</td>
<td>1268</td>
<td>1624</td>
<td>2263</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>234</td>
<td>274</td>
<td>316</td>
<td>369</td>
<td>458</td>
<td>544</td>
<td>684</td>
<td>868</td>
<td>1252</td>
<td>1906</td>
</tr>
<tr>
<td>2500</td>
<td>ARL2FA</td>
<td>2534</td>
<td>2513</td>
<td>2593</td>
<td>2564</td>
<td>2753</td>
<td>2936</td>
<td>3186</td>
<td>3906</td>
<td>4854</td>
<td>6964</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>781</td>
<td>816</td>
<td>942</td>
<td>1026</td>
<td>1249</td>
<td>1564</td>
<td>1906</td>
<td>2583</td>
<td>3600</td>
<td>5727</td>
</tr>
<tr>
<td>25000</td>
<td>ARL2FA</td>
<td>24775</td>
<td>25057</td>
<td>25538</td>
<td>25310</td>
<td>25463</td>
<td>26133</td>
<td>29749</td>
<td>32743</td>
<td>39357</td>
<td>60209</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>7705</td>
<td>8302</td>
<td>8638</td>
<td>8773</td>
<td>10078</td>
<td>11927</td>
<td>14549</td>
<td>19502</td>
<td>26699</td>
<td>49032</td>
</tr>
</tbody>
</table>