The Annals of Statistics
2021, Vol. 49, No. 1, 557–567
https://doi.org/10.1214/20-AOS1968

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

BY MOSHE POLLAK

Statistics Department, The Hebrew University of Jerusalem, moshe.pollak@mail.huji.ac.il

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 ARLs 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 measurements 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);

Received January 2019; revised February 2020.
MSC2020 subject classifications. Primary 62L10; secondary 94A13.
Key words and phrases. Average run length, Cusum, Shiryaev–Roberts, exponential distribution, p-value.

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 $\to \infty$) 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 coordinates is not too strong, the "blip" will most likely be weaker than the "spurt" (recall regression toward 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 signaled 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 the 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 density of X (with respect to a sigma-finite measure μ_X) is $f_\theta^X(x) = \exp(\theta x - \Psi_X(\theta))$; $-\infty < \theta < \infty$, where the difference between in-control and out-of-control manifests itself in a change in the parameter θ and the marginal density of Y (with respect to a sigma-finite measure μ_Y) is $f_\lambda^Y(y) = \exp(\lambda y - \Psi_Y(\lambda))$; $-\infty < \lambda < \infty$, where the difference between in-control and out-of-control manifests itself in a change in the parameter λ. (By abuse of notation, we use f_0, f_1, f_θ, f_λ 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 θ, λ. Thus, denoting $H_0^X : \theta = 0$ and $H_0^Y : \lambda = 0$ 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 θ and λ are such that $Z^X(\theta)$ and $Z^Y(\lambda)$ are nonlattice. 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: $H_0^X : \theta = 0$, $H_0^Y : \lambda = 0$, $H_1^X : \theta \sim G_X$, $H_1^Y : \lambda \sim G_Y$.

We assume that the $\{H_0^X, H_0^Y\}$-distribution of $Z^X(\theta)$ conditional on $Z^Y(\lambda)$ is not degenerate and neither is the $\{H_0^X, H_0^Y\}$-distribution of $Z^Y(\lambda)$ conditional on $Z^X(\theta)$, for all θ, λ. (If $\text{correlation}_{H_0^X, H_0^Y}(Z^X, Z^Y) = -1$, 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, $E(Z^X(\theta)|Z^Y(\lambda))$, $E(Z^Y(\lambda)|Z^X(\theta))$, $\text{Var}(Z^X(\theta)|Z^Y(\lambda))$, $\text{Var}(Z^Y(\lambda)|Z^X(\theta))$ are continuous in $\lambda \in \text{support}(G_Y)$, $\theta \in \text{support}(G_X)$.

Denote the (separate) Cusum sequences by

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

and

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

and the Cusum stopping times by

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

Denote the (separate) Shiryaev–Roberts sequences by

$$R_n^X = \int \sum_{k=1}^n e^{\sum_{i=k}^n Z_i^X(\theta)}\, dG_X(\theta) \quad \text{and} \quad R_n^Y = \int \sum_{k=1}^n e^{\sum_{i=k}^n Z_i^Y(\lambda)}\, dG_Y(\lambda)$$

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 θ, λ 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 $(\mathcal{E}_X, \mathcal{E}_Y)$ where $\mathcal{E}_X$ and $\mathcal{E}_Y$ 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 $(\mathcal{E}_X, \mathcal{E}_Y)$ where $\mathcal{E}_X$ and $\mathcal{E}_Y$ 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

$$\frac{1}{\frac{1}{E_{H_0^X}(N_X)} + \frac{1}{E_{H_0^Y}(N_Y)}} \times \big(1 + o(1)\big)$$

and

$$\frac{1}{\frac{1}{E_{H_0^X}(M_X)} + \frac{1}{E_{H_0^Y}(M_Y)}} \times \big(1 + o(1)\big)$$

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

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

and

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

Two other results that are used as lemmas in the proof of the theorems (they appear and are proved in the Supplementary Material) may be of interest in their own right and are presented here. Lemma 4 is a probabilistic statement that can be interpreted as a generalization of regression toward the mean. Lemma 5 is a blueprint of the anatomy of a false alarm. Other lemmas and proofs are relegated to the Supplementary Material (Pollak (2020)).

LEMMA 4. *Suppose G_X and G_Y are concentrated on θ and on λ, respectively. Let H_1^X : $\{X_i \sim P_\theta^X\}$ and $H_1^Y : \{Y_i \sim P_\lambda^Y\}$. Then*

$$E_{H_1^Y}\big(Z_i^X(\theta)\big) < E_{H_1^Y}\big(Z_i^Y(\lambda)\big) \quad \textit{and} \quad E_{H_1^X}\big(Z_i^Y(\lambda)\big) < E_{H_1^X}\big(Z_i^X(\theta)\big).$$

For the second result, define and denote

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\hat{\theta}_n \text{ via } \bar{X}_n = \Psi'_X(\hat{\theta}_n),$$

$$\tau_b^X = \min\left\{n \,\middle|\, \int \exp\left(\sum_{i=1}^{n} Z_i^X(\theta)\right) dG_X(\theta) \geq \exp(b)\right\},$$

$$\gamma_X(\theta) = \lim_{b\to\infty} E_\theta \exp\Biggl\{-\Biggl(\sum_{i=1}^{\tau_b^X} Z_i^X(\theta) - b\Biggr)\Biggr\} \quad \text{for } G_X \text{ concentrated on } \{\theta\},$$

$$\gamma_{G_X} = \int \gamma_X(\theta)\, dG_X(\theta) \quad \text{for general } G_X.$$

Regard the equation $b_X = \int e^{\theta h(t) - t\Psi(\theta)}\, dG_X(t)$. 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 $\sum_{i=1}^n X_i$ must cross for τ_b^X to be finite. Denote the excess of $\sum_{i=1}^{\tau_b^X} X_i$ over/under the boundary by ς. Since $O_{\tau_b^X}^X$ is stochastically bounded, so is ς.

LEMMA 5. *Conditional on* $\{\tau_b^X < \infty\}$:

(a) *The* P_0*-stochastic behavior of the trajectory* $X_1, X_2, \dots, X_{\tau_b^X}$ *can be obtained by first randomly procuring* θ *and then obtaining observations* $X_1, X_2, \dots, X_{\tau_b^X - 1}, (X_{\tau_b^X} - \varsigma), \varsigma$. *The stochastic behavior of the observations* $X_1, X_2, \dots, X_{\tau_b^X - 1}, (X_{\tau_b^X} - \varsigma)$ *is* P_θ.

(b) *Given* θ, $\hat{\theta}_{\tau_b^X} \to \theta$ *in* P_θ*-probability as* $b \to \infty$.

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

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 ARL2FA $= B$ 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 ARL2FA 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 ρ 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 $\rho = 0.8, 0.6, 0.4, 0.2, -0.4, -1$. 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 ρ and A of the rate that the asymptotics kicks in: for a given value of ρ, the larger the cut-off level the better the approximation and for a given cut-off level, the larger the value of ρ the worse the approximation. The approximation indicated by Theorem 1 seems to work reasonably well for standard ARLs to false alarm if $\rho < 0.5$. 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 $\rho \le 0.4$) are consistent with independence.

Note that when the correlations are negative, things behave as expected by Theorem 2 even if the cut-off 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.)

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

	Threshold $A =$	100	500	1000	100	500	1000	100	500	1000
	correlation(X,Y)		$\rho = 0.8$			$\rho = 0.6$			$\rho = 0.4$	
M_X	mean	183	863	1704	172	890	1806	179	862	1793
M_X	sd	177	885	1812	168	872	1749	170	864	1824
M_Y	mean	180	948	1697	183	892	1789	167	892	1785
M_Y	sd	173	974	1777	162	920	1828	174	850	1809
$\min\{M_X, M_Y\}$	mean	119	567	1007	98	495	1002	100	457	922
$\min\{M_X, M_Y\}$	sd	114	596	1015	90	484	1028	96	440	957
M_X, M_Y	corr	0.307	0.266	0.200	0.115	0.099	0.136	0.1000	0.020	0.044
$M_{\max}, M_{\max} - M_{\min}$	corr	0.014	−0.024	0.050	0.009	0.003	0.005	0.009	−0.008	−0.009
	correlation(X,Y)		$\rho = 0.2$			$\rho = -0.4$			$\rho = -1$	
M_X	mean	181	856	1784	167	869	1785	176	880	1787
M_X	sd	172	817	1772	155	881	1753	172	859	1755
M_Y	mean	177	921	1814	174	922	1745	185	843	1697
M_Y	sd	175	879	1807	161	906	1743	178	854	1748
$\min\{M_X, M_Y\}$	mean	94	466	886	88	440	859	91	432	825
$\min\{M_X, M_Y\}$	sd	87	459	875	81	451	828	87	413	770
M_X, M_Y	corr	0.026	0.065	−0.028	−0.013	−0.013	−0.067	−0.018	−0.004	−0.069
$M_{\max}, M_{\max} - M_{\min}$	corr	0.023	0.035	−0.012	−0.007	−0.017	−0.046	−0.005	0.030	0.026

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 ρ. 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 ~ 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 0.45γ. 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 ~ 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 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 $E(X_i) = 1, i = 1, \dots, 5$, *when OOC by separate Shiryaev–Roberts control charts, stopping after two streams have signaled; nominal* ARL2FA = 741

$\rho = \text{corr}(X_i, X_j)$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
ARL2FA	745	744	760	772	787	802	833	876	964	1077	1661

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 ρ 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 cut-off value A for a given nominal ARL2FA is again calculated via $A = \text{ARL2FA}/(1.7845 \sum_{i=1}^{10} [1/(101 - i)]$ and the nominal standard deviation via $\text{SD} = 1.7845 A \sqrt{\sum_{i=1}^{10} [1/(101 - i)]^2}$. 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 ρ 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).

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)	$\rho = 0.2$			$\rho = 0.4$			$\rho = 0.6$		
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	0.095	0.037	0.002	0.141	0.032	0.038	0.192	0.100	0.076
$M_{\max}, M_{\max} - M_{\min}$	corr	−0.009	−0.050	0.002	0.045	0.04	−0.011	0.056	−0.016	−0.067
$\frac{1}{\frac{1}{\text{mean}\{M_X\}} + \frac{1}{\text{mean}\{M_Y\}}}$		74	357	738	77	363	756	74	362	771

TABLE 4
Simulations: *monitoring* $E_{\mathrm{IC}}(\mathbf{X_i}) = \mathbf{0}$ *vs.* $E_{\mathrm{OOC}}(\mathbf{X_i}) = \mathbf{1}$, $i = 1, \dots, 100$, *where* $\rho = \rho((\mathbf{X_i})_\mathbf{j}, (\mathbf{X_i})_\mathbf{k}) = \mathrm{correlation}((\mathbf{X_i})_\mathbf{j}, (\mathbf{X_i})_\mathbf{k})$, *by separate Shiryaev–Roberts control charts*, *stopping after* 10 *of* 100 *streams have signaled*, *for various nominal values of ARL2FA*

nominal	simulated	$\rho=0$	$\rho=0.1$	$\rho=0.2$	$\rho=0.3$	$\rho=0.4$	$\rho=0.5$	$\rho=0.6$	$\rho=0.7$	$\rho=0.8$	$\rho=0.9$
741	ARL2FA	763	760	769	836	871	938	1063	1268	1624	2263
234	SD	234	274	316	369	458	544	684	868	1252	1906
2500	ARL2FA	2534	2513	2593	2564	2753	2936	3186	3906	4854	6964
791	SD	781	816	942	1026	1249	1564	1906	2583	3600	5727
25,000	ARL2FA	24,775	25,057	25,538	25,310	25,463	26,133	29,749	32,743	39,357	60,209
7909	SD	7705	8302	8638	8773	10,078	11,927	14,549	19,502	26,699	49,032

4. Remarks.

REMARK 1. The multivariate case follows from the bivariate case. Suppose $N_1, N_2, \dots, N_p$ stopping times were run separately on each of p streams $\{\mathbf{X_i^{(1)}}\}, \{\mathbf{X_i^{(2)}}\}, \dots, \{\mathbf{X_i^{(p)}}\}$. Setting up blocks of size η_{b_X} 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 $P(K_{X_i^*} = K_{X_j^*}) = o(1)$. The independence between blocks accounts for the asymptotic independence of the stopping times, which are exponentially distributed.

REMARK 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 $\mathrm{correlation}(Z_\theta^X, Z_\lambda^Y) = -1$). 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.

REMARK 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 $\mathrm{correlation}(Z^X, Z^Y) = -1$, care must be taken, as the result may not be valid. For example, if $X \sim N(0, 1)$ under H_0^X and $Y = -X$, if $G_X = G_Y$ are standard normal then $\int e^{\sum_{i=k}^n Z_i^X(\theta)}\, dG_X(\theta)$ and $\int e^{\sum_{i=k}^n Z_i^Y(\lambda)}\, dG_Y(\lambda))$ are identical.

REMARK 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 Cauchy$(0, 1)$ when in control and 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 loglikelihood ratios $Z_i^X = \log(\frac{1+X_i^2}{1+(X_i-1)^2})$, $Z_i^Y = \log(\frac{1+Y_i^2}{1+(Y_i-1)^2})$ that have the same ARL2FA will stop together if the ARL2FA is large enough.

REMARK 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 δ is known (considered to be a representative of a post-change increase in mean) but μ is unknown. An example of a procedure based on invariance calls for defining $Y_i = X_i - X_1$ and monitoring the sequence $Y_1, Y_2, \dots$ by Cusum or Shiryaev–Roberts (cf. Pollak and Siegmund (1991)). A standard calculation obtains that the log-likelihood ratio of $Y_2, Y_3, \dots, Y_n$ for $\nu = k$ vs. $\nu = \infty$ (when translated back into the X's, ν 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 μ, so without loss of generality assume that $\mu = 0$. In this case, when it is known that $\mu = 0$, the log-likelihood ratio of $X_1, X_2, \dots, X_n$ for $\nu = k$ versus $\nu = \infty$ 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 $(\log A)/n + \sqrt{1/n}$.

Hence, for $n > o(\sqrt{A})$ and $k > n - O(\log A)$, the difference $[(n-k+1)[\frac{1}{2}\delta^2 \frac{n-k+1}{n} - \delta \bar{X}_n]$ 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 $E_{\mathrm{IC}} R_j = j$. Therefore,

$$\begin{aligned} &P_{\mathrm{IC}}\{\text{Shiryaev–Roberts stops before } m\} \\ &\quad = P_{\mathrm{IC}}\{R_j > A \text{ for some } j < m\} \\ &\quad \le \sum_{j=1}^{m} P_{\mathrm{IC}}\{R_j > A\} \le \sum_{j=1}^{m} \frac{E_{\mathrm{IC}}(R_j)}{A} = \sum_{j=1}^{m} \frac{j}{A} = O\left(\frac{m^2}{A}\right). \end{aligned}$$

REMARK 6. A reference for the exponentiality of Shiryaev–Roberts is Yakir (1995, 1998). A reference for the classical (simple) Cusum is Pollak and Tartakovsky (2008). The statement for Cusum with a prior for the unknown post-change parameter seems to be novel.

REMARK 7. One of the deficiencies of change-point 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 θ. 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.

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

REMARK 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 4.) 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 (Pollak (2020)).

Acknowledgments. The author is affiliated at present also with the Israel Central Bureau of Statistics.

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.

SUPPLEMENTARY MATERIAL

Supplement to "A rule of thumb: Run lengths to false alarm of many types of control charts run in parallel on dependent streams are asymptotically independent" (DOI: 10.1214/20-AOS1968SUPP; .pdf). Supplementary information.

REFERENCES

EPPRECHT, E. K. (2015). Statistical control of multiple-stream processes: A literature review. In *Frontiers in Statistical Quality Control* 11 (S. Knoth and W. Schmid, eds.) 29–47. Springer-Verlag, New York.

LUZIA, N. (2018). A simple proof of the strong law of large numbers with rates. *Bull. Aust. Math. Soc.* **97** 513–517. MR3802466 https://doi.org/10.1017/S0004972718000059

MEI, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika* **97** 419–433. MR2650748 https://doi.org/10.1093/biomet/asq010

MENECES, N. S., OLIVERA, S. A., SACCONE, C. D. and TESSORE, J. (2008). Statistical control of multiple stream processes: A Shewhart control chart for each stream. *Qual. Engin.* **20** 185–194.

POLLAK, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist.* **15** 749–779. MR0888438 https://doi.org/10.1214/aos/1176350373

POLLAK, M. (2020). Supplement to "A rule of thumb: Run lengths to false alarm of many types of control charts run in parallel on dependent streams are asymptotically independent." https://doi.org/10.1214/20-AOS1968SUPP

POLLAK, M. and SIEGMUND, D. (1991). Sequential detection of a change in a normal mean when the initial value is unknown. *Ann. Statist.* **19** 394–416. MR1091859 https://doi.org/10.1214/aos/1176347990

POLLAK, M. and TARTAKOVSKY, A. G. (2008). Asymptotic exponentiality of the distribution of first exit times for a class of Markov processes with applications to quickest change detection. *Teor. Veroyatn. Primen.* **53** 500–515. MR2759707 https://doi.org/10.1137/S0040585X97983742

RÉNYI, A. (1957). On the asymptotic distribution of the sum of a random number of independent random variables. *Acta Math. Acad. Sci. Hung.* **8** 193–199. MR0088093 https://doi.org/10.1007/BF02025242

SIEGMUND, D. (1985). *Sequential Analysis*: *Tests and Confidence Intervals*. *Springer Series in Statistics*. Springer, New York. MR0799155 https://doi.org/10.1007/978-1-4757-1862-1

STONE, C. (1965). On moment generating functions and renewal theory. *Ann*. *Math*. *Stat*. **36** 1298–1301. MR0179857 https://doi.org/10.1214/aoms/1177700003

YAKIR, B. (1995). A note on the run length to false alarm of a change-point detection policy. *Ann*. *Statist*. **23** 272–281. MR1331668 https://doi.org/10.1214/aos/1176324467

YAKIR, B. (1998). On the average run length to false alarm in surveillance problems which possess an invariance structure. *Ann*. *Statist*. **26** 1198–1214. MR1635389 https://doi.org/10.1214/aos/1024691095