Efficient scalable schemes for monitoring a large number of data streams

BY Y. MEI

School of Industrial and Systems Engineering, Georgia Institute of Technology,
Atlanta, Georgia 30329-0205, U.S.A.
ymei@isye.gatech.edu

SUMMARY

The sequential changepoint detection problem is studied in the context of global online monitoring of a large number of independent data streams. We are interested in detecting an occurring event as soon as possible, but we do not know when the event will occur, nor do we know which subset of data streams will be affected by the event. A family of scalable schemes is proposed based on the sum of the local cumulative sum, CUSUM, statistics from each individual data stream, and is shown to asymptotically minimize the detection delays for each and every possible combination of affected data streams, subject to the global false alarm constraint. The usefulness and limitations of our asymptotic optimality results are illustrated by numerical simulations and heuristic arguments. The Appendices contain a probabilistic result on the first epoch to simultaneous record values for multiple independent random walks.

Some key words: Change detection; CUSUM; Renewal theory; Scalability; Sequential detection; Simultaneous record.

1. INTRODUCTION

Suppose that one is monitoring a large number of independent data streams over time, say, observing $X_{k,n}$ at the $k$th data stream over time $n = 1, 2, \ldots$ for $k = 1, \ldots, K$ with large $K$. At some unknown time, an unusual event changes the distributions of the observations $X_{k,n}$ at some data streams. It is desirable to utilize the observed streaming data $X_{k,n}$s to raise an alarm as soon as the event occurs while keeping false alarms as infrequent as possible. This sequential changepoint detection problem arises in areas such as quality control, surveillance, health care, security and environmental science.

One naive approach is to monitor each local data stream individually. For each local data stream, many efficient local monitoring schemes are available in the literature, and a partial list includes Shewhart’s control chart; moving average charts; Page’s cumulative sum, or CUSUM, procedure; and the Shiryaev–Roberts procedure; see, for example, Shewhart (1931), Page (1954), Shiryaev (1963) and Roberts (1966). For recent reviews, we refer readers to Basseville & Nikiforov (1993), Lai (1995, 2001) and Poor & Hadjiliadis (2008). Unfortunately, the local monitoring approach does not take advantage of global information, and may lead to large detection delays if several data streams provide information about the occurring event. More importantly, even if the local false alarm rate is well controlled at each data stream, the global false alarm rate can be severe when the number of data streams is large, leading to obvious costs and the classic boy who cried wolf phenomenon.

A better approach is to monitor these data streams globally. In the literature, two scenarios of global monitoring have gained a lot of attention. The first is consensus detection, in which the
distributions of all data streams are assumed to change simultaneously. The second is when the change is assumed to affect exactly one data stream, e.g. multi-channel detection in signal processing. See, for example, Montgomery (1991, Ch. 8-3); Veeravalli (2001); and Tartakovsky et al. (2006). In other words, if we denote by \( m \) the number of affected streams out of a total of \( K \) data streams, then the existing research focuses on two specific scenarios: \( m = K \) or \( m = 1 \).

In this article, we are interested in global monitoring when we do not know which data streams are affected by the change and when the number of affected data streams is either moderately large or completely unknown. These scenarios can happen to many real-world applications. For example, a security system may monitor many different data streams from different information sources, and an enemy target or intrusion may be detectable by some but not all sources. Another example occurs in biosurveillance, where each city or county may run a local surveillance system to monitor disease outbreaks, whereas a community disease outbreak may affect several cities or counties in a community, area or region. The example can also occur when a local syndromic surveillance system monitors a list of pre-identified sets of symptoms among certain medical care providers, and the outbreak of an emerging infectious disease increases the frequencies of several unknown sets of symptoms.

From the theoretical point of view, no matter what value of the number of affected data streams we have, efficient global online monitoring schemes can be developed by the standard statistical methods such as generalized likelihood ratios or mixture likelihood ratios that combine likelihood ratios for all possible post-change hypotheses; see, for example, from Lorden (1971) to Pollak (1987) to Lai (2001). However, from the computational or practical point of view, these standard statistical methods are not applicable for online monitoring of a large number of data streams when the number \( m \) of affected data streams is either moderately large or completely unknown. This is because the number of all possible post-change hypotheses, i.e. possible combinations for affected data streams, is \( C_m^K \) and \( 2^K - 1 \) in these cases, and both can be too huge to do the exhaustive search even when \( m \) and \( K \) are only moderately large.

The goal of this article is to develop a family of efficient global monitoring schemes that is scalable and robust for global online monitoring when the number of affected data streams is either moderately large or completely unknown. Here a scalable scheme is one that can be easily implemented online in the system with a large number of data streams over a long time period, and a robust scheme has the ability to detect different kinds of post-change scenarios regardless of the different combinations of affected data streams. In \( \S \) 4, such schemes are developed by using the sum of the local CUSUM statistics from each data stream, and are shown to be asymptotically optimal in a suitable sense.

It is useful to point out that the global monitoring problem we considered is the natural online version of multiple hypothesis testing problems. While classical off-line multiple hypothesis testing problems have been well studied in statistics, theoretical research on the corresponding online version is rather limited, partly because there is a fundamental difference between the online and off-line decision problems. In particular, the Type I and II error probabilities criteria of the off-line hypothesis testing problem have been replaced by the average run length of a scheme in the online changepoint detection problem. Moreover, while the term \( p \)-value may have been overused in the off-line hypothesis testing problem, it has been an open problem to define an analogue of a \( p \)-value that is meaningful in the online changepoint detection problem. We leave to one side the fascinating and extremely important problem of identifying the affected data streams after a monitoring scheme raises an alarm. Instead, we simply assume that decision makers will take a further look and act appropriately. Our interest here will centre on quickest detection, which is already challenging and is usually the essential step.
There are a few other approaches to the formulation and solution of problems involving global online monitoring, and most, if not all, have some twists to significantly reduce the number of possible post-change hypotheses so that the standard statistical methods are feasible in their respective contexts. For instance, Sonesson (2007) considered the problem in the context of detecting emerging clusters of disease in space-time disease surveillance and made assumptions on the shape and size of the disease cluster. He gave no optimality results, but did report numerical simulation results under several different specific assumptions. An interesting asymptotic theory developed in Siegmund & Yakir (2008) was motivated by image analysis and assumed that the shape of the signal/image is partially specified.

Throughout this article, we make the following standard assumptions.

Assumption 1. The observations \( X_{k,n} \) are independent over time as well as among different data streams.

Assumption 2. For each \( k = 1, \ldots, K \), the probability density function of the observations at the \( k \)th data stream is either \( f_k \) or \( g_k \) with respect to some sigma-finite measure \( \mu(x) \), where the \( f \)s and \( g \)s are given. For each \( k \), the Kullback–Leibler information number

\[
I(g_k, f_k) = \int \log \left( \frac{g_k(x)}{f_k(x)} \right) g_k(x) d\mu(x)
\]

is finite and positive. Moreover, exchanging the roles of \( f_k \)s and \( g_k \)s in (1), the Kullback–Leibler information number \( I(f_k, g_k) \) is also assumed to be finite and positive. In addition, we also assume that

\[
\int \log^2 \left( \frac{g_k(x)}{f_k(x)} \right) g_k(x) d\mu(x) < \infty.
\]

These two assumptions, although restrictive, are reasonable if one treats the \( X_{k,n} \)s as the residuals of some spatio-temporal models rather than the original raw observations, and if one treats the known \( f \)s and \( g \)s as representative pre-change and post-change distributions. In addition, these assumptions are a natural starting point for more sophisticated statistical analysis. Assumption 2 guarantees that the pre-change and post-change distributions at each affected data stream, \( f_k \) and \( g_k \), are separated.

2. PROBLEM FORMULATION

Suppose we are monitoring \( K \) data streams, observing the \( k \)th data stream \( X_{k,n} \) over time \( n = 1, 2, \ldots \). Initially, the \( X_{k,n} \)s are distributed according to the density \( f_k \) for \( k = 1, \ldots, K \). At some unknown time \( \nu \), an unusual event occurs and affects an unknown subset of data streams in the sense that if the \( k \)th data stream is affected, the density function of its local observations \( X_{k,n} \) changes from \( f_k \) to \( g_k \) at time \( \nu_k \geq \nu \). The problem is to raise an alarm as quickly as possible after the event occurs. For simplicity, here we assume \( \nu_k \equiv \nu \), i.e. an instantaneous change if the \( k \)th data stream is affected, but our proposed methods are also applicable when the \( \nu_k \)s are different.

In changepoint problems, a scheme for detecting that an event has occurred is defined as a stopping time \( T \) with respect to \( (X_{1,n}, \ldots, X_{K,n})_{n \geq 1} \). The interpretation of \( T \) is that, when \( T = n \), we stop at time \( n \) and declare that an event has occurred in some data streams somewhere in the first \( n \) time steps. We want to find a scheme \( T \), which will stop as soon as an event occurs but will continue sampling as long as possible if no event occurs. Thus, the performance of a scheme \( T \) is evaluated by two criteria: the detection delay and the false alarm rate.
Denote by \( P^{(k_1, \ldots, k_m)}_v \) and \( E^{(k_1, \ldots, k_m)}_v \) the probability measure and expectation when at time \( v \), the density of observations \( X_{k,v} \), changes from \( f_k \) to \( g_k \) only at the \( k \)th data stream for \( k = k_1, \ldots, k_m \), and there are no changes at other data streams. Also, denote by \( P_{\infty} \) and \( E_{\infty} \) the same when no events occur. Then the detection delay of a scheme \( T \) can be evaluated by the following worst case detection delay defined in Lorden (1971):

\[
\bar{E}^{(k_1, \ldots, k_m)}(T) = \sup_{v=1,2,\ldots} \text{ess sup} E^{(k_1, \ldots, k_m)}_v \{ (T - v + 1)^+ \mid F_{v-1} \},
\]

where \( F_{v-1} = (X_{1,1,v-1}, \ldots, X_{K,1,v-1}) \) denotes past global information at time \( v \) and \( X_{k,1,v-1} = (X_{k,1}, \ldots, X_{k,v-1}) \) denotes past local information for the \( k \)th data stream. Here \( u^+ = \max(u, 0) \) and ess sup denotes the essential supremum which takes the least-favorable observations before the change. Often but not always the worst case detection delay of a scheme occurs at \( v = 1 \), i.e. it is more difficult to detect when a change occurs at earlier stages rather than at later stages.

It turns out (Lorden, 1971) that if \( \bar{E}^{(k_1, \ldots, k_m)}(T) \) is finite, then \( P_{\infty}(T < \infty) = 1 \), implying that we will have a false alarm sooner or later with probability 1 even when no event occurs. Thus, in the changepoint detection literature, the false alarm rate is usually measured by \( 1 / E_{\infty}(T) \), where \( E_{\infty}(T) \) is often called the average run length to false alarm.

Our problem can then be formulated as finding a scheme \( T \) that minimizes the detection delay \( \bar{E}^{(k_1, \ldots, k_m)}(T) \) for each and every possible subset \( (k_1, \ldots, k_m) \) of \( 1, \ldots, K \) subject to the average run length to false alarm constraint

\[
E_{\infty}(T) \geq \gamma,
\]

where \( \gamma \) is a prespecified constant to control the global false alarm rate.

3. Likelihood Ratio Procedures

A standard tool to construct statistical tests or procedures is the likelihood ratio. In this section we present likelihood ratio procedures in our context to provide a better understanding of their usefulness and limitations. On the one hand, they are computationally infeasible under our setting. On the other hand, they shed light on developing new desired schemes.

The changepoint problem can be thought of as sequentially testing the simple pre-change hypothesis, \( H_0 : P_{\infty} \) is true, against the composite post-change hypothesis

\[
H_1 : P^{(k_1, \ldots, k_m)}_v \text{ is true for some finite } v = 1, 2, \ldots \text{ and for some } (k_1, \ldots, k_m) \in \Delta,
\]

where \( \Delta \) represents the set of all possible subsets of affected data streams, which of course depends on our prior knowledge about the post-change scenario. Denote by \( X_{[1.1,n]} = (X_{1,1,n}, \ldots, X_{K,1,n}) \) all information up to time \( n \). Then the logarithm of the likelihood ratio statistic up to time \( n \) under our setting is

\[
G_n = \max_{v=1,2,\ldots} \max_{(k_1, \ldots, k_m) \in \Delta} \log \frac{dP^{(k_1, \ldots, k_m)}_v}{dP_{\infty}}(X_{[1,k],1,n}),
\]

and the corresponding likelihood ratio procedure will raise an alarm at the time

\[
T_{LR}(a) = \inf \{ n \geq 1 : G_n \geq a \},
\]

where the constant \( a > 0 \) is chosen to satisfy the false alarm constraint in (2). As is conventional, the stopping time in (4) and also other stopping times below are defined to be \( \infty \) if no such \( n \)
exists. Observing that the $P_{V}^{(k_1, \ldots, k_m)}$-distribution of $X_{[1,k], [1,n]}$ is the same as $P_{\infty}$-distribution of $X_{[1,k], [1,n]}$ when $n < \nu$, it is easy to see that expression (3) can be simplified as

$$G_n = \max \left\{ 0, \max_{i=1, \ldots, n} \max_{(k_1, \ldots, k_m) \in \Delta} \sum_{i=1}^{n} \sum_{j=1}^{m} \log \frac{g_{k_i}(X_{k_j,i})}{f_{k_j}(X_{k_j,i})} \right\}.$$  

If one has perfect prior knowledge about the subset of affected data streams, i.e. if $\Delta$ only contains a single point, say, $\Delta = \{(k_1, \ldots, k_m)\}$, then the statistic $G_n$ in (3) is just the CUSUM statistic, and the procedure $T_{\text{LR}}(a)$ in (4) becomes the well-known CUSUM procedure (Page, 1954). To emphasize this fact, we will use new notations $W$ and $T_{\text{CM}}$ to denote the statistic $G_n$ and the procedure $T_{\text{LR}}$ for this scenario. Specifically, the statistic $G_n$ in (3) becomes

$$W^{(k_1, \ldots, k_m)}_n = \max \left\{ 0, \max_{i=1, \ldots, n} \sum_{j=1}^{m} \log \frac{g_{k_i}(X_{k_j,i})}{f_{k_j}(X_{k_j,i})} \right\},$$  

and $T_{\text{LR}}(a)$ in (4) becomes

$$T_{\text{CM}}^{(k_1, \ldots, k_m)}(a) = \inf \left\{ n \geq 1 : W^{(k_1, \ldots, k_m)}_n \geq a \right\}.$$  

A nice property of the CUSUM statistic $W^{(k_1, \ldots, k_m)}_n$ in (5) is that it can be calculated recursively via

$$W^{(k_1, \ldots, k_m)}_n = \max \left\{ 0, W^{(k_1, \ldots, k_m)}_{n-1} + \sum_{j=1}^{m} \log \frac{g_{k_i}(X_{k_j,n})}{f_{k_j}(X_{k_j,n})} \right\}$$

for $n \geq 1$ and $W^{(k_1, \ldots, k_m)}_0 = 0$. This recursive form is very useful in online monitoring since it significantly reduces computation complexity and memory. Moreover, it is well known that $T_{\text{CM}}^{(k_1, \ldots, k_m)}(a)$ in (6) is exactly the optimal scheme to detect the change when $(k_1, \ldots, k_m)$ is indeed the subset of actual affected data streams; see Moustakides (1986).

Let us now focus on more realistic scenarios when one does not have a perfect prior knowledge about the subset of affected data streams, i.e. $|\Delta| \geq 2$. While the procedure $T_{\text{LR}}(a)$ in (4) seems to be still efficient from the theoretical viewpoint, its naive implementation is generally infeasible for online monitoring, as (3) generally no longer enjoys the above nice recursive form for $W^{(k_1, \ldots, k_m)}_n$. To overcome this, one alternative implementation of the procedure $T_{\text{LR}}(a)$ is based on the fact that it can be rewritten as

$$T_{\text{LR}}(a) = \inf \left\{ n \geq 1 : \max_{(k_1, \ldots, k_m) \in \Delta} W^{(k_1, \ldots, k_m)}_n \geq a \right\},$$

or equivalently,

$$T_{\text{LR}}(a) = \min_{(k_1, \ldots, k_m) \in \Delta} T_{\text{CM}}^{(k_1, \ldots, k_m)}(a),$$

where $W^{(k_1, \ldots, k_m)}_n$ in (5) and $T_{\text{CM}}^{(k_1, \ldots, k_m)}(a)$ in (6) can be conveniently implemented online. In other words, we can implement the procedure $T_{\text{LR}}(a)$ in (4) as simultaneously monitoring a total of $|\Delta|$ schemes with each scheme designed for a specific possible post-change scenario.

This alternative implementation is very efficient when $|\Delta|$, the number of possible post-change hypothesis, is not too large, and it has been well documented particularly when one knows that exactly $m = 1$ out of these $K$ data streams is affected, i.e. when $\Delta = \{1, \ldots, K\}$; see, for example, Tartakovsky et al. (2006). Specifically, at time $n$, for each data stream, one can calculate its local
CUSUM statistic recursively by

\[ W^{(k)}_n = \max \left\{ 0, \ W^{(k)}_{n-1} + \log \frac{g_k(X_{k,n})}{f_k(X_{k,n})} \right\} \tag{7} \]

for \( n \geq 1 \), and \( W^{(k)}_0 = 0 \). Then one will raise an alarm at the global level at time

\[ N_{\text{max}}(a) = \inf \left\{ n \geq 1 : \ \max_{k=1}^{K} W^{(k)}_n \geq a \right\}. \tag{8} \]

Here for the purpose of comparison with our proposed scheme, we use \( N_{\text{max}} \) to emphasize that this procedure is based on the maximum of local CUSUM statistics, although it is in fact the likelihood ratio procedure corresponding to the scenario when one knows that exactly \( m = 1 \) out of the \( K \) data streams is affected.

Unfortunately, when the number \( m \) of affected data streams is either known to be moderately large or completely unknown, this alternative implementation is also infeasible, as it requires us to monitor \( C_j^m \) or \( 2^K - 1 \) schemes simultaneously. Hence, new implementations or new methodologies are needed.

### 4. Our Proposed Methodology

Our proposed scheme for global online monitoring of multiple data streams is as follows. At time \( n \), for each data stream, we calculate its local CUSUM statistic \( W^{(k)}_n \) recursively as in (7). Then we will raise an alarm at the global level and declare that a change has occurred as soon as the sum of all local CUSUM statistics \( W^{(k)}_n \) exceeds some prespecified constant threshold. The proposed stopping time is

\[ N_{\text{sum}}(a) = \inf \left\{ n \geq 1 : \ \sum_{k=1}^{K} W^{(k)}_n \geq a \right\}, \tag{9} \]

where the constant \( a > 0 \) is chosen to satisfy the false alarm constraint (2). To shorten the notation, below we will just refer to the stopping time \( N_{\text{sum}}(a) \) in (9) as our proposed scheme.

The intuition behind our proposed scheme \( N_{\text{sum}}(a) \) in (9) is simple: a large value of the local CUSUM statistic indicates a possible change at the specific local data stream, and thus it is natural to explore the sum of local CUSUM statistics if the number \( m \) of affected data streams is either moderately large or completely unknown, since a large sum may imply that some individual data streams are affected by the change. A comparison between (9) and (8) shows that we simply replace the maximum of local CUSUM statistics by their sum. Intuitively, one would expect that in the finite sample simulations, \( N_{\text{max}}(a) \) should work better when a small number of data streams are affected, whereas our proposed scheme \( N_{\text{sum}}(a) \) is better otherwise. Our simulation studies below indeed confirm these intuitions.

A surprising theoretical result for our proposed scheme \( N_{\text{sum}}(a) \) in (9) is that as \( a \) goes to \( \infty \), it is asymptotically optimal up to first-order to detect each and every possible combination of affected data streams when the data streams are independent, no matter how many data streams are affected. The main result is stated in the next theorem and its corollary, and the proof of the theorem is given in Appendix B.

**Theorem 1.** As \( a \to \infty \),

\[ E_{\infty} \{ N_{\text{sum}}(a) \} \geq \left\{ 1 + o(1) \right\} \frac{\exp(a)}{1 + a + a^2/2! + \cdots + a^{K-1}/(K-1)!}. \tag{10} \]
Moreover, for any possible subset \((k_1, \ldots, k_m)\) of \((1, \ldots, K)\), as \(a \to \infty\),
\[
\tilde{E}^{(k_1, \ldots, k_m)}(N_{\text{sum}}(a)) \leq \frac{a}{\sum_{j=1}^{m} I(g_{k_j}, f_{k_j})} + O(1),
\]
where \(I(g_k, f_k)\) is the Kullback–Leibler information number defined in (1).

**Corollary 1.** For a given \(K\), let
\[
a_\gamma = \log \gamma + (K - 1) \log \log \gamma + o(\log \log \gamma),
\]
as \(\gamma \to \infty\). Then \(N_{\text{sum}}(a_\gamma)\) satisfies the false alarm constraint (2) and
\[
\tilde{E}^{(k_1, \ldots, k_m)}(N_{\text{sum}}(a_\gamma)) \leq \frac{\log \gamma + (K - 1) \log \log \gamma + o(\log \log \gamma)}{\sum_{j=1}^{m} I(g_{k_j}, f_{k_j})}.
\]
Thus, as \(\gamma \to \infty\), our proposed scheme \(N_{\text{sum}}(a_\gamma)\) asymptotically minimizes the detection delay up to the order \(O(\log \log \gamma)\) for each and every possible post-change hypothesis subject to the false alarm constraint (2).

**Proof.** The choice of \(a_\gamma\) in (12) follows from (10) and the fact that
\[
1 + a + a^2/2! + \cdots + a^{K-1}/(K-1)! \sim a^{K-1}/(K-1)!
\]
as \(a\) goes to \(\infty\). Moreover, relation (13) follows at once from (11) and the definition of \(a_\gamma\) in (12). Finally, the asymptotic optimality properties of \(N_{\text{sum}}(a_\gamma)\) is a simple application of the classical well-known relationship between the average run length to false alarm and the detection delay: there exists a constant \(M\) such that for any stopping time \(\tau\),
\[
\tilde{E}^{(k_1, k_2, \ldots, k_m)}(\tau) \geq \frac{\log E_\infty(\tau)}{\sum_{j=1}^{m} I(g_{k_j}, f_{k_j})} + M,
\]
where \(M\) depends only on the \(f_k\)s and \(g_k\)s; see, for example, Proposition 2.1 of Mei (2006). \(\square\)

To better understand our proposed scheme \(N_{\text{sum}}(a)\) in (9) and its properties in Theorem 1 and Corollary 1, we add some remarks. First, \(N_{\text{sum}}(a)\) is scalable and can be easily implemented for large \(K\) over long time periods, since it only requires \(O(K)\) computations and memory allocations at every time step \(n\). Moreover \(N_{\text{sum}}(a)\) becomes the well-known local CUSUM procedure in the case of a single data stream, i.e. when \(K = 1\).

Second, \(N_{\text{sum}}(a)\) is robust and omnibus in the sense that it can detect many kinds of changes including not only different combinations of affected data streams but also different changepoint \(\nu_k\)s in different data streams. To see this, note that when the observations are independent between data streams, our proposed scheme \(N_{\text{sum}}(a)\) in (9) is based on the statistic \(\sum_{k=1}^{K} W_n^{(k)}\), which is just the likelihood ratio statistic in the problem when the changepoint \(\nu_k\) at the \(k\)th data stream is assumed to be unknown for all \(k\). This feature suggests that our proposed scheme is applicable to other scenarios when a common event triggers a different onset time of changes at different data streams, although the asymptotic optimality property may depend on how fast the changes occur relative to the data accumulation.

Of course, like any other omnibus schemes or tests, our proposed scheme \(N_{\text{sum}}(a)\), despite being asymptotically optimal, may not be efficient when precise prior knowledge about possible changes is available, especially when the average run length to false alarm constraint \(\gamma\) is only moderately large. For example, if we know that the change occurs simultaneously at all data streams, or if we know that the change occurs only at one data stream, then better detection schemes can be defined.
Third, the idea of using the sum of some suitable statistics to construct detection schemes has been well documented in the literature, especially the celebrated Shiryaev–Roberts procedure; see also Pollak (1985, 2006). A fundamental difference between our proposed scheme and the Shiryaev–Roberts procedure can be found from the spatial-temporal detection viewpoint: our proposed scheme sums the CUSUM statistics over the spatial domain whereas the Shiryaev–Roberts procedure sums the likelihood ratio statistics over the time domain.

Fourth, the main difference between our proposed schemes and those theoretically efficient but practically infeasible global monitoring schemes can also be easily seen from the spatial-temporal detection viewpoint: our proposed schemes look first at the time domain to derive individual local CUSUM statistics for each individual data stream, and then look in the spatial domain by taking advantage of independence between data streams. Schemes such as the likelihood ratio procedure $T_{LR}(a)$ in (4) or the extension of Shiryaev–Roberts procedures first look at the spatial domain to find out all possible post-change scenarios, and then consider the time domain to detect possible changes. Switching the spatial and temporal roles allows us to reduce the order of computational complexity to $K$ from $C^m_K$ or $2^K - 1$, leading to scalable schemes.

The likelihood ratio scheme $T_{LR}(a)$ in (4) or extensions of Shiryaev–Roberts procedures have good theoretical performance. It is not difficult to show that subject to the constraint (2), the second-order term of their detection delays is $K \log 2$, which is smaller than $(K - 1) \log \log \gamma$ in (13) for our proposed scheme $N_{\text{sum}}(a)$ in (9). This is the price we pay for scalable schemes.

Fifth, the log $\log \gamma$ term in (13) is not surprising, as it has also arisen in other contexts such as the changepoint problem with composite post-change distributions for one-parametric exponential families; see, for example, Theorem 4 of Pollak (1987), or § 3.2 of Lai (1995). The only difference is that the coefficient of the log log $\gamma$ term for one-parametric exponential families is $1/2$, whereas the coefficient in our context is $K - 1$. It remains an open problem whether we can find another family of scalable schemes with smaller second-order term in their detection delays.

Finally, while relation (10) is proved only for fixed $K$ and large $a$, it also seems to be useful to guide us to choose the threshold $a = a_\gamma$ so that $N_{\text{sum}}(a)$ satisfies (2) for large $K$ and moderately large $\gamma$. To see this, suppose we want to choose $a$ such that

$$C \times \frac{\exp(a)}{1 + a + a^2/2! + \cdots + a^{K-1}/(K-1)!} = \gamma,$$

for some constant $C$. If we denote by $X_K$ the sum of $K$ independent exponential random variables with mean 1, then $X_K$ has a Gamma distribution and it is equivalent to find $a = a_\gamma$ so that

$$P(X_K \geq a) = e^{-a} \sum_{k=0}^{K-1} \left(\frac{a^k}{k!}\right) = C/\gamma.$$

By the central limit theorem, we have $(X_K - K)/K^{1/2} \sim Z = N(0, 1)$ for large $K$. Hence a heuristic choice of the threshold $a$ is

$$a_\gamma \approx K + K^{1/2} Z_{C/\gamma}, \quad (14)$$

where $Z_\alpha$ is the value $z$ such that $P\{N(0, 1) > z\} = \alpha$. In our simulations below, relation (14) is used to simplify our computation: we first use (14) to guess the range of possible values of $a_\gamma$, and then conduct Monte Carlo simulations to determine the actual value of $a_\gamma$. 

5. NUMERICAL SIMULATIONS

We now present a numerical illustration of our asymptotic results. Suppose that one monitors $K = 100$ independent data streams. For $k = 1, \ldots, K$, the observations at the $k$th data stream are independent and identically normally distributed with mean $\mu_0 = 0$ and variance $\sigma_k^2$ before the change and with mean $\mu_1 = 0.5$ and variance $\sigma_k^2$ after the possible change.

As an illustration, two scenarios are considered: a homogeneous scenario where $\sigma_k \equiv 1$ for all $k = 1, \ldots, K$; and a nonhomogeneous scenario where $\sigma_1 = 0.25$ and $\sigma_k = 1$ for all other $k = 2, \ldots, K$; that is, the first data stream is the most sensitive.

For the purpose of comparison, besides our proposed scheme $N_{\text{sum}}(a)$ in (9), we also consider the scheme $N_{\text{max}}(a)$ in (8), which is designed for the scenario when exactly $m = 1$ out of these $K$ data streams are affected. Another natural candidate may be Page’s CUSUM procedure for detecting simultaneous changes at all data streams; however, if the event actually affects less than half of the homogeneous data streams, the detection delays of such a procedure are computationally expensive to simulate, due to heavy-tail distributions with very large mean; see also Tartakovsky et al. (2006). We also report the smallest possible detection delays when the subset of affected data streams is known. These best values are attained by the optimal CUSUM procedure $T_{(k_1, \ldots, k_m)}(a)$ in (6) when $(k_1, \ldots, k_m)$ is the subset of affected data streams. We want to emphasize that these best values require perfect prior knowledge about the affected data streams and thus are unattainable in practical situations when an unknown group of data streams is affected by the occurring event.

For each of these schemes $T(a)$, a Monte Carlo simulation with 1000 replicates was first performed to determine the appropriate values of the threshold $a$ to satisfy $E_\infty\{T(a)\} \approx \gamma$ within the range of sampling error. Next, using the obtained threshold value $a$, we ran 1000 replicates to simulate the detection delays when the changepoint occurs at time $\nu = 1$ under several different post-change scenarios, i.e. with a different number of affected data streams, thereby providing the estimated values of the worst case detection delays. This is because, for each detection scheme we considered, the worst case detection delay occurs when the changepoint $\nu = 1$.

Our simulation results are summarized in Tables 1 and 2 with the false alarm constraint $\gamma = 10^3$ or $10^4$. Table 1 reports the homogeneous scenario, and considers different sub-scenarios of post-change depending on how many data streams are actually affected. Table 2 summarizes two sub-scenarios of the nonhomogeneous scenario, depending on whether or not the most sensitive data stream is affected by the event. In each table, the Monte Carlo estimate of the detection delay is recorded, and the largest standard error is given in the caption of the table.

Table 1 shows that in the homogeneous scenario, our proposed scheme $N_{\text{sum}}(a)$ performs worse than the scheme $N_{\text{max}}(a)$ when the change occurs in less than five out of 100 data streams, but it
Table 2. Detection delays for the nonhomogeneous scenario. Top: different post-change scenarios when the most sensitive data stream is not affected, and the largest standard error for the simulation results is 0.4, except when the $m = 1$ data stream is affected in which the corresponding largest standard error is 2. Bottom: when the most sensitive data stream is affected, and the largest standard error is 0.1.

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>Detection scheme</th>
<th>80</th>
<th>50</th>
<th>20</th>
<th>10</th>
<th>5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affected subset known</td>
<td>$N_{\text{max}}(a = 8.78)$</td>
<td>23.3</td>
<td>24.7</td>
<td>28.8</td>
<td>33.1</td>
<td>38.5</td>
<td>66</td>
</tr>
<tr>
<td>$N_{\text{sum}}(a = 101.09)$</td>
<td>6.5</td>
<td>9.1</td>
<td>17.3</td>
<td>27.6</td>
<td>43.5</td>
<td>128</td>
<td></td>
</tr>
<tr>
<td>Affected subset known</td>
<td>$N_{\text{max}}(a = 11.15)$</td>
<td>32.8</td>
<td>34.8</td>
<td>40.0</td>
<td>45.5</td>
<td>52.6</td>
<td>86</td>
</tr>
<tr>
<td>$N_{\text{sum}}(a = 110.49)$</td>
<td>7.3</td>
<td>10.2</td>
<td>20.0</td>
<td>33.3</td>
<td>54.8</td>
<td>193</td>
<td></td>
</tr>
</tbody>
</table>

The first data stream not affected

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>Detection scheme</th>
<th>100</th>
<th>50</th>
<th>20</th>
<th>10</th>
<th>5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affected subset known</td>
<td>$N_{\text{max}}(a = 8.78)$</td>
<td>5.1</td>
<td>5.1</td>
<td>5.1</td>
<td>5.1</td>
<td>5.1</td>
<td>5.1</td>
</tr>
<tr>
<td>$N_{\text{sum}}(a = 101.09)$</td>
<td>5.0</td>
<td>7.6</td>
<td>11.7</td>
<td>14.7</td>
<td>17.2</td>
<td>20.1</td>
<td></td>
</tr>
<tr>
<td>Affected subset known</td>
<td>$N_{\text{max}}(a = 11.15)$</td>
<td>6.2</td>
<td>6.2</td>
<td>6.2</td>
<td>6.2</td>
<td>6.2</td>
<td>6.2</td>
</tr>
<tr>
<td>$N_{\text{sum}}(a = 110.49)$</td>
<td>5.5</td>
<td>8.4</td>
<td>13.2</td>
<td>17.0</td>
<td>20.0</td>
<td>23.5</td>
<td></td>
</tr>
</tbody>
</table>

is better than $N_{\text{max}}(a)$ when the change occurs in more than five data streams. This is consistent with our intuition. Moreover, as we expected, both $N_{\text{sum}}(a)$ and $N_{\text{max}}(a)$ have larger detection delays than the optimal CUSUM procedure designed for the specific post-change hypothesis when one knew the subset of affected data streams.

Table 2 shows that in the nonhomogeneous scenario, our proposed $N_{\text{sum}}(a)$ performs better than the scheme $N_{\text{max}}(a)$ when the most sensitive data stream is not affected and when the change occurs in more than five out of 100 data streams. However, the scheme $N_{\text{max}}(a)$ has a smaller detection delay when the most sensitive data stream is affected. When the most sensitive data stream is affected, the detection delays of the scheme $N_{\text{max}}(a)$ are similar no matter how many other data streams are affected when the false alarm constraint $\gamma$ is only moderately large.

From Tables 1 and 2, we can see that the heuristic relation (14), derived from the Gamma distribution and the central limit theorem, seems to be reasonable for large $K$ and a moderately large value of $\gamma$. All these simulation results are consistent with our asymptotic theory. In addition, our results also have implications for applications: monitoring more data streams does not necessarily mean that one can detect the occurring event earlier, especially if one is not sure whether a data stream will provide information to the event. Indeed, practitioners may need to find better data sources instead of more noisy data sources to achieve required system performance in specific applications.

Acknowledgement

I would like to thank the referees and the associate editor for careful reading and valuable suggestions. This research was supported by grants from the Air Force Office of Scientific Research and from the National Science Foundation, U.S.A.
APPENDIX A

On first epoch to simultaneous record values

Here we present a probabilistic result in renewal theory that is of interest on its own merit and is used to prove our theorem. Suppose there are $K$ independent sequences of random variables, say, $\{\xi_{k,n}\}_{n=1}^{\infty}$, where $\xi_{k,1}, \xi_{k,2}, \ldots$ are independently and identically distributed random variables with mean $\mu_k < 0$. For each $k = 1, \ldots, K$, consider a one-dimensional random walk

$$S_{k,0} = 0, \quad S_{k,n} = \xi_{k,1} + \cdots + \xi_{k,n}.$$ 

Let $T$ be the first epoch where all $K$ random walks reach record minimum values simultaneously, i.e.

$$T = \inf \{n \geq 1 : S_{k,n} = \min_{i=0,\ldots,n} S_{k,i}, \quad k = 1, \ldots, K\}.$$

**Proposition 1.** Assume $E(\xi_{k,n}) = \mu_k < 0$ for all $k = 1, \ldots, K$. Then

$$E(T) = \prod_{k=1}^{K} E(T_k)$$

is finite, where $T_k$ is the first weak descending ladder epoch for the $k$th random walk $S_{k,n}$, i.e.

$$T_k = \inf \{n \geq 1 : S_{k,n} = \min_{i=0,\ldots,n} S_{k,i}\}.$$ 

The same assertions are true if $E(\xi_{k,n}) = \mu_k > 0$ and the minimum value and the weak descending ladder epoch are replaced by the maximum value and the weak ascending ladder epoch, respectively.

**Proof.** Since $S_{k,n}$ drifts to $-\infty$, the finiteness of $E(T_k)$ is well known from renewal theory; see, for example, Theorem 3 of Feller (1971, p. 416). Thus, it suffices to prove relation (A1). The main idea in the proof is to consider all $K$ renewal processes formed by the sequences of weak descending ladder points at each random walk, and then investigate the probability that all $K$ processes have simultaneous renewals at a common time $n$. Let $p_{k,n}$ be the probability that the $k$th renewal process has a renewal at time $n$, i.e.

$$p_{k,n} = \Pr(S_{k,n} = \min_{i=0,\ldots,n} S_{k,i}).$$ 

Since $S_{k,n}$ drifts to $-\infty$, it is well known from renewal theory that, as $n$ goes to $\infty$, $p_{k,n}$ converges to a constant $p_k^* > 0$, where

$$p_k^* = \Pr(S_{k,n} > 0, n = 1, 2, \ldots) = E(T_k)^{-1};$$

see, for example, Corollary 8.39 of Siegmund (1985, p. 173), or Theorem 3 of Feller (1971, p. 416).

Denote by $N_t$ the number of simultaneous renewals for these $K$ processes up to time $t$. On the one hand, it is well known from renewal theory that as $t \to \infty$,

$$N_t \overset{\text{a.s.}}{\to} E(T)^{-1}$$

see, for example, Theorems 4.1 and 4.2 of Durrett (1996, p. 204). On the other hand, by independence, the probability that all $K$ renewal processes have renewals at time $n$ is

$$\Pr(S_{k,n} = \min_{i=0,\ldots,n} S_{k,i}, k = 1, \ldots, K) = \prod_{k=1}^{K} p_{k,n},$$

which tends to the constant $\prod_{k=1}^{K} p_k^* > 0$ as $n \to \infty$, and the expected number of simultaneous renewals up to time $t$ is

$$E(N_t) = \sum_{n=1}^{t} \left( \prod_{k=1}^{K} p_{k,n} \right).$$
Combining the above relations together yields that
\[
E(T)^{-1} = \lim_{t \to \infty} \frac{E(N_t)}{t} = \lim_{t \to \infty} \frac{\sum_{n=1}^{t} (\prod_{k=1}^{K} p_{kn})}{t} = \prod_{k=1}^{K} p_k^* > 0.
\]
Relation (A1) then follows at once from this and relation (A2).

**APPENDIX B**

**Proof of Theorem 1**

We first prove relation (11) on the detection delay of our proposed scheme \(N_{\text{sum}}(a)\) in (9), as its proof is simpler. For any given subset \((k_1, \ldots, k_m)\) of \((1, 2, \ldots, K)\), it is evident from the definitions in (5) and (7) that the CUSUM statistics \(W_{n}^{(k_1, \ldots, k_m)}\) are dominated by the statistic \(\sum_{k=1}^{K} W_{n}^{(k)}\) for all \(n\). Thus as stopping times, the proposed scheme \(N_{\text{sum}}(a)\) is dominated by Page’s CUSUM procedure \(T_{CM}^{(k_1, \ldots, k_m)}(a)\) in (6), and so are the detection delays, since the worst-case detection delays occur at time \(\nu = 1\) for both schemes. By the classical result on the CUSUM procedure, the detection delay of Page’s CUSUM scheme \(T_{CM}^{(k_1, \ldots, k_m)}(a)\) is characterized by the right-hand side of (11); see, for example, equation (2.53) on p. 26 of Siegmund (1985). Hence, relation (11) also holds for our proposed scheme \(N_{\text{sum}}(a)\).

Now let us focus on the proof of relation (10) on the average run length to false alarm of our proposed scheme \(N_{\text{sum}}(a)\). The proof decomposes into two main parts. The easiest part relates the sum of local CUSUM statistics, \(\sum_{k=1}^{K} W_{n}^{(k)}\), to a Gamma random variable; see Lemma B1 below. This relation is an extension of a well-known upper bound on the tail distribution of CUSUM statistics. The most original and tough part is Lemma B2 below, which uses our probabilistic result in Appendix A and regenerative arguments to show that the proposed scheme \(N_{\text{sum}}(a)\) is asymptotically exponentially distributed under \(P_{\infty}\) when there are no changes. It is well known that the asymptotic exponential distribution property is held by many classical changepoint schemes such as CUSUM or Shiryaev–Roberts; however, the proof to \(N_{\text{sum}}(a)\) in (9) by no means is either straightforward or obvious.

To simplify our notation below, let
\[
B = B_a = e^{\frac{\theta}{K-1}} \prod_{j=0}^{K-1} \frac{a^j}{j!}.
\]
Since \(N_{\text{sum}}(a) \geq 1\) is an integer-valued random variable,
\[
E_{\infty}\{N_{\text{sum}}(a)\} = \sum_{n=1}^{\infty} P_{\infty}\{N_{\text{sum}}(a) \geq n\} \geq \sum_{n=1}^{\infty} \int_{n}^{n+1} P_{\infty}\{N_{\text{sum}}(a) \geq x\} dx = \int_{1}^{\infty} P_{\infty}\{N_{\text{sum}}(a) \geq x\} dx = B \int_{1/B}^{\infty} P_{\infty}\{N_{\text{sum}}(a) \geq tB\} dt.
\]
By Lemma B2 below and Fatou’s lemma,
\[
\liminf_{a \to \infty} \frac{E_{\infty}\{N_{\text{sum}}(a)\}}{B} = \int_{0}^{\infty} \liminf_{a \to \infty} P_{\infty}\{N_{\text{sum}}(a) \geq tB\} I(t \geq 1/B) dt = \int_{0}^{\infty} \liminf_{a \to \infty} P_{\infty}\{N_{\text{sum}}(a) \geq tB\} dt \geq \int_{0}^{\infty} \exp(-t) dt = 1
\]
and thus (10) holds.

To complete the proof, we need the following two lemmas.
Lemma B1. For any integers $K \geq 1$ and $n \geq 1$, and for any real number $a \geq 0$,

$$P_\infty \left( \sum_{k=1}^{K} W_n^{(k)} \geq a \right) \leq e^{-a} \sum_{j=0}^{K-1} \frac{a^j}{j!}.$$ 

Proof. Before we present a rigorous proof, it is informative to restate Lemma B1 under the context that $X$ is stochastically smaller than $Y$ when $\Pr(X \geq t) \leq \Pr(Y \geq t)$ for all $t$. Essentially, Lemma B1 claims that $\sum_{k=1}^{K} W_n^{(k)}$ is stochastically smaller than the sum of $K$ standard exponential random variables, which has a Gamma distribution. Thus we prove Lemma B1 by induction.

If $K = 1$, the lemma follows from a well-known property of the CUSUM statistic that for any real number $a \geq 0$, $P_\infty(W_n^{(1)} \geq a) = \exp(-a)$; see, for example, Appendix 2 on p. 245 of Siegmund (1985), or Lemma 3 of Mei (2005). Hence the lemma holds for $K = 1$.

Now suppose $K \geq 2$ and the lemma holds for $K - 1$. Since the observations are independent between different data streams and the $W_n^{(k)}$s are nonnegative by their definitions in (7),

$$P_\infty \left( \sum_{k=1}^{K} W_n^{(k)} \geq a \right) \leq P_\infty \left( \sum_{k=1}^{K-1} W_n^{(k)} \geq a \right) + \int_0^a P_\infty \left( \sum_{k=1}^{K-1} W_n^{(k)} \leq dy \right) P_\infty \left( W_n^{(K)} \geq a - y \right).$$

Using the lemma for the case of $K = 1$, we have $P_\infty(W_n^{(1)} \geq a - y) = \exp(-(a - y))$ for any $0 \leq y \leq a$. Thus, the second term of the right-hand side of the above inequality is bounded by

$$\int_0^a e^{-(a-y)} P_\infty \left( \sum_{k=1}^{K-1} W_n^{(k)} \leq dy \right),$$

which, using integration by parts, equals

$$-e^{-(a-y)} P_\infty \left( \sum_{k=1}^{K-1} W_n^{(k)} \geq y \right) \bigg|_{y=0}^{y=a} + \int_0^a P_\infty \left( \sum_{k=1}^{K-1} W_n^{(k)} \geq y \right) e^{-(a-y)} dy$$

$$= -P_\infty \left( \sum_{k=1}^{K-1} W_n^{(k)} \geq a \right) + e^{-a} + \int_0^a P_\infty \left( \sum_{k=1}^{K-1} W_n^{(k)} \geq y \right) e^{-(a-y)} dy.$$

Using the lemma for the case of $K - 1$ and combining all above relations, we have

$$P_\infty \left( \sum_{k=1}^{K} W_n^{(k)} \geq a \right) \leq e^{-a} + \int_0^a \left( e^{-(K-1)y} \sum_{j=0}^{K-1} \frac{y^j}{j!} \right) e^{-(a-y)} dy = e^{-a} \sum_{j=0}^{K-1} \frac{a^j}{j!},$$

which completes the induction step. Hence, the lemma holds for all integer $K \geq 1$. \qed

Lemma B2. For given $t > 0$, our proposed scheme $N_{sum}(a)$ in (9) satisfies

$$\lim_{a \to \infty} \inf P_\infty \{ N_{sum}(a) \geq \tau \} \geq \exp(-t),$$

where $B = B_a$ is defined in (B1).

Proof. Our proof is inspired by Iglehart (1972) and Siegmund & Venkatraman (1995). The key idea is to explore the renewal properties of the CUSUM statistics and to break the time axis $[1, tB]$ into some subintervals so that $W_n^{(k)}$s are independent, or approximately independent, between these subintervals. Then the lemma can be proved by applying Lemma B1 to each subinterval.

To prove the lemma rigorously, define stopping times $\tau_0 \equiv 0 < \tau_1 < \tau_2 < \cdots$ as follows: for $j = 0, 1, \ldots$, the stopping time $\tau_{j+1}$ is the smallest $n$ such that $n > \tau_j$ and

$$W_n^{(1)} = W_n^{(2)} = \cdots = W_n^{(K)} = 0,$$
and \( \tau_{j+1} = \infty \) if there is no such \( n \). Define \( S_{k,n} = X_{k,1} + X_{k,2} + \cdots + X_{k,n} \). Then

\[
W^{(k)}_n = S_{k,n} - \min_{i=1,\ldots,n} S_{k,i}, \quad W^{(k)}_{\tau_j} = 0.
\]

Hence an equivalent definition of \( \tau_{j+1} \) is the smallest \( n > \tau_j \) such that

\[
S_{k,n} = \min_{i=\tau_j+1,\ldots,n} S_{k,i} \quad (k = 1, \ldots, K).
\]

Therefore, the stopping times \( \tau_j \)'s correspond to the time when all \( K \) random walks \( S_{k,n} \)'s reach minimum record values simultaneously, and thus they can be thought of as ladder epochs in multi-dimensional random walks \( (S_{1,n}, \ldots, S_{k,n}) \). Let \( X_{[1,K],i} = (X_{1,i}, \ldots, X_{K,i}) \) denote the current vector observations at time \( i \). Then it is easy to see that the sequences

\[
\{ (\tau_j - \tau_{j-1}, X_{[1,K],\tau_{j-1}}, \ldots, X_{[1,K],\tau_j}) : j = 1, 2, \ldots \}
\]

are independent and identically distributed.

First, by (B2), \( \tau_j - \tau_{j-1} \) has the same distribution as \( \tau_1 \). Moreover, under \( P_\infty \), \( \log g_k(X_{k,n})/f_k(X_{k,n}) \) has negative mean by Assumption 2, implying that with probability 1, \( S_{k,n} \) drifts to \(-\infty\) for every \( k = 1, \ldots, K \). Then by the renewal theorem we proved in (A1), we have \( E_\infty(\tau_1) < \infty \).

Next, define by \( M_+(j) \) the maximum of \( \sum_{k=1}^{K} W_i^{(k)} \) over the \( j \)th subinterval \( [\tau_{j-1}, \tau_j) \) for \( j = 1, 2, \ldots \). Then by the definition of \( \tau_{j-1} \), it is evident that \( M_+(j) \) can be defined in terms of the vector variables in (B2), and the sequence \( \{M_+(j) : j \geq 1\} \) is also independent and identically distributed. Hence, for any real value \( a > 0 \),

\[
P_\infty \{M_+(j) \geq a\} = P_\infty \{M_+(1) \geq a\} = E_\infty \left[ P_\infty \left\{ \max_{i=1,1}^{K} \sum_{k=1}^{K} W_i^{(k)} \geq a | \tau_1 \right\} \right] \\
\leq E_\infty \left[ \sum_{i=1}^{\tau_1} P_\infty \left\{ \sum_{k=1}^{K} W_i^{(k)} \geq a \right\} \right] \leq E_\infty \left( \tau_1 \frac{1}{B} \right) = \frac{E_\infty(\tau_1)}{B},
\]

where \( B = B_a \) is defined in (B1). Here the first inequality follows from the fact that \( P(\cup A_i) \leq \sum P(A_i) \), and the second inequality follows from Lemma B1.

Now let us return to the proof of the lemma. Let \( \{l(n) : n \geq 0\} \) be the discrete renewal process associated with the independent and identically distributed sequence \( \{\tau_j - \tau_{j-1} : j \geq 1\} \), or more precisely, \( l(n) = 0 \) for \( n < \tau_1 \) and otherwise \( l(n) = \max \{k : \tau_k \leq n\} \). Then the maximum of the sum of the CUSUM statistics in the first \( n \) observations satisfies

\[
\max_{j=1,\ldots,l(n)} M_+(j) \leq \max_{i=1,\ldots,n} \sum_{k=1}^{K} W_i^{(k)} \leq \max_{j=1,\ldots,l(n)+1} M_+(j).
\]

Hence, given \( t \geq 0 \), we have

\[
P_\infty \{N_{\text{sum}}(a) \geq tB\} = P_\infty \left\{ \max_{i < B} \sum_{k=1}^{K} W_i^{(k)} < a \right\} \geq P_\infty \left\{ \max_{j=1,\ldots,l(B)+1} M_+(j) < a \right\} \\
= E_\infty \left[ P_\infty \left\{ \max_{j=1,\ldots,l(B)+1} M_+(j) < a \right| l(tB) \right\} \right] \\
= E_\infty \left[ P_\infty \{M_+(1) < a\} l(tB) + 1 \right] \geq E_\infty \left[ \left( 1 - \frac{E_\infty(\tau_1)}{B} \right) l(tB) + 1 \right] \\
= E_\infty \exp \left( l(tB) + 1 \right) \log \left\{ 1 - \frac{E_\infty(\tau_1)}{B} \right\}.
\]
example, Theorem 4.1 on p. 204 of Durrett (1996). Thus, as \( a \to \infty \),
\[
\exp \left[ \left\{ l(tB) + 1 \right\} \log \left\{ 1 - \frac{E_\infty(\tau_1)}{B} \right\} \right] \to \exp(-t) \quad \text{almost surely.}
\]
Furthermore, as a probability, this term is dominated by the constant 1. Hence, by the dominated convergence theorem, as \( a \) goes to \( \infty \), the last term of (B3) is bounded below by \( \exp(-t) \), completing the proof.

\[\Box\]

REFERENCES


[Received June 2009. Revised October 2009]