Quick Detection of Changes in Traffic Statistics: Application to Variable Rate Compression [1]

Ivy Hsu and Jean Walrand [2]

Department of EECS, University of California, Berkeley, CA 94720

presented at the 32nd Annual Allerton Conference on Communication, Control, and Computing, Urbana-Champion, IL, Sept. 1994

Abstract

In many communication applications, sources generate data at a variable rate. This variation can be captured by a two-level model. At a larger time scale, the source alternates between a finite number of models at the higher level, while each model represents the local fluctuation at the smaller time scale. Typically the data goes through a lossy compression at the encoder, and then is entered into a buffer before being transmitted over a fixed-rate or a regulated-rate (e.g., Leaky Bucket) channel. Since the buffer size is finite, the choice of compression parameters must balance the tradeoff between the distortion due to lossy compression and the distortion caused by buffer overflow. As the optimal compression parameters may be different for each individual model, a method is needed to detect changes from one model to another.

We show via examples that for many common traffic models, the problem of detecting changes in traffic statistics can be reduced to the following general change point detection problem: Let $A_1, A_2, \ldots$ be a series of independent observations such that $A_1, A_2, \ldots, A_{m-1}$ are i.i.d. distributed according to an unknown distribution F, and $A_m, A_{m+1}, \ldots$ are i.i.d. distributed according to another unknown distribution G. m is the unknown change point. The objective is to determine that a change has occurred as soon as possible, while maintaining a low rate of false alarms. In this paper we propose a sequential nonparametric change point detection algorithm using Kolmogorov-Smirnov statistics, and demonstrate that it is *asymptotically optimal*, in that under the constraint E[time until a false alarm] $\geq T$, our algorithm is such that E[detection delay] $= O(\log(T))$, as $T \to \infty$. This property is known to be the best that can be achieved. Furthermore, unlike other nonparametric schemes, the performance of the algorithm is independent of the distributions F and G.

1 Introduction

Consider communication applications such as MPEG video, in which the output stream generated by the source has a variable rate due to statistical variations in the data. The output stream is released onto a channel at a constant rate or at a regulated rate that allows some burstiness. Fixed bandwidth radio channels for wireless communication are

[1]This research is supported in part by a grant of the Networking and Communications Research Program of the National Science Foundation.

[2]Email addresses: ihsu@eecs.berkeley.edu, wlr@eecs.berkeley.edu

an example of the former. The Leaky Bucket scheme proposed for the Asynchronous Transfer Mode (ATM) networks represents the latter. In both cases, unless the transmission rate exceeds the peak rate of the source, it is necessary to place a buffer between the source encoder and the channel to store data when they are generated faster than what the channel can transmit.

First suppose that the source can be characterized by a time invariant model. Then for any given compression scheme, the quantities of interest are its mean rate and mean-square distortion measure for the given source model. Since the buffer size is finite, one must consider the tradeoff between this distortion measure and the distortion caused by lost data when the buffer overflows. A finer compression reduces distortion, but generates more data and thus leads to higher overflow probability.

Define the distortion-rate curve as the minimum distortion achievable by any scheme under an average rate constraint. Tse et al. [8] suggest that if the distortion-rate curve is concave at the channel rate R_c, then one can achieve a lower distortion by time-sharing between two compression schemes. Define $D_T(R_c)$ as the minimum distortion achievable by any time-sharing scheme under an average rate constraint R_c. Then the objective is for the buffer overflow probability and the steady state average distortion when the buffer is not full to approach zero and $D_T(R_c)$, respectively, as the buffer size becomes large. The control scheme proposed in [8] places a threshold at the half point of the buffer, and uses a coarse compression with average rate $R_1 < R_c$ when the buffer occupancy exceeds the threshold and a fine compression with rate $R_2 > R_c$ otherwise. Under this scheme, both the overflow probability and the average distortion approach their asymptotic values exponentially fast as the buffer size goes to infinity.

Under a more refined assumption, the source can be viewed to be alternating between a finite number of models, at a time scale much larger than the time scale of fluctuation captured by each individual model. For example, a video sequence can contain active and quiet scenes, with a mean scene duration of several seconds. The active scenes in general cannot be compressed as well as the quiet ones. Therefore the optimal parameters of the compression scheme would be different in each model.

Thus the encoder must be able to respond quickly when a change in the source model occurs and identify the new model. One approach is to start with a fixed set of source models and parameters for which the control parameters are known, and conduct a *parametric* detection, i.e., compare the incoming traffic with all existing models, and determine the model that is closest. Then apply the corresponding control scheme. This approach, although simple in nature, has a limited accuracy and requires extensive analyses of all possible traffic statistics. A more general approach is to make no prior assumption about source models, and adaptively determine the control parameters to use for each model. Thus the algorithm for change point detection must be *nonparametric*. The system maintains a library of past source statistics and corresponding control parameters. Once a change is detected, the new statistics are compared with entries in the library in order to determine if it belongs to one of the existing models. If so, the control parameters associated with the identified model are used. Otherwise, a new entry is created in the library, and the control parameters for the closest model are selected as its starting point. In both cases, these parameters are then adjusted over time, as the measured distortion is compared with the target.

As an example, Fig. 1 illustrates the block diagram of a possible control scheme

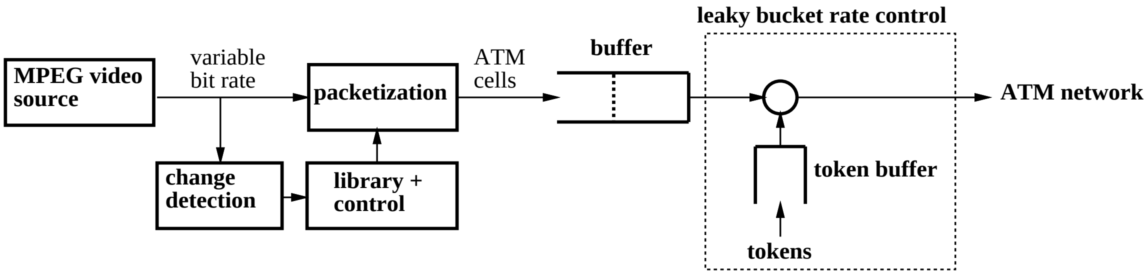

Figure 1: Transporting a variable rate MPEG sequence over an ATM network with rate constraint.

for variable rate video applications. The packetization block segments data generated by the video source into ATM cells of two priorities, in a way that incorporates the multiresolution principle: decoding with the high priority cells alone results in a coarser quality of video, while decoding the combined priorities results in a finer quality. The proportion of the two priorities is controlled based on the source statistics as mentioned above. The buffer operates under a *partial buffer sharing* mechanism [5], in which only high priority cells can enter the buffer when the occupancy is above a threshold. The data is emptied from the buffer through a leaky bucket before entering an ATM network.

In this paper we address the issue of detecting changes in traffic statistics. The general change point (or disorder) detection problem has undergone much investigation since 1950s. For the application at hand we are interested in a sequential nonparametric detection method. The main criterion of an effective sequential scheme is that it should respond to changes as soon as possible, while maintaining a low frequency of false alarms. In addition, it is desirable for the decision rule to be *distribution-free*, i.e., its effectiveness should not vary for different pairs of statistics before and after the change point. Finally, from the implementation point of view, the scheme should require only a bounded number of operations and memory at each decision point.

Several sequential nonparametric detection methods have been proposed previously. For a comprehensive review, see [2]. Most of these results are concentrated on the case in which the observed sequence is a concatenation of two i.i.d. random sequences, with some limited extension to n-dependent random sequences. Their common drawback, however, is the lack of the distribution-free property. For example, the detection scheme in [1] compares the empirical means before and after each decision point. Consequently the false alarm frequency depends on the variance of the distribution before the change.

In §2, we present a general detection algorithm based on comparisons of empirical distributions. We show via examples how this algorithm can be applied in the context of detecting changes in incoming traffic for many common models. The statistic used in this case is the Kolmogorov-Smirnov distance, which has the desirable property of being distribution-free [4]. This statistic has been used in [3] for the *a posteriori* change point problem, in which the emphasis is on accurate identification of the most likely point of change within a given set of observations. §3 presents the main theorem of the paper. We show that the ratio of the expected delay of a correct detection to the logarithm of the expected time until a false alarm is upper bounded by a constant, as the memory size of the algorithm increases. This logarithmic relation implies that the proposed scheme belongs to the class of asymptotically optimal algorithms, as indicated in [1, 6]. We present some preliminary numerical results in §4 and conclude in §5.

2 Change Point Detection Algorithm

2.1 Assumptions

Let $\{A_n\}$ be the observed sequence and m the unknown change point. Denote $1_{(\cdot)}$ as the indicator function. Assume that $A_n = \eta_n 1_{(n<m)} + \xi_n 1_{(n\geq m)}, n \geq 1$, where η_n's (resp. ξ_n's) are i.i.d. random variables with a continuous distribution F (resp. G). Furthermore, assume that there exists a constant $\varepsilon > 0$ such that the set $\mathcal{A} = \{x \in \mathbf{R} : |F(x) - G(x)| \geq \varepsilon\}$ satisfies either $\int_{\mathcal{A}} dF(x) > 0$ or $\int_{\mathcal{A}} dG(x) > 0$.

At the first glance the i.i.d. assumption may appear restrictive. But as the following examples demonstrate, A_n can be derived in many complex traffic models:

Example 1 Markov Modulated Sources *One common model for bursty sources is the Markov modulated fluid, in which the instantaneous arrival rate from a source depends on the state of an underlying Markov chain. Consider, for example, an ON-OFF process, in which the Markov chain alternates between two states, and the source generates data at a constant rate when the chain is ON and no data otherwise. The source statistics can differ when there is a change in the transition rates of the Markov chain, or in the output rate, or both. Choose the regenerative points t_n as the successive times at which the process returns to the OFF state. Then define $A_n \equiv (\text{total output in } [t_n, t_{n+1}))/(t_{n+1} - t_n)$.*

Example 2 Periodic Sources *Periodic sources arise when input data is encoded at fixed time intervals (e.g., PCM of voice, frame-by-frame encoding of video). Thus A_n can be taken as the data generated over any constant multiple of a period. §4 gives a more complex example in the case of MPEG video, where the source can be treated as a superposition of several periodic streams.*

2.2 Algorithm

The algorithm is characterized by three parameters: the memory size N, the windowing parameter a, and the threshold c. At each time n, only the last N observations need to be stored in memory. The parameter a is a constant in $(0, \frac{1}{2}]$. For simplicity in notation, let a be chosen such that $aN \in \mathcal{N}$. Given the memory of N points, a defines a window such that each k in $\{aN, aN+1, \ldots, (1-a)N\}$ is a candidate change point. Define, for $n \geq N$ and $k \in [aN, (1-a)N]$,

$$D_N(k,n) = \max_{n-N+1 \leq i \leq n} \left| \frac{1}{k} \sum_{j=n-N+1}^{n-N+k} 1_{(A_j \leq A_i)} - \frac{1}{N-k} \sum_{j=n-N+k+1}^{n} 1_{(A_j \leq A_i)} \right|, \qquad (1)$$

$$E_N(n) = \max_{aN \leq k \leq (1-a)N} D_N(k,n). \qquad (2)$$

Let the decision rule be $d_N(n) = 1_{(E_N(n) > c)}$, where $0 < c < 1$. The detection time is then $\tau(N) = \min\{n \geq N : d_N(n) = 1\}$.

The algorithm can be interpreted as follows: $D_N(k,n)$ is the Kolmogorov-Smirnov distance, namely the maximum absolute difference, between the empirical distributions of two sequences, $(A_{n-N+1}, \ldots, A_{n-N+k})$ and $(A_{n-N+k-1}, \ldots, A_n)$. For each $aN \le k \le (1-a)N$, we calculate the Kolmogorov-Smirnov distance and declare that a change has occurred if at any point k the distance exceeds the threshold c.

3 Theorem

Let $P_\infty(\cdot)$ (resp. $E_\infty(\cdot)$) represent the conditional probability (resp. expectation) given that no change point ever occurs. Similarly, let $P_m(\cdot)$ (resp. $E_m(\cdot)$) be the conditional probability (resp. expectation) given that change occurs at time m. As mentioned earlier, we are interested in two quantities:

(1) $E_\infty[\tau(N) - N] \equiv$ expected time until a false alarm

(2) $E_m[(\tau(N) - m)|\tau(N) \ge m] \equiv$ expected detection delay

The objective is to detect the change as fast as possible, while maintaining a large expected time between false alarms. The following theorem relates the two quantities under the above algorithm.

Theorem 1 *Given that*

$$\varepsilon \ge c + \left(\frac{-1}{2a(1-a)} \log(1 - c/2) \right)^{1/2}, \tag{3}$$

$$\limsup_{N\to\infty} \sup_{m \ge N} \frac{E_m[(\tau(N) - m)|\tau(N) \ge m]}{\log E_\infty[\tau(N) - N]} \le Constant. \tag{4}$$

It is pointed out in [1, 6] that for the problem of minimizing $\sup_m E_m[(\tau(N) - m)|\tau(N) \ge m]$ under the constraint $E_\infty[\tau(N) - N] \ge T$, the optimal sequential detection rule admits an asymptotic behavior $\sup_m E_m[(\tau(N) - m)|\tau(N) \ge m] = O(\log(T))$, as $T \to \infty$. Thus by virtue of theorem 1, the proposed algorithm is asymptotically optimal.

Theorem 1 follows from Lemmas 2 and 3 below. We first derive an upper bound on the probability of false alarm at time $N + k$, for all $k \ge 0$.

Lemma 1

$$P_\infty(\tau(N) - N = k) \le L_1(N) N e^{-L_2(N)N}, \quad \forall k \ge 0, \tag{5}$$

where $L_1(N)$ and $L_2(N)$ are defined below.

Proof Define $\beta_N(n) = P_\infty(d_N(n) = 1)$ for all $n \ge N$, then $P_\infty(\tau(N) - N = k) \le \beta_N(N + k)$. Given that no change takes place, the samples are i.i.d. $\sim F$. Thus $\beta_N(n) \equiv \beta_N(N)$, for all $n \ge N$. Then

$$\beta_N(N) = P_\infty \left(\max_{aN \le k \le (1-a)N} D_N(k, N) > c \right) \le \sum_{aN \le k \le (1-a)N} P_\infty(D_N(k, N) > c). \tag{6}$$

$$
\begin{aligned}
& P_\infty(D_N(k,N) > c) \\
\overset{(a)}{\leq} \;& P_\infty\left(\max_{1\leq i\leq N} |\frac{1}{k}\sum_{j=1}^{k} 1_{(A_j\leq A_i)} - F(A_i)| + |\frac{1}{N-k}\sum_{j=k+1}^{N} 1_{(A_j\leq A_i)} - F(A_i)| > c \right) \\
\overset{(b)}{\leq} \;& P_\infty\left(\max_{1\leq i\leq N} |\frac{1}{k}\sum_{j=1}^{k} 1_{(A_j\leq A_i)} - F(A_i)| + \max_{1\leq i\leq N} |\frac{1}{N-k}\sum_{j=k+1}^{N} 1_{(A_j\leq A_i)} - F(A_i)| > c \right) \\
\overset{(c)}{\leq} \;& P_\infty\left(\sup_x |\frac{1}{k}\sum_{j=1}^{k} 1_{(A_j\leq x)} - F(x)| > c/2 \right) + \\
+ \;& P_\infty\left(\sup_x |\frac{1}{N-k}\sum_{j=k+1}^{N} 1_{(A_j\leq x)} - F(x)| > c/2 \right), \qquad (7)
\end{aligned}
$$

where (a) is due to the triangular inequality, and (b) follows from the fact that the maximum of the sum is less than or equal to the sum of the maxima. (c) results from $P(X+Y>c) \leq P(X>c/2 \text{ or } Y>c/2) \leq P(X>c/2) + P(Y>c/2)$, and that the maximum distance between an empirical distribution and the actual distribution at the sample points is upper bounded by the supremum distance.

$D_k \equiv \sup_x |\frac{1}{k}\sum_{j=1}^{k} 1_{(A_j\leq x)} - F(x)|$ is the Kolmogorov-Smirnov one-sample statistic based on k sample points. It can be expressed as the maximum of two non-negative one-sided Kolmogorov-Smirnov statistics D_k^+ and D_k^-, defined as

$$
D_k^+ \equiv \sup_x \left[\frac{1}{k}\sum_{j=1}^{k} 1_{(A_j\leq x)} - F(x)\right] = \max_{1\leq i\leq k}\left[\frac{i}{k} - F(A_{(i)}^k)\right], \qquad (8)
$$

and

$$
D_k^- \equiv \sup_x \left[F(x) - \frac{1}{k}\sum_{j=1}^{k} 1_{(A_j\leq x)}\right] = \max_{1\leq i\leq k}\left[F(A_{(i)}^k) - \frac{i-1}{k}\right], \qquad (9)
$$

where $A_{(1)}^k < A_{(2)}^k < \cdots < A_{(k)}^k$ are the order statistics of $A_1,\ldots,A_k$. The superscript represents the total number of sample points.

Since F is continuous, by the probability-integral transformation theorem, given a random variable $A \sim F$, the random variable U generated by the transformation $U = F(A)$ has a Uniform(0, 1) distribution. Therefore, $U_{(i)}^k = F(A_{(i)}^k)$ is the i^{th} order statistic of k i.i.d. Uniform(0, 1) random variables. Thus for any fixed $\epsilon > 0$,

$$
P(\max_{1\leq i\leq k}\frac{i}{k} - F(A_{(i)}^k) > \epsilon) = P(\max_{1\leq i\leq k}\frac{i}{k} - U_{(i)}^k > \epsilon) \leq \sum_{i=1}^{k} P(U_{(i)}^k < \frac{i}{k} - \epsilon). \qquad (10)
$$

Note that for $i/k < \epsilon$, $P(U_{(i)}^k < \frac{i}{k} - \epsilon) = 0$. For $i/k \geq \epsilon$, $P(U_{(i)}^k < \frac{i}{k} - \epsilon)$ is the probability that out of k i.i.d. Uniform(0, 1) random variables, at least i of them are less than $i/k - \epsilon$. Let $W_j \equiv 1(U_j < i/k - \epsilon)$. Then $P(U_{(i)}^k < \frac{i}{k} - \epsilon) = P(\sum_{j=1}^k W_j \geq i)$, which is the probability that the sample average is greater than i/k, while $E[W_1] = i/k - \epsilon$.

Define $\varphi(\theta) \equiv \log Ee^{\theta W_1}$. Then by Markov inequality, $P(\sum_{j=1}^k W_j \geq \frac{i}{k}\cdot k) \leq \exp(-k\cdot\frac{i}{k}\theta)\exp(k\cdot\varphi(\theta))$. In particular, $P(\sum_{j=1}^k W_j \geq \frac{i}{k}\cdot k) \leq \exp(-k\sup_\theta[\frac{i}{k}\theta - \varphi(\theta)]) \equiv \exp(-kI_+(i,k,\epsilon))$, where

$$
I_+(i,k,\epsilon) = \begin{cases} \frac{i}{k}\log\left(\frac{i/k}{i/k-\epsilon}\right) + (1-\frac{i}{k})\log\left(\frac{1-i/k}{1-i/k+\epsilon}\right), & \text{for } \frac{i}{k} > \epsilon \\ \infty, & \text{for } \frac{i}{k} \leq \epsilon. \end{cases} \qquad (11)
$$

Similarly, we can show that $P(U^k_{(i)} - \frac{i-1}{k} > \epsilon) \leq \exp(-kI_-(i,k,\epsilon))$, and

$$I_-(i,k,\epsilon) = \begin{cases} \frac{i-1}{k}\log\left(\frac{(i-1)/k}{(i-1)/k+\epsilon}\right) + (1-\frac{i-1}{k})\log\left(\frac{1-(i-1)/k}{1-(i-1)/k-\epsilon}\right), & \text{for } \frac{i-1}{k} < 1-\epsilon \\ \infty, & \text{for } \frac{i-1}{k} \geq 1-\epsilon. \end{cases} \tag{12}$$

Define $I_N(\epsilon) = \min_{aN\leq k\leq(1-a)N}\min_{1\leq i\leq k}\min\{I_+(i,k,\epsilon), I_-(i,k,\epsilon)\}$. Then $P_\infty(D_k > c/2) \leq P(D^+_k > c/2) + P(D^-_k > c/2) \leq 2k\cdot\exp(-kI_N(c/2))$. Substituting into Eqn. 6,

$$\begin{aligned} \beta_N(N) &\leq \sum_{aN\leq k\leq(1-a)N} P_\infty(D_k > c/2) + P_\infty(D_{N-k} > c/2) \\ &\leq \sum_{aN\leq k\leq(1-a)N} 2\left[k\exp(-kI_N(c/2)) + (N-k)\exp(-(N-k)I_N(c/2))\right] \\ &\leq \frac{4(1-a)}{1-\exp(-I_N(c/2))} N e^{-aI_N(c/2)N} \equiv L_1(N)Ne^{-L_2(N)N}. \end{aligned} \tag{13}$$

Note that if $P_\infty(\tau(N)-N=j) \leq \beta$ for all $j \geq 0$, then $E_\infty[\tau(N)-N]$ is minimized if

$$P_\infty(\tau(N)-N=j) = \begin{cases} \beta, & j \leq j^* = \lfloor 1/\beta \rfloor, \\ 1-j^*\beta, & j = j^*+1, \\ 0, & \text{otherwise.} \end{cases}$$

Thus $E_\infty[\tau(N)-N] \geq \frac{1}{2\beta}(1-o(1))$, where $o(1)\to 0$ as $\beta \to 0$. Since $\beta_N(N)\to 0$ as $N\to\infty$, applying lemma 1, we obtain a lower bound on the expected time until a false alarm:

Lemma 2

$$E_\infty[\tau(N)-N] \geq \frac{1-\exp(-I_N(c/2))}{8(1-a)N} e^{aI_N(c/2)N}(1-o(1)), \tag{14}$$

where $o(1)\to 0$ *as* $N\to\infty$.

Before giving the upper bound on the expected detection delay, we first state the Hoeffding's inequality:

Theorem 2 (Hoeffding's Inequality) *[7] Let* $Y_1, Y_2, \ldots, Y_N$ *be independent random variables with zero means and bounded ranges:* $a_i \leq Y_i \leq b_i$*. For each* $\eta > 0$*,*

$$P(|Y_1+\cdots+Y_N| \geq \eta) \leq 2\exp\left[\frac{-2\eta^2}{\sum_{i=1}^N (b_i-a_i)^2}\right]. \tag{15}$$

Lemma 3 *Given condition 3,*

$$E_m[\tau(N)-m|\tau(N)\geq m] \leq N\left(1+\frac{2}{1-aN\beta_N}\right). \tag{16}$$

Proof Define $p_N \equiv P_m(d_N(m+(1-a)N-1)=0)$, and $q_N \equiv P_m(d_N(m+N)=1)$. Then the following inequality is given in [1]:

$$E_m[\tau(N)-m|\tau(N)\geq m] \leq N\left(1+\frac{p_N/q_N}{1-aN\beta_N}\right). \tag{17}$$

Denote $F_i(\cdot)$ (resp. $G_i(\cdot)$) as the empirical distribution function generated by i i.i.d. samples with distribution F (resp. G). Then

$$D_N(k,m+(1-a)N-1) = \sup_x |\frac{aN}{k}F_{aN}(x)+\frac{k-aN}{k}G_{k-aN}(x)-G_{N-k}(x)|. \tag{18}$$

Therefore, for any fixed $x_0 \in \mathcal{A}$,

$$p_N \leq P(\sup_x |F_{aN}(x)-G_{(1-a)N}(x)| \leq c) \leq P(|F_{aN}(x_0)-G_{(1-a)N}(x_0)| \leq c). \tag{19}$$

By the triangular inequality,

$$|F_{aN}(x_0)-G_{(1-a)N}(x_0)-F(x_0)+G(x_0)|+|F_{aN}(x_0)-G_{(1-a)N}(x_0)| \geq |F(x_0)-G(x_0)| > \varepsilon. \tag{20}$$

Since the condition implies that $\varepsilon \geq c$,

$$\begin{aligned} & p_N \leq P(|F_{aN}(x_0)-F(x_0)+G(x_0)-G_{(1-a)N}(x_0)| > \varepsilon - c) \\ = \; & P(|\frac{1}{a}\sum_{j=1}^{aN}[1_{(\eta_j\leq x_0)}-F(x_0)]+\frac{1}{1-a}\sum_{j=aN+1}^{N}[G(x_0)-1_{(\xi_j\leq x_0)}]| > N(\varepsilon-c)). \end{aligned} \tag{21}$$

Applying the Hoeffding's Inequality to Eqn. 21, we get

$$b_j = \begin{cases} \frac{1}{a}[1-F(x_0)], & \text{for } 1\leq j\leq aN, \\ \frac{1}{1-a}[G(x_0)], & \text{for } aN+1\leq j\leq N, \end{cases}$$

$$a_j = \begin{cases} \frac{1}{a}[-F(x_0)], & \text{for } 1\leq j\leq aN, \\ \frac{1}{1-a}[G(x_0)-1], & \text{for } aN+1\leq j\leq N. \end{cases}$$

Thus $\sum_{j=1}^N (b_j-a_j)^2 = aN/a^2+(1-a)N/(1-a)^2 = N/a(1-a)$, and $p_N \leq 2\exp[-2Na(1-a)(\varepsilon-c)^2]$.

On the other hand, we obtain the following lower bound for q_N:

$$\begin{aligned} q_N &\geq P(\sup_x |G_{aN}(x)-G(x)+G(x)-G_{(1-a)N}(x)| > c) \\ &\geq P(\sup_x(G_{aN}(x)-G(x)) - \sup_x(G_{(1-a)N}(x)-G(x)) > c) \\ &\geq P(\sup_x(G_{aN}(x)-G(x)) > c/2)\cdot P(\sup_x(G_{(1-a)N}(x)-G(x)) < -c/2) \\ &= P(D^+_{aN} > c/2)\cdot P(D^-_{(1-a)N} > c/2) \\ &= P(\max_{1\leq i\leq aN}\frac{i}{aN}-U^{aN}_{(i)} > c/2)\cdot P(\max_{1\leq i\leq (1-a)N} U^{(1-a)N}_{(i)} - \frac{i-1}{(1-a)N} > c/2) \\ &\geq P(U^{aN}_{(aN)} < 1-c/2)\cdot P(U^{(1-a)N}_{(1)} > c/2) = (1-c/2)^N. \end{aligned} \tag{22}$$

$p_N \leq 2q_N$ for all N if $e^{-2a(1-a)(\varepsilon-c)^2} \leq 1-c/2$, or equivalently, if condition 3 holds.

Note that $I_N(c/2) \geq \inf_{x>c/2}[x\log\frac{x}{x-c/2}+(1-x)\log\frac{1-x}{1-x+c/2}], \forall N$, and thus Theorem 1 follows.

4 Numerical Example

In the variable-rate video coding schemes such as MPEG, the number of bits generated by each frame depends on both its information content and the encoding technique used. In MPEG, the choices of encoding modes are intraframe (I), predictive (P), and interpolative (B), in the order of increasing coding efficiency. A typical sequence consists of periodic interleaving of frames from each mode. For example, the sequence in our example has a period of p =15 frames in the following pattern: {I, B, B, P, B, B, P, B, B, P, B, B, P, B, B}.

One approach is to define A_n as the total amount of data generated over a 15-frame period. However, using the knowledge of the framing pattern, one can expect to achieve a faster detection by modifying the algorithm as follows. First fix the windowing parameter $a = 0.5$ and the memory size N = even multiple of p. We can treat frames from the three modes as three separate streams. Define $D_N^i(aN, n)$ as the signed Kolmogorov-Smirnov distance of the i^{th} stream, $i \in \{I, P, B\}$. Let $E_N(n)$ be the absolute value of the weighted average of the three distances. That is, $E_N(n) = |\frac{1}{15}D_N^I(aN, n) + \frac{4}{15}D_N^P(aN, n) + \frac{10}{15}D_N^B(aN, n)|$.

Fig. 2 illustrates the output of a broadcast-quality MPEG sequence, in Kb/frame. Note that the data rates of I, P, and B frames form three distinct groups. We apply the algorithm with memory size $N = 120$ and threshold $c = 0.75$. The circles at the top of the figure indicate the resulting detection points. They correspond well with visual partitions of the sequence. Our preliminary experiment using the scheme in Fig. 1 shows that changing the proportion of the two priorities at these points result in a smaller distortion than a fixed-priority policy.

5 Conclusion

In this paper we proposed a sequential change point detection algorithm that requires no assumption of the distributions before and after the change. We proved that it is asymptotically optimal as the memory size becomes large. The algorithm is general enough to be applicable in many contexts. It is presented here for the application of detecting changes in traffic models for variable-rate compression. By catering the compression parameters to each traffic model, the system can achieve better utilization of the resources.

Acknowledgements The authors are grateful to Profs. Peter Glynn and Vivek Borkar for their suggestions and Dr. Daniel Reininger for supplying the MPEG sequence.

References

[1] B. S. Darkovsky and B. E. Brodsky (1987) "A Nonparametric Method for Fastest Detection of a Change in the Mean of a Random Sequence", *Theory of Probability and its Applications*, 1987, pp. 640-648.

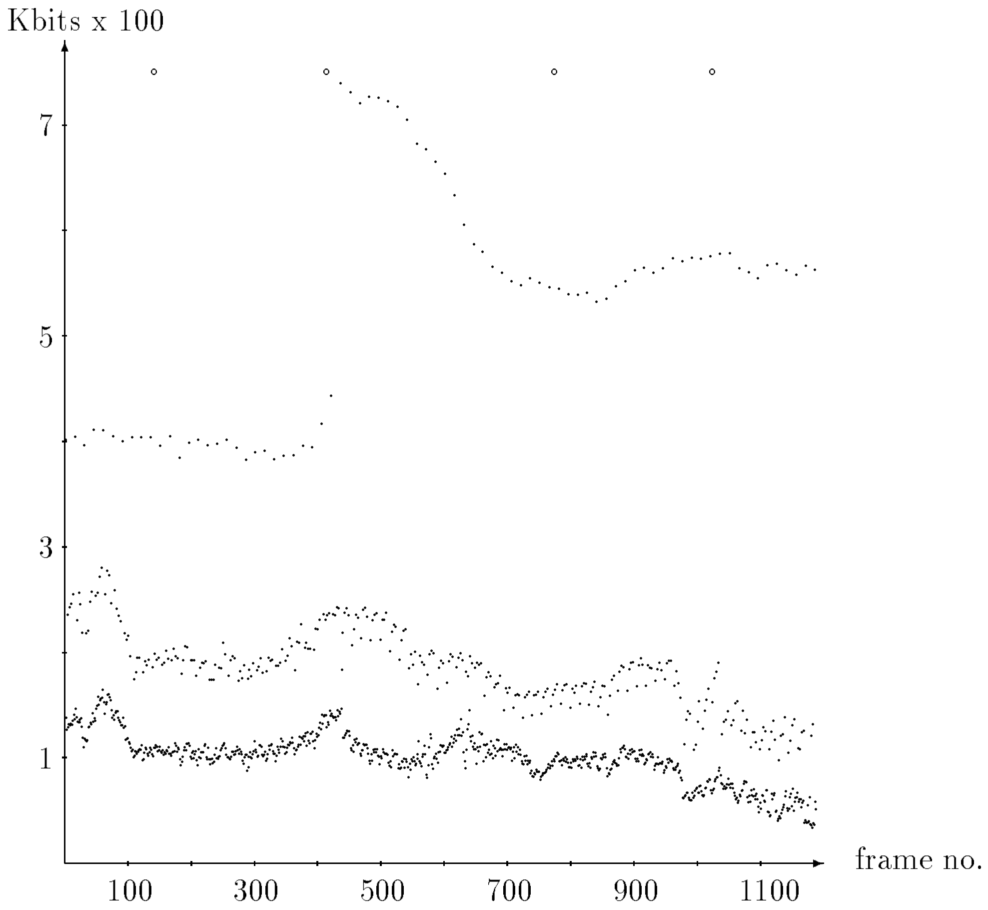

Figure 2: Output per frame of an MPEG sequence. The open circles represent the detection points.

[2] B. E. Brodsky and B. S. Darkhovsky (1993) *Nonparametric Methods in Change-Point Problems* Kluwer Academic Publishers.

[3] E. Carlstein (1988) "Nonparametric Change-Point Estimation", *Annals of Statistics*, 1988, vol. 16, no. 1, pp. 188-197.

[4] J. D. Gibbons and S. Chakraborti (1992) *Nonparametric Statistical Inference*, Marcel Dekker.

[5] H. Kroner, G. Hebuterne, P. Boyer, and A. Gravey (1991) "Priority Management in ATM Switching Nodes", *IEEE Journal on Selected Areas in Communications*, April 1991, vol. 9, no. 3, pp. 418-427.

[6] M. Pollak (1985) "Optimal Detection of a Change in Distribution", *Annals of Statistics*, 1985, pp. 206-227.

[7] D. Pollard (1984) *Convergence of Stochastic Processes*, Springer-Verlag.

[8] D. Tse, R. Gallager and J. Tsitsiklis (1993) "Optimal Buffer Control for Variable-Rate Lossy Compression", *31st Allerton Conference.*