This article has been accepted for publication in IEEE Signal Processing Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LSP.2021.3075919

Sequential Nonparametric Detection of Anomalous Data Streams

Sreeram C. Sreenivasan and Srikrishna Bhashyam

***Abstract*—We study a nonparametric search problem to detect L anomalous streams from a finite set of S data streams. The L anomalous streams are real-valued independent and identically distributed (i.i.d.) sequences drawn from the distribution q, while the remaining $S-L$ data streams are i.i.d. sequences drawn from the distribution p. The distributions p and q are assumed to be *arbitrary* and *unknown*, but distinct. We consider two cases: one where $L=1$, and the other where $0 \leq L \leq A$. In both cases, we propose universal distribution-free sequential tests that are consistent. For the first case, we also: (1) show that the test is universally exponentially consistent and stops in finite time almost surely, and (2) bound the limiting growth rate of the expected stopping time as the probability of error decreases to zero. Simulations show that the performance of the proposed test is better than that of the fixed sample size test.**

I. Introduction

Algorithms for detection of anomalous data streams find applications in search problems, sensor networks, fraud detection, environmental monitoring, etc. [4], [8]. In this work, we consider the anomalous or outlier hypothesis testing problem of detecting L anomalous data streams among a finite collection of S data streams. The typical data streams are i.i.d. sequences drawn according to p, while the L anomalous data streams are i.i.d. sequences drawn according to q. Various results in the setting where p and q are known are discussed in [7], [8]. In this work, we focus on the nonparametric setting where no information is known about either p or q, except that they are distinct. When $S=2$, the problem reduces to the two-sample testing problem in [3].

Fixed sample size (FSS) generalized likelihood-based tests were derived in [2], [5] for the case when p and q are discrete distributions over finite alphabets. The error exponents were analyzed and the tests were shown to be universally consistent. In [1], continuous distributions were considered, but the typical distribution p is assumed to be known. FSS tests were derived for the more general case of unknown p and q without the finite alphabet restriction in [11]. The decision statistic is based on the maximum mean discrepancy (MMD) introduced in [3].

Universal *sequential* outlier hypothesis tests were proposed in [6] for the case of discrete p and q over finite alphabets. The tests were shown to be universally consistent and to stop in finite time almost surely. In this letter, we propose an universal sequential outlier hypothesis test for the hitherto open general nonparametric case of unknown arbitrary distributions without the finite alphabet restriction. Our test also uses the MMD in [3], [11]. We propose *distribution-free consistent sequential* tests for two cases: (1) $L=1$[1], and (2) $0 \leq L \leq A < S/2$, where A is a given constant. For the $L=1$ case, the proposed test is also universally exponentially consistent, and stops almost surely in finite time. The limiting growth rate of the expected delay, as the probability of error decreases to zero, is also bounded. While our tests use the MMD in [3], our work is different: we consider the sequential setting, generalize the method in [6] to detect anomalous streams in continuous valued data streams, use a max-min MMD distance between groups of sequences to do this, and, furthermore, the test and consistency analysis depends on the assumptions on L.

This work was supported by the Science and Engineering Research Board, Department of Science and Technology [grant no. EMR/2016/002503].
Sreeram C. Sreenivasan and Srikrishna Bhashyam are with the Department of Electrical Engineering, IIT Madras, Chennai, India.
The authors acknowledge fruitful discussions with Rajesh Sundaresan.

II. Problem Statement

Consider the following anomaly detection problem. There are S data streams $\{X_1, \ldots, X_S\}$, each being a sequence of i.i.d. samples drawn from either p (typical distribution) or q (anomalous distribution). Streams X_i and X_j are independent for $i \neq j$. The distributions p and q are unknown and arbitrary. The number of anomalous streams is denoted by L. We consider two cases: $L=1$, and $0 \leq L \leq A < S/2$, where A is a given constant. The set of indices of the anomalous streams is denoted by $\mathscr{I}$. Our objective is to design a consistent, universal sequential test to detect the indices of the anomalous streams with as few samples as possible for a given accuracy. A *universal* test is distribution-free, i.e., it works for every p and q. In the *sequential* setting, at each time n, one sample from each stream is available to the detection algorithm. At each instant, a sequential test decides whether to stop sampling and make a decision or continue sampling. When the test stops it outputs a decision δ for the set of indices of the anomalous streams. The maximal error probability of a test P_{max} is the maximum of $\mathbb{P}_{\mathscr{I}}[\delta \neq \mathscr{I}]$ over $\mathscr{I}$, where $\mathbb{P}_{\mathscr{I}}$ denotes the probability measure under the hypothesis that the set of indices of the anomalous streams is $\mathscr{I}$ (When $\mathscr{I}$ is the null set, we use the notation $\mathbb{P}_0$). A sequence of tests is said to be *universally consistent* if the maximal error probability goes to zero for any distinct pair of distributions p and q, and *universally exponentially consistent* if $P_{\max}$ decays exponentially in expected stopping time [6].

Our test should distinguish the anomalous streams using a finite number of samples from each stream without knowning p and q. To do this, we rely on the maximum mean discrepancy (MMD) formulated in [3]. The MMD can be estimated

[1]The $L=1$ case is closely related to the *slippage* problem studied in statistics under some paramteric settings. See [6] for more details regarding this connection.

This article has been accepted for publication in IEEE Signal Processing Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LSP.2021.3075919

effectively from samples and the estimates converge to the true MMD. Furthermore, MMD-based tests have been observed to be more effective that several others in [11] for the FSS setting. Consider a class of measurable functions $\mathscr{F}$, to be defined soon, with a common metric domain $(\mathscr{X}, d)$ and range $\mathbb{R}$. Also consider distributions p and q on $\mathscr{X}$ $(p \neq q)$ and $\mathscr{X}$-valued random variables X, Y distributed according p and q, respectively. The MMD is defined as:

$$\text{MMD}\,[p,q] = \sup_{f \in \mathscr{F}} \left(\mathbb{E}_p\left[f(X)\right] - \mathbb{E}_q\left[f(Y)\right]\right),$$

where $\mathbb{E}_p$ and $\mathbb{E}_q$ denote the expectation under the probability distributions p and q, respectively. In [3], the authors propose the unit ball in a reproducing kernel Hilbert space $\mathscr{H}$ as their choice of $\mathscr{F}$. With this choice for $\mathscr{F}$, they also give finite-sample estimators for the MMD. We use the unbiased estimator M_u^2 of the square of MMD in our work: Given two sequences $X_i^n = \{x_{i1}, x_{i2}, \cdots, x_{in}\}$, and $X_j^n = \{x_{j1}, x_{j2}, \cdots, x_{jn}\}$, each of n i.i.d samples, the unbiased estimator of the square of the MMD is given by [3, Lemma 6]:

$$M_u^2(i,j,n) = \frac{1}{n(n-1)} \sum_{l \neq m} h(z_l, z_m),$$

where $z_l = (x_{il}, x_{jl})$, $h(z_l, z_m) = k(x_{il}, x_{im}) + k(x_{jl}, x_{jm}) - k(x_{il}, x_{jm}) - k(x_{im}, x_{jl})$, and $k(x,y)$ is the kernel function[2]. This estimate $M_u^2(i,j,n+1)$ can be computed sequentially from $M_u^2(i,j,n)$ as:

$$\frac{n-1}{n+1} M_u^2(i,j,n) + \frac{1}{n(n+1)} \sum_{l=1}^{n} [h(z_{n+1}, z_l) + h(z_l, z_{n+1})].$$

The sequential update requires $O(n)$ computations while full computation requires $O(n^2)$ computations. Suppose X_i^n is i.i.d. according to p and X_j^n is i.i.d. according to q. From the large deviation bound in [3, Lemma 10] for $M_u^2(i,j,n)$ and the summability of this bound, we get almost sure convergence of $M_u^2(i,j,n)$ to $\text{MMD}^2(p,q)$.

III. Proposed Nonparametric Sequential Tests

In this section we present our proposed nonparametric sequential tests. We use $\mathscr{S} = \{1, \ldots, S\}$ and $\mathscr{I}^c = \mathscr{S} \setminus \mathscr{I}$. Let $\mathscr{S}_A$ denote the set of all possible (non-empty) anomalous subsets $\mathscr{I}$. For example, when $L = 1$, $\mathscr{S}_A$ is the set of single element subsets of $\mathscr{S}$. At each time $n \geq 2$, our proposed test identifies a candidate anomalous subset as

$$\hat{\mathscr{I}}_n = \underset{\mathscr{A} \in \mathscr{S}_A}{\operatorname{argmax}} \left\{ \min_{i \in \mathscr{A}} \min_{j \in \mathscr{S} \setminus \mathscr{A}} M_u^2(i,j,n) \right\}, \tag{1}$$

i.e., for each subset $\mathscr{A}$, the minimum distance between a stream in $\mathscr{A}$ and another in $\mathscr{S} \setminus \mathscr{A}$ is determined. Then, the subset of streams which has the maximum such minimum distance is chosen as the candidate anomalous subset. Thus, we employ the max-min distance

$$\Gamma_n = \min_{i \in \hat{\mathscr{I}}_n} \min_{j \in \mathscr{S} \setminus \hat{\mathscr{I}}_n} M_u^2(i,j,n), \tag{2}$$

as our test statistic. We propose the test NP-SEQ(α, T_0) given by the following stopping and decision rules:

$$N = \min\{\tilde{N}, T_0\}, \tag{3}$$

$$\delta = \begin{cases} \hat{\mathscr{I}}_{\tilde{N}}, & \tilde{N} < T_0 \\ \varnothing, & \text{otherwise} \end{cases}, \tag{4}$$

where N is the stopping time, $\varnothing$ denotes the null set, T_0 is a time-out parameter and $\tilde{N}$ is given by $\tilde{N} = \operatorname{argmin}_{n \geq 2} \{\Gamma_n > T_n\}$, and $T_n = C/n^\alpha$ is a time-varying threshold with $C \geq C_0 > 0$ and $\alpha > 0$, where C_0 is a strictly positive constant.

A. Exactly one anomaly ($L = 1$)

For this case, we have S hypotheses with hypothesis i corresponding to $\mathscr{I} = \{i\}$, for $1 \leq i \leq S$. We set $\alpha = 1$ and $T_0 = \infty$ in NP-SEQ(α, T_0). We denote this test NP-SEQ-1. Observe that the max-min distance converges almost surely to $\text{MMD}^2(p,q) > 0$. This, along with the choice of monotonically-decreasing T_n, ensures that the test stops in finite time almost surely.

Theorem 1: The stopping time N is almost surely finite, i.e., under each hypothesis i, $1 \leq i \leq S$, $\mathbb{P}_i\left[N < \infty\right] = 1$.

Proof: See section VI-A of the appendix. ∎

Theorem 2: The test is universally consistent. That is, for all distinct p, q, the threshold $T_n = C/n$ is independent of p and q, and the error probability satisfies $\lim_{C \to \infty} P_{\text{max}} = 0$.

Proof: See section VI-B of the appendix. ∎

To show universal consistency, we show that the error probability goes to zero for a sequence of tests with threshold value increasing to infinity.

Remark 1: Theorems 1 and 2 also hold for a sequential test with threshold C/n^α for any $\alpha > 0$ (by proper choice of n_0 to satisfy (8)). We observed in the performance simulations that the choice of $\alpha = 1$ gives the best performance in terms of expected delay.

Theorem 3: Under each hypothesis i, the stopping time N satisfies

$$\lim_{C \to \infty} \mathbb{E}_i \left[\left| \frac{N}{C} - \frac{1}{\text{MMD}^2(p,q)} \right| \right] = 0,$$

where $\mathbb{E}_i$ denotes expectation under hypothesis i. Consequently, we have, for each hypothesis i and any p, q, that

$$\mathbb{E}_i\left[N\right] \leq \frac{-32K^2 \log P_{\max}}{\text{MMD}^4(p,q)} (1 + o(1)). \tag{5}$$

Thus, the test is universally exponentially consistent.

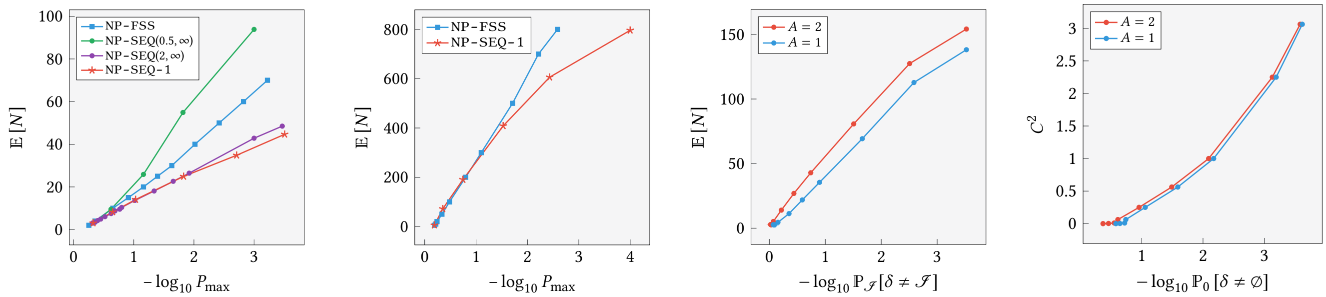
Proof: See section VI-C of the appendix. ∎

The proof uses the almost sure convergence of the estimate of the MMD^2 to the true MMD^2 as $n \to \infty$, and the observation that $N \to \infty$ as $C \to \infty$. From (5), the growth rate of expected stopping time, i.e., $\frac{\mathbb{E}_i[N]}{\log(1/P_{\max})}$, is bounded by $32K^2/\text{MMD}^4(p,q)$ asymptotically (for small $P_{\max}$).

Note that the case where the number of anomalous sequences is one or zero is a special case of the next subsection with A set to 1.

[2]We assume that the kernel function is bounded by K, i.e., $0 \leq k(x,y) \leq K < \infty$. This also implies that $\text{MMD}(p,q)$ and $M_u^2(i,j,n)$ are bounded.

This article has been accepted for publication in IEEE Signal Processing Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LSP.2021.3075919



(a) $\mathscr{N}(1.2, 1)$ (b) $\mathscr{L}(0, 1/\sqrt{2})$ (c) Mis-identification error (d) False-alarm error

Fig. 1: Performance of NP-SEQ-1 and of NP-SEQ-UA

B. At most A anomalies ($0 \le L \le A$)

We assume $1 \le A < S/2$. The set of all possible anomalous subsets $\mathscr{S}_A$ is the set of all non-empty subsets of $\mathscr{S}$ with number of elements less than or equal to A. We propose the sequential test NP-SEQ(α, T_0) with $\alpha = 0.5$ and $T_0 = BC^2$, where B is a sufficiently large constant chosen such that $T_0 \ge \left\lceil \frac{4C^2}{\text{MMD}^4(p,q)} \right\rceil + 1$. We denote this test NP-SEQ-UA.

Theorem 4: The test NP-SEQ-UA is universally consistent, i.e., the error probability satisfies $\lim_{C\to\infty} P_{\max} = 0$.

Proof: See appendix. We show that, for any $\mathscr{I} \in \mathscr{S}_A$,

$$\lim_{C\to\infty} \mathbb{P}_0[\delta \neq \varnothing] = 0 \text{ and } \lim_{C\to\infty} \mathbb{P}_{\mathscr{I}}[\delta \neq \mathscr{I}] = 0. \quad \blacksquare$$

IV. Simulation Results

For all our experiments we used the Gaussian kernel [3], [11], which makes the MMD a metric, and chose $\sigma^2 = 0.5$. We chose $S = 5$, and set the typical distribution to be $p = \mathscr{N}(0, 1)$. The results are averaged across 100000 realizations. In Figs. 1a and 1b, we compare the sequential test NP-SEQ-1 with the nonparametric fixed sample size (NP-FSS) test[3] in [11], NP-SEQ$(0.5, \infty)$ (a test with a slower decay in the threshold) and NP-SEQ$(2, \infty)$ (a test with a faster decay in the threshold). Figs. 1a and 1b show the average stopping delay versus $-\log_{10} P_{\max}$ for two choices of q: $\mathscr{N}(1.2, 1)$, and $\mathscr{L}(0, 1/\sqrt{2})$, where $\mathscr{L}(0, 1/\sqrt{2})$ denotes a Laplacian distribution with zero mean and unit variance. For the NP-FSS test of [11], we plot the sample size versus $\log_{10}(1/P_{\max})$. We observe that: (1) the sequential test (NP-SEQ-1) performs better than the NP-FSS test in both cases, (2) NP-SEQ-1 performs better than both NP-SEQ$(0.5, \infty)$ and NP-SEQ$(2, \infty)$, and (3) the slope from the plots is upper bounded by the theoretical upper bound in (5). The $\text{MMD}^2(p, q)$ for Figs. 1a and 1b are 0.2288 and 0.01357, respectively.

Figs. 1c and 1d show the performance of NP-SEQ-UA (with q being $\mathscr{N}(1.2, 1)$, and $A = 1, 2$) under hypothesis $\mathscr{I} = \{5\}$ and null hypothesis, respectively. It can be observed that: (1) the false alarm error goes to 0 as $C \to \infty$, and (2) the growth of expected delay versus log of the mis-identification error is similar to the case with exactly one anomaly.

[3]The NP-FSS test for n samples and known L declares the L sequences with largest $M_u^2(i, \bar{i})$ to be the anomalous sequences, where $M_u^2(i, \bar{i})$ is the unbiased estimate of the square of the MMD between the i^{th} sequence X_i^n and the sequence $\bar{X}_i^n = \{X_1^n, X_2^n, \ldots, X_{i-1}^n, X_{i+1}^n, \ldots, X_S^n\}$ which is a concatenation of all sequences except the i^{th} sequence.

V. Summary and Future Work

We proposed universal, distribution-free, consistent sequential tests for detecting L anomalous data streams among a finite collection of S data streams for two cases: $L = 1$ and $0 \le L \le A < S/2$. This nonparametric setting is more general than the setting in [6] where the p and q are assumed to be over finite alphabets. The performance is better the FSS test for this setting in [11]. Some interesting directions for future work are: finding second-order asymptotically optimal tests as in [10] for our setting, and considering the setting with restrictions on the number of observed streams at each time [4], [9] since the sampling strategy is to be optimized.

VI. Appendix

A. Proof of Theorem 1

Proof: Let N_i be defined as:

$$N_i = \operatorname*{argmin}_{n\ge 2} \left\{ \min_{j\ne i} M_u^2(i, j, n) > T_n \right\}. \quad (6)$$

The actual stopping time of the proposed algorithm N satisfies, $N \le N_i$ for all i. Let the i^{th} data stream be anomalous. Let

$$n_0 = \left\lceil \frac{2C}{\text{MMD}^2(p, q)} \right\rceil + 1. \quad (7)$$

Note that for $n > n_0$, we have

$$T_{n-1} - \text{MMD}^2(p, q) \le -\frac{\text{MMD}^2(p, q)}{2} < 0. \quad (8)$$

For any $n > n_0$, we have

$$\begin{aligned}
&\mathbb{P}_i[N \ge n] \le \mathbb{P}_i[N_i \ge n] \\
&= \mathbb{P}_i\left[\min_{j\ne i} M_u^2(i, j, n_1) \le T_{n_1}, \forall\, n_1 < n\right] \\
&\le \mathbb{P}_i\left[\min_{j\ne i} M_u^2(i, j, n-1) \le T_{n-1}\right] \\
&\le \sum_{j\ne i} \mathbb{P}_i\left[M_u^2(i, j, n-1) \le T_{n-1}\right] \\
&= \sum_{j\ne i} \mathbb{P}_i\Big[M_u^2(i, j, n-1) - \text{MMD}^2(p, q) \\
&\qquad\qquad \le T_{n-1} - \text{MMD}^2(p, q)\Big] \\
&\le \sum_{j\ne i} \exp\left\{-\frac{\left(T_{n-1} - \text{MMD}^2(p, q)\right)^2 (n-1)}{16K^2}\right\} \quad (9)
\end{aligned}$$

$$\leq (S-1)\, e^{-a(n-1)} \tag{10}$$

where $a = \text{MMD}^4(p,q)/64K^2$ and inequality (9) follows from [3, Thm.10]. Thus, $\mathbb{P}_i\left[N \geq n\right] \to 0$ as $n \to \infty$. Since $\mathbb{P}_i\left[N_i = \infty\right] \leq \mathbb{P}_i\left[N_i \geq n\right] \forall n$, we have $\mathbb{P}_i\left[N_i = \infty\right] = 0$. ■

B. Proof of Theorem 2

Proof: From Thm. 1, the algorithm stops in finite time with probability 1. We assume without loss of generality that the first stream is anomalous, i.e., $\mathscr{I} = \{1\}$, and consider error terms, $\mathbb{P}_1\left[\delta = j\right]$. where $j \neq 1$. For $j \neq 1$, we have

$$\mathbb{P}_1\left[\delta = j\right] = \sum_{n\geq 2}^{n_0} \mathbb{P}_1\left[N = n, \delta = j\right] + \sum_{n>n_0}^{\infty} \mathbb{P}_1\left[N = n, \delta = j\right],$$

where n_0 is as defined in (7). Since, $\mathbb{P}_1\left[N = n, \delta = j\right] \leq \mathbb{P}_1\left[N = n\right]$, it follows from (10) that

$$\begin{aligned} \sum_{n>n_0}^{\infty} \mathbb{P}_1\left[N = n, \delta = j\right] &\leq \sum_{n>n_0}^{\infty} (S-1)\, e^{-a(n-1)} \\ = \frac{(S-1)\, e^{-an_0}}{1-e^{-a}} &\leq \frac{(S-1)\exp\left\{-\frac{\text{MMD}^2(p,q)C}{32K^2}\right\}}{1-e^{-a}}, \end{aligned} \tag{11}$$

since $n_0 > \frac{2C}{\text{MMD}^2(p,q)}$. For $n \leq n_0$, we have

$$\mathbb{P}_1\left[N = n, \delta = j\right] \leq \mathbb{P}_1\left[N = n, \min_{l\neq j} M_u^2(j,l,n) > T_n\right]$$

$$\leq \mathbb{P}_1\left[\min_{l\neq j} M_u^2(j,l,n) > T_n\right] \leq \mathbb{P}_1\left[M_u^2(j,l,n) > T_n\right] \tag{12}$$

$$\leq \exp\left(-\frac{{T_n}^2(n-1)}{16K^2}\right) = \exp\left(-\frac{C^2(n-1)}{16K^2n^2}\right) \tag{13}$$

$$\leq \exp\left(-\frac{C^2(n_0-1)}{16K^2n_0^2}\right) \leq \exp\left(-\frac{\text{MMD}^2(p,q)Cg(C)}{32K^2}\right), \tag{14}$$

where l is chosen such that $l \neq 1$ in (12) and $g(C) = \frac{\frac{2}{\text{MMD}^2(p,q)}\left(\frac{2}{\text{MMD}^2(p,q)}+\frac{1}{C}\right)}{\left(\frac{2}{\text{MMD}^2(p,q)}+\frac{2}{C}\right)^2}$. Since both j^{th} and l^{th} streams are not anomalous, $M_u^2(j,l,n) \to 0$ as $n \to \infty$, and inequality (13) follows from the Hoeffding bound for M_u^2 in [3, Thm. 10]. Combining (14) and (11), and using (7), we get

$$\begin{aligned} \mathbb{P}_1\left[\delta = j\right] \leq &\left[\frac{2C}{\text{MMD}^2(p,q)} + 1\right] \exp\left\{-\frac{\text{MMD}^2(p,q)Cg(C)}{32K^2}\right\} \\ &+ \frac{(S-1)\exp\left\{-\frac{\text{MMD}^2(p,q)C}{32K^2}\right\}}{1-e^{-a}}, \end{aligned}$$

which goes to 0 as $C \to \infty$. The maximum error probability is $P_{\max} = \max_{i=1,\ldots,S} \mathbb{P}_i\left[\delta \neq i\right]$. Since $P_{\max} \to 0$ as $C \to \infty$, the test is *universally consistent*. Furthermore, for sufficiently large C, we can bound $P_{\max}$ as

$$P_{\max} \leq \exp\left\{-\left(\frac{\text{MMD}^2(p,q)}{32K^2} - \epsilon\right)C\right\}. \tag{15}$$

■

C. Proof of Theorem 3

Proof: We know from Theorem 1 that the stopping time is finite almost surely. Therefore with probability one under each non-null hypothesis, we have that $\Gamma_N > \frac{C}{N}$ and $\Gamma_{N-1} \leq \frac{C}{N-1}$. Equivalently, we have

$$\frac{1}{\Gamma_N} < \frac{N}{C} \leq \frac{1}{\Gamma_{N-1}} + \frac{1}{C}. \tag{16}$$

Since the kernel is bounded between 0 and K, the statistic Γ_n is bounded between $-2K$ and $2K$. Therefore, for $n < C/2K$, $\Gamma_n < C/n$, and $\mathbb{P}\left[N < n\right] = 0$. This implies that as $C \to \infty$, the stopping time $N \to \infty$. As $n \to \infty$, we have $\Gamma_n \to \text{MMD}^2(p,q)$ almost surely using [3, Thm. 10], Borel-Cantelli Lemma and the fact that the series involving the upper bound in [3, Thm. 10] is summable. Thus, from (16), we have $\frac{N}{C} \to \frac{1}{\text{MMD}^2(p,q)}$, almost surely as $C \to \infty$ under $\mathbb{P}_i\left[\cdot\right]$, and hence in probability. Finally, we get the convergence stated in the theorem since the collection of random variables $\{N/C, C \geq C_0\}$ is uniformly integrable. Universal exponential consistency is obtained by combining the convergence result for N/C with the exponential bound for $P_{\max}$ in terms of C in (15). ■

D. Proof of Theorem 4

Proof: First, consider false alarm probability $\mathbb{P}_0\left[\delta \neq \phi\right]$.

$$\mathbb{P}_0\left[\delta \neq \phi\right] = \sum_{n=2}^{T_0-1} \mathbb{P}_0\left[\tilde{N} = n\right] \leq \sum_{n=2}^{T_0-1} \mathbb{P}_0\left[\Gamma_n > T_n\right]$$

$$\leq \sum_{n=2}^{T_0-1} \mathbb{P}_0\left[\min_{i\in\mathscr{A}} \min_{j\in\mathscr{S}\setminus\mathscr{A}} M_u^2(i,j,n) > T_n \text{ for some } \mathscr{A} \in \mathscr{S}_A\right]$$

$$\leq |\mathscr{S}_A| \sum_{n=2}^{T_0-1} \mathbb{P}_0\left[M_u^2(i,j,n) > T_n\right] \text{(with } i \neq j)$$

$$\leq |\mathscr{S}_A| \sum_{n=2}^{T_0-1} \exp\left(-\frac{T_n^2(n-1)}{16K^2}\right) \tag{17}$$

$$\leq |\mathscr{S}_A|(T_0-2)\exp\left(-\frac{C^2}{16K^2}\left(1-\frac{1}{T_0-1}\right)\right),$$

which goes to 0 as $C \to \infty$. Next, consider $\mathbb{P}_{\mathscr{I}}\left[\delta \neq \mathscr{I}\right] = \mathbb{P}_{\mathscr{I}}\left[\delta = \phi\right] + \mathbb{P}_{\mathscr{I}}\left[\delta \neq \mathscr{I}, \delta \neq \phi\right]$. Similar to (6), define:

$$N_{\mathscr{I}} = \underset{n\geq 2}{\text{argmin}} \left\{\min_{i\in\mathscr{I}} \min_{j\in\mathscr{S}\setminus\mathscr{I}} M_u^2(i,j,n) > C/\sqrt{n}\right\}.$$

The stopping time $\tilde{N}$ satisfies $\tilde{N} \leq N_{\mathscr{I}}$ for all $\mathscr{I}$. Under any non-null hypothesis $\mathscr{I} \in \mathscr{S}_A$, we have

$$\begin{aligned} &\mathbb{P}_{\mathscr{I}}\left[\delta = \phi\right] = \mathbb{P}_i\left[\tilde{N} \geq T_0\right] \leq \mathbb{P}_{\mathscr{I}}\left[N_{\mathscr{I}} > T_0\right] \\ &\overset{(a)}{\leq} |\mathscr{I}|(S-|\mathscr{I}|)e^{-a(T_0-1)} \overset{(b)}{\leq} |\mathscr{I}|(S-|\mathscr{I}|)e^{-\frac{C^2}{16K^2}}, \end{aligned}$$

where (a) is obtained following the steps from (8)-(10) in Theorem 1, and (b) is obtained by choosing $T_0 \geq n_0 = \left\lceil\frac{4C^2}{\text{MMD}^4(p,q)}\right\rceil + 1$. Thus, $\mathbb{P}_{\mathscr{I}}\left[\delta = \phi\right] \to 0$ as $C \to \infty$. The term $\mathbb{P}_{\mathscr{I}}\left[\delta \neq \mathscr{I}, \delta \neq \phi\right]$ can also be shown to go to zero as $C \to \infty$ as follows: (1) Bound this term by the probability of error under hypothesis $\mathscr{I}$ without a timeout ($T_0 = \infty$), (2) then follow the steps in the proof of Theorem 2 with $T_n = C/\sqrt{n}$. The resulting bound for $\mathbb{P}_{\mathscr{I}}\left[\delta \neq \mathscr{I}, \delta \neq \phi\right]$ is of the form $k_1\exp(-C^2/32K^2)$, where k_1 is a constant, and goes to 0 as $C \to \infty$. ■

REFERENCES

[1] Y. Bu, S. Zou, Y. Liang, and V. V. Veeravalli, "Universal outlying sequence detection for continuous observations," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China*, March 2016, p. 4254–4258.

[2] Y. Bu, S. Zou, and V. V. Veeravalli, "Linear-complexity exponentially-consistent tests for universal outlying sequence detection," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, p. 2115–2128, April 2019.

[3] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, no. Mar, p. 723–773, 2012. [Online]. Available: http://www.jmlr.org/papers/v13/gretton12a.html

[4] L. Lai, H. V. Poor, Y. Xin, and G. Georgiadis, "Quickest search over multiple sequences," *IEEE Transactions on Information Theory*, vol. 57, no. 8, p. 5375–5386, Aug 2011.

[5] Y. Li, S. Nitinawarat, and V. V. Veeravalli, "Universal outlier hypothesis testing," *IEEE Transactions on Information Theory*, vol. 60, no. 7, p. 4066–4082, July 2014.

[6] ——, "Universal sequential outlier hypothesis testing," *Sequential Analysis*, vol. 36, no. 3, p. 309–344, 2017. [Online]. Available: https://doi.org/10.1080/07474946.2017.1360086

[7] A. Tajer and H. V. Poor, "Quick search for rare events," *IEEE Transactions on Information Theory*, vol. 59, no. 7, p. 4462–4481, July 2013.

[8] A. Tajer, V. V. Veeravalli, and H. V. Poor, "Outlying sequence detection in large data sets: A data-driven approach," *IEEE Signal Processing Magazine*, vol. 31, no. 5, p. 44–56, Sep. 2014.

[9] N. K. Vaidhiyan and R. Sundaresan, "Learning to detect an oddball target," *IEEE Transactions on Information Theory*, vol. 64, no. 2, p. 831–852, Feb 2018.

[10] L. Zhou, Y. Wei, and A. Hero, "Second-order asymptotically optimal universal outlying sequence detection with reject option," 2020.

[11] S. Zou, Y. Liang, H. V. Poor, and X. Shi, "Nonparametric detection of anomalous data streams," *IEEE Transactions on Signal Processing*, vol. 65, no. 21, p. 5785–5797, Nov 2017.