Separating an Outlier from a Change

Deniz Sargun
Department of Electrical and Computer Engineering
The Ohio State University
Columbus, OH, USA
sargun.1@osu.edu

C. Emre Koksal
Department of Electrical and Computer Engineering
The Ohio State University
Columbus, OH, USA
koksal.2@osu.edu

Abstract—We study the quickest change detection problem with an unknown post-change distribution. In this scenario, the unknown change in the distribution of observations may occur in many ways without much structure, while, before change, an outlier (a false alarm event) is highly structured, following a particular sample path. We first characterize these likely events for the deviation of finite strings and propose a method to test the deviation, relative to the most likely way for it to occur as an outlier. Our method works along with other change detection schemes to substantially reduce the false positive rates associated with the generalized likelihood ratio test. Finally, we apply our method on economic market indicators and climate data. Our method successfully captures the regime shifts during times of historical significance and identifies the current climate change phenomenon to be a highly likely regime shift.

Index Terms—hypothesis testing, quickest change detection, time complexity, generalized likelihood ratio test, large deviations; information projection; KL divergence

I. INTRODUCTION

Not every long-term deviation from the norm should be considered as a result of change. Such occurrences may also be caused by rare events driven by the system. To that end, this paper focuses on a problem that is an instance of non-Bayesian quickest change detection [1] (also see [2] for a wide treatment of change detection and refer to chapter 8 for nonadditive change models and existing approaches) of classifying deviations observed in a time-series without a parametric model for the post-change distribution. Quickest change detection with unknown post-change models have been studied widely, both in theory [3]–[5] and in application [6], [7]. More variants of the change detection problem and applications are discussed in [8].

For some non-Bayesian quickest change detection problems where pre- and post-change distributions have finitely many alternatives it is known that a version of the cumulative sum algorithm (CUSUM) is optimal or asymptotically optimal [9], [10]. For other problems where the pre- or post-change distribution is unknown or not parametrized most of the CUSUM-like algorithms that form a likelihood ratio test (GLRT) over the set of alternative distributions and then use it as a drift term [11], which can be computationally demanding or without any optimality guarantees. For this problem, we propose a computationally efficient method, called information projection test (IPT), that works in conjunction with an existing change detection algorithm as an additional filter to separate outliers from a change. Used for hypothesis testing, IPT reduces the false positive probability without much effect on the power of the test. IPT assumes a standard online change detection scheme (similar to CUSUM) at the first level and comparison of the relative entropy of the empirical distribution with the distribution of the most likely outcome for the outlier.

We show that IPT reduces the probability of false positive exponentially with the threshold selected, at the expense of a reduction in the probability of correct detection only with a polynomial order of the same value. This, in turn, substantially improves the operating characteristics (OC) of the underlying hypothesis testing scheme. We also share our simulation results for the detection performance in terms of detection delay vs false alarm rate which empirically show that IPT performs similar to GLRT with a few orders of lower complexity (99.86% reduction in test time for an alphabet with ~ 6500 letters and same order of samples). Thus, under certain assumptions, our algorithm achieves a better trade-off between computation time and performance compared to other methods.

Contributions of this paper can be summarized as follows:

• We introduce a very simple and novel algorithm for detecting changes with known pre-change distribution but unknown post-change distribution.
• Theorems given prove detection performance bounds for the finite strings length regime.
• We empirically benchmark our method using Lorden’s criteria with the worst case post-change assumption.
• The time complexities of our approach is compared against existing methods and is empirically tested.
• We draw new insights by applying our idea in analyzing economic and climate time series data as to understand whether periods of long deviations from the norm are outliers or change. We identify shifts from the average behavior of market indices and company returns to show that our scheme successfully manages to detect the periods with regime changes like crises. One of the highlights of our analysis is that the climate change phenomenon that is observed in the last 30 years is highly unlikely to be an outlier, giving credence to the hypothesis that it is
caused by exogenous factors.

II. MODEL AND PROBLEM STATEMENT

Consider \( X_1^\infty = (X_1, X_2, \ldots) \), a sequence of observations, where each random variable \( X_i \) is independent and identically distributed (i.i.d.) with the pre-change distribution \( Q \) except for \( t \in [t_1, t_2, \ldots] \), where unknown nonrandom change times \( t_i = (t_1, t_2, \ldots) \) satisfy \( t_0 = 1 < t_1 < \ldots < t_c \leq \infty \) and \( d \in [1, \infty] \) denotes the known duration of changes (deviations). Samples of the \( i^{th} \) deviation, \( X_1^{t_1+d} = (X_{t_1}, X_{t_1+1}, \ldots, X_{t_1+d-1}) \) are i.i.d. with \( R_t \), independent of previous observations \( X_1^{t-1} \). We will assume the pre- and post-change distributions have finite mean and, without loss of generality, that \( \mathbb{E} X_1 = \int x dF_{X_1}(x) = 0 \). To simplify our treatment, assume the post-change distributions \( R_t \) are unknown up to \( R_t \in \Delta(\hat{\mu}) \) for some known \( \hat{\mu} > 0 \) where \( \Delta(\mu) \) denotes the probability simplex and

\[
\Delta(\mu) = \left\{ P \in \Delta \middle| \sum x P(x) \geq \mu \right\} \tag{1}
\]

is the set of distributions with mean at least \( \mu \) for any real \( \mu \).

Both the initial distribution, \( Q \), and the post-change distributions \( R_t \) are over the same finite alphabet \( A = \{a_1, a_2, \ldots, a_m\} \) where \( a_i < a_{i+1} \) for \( i = 1, 2, \ldots, m-1 \) and \( m \) denotes the size of the alphabet. We also use \( P_{x_{i}} \) for the empirical distribution of realizations \( x_{i} = (x_i, x_{i+1}, \ldots, x_j) \), ie., \( P_{x_{i}}(x) = \frac{1}{j-i+1} \sum_{k=i}^{j} 1_{x_{i}}(x) \) and \( P_{d} \) for the discrete set of empirical distributions that can be realized with \( d \) samples from \( A \).

We will first consider the offline binary hypothesis testing problem under multiple change points. Given \( t \), the null hypothesis, \( H_0 \), is true if \( X_1^{t-d} \) are i.i.d \( Q \), ie., \( \forall i = 1, 2, \ldots, r \) either \( t_i < t-2d+2 \) or \( t_i > t \) and the alternative hypothesis, \( H_1 \), is true otherwise. Our goal is to minimize the maximum probability of misdetection given an upper bound on the probability of false alarm. An example decision rule region for the problem is given in Fig. [13] for \( m = 3 \). Once we define the decision regions \( \Gamma_0 = \{H = H_0\} \) and \( \Gamma_1 = \{H = H_1\} \) the problem can be stated as follows.

\[
\min_{H \in \Delta(\mu)} \max \left( P_{X_{1-d}^{t} \in \Gamma_0 | H_1} \right) \tag{2}
\]

subject to

\[
\min_{H \in \Delta(\mu)} \max \left( P_{X_{1-d}^{t} \in \Gamma_1 | H_0} \right) \leq \alpha \tag{3}
\]

where

\[
WADD(t_a) = \sup_{R,t_0} \mathbb{E}((t_a - t_0 + 1)^d | R, t_0, X_{t_0+1}^{t_0}) \tag{7}
\]

\[
ARL(t_a) = \mathbb{E}(t_a | t_0 = \infty) \tag{8}
\]

Optimal solution to this problem is not known under the assumption that post-change distribution is unknown. Therefore, we only compare our solution of [5],[6] with existing methods in the literature.

III. MOTIVATION

The first problem is an instance of composite hypothesis testing [15], without a parametric model after the change point under \( H_1 \). A typical detector in this case will pick \( H_1 \) if the empirical distribution, \( P = P_{x_{1-d+1}^{t}} \) is closer to a distribution in region \( \Delta(\mu) \) than to \( Q \). However, if \( P \) is close to \( \Delta(\mu) \), it is not necessarily true that \( x_{1-d+1}^{t} \) is drawn from some distribution in \( \Delta(\mu) \). It may be the case that it is drawn from \( Q \), yet the empirical distribution looks as if it is drawn from some \( R \in \Delta(\mu) \), leading to a false positive. We call such strings that are drawn from \( Q \) but satisfy \( P_{x_{1-d+1}^{t}} \in \Delta(\mu) \) an outlier.

For clarity we provide insight from large deviations theory, where the main results are asymptotic in the large string regime, but our main contribution is for finite strings. Next, we study the way outliers occur, when they occur using Sanov’s theorem. We will use simplified versions of the theorem statements, for the sake of clarity of exposition. (see [14] for a more detailed treatment)

Theorem 1 (Sanov): Given a distribution \( Q \) and \( \mu > \int xQ(x)dx \) and \( X_1^\infty \) i.i.d. with \( Q \),

\[
\lim_{T \to \infty} -\frac{1}{T} \log P \left( \frac{1}{T} \sum_{t=1}^{T} X_t \geq \mu \right) = \inf_{P \in \Delta(\mu)} D(P||Q) \tag{9}
\]

Note that we have the equality in Sanov’s theorem, since \( \Delta(\mu) \) is the closure of its interior as shown in Fig. [16] where we also illustrate Sanov’s theorem. The I-projection \( Q^\ast \) is the solution of the linear constrained convex optimization \( \arg\min_{P \in \Delta(\mu)} D(P||Q) \). The proof of the most general form can be found in [15]. The theorem gives an asymptotic result in the length of the string and provides the exact characterization of the exponent at which the probability of an outlier in the form of sample mean exceeding \( \mu \) with \( i \). This implies that the most likely way for a sustained deviation to occur is the empirical distribution of the associated string to look as if it is drawn from the I-projection, \( Q^\ast \). The probability of this particular way of deviation dominates the probability of all others. An approximate version of this result holds for a finite string of length \( d \) if Stirling’s approximation is satisfied. Note that the application will be for testing finite string lengths. In the next section, we give our algorithm, which is based on using the projected distribution as the most likely outlier distribution. Subsequently, we evaluate the detection performance, where our analysis also takes the finiteness of the strings into account to justify the applicability in the non-asymptotic region.
IV. INFORMATION PROJECTION TEST

The information projection test exploits the observation that outliers occur in a particular way with high probability, while there is no such structure for the deviations. To run the algorithm, initially we pick a mean $\mu \in [0, \bar{\mu}]$, a radius $\delta > 0$ and compute $Q^* = \arg\min_{P \in \Delta(\mu)} D(P\|Q)$ and follow the steps below:

1. For a given string $x_{t-d+1}^t$, check if $P = P_{x_{t-d+1}^t} \in \Delta(\mu)$ (or in the region outside the decision threshold for a deviation). This is the standard step for the first level identification.

2. If Step 1 holds, check the empirical distribution $P$ against the most likely distribution of an outlier, $Q^*$. Then claim $H_1$ if and only if $D(P\|Q^*) > \delta$.

Step 2 checks whether the empirical distribution is close to the tilted distribution $Q^*$, which is the most likely way for an outlier. This process is illustrated in Fig. 1C for the decision threshold right on the border of $\Delta(\mu)$. In the rest of the paper, we refer to the expected log-likelihood ratio $D(P\|Q^*) = E_{Q^*} \log \frac{P}{Q}$ as the relative log likelihood function (RLLF), referring to the comparison of the likelihood of the projection distribution with that of the empirical distribution of the observation at hand.

V. BOUNDS ON DETECTION PERFORMANCE

In the following sequence of theorems we characterize the trade-off between the decrease in the false positive probability and the increase in the misdetection probability associated with the addition of Step 2 to plain detection rule specified in Step 1. For a finite alphabet we will show that the false positive probability decays exponentially with $\delta$. To quantify the negative effect on detecting deviations we assume a uniform distribution for $R$ on $\Delta(\tilde{\mu})$. Then, the probability of misdetection increases in $\delta$ only as a polynomial of degree $m/2$.

We will prove the theorems for the threshold of the decision region on the boundary of $\Delta(\mu)$, but generalizations to other decision regions is straightforward.

We first show the the probability of false alarm under the pre-change distribution decreases exponentially with the number of samples $d$ at a rate that converges to $\delta$.

Theorem 2: Given $X_i^d$ i.i.d. $Q$, threshold $\mu > E_X$ and radius $\delta > 0$, the probability of a false alarm given empirical mean is greater than $\mu$ is

$$\mathbb{P}(D(P\|Q^*) \geq \delta | P \in \Delta(\mu)) = \exp(-\kappa d)$$

where $\kappa \geq \delta - c \frac{\log d}{d}$ for some constant $c > 0$.

Proof: Consider the closed and convex set $\Delta(\mu)$. Define $Q^* = \arg\min_{P \in \Delta(\mu)} D(P\|Q)$ and $\Delta(\mu, \delta) = \{ P \in \Delta(\mu) | D(P\|Q^*) > \delta \}$. By the Pythagorean theorem for relative entropy [16], since $\Delta(\mu)$ is a closed convex set, for any $P \in \Delta(\mu, \delta)$ we have $D(P\|Q) \geq D(Q^*\|Q) + \delta$.

$$\mathbb{P}(P_{X_i^d} \in \Delta(\mu, \delta)) = \sum_{P \in \mathcal{P}_{d \cap \Delta(\mu, \delta)}} \mathbb{P}(P_{X_i^d} = P) = \sum_{P \in \mathcal{P}_{d \cap \Delta(\mu, \delta)}} |T(P)| \exp(-d(H(P) + D(P\|Q))) \leq \sum_{P \in \mathcal{P}_{d \cap \Delta(\mu, \delta)}} \exp(-d(D(Q^*\|Q) + \delta)) \leq (d + 1)^m \exp(-d(D(Q^*\|Q) + \delta))$$

We also have,

$$\mathbb{P}(P_{X_i^d} \in \Delta(\mu)) = \sum_{P \in \mathcal{P}_{d \cap \Delta(\mu)}} \mathbb{P}(P_{X_i^d} = P) = \sum_{P \in \mathcal{P}_{d \cap \Delta(\mu)}} |T(P)| \exp(-d(H(P) + D(P\|Q))) \geq (d + 1)^{-m} \exp(-d(D(Q^{**}\|Q)))$$

where $Q^{**} = \arg\min_{P \in \mathcal{P}_{d \cap \Delta(\mu)}} D(P\|Q)$. The last inequality follows from the fact that $P = Q^{**}$ appears in the summation. Then,

$$\mathbb{P}(D(P\|Q^*) \geq \delta | P \in \Delta(\mu)) \leq (d + 1)^{2m} \exp(-d(D(Q^*\|Q) + \delta - D(Q^{**}\|Q)))$$

and one can choose $\kappa \geq \delta + D(Q^*\|Q) - D(Q^{**}\|Q) -$
Define $Q$ with respect to the operating characteristics of the VMT. Therefore, Theorems 2 and 3 bound the operating point of the IPT with respect to the operating characteristics of the VMT.

The next theorem shows that the misdetection probability increases as a polynomial of $\delta$.

**Theorem 3:** Given $X^d_i \sim R_i$ for some $i = 1, 2, \ldots, r$ where $R_i \in \Delta(\bar{\mu})$ is unknown, the increase in the worst case misdetection probability is bounded as follows.

$$\mathbb{P}(P_{X^d_i} \in \Delta(\mu) \setminus \Delta(\mu, \delta)|X^d_i \sim R) \leq \exp(-\eta d)$$

where $\eta \geq c_1 \sqrt{\delta} + c_2 \frac{\log d}{d}$ for some constants $c_1, c_2 > 0$.

**Proof:** We select the pair $(\mu, \delta)$ so that we can select an even larger mean filter $\mu < \mu' < \bar{\mu}$ that strictly includes our new decision region with the KL ball around $Q^*$. Then we I-project the worst case post-change distribution $Q^*$ on this new mean filter. Finally we upper bound the misdetection probability using the method of types and the property of I-projection.

Define $\mu' = \frac{\bar{\mu} + \mu}{2} + \sqrt{\frac{\log 2}{d} (a_m - a_1)}\sqrt{\delta}$ where we have selected $\mu$ and $\delta$ to satisfy $\mu < \mu' < \bar{\mu}$. Let $\Delta'((\mu'))$ denote $\{P \in \Delta | \sum x P(x) \leq \mu'\}$ and for any $R \in \Delta(\bar{\mu})$, $R^* = \arg\min_{P \in \Delta'((\mu'))} D(P || R)$. Then we have the following, for any $R \in \Delta((\mu'))$

$$\mathbb{P}(P_{X^d_i} \in \Delta(\mu) \setminus \Delta(\mu, \delta)|X^d_i \sim R) \leq (d + 1)^m \max_P \exp(-d D(P || R^*) + D(R^* || R))$$

$$\leq (d + 1)^m \exp \left(-d \left(\frac{\bar{\mu} - \mu'}{2}\right)\right)$$

$$\leq \exp \left(-d \left(\frac{\sqrt{\frac{\log 2}{d} (a_m - a_1)} + m \frac{\log(d + 1)}{d}\right)\right)$$

where we have used a similar method of types argument and Pythagorean theorem for relative entropy.

Therefore, Theorems 2 and 3 bound the operating point of the IPT with respect to the operating characteristics of the VMT.

VI. COMPARISON TO OTHER DETECTION SCHEMES

We compare our change detection algorithm with various quickest change detection algorithms proposed in the literature in terms of operating characteristics and detection for delay given mean time between false alarms.

A. Operating characteristics

**Numerical Example 1:** The operating characteristics (OC) associated with our method is illustrated in Fig. 2a and compared with the OC of a (CUSUM-like) mean filter. We use a ternary alphabet $\{-1, 0, 1\}$, define the pre-change distribution to be uniform and $\bar{\mu} = 0.25$. The deviations are chosen to have length $d = 25$. At this stage, we determine the I-projection of $Q$ on $\Delta(\mu)$. To find the optimum performance of our algorithm, we change $\delta$ over $[2^{-8}, 2^{-3}]$. The curves in Fig. 2a show that, our proposed method is able to decrease the false positive rate compared to CUSUM-like mean filter for the same misdetection rate and performs similar to the GLRT.

IPT’s average misdetection performance is within 27% of that of GLRT but performs 34% better than VMT over the measured region of the OC. The area under the misdetecion-false alarm curves are 0.343, 0.224 and 0.184 for VMT, IPT and GLRT respectively.

B. Delay characteristics

**Numerical Example 2:** We consider the minimum worst case mean delay vs mean time between failures of the methods compared under the same setting as in VI-A. IPT provides a tradeoff between the mean test and GLRT with slightly worse detection delay compared to the latter. But in a scenario where computational costs would not allow a user to choose only according to the delay characteristics IPT may outperform both VMT and GLRT. A fast stream of data tracked for change at a central unit is such an example.

VII. COMPLEXITY

In this section we give the complexity of the change detection algorithms: IPT, VMT and GLRT. We compare the computational time complexity as a function of sample size $d$, alphabet size $m$ and error parameter $\epsilon$. The mean test using a rolling window has $O(d)$ complexity to decide on the alarm time during execution. The GLRT forms a likelihood ratio for each new window $x^{(k+1)d}_{kd+1}$ and therefore solves the following optimization problem:

$$\min_{R} D(P_{x^{(k+1)d}_{kd+1}} || R)$$

subject to $R \in \Delta((\mu'))$

which is a linearly constrained convex optimization problem is computationally equivalent to binary searching two Lagrange multipliers $\nu$ and $\lambda$ using an equation of $m$ rational terms, thus it has $O((d + m \log 2 \log \log 2)$ complexity. IPT omits this projection step by initially finding the most likely deviation and then comparing any realization to this distribution. Therefore, it only computes the KL divergence $D(P_{x^{(k+1)d}_{kd+1}} || Q^*)$ in $O(d + m)$ time.

**Numerical Example 3:** Using Ohio Supercomputer Center Owens cluster’s single node of 128GB memory and 2.40GHz clock the test times varied as in Fig. 2c [13]. We increased the alphabet and sample sizes proportionally and averaged over the total number of tests. IPT performed 25 times faster than GLRT for smaller alphabets and the difference grew with the alphabet size.
VIII. APPLICATIONS TO ECONOMICS AND CLIMATE DATA

In this section, we apply our outlier detection scheme, based on checking the RLLF of the most likely outlier distribution to the empirical distribution of the given period. We would like to test whether shifts of historic value are merely outliers or it happens due to an exogenous factor. To obtain the ground truth, we will use the long-term empirical distribution. Thus, the underlying assumption we make in the following data analysis is that, for \( r \) shifts of length \( d \) the overall cumulative duration of deviations is much smaller compared to the size of the data: \( rd \ll T \). To analyze a period of interest, we pick a duration \( d \) and a threshold, \( \mu \), that matches the mean of the underlying data over that period.

A. Daily Returns of Portfolios by Size

We focus on portfolio returns with different market caps [19] to identify segments with a worse performance compared to the average behavior, sustained over a substantial duration of time. The primary question we answer is the following: “is the identified negative sequence a rare event generated by the statistical nature of the market or is it driven by some exogenous factors with the potential to cause a financial crisis?” We quantized the daily percentage portfolio returns for different market caps over July 1926-March 2019 and selected a threshold \( \mu \) below the average returns. We then computed the information projection of the long term empirical distribution over the distributions with mean less than the threshold. The \( d = 6 \) month moving average and the KL divergence of the observation window against the I-projection is given in Fig. 3a.

With the right choice of \((d, \mu, \delta)\) we were able to identify historically significant events like the great depression or the 2009 financial crisis.

B. Oxygen Isotope Data

We analyze climate data collected from a polar cap that spans ~ 1800 years to gain insight of our method and historic climate data [20]. Fig. 3b uses a moving average window of \( d = 20 \) years of quantized data of oxygen isotope density over the years 226 – 2009. We then select thresholds \( \mu \), half a standard deviation above the mean, and \( \delta \) and determine the I-projections. Note that the only point IPT classifies as change is the last 20-year period of above norm d18O levels. Although the threshold selections are not unique this intuitive approach with these hyperparameters gives insight that the last 20-year period is less likely to be an outlier rather than an effect of an exogenous change, ie. man made.

IX. CONCLUSION

In a variety of applications, it is important to identify shifts from the typical behavior. However, not every shift from the norm marks a regime change or a deviation; instead it could be an outlier. In this paper, we developed a method that differentiates between outliers and change. To achieve that, we used results from large deviations theory, which gives us statistical characterization for the outliers. Our method tests a given string against the most likely way and marks deviations using detection-theoretic tools. Finally, we applied our algorithm to a variety of applications and drew insights from the observed time series data.

REFERENCES

[1] H. V. Poor and O. Hadjiliadis, Quickest Detection. Cambridge University Press, 2008.
[2] M. Basseville and I. V. Nikiforov, Detection of Abrupt Changes: Theory and Application. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
[3] T. S. Lau, W. P. Tay, and V. V. Veeravalli, “A binning approach to quickest change detection with unknown post-change distribution,” IEEE Transactions on Signal Processing, vol. 67, no. 3, pp. 609–621, Feb 2019.
[4] A. Tartakovsky, “Asymptotically Optimal Quickest Change Detection In Multistream Data - Part 1: General Stochastic Models,” arXiv e-prints, p. arXiv:1807.08971, Jul 2018.
[5] D. Egea-Roca, J. A. Lpez-Salcedo, G. Seco-Granados, and H. V. Poor, “Performance bounds for finite moving average tests in transient change detection,” IEEE Transactions on Signal Processing, vol. 66, no. 6, pp. 1594–1606, March 2018.
Fig. 3. Our method on economics and climate time series data sets.

[6] I. Nikiforov, “Quickest multidecision abrupt change detection with some applications to network monitoring,” in Distributed Computer and Communication Networks, V. Vishnevsky and D. Kozyrev, Eds. Cham: Springer International Publishing, 2016, pp. 94–101.

[7] L. Xie, G. V. Moustakides, and Y. Xie, “First-order optimal sequential subspace change-point detection,” arXiv e-prints, p. arXiv:1806.10760, Jun. 2018.

[8] V. V. Veeravalli and T. Banerjee, “Quickest Change Detection,” arXiv e-prints, p. arXiv:1210.5552, Oct. 2012.

[9] O. Hadjiliadis and V. Moustakides, “Optimal and asymptotically optimal cusum rules for change point detection in the brownian motion model with multiple alternatives,” Theory of Probability & Its Applications, vol. 50, no. 1, pp. 75–85, 2006. [Online]. Available: https://doi.org/10.1137/S0040585X97981494

[10] T. Banerjee and V. V. Veeravalli, “Data-Efficient Minimax Quickest Change Detection with Composite Post-Change Distribution,” arXiv e-prints, p. arXiv:1410.3450, Oct 2014.

[11] T. S. Lau, W. P. Tay, and V. V. Veeravalli, “Quickest change detection with unknown post-change distribution,” in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2017, pp. 3924–3928.

[12] G. Lorden, “Procedures for reacting to a change in distribution,” Ann. Math. Statist., vol. 42, no. 6, pp. 1897–1908, 12 1971. [Online]. Available: https://doi.org/10.1214/aoms/1177693055

[13] H. V. Poor, An Introduction to Signal Detection and Estimation (2Nd Ed.). Berlin, Heidelberg: Springer-Verlag, 1994.

[14] R. G. Gallager, Stochastic processes: theory for applications. Cambridge University Press, 2016.

[15] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications. Springer Berlin Heidelberg, 2010.

[16] T. M. Cover and J. A. Thomas, Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). New York, NY, USA: Wiley-Interscience, 2006.

[17] I. Sason and S. Verdú, “Upper bounds on the relative entropy and rnyi divergence as a function of total variation distance for finite alphabets,” in 2015 IEEE Information Theory Workshop - Fall (ITW), Oct 2015, pp. 214–218.

[18] O. S. Center, “Ohio supercomputer center,” 1987. [Online]. Available: http://osc.edu/ark:/19495/f5s1ph73

[19] K. R. French, “Data library,” Tuck School of Business, Dartmouth University, Tech. Rep., 2019. [Online]. Available: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

[20] L. Thompson, E. Mosley-Thompson, M. Davis, V. Zagorodnov, I. Howat, V. Mikhaleenko, and P.-N. Lin, “Annually resolved ice core records of tropical climate variability over the past 1800 years,” Science, vol. 340(6135), pp. 945–950, 2013.