Non-Parametric Approach to Change Detection and Estimation in Large Scale Sensor Networks

Shai Ben-David, Ting He, and Lang Tong

School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853
Email: {shai@ece.,ltong@ece.,th255@}cornell.edu

Abstract — **We consider a non-parametric, spatial sample-based scheme for the detection and estimation of changes of a random field by collecting packets from randomly distributed sensors. We assume that each sensor has a fixed probability of successfully sending to a mobile access point a packet containing its local state—either "excited" or "baseline" and its location. The task we are concerned with here is as follows: Given two sets of packets collected over two nonoverlapping time-windows, construct a test to determine if the distribution generating the sensor's states has changed between these two time windows. Furthermore, if a change of distribution has occurred, we wish to estimate the distribution of the change.**

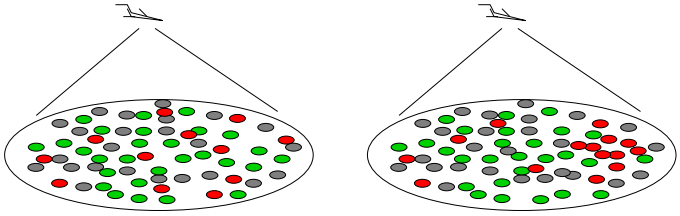
Keywords: **Non-parametric statistical methods, Change detection, Sensor Networks.**

I. Introduction

We consider the problem of detection and estimation of changes in a large scale sensor network in which each sensor is in either an "excited state" (RED) or a "baseline state" (GREEN). One example is that each sensor detects locally the presence of targets or certain chemical/biological agents. A sensor is RED if has detection and GREEN otherwise. We assume the architecture of SENMA—Sensor Networks with Mobile Access [5]—where a mobile Access Point (AP) collects data directly from sensors via random access (such as ALOHA). Each sensor transmits its local detection in the form of a packet, and the process of data collection by the mobile AP can then be modeled as a random sampling of the sensor field. We assume that at the mobile AP, the location of each received local decision is known. This can be implemented either by requiring sensors transmit their location information or letting the mobile AP poll sensors at specific locations.

The problem of change detection and estimation in SENMA, illustrated in Figure 1, is to decide, from two consecutive records, whether there is a change of the underlying probability distribution (detection) and the distribution of the change (estimation) if changes occur. Since random access is used by the mobile AP, there is no guarantee that a measurement of one sensor in the first data collection will appear in the second, nor can we be assured that the same number of samples are collected. For many applications, there is no prior knowledge about the the distribution of the sensor states. The change detection and estiamtion is therefore non-parametric

[1]This work was supported in part by the Army Research Office under Grant ARO-DAAB19-00-1-0507 and the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564.



First data collection Secod data collection

Figure 1: Sensor Networks with Mobile Access. RED: sensors with detection. GREEN: sensors with no detection. Gray: Data not collected.

in which we make no assumption about the specific form of the probability distribution of the binary random field.

Two factors must be considered for change detection and estimation in a large scale sensor networks: the time required or the number of packets required for data collection and the complexity of the detection and estimation algorithm. In practice, minimizing the number of samples required for detection reduces the collection time of a mobile AP and, more importantly in terms of energy consumption, the number of transmissions from sensors. In this paper, we aim to provide a mathematical characterization of the required sample size for which, with high probability, the distance of any two probability distributions can be estimated accurately from empirical distance between samples generated by these distributions. The basic tool used here is the Vapnik-Chervonenkis Theory [1]. We are also concerend with the detection and estimation algorithm. We present three algorithms with varying complexity, scalability, and levels of confidence.

Non-parametric change detection is a classical problem [6]. Classical techniques include permutation tests, Kolmogorov-Smirnov test, and Cramér-von Mises tests [7]. In the context of large scale sensor networks, however, it is not obvious that these classical techniques are applicable. Specifically, the samples are two dimensional, and we are interested in not only whether changes have occurred but also possible locations of changes. Furthermore, we are also interested in the asymptotic behavior as the number of samples increases and whether the detection-estimation algorithm scales.

II. A spatial notion of distance for probability distributions

Definition 1 *Fix a measure space and let $\mathcal{A}$ be a collection of measurable sets. Let P and P' be probability distributions over this space.*

The (A)-distance *between* P *and* P' *is defined as*

$$d_{(A)}(P, P') = 2 \sup_{A \in \mathcal{A}} |P_F(A) - P_G(A)|$$

We say that P, P' *are* ϵ-close with respect to $\mathcal{A}$ *if* $d_{(A}(P, P') \leq \epsilon$.

For finite domain subsets, S_1 *and* S_2*, we define the* empirical distance *to be*

$$d_{\mathcal{A}}(S_1, S_2) = 2 \sup_{A \in \mathcal{A}} \left| \frac{|S_1 \cap A|}{|S_1|} - \frac{|S_2 \cap A|}{|S_2|} \right|$$

.

The intuitive meaning of $\mathcal{A}$-distance is that it is the largest change in probability of a set that the user cares about. In particular, if we consider the scenario of a sensor network spread over some geographical area, one may assume that the changes that are of interest will be noticeable in some local rectangles or circles. This notion of $\mathcal{A}$-distance is a relaxation of the total variation distance, which is defined as $2 \sup_E |P(E) - P'(E)|$ (where the sup is taken over all measurable sets). When P and P' both have densities, then the total variation is equal to the L^1 distance. Thus it is not hard to see that $d_{(A)}$-distance is always $\leq$ the total variation or L^1 norm (if it exists) and therefore is less restrictive. This point helps get around the statistical difficulties associated with the L^1 norm[1]. If (and only if) the VC-dimension of $\mathcal{A}$ is finite, there exists a test t that can distinguish (with high probability) if any two distributions are ϵ-close (with respect to $\mathcal{A}$) using a sample size that is independent of the domain size.

For the case where the domain set is the real line, the Kolmogorov-Smirnov statistics considers $\sup_x |F_1(x) - F_2(x)|$ as the measure of difference between two distributions (where $F_i(x) = P_i(\{y : y \leq x\})$). By setting $\mathcal{A}$ to be the set of all the one-sided intervals $(-\infty, x)$ the $\mathcal{A}$ distance becomes the Kolmogorov-Smirnov statistic. Thus our notion of distance, $d_{\mathcal{A}}$ can be viewed as a generalization of this classical statistics.

The $\mathcal{A}$-distance reflects the relevance of locally centered changes. However, having adopted the concept of determining distance by focusing on a family of relevant subsets, there are different ways of quantifying such a change. The $\mathcal{A}$ measure defined above is additive - the significance of a change is measured by the *difference* of the weights of a subset between the two distributions. Alternatively, one could argue that changing the probability weight of a set from 0.5 to 0.4 is less significant than the change of a set that has probability weight of 0.1 under P_1 and weight 0 under P_2.

The following notion of *relativized discrepancy* offers a variation of the $\mathcal{A}$ distance that takes the relative magnitude of a change into account.

[1]For example, Batu et al, [3], show that over a finite domain of size N, any algorithm that can distinguish (with high probability) between between any two distributions that have L^1 distance $> 1/2$, requires (for its worst-case pair of distributions) at least $O(N^{2/3})$ sample points. This has serious implications for infinite domains (such as $\mathbb{R}^2$). For any algorithm for testing if two distributions have L^1 distance $> \epsilon$ (on the basis of viewing i.i.d. samples generated by these distributions), for any number M, there exist two distributions P_1 and P_2 such that the L^1 distance between them is, say $1/2$, and yet the algorithm needs more than M sample points to distinguish between P_1 and P_2.

Definition 2 (Relativized Discrepancy) *Let* P_1, P_2 *be two probability distributions over the same measure space, let* $\mathcal{A}$ *denote a family of measurable subsets of that space, and* A *a set in* $\mathcal{A}$.

$$W_{P_1,P_2}(A) = \frac{P_1(A) + P_2(A)}{2}$$

$$\phi_{\mathcal{A}}(P_1, P_2) = \sup_{A \in \mathcal{A}} \frac{|P_1(A) - P_2(A)|}{\sqrt{\min\{W_{P_1,P_2}(A), (1 - W_{P_1,P_2}(A))\}}}$$

For finite samples S_1, S_2*, we define* $\phi_{\mathcal{A}}(S_1, S_2)$ *similarly, by replacing* $P_i(A)$ *in the above definition by the empirical measure* $S_i(A) = |S_i \cap A|/|S_i|$.

III. Technical preliminaries

Our basic tool for sample based estimation of the $\mathcal{A}$ distance between probaility distributions is based on the Vapnik-Chervonenkis theory.

Let $\mathcal{A}$ denote a family of subsets of some domain set X. We define a function $\Pi_{\mathcal{A}} : \mathcal{N} \mapsto \mathcal{N}$ by

$$\Pi_{\mathcal{A}}(n) = \max\{|\{A \cap B : A \in \mathcal{A}\}| : B \subseteq X \text{ and } |B| = n\}$$

Clearly, for all n, $\Pi_{\mathcal{A}} \leq 2^n$. For example, if $\mathcal{A}$ is the family of all intervals over the real line, then $\Pi_{\mathcal{A}}(n) = O(n^2)$, $(0.5n^2 + 1.5n$, to be precise).

Definition 3 (VC-Dimension) *The Vapnik-Chervonenkis dimension of a collection* $\mathcal{A}$ *of sets is*

$$\textit{VC-dim}(\mathcal{A}) = \sup\{n : \Pi_{\mathcal{A}}(n) = 2^n\}$$

The following combinatorial fact, known as Sauer's Lemma, is a basic useful property of the function $\Pi_{\mathcal{A}}$.

Lemma III.1 (Sauer, Shelah) *If* $\mathcal{A}$ *has a finite VC-dimension,* d*, then for all* n*,*

$$\Pi_{\mathcal{A}}(n) \leq \Sigma_{i=0}^{d} \binom{n}{i}$$

Definition 4 (ϵ - sample) *Let* X*,* $\mathcal{A}$ *be as above and let* S *be a finite subset of* X*. For* $\epsilon \in (0, 1)$ *and a probability distribution* P *over* X*, we say that* S *is an* ϵ - sample *w.r.t.* $(\mathcal{A}, P)$ *if*

$$\forall A \in \mathcal{A}, \ \left| \frac{|S \cap A|}{|S|} - P(A) \right| \leq \epsilon$$

The following fundamental result is due to Vapnik and Chervonenkis [1].

Theorem III.1 (Vapnik-Chervonenkis) *For* X*,* $\mathcal{A}$ *as above, for any probability distribution* P *over* X *and any* ϵ*,* $\delta \in (0, 1)$*, if* S *is an* m*-size i.i.d.* P*-sample and*

$$m = O\left(\frac{1}{\epsilon^2}(\textit{VC-dim}(\mathcal{A}) \log(1/\epsilon) + \log(1/\delta))\right)$$

then with probability $\geq 1 - \delta$*,* S *is an* ϵ*-sample w.r.t.* $(\mathcal{A}, P)$.

M. Talagrand [2] improved the sample size bound above to

$$m = O\left(\frac{1}{\epsilon^2}(\text{VC-dim}(\mathcal{A}) + \log(1/\delta))\right)$$

IV. The case of sensor networks

For the rest of the paper, we let our domain set be the two dimensional plane $\mathbb{R}^2$. We shall set our collection of 'sets of interest', $\mathcal{A}$, to consist of some semi-algebraic family with a small (finite) VC-dimension, like the set of all two dimensional balls, or the set of all axis aligned rectangles.

Our basic task is to detect change between two probability distributions over $\mathbb{R}^2 \times \{RED, GREEN\}$. We consider the following scenario: P_1, P_2 are two probability distributions over the same domain X, and $\mathcal{A}$ is a family of subsets of that domain. Given two finite sets S_1, S_2 that are i.i.d. samples of P_1, P_2 respectively, we wish to estimate the $\mathcal{A}$ distance between the two distributions , $d_{\mathcal{A}}(P_1, P_2)$. Recall that, for any subset A of the domain set, and a finite sample S, we define the S- empirical weight of A by $S(A) = \frac{|S \cap A|}{|S|}$.

Claim IV.1 *Let P_1, P_2 be any probability distributions over some domain X and let $\mathcal{A}$ be a family of subsets of X and ϵ_1, $\epsilon_2 \in (0,1)$. If S_1 is an ϵ_1 sample w.r.t. $(\mathcal{A}, P_1)$ and S_2 is an ϵ_2 sample w.r.t. $(\mathcal{A}, P_2)$ then, for every $A \in \mathcal{A}$,*

$$||P_1(A) - P_2(A)| - |S_1(A) - S_2(A)|| \leq \epsilon_1 + \epsilon_2$$

In particular, we get

$$|d_{\mathcal{A}}(S_1, S_2) - d_{\mathcal{A}}(P_1, P_2)| \leq \epsilon_1 + \epsilon_2$$

Combining this claim with Theorem III.1, we conclude that for $\mathcal{A}$'s of finite VC-dimension, d, sample sizes of order $\frac{d}{\epsilon^2}$ suffice to detect the existence of a set $A \in \mathcal{A}$ on which the distributions generating these samples differ by more than ϵ.

Theorem IV.1 *Let P_1, P_2 be any probability distributions over some domain X and let $\mathcal{A}$ be a family of subsets of X and $\epsilon \in (0,1)$. If S_1, S_2 are i.i.d m samples drawn by P_1, P_2 respectively, then,*

$$P\left[\exists A \in \mathcal{A} \, ||P_1(A) - P_2(A)| - |S_1(A) - S_2(A)|| \geq \epsilon\right]$$

$$< \Pi_{\mathcal{A}}(2m) 4e^{-m\epsilon^2/4}$$

Where P in the above inequality is the probability over the pairs of samples (S_1, S_2) induced by the sample generating distributions (P_1, P_2).

One should note that if $\mathcal{A}$ has a finite VC-dimension, d, then by Sauer's Lemma, $\Pi_{\mathcal{A}}(n) < n^d$ for all n.

We thus have bounds on the probabilities of both missed detections and false alarms of our change detection tests.

The rate of growth of the needed sample sizes as a function of the sensitivity of the test can be further improved by considering a notion of *relativized discrepancy*. We can get results similar to Theorem IV.1 for the distance $\phi_{\mathcal{A}}(P_1, P_2)$. We start with the following result of Anthony and Shawe-Taylor [4].

Theorem IV.2 (Anthony, Shawe-Taylor) *Let $\mathcal{A}$ be a collection of subsets of a finite VC-dimension d. Let S be a sample of size n each, drawn i.i.d. by a probability distribution, P (over X), then*

$$P^n(\phi_{\mathcal{A}}(S, P) > \epsilon) \leq (2n)^d e^{-n\epsilon^2/4}$$

(Where P^n is the n'th power of P - the probability that P induces over the choice of samples).

Imitating the proof of theorem IV.2 we obtain the following bound on the probability of false alarm for the $\phi_{\mathcal{A}}(S_1, S_2)$ test.

Theorem IV.3 *Let $\mathcal{A}$ be a collection of subsets of a finite VC-dimension d. If S_1 and S_2 are samples of size n each, drawn i.i.d. by the same distribution, P (over X), then*

$$P^{2n}(\phi_{\mathcal{A}}(S_1, S_2) > \epsilon) \leq (2n)^d e^{-n\epsilon^2/4}$$

(Where P^{2n} is the $2n$'th power of P - the probability that P induces over the choice of samples).

Another corollary of theorem IV.2 is the following bound on the probability that the empirical distance of samples drawn by two distributions deviates from the true distance between these distributions.

Corollary 1 *Let P_1, P_2 be any probability distributions over some domain X and let $\mathcal{A}$ be a family of subsets of X and $\epsilon \in (0,1)$. If S_1, S_2 are i.i.d m samples drawn by P_1, P_2 respectively, then,*

$$P\left[\exists A \in \mathcal{A} \frac{||P_1(A) - P_2(A)| - |S_1(A) - S_2(A)||}{\sqrt{P_1(A)} + \sqrt{P_2(A)}} \geq \epsilon\right]$$

$$< \Pi_{\mathcal{A}}(2m) 4e^{-m\epsilon^2/4}$$

Where P in the above inequality is the probability over the pairs of samples (S_1, S_2) induced by the sample generating distributions (P_1, P_2).

To appreciate the potential benefits of using this relative discrepancy approach, consider the case where $\mathcal{A}$ is the collection of all 'cylinders' in $\mathbb{R}^2 \times \{RED, GREEN\}$. That is, each $A \in \mathcal{A}$ is of the form $A = I \times C$ where I is one of $\{RED\}$, $\{GREEN\}$, $\{RED, GREEN\}$ or $\emptyset$, and C is a planar disk. It is easy to verify that the VC-dimension of this family $\mathcal{A}$ is 4 (the CV-dimension of the family of planar disks is 3). Let us estimate what sample sizes are needed to be 99% sure that a disk D, that changed from having no RED readings to having η fraction of the detected readings being REDs in this disk, indicate a real change in the measured field. Note that for such a disk, $\frac{S_1(D) - S_2(D)}{\sqrt{0.5(S_1(D) + S_2(D))}} = \sqrt{2\eta}$. We can now apply Theorem IV.3 to see that $m = 30/\eta$ should suffice. Note that if we used the $d_{\mathcal{A}}$ measure and theorem IV.1, the bound we could guarantee would be in the order of $1/\eta^2$.

V. Algorithms

In this section, we describe three algorithms with varying degree of completeness, complexity, flexibility, all based on the empirical probability distribution to detect the changes in the actual distribution. Specifically, we rely on Claim IV.1, Theorem IV.1 and Theorem IV.3 to provide a guaranteed level of accuracy.

Practical implementations require that the distance between two empirical distributions $d_{\mathcal{A}}(S_1, S_2)$ be computed within a small finite number of operations. We address this consideration by reducing the search in $\mathcal{A}$ to a search in a *finite* subset $\mathcal{H} \subset \mathcal{A}$, possibly a function of sample S, and replacing $d_{\mathcal{A}}(S_1, S_2)$ by $d_{\mathcal{H}}(S_1, S_2)$. If $\mathcal{H}$ is not chosen properly, such a reduction of the search domain may lead to a loss of performance. We call the collection $\mathcal{H}$ *complete with respect to sample S* if $\forall A \in \mathcal{A}$, there exists a $B \in \mathcal{H}$ such that $S \cap A = S \cap B$. Thus, if $\mathcal{H}$ is complete w.r.t. $S_1 \cup S_2$, then $d_{\mathcal{A}}(S_1, S_2) = d_{\mathcal{H}}(S_1, S_2)$.

When a sensor network is used for time critical monitoring, the computational complexity of the data analyzing algorithms is critical. Specifically, we focus on how the computational complexity scales with respect to the sample size $M = |S_1 \cup S_2|$. The key to complexity reduction is the reuse of previous computations.

Algorithm design also depends on the distance metric used: the empirical distance $d_{\mathcal{A}}(S_1, S_2)$ vs. the relativized discrepancy $\phi_{\mathcal{A}}(S_1, S_2)$. It is advantageous that algorithms are amendable to both metrics, but such flexibility may lead to an increase in complexity. We will present our algorithms using $d_{\mathcal{A}}(S_1, S_2)$ and comment on their flexibility.

Algorithm 1: Exhaustive Search in Planar Disks

Let $\mathcal{A}$ be the collection of two dimensional disks. For a finite subset, $S \subseteq X$, consider the finite collection of disks $\mathcal{H}_S \subset \mathcal{A}$ defined by

$$\mathcal{H}_1(S) \triangleq \{D(s_i, s_j, s_k) : s_i, s_j, s_k \in S\} \quad (1)$$

where $D(s_i, s_j, s_k)$ is the circle with s_i, s_j, and s_k on its boundary. See Figure 2.

Claim V.1 *Let $\mathcal{A}$ be the collection of two dimensional disks. Given S_1 and S_2 The finite collection $\mathcal{H}_1 = \mathcal{H}_1(S_1 \cup S_2)$ in (1) is complete with respect to $S_1 \cup S_2$.*

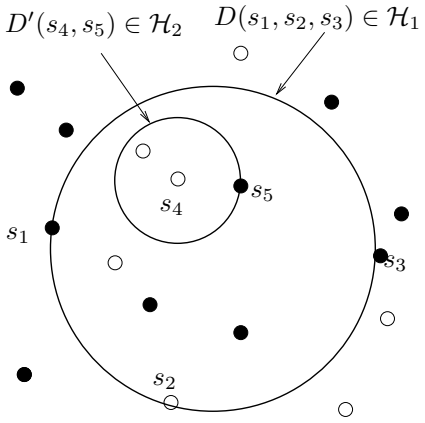


Figure 2: Members of $\mathcal{H}_1$ and $\mathcal{H}_2$

Given that $\mathcal{H}_1$ is complete, the first algorithm performs the following exhaustive search

$$\max_{A \in \mathcal{H}_1} \left| \frac{|S_1 \cap A|}{|S_1|} - \frac{|S_2 \cap A|}{|S_2|} \right|.$$

The algorithm includes (i) generating elements of $\mathcal{H}_1$; (ii) the computation of $\left| \frac{|S_1 \cap A|}{|S_1|} - \frac{|S_2 \cap A|}{|S_2|} \right|$, and (iii) finding the maximum.

The complexity of Algorithm 1 for sample size $M = |S_1 \cup S_2|$ is $O(M^4)$. The dominating term is the computation of empirical distance for each disk. This algorithm is flexible, i.e., it can be easily modified for a different metric, $\phi_{\mathcal{A}}(S_1, S_2)$, in particular.

Algorithm 2: Search in Sample Centered Disks

Let $\mathcal{A}$ be the collection of two dimensional disks. For a finite $S \subseteq X$, consider the finite collection of disks $\mathcal{H}_2(S) \subset \mathcal{A}$ defined by

$$\mathcal{H}_2(S) \triangleq \{D'(s_i, s_j) : s_i, s_j \in S\} \quad (2)$$

where $D'(s_i, s_j)$ is the circle with s_i at the center and s_j on the boundary. See Figure 2.

Algorithm 2 carries out the search for a change in the empirical weight only over disks in $\mathcal{H}_2 = \mathcal{H}_2(S_1 \cup S_2)$. The search for a disk in $\mathcal{H}_2$ that has the maximum change in empirical distributions proceeds by searching in $\mathcal{H}_2$ for fixed center s_i. Define

$$f_i^k(j) \triangleq \frac{|S_k \cap D'(s_i, s_j)|}{|S_k|}, k = 1, 2 \quad (3)$$

$$F_i(j) = |f_i^1(j) - f_i^2(j)| \quad (4)$$

where $F_i(j)$ is the change in empirical probability of $D'(s_i, s_j)$.

For fixed s_i, computing $F_i(j)$ is done by first sorting the sample points according to their distance to s_i[2], and then count $|S_k \cap D'(s_i, s_j)|, k = 1, 2$ from the inner most to the outer most disk centered at s_i. Then compute

$$i_{max} = \arg\max_j F_i(j) \quad (5)$$

The optimal disk in $\mathcal{H}_2$, for fixed s_i, is given by $D'(s_i, s_{i_{max}})$. The search repeats for all possible s_i.

Given sample sets S_1, S_2, the collection of disks $\mathcal{H}_2(S_! \cup S_2$ may not complete w.r.t. $S_1 \cup S_2$ (for the set $\mathcal{A}$ of all planar disks). However, the complexity of this algorithm, comparing with Algorithm 1, reduces to $O(M^2 \log M)$. The dominating term is the sorting of the sample points according to the distances to a certain sample point. Furthermore, this algorithm can also be modified for the relative discrepancy distance metric.

Algorithm 3: Search in Axis-aligned Rectangles

Let $\mathcal{A}$ be the collection of two dimensional axis-aligned rectangles. Given Samples S_1 and S_2, let $S = S_1 \cup S_2 = \{(x_1, y_1), \cdots, (x_M, y_M)\}$. We assume[3] that, $x_1 \leq x_2 \leq \cdots \leq x_M$. Consider the finite collection of axis-aligned rectangles $\mathcal{H}_3$ defined by

$$\mathcal{H}_3(S) \triangleq \{R(y_i, y_j, x_m, x_n) : (x_k, y_k) \in S, k = i, j, m, n\} \quad (6)$$

where $R(y_i, y_j, x_m, x_n)$ is the rectangle defined by the four lines $y = y_i, y = y_j, x = x_m, x = x_n$. See Figure 3.

Claim V.2 *Let $\mathcal{A}$ be the collection of two dimensional axis-aligned rectangles. Given S_1 and S_2 The finite collection $\mathcal{H}_3(S_1 \cup S_2)$ in (6) is complete with respect to $S_1 \cup S_2$.*

The search for the rectangle that has the maximum change in empirical distributions proceeds by searching in $\mathcal{H}_3$ with fixed y_i and y_j. Define

$$f_{ij}^k(n) \triangleq |S_k \cap R(y_i, y_j, x_1, x_n)|, k = 1, 2 \quad (7)$$

$$F_{ij}(n) = f_{ij}^1(n) - f_{ij}^2(n) \quad (8)$$

[2]This sort is at the cost of $O(M \log M)$.

[3]A sort of S may be needed at the cost of $O(M \log M)$.

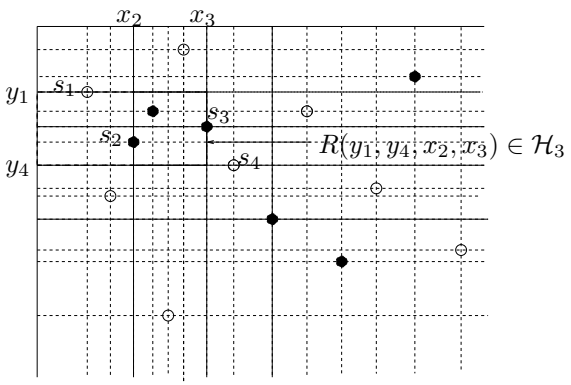


Figure 3: Members of $\mathcal{H}_3$

where $F_{ij}(n)$ is the accumulative difference in $R(y_i, y_j, x_1, x_n)$.

Compute

$$i_{max} = \arg\max_n F_{ij}(n), \quad i_{min} = \arg\min_n F_{ij}(n) \tag{9}$$

$$l \triangleq \min\{i_{max}, i_{min}\} + 1, \quad u \triangleq \max\{i_{max}, i_{min}\} \tag{10}$$

The optimal rectangle, for fixed y_i and y_j, is then given by $R(y_i, y_j, x_l, x_u)$. The search repeats for all possible pairs of y_i and y_j.

The family $\mathcal{H}_3$ that Algorithm 3 searches is complete for the family of all planar axis aligned rectangles, and its complexity $O(M^3)$ falls in between those of Algorithm 1 and Algorithm 2. Note that the VC dimension of the collection of axis-aligned rectangles is 4 whereas the VC dimension of the collection of planar disks is 3, which results in an increase in sample size M. It is also not obvious that this algorithm can be amended for the relativized discrepancy.

References

[1] V.N. Vapnik and A. Ya. Chervonenkis "On the uniform convergence of relative frequencie of events to their probabilities"" in Theory of Probability and its Applications, Vol. 16, pp 264-280, 1971.

[2] M. Talagrand, "Majorazing maesures: the generic chaining", in Annals of Probability, vol. 24, pp. 1049-1103, 1996.

[3] T. Batu, L. Fortnow, R. Rubinfeld, W.D. Smith, and P. White "Testing that distributions are close", in IEEE Symposium on foundations of Computer Science, pp. 259-269, 2000.

[4] M. Anthony and J. Shawe-Taylor, "A result of Vapnik with applications", in Discrete and Applied Mathematics, vol. 47(2), pp. 207-217, 1993.

[5] L. Tong, Q. Zhao, and S. Adireddy, "Sensor Networks with Mobile Agents," in Proc. IEEE 2003 MILCOM, (Boston, MA), Oct. 2003.

[6] B. Brodsky and B. Darkovsky, *Non-Parametric Methods in Change-Point Problems,* Kluwer Academic, The Netherlands, 1993.

[7] J. Shao, *Mathematical Statistics,* Springer, 1999.