NONPARAMETRIC CHANGE ESTIMATION IN 2D RANDOM FIELDS

Ting He, and Lang Tong[†]

School of Electrical and Computer Engineering

Cornell University

Email: {th255@,ltong@ece.}cornell.edu

and

Ananthram Swami

Army Research Laboratory

Adelphi, MD 20783, USA

Email: aswami@arl.army.mil

ABSTRACT

We consider the problem of change estimation in a 2D random field. Our goal is to estimate the locations of changes when they occur. We propose a nonparametric change estimator which can be implemented efficiently. Consistency result is derived for the proposed estimator under certain conditions. We analyze the performance of the proposed algorithms under several new estimation criteria.

Index Terms - Nonparametric change estimation, Consistent change estimation, Change estimation algorithms, Estimation performance criterion.

1. INTRODUCTION

We consider the problem of change estimation in a large-scale randomly deployed sensor field. For example, as illustrated in Fig. 1, sensors are deployed to detect certain chemical or biological components in some field. When the sensor measurement exceeds a certain threshold, the sensor is "alarmed". The state of a sensor is random, and it depends on its location and the time epoch when that measurement is taken. At the time of fusion, some (not necessarily all) of the alarmed sensors report to a fusion center.

For change detection and estimation, we consider two consecutive data collections, assuming that each report contains the location of the alarmed sensor. Note that due to duty cycling, mobility, multiple access and fading, and the randomness in sensing, the reporting sensors heard at the fusion center will be random and will not be the same in different data collection periods. The problem of change detection is to decide if the underlying distribution of the alarmed sensor has changed. Such a change may be an indication of the occurrence of certain events. The problem of change estimation, on the other hand, is to estimate the location of changes when there is a change of underlying distribution. In other words we would like to find the geographical locations where the distribution of alarmed sensors has changed. In this paper, specifically, we are interested in the region in which the change is the maximum.
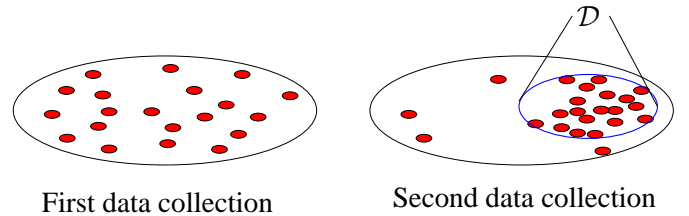

Fig. 1. Reported alarmed sensors (red) in two collections. $\mathcal{D}$: abnormal area

Applications of such change estimation could be found in infection control, pollution control as well as military needs such as defense against biochemical weapons. If there is an outbreak of disease, an accident causing environmental pollution, or existence of unusual biochemical components in a battle field, sensors can sense it and report to a fusion center. The fusion center aggregates the sensor reports to a best guess of the most abnormal area (*e.g.,* the area where the disease or pollutant is spreading fastest, or the en-

emy is using biochemical weapons most intensively), and immediate action is taken in this area.

In practice it is difficult to get accurate and updated information about the underlying distribution of alarmed sensors. Therefore, the change estimation rule is preferred to be nonparametric, thus this paper adopts a nonparametric framework.

1.1. Related Work

In the classical change-point estimation framework, the system changes at an unknown epoch called change-point, and we want to estimate this change-point. Such a setup belongs to the general family of point estimation because the estimand is a point in some vector space. In point estimation, the point-to-point divergence serves as a natural performance criterion; examples include squared error, absolute error, uniform cost, etc [1]. For nonparametric methods in change-point problem, see [2].

The change estimation considered in this paper belongs to another family called set estimation problems, where the estimand is a set in some space $\mathbb{R}^d$. The classical set estimation in the parametric framework is one of estimating the confidence set for some parameter. Assume that we have a family of distributions for the random observation Y, indexed by a parameter θ; *i.e.,* $\{P_\theta; \theta \in \Theta\}$. A $(1 - \alpha)$ confidence set for θ is a mapping S from the observation space $\mathcal{Y}$ to subsets of Θ such that

$$\inf_{\theta \in \Theta} \Pr\{\theta \in S(Y)|\theta\} = 1 - \alpha,$$

i.e., for every $\theta \in \Theta$, the estimated confidence set $S(Y_\theta)$ covers the true θ with probability at least $(1-\alpha)$.

In the nonparametric framework, set estimation has been applied in but not limited to statistical learning [3], change detection [4], as well as cluster analysis [5]. In statistical learning (also called classification or pattern recognition), we observe data-label pairs (X_i, Y_i) $(i = 1, \ldots, n)$ drawn i.i.d. from unknown joint distribution π, where $X \in \mathbb{R}^d$ and $Y \in \{0, 1\}$. Then given another X, the goal is to predict the corresponding $Y \in \{0, 1\}$. The classifier $\hat{Y}(x)$ is uniquely determined by a set $G = \{x \in \mathbb{R}^d : \hat{Y}(x) = 1\}$. If π is known, then the optimal classifier minimizing classification error $\Pr(Y \neq I(X \in G))$ ($I(\cdot)$ is the indicator

function) is

$$G^* = \{x : \Pr(Y = 1|X = x) \geq \frac{1}{2}\}.$$

Now that π is unknown, the problem boils down to estimating G^* from (X_i, Y_i), $i = 1, \ldots, n$.

Set estimation is utilized in change detection through the Devroye-Wise method [4]. Specifically, given i.i.d. observations $X_1, \ldots, X_n$ drawn from an unknown pdf f on $\mathbb{R}^d$, the problem is to decide whether a new observation X_{n+1} comes from f. The Devroye-Wise method is to construct a support estimator $\hat{S}_n$ estimating the set $\{f > 0\}$, and decide there is a change in the distribution of X_{n+1} if

$$X_{n+1} \notin \hat{S}_n.$$

Cluster analysis is the study of clusters, where a c-cluster (for constant $c > 0$) is defined as the connected components of the level set $\{f > c\}$, where f is the pdf. Hence the estimation of level sets can be an intermediate step [5].

In the change estimation problem, we compare two distributions and try to construct a measurable subset of the sample space, which exhibits the largest change[1] between the two distributions. It is an application of the nonparametric set estimation; its specific application, however, leads to a distinct structure of the estimator and a new guideline to performance criteria as will be seen later.

1.2. Summary of Results and Organization

In this paper we define two pseudo metrics on probability distributions. We formulate the change estimation problem as one of estimating the set that gives the maximum distance between two probability distributions according to the metric used. We define cost functions to evaluate the performance of change estimators. A nonparametric change estimator is proposed and its implementation is discussed. The proposed estimator is proved to be consistent under certain conditions, and its finite-sample performance is evaluated under the proposed cost functions.

---

[1]It depends on what metric on probabilities is used, as will be made clear later.

This paper is organized as follows. Section 2 specifies the model. Section 3 defines two notions of distances between probability distributions and the corresponding estimands in change estimation. It also defines cost functions for each proposed estimand. Section 4 defines estimators for the corresponding estimands and gives consistency results on the estimators. Section 5 analyzes the finite-sample performance of the proposed estimators by deriving upper bounds on the average costs for cost functions defined in Section 3. The paper is concluded with summary and comments on the estimator design.

2. PROBLEM STATEMENT

We consider the following setup: probability space $(X, \mathcal{F}, P_i)$[2] models the ith random collection of the locations of alarmed sensors. Denote S_i as the set of locations of alarmed sensors in the ith collection. We assume that, in each collection, sample points in S_i are drawn i.i.d. according to P_i and the drawings in different collections are independent. The specific form of P_i is unknown. We introduce a collection $\mathcal{A} \subseteq \mathcal{F}$ of measurable sets to model the set of geographical areas in which events of interest are observed.

Given a collection $\mathcal{A} \subseteq \mathcal{F}$, we answer in the sequel which set in $\mathcal{A}$ has the largest change in probability between P_1 and P_2, and how to find it.

3. ESTIMATION FORMULATION

We define two metrics[3] on $[0, 1]$:
$f_d : [0, 1] \times [0, 1] \to [0, 1]$,

$$f_d(x, y) = |x - y|,$$

and $f_\phi : [0, 1] \times [0, 1] \to [0, \sqrt{2}]$,

$$f_\phi(x, y) = \begin{cases} \frac{|x-y|}{\sqrt{\frac{x+y}{2}}} & \text{if } x \neq y \\ 0 & \text{if } x = y. \end{cases}$$

Based on them, we define the following pseudo metrics on probability distributions.

---

[2]The notation $(X, \mathcal{F}, P_i)$ is standard: X is the sample space, $\mathcal{F}$ the σ-field, P_i the probability measure.

[3]f_d is clearly a metric. For the proof that f_ϕ is a metric, see [6].

Definition 1 ($\mathcal{A}$-distance and Relative $\mathcal{A}$-distance) *Given probability spaces $(X, \mathcal{F}, P_i)$ and a collection $\mathcal{A} \subseteq \mathcal{F}$, the $\mathcal{A}$-distance between P_1 and P_2 is defined as*

$$d_\mathcal{A}(P_1, P_2) = \sup_{A \in \mathcal{A}} f_d(P_1(A), P_2(A)). \quad (1)$$

The relative $\mathcal{A}$-distance $\phi_\mathcal{A}(P_1, P_2)$ *is defined as*

$$\phi_\mathcal{A}(P_1, P_2) = \sup_{A \in \mathcal{A}} f_\phi(P_1(A), P_2(A)). \quad (2)$$

Empirical distances $d_\mathcal{A}(S_1, S_2)$ and $\phi_\mathcal{A}(S_1, S_2)$ are defined similarly by replacing $P_i(A)$ with the empirical probability measure

$$S_i(A) \triangleq \frac{|S_i \cap A|}{|S_i|}. \quad (3)$$

where $|\cdot|$ denotes the cardinality of a set.

Now we are ready to formulate the change estimation problem formally. Given a class $\mathcal{A}$ and a metric $d(x, y)$ on $[0, 1]$, we say that the set $A^* \in \mathcal{A}$ has the largest d-distance in $\mathcal{A}$ between distributions P_1, P_2 if

$$d(P_1(A^*), P_2(A^*)) = \sup_{A \in \mathcal{A}} d(P_1(A), P_2(A)). \quad (4)$$

The change estimation considered in this paper is the one of estimating A^*. Specifically, if $d = f_d$, the estimand is $A^* = A^*_{d_\mathcal{A}}$, where

$$f_d(P_1(A^*_{d_\mathcal{A}}), P_2(A^*_{d_\mathcal{A}})) = d_\mathcal{A}(P_1, P_2).$$

If $d = f_\phi$, then $A^* = A^*_{\phi_\mathcal{A}}$, where

$$f_\phi(P_1(A^*_{\phi_\mathcal{A}}), P_2(A^*_{\phi_\mathcal{A}})) = \phi_\mathcal{A}(P_1, P_2).$$

Given an estimator, we want to evaluate its performance. To this end, we define the following cost functions.

Definition 2 *Let $D \in \mathcal{A}$. To estimate $A^*_{d_\mathcal{A}}$, we define the following costs:*

1. Uniform cost: given Δ, $0 < \Delta < 1$,

$$R^{(U)}_{d; \Delta}(D) = I_{\{f_d(P_1(D), P_2(D)) \leq d_\mathcal{A}(P_1, P_2) - \Delta\}} \quad (5)$$

2. Linear cost: given $a > 0$,

$$R_{d;\,a}^{(L)}(D) = a\left[d_{\mathcal{A}}(P_1, P_2) - f_d(P_1(D), P_2(D))\right] \tag{6}$$

To estimate $A_{\phi_{\mathcal{A}}}^*$, we have similarly defined uniform cost and linear cost

$$R_{\phi;\,\Delta'}^{(U)}(D) = I_{\{f_\phi(P_1(D), P_2(D)) \le \phi_{\mathcal{A}}(P_1, P_2) - \Delta'\}} \tag{7}$$

$$R_{\phi;\,a'}^{(L)}(D) = a'\left[\phi_{\mathcal{A}}(P_1, P_2) - f_\phi(P_1(D), P_2(D))\right] \tag{8}$$

where $0 < \Delta' < \sqrt{2}$, $a' > 0$.

Note that both cost functions are nonnegative. Uniform cost is an "estimation error" with tolerance Δ, and linear cost is a linear penalty to the difference between the captured change and the maximum change.

The definition of cost functions depends on applications. In statistical learning, the goal of application is to minimize classification error, so a widely-used cost is the difference in the classification error probabilities between the estimate and the estimand [3]. In level set and support set estimation, the goal is to find the set itself, so costs with direct geometric interpretation are preferred; examples include measure-based costs such as the symmetric difference metric [4] and Euclidean distance-based costs such as minimal distance, maximal distance, maximin distance and Hausdorff distance [7, 8]. In the change estimation considered in this paper, great interest is usually paid to any set with a significant change (*e.g.,* investigators should be sent to areas with substantial increase in the pollutant level), and thus it is preferred that the sets are evaluated by their changes between probability distributions. Uniform cost and linear cost both serve this purpose.

4. CHANGE ESTIMATOR AND CONSISTENCY

Assuming that a change in the probabilities on $\mathcal{A}$ has occurred, we define the change estimator as follows:

Definition 3 (Estimator) *Given a class $\mathcal{A}$ and two collections of sample points S_1 and S_2, drawn i.i.d from*

probability distributions P_1 and P_2 respectively, the estimators for $A_{d_{\mathcal{A}}}^*$ and $A_{\phi_{\mathcal{A}}}^*$ are respectively defined as[4]

$$\hat{A}_{d_{\mathcal{A}}}^*(S_1,\ S_2) = \underset{A \in \mathcal{A}}{\arg\max}\, f_d(S_1(A),\ S_2(A)), \tag{9}$$

$$\hat{A}_{\phi_{\mathcal{A}}}^*(S_1,\ S_2) = \underset{A \in \mathcal{A}}{\arg\max}\, f_\phi(S_1(A),\ S_2(A)) \tag{10}$$

We point out that there are no general solutions to (9, 10). Implementation of the estimators depends on choices of $\mathcal{A}$. For several regular classes such as planar disks, axis-aligned rectangles and stripes, there are known algorithms to compute (9, 10). See [9].

We present next in this section the consistency results for the estimators defined above, whose proof is in [10].

Theorem 1 *Given probability spaces $(X, \mathcal{F}, P_i)$, $i = 1, 2$, and a collection $\mathcal{A} \subseteq \mathcal{F}$ with finite VC-dimension*[5]. *Assume $S_i\ (i = 1,\ 2)$ is a collection of n sample points drawn i.i.d. from P_i, and $S_1,\ S_2$ are independent. If $A_{d_{\mathcal{A}}}^*$ is unique, then*

$$\lim_{n \to \infty} \Pr\{\hat{A}_{d_{\mathcal{A}}}^*(S_1, S_2) = A_{d_{\mathcal{A}}}^*\} = 1.$$

If $A_{\phi_{\mathcal{A}}}^*$ is unique, then

$$\lim_{n \to \infty} \Pr\{\hat{A}_{\phi_{\mathcal{A}}}^*(S_1, S_2) = A_{\phi_{\mathcal{A}}}^*\} = 1.$$

5. PERFORMANCE ANALYSIS FOR FINITE SAMPLE SIZE

By Theorem 1 we see that both the average uniform cost and the average linear cost go to 0 as the sample size $n \to \infty$. In the propositions below we derive upper bounds on the average costs for finite n.

5.1. Average Uniform Cost

Proposition 1 *Given error tolerance Δ, $0 < \Delta < 1$, the average uniform cost of estimator $\hat{A}_{d_{\mathcal{A}}}^*$ is bounded by*[6]

$$\mathbb{E}[R_{d;\,\Delta}^{(U)}(\hat{A}_{d_{\mathcal{A}}}^*)] \le 8\Pi_{\mathcal{A}}(2n)e^{-n\Delta^2/50}.$$

---

[4] A predefined rule such as the smallest index rule is used to solve ties.

[5] The VC-dimension of $\mathcal{A}$ is the cardinality of the largest set shatterable by $\mathcal{A}$. See [9, 11]

[6] $\Pi_{\mathcal{A}}(n)$ is the *shatter coefficient* [12]. If $\mathcal{A}$ has a finite VC-dimension d, then by Sauer's Lemma, $\Pi_{\mathcal{A}}(n) < (n+1)^d$ for all n.

To prove Proposition 1, we first introduce a lemma on $|P(A) - S(A)|$.

Lemma 1 *For a measurable set A, $0 < \epsilon < 1$,*

$$\Pr\{|S(A) - P(A)| > \epsilon\} \leq 2e^{-2n\epsilon^2}.$$

Proof: Let $X_i, i = 1, \ldots, n$ be i.i.d sample points drawn from distribution P. Define

$$Z_n = \sum_{i=1}^{n} I_{\{X_i \in A\}}.$$

Then $\mathbb{E}[Z_n] = nP(A)$.

By Hoeffding's Inequality [13],

$$\Pr\{S(A) - P(A) > \epsilon\} = \Pr\{Z_n - \mathbb{E}[Z_n] > n\epsilon\}$$
$$\leq e^{-2n^2\epsilon^2/\sum_{i=1}^{n} 1^2} = e^{-2n\epsilon^2},$$
$$\Pr\{S(A) - P(A) < -\epsilon\} = \Pr\{Z_n - \mathbb{E}[Z_n] < -n\epsilon\}$$
$$\leq e^{-2n^2\epsilon^2/\sum_{i=1}^{n} 1^2} = e^{-2n\epsilon^2}.$$

Hence, $\Pr\{|S(A) - P(A)| > \epsilon\} \leq 2e^{-2n\epsilon^2}$. $\blacksquare$

Proof: [Proposition 1] Let $D = \hat{A}_{d_{\mathcal{A}}}^{*}$.

$$\mathbb{E}[R_{d;\,\Delta}^{(U)}(D)]$$
$$= \Pr\{|P_1(D) - P_2(D)| \leq d_{\mathcal{A}}(P_1, P_2) - \Delta\}$$
$$\leq \Pr\{(d_{\mathcal{A}}(P_1, P_2) - d_{\mathcal{A}}(S_1, S_2))$$
$$\quad + |P_1(D) - S_1(D)| + |P_2(D) - S_2(D)| \geq \Delta\}$$
$$\leq \Pr\{d_{\mathcal{A}}(P_1, P_2) - d_{\mathcal{A}}(S_1, S_2) \geq \frac{4\Delta}{5}\}$$
$$\quad + \Pr\{|P_1(D) - S_1(D)| \geq \frac{\Delta}{10}\}$$
$$\quad + \Pr\{|P_2(D) - S_2(D)| \geq \frac{\Delta}{10}\}$$
$$\leq \Pr\{\sup_{A \in \mathcal{A}} |f_d(P_1(A), P_2(A))$$
$$\quad - f_d(S_1(A), S_2(A))| \geq \frac{4\Delta}{5}\}$$
$$\quad + \Pr\{|P_1(D) - S_1(D)| \geq \frac{\Delta}{10}\}$$
$$\quad + \Pr\{|P_2(D) - S_2(D)| \geq \frac{\Delta}{10}\}$$
$$\leq 8\Pi_{\mathcal{A}}(2n)e^{-n\Delta^2/50} + 2e^{-n\Delta^2/50} + 2e^{-n\Delta^2/50}$$

$$\text{(11)}$$

$$\approx 8\Pi_{\mathcal{A}}(2n)e^{-n\Delta^2/50}$$

where (11) follows from Lemma 1 and the following inequality from [14]

$$\Pr\{\sup_{A \in \mathcal{A}} |f_d(P_1(A), P_2(A)) - f_d(S_1(A), S_2(A))| \geq \epsilon\}$$
$$< 8\Pi_{\mathcal{A}}(2n)e^{-n\epsilon^2/32}.$$
$\blacksquare$

For the estimator $\hat{A}_{\phi_{\mathcal{A}}}^{*}$, we have similar results:

Proposition 2 *Given error tolerance $\Delta' \in (0, \sqrt{2})$, the average uniform cost of estimator $\hat{A}_{\phi_{\mathcal{A}}}^{*}$ is bounded by*

$$\mathbb{E}[R_{\phi;\,\Delta'}^{(U)}(\hat{A}_{\phi_{\mathcal{A}}}^{*})] \leq 16\Pi_{\mathcal{A}}(2n)e^{-n\Delta'^2/64}.$$

Lemma 2 *For a measurable set A and $\epsilon \in (0, \sqrt{2})$*

$$\Pr\{f_{\phi}(P(A), S(A)) > \epsilon\} \leq 8e^{-n\epsilon^2/4}.$$

Proof: Define

$$Q = \left\{ S \in X^n : \frac{P(A) - S(A)}{\sqrt{\frac{P(A)+S(A)}{2}}} > \epsilon \right\}.$$

$$R = \left\{ SS' \in X^{2n} : \frac{S'(A) - S(A)}{\sqrt{\frac{S'(A)+S(A)}{2}}} > \epsilon \right\}.$$

It is easy to see that if $S \in Q$ and $S'(A) \geq P(A)$, then $SS' \in R$. It is known from [14] that for $n > \frac{4}{\epsilon^2}$, $\Pr\{S'(A) \geq P(A)\} \geq \frac{1}{4}$.

Thus for $n > \frac{4}{\epsilon^2}$, $\Pr(Q) \leq 4\Pr(R)$.

In [15] it was proved that $\Pr(R) \leq e^{-n\epsilon^2/4}$. Thus $\Pr(Q) \leq 4e^{-n\epsilon^2/4}$.

By symmetry, we have

$$\Pr\{f_{\phi}(P(A), S(A)) > \epsilon\} \leq 8e^{-n\epsilon^2/4}.$$

The proof completes by noting that the bound trivially holds if $n \leq \frac{4}{\epsilon^2}$. $\blacksquare$

The proof of Proposition 2 is similar to that of Proposition 1. The key is to use Lemma 2 and the following inequality from [14]

$$\Pr\{|\phi_{\mathcal{A}}(P_1, P_2) - \phi_{\mathcal{A}}(S_1, S_2)| \geq \epsilon\}$$
$$< 16\Pi_{\mathcal{A}}(2n)e^{-n\epsilon^2/16}.$$

We have derived upper bounds on the average uniform costs for $\hat{A}_{d_{\mathcal{A}}}^{*}$ and $\hat{A}_{\phi_{\mathcal{A}}}^{*}$. Note that both upper bounds decay exponentially with respect to sample size n if $\mathcal{A}$ has finite VC-dimension.

5.2. Average Linear Cost

Since a, $a' > 0$ are just scaling factors, we let $a = a' = 1$ without loss of generality.

Proposition 3 *The average linear cost of $\hat{A}_{d_{\mathcal{A}}}^*$ is bounded by*

$$\mathbb{E}[R_{d;\,1}^{(L)}(\hat{A}_{d_{\mathcal{A}}}^*)] \leq \sqrt{\frac{2\pi}{n}}.$$

Lemma 3 *For any measurable set D,*

$$\mathbb{E}[f_d(P(D), S(D))] \leq \sqrt{\frac{\pi}{2n}}.$$

Proof: $f_d(P(D), S(D))$ is a nonnegative random variable on $[0, 1]$. Therefore,

$$
\begin{aligned}
\mathbb{E}[f_d(P(D), S(D))] &= \int_0^1 \Pr\{|S(D) - P(D)| > x\} dx \\
&\leq \int_0^1 2e^{-2nx^2} dx \quad (12) \\
&\leq 2 \int_0^\infty e^{-2nx^2} dx \\
&= \frac{2}{\sqrt{2n}} \int_0^\infty e^{-x^2} dx \\
&= \sqrt{\frac{\pi}{2n}}.
\end{aligned}
$$

where (12) is from Lemma 1. $\blacksquare$

Proof: [Proposition 3] Let $D = \hat{A}_{d_{\mathcal{A}}}^*$.

$$
\begin{aligned}
&\mathbb{E}[R_{d;\,1}^{(L)}(D)] \\
=\ & d_{\mathcal{A}}(P_1, P_2) - \mathbb{E}[f_d(P_1(D), P_2(D))] \\
\leq\ & d_{\mathcal{A}}(P_1, P_2) - \mathbb{E}[d_{\mathcal{A}}(S_1, S_2)] + \mathbb{E}[f_d(P_1(D), S_1(D))] \\
& + \mathbb{E}[f_d(P_2(D), S_2(D))]
\end{aligned}
$$

Note that by convexity of $\sup$ and $|\cdot|$,

$$
\begin{aligned}
\mathbb{E}[d_{\mathcal{A}}(S_1, S_2)] &= \mathbb{E}[\sup_{A \in \mathcal{A}} |S_1(A) - S_2(A)|] \\
&\geq \sup_{A \in \mathcal{A}} \mathbb{E}[|S_1(A) - S_2(A)|] \\
&\geq \sup_{A \in \mathcal{A}} |\mathbb{E}[S_1(A) - S_2(A)]| \\
&= \sup_{A \in \mathcal{A}} |P_1(A) - P_2(A)| \\
&= d_{\mathcal{A}}(P_1, P_2)
\end{aligned}
$$

So $d_{\mathcal{A}}(P_1, P_2) - \mathbb{E}[d_{\mathcal{A}}(S_1, S_2)] \leq 0$.

By Lemma 3, $\mathbb{E}[f_d(P_i(D), S_i(D))] \leq \sqrt{\frac{\pi}{2n}}$, $i = 1, 2$.

Hence $\mathbb{E}[R_{d;\,1}^{(L)}(D)] \leq \sqrt{\frac{2\pi}{n}}$. $\blacksquare$

Now let us consider the linear cost of $\hat{A}_{\phi_{\mathcal{A}}}^*$.

Lemma 4 *For any measurable set D,*

$$\mathbb{E}[f_\phi(P(D), S(D))] \leq 8\sqrt{\frac{\pi}{n}}.$$

The proof is similar to that of Lemma 3 (using Lemma 2 instead), and is omitted here.

Proposition 4 *The average linear cost of $\hat{A}_{\phi_{\mathcal{A}}}^*$ is bounded by*

$$\mathbb{E}[R_{\phi;\,1}^{(L)}(\hat{A}_{\phi_{\mathcal{A}}}^*)] \leq 32\sqrt{\frac{\pi}{n}}.$$

Proof: Let $D = \hat{A}_{\phi_{\mathcal{A}}}^*$.

$$
\begin{aligned}
\mathbb{E}[R_{\phi;\,1}^{(L)}(D)] \leq\ & \phi_{\mathcal{A}}(P_1, P_2) - \mathbb{E}[\phi_{\mathcal{A}}(S_1, S_2)] \\
& + \mathbb{E}[f_\phi(P_1(D), S_1(D))] + \mathbb{E}[f_\phi(P_2(D), S_2(D))] \quad (13)
\end{aligned}
$$

We then need to bound $\mathbb{E}[\phi_{\mathcal{A}}(S_1, S_2)]$ from below.

$$
\begin{aligned}
\mathbb{E}[\phi_{\mathcal{A}}(S_1, S_2)] &\geq \sup_{A \in \mathcal{A}} \mathbb{E}[f_\phi(S_1(A), S_2(A))] \\
&\geq \sup_{A \in \mathcal{A}} f_\phi(P_1(A), P_2(A)) - 16\sqrt{\frac{\pi}{n}}
\end{aligned}
$$
$$(14)$$

where (14) is because for a fixed $A \in \mathcal{A}$,

$$
\begin{aligned}
&\mathbb{E}[f_\phi(S_1(A), S_2(A))] \\
\geq\ & \mathbb{E}[f_\phi(P_1(A), P_2(A)) - f_\phi(S_1(A), P_1(A)) \\
& - f_\phi(S_2(A), P_2(A))] \quad (15) \\
\geq\ & f_\phi(P_1(A), P_2(A)) - 16\sqrt{\frac{\pi}{n}} \quad (16)
\end{aligned}
$$

where (15) is by the triangle inequality [6], and (16) is from Lemma 4, and since this holds for any $A \in \mathcal{A}$, taking $\sup$ on both sides yields (14).

Then by plugging Lemma 4 and (14) into (13) we have

$$\mathbb{E}[R_{\phi;\,1}^{(L)}(D)] \leq 32\sqrt{\frac{\pi}{n}}.$$

$\blacksquare$

We have shown upper bounds on the average linear costs, which decay at the rate of $O(\frac{1}{\sqrt{n}})$. We point out that the actual decay rates of the average costs may be higher.

6. CONCLUSION

We[7] consider in this paper the problem of nonparametric change estimation in 2D random field, which is formulated as a nonparametric set estimation problem. Our goal here is to locate a data set that exhibits the most change in probabilities. Another type of change estimation is that of change value estimation, where based on some metrics on probability distributions, the goal is to estimate the distance between two distributions.

As for the estimator design strategy, we note that the proper choice of class $\mathcal{A}$ is crucial. Although given $\mathcal{A}$ our estimators are independent of the distributions, the knowledge about the distributions before and after the change may help in defining $\mathcal{A}$ to capture such a change properly and thus facilitate the change estimation.

7. REFERENCES

[1] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.

[2] B. E. Brodsky and B. S. Darkovsky, *Non-Parametric Methods in Change-Point Problems*. The Netherlands: Kluwer Academic Publishers, 1993.

[3] A. B. Tsybakov, "Optimal aggregation of classifiers in statistical learning," *Annals of Statistics*, vol. 32, no. 1, 2004.

[4] A. Baillo, A. Cuevas, and A. Justel, "Set estimation and nonparametric detection," *The Canadian Journal of Statistics*, vol. 28, pp. 765–782, 2000.

[5] W. Polonik, "Measuring mass concentrations and estimating density contour clusters. An excess mass approach," *Annals of Statistics*, vol. 23, pp. 855–882, 1995.

[6] S. Ben-David, J. Gehrke, and D. Kifer, "Detecting Change in Data Streams," in *Proc. 2004 VLDB Conference*, (Toronto, Canada), 2004.

[7] A. Trybulec, "On the minimal distance between sets in Euclidean space," *Journal of Formalized Mathematics*, vol. 14, 2002.

[8] A. Grabowski, "On the Hausdorff distance between compact subsets," *Journal of Formalized Mathematics*, vol. 15, 2003.

[9] T. He, S. Ben-David, and L. Tong, "Nonparametric Change Detection and Estimation in Large Scale Sensor Networks." submitted to IEEE Trans. on Signal Processing, December 2004.

[10] T. He and L. Tong, "Consistent Change Estimation in 2D Random Fields," Tech. Rep. ACSP-TR-08-05-01, Cornell University, August 2005. http://acsp.ece.cornell.edu/pubR.html.

[11] V. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencie of events to their probabilities," *Theory of Probability and its Applications*, vol. 16, pp. 264–280, 1971.

[12] L. Gyorfi, *Principles of Nonparametric Learning*. New York, NY: Springer Wien New York, 2002.

[13] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, pp. 13–30, March 1963.

[14] T. He and L. Tong, "On $\mathcal{A}$-distance and Relative $\mathcal{A}$-distance," Tech. Rep. ACSP-TR-08-04-02, Cornell University, August 2004. http://acsp.ece.cornell.edu/pubR.html.

[15] M. Anthony and J. Shawe-Taylor, "A result of Vapnik with applications," *Discrete and Applied Mathematics*, vol. 47, no. 2, pp. 207–217, 1993.

---

[7]The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U. S. Government.