ADAPTIVE COMMUNICATIONS AND SIGNAL PROCESSING LABORATORY
CORNELL UNIVERSITY, ITHACA, NY 14853

Consistent Change Estimation in 2D Random Fields

Ting He and Lang Tong

Technical Report No. ACSP-TR-08-05-01

August 2005

I. INTRODUCTION

We formulate the problem of change location estimation in 2D random fields as a nonparametric set estimation problem. We define nonparametric estimators to solve this problem, and prove that our estimators are consistent under certain conditions.

II. PROBLEM FORMULATION

We consider the following setup: $P_1,\ P_2$ are probability measures on a common measurable space $(X,\ \mathcal{F})$[1]. $S_i,\ i = 1,\ 2$ is the set of realizations drawn i.i.d. according to P_i, and $S_1,\ S_2$ are independent. The specific form of P_i is unknown. Assume $P_1 \neq P_2$. Given $S_1,\ S_2$, our goal is to estimate the subset of X that has the largest change in probability between P_1 and P_2.

To this end, we introduce a collection $\mathcal{A} \subseteq \mathcal{F}$ of measurable sets to model the subsets of X in which events of interest are observed. Given the collection $\mathcal{A}$, the problem becomes which set in $\mathcal{A}$ has the largest change in probability between P_1 and P_2.

We define two metrics[2] on $[0,\ 1]$:

$f_d : [0, 1] \times [0, 1] \rightarrow [0, 1]$,

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

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

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

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

$\mathcal{A}$-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

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

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

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

[1] X is the sample space, and $\mathcal{F}$ is the σ-field.

[2] f_d is clearly a metric. For the proof that f_ϕ is a metric, see [1].

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

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

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

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

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

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

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

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

III. Change Estimator and Consistency

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

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[3]

$$\hat{A}^*_{d_{\mathcal{A}}}(S_1,\ S_2) = \arg\max_{A\in\mathcal{A}} f_d(S_1(A),\ S_2(A)), \tag{5}$$

$$\hat{A}^*_{\phi_{\mathcal{A}}}(S_1,\ S_2) = \arg\max_{A\in\mathcal{A}} f_\phi(S_1(A),\ S_2(A)). \tag{6}$$

We point out that there are no general solutions to (5, 6). 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 (5, 6). See [2].

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

We present next in this section the consistency results for the estimators defined above, which are stated in the following theorem.

Theorem 3.1: Given probability spaces $(X, \mathcal{F}, P_i)$, $i = 1, 2$, and a collection $\mathcal{A} \subseteq \mathcal{F}$ with finite VC-dimension[4], if $A^*_{d_\mathcal{A}}$ is unique, then

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

Similarly, if $A^*_{\phi_\mathcal{A}}$ is unique, then

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

Proof: Let $A = A^*_{d_\mathcal{A}}$. Let $VC\text{-}d(\mathcal{A}) = d < \infty$. Define η to be

$$\eta \triangleq f_d(P_1(A),\ P_2(A)) - \sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_d(P_1(B),\ P_2(B))$$

The uniqueness of A guarantees that $\eta > 0$.

By results of [4], we have

$$\Pr\{\sup_{B\in\mathcal{A}} |f_d(P_1(B),\ P_2(B)) - f_d(S_1(B),\ S_2(B))| \leq \frac{\eta}{3}\}$$
$$\geq 1 - 8(2n+1)^d e^{-n\eta^2/288}.$$

So with probability $\geq 1 - 8(2n+1)^d e^{-n\eta^2/288}$,

$$\begin{aligned}
& f_d(S_1(A),\ S_2(A)) - \sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_d(S_1(B),\ S_2(B)) \\
\geq\ & f_d(P_1(A),\ P_2(A)) - \sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_d(P_1(B),\ P_2(B)) \\
& -|f_d(S_1(A),\ S_2(A)) - f_d(P_1(A),\ P_2(A))| \\
& -\Big| \sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_d(S_1(B),\ S_2(B)) - \sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_d(P_1(B),\ P_2(B)) \Big| \qquad (7) \\
\geq\ & \eta - 2\sup_{B\in\mathcal{A}} |f_d(P_1(B),\ P_2(B)) - f_d(S_1(B),\ S_2(B))| \qquad (8) \\
\geq\ & \frac{\eta}{3} \qquad (9)
\end{aligned}$$

[4]The VC-dimension of $\mathcal{A}$ is the cardinality of the largest set shatterable by $\mathcal{A}$. See [2], [3]

That is,

$$\Pr\{\hat{A}^*_{d_{\mathcal{A}}}(S_1,S_2)=A\} \geq 1-8(2n+1)^d e^{-n\eta^2/288}.$$

Now let $n\to\infty$,

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

For relative $\mathcal{A}$-distance, the proof is similar. Let $A=A^*_{\phi_{\mathcal{A}}}$ and

$$\eta \triangleq f_\phi(P_1(A),\ P_2(A)) - \sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_\phi(P_1(B),\ P_2(B)).$$

The condition $A^*_{\phi_{\mathcal{A}}}$ is unique implies that $\eta>0$.

By [4] we have

$$\Pr(\sup_{B\in\mathcal{A}} f_\phi(S_i(B),\ P_i(B)) \leq \frac{\eta}{5}) \geq 1-8(2n+1)^d e^{-n\eta^2/100}.$$

Since $f_\phi(x,y)$ satisfies triangle inequality ([1]), with probability $\geq [1-8(2n+1)^d e^{-n\eta^2/100}]^2$, we have

$$\begin{aligned}
& f_\phi(S_1(A),\ S_2(A)) - \sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_\phi(S_1(B),\ S_2(B)) \\
\geq\ & f_\phi(P_1(A),\ P_2(A)) - \sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_\phi(P_1(B),\ P_2(B)) \\
& -|f_\phi(P_1(A),\ P_2(A)) - f_\phi(S_1(A),\ S_2(A))| \\
& -|\sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_\phi(P_1(B),\ P_2(B)) - \sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_\phi(S_1(B),\ S_2(B))| \qquad (10)\\
\geq\ & \eta - f_\phi(P_1(A),\ S_1(A)) - f_\phi(P_2(A),\ S_2(A)) \\
& -\sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_\phi(P_1(B),\ S_1(B)) - \sup_{\substack{B\in\mathcal{A}\\ B\neq A}} f_\phi(P_2(B),\ S_2(B)) \qquad (11)\\
\geq\ & \eta - 2\sup_{B\in\mathcal{A}} f_\phi(P_1(B),\ S_1(B)) - 2\sup_{B\in\mathcal{A}} f_\phi(P_2(B),\ S_2(B)) \qquad (12)\\
\geq\ & \frac{\eta}{5} \qquad (13)
\end{aligned}$$

Hence

$$\Pr\{\hat{A}^*_{\phi_{\mathcal{A}}}(S_1,S_2)=A\} \geq [1-8(2n+1)^d e^{-n\eta^2/100}]^2.$$

Now let $n\to\infty$, the proof completes.

■

REFERENCES

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

[2] 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.

[3] 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.

[4] 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`.