ON CONSISTENT ESTIMATION OF CLASSES IN R^2 IN THE CONTEXT OF CLUSTER ANALYSIS

C. A. MURTHY

Thesis submitted to Indian Statistical Institute in partial fulfilment of the requirements for the award of the degree of
DOCTOR OF PHILOSOPHY, CALCUTTA, 1988.

Acknowledgements

I am grateful to Professor D. Dutta Majumder, my supervisor, for his help and encouragement during the period of work. This work would not have been possible without him.

I wish to acknowledge Ms. L. Lalitha for her help in writing the computer programme. I also like to thank Dr. S.K. Parui, Mr. S.E.Sarma, Mr. B. Chanda, Dr. S.K. Pal, Dr. B.B. Chaudhuri and Prof. A.K. Datta for their encouragement at various stages of work. I am grateful to Mr. M. Deekshitulu, Mr. M. Vijaykumar, Mr. B. Rajeev, Dr. V.V. Srivatsa, Mr. P. Gajapati, Dr. A.P. Reddy, Dr. R. Barua, Dr. V.R. Padmawar and Mr. M. Chakraborty for their help and encouragement.

I also like to thank Mr. N. Chatterjee for typing the manuscript.

Above all, I like to thank Dr. A. Maitra, Dr. J.K. Ghosh and Dr. B.V. Rao. They have always been my sources of inspiration.

Lastly I am grateful to my family members for their patience.

Contents

I. INTRODUCTION

1.1 Brief review of cluster analysis

The literature on cluster analysis is basically oriented towards the development of algorithms [1,2,3,4,5]. Anderberg [1] gives the various steps in cluster analysis starting from the choice of data points to interpreting the results. Hartigan [5] describes various clustering algorithms in his book. He also states the uses of those methods in various fields. Jardine and Sibson [6] develops measures of dissimilarity and regards a cluster method as a function from dissimilarity matrices to trees.

Clustering techniques can be broadly divided into two categories, (i) Hierarchical (ii) Non hierarchical [1]. Hierarchical clustering techniques begin with a triangular similarity matrix, whose rows and columns correspond to patterns and whose entries measure similarity, the larger the entries, the more similar the patterns. The output of an hierarchical clustering program is a dendrogram, which is a tree showing a sequence of nested clusterings. This graphical output is the main feature of such programs since several clusterings are represented on the same picture.

Hierarchical methods may again be divided into two categories namely (i) agglomerative and (ii) divisive [1]. The general procedure for agglomerative clustering on a data consisting of n entities is as follows :-

1. Begin with n clusters each consisting of exactly one entity. Let the clusters be labelled with the numbers 1 to n.

2. Search the similarity matrix for the most similar pair of clusters. Let the chosen clusters be labelled p and q and let their associated similarity be s_{pq}, $p > q$.

3. Reduce the number of clusters by 1 through the merger of clusters p and q. Label the product of the merger q and update the similarity matrix entries in order to reflect the revised similarities between cluster q and all other existing clusters. Delete the row and column of S (similarity matrix) pertaining to cluster q.

4. Perform steps 2 and 3 a total of (n-1) times (at which point all entities will be in one cluster). At each stage record the identity of the clusters which are merged along with the value of similarity between them in order to have a complete record of them.

Different agglomerative methods are implemented by varying the procedures used for defining the most similar pair at step 2 and updating the similarity matrix at step 3. Some of the methods are single linkage, complete linkage, average linkage.

Agglomerative techniques assume in the beginning that there are n clusters and end at a point where there is only one cluster. The divisive techniques are focused on finding the groups which are the best separated from each other or most distinctive as opposed to the agglomerative notion of putting together the entities which are most alike.

Hierarchical techniques are useful when the number of clusters is unknown. On the other hand, non hierarchical techniques are useful when the number of clusters is known. The central idea in most of these methods [1] is to choose some initial partition of the data units and then alter cluster memberships so as to obtain a 'better' partition. The various algorithms which have been proposed differ as to what constitutes a 'better' partition and what methods may be used for achieving improvements.

Many of the non-hierarchical methods concentrate on minimizing the squared error, which try to define clusters that are hyper ellipsoidal in shape [7]. To put it mathematically, let the ith entity, $i=1,\ldots n$ from the data set be written as x_i. A clustering is a partition $c_1, c_2, \ldots c_k$ of the integers $1,2,\ldots n$ that assigns each entity a single cluster label. Let n_j be the number of elements of c_j, $j=1,\ldots k$. Let $\bar{x}_j = \frac{1}{n_j} \sum_{i \in c_j} x_i$. Then the squared error for the cluster c_j is $e_j^2 = \sum_{i \in c_j} (x_i - \bar{x}_j)'(x_i - \bar{x}_j)$ and the squared error for the clustering is $E^2 = \sum_{j=1}^{k} e_j^2$. The notable methods here are Forgy's method [1,8] ISODATA [1,9],

Macqueen's k-means method [10] is also a clustering method which is different from the previous method [1]. The key implication in this process is that the cluster centroid is calculated on the basis of the clusters' current membership rather than its membership at the end of the last reallocation. Wishart's variant on k means [1,11] is a squared error minimizing technique, though hierarchical. Diday's dynamic clustering [15] is closely related to k-means and a generalization of ISODATA.

Among the other techniques of clustering, minimal spanning tree is used to separate clusters that are spaced along a straight line or on a sheet [1,7,12] . This procedure is equivalent to cutting the dendogram generated by the single link hierarchical clustering procedure at a level equal to the threshold.

Another clustering procedure is that of R.F. Ling [13,14] , the input of which is ordinal data. It may be looked as a generalization of single linkage method with the condition that each cluster will contain at least a certain number of entities. J.P. Rasson et al suggested an admissible clustering procedure [46], admissibility in the sense of Fisher and Van Ness [24]. They showed that the maximum likelihood solution for the clustering problem is constituted by k groups of points such that the sum of the lebesgue measures of their convex hulls is the minimum. But the convex admissibility condition [24] need not be satisfied by the data points as illustrated in fig. 1.1.1. Simultaneously, it may be mentioned here that it may not be possible to characterize the best clustering procedure [24,46]. The admissibility conditions provide only a way of evaluating clustering criteria.

1.2 Comparison of the clustering methods and definitions of clusters

One of the main questions faced by a user of clustering algorithms is that, which algorithm is to be applied for a given set of data. In order to answer this question the user must be acquainted with (i) various properties of the algorithms and (ii) comparisons among them. L. Fisher et al [24] defined admissibility criteria for clustering methods. Gnanadesikan et al [25], Sidak [26], Rand [27], Dubes and Jain [7] dealt with this question in their papers. There are other references too in this connection [28, 29, 30, 31] .

Another question in this regard is 'How do you know when you have a good set of clusters ?' [1]. Anderberg [1] gives three possible answers to this question. "(i) The clusters are like statistics calculated from the data like mean and variance. There is no room to assess whether the clusters are 'good' or 'bad'. (ii) The clustering procedure is to be so well defined as to ensure that the clusters have the desired properties inherent in their method of generation. The clusters themselves cannot be subject to question as long as the procedure is accepted as valid. (iii) Many times, too little is known about the data set to say whether a given clustering procedure should produce a set of clusters with any relevance .....' once a satisfactory structure is known and defended on its own merits, any cluster analysis that contributed to its discovery is only of historical interest. But without such a structure in abstract, there is no criterion for judging goodness of clusters in numerical data".

1.2.a The main question that is treated in this thesis is, whether such an abstract structure can be defined. In case it can be defined, how to reach it on the basis of finitely many observations. The observations are assumed to be in R^2 and the euclidean distance is used as dissimilarity measure.

Many authors dealt with this question in various contexts. Everitt [30] called for more critical approach to cluster analysis instead of developing new algorithms. Cormack [32] stressed the need for methods based on well defined mathematical formulation. He clearly says 'The growing tendency to regard numerical taxonomy, as a satisfactory alternative to clear thinking is condemned'.

Hubert [33] defines subsets of data items that satisfy specific restrictions as 'perfect clusters'. Definitions of perfect clusters usually involve an index of compactness which measures cohesiveness among data items and/or an index of isolation which measures separation among clusters. Mcquitty [34] defined a 'comprehensive type' cluster as a cluster in which each data item is more like every other data item in the cluster than it is like any data item not in the cluster. In a 'restricted type' cluster each data item is more like some other data item in the clusters than it is like any data item outside the cluster.

Van Rijsbergen's perfect cluster [35] requires that the least similar pair of data items in a cluster be more alike than the most similar pair comprised of one data item from the cluster and one from outside it. This idea of perfect cluster is more restrictive than Mcquitty's perfect cluster. Day [36] extended this approach to the overlapping case in which data items can belong to more than one cluster. He defines consistency and authenticity properties and determines whether classes of clustering methods have these properties. But all these definitions are based on data alone. There is no assumption on the abstract structure. Even though these definitions indirectly give rise to such an abstract structure in the population, proper mathematical formalization has not been done. A few more papers in this context are by Dubes et al [37,38] in which clustering tendency, compactness and isolation indices are considered as well as hierarchical methods are suggested for validating clusters. But, again, no abstract theoretical set up on the population was considered in them.

We shall review briefly the literature on statistical pattern classification before attempting to understand the abstract structure of the data.

1.3 Brief review of statistical pattern classification

Statistical pattern recognition [16,17,18] consists of feature selection and classification. Clustering is called as non supervised learning in the context of pattern recognition. Here we will review classification.

The problem of classification is an age old problem in statistics. There are quite a few standard methods in the literature. Notable among them are Bayes decision rules for minimum risk, minimum error rate and minimax rule [18]. If the measurement process is sequential then at each step the observer is to decide whether one more observation is to be taken or the decision is to be made on the basis of the available observations. Sequential decision theory was pioneered by Wald [19].

An alternative to Bayes decision rule is Neyman Pearson decision rule which is thoroughly treated in many statistical text books [48,49] .

The above classification techniques rely heavily on the assumed knowledge of the underlying statistical distribution. By contrast, a family of decision rules, namely the nearest neighbour rules, apparently ignore these distributions. Roughly speaking, nearest neighbour rules exchange the need to know the underlying distributions for that of knowing a large number of correctly classified patterns. If there are n correctly classified observations from m classes, each class being represented by at least one observation, then a new observation x will be put in that class j if $d(x,x') = \min_{i=1,\dots,n} d(x,x_i)$ where x' is in the jth classand d is a metric on the observation space, x_i's are the n observations. This decision rule is 1-NNR. (one nearest neighbour rule). A natural extension to it is k-NNR which consists of searching k nearest neighbours for x. The nearest neighbour rule first appeared in two important reports by Fix and Hodges [20,21]. The proof of convergence is due to Cover and Hart [21].

Another approach for classification is to consider discriminant functions. A functional form for the discriminant function is to be selected and the methods for estimating the unknowns are to be developed. The statistical theory of discriminant analysis finds its origin in a paper by R.A. Fisher [23]. Since then the published literature on discriminant functions has grown to the point where even specialized bibliographies can contain hundreds of references.

Observe that in the above mentioned methods either class conditional densities are assumed to be known - which is a stringent assumption - or the word 'class' is not defined mathematically. The main thrust of this thesis is to define a 'class' in R^2 and find out a method of classification and estimation of 'classes'.

1.4 Finding the shape of a set of points

The problem of finding the 'shape' of a finite set of points has been considered in the literature. 'Shape' is not a well defined idea. Naturally there are various methods in the literature to find the rough shape of a finite set of points [39, 40, 41, 42, 43]. The methods suggested are basically trial and error methods.

The method suggested by Edelsbrunner et al [43] is the following. Remove from the whole space open discs of radius $r > 0$, r is to be chosen suitably, such that they do not have any intersection with the set given. Grenander [44] arrived at the same conclusion earlier. He, by using the method of sieves, showed that a consistent maximum likelihood estimator exists for the set under the assumption of uniform distribution. But there is no automatic way of finding r. Grenander [45] in the same book, in another place has shown that a consistent estimate for the set can be obtained by drawing discs of radius r around every given point under the assumption of uniform distribution. Again there is no automatic way of getting the radius r.

Another contribution in this context is that of S.K. Parui et al [47,50,51,52]. He defined shape of an object as that property which is invariant under translation, dilation and rotation. Shape distances were also found and used in the processing of binary pictures. But, the problem of finding the set given a few random points in it is not considered.

In this thesis the classes will be defined and consistent estimates are found in such a way that this difficulty is resolved. In the process the 'shape' of the set of points is also obtaindd. The procedure is such that it can be implemented on a computer.

1.5 Scope of the thesis

As it was mentioned in the previous sections, the problem of defining and estimating classes will be treated in this thesis. The space in which the classes will be defined is considered to be R^2, because any structure imposed on R^2 can be visualized. The scope for generalization is also discussed in the thesis.

The partitionings in the population are called classes and the same in the sample as clusters. This distinction is made for the sake of clarity.

In many problems, the number of classes may be unknown. Hierarchical clustering methods are particularly helpful in these cases. Here it is assumed that the number of classes is known.

The measure of dissimilarity between two observations is assumed to be the euclidean distance between them. Cormack [32] gives a list of dissimilarity measures generally used in practice. The problems faced by the user in the selection of 'right' dissimilarity measures are discussed by Anderberg [1]. The (dis)similarity measure must be chosen in such a way that it must reflect the associations in the population well. On the basis of the assumptions on the classes (Chapter II) the choice of euclidean distance will be clear.

Now the problems considered in this thesis are as follows

(i) Definition of classes in R^2

(ii) Dividing the given set of points into specific number of groupings and

(iii) Estimation of classes.

The probability distributions on the classes may vary the estimators. Densities may not be known in many cases. If the method developed for estimation is such that it is independent of the density of random variables, then it can be applied in many problems. The crux of the thesis is that whichever may be the class under consideration in R^2 and whatever may be the density (chapter II) the same method of estimation is applicable.

The chapterwise break up of the thesis is as follows

Chapter II It deals with the definitions of classes and the distributions on them.

Basically the classes are assumed to be bounded and disjoint. The assumption of boundedness is justified because in many practical problems, the observations are bounded. There are of course obvious examples of unbounded classes such as a bivariate normal population.

Non-overlapping classes is considered to simplify the calculations. There are many examples of overlapping classes such as two truncated bivariate normal populations where the ranges intersect.

The distributions of random vectors on these classes are also defined in this chapter. The method presented in chapters IV, V and VI is applicable for any distribution defined in this chapter.

Chapter III Minimal spanning tree has been used repeatedly in chapters IV, V and VI in the estimation of classes. The definition of it - assuming the edge weight to be the euclidean distance between the corresponding nodes - is given in this chapter along with a few properties.

Chapter IV The method of estimation of 'compact region' - which is defined as a class in chapter II - is given here under the assumption of uniform distribution. Later it is generalized to any distribution. Division of the observations into specific number of groupings is also stated. It is shown that the estimation procedure is consistent.

Chapter V The methods of estimation of 'bounded line classes' - which is defined in chapter II - is given here. Two methods of estimation are developed under two definitions of consistency. One method is applicable when it is known a-priori that the class is a bounded line class. The second method is applicable when no information is known. The word 'bounded' may seem to be redundant since a class is assumed to be a bounded set. But a class may also be an unbounded set as discussed earlier. 'Unbounded classes' are discussed in chapter VII.

Chapter VI The method of estimation of 'bounded mixture classes' - which is defined in chapter II - is discussed in this chapter. Again, it is shown that the estimation procedure is consistent.

Chapter VII The method developed in chapters IV, V and VI is applicable in estimating the classes consistently irrespective of whichever class it is and whatever may be the underlying distribution defined in chapter II. The computer implementation of it is discussed in this chapter along with an example.

Chapter VIII It deals with the conclusion of the thesis and scope for further work. The methods developed in the previous chapters are useful only for R^2. Generalization to R^n is discussed here. And also, relaxing the assumption of boundedness is discussed here.

1.6 Notations

The notations used in this manuscript are the following

(1) For two sets $A, B \subseteq R^n$

$$d(A,B) = \inf_{x \in A,\ y \in B} d(x,y) \quad \text{where}$$

$d(x,y)$ is the euclidean distance between x and y.

At some places it is also referred as $\|x-y\|$.

(2) For two sets $A,B \subseteq R^n$, the hausdorff metric, denoted by $D(A,B)$, is defined as follows

$$D(A,B) = \text{Max} \left\{ \sup_{x \in B} d(x,A), \quad \sup_{x \in A} d(x,B) \right\}$$ [53]

(3) For two sets $A,B \subseteq R^n$, the minkowski addition of A and B, denoted by $A \oplus B$ is $\{x + y : x \in A, y \in B\}$.

(4) Lebesgue measure of $A \subseteq R^2$ is denoted by $\lambda(A)$.

(5) In chapter II, a condition is defined. It is denoted by C1.

(6) Equation k of section j of chapter i is denoted by i.j.k.
Figure k of section j of chapter i is denoted by 'fig.i.j.k.'

(7) Some statements, which will be quoted in the manuscript are serialized as a, b, c, ... That is, the first statement in the first section of the second chapter will be denoted by 2.1.a, the second statement by 2.1.b and so on.

(8) In appendix, a few examples are stated and a few small results are proved. They are denoted by A1, A2, A3, ...

(9) Theorem k of section j of Chapter i is represented by Thm i.j.k. Similarly lemma/proposition k of section j of Chapter i is represented by lemma i.j.k/Prop i.j.k.

(10) int (A) represents interior of A.

cl (A) represents closure of A.

δA represents boundary of A, i.e. $\delta A = cl(A) \cap cl(A^c)$

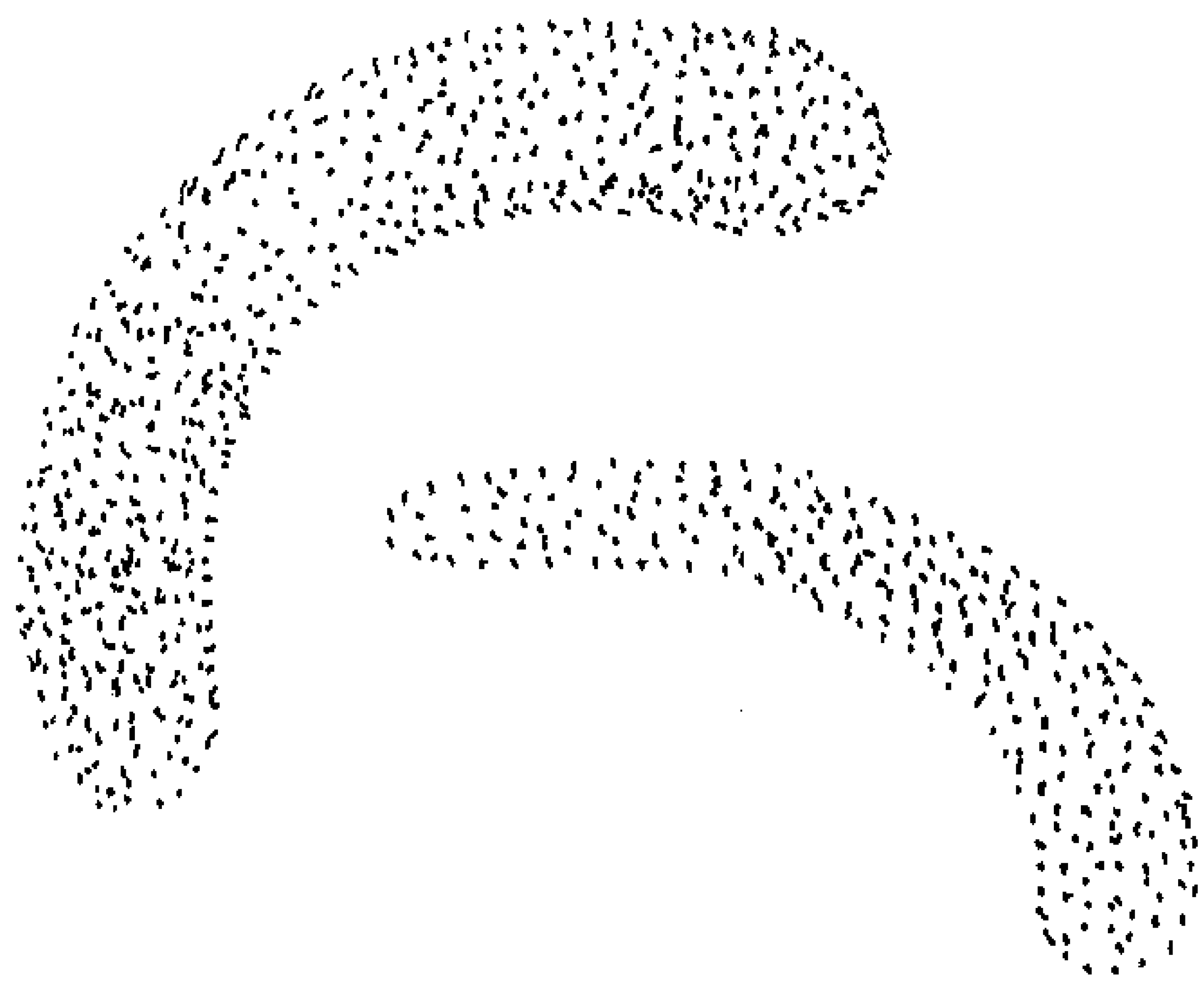


Fig.1.1.1: An example of nonconvex clusters.

II. Definitions of Classes in R^2

2.1 Introduction

Anderberg [1] stated that there is no criterion for judging 'goodness' of clusters in numerical data without an abstract structure in the population. Cormack [32] called for development of methods based on well defined mathematical formulation of the problem. Everitt [30] also emphasized the need for a more critical approach to cluster analysis.

But any mathematical modelling of the problem cannot take care of all possible data sets. Binder [54] studied Bayesian clustering rules under a general set up and provided an algorithm for approximating the similarity matrix and getting an 'optimal' partition. But the definition of a 'class' theoretically is not given in his paper. Optimization methods in cluster analysis were also studied by Marriott [55] and Dubes et al [7]. Strauss [56] studied the intensity of clustering, if it exists, under certain assumptions, the most severe of which is of Markov type. Testing for uniformity has also been studied by Smith et al [57].

The definitions we are going to suggest are not a panacea of all problems. In this thesis the definitions are given only for R^2. The generalization to R^n is discussed in Chapter VIII. The definitions are based on intuitive ideas. The main purpose of this dissertation is to estimate the classes consistently on the basis of the definitions given so that the method can be implemented on a computer.

2.2 Main Assumptions

A few assumptions are stated before the definitions are given.

Assumption (1) If A and B are subsets of R^2 and they can be considered as two classes then $A \cap B = \emptyset$.

The above assumption need not be true in all the cases. In fact, in the case of two normal populations the two classes are one and the same. This assumption assures disjoint classes.

Assumption (2) Every class is a path connected [58] closed set.

Closure is assumed for mathematical convenience. Path connectivity is the main point which is intuitively valid. That is, if two points belong to the same class then there must exist a path which is completely contained in it and also if there exists no path between two points, then those two should belong to two different classes.

If α represents a class and λ represents the lebesgue measure on R^2 then the following four cases arise.

(1) α bounded, $\lambda(\alpha) > 0$.

(2) α bounded, $\lambda(\alpha) = 0$.

(3) α unbounded, $\lambda(\alpha) = 0$.

(4) α unbounded, $\lambda(\alpha) > 0$.

Assumption (3) It is also assumed that α is bounded. There are many examples of unbounded classes. But in many practical problems the assumption of boundedness is justified.

Cases (3) and (4) above will be discussed in section 2.6.

2.3 Compact Regions

α bounded, $\lambda(\alpha) > 0$.

(a) Given that $\lambda(\alpha) > 0$, $Int(\alpha)$ may be a null set. An example is stated in A1 of appendix of a set $\alpha \subseteq R^2$ which is path connected, compact, $\lambda(\alpha) > 0$ but $Int(\alpha) = \emptyset$. These type of sets need not be considered as classes. So we will assume that $Int(\alpha) \neq \emptyset$. But this is not exactly a panacea of the problems because α could still be equal to B$\cup$C where B is the set considered in A1 and $C = [1,2] \times [0,1]$. So we shall assume that $cl(Int(\alpha)) = \alpha$. Apostol [58] considered region to be a union of open connected set with some, none or all the limit points. We are interested in regions which are compact. Also $Int(\alpha)$ is assumed to be connected. An example is stated in A2 in this regard. To facilitate the calculations, it is also assumed that $\delta\alpha$ consists of finitely many rectifiable curves.

So $\mathcal{E} = \{\alpha : \alpha \subseteq R^2,$ α is path connected, compact, $cl(Int(\alpha)) = \alpha$, $Int(\alpha)$ is connected, $\delta\alpha$ consists of finitely many rectifiable curves$\}$ is the parametric space.

We will call any $\alpha \in \mathcal{E}$ as compact region.

(b) By imposing the condition that $cl(Int(\alpha)) = \alpha$, the set $\beta = [0,1] \times [0,1] \cup \{(x,0) : 1 \leq x \leq 2\}$ cannot be considered as class. This case is treated in 2.5.

2.3.c The class of continuous distributions on compact regions have the following properties.

2.3.c.1 Q_α is a probability measure with support α

2.3.c.2 $Q_\alpha << \lambda$. [59,60]

2.3.c.3 The density f is continuous on $Int(\alpha)$ [59,60].

2.3.c.4 $Int\left\{x : f(x) \geq \frac{1}{n}\right\}$ is path connected for sufficiently large n

2.3.c.5 $\inf_{x \in \alpha_\varepsilon^{in}} f(x) > 0$ for all $\varepsilon > 0$ where

$$\alpha_\varepsilon^{in} = \alpha \cap \left[\alpha^c \oplus D(\varepsilon)\right]^c .$$

Let $\mathcal{G}_\alpha = \{Q : Q$ is a probability measure on α satisfying 2.3.c$\}$ where α is a compact region.

$\mathcal{G}_\alpha$ is a collection probability measures on α. This will be used repeatedly in the manuscript.

2.4 Bounded line classes

Let $\alpha \subseteq R^2$ be path connected compact with $\lambda(\alpha) = 0$.

Then (2.4a) $Int(\alpha) = \emptyset$

(2.4b) If $\lambda(\alpha) = 0$ then α may be assumed to be a union of either (i) finitely many, (ii) denumerably many or (iii) uncountably many distinct curves. But cases (ii) and (iii) are not assumed to happen in real life on the basis of the examples A2 and A3 in appendix.

2.4.c α is path connected, compact and α is a union of finitely many curves. But the length of some of those curves may not be finite. An example is given A4 of appendix of a curve which is not rectifiable. As it can be seen from the construction itself that such a set need not be considered as class. But even the condition of rectifiability would not avoid a curve shown in A5 of Appendix .

2.4.d In order to avoid the curve shown in A5, a few more conditions are imposed on these sets.

Definition 2.4.1 Let $\alpha \subseteq R^2$ be a Jordan arc or Jordan curve. α is said to satisfy C1 if for every $x \in \alpha$, if there exists $\varepsilon_{x_i} > 0$ such that for every $0 < \varepsilon < \varepsilon_{x_i}$, $\mathcal{V}_\varepsilon \cap \alpha$ is the component of x in $\mathcal{V}_\varepsilon \cap \alpha$ where $\mathcal{V}_\varepsilon$ is the open disc of radius ε at x and $\varepsilon_{x_i} < \sup\limits_{x_1, y_1 \in \alpha} \| x_1 - y_1 \|$

If α satisfies C1 then there are many such ε_{x_i}'s for every x. Define connectivity order of x to be $\frac{1}{2}$ [Supremum of such ε_{x_i}'s].

Definition 2.4.2 Let $f : [a,b] \longrightarrow R^2$ describe a Jordan arc or Jordan Curve, $a < b$.

Let $g_x(t) = \|f(t) - f(x)\|$, $x, t \in [a,b]$.

(i) f is said to have a concavity at $x \in [a,b)$ or at $f(x)$, in the positive direction if there exist y_1, y_2, y_3 such that $x < y_1 < y_2 < y_3 \leq b$ and either $g_x(t)$ increases for $t \in [y_1, y_2]$ and decreases for $t \in [y_2, y_3]$ or $g_x(t)$ decreases for $t \in [y_1, y_2]$ and increases for $[y_2, y_3]$.

(ii) f is said to have a concavity at $x \in (a,b]$ or at f(x) in the negative direction if there exist y_1, y_2, y_3, $a \leq y_1 < y_2 < y_3 \leq x$ such that either $g_x(t)$ increases in $[y_1, y_2]$ and decreases in $[y_2, y_3]$ or $g_x(t)$ decreases in $[y_1, y_2]$ and increases in $[y_2, y_3]$.

(iii) f is said to have k concavities at x, $x \in [a,b)$ or at f(x) in the positive direction if there exists y_{1i}, y_{2i}, y_{3i}, $i=1,\dots k$ such that $x < y_{11}$, $y_{3i} \leq y_{1(i+1)}$ for $i=1,\dots(k-1)$, $y_{3k} \leq b$ such that f has a concavity at x in the positive direction for each i.

(iv) Similar definition for k concavities in the negative direction can be given.

(v) f is said to have (at most) k concavities at $x \in [a,b]$ or at f(x) if the number of concavities in the positive as well as negative directions add up to (at most) k.

(vi) A Jordan arc or Jordan curve α is said to have at most k concavities if for every f describing α has at most k concavities for every f(x).

Definition 2.4.3 α is said to be a bounded line class if it is a path connected compact subset of R^2 such that $\alpha = \bigcup_{i=1}^{k} \alpha_i$ where each α_i is a rectifiable Jordan arc or Jordan curve having at most finitely many concavities.

Observe that a curve satisfies Def.2.4.2 $\Rightarrow$ it satisfies Def.2.4.1.

A proposition is proved before probability measures on bounded line classes are defined.

Preposition 2.4.1 Let $\beta \subseteq R^2$ be a rectifiable Jordan arc or Jordan curve satisfying $\mathcal{C}1$ and let A be an open disc. Then $\beta \cap A$ can have atmost countably many components.

Proof Let $\beta \cap A \neq \emptyset$. Let $x_0 \in \beta \cap A$.

$x_0 \in A \Rightarrow$ there exists open disc $\mathcal{V}_r$ of radius r around x_0 where $r > 0$ such that $\mathcal{V}_r \subseteq A$.

$x_0 \in \beta \Rightarrow$ there exists $y \in \beta$, $y \neq x$ such that $\|x_0 - y\| \leq \operatorname{Min}\left(\frac{r}{2}, \varepsilon_{x_0}\right) = \eta$ and $\|x_0 - z\| \leq \|x_0 - y\|$ for all z where z is a point on that portion of the curve joining x_0 and y and ε_{x_0} is the connectivity order of x_0. The length of the curve between x_0 and y is greater than or equal to $\|x_0 - y\| > 0$. So the component of $x \in \beta \cap A$ [call it c_{x_0}] is a portion of the curve β such that the arc length of the curve is greater than zero. Since c_{x_0} is not a closed set, define the arc length of c_{x_0} [$\mu(c_{x_0})$] to be the length of cl (c_{x_0}). [58].

Let $A_n = \left\{ c_x : \text{Arc length of } c_x > \frac{1}{n} \right\}$

A_n is a finite set since arc length of each $c_x > 0$ and β is rectifiable. $\bigcup_{n=1}^{\infty} A_n = \beta \cap A$ and $\bigcup_{n=1}^{\infty} A_n$ is a countable set. So $\beta \cap A$ has at most countably many components.

Definition 2.4.4 Let $\beta \subseteq R^2$ be rectifiable Jordan arc or Jordan curve satisfying $\mathcal{C}1$. Define for every open disc $\mathcal{V}$,

$\mu(\mathcal{V})$ = sum of the arc lengths of components of $\beta \cap \mathcal{V}$

and $\mu\left(\bigcup_{i=1}^{\infty} \mathcal{V}_i\right) = \sum_{i=1}^{\infty} \mu(\mathcal{V}_i)$ if $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$ for $i \neq j$.

It can be shown that the above μ can be extended uniquely to σ - field of open sets which is borel σ - field and it is a measure on it [59].

Notation : $\mu(\mathcal{V})$ is also denoted by $\mu(\mathcal{V} \cap \beta)$ for the curve β.

Definition 2.4.5 Let $\alpha = \bigcup_{i=1}^{n} \alpha_i$ be a bounded line class where each α_i is a rectifiable Jordan arc or Jordan curve satisfying C1. Define for every open disc $\mathcal{V}$,

$$\tau(\mathcal{V}) = \sum_{i=1}^{n} \mu(\mathcal{V} \cap \alpha_i) \quad \text{and}$$

$$\tau\left(\bigcup_{1}^{\infty} \mathcal{V}_i\right) = \sum_{i=1}^{\infty} \tau(\mathcal{V}_i) \quad \text{if } \mathcal{V}_i \cap \mathcal{V}_j = \emptyset \text{ for } i \neq j .$$

and $\mathcal{V}_i$'s are open discs.

It can be shown that the above τ can be uniquely extended to borel σ-field and it is a measure on α. The above τ shall be denoted by μ. The probability measures on bounded line classes are defined with the help of the measure μ.

Notation Let Q be a probability measure on α and Q satisfies the following properties.

f is a measurable function from $R^2 \longrightarrow R^+$ such that

2.4.e.1 $f(x) \geq 0 \quad \forall x \in \alpha,$

2.4.e.2 $f(x) = 0$ for all $x \in \alpha^c$

2.4.e.3 f is continuous on α.

2.4.e.5 $Q(A) = \int_A f d\mu$ for every borel set A and

2.4.e.5 $Q(\alpha) = \int_\alpha f d\mu = 1.$ Then denote by

$\mathcal{H}_\alpha = \{ Q :$ Q is a probability measure on α satisfying the above properties 2.4.e.1 to 2.4.e.5 $\}$ where α is a bounded line class.

$\mathcal{H}_\alpha$ will be repeatedly used in the following chapters.

2.5 Bounded Mixture Classes

Compact regions and bounded line classes do not take into consideration set α with area > 0 but $cl(Int(\alpha)) \neq \alpha$ as stated in the example given below.

Example 2.5.1 Let $\beta = [0,1] \times [0,1] \cup \{ (x,0) : 1 \leq x \leq 2 \}$.

$\lambda(\beta) = 1$ but $cl(Int(\beta)) = [0,1] \times [0,1] \subseteq \beta$. Observe that β can be considered as a union of a compact region and a bounded line class. Mathematical formulation of this idea is given below.

Definition 2.5.1 $\alpha \subseteq R^2$ is called a bounded mixture class if $\alpha = (\bigcup_{i=1}^{k_1} A_i) \cup (\bigcup_{j=1}^{k_2} B_j)$ where each A_i is a compact region, each B_j is a bounded line class, α is path connected $A_i \cap A_j = \emptyset \quad \forall i \neq j$ and $B_i \cap B_j = \emptyset \quad \forall i \neq j$.

Probability measures on bounded mixture classes are also mixtures of probability measures on compact regions and bounded line classes. They are defined below.

Definition 2.5.2 Let $\alpha = (\bigcup_{i=1}^{k_1} A_i) \cup (\bigcup_{j=1}^{k_2} B_i)$ where each A_i is a compact region, each B_i is a bounded line class and α is a bounded mixture class. Let $q_1, q_2, \ldots q_{k_1}, r_1, r_2, \ldots r_{k_2}$ be such that $q_i > 0 \;\; \forall i = 1, \ldots k_1$, $r_i > 0 \;\; \forall i = 1, \ldots k_2$,

$\sum_1^{k_1} q_i + \sum_1^{k_2} r_i = 1$ and $Q = \sum_{i=1}^{k_1} q_i Q_i + \sum_{i=1}^{k_2} r_i S_i$ where

$Q_i \in \mathcal{G}_{A_i} \;\; \forall i$ and $S_i \in \mathcal{H}_{B_i} \;\; \forall i$.

As it can be seen Q is a probability measure on α.

2.6 Unbounded Classes

As it is mentioned earlier we shall be dealing with bounded sets for classification and estimation purposes in this thesis. Nevertheless unbounded classes are discussed in this section.

As it is the case with bounded classes, there are three types of unbounded classes. They may be defined much on the same lines as bounded classes.

Definition 2.6.1 $\beta \subseteq R^2$ is said to be one side unbounded curve if there exists a function $f: [0,\infty) \to \beta$, f is continuous, onto and $x_n \to \infty \Rightarrow f(x_n)$'s are unbounded. β is said to be σ-finite if for every bounded interval A of $[0,\infty)$, f(A) has finite length. β is said to satisfy Cl if every bounded path connected subset of β satisfies Cl.

Though Cl is defined for Jordan arcs and curves, it may be generalized for every bounded path connected set.

Definition 2.6.2 $\beta \subseteq R^2$ is said to be two side unbounded curve if there exists a function $f: (-\infty,\infty) \to \beta$, f is onto continuous and

$$x_n \to \infty \Rightarrow f(x_n)\text{'s are unbounded} \quad \text{and}$$
$$y_n \to -\infty \Rightarrow f(y_n)\text{'s are unbounded.}$$

β is said to be σ-finite if for every bounded interval A of A of $(-\infty,\infty)$, f(A) has finite length. β is said to satisfy Cl if every bounded path connected subset of β satisfies Cl.

Definition 2.6.3 $\alpha \subseteq R^2$ is said to be an unbounded closed region if

(i) α is path connected, closed, $cl(Int(\alpha)) = \alpha$

(ii) $Int(\alpha)$ is path connected and

(iii) $\delta\alpha$ consists of finitely many σ-finite unbounded curves (one side or both sided) or rectifiable curves.

Definition 2.6.4 $\alpha \subseteq R^2$ is said to be an unbounded line class if

$$\alpha = \bigcup_{i=1}^{k} \alpha_i \quad \text{where}$$

(i) at least one α_i is an unbounded σ-finite curve,

(ii) every rectifiable Jordan arc or Jordan curve has atmost finitely many concavities,

(iii) α is path connected and

(iv) for every unbounded curve α_1 and for every Jordan arc or Jordan curve $\beta \subseteq \alpha_i$ has finitely many concavities.

Definition 2.6.5 $\alpha \subseteq R^2$ is said to be an unbounded mixture class if

$$\alpha = \left(\bigcup_{i=1}^{k_1} A_i \right) \cup \left(\bigcup_{j=1}^{k_2} B_j \right) \text{ such that}$$

(i) at least one of A_i's is an unbounded closed region and the rest are compact regions or at least one of B_j's is an unbounded line class and the rest are bounded line classes.

(ii) $A_i \cap A_j = \emptyset \; \forall i \neq j$ and $B_i \cap B_j = \emptyset \; \forall i \neq j$ and

(iii) α is path connected.

Probability measures may also be defined on these classes similar to that of bounded classes presented in sections 2.3, 2.4 and 2.5.

2.7 Conclusions and remarks

Definitions of classes are given in earlier sections of this chapter. But the generalization to any R^n creates problems at this stage because the effect of these assumption on the estimation procedure has not yet been studied. In fact it will be seen in Chapter VIII that some assumptions on bounded line classes were not used in the estimation procedure.

Probable definition for compact region in R^n may be given in the following way.

Definition 2.7.1 $\alpha \subseteq R^n$ is called as compact region in R^n if α is compact, path connected, $cl(Int(\alpha)) = \alpha$, $Int(\alpha)$ is path connected and $\lambda(\delta\alpha) = 0$ where λ is the lebesgue measure on R^n.

But a definition of bounded line class in R^n is to be given recursively. For the space R^3, path connected sets with volume zero are curves as well as planes. They may be assumed to be compact regions in R^1 and R^2. Similarly for any R^n, path connected compact sets with hyper volume zero may be assumed to be a set in R^k, $k \leq n-1$. The questions related to definitions and estimation of classes in R^n are discussed in Chapter VIII.

The material for this chapter is basically taken from a paper by C.A. Murthy and D. Dutta Majumder [64].

III. Minimal Spanning Tree [MST]

3.1 Introduction

Minimal spanning tree has been in use [1,2,5,12] to cluster the given set of data. It is also used for the generalization of wald-wolfonitz test [62] as well as for a test for clustering [57].

The computational problems involved in the construction of MST have been extensively studied. The classical algorithms are due to Kruskal [63] and Prim [64]. Prim's principles can be embodied into a computer algorithm [65] that requires computation time $O(n^2)$ where n is the number of data points. Shamos and Huey [66] have developed an algorithm with computation time never greater than $O(n \log n)$ for data sets in the plane. For higher dimensions, Bentley and Friedman [67] and Rohlf [68] have presented algorithms for which the computation time has been measured to be on the average $O(n \log n)$.

In this manuscript, MST is used in the estimation of sets as well as for classification. It will be proved in Chapters IV, V and VI that certain sets based on MST are consistent estimators of classes. We will be only bothered about its implementation on computer. Much stress has been laid on the implementation on computer, but not on faster implementation. In Chapter VII, this aspect is discussed.

3.2 Definition of MST

We begin by reviewing some terms of graph theory. A graph consists of a set of nodes and a set of node pairs called edges. We say that an edge links the two nodes defining it and it is incident on both of them. The degree of a node is number of edges incident on it. A path between two prescribed nodes is an alternating sequence of nodes and edges with the prescribed nodes as first and last elements, all other nodes distinct, and each edge linking the two nodes adjacent to it in the sequence. A connected graph has a path between any two distinct nodes. A cycle is a path beginning and ending with the same node. A tree is a connected graph with no cycles. A subgraph of a given graph is a graph with all of its nodes and edges in the given graph. A spanning subgraph of a given graph is a subgraph with node set identical to the node set of the given graph. A spanning tree of a graph is a spanning graph that is a tree. Note that there is a unique path between every two nodes in a tree and thus a spanning tree of a connected graph provides a path between every two nodes of the graph.

An edge weighted graph is a graph with a real number assigned to each edge. A minimal spanning tree (MST) of an edge weighted graph is a spanning tree for which the sum of edge weights is minimum (The terminology of graph theory has not been standardized, for a general discussion, see Harary [69]).

Note that a graph (collection of nodes and edges) is assumed to be given for MST. But here, MST will be constructed when a set of nodes is given with the edge weight being the euclidean distance between the corresponding nodes. The definition is given below.

Definition 3.2.1 For two points $x, y \in R^2$,

$$\text{let } [x, y] = \{\tau x + (1-\tau) y : 0 \leq \tau \leq 1\}$$

Definition 3.2.2 Let $S = \{x_1, x_2, \ldots x_n\} \subseteq R^2$.

(i) $G \subseteq R^2$ is a graph of S if $G = \bigcup_{i=1}^{k} [x_{i1}, x_{i2}]$,

$\bigcup_{i=1}^{k} \{x_{i1}, x_{i2}\} = S_1 \subseteq S$. Each $[x_{i1}, x_{i2}]$ is an edge of G.

(ii) A graph G of S is said to be connected if for every $y_1, y_2 \in S_1$, $y_1 \neq y_2$ there exists $z_1, z_2, \ldots z_m \in S_1$ such that

$\bigcup_{i=0}^{m} [z_i, z_{i+1}] \subseteq G$ where $z_0 = y_1$ and $z_{m+1} = y_2$.

(iii) A graph G of S is said to be a spanning tree of S if

$G = \bigcup_{i=1}^{n-1} [x_{i1}, x_{i2}]$, $\bigcup_{i=1}^{n-1} \{x_{i1}, x_{i2}\} = S$ and G is connected.

(iv) Length of a spanning tree G of S is ℓ where

$$\ell = \sum_{i=0}^{n-1} \| x_{i1} - x_{i2} \| \quad \text{where} \quad G = \bigcup_{i=1}^{n-1} [x_{i1}, x_{i2}].$$

(v) A spanning tree G_0 of S is said to be minimal if length of $G_0 \leq$ length of G for every spanning tree G of S.

Note that for a given set $S = \{x_1, \ldots x_n\}$, there may exist more than one MST. That is, MST is not unique. But the results to be proved in Chapters IV, V and VI hold good for whatever MST is chosen. Therefore, it is not going to matter which set is considered as MST among them. MST can be made unique if the following conventions are adapted in any process of generation of it.

(i) If for a particular point x_i, there are two nearest neighbours x_{j_1} and x_{j_2} then choose x_k as the nearest neighbour of x_i where $k = \min(j_1, j_2)$.

(ii) If for a particular subset S_1 of $S = \{x_1, x_2, \ldots x_n\}$ there are two nearest neighbours x_{i_1} and x_{i_2}, thus correspondingly there are two points x_{j_1} and x_{j_2} in S_1 such that $d(x_{j_1}, x_{i_1}) = d(x_{j_2}, x_{i_2})$ then choose x_j from S_1 where $j = \min(j_1, j_2)$.

(iii) If the number of nearest neighbours either for a point or for a subset of S is greater than or equal to 3, follow similar procedure (i.e. always choose the one minimum suffix) stated in the above conventions.

(iv) Follow similar conventions if there are two nearest subsets S_{j_1} and S_{j_2} of S for a subset S_i of S

(v) If $d(x_{i_1}, x_{i_2}) = d(x_{j_1}, x_{j_2})$ or $d(S_{i_1}, S_{i_2}) = d(S_{j_1}, S_{j_2})$ and one pair must be joined (either i_1, i_2 or j_1, j_2) then choose $i = \min(i_1, i_2, j_1, j_2)$ and join it with the corresponding point/set whatever may be the case.

3.3 Properties of MST

Minimal spanning trees have two important properties that make them appropriate for applications (1) they connect all of the nodes with n-1 edges and (2) the node pairs defining the edges represent points that tend to be close together (small distance or dissimilarity). The first property follows from the fact that MST is a spanning tree and the second from the requirement that the sum of the edge weights be a minimum. This is amplified by the following two theorems from Prim [64].

Theorem 3.3.1 An MST contains as a subgraph the 'nearest neighbour graph'. That is, there is an edge linking each node and the node closest to it (or one of them if there are ties).

Theorem 3.3.2 If any edge of an MST is deleted, thereby dividing the graph into two disjoint connected subgraphs, and thus the points into two disjoint subsets, the deleted edge weight is the smallest interpoint distance between the two subsets.

The above two properties will be exploited in chapters IV, V and VI. Apart from these two, a few more results which will be used in later chapters are proved below.

Proposition 3.3.1 : Let $A \subseteq R^2$. Let $A^n = A \times A \times \ldots \times A$ (n times).

Let for any $a = (x_1, \ldots x_n) \in A^n$, $\bar{a} = \{x_1, \ldots x_n\}$.

Let $B_n(A) = \{\bar{a} : a \in A^n\}$.

Let $\beta_n(\bar{a})$ = Length of MST of $\bar{a}$

Let $\tau_n(A) = \sup \{ \beta_n(\overline{a}) : \overline{a} \in B_n(A) \}$.

Then $\tau_n(A) \leq \tau_{n+1}(A)$ for $n = 2,3,\ldots$ and for any $A \subseteq R^2$.

<u>Proof</u> $B_n(A) \subseteq B_{(n+1)}(A)$

So $\sup \{ \beta_n(\overline{a}) : a \in A^n \} \leq \sup \{ \beta_{n+1}(\overline{a}) : a \in A^{n+1} \}$.

i.e., $\tau_n(A) \leq \tau_{n+1}(A)$.

Similarly the following proposition can be proved.

Proposition 3.3.2 Let $A \subseteq B$.

Let for any $a = (x_1, \ldots x_n) \in A^n$, $\overline{a} = \{x_1, \ldots x_n\}$ and

$b = (y_1, \ldots y_n) \in B^n$, $\overline{b} = \{y_1, \ldots y_n\}$

Let $B_n(A) = \{ \overline{a} : a \in A^n \}$ and

$B_n(B) = \{ \overline{b} : b \in B^n \}$.

$\tau_n(A) = \sup \{ \beta_n(\overline{a}) : \overline{a} \in B_n(A) \}$ and

$\tau_n(B) = \sup \{ \beta_n(\overline{b}) : \overline{b} \in B_n(B) \}$ where

$\beta_n(\overline{a})$ = length of MST of $\{x_1, \ldots x_n\}$ where $\overline{a} = (x_1, \ldots x_n)$.

Then $\tau_n(A) \leq \tau_n(B)$.

<u>Proof</u> Trivial since $A \subseteq B$.

A few more definitions regarding MST which will be used in the following chapters are given below.

Let $S = \{x_1, \ldots x_n\}$ and $G = \bigcup_{i=1}^{n-1} [x_{i1}, x_{i2}]$ be a MST of S.

Definition 3.3.1 Maximum edge length of G is denoted by m and is defined as $\sup \| x_{i1} - x_{i2} \|$ $i = 1,2,\ldots(n-1)$

Definition 3.3.2 For any set $\alpha \supsetneq S$, G is a spanning tree of S, define $\mathcal{U}_{\alpha} = \sup_{x \in G} \inf_{y \in \alpha} \| x-y \|$.

3.4 Motivation for using MST in estimation of classes

We shall briefly state by citing several examples why MST may be helpful in estimating the classes.

Example 3.4.1 : $X_1, X_2, \ldots X_n$ are independent and identically distributed random variables following uniform distribution in the range 0 to θ, $\theta > 0$ unknown. Then the maximal likelihood estimate of θ is the nth order statistic $X_{(n)}$. $\left[X_{(n)} = \text{Max} \{ X_1, \ldots X_n \} \right]$. Observe that by estimating θ, the whole set 0 to θ is estimated. The estimate of the set is $[0, X_{(n)}]$ which may be looked upon as an MST of $\{0, X_1, \ldots X_n\}$.

Example 3.4.2 $X_1, X_2, \ldots X_n$ are independent and identically distributed random variables following uniform distribution in the range θ_1 to θ_2, $\theta_1 < \theta_2$, θ_1 and θ_2 are unknown. The estimator for θ_1 is $X_{(1)} = \text{Min}\{X_1, \ldots X_n\}$ and for θ_2 is $X_{(n)}$. So the estimate for the whole set is $[X_{(1)}, X_{(n)}]$ which may be looked upon as an MST of $\{X_1, \ldots X_n\}$.

Example 3.4.3 $X_1, X_2, \ldots X_n$ are independent and identically distributed random vectors following uniform distribution over a compact convex set $A \subseteq R^2$. Then the maximum likelihood estimator for A is the minimal convex set generated from $X_1, \ldots X_n$ which in other words is the convex hull of $X_1, \ldots X_n$. [Minimum in the sense of area].

Example 3.4.4 $X_1, X_2, \ldots X_n$ are independent and identically distributed random vectors following uniform distribution over a compact path connected subset A of R^2. Then the maximum likelihood estimator for A does not exist. Because there exists no minimal path connected set (minimum in the sense of area) generated from $X_1, \ldots X_n$. But there is a minimal path connected set, minimum in the sense length, generated from $X_1, \ldots X_n$ which is MST. But MST has area zero. Intuitively, the set must be concentrated around MST, i.e., discs of radius ε_n may be drawn around every point on MST in such a way that $\varepsilon_n \to 0$ as $n \to \infty$ so that the estimated set may actually converge to the original set. This idea is exploited in Chapters IV, V and VI.

But the distribution considered in the above examples is uniform. Generalization to any continuous distribution defined in Chapter II is also covered in the later chapters. The convergence of a sequence of sets in terms of probability has been defined in Chapter IV.

The property stated in theorem 3.3.2 is utilized in the classification because the classes are closed and disjoint. If there are two classes and for every n (n is the number of observations) the edge with maximum length in the MST is deleted then as $n \to \infty$, the original classification will be achieved. After deleting the edge, the sets may be estimated by drawing discs around the two MST's. The choice of ε_n's is discussed in the later chapters. But this procedure will fail if the classes are overlapping.

3.5 Further Remarks

Minimal spanning tree tends to give clusters which are line shaped. It may successfully separate clusters shown in fig. 3.5.1 but may not be able to separate clusters in fig. 3.5.2. For the separation of clusters shown in fig. 3.5.3 MST is helpful. Basically there is a 'chaining effect' inherent in MST. The draw-backs of MST in clustering (MST is same as single linkage method) are listed by Dubes and Jain [7]. But in the estimation of path connected sets, it will be proved, MST's are helpful. The classes, if they are distinct, will be apparent if the number of observations is sufficiently large. Then the classification procedure using MST as well as estimation procedure using MST may both be applied for the finitely many observations. The results to be proved are all asymptotic results. One has to be watchful in applying MST for classification.

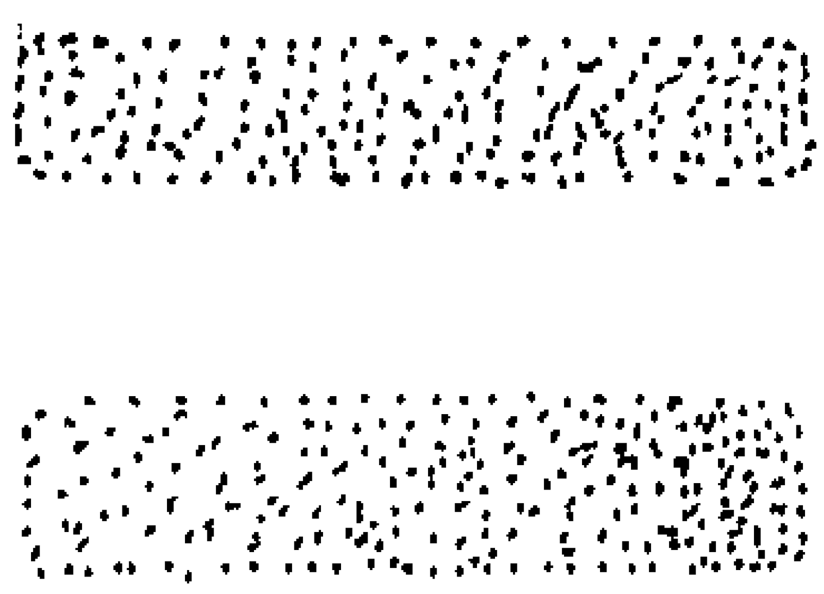
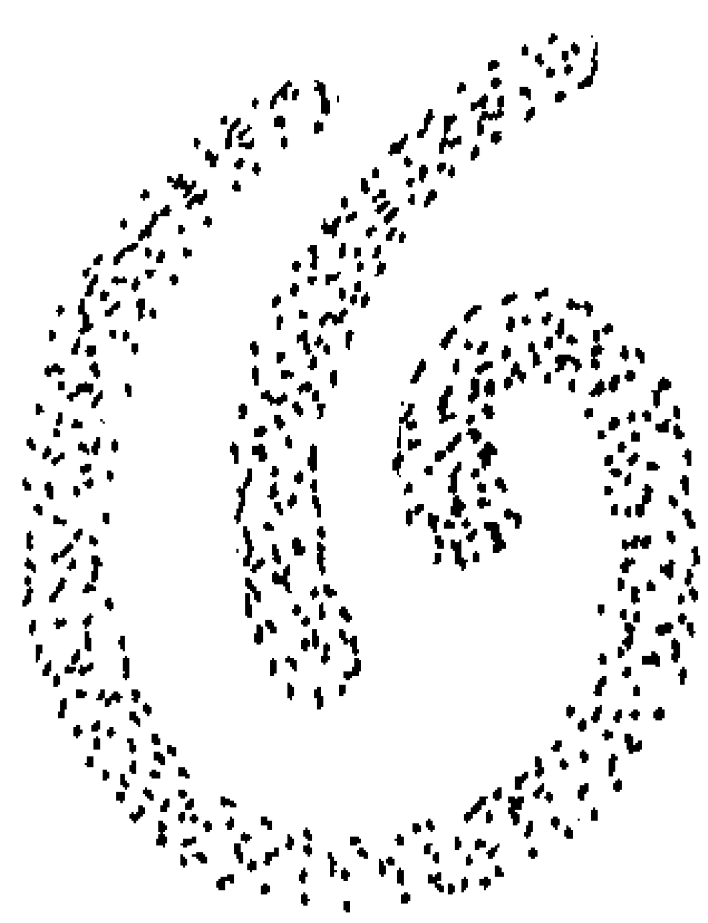


Fig. 3.5.1

Fig 3.5.2



Fig, 3.5. 3

MST is likely to separate clusters in figures 3.5.1 and 3.5.3. But it is unlikely to do well in figure 3.5.2.

IV. Consistent Estimation of Compact Regions

4.1 Consistency

The definition for consistent estimator of a set, that will be followed here is given below.

Definition 4.1.1 [45]: Let $X_1, X_2, \ldots X_n \ldots$ be independent and identically distributed random vectors from a probability measure P_α whose support is α. Let α_n^* be an estimator based on $X_1, X_2, \ldots X_n$.
α_n^* is said to be consistent estimator of α if

$$E_\alpha\left[\mu(\alpha_n^* \Delta \alpha)\right] \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty,$$

where μ is some given σ-finite measure, Δ represents symmetric difference and 'E' represents expectation.

Grenander [45] also proved a result in this connection which is stated below.

Theorem 4.1.1 [45]: Let $\alpha \subseteq R^2$ be a bounded borel set whose boundary has lebesgue measure zero. Let $\varepsilon_n > 0$ and $n\,\varepsilon_n^2 \longrightarrow \infty$.
Let $\alpha_n^* = \bigcup_{i=1}^{n} \left\{ x : \|X_i - x\| < \varepsilon_n \right\}$ where $X_1, X_2, \ldots$ are independent and identically distributed random vectors from uniform distribution over α.
Then $E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right] \longrightarrow 0$ as $n \longrightarrow \infty$ where λ is the lebesgue area.

It has also been mentioned that extension of the above theorem to higher dimensional euclidean spaces is easy. The corresponding theorem in higher dimensional euclidean spaces is given below.

Theorem 4.1.2 Let $\alpha \subseteq R^k$ be a bounded borel set whose boundary has lebesgue measure zero [lebesgue measure in R^k]. Let $\varepsilon_n > 0$ and $n\,\varepsilon_n^{\,k} \dashrightarrow \infty$. Let $\alpha_n^* = \bigcup_{i=1}^{n} \{x : \|X_i - x\| < \varepsilon_n\}$ where $x_1, X_2, \ldots$ are *independent* and identically distributed random vectors following uniform distribution over α. Then $E_\alpha\left[\mu(\alpha_n^* \Delta\, \alpha)\right] \dashrightarrow 0$ as $n \dashrightarrow \infty$ where μ is the lebesgue measure in R^k.

But the above theorems cannot be applied on a given data set. Without loss of generality, let us assume that the observations are from R^2. There are infinitely many sequences of ε_n's with the properties that $\varepsilon_n > 0$ and $n\,\varepsilon_n^{\,2} \dashrightarrow \infty$. In addition to, in any sequence the first finitely many terms don't make any impact on the limit and we almost always have finitely many observations. *If* ε_n's *are made to be a function of random vectors* $X_1, X_2, \ldots X_n$ then there will be an automatic way of deriving the value of ε_n. But, the proof given by Grenander [45] fails if ε_n is a function of $X_1, X_2, \ldots X_n$.

In the next section, ε_n's are made functions of $X_1, X_2, \ldots X_n$ in the context of compact regions and a consistent estimator for α is derived under the assumption of uniform distribution. Later it is generalized to any continuous distribution defined in Chapter II.

4.2 Consistent estimation of compact regions for uniform distribution

'*Compact region*' is *defined* in 2.3. Any $\alpha \in \mathcal{E}$ is called as compact region where

$\mathcal{E} = \{\beta : \beta \subseteq R^2$, β is path connected, compact, $cl(\text{Int}(\beta)) = \beta$, $\text{Int}(\beta)$ is connected and $\delta\beta$ consists of finitely many rectifiable curves$\}$.

A consistent estimate for any $\alpha \in \mathcal{E}$ under the assumption of uniform distribution is obtained with the help of the following theorems.

Theorem 4.2.1: Let $\alpha \in \mathcal{E}$, α unknown. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors from $(\Omega, \mathcal{A}, P_\alpha)$ to R^2, taking values in α and following uniform distribution on α.

Let

$$G_n(w) = \text{MST of } \{X_1(w), \ldots X_n(w)\}, \quad w \in \Omega.$$

$$\left.\begin{aligned} \ell_n(w) &= \text{Length of } G_n(w), \\ m_n(w) &= \text{Maximum edge length of } G_n(w), \end{aligned}\right\} \text{ defined in Chapter II.}$$

$$h_n(w) = \sqrt{[\ell_n(w)/n]}\,,$$

and

$$\alpha^*_{n1}(w) = \bigcup_{i=1}^{n} \{x \in R^2 : \|X_i(w) - x\| \leq h_n(w)\}.$$

Then $E_\alpha\left[\lambda(\alpha^*_{n1} \Delta \alpha)\right] \longrightarrow 0$ as $n \longrightarrow \infty$.

<u>Proof</u> We shall give the proof using a few lemmas stated below.

Lemma 4.2.1 $\ell_n(w) \longrightarrow \infty$ in probability.

<u>Proof</u> $\ell_n(w)$ is shown to be a measurable function in appendix [A6]. It is to be shown that $P_\alpha\{w : \ell_n(w) > M\} \longrightarrow 1$ as $n \longrightarrow \infty$ for any $M > 0$. Now $\text{Int}(\alpha) \neq \varphi$. So there exists an open disc $\mathcal{V}$ of diameter 'a' > 0 such that $\mathcal{V} \subseteq \alpha$. (fig.4.2.1). Consider open discs $F_1, F_2, \ldots F_7$ of diameter 'a/5' as shown in Fig.4.2.1.

Then $Pr(X_1 \notin F_1, X_2 \notin F_1, \ldots) = \lim_{n \to \infty} (1-(\lambda(F_1)/\lambda(\alpha)))^n = 0$

So $Pr\{w :$ There exist a positive integer M_1 such that $X_{M_1}(w) \in F_1\} = 1$.

Similarly $Pr\{w:$ there exists positive integer M_i such that

$$X_{M_i}(w) \in F_i\} = 1 \text{ for all } i = 1,2,\ldots 7.$$

So $Pr\{w :$ There exists positive integers $M_1, M_2, \ldots M_7$ such that

$$X_{M_i}(w) \in F_i \quad \text{for all } i = 1,2,\ldots 7\} = 1.$$

i.e., $Pr\{w : \ell(\text{MST Generated by } X_1(w), X_2(w), \ldots X_n(w)) \geq \frac{6a}{5}\} \longrightarrow 1$ as $n \longrightarrow \infty$.

Similarly it can be proved by considering 7 discs of diameter a/25 in each one of $F_1, F_2, \ldots F_7$ that $Pr\{w: \ell_n(w) \geq \frac{6a}{5} + 7.\frac{6a}{5^2}\} \longrightarrow 1$ as $n \longrightarrow \infty$

Similarly $Pr\{w: \ell_n(w) \geq \frac{6a}{5} (1 + \frac{7}{5} + (\frac{7}{5})^2\} \longrightarrow 1$ as $n \longrightarrow \infty$

i.e., $Pr\{w : \ell_n(w) \geq M\} \longrightarrow 1$ as $n \longrightarrow \infty$ for any $M > 0$ since

$'a' > 0$ and $1 + \frac{7}{5} + (\frac{7}{5})^2 + \ldots \longrightarrow \infty$.

Lemma 4.2.2 Let $G_n(w)$ = MST generated by $X_1(w), \ldots X_n(w)$.

Let $m_n(w)$ = the maximum edge length of G_n [Def.3.3.1]

Then $Pr(m_n > \varepsilon) \longrightarrow 0$ as $n \longrightarrow \infty$ for every $\varepsilon > 0$.

[It is shown in appendix A6 that m_n is a measurable function].

Proof α is bounded. So there exists a square B of side $b > 0$ such that $\alpha \subseteq B$. Divide the square B into k^2 squares, each square of side $\frac{b}{k}$.

$\sup\limits_{x\in A_1,\ y\in A_2} d(x,y) = 2\sqrt{2}\,\frac{b}{k}$, where A_1 and A_2 are adjacent squares (Fig.4.2.2)

$\frac{2\sqrt{2}b}{k} \longrightarrow 0$ as $k \longrightarrow \infty$.

We are to show that

$Pr(m_n > \varepsilon) \longrightarrow 0$ as $n \longrightarrow \infty$ for every $\varepsilon > 0$.

Let $\varepsilon > 0$. Let k be a positive integer such that $\frac{2\sqrt{2}b}{k} < \varepsilon$. Cover B by k^2 squares $(A_1, A_2, \ldots A_{k^2})$, each square of side $\frac{b}{k}$.

Let M be such that M is an integer, $1 \leq M \leq k^2$, $\bigcup\limits_{i=1}^{M} A_i \supseteq \alpha$

and $Int(A_i) \cap \alpha \neq \varphi$ for all $i = 1,2,\ldots M$.

As $n \longrightarrow \infty$, there exists at least one j_i such that $X_{j_i} \in A_i$ for $i=1,\ldots M$ with probability one. (The argument is similar to that of lemma 4.2.1).

So $Pr\ (m_n > \frac{2\sqrt{2}b}{k}) \longrightarrow 0$ as $n \longrightarrow \infty$

$(m_n > \varepsilon) \subseteq (m_n > \frac{2\sqrt{2}b}{k})$ for all n.

So $Pr(m_n > \varepsilon) \longrightarrow 0$ as $n \longrightarrow \infty$.

Lemma 4.2.3 Let $h_n(w) = (\frac{\ell_n(w)}{n})^{1/2}$

then $h_n(w) \longrightarrow 0$ in probability and $nh_n^2(w) \longrightarrow \infty$ in probability. [It is shown in appendix A6 that h_n is a measurable function].

Proof $nh_n^2(w) = \ell_n(w) \longrightarrow \infty$ in probability from lemma 4.2.1.

$\ell_n(w) \leq n\, m_n(w)$ for all n.

So $\frac{\ell_n(w)}{n} \leq m_n(w)$.

$m_n(w) \longrightarrow 0$ in probability from lemma 4.2.2.

So $h_n(w) \longrightarrow 0$ in probability.

Lemma 4.2.4 $\varepsilon_n \longrightarrow 0, \quad n\varepsilon_n^2 \longrightarrow \infty \Rightarrow (1 - \varepsilon_n^2)^n \longrightarrow 0.$

Proof Take any $\varepsilon > 0$.

Choose $M > 0$ such that $e^{-M} < \varepsilon$.

Choose $N_1 > 0$ such that $n > N_1 \Rightarrow (1 - \frac{M}{n})^n < \varepsilon$.

Choose $N_2 > 0$ such that $n > N_2 \Rightarrow n\,\varepsilon_n^2 > M$.

Let $N = \text{Max}\,(N_1, N_2)$

$(1 - \varepsilon_n^2)^n < \varepsilon \iff 1 - \varepsilon_n^2 > \varepsilon^{1/n} \iff 1 - \varepsilon^{1/n} > \varepsilon_n^2$.

Observe that $n > N \Rightarrow \varepsilon_n^2 > \frac{M}{n}$.

It suffices to show that $\frac{M}{n} > 1 - \varepsilon^{1/n}$.

Now $\frac{M}{n} > 1 - \varepsilon^{1/n} \iff \varepsilon^{1/n} > 1 - \frac{M}{n}$ or $\varepsilon > (1 - \frac{M}{n})^n$ which is true for $n > N$. Hence the lemma.

Lemma 4.2.5 Let $\alpha_{n1}^*(w) = \bigcup_{i=1}^{n} \left\{ x \in R^2 : \|X_i(w) - x\| \leq h_n(w) \right\}$

Then α_{n1}^* is a consistent estimate of α.

Proof Let ε_n be a sequence of positive real numbers such that $1 > \varepsilon_n$ for all n and $\varepsilon_n \downarrow 0$.

Since $nh_n^2 \longrightarrow \infty$ in probability, there exists $N_{1k} > 0$

such that for all $n \geq N_{1k}$, $Pr(nh_n^2 \geq k) \geq 1 - \frac{\varepsilon_k}{4}$ for all $k = 1,2,\ldots$

Similarly since $h_n \longrightarrow 0$ in probability, there exists $N_{2k} > 0$ such that for all $n > N_{2k}$, $Pr(h_n \leq \varepsilon_k) \geq 1 - (\varepsilon_k/4)$ for all $k = 1,2,\ldots$

Let $N_k = \text{Max}\,(N_{1k}, N_{2k})$ for all k.

Let $s_i = 1$ for $i = 1,\ldots N_2$

$= k$ for $i = N_k + 1,\ldots N_{k+1}$ for $k = 2,3,\ldots$

It is clear that $s_n \longrightarrow \infty$.

Let $\gamma_n = (s_n/n)^{1/2}$

Let $t_i = \varepsilon_1$ for $i = 1,\ldots N_2$

$= \varepsilon_k$ for $i = N_k + 1,\ldots, N_{k+1}$ for $k = 2,3,\ldots$

It is clear that $t_n \longrightarrow 0$.

Now $Pr(t_n > h_n > \gamma_n) \longrightarrow 1$ as $n \longrightarrow \infty$.

So $\gamma_n \longrightarrow 0$ since $t_n \longrightarrow 0$.

Similarly $Pr(nt_n^2 > nh_n^2 > n\gamma_n^2) \longrightarrow 1$ as $n \longrightarrow \infty$

$n\gamma_n^2 \longrightarrow \infty$ as $n \longrightarrow \infty$. So $nt_n^2 \longrightarrow \infty$.

Let $D(t_n)$ be the closed disc of radius t_n around origin.

Let
$$\left.\begin{aligned} \alpha_n^{in} &= \alpha \cap \left[\alpha^c \oplus D(t_n)\right]^c \\ \alpha_n^{out} &= \left[\alpha \oplus D(t_n)\right]^c \end{aligned}\right\}$$
where $\oplus$ represents minkowski addition.

(Fig.4.2.3)

Let $I_{1n}(x,\omega) = 1$ it $x \in \alpha_{n1}^*(\omega)$ [For the measurability of I_{1n}, see
$= 0$ otherwise. appendix A7].

$I_1(x,\omega) = 1$ if $x \in \alpha$
$= 0$ otherwise.

Let $A_n = \{\omega : h_n(\omega) \geq \gamma_n\}$

$Pr(A_n) \longrightarrow 1$ as $n \longrightarrow \infty$ (From lemma 4.2.3). So $Pr(A_n^c) \longrightarrow 0$ as $n \longrightarrow \infty$.

Let $J_n = E_\alpha\left[\lambda(\alpha_{n1}^* \Delta \alpha)\right]$

$$= \int_\Omega \left(\int_{R^2} |I_{1n} - I_1| \, d\lambda \right) dP_\alpha$$

$$= \int_{R^2} \left(\int_\Omega |I_{1n} - I_1| \, dP_\alpha \right) d\lambda \text{ by Fubini's theorem.}$$

$$= \int_{\alpha_n^{in}} \left(\int_{A_n} (1 - I_{1n}) \, dP_\alpha \right) d\lambda$$

$$+ \int_{\alpha_n^{in}} \left(\int_{A_n^c} (1 - I_{1n}) \, dP_\alpha \right) d\lambda + \int_{\alpha_n^{out}} \left(\int_\Omega I_{1n} \, dP_\alpha \right) d\lambda$$

$$+ \int_{(\alpha_n^{in} \cup \alpha_n^{out})^c} \left(\int_\Omega |I_{1n} - I_1| \, dP_\alpha \right) d\lambda$$

$= J_{1n} + J_{2n} + J_{3n} + J_{4n}$ where J_{in}'s are the corresponding integrals.

$$(\alpha_n^{in} \cup \alpha_n^{out})^c = \left[\left\{ \alpha \cap [\alpha^c \oplus D(t_n)]^c \right\} \cup \left\{ \alpha \oplus D(t_n) \right\}^c \right]^c$$

$$= \left\{ \alpha \cap [\alpha^c \oplus D(t_n)]^c \right\}^c \cap [\alpha \oplus D(t_n)]$$

$$= \left\{ \alpha^c \cup [\alpha^c \oplus D(t_n)] \right\} \cap [\alpha \oplus D(t_n)]$$

$$= \left\{ \alpha^c \cap [\alpha \oplus D(t_n)] \right\} \cup \left\{ [\alpha^c \oplus D(t_n)] \cap [\alpha \oplus D(t_n)] \right\}$$

$$\left[\alpha \oplus D(t_n)\right] \longrightarrow cl(\alpha) = \alpha$$

$$\left[\alpha^c \oplus D(t_n)\right] \longrightarrow cl(\alpha^c)$$

So $\left\{\left[\alpha^c \oplus D(t_n)\right] \cap \left[\alpha \oplus D(t_n)\right]\right\} \longrightarrow \alpha \cap cl(\alpha^c) = \delta\alpha$

$$\alpha^c \cap \left[\alpha \oplus D(t_n)\right] \longrightarrow \alpha^c \cap \alpha = \emptyset$$

So $\lambda\left\{\left[\alpha_n^{in} \cup \alpha_n^{out}\right]^c\right\} \longrightarrow \lambda\ (\delta\alpha) = 0$.

So $J_{4n} \leq \lambda\left[(\alpha_n^{in} \cup \alpha_n^{out})^c\right] \longrightarrow 0 \Rightarrow J_{4n} \longrightarrow 0$.

$$J_{3n} = \int_{\alpha_n^{out}} \left(\int_{\Omega} I_{\bigcup_{i=1}^{n} \{(x,w):\ \|X_i(w) - x\| \leq h_n\}}\, dP_\alpha \right) d\lambda$$

[where $I_{(\cdot)}$ indicates the indicator function of '.']

$$= \int_{\alpha_n^{out}} \left(\int_{(h_n \leq t_n)} I_{\bigcup_{i=1}^{n} \{(x,w):\ \|X_i(w) - x\| \leq h_n\}}\, dP_\alpha \right) d\lambda$$

$$+ \int_{\alpha_n^{out}} \left(\int_{(h_n > t_n)} I_{\bigcup_{i=1}^{n} \{(x,w):\ \|X_i(w) - x\| \leq h_n\}}\, dP_\alpha \right) d\lambda$$

$= J_{31n} + J_{32n}$ where J_{31n} and J_{32n} are the respective integrals.

$$J_{31n} \leq \int_{\alpha_n^{out}} \left(\int_{(h_n \leq t_n)} I_{\bigcup_{i=1}^{n} \{(x,w):\ \|X_i(w) - x\| \leq t_n\}}\, dP_\alpha \right) d\lambda$$

$$\leq \int_{\alpha_n^{out}} \left(\int_{\Omega} I_{\bigcup_{i=1}^{n} \{(x,w):\ \|X_i(w) - x\| \leq t_n\}}\, dP_\alpha \right) d\lambda .$$

$$x \in \alpha_n^{out} \Rightarrow I_{\bigcup_{i=1}^{n} \{(x,w):\ \|X_i(w) - x\| \leq t_n\}} = 0 \quad \text{a.e.}(w)$$

since $X_i(w) \in \alpha$ a.e.(w) and for all i.

So $J_{31n} \to 0$ as $n \to \infty$

Let $a = \sup\{\|x-y\| : x, y \in \alpha\}$

'a' is finite since α is bounded. Observe that $h_n \leq a$ a.e.(w) and for all n.

Then $\alpha_n^{out} = [\alpha \oplus D(a)]^c \cup [\alpha_n^{out} \cap (\alpha \oplus D(a))]$ where D(a) is the closed disc of radius 'a' around origin.

$$J_{32n} = \int_{(\alpha \oplus D(a))^c} \left(\int_{(h_n > t_n)} I_{\bigcup_{i=1}^{n} \{(x,w) : \|X_i(w)-x\| \leq h_n\}} dP_\alpha \right) d\lambda$$

$$+ \int_{(\alpha \oplus D(a)) \cap \alpha_n^{out}} \left(\int_{(h_n > t_n)} I_{\bigcup_{i=1}^{n} \{(x,w) : \|X_i(w)-x\| \leq h_n\}} dP_\alpha \right) d\lambda$$

$$\leq \int_{(\alpha \oplus D(a))^c} \left(\int_{(h_n > t_n)} I_{\bigcup_{i=1}^{n} \{(x,w) : \|X_i(w)-x\| \leq a\}} dP_\alpha \right) d\lambda$$

$$+ \int_{(\alpha \oplus D(a))} \left(\int_{(h_n > t_n)} I_{\bigcup_{i=1}^{n} \{(x,w) : \|X_i(w)-x\| \leq h_n\}} dP_\alpha \right) d\lambda$$

$$x \in (\alpha \oplus D(a))^c \Rightarrow I_{\bigcup_{i=1}^{n} \{(x,w) : \|X_i(w)-x\| \leq a\}} = 0 \quad \text{a.e.}(w).$$

So

$$\int_{(\alpha \oplus D(a))^c} \left(\int_{(h_n > t_n)} I_{\bigcup_{i=1}^{n} \{(x,w) : \|X_i(w)-x\| \leq a\}} dP_\alpha \right) d\lambda$$

$$\to 0 \text{ as } n \to \infty.$$

$$\int_{(\alpha \oplus D(a))} \left(\int_{(h_n > t_n)} I_{\bigcup_{i=1}^{n} \{(x,w):\ \|X_i(w)-x\| \le h_n\}} dP_\alpha \right) d\lambda$$

$\le \lambda(\alpha \oplus D(a)) Pr(h_n > t_n) \longrightarrow 0$ as $n \longrightarrow \infty$ since $\lambda\left[\alpha \oplus D(a)\right]$ is finite.

i.e., $J_{3n} \longrightarrow 0$ as $n \longrightarrow \infty$.

$J_{2n} \le \lambda(\alpha_n^{in})\ Pr(A_n^c) \longrightarrow 0$ as $n \longrightarrow \infty$ since $\lambda(\alpha_n^{in}) \le \lambda(\alpha)$ which is finite.

$$J_{1n} = \int_{\alpha_n^{in}} \left(\int_{A_n} (1 - I_{1n}) dP_\alpha \right) d\lambda$$

$$= \int_{\alpha_n^{in}} \left(\int_{A_n} I_{\bigcap_{i=1}^{n} \{(x,w):\ \|X_i(w)-x\| > h_n(w)\}} dP_\alpha \right) d\lambda$$

$$\le \int_{\alpha_n^{in}} \left(\int_{A_n} I_{\bigcap_{i=1}^{n} \{(x,w):\ \|X_i(w)-x\| > \gamma_n\}} dP_\alpha \right) d\lambda$$

$$\le \int_{\alpha_n^{in}} \left(\int_{\Omega} I_{\bigcap_{i=1}^{n} \{(x,w):\ \|X_i(w)-x\| > \gamma_n\}} dP_\alpha \right) d\lambda$$

$\le \lambda(\alpha) \left(1 - \frac{\pi \gamma_n^2}{\lambda(\alpha)}\right)^n \longrightarrow 0$ as $n \longrightarrow \infty$ [$\lambda(\alpha)$ is finite. Lemma 4.2.4 is applied]

So α_{n1}^* is a consistent estimate of α.

Observe that the estimator α_{n1}^* may not be connected. In the following theorem a path connected estimator α_{n2}^* is defined under the assumption of uniform distribution and it is proved to be consistent.

Theorem 4.2.2 Let $\alpha \in \mathcal{C}$, α unknown. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors from $(\Omega, \mathcal{A}, P_\alpha)$ to R^2, taking values in R^2 and following uniform distribution on α.

$$\text{Let } G_n(w) = \text{MST of } \{X_1(w), X_2(w), \ldots X_n(w)\}, \quad w \in \Omega .$$

$$\ell_n(w) = \text{Length of } G_n(w), \quad w \in \Omega ,$$

$$h_n(w) = (\ell_n(w)/n)^{1/2}, \quad w \in \Omega ,$$

$$m_n(w) = \text{maximum edge length of } G_n(w), \quad w \in \Omega$$

$$\text{and } \alpha^*_{n2}(w) = \{x : d(x, G_n(w)) \leq h_n(w)\}, \quad w \in \Omega .$$

Then α^*_{n2} is a consistent estimator of α.

Proof This theorem will be proved using the lemmas 4.2.1, 4.2.2, 4.2.3, 4.2.4 and 4.2.5 and the following lemmas.

Lemma 4.2.6 Let $u_n(w) = \sup_{x \in G_n} \inf_{y \in \alpha} \|x-y\|$. [Def. 3.3.2]

Then $u_n(w) \longrightarrow 0$ in probability.

Proof u_n is proved to be measurable in appendix [A6]. It is to be shown that for any $\varepsilon > 0$, $\Pr(u_n > \varepsilon) \longrightarrow 0$ as $n \longrightarrow \infty$. The proof is similar to the proof of lemma 4.2.2.

α is bounded. So there exists a square of side $b > 0$ such that $\alpha \subseteq B$. Cover α by k^2 squares of side $\frac{b}{k}$, represented by $A_1, A_2, \ldots A_{k^2}$.

Then $\sup\limits_{x\in A_i \;\; y\in A_j} d(x,y) = 2\sqrt{2}\frac{b}{k}$ where A_i and A_j are two adjacent squares.

Let $\varepsilon > 0$. Let k be a positive integer such that $\sqrt{2}\frac{b}{k} < \varepsilon$. Cover B by k^2 squares $A_1, A_2, \dots A_{k^2}$ such that each square is of side $\frac{b}{k}$. Let $M > 0$ be a positive integer such that $1 \leq M \leq k^2$, $\bigcup\limits_{i=1}^{M} A_i \supseteq \alpha$ and $Int(A_i) \cap \alpha \neq \emptyset$ for all $i = 1, 2, \dots M$.

As $n \to \infty$, there exists at least one j_i such that $X_{j_i} \in A_i$ for $i = 1, \dots M$ with probability one (The argument is similar to that of lemma 4.2.1).

Let $x \in G_n(w)$. There exists $X_i(w)$ and $X_j(w)$ such that

$$x \in \left[X_i(w),\ X_j(w)\right].$$

$$d(x,\alpha) \leq d(X_i(w), X_j(w)) \leq m_n(w)$$

$$\sup_{x\in G_n(w)} d(x,\alpha) \leq m_n(w)$$

$$\text{i.e.,}\quad u_n(w) \leq m_n(w).$$

So $u_n(w) \to 0$ in probability since $m_n(w) \to 0$ inprobability (Lemma 4.2.2).

Lemma 4.2.7 α^*_{n2} is a consistent estimator of α.

<u>Proof</u> Let ε_n be a sequence of positive real numbers such that $1 > \varepsilon_n$ for all n and $\varepsilon_n \downarrow 0$.

Since $n h_n^2 \to \infty$ in probability (Lemma 4.2.1), there exists $N_{1k} > 0$ such that for all $n \geq N_{1k}$,

$Pr\,(n h_n^2 \geq k) \geq 1 - \frac{\varepsilon_k}{4}$ for all $k = 1, 2, \dots$

Similarly, since $h_n \rightarrow 0$ in probability (Lemma 4.2.3), there exists $N_{2k} > 0$ such that for all $n \geq N_{2k}$, $Pr(h_n \leq \varepsilon_k) \geq 1 - (\varepsilon_k/4)$ for all $k=1,2,\ldots$ It is also known that $\psi_n(w) \rightarrow 0$ in probability.

So there exists $N_{3k} > 0$ such that $Pr(\psi_n \leq \varepsilon_k) \geq 1 - (\varepsilon_k/4)$ for all $n \geq N_{3k}$.

Let $N_k = \text{Max}\,(N_{1k}, N_{2k}, N_{3k})$ for all k.

Let $s_i = 1$ for $i = 1,\ldots N_2$

$= k$ for $i = N_k + 1,\ldots N_{k+1}$ for $k = 2,3,\ldots$

It is clear that $s_n \rightarrow \infty$. Let $\gamma_n = (s_n/n)^{1/2}$.

Let $t_i = \varepsilon_1$ for $i = 1,2,\ldots N_2$

$= \varepsilon_k$ for $i = N_k + 1,\ldots N_{k+1}$ for $k = 2,3,\ldots$

It is clear that $t_n \rightarrow 0$.

Now $Pr(t_n > h_n > \gamma_n) \rightarrow 1$ as $n \rightarrow \infty$. So $\gamma_n \rightarrow 0$ since $t_n \rightarrow 0$. Similarly $Pr(nt_n^2 > n\,h_n^2 > n\,\gamma_n^2) \rightarrow 1$ as $n \rightarrow \infty$.

$n\,\gamma_n^2 \rightarrow \infty \Rightarrow n\,t_n^2 \rightarrow \infty$. Let $D(t_n)$ be the closed disc of radius t_n around the origin. $D(2t_n)$ be the closed disc of radius $2t_n$ around the origin.

$$\text{Let } \alpha_n^{out} = \left[\alpha \oplus D(2t_n)\right]^c$$

$$\alpha_n^{in} = \alpha \cap \left[\alpha^c \oplus D(t_n)\right]^c$$

Let $I_{n2}(x,w) = 1$ if $x \in \alpha_{n2}^*(w)$, $w \in \Omega$.

$= 0$ otherwise.

$I(x,w) = 1$ if $x \in \alpha$

$= 0$ otherwise.

Measurability of I_{n2} is shown in appendix A8.

Let $J_{3n} = \int_{\alpha_n^{out}} \left(\int_{\Omega} \left| I_{n2} - I_1 \right| dP_\alpha \right) d\lambda$

$$= \int_{\alpha_n^{out}} \left(\int_{(h_n \leq t_n) \cap (u_n \leq t_n)} I_{\{(x,w):\ d(x, G_n(w)) \leq h_n\}} dP_\alpha \right) d\lambda$$

$$+ \int_{\alpha_n^{out}} \left(\int_{(h_n > t_n) \cap (u_n \leq t_n)} I_{\{(x,w):\ d(x, G_n(w)) \leq h_n\}} dP_\alpha \right) d\lambda$$

$$+ \int_{\alpha_n^{out}} \left(\int_{(h_n \leq t_n) \cap (u_n > t_n)} I_{\{(x,w):\ d(x, G_n(w)) \leq h_n\}} dP_\alpha \right) d\lambda$$

$$+ \int_{\alpha_n^{out}} \left(\int_{(h_n > t_n) \cap (u_n > t_n)} I_{\{(x,w):\ d(x, G_n(w)) \leq h_n\}} dP_\alpha \right) d\lambda$$

$= J_{31n} + J_{32n} + J_{33n} + J_{34n}$ where J_{3in}'s are the respective integrals.

$$J_{31n} \leq \int_{\alpha_n^{out}} \left(\int_{(h_n \leq t_n) \cap (u_n \leq t_n)} I_{\{(x,w):\ d(x, G_n(w)) \leq t_n\}} dP_\alpha \right) d\lambda$$

$$x_0 \in \alpha_n^{out} \Rightarrow I_{\{(x_0,w):\ d(x_0, G_n(w)) \leq t_n\}} = 0$$

for $w \in (h_n \leq t_n) \cap (u_n \leq t_n)$

So $J_{31n} \longrightarrow 0$ as $n \longrightarrow \infty$.

$J_{32n} + J_{33n} + J_{34n}$

$$\leq \int_{\alpha_n^{out}} \left(\int_{(h_n > t_n)} I_{\{(x,w) : d(x, G_n(w)) \leq h_n(w)\}} \, dP_\alpha \right) d\lambda$$

$$+ \int_{\alpha_n^{out}} \left(\int_{(u_n > t_n)} I_{\{(x,w) : d(x, G_n(w)) \leq h_n(w)\}} \, dP_\alpha \right) d\lambda$$

$$= \int_{[\alpha \oplus D(2a)]^c} \left(\int_{(h_n > t_n)} I_{n2} \, dP_\alpha \right) d\lambda \quad \text{where } a = \sup_{x, y \in \alpha} d(x, y)$$

$$+ \int_{\alpha_n^{out} \cap (\alpha \oplus D(2a))} \left(\int_{(h_n > t_n)} I_{n2} \, dP_\alpha \right) d\lambda$$

$$+ \int_{[\alpha \oplus D(2a)]^c} \left(\int_{(u_n > t_n)} I_{n2} \, dP_\alpha \right) d\lambda$$

$$+ \int_{[\alpha \oplus D(2a)] \cap \alpha_n^{out}} \left(\int_{(u_n > t_n)} I_{n2} \, dP_\alpha \right) d\lambda$$

$$0 + \lambda(\alpha \oplus D(2a)) Pr(h_n > t_n) + 0 + \lambda(\alpha \oplus D(2a)) Pr(u_n > t_n)$$

$\longrightarrow 0$ as $n \longrightarrow \infty$.

So $J_{3n} \longrightarrow 0$.

Let $J_n = E_\alpha[\lambda(\alpha_{n2} * \Delta\alpha)] = \int_{\alpha_n^{in}} \left(\int_{(h_n > \gamma_n)} (1 - I_{n2}) dP_\alpha \right) d\lambda$

$$= \int_{\alpha_n^{in}} \left(\int_{(h_n > \gamma_n)} (1 - I_{n2}) dP_\alpha \right) d\lambda + \int_{\alpha_n^{out}} \left(\int_{\Omega} I_{n2} \, dP_\alpha \right) d\lambda$$

$$+ \int_{(\alpha_n^{in} \cup \alpha_n^{out})^c} \left(\int_{\Omega} |I_1 - I_{n2}| \, dP_\alpha \right) d\lambda$$

$= J_{1n} + J_{2n} + J_{3n} + J_{4n}$ where J_{1n}, J_{2n}, J_{3n} and J_{4n} are the respective integrals.

$J_{4n} \leq \lambda\ (\alpha_n^{in} \cup \alpha_n^{out})^c \longrightarrow \lambda(\delta\alpha) = 0.$

$J_{3n} \longrightarrow 0$ as proved earlier.

$J_{2n} \leq Pr\ (h_n > \gamma_n)\ \lambda(\alpha) \longrightarrow 0$ as $n \longrightarrow \infty$.

$$J_{1n} = \int_{\alpha_n^{in}} \left(\int_{(h_n > \gamma_n)} I_{\{(x,w):\ d(x, G_n(w)) > h_n\}}\ dP_\alpha \right) d\lambda$$

$$\leq \int_{\alpha_n^{in}} \left(\int_{(h_n > \gamma_n)} I_{\{(x,w):\ d(x, G_n(w)) > \gamma_n\}}\ dP_\alpha \right) d\lambda$$

$$\leq \int_{\alpha_n^{in}} \left(\int_{(h_n > \gamma_n)} I_{\bigcap_{i=1}^{n}\{(x,w)\ :\ \| X_i(w) - x \| > \gamma_n\}}\ dP_\alpha \right) d\lambda$$

$$\leq \int_{\alpha_n^{in}} \left(\int_{\Omega} I_{\bigcap_{i=1}^{n}\{(x,w):\ \| X_i(w) - x \| > \gamma_n\}}\ dP_\alpha \right) d\lambda$$

$\leq \lambda(\alpha)\ (1 - (\pi\gamma_n^{\ 2}/\lambda(\alpha)))^n \longrightarrow 0$ as $n \longrightarrow \infty$. Hence the theorem.

In the next section, generalization to any continuous distribution defined in Chapter II is provided for the same $\alpha_{n2}^{\ *}$.

4.3 Consistent estimation for compact region where the distribution need not be uniform.

The properties of continuous distributions on compact regions have been defined in Chapter II (2.3.c). They are the following.

2.3.c.1 Q_α is a probability measure with support $\alpha \in \mathcal{E}$.

2.3.c.2 $Q_\alpha << \lambda$.

2.3.c.3 The density f is continuous on $\text{Int}(\alpha)$

2.3.c.4 $\text{Int}\left\{x : f(x) \geq \frac{1}{n}\right\}$ is path connected for sufficiently large n.

2.3.c.5 $\inf_{x \in \alpha^{in}_\varepsilon} f(x) > 0$ for every $\varepsilon > 0$ where

$$\alpha^{in}_\varepsilon = \alpha \cap \left[\alpha^c \oplus D(\varepsilon)\right]^c .$$

$$\mathcal{G}_\alpha = \left\{Q : Q \text{ is a probability measure on } \alpha \text{ satisfying the above properties}\right\} \text{ where } \alpha \text{ is a compact region.}$$

4.3.a Observe that in all the lemmas proved in the previous section, the fact that the distribution is uniform, is used only when the integrals on α^{in}_n are calculated. Essentially, the fact that was used was "$Pr(v \cap \alpha) \neq 0$ for all v open, $v \cap \alpha \neq \emptyset$ ", which holds here also because of 2.3.c.5.

A theorem is stated below for the generalization to continuous distributions.

Theorem 4.3.1 Let $\alpha \in \mathcal{E}$ and $Q_\alpha \in \mathcal{G}_\alpha$. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors on $(\Omega, \mathcal{A}, P_\alpha)$ with the induced measure Q_α.

Let $G_n(w)$ = MST of $\{X_1(w), \ldots X_n(w)\}$, $w \in \Omega$;

$\ell_n(w)$ = Length of $G_n(w)$, $w \in \Omega$;

$m_n(w)$ = maximum edge length of $G_n(w)$, $w \in \Omega$;

$h_n(w) = (\ell_n(w)/n)^{1/2}$, $w \in \Omega$, and

$$\alpha^*_{n1}(w) = \bigcup_{i=1}^{n} \{x : \|X_i(w) - x\| \le h_n(w)\}, \quad w \in \Omega$$

Then α^*_{n1} is a consistent estimator of α.

<u>Proof</u> It can be easily proved, after the observation made in 4.3.a, that

(i) $\ell_n(w) \dashrightarrow \infty$ in probability

(ii) $m_n(w) \dashrightarrow 0$ in probability and

(iii) $h_n(w) \dashrightarrow 0$ in probability.

Let $1 > \epsilon_n$ for all n and $\epsilon_n \downarrow 0$. Define t_n and γ_n similar to lemma 4.2.5 of theorem 4.2.1.

Then $t_n \ge \gamma_n$, $nt_n^2 \dashrightarrow \infty$, $t_n \dashrightarrow 0$, $n\gamma_n^2 \dashrightarrow \infty$ and $\gamma_n \dashrightarrow 0$ and of course $Pr(t_n \ge h_n \ge \gamma_n) \dashrightarrow 1$.

Define $\alpha_n^{in} = \alpha \cap [\alpha^c \oplus D(t_n)]^c$

$$\alpha_n^{out} = [\alpha \oplus D(t_n)]^c$$

$$J_n = E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right] = \int_{\alpha_n^{in}} \left(\int_{(h_n \ge \gamma_n)} (1 - I_{\alpha_{n1}^*})\, dP_\alpha\right) d\lambda$$

$$+ \int_{\alpha_n^{in}} \int_{(h_n < \gamma_n)} (1 - I_{\alpha_{n1}^*})\, dP_\alpha\, d\lambda + \int_{\alpha_n^{out}} \left(\int_\Omega I_{\alpha_{n1}^*}\, dP_\alpha\right) d\lambda$$

$$+ \int_{(\text{in}_{\alpha_n} \cup \text{out}_{\alpha_n})^c} \left(\int_\Omega \left| I_\alpha - I_{\alpha_{n1}}^* \right| dP_\alpha \right) d\lambda = J_{1n} + J_{2n} + J_{3n} + J_{4n}$$

The proofs of $J_{2n} \dashrightarrow 0$, $J_{3n} \dashrightarrow 0$ and $J_{4n} \dashrightarrow 0$ are not being given. They are same proofs given in lemma 4.2.5 of theorem 4.2.1. The proof of $J_{1n} \dashrightarrow 0$ will be given below.

Proposition 4.3.1 : $J_{1n} \dashrightarrow 0$ as $n \dashrightarrow \infty$.

<u>Proof</u> $n \gamma_n^2 \dashrightarrow \infty$ as $n \dashrightarrow \infty$.

So $n \gamma_n^2 > k$ for $n \geq M_k$ for $k = 1, 2, \ldots$

Let $\theta_i = \gamma_1$ for $i = 1, \ldots M_2$

$= \gamma_k$ for $i = M_k + 1, \ldots M_{k+1}$ for $k = 2, 3, \ldots$

Observe that $n \gamma_n^2 \theta_n \dashrightarrow \infty$ and $\theta_n \dashrightarrow 0$ as $n \dashrightarrow \infty$.

Let $B_n = \{ x : f(x) \geq \theta_n \}$. $\text{Int}(B_n)$ is path connected for sufficiently large n [2.3.c.4]. $\lambda(B_n^c \cap \alpha) \dashrightarrow 0$ as $n \dashrightarrow \infty$.

Let $B_n^{in} = B_n \cap \left[B_n^c \oplus D(\gamma_n) \right]^c$

$$J_{1n} = \int_{\text{in}_{\alpha_n}} \left(\int_{(h_n \geq \gamma_n)} \left(1 - I_{\bigcup_{i=1}^{n} \{(x, w) : \| X_i(w) - x \| \leq h_n(w)\}}\right) dP_\alpha \right) d\lambda$$

$$\leq \int_{\text{in}_{\alpha_n}} \left(\int_\Omega I_{\bigcap_{i=1}^{n} \{(x, w) : \| X_1(w) - x \| > \gamma_n \}} dP_\alpha \right) d\lambda$$

$$= \int_{\text{in}_{\alpha_n}} \left[1 - \Pr(X_1 \in (x \oplus D(\gamma_n))) \right]^n d\lambda(x)$$

$$= \int_{\alpha_n^{in} \cap B_n^{in}} \left[1 - Pr(X_1 \in (x \oplus D(\gamma_n)))\right]^n d\lambda(x)$$

$$+ \int_{\alpha_n^{in} \cap (B_n^{in})^c} \left[1 - Pr(X_1 \in (x \oplus D(\gamma_n)))\right]^n d\lambda(x)$$

$= J_{11n} + J_{12n}$ where J_{11n}, J_{12n} are the corresponding integrals.

$J_{11n} \leq \lambda(\alpha)\left[1 - \pi\, \gamma_n^2\, \theta_n\right]^n$ because eventually

$x \in \alpha_n^{in} \cap B_n^{in} \Rightarrow f(y) \geq \theta_n$ for $y \in (x \oplus D(\gamma_n))$ } From 2.3.c.4 and 2.3.c.5.

So $J_{11n} \to 0$ as $n \to \infty$, since $n\, \gamma_n^2\, \theta_n \to \infty$

$$J_{12n} \leq \lambda\left[\alpha_n^{in} \cap (B_n^{in})^c\right]$$

$$\leq \lambda\left[(B_n^{in})^c \cap \alpha\right] = \lambda\left\{\left[(B_n^c \cup \left[B_n^c \oplus D(\gamma_n)\right])\right] \cap \alpha\right\}$$

$$\leq \lambda(B_n^c \cap \alpha) + \lambda\left[(B_n^c \oplus D(\gamma_n)) \cap \alpha\right] \to \lambda(\delta\alpha) = 0$$

So $J_{1n} \to 0$. That proves the theorem.

Similarly the following theorem can be proved.

Theorem 4.3.2 Let $\alpha \in \mathcal{E}$ and $Q_\alpha \in \mathcal{G}_\alpha$. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors on $(\Omega, \mathcal{A}, P_\alpha)$ with the induced measure Q_α.

Let $G_n(\omega)$ = MST of $\{X_1(\omega), \ldots X_n(\omega)\}$, $\omega \in \Omega$;

$\ell_n(\omega)$ = length of $G_n(\omega)$, $\omega \in \Omega$,

$m_n(\omega)$ = Maximum edge length of $G_n(\omega)$, $\omega \in \Omega$;

$h_n(\omega)$ = $(\ell_n(\omega)/n)^{1/2}$, $\omega \in \Omega$;

$u_n(\omega) = \sup_{x \in G_n(\omega)} \inf_{y \in \alpha} \|x-y\|$ and

$\alpha^*_{n2}(\omega) = \{x : d(x, G_n(\omega)) \leq h_n(\omega)\}$, $\omega \in \Omega$.

Then α^*_{n2} is a consistent estimator of α.

4.4 Classification and estimation of compact regions.

In the previous sections, path connected consistent estimator for compact region is found for continuous distributions. Automatic classification and consistent estimation of compact regions is going to be dealt with in this section. A theorem is stated below where the division is also achieved using MST.

Theorem 4.4.1 Let $A, B \in \mathcal{E}$, $A \cap B = \emptyset$. Let $Q_1 \in \mathcal{G}_A$ and $Q_2 \in \mathcal{G}_B$. Let $0 < q < 1$ and $Q = qQ_1 + (1-q)Q_2$. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ to R^2 inducing the measure Q on α where $\alpha = A \cup B$. Let $G_n(\omega)$ be the MST of $X_1(\omega), \ldots X_n(\omega)$, $\omega \in \Omega$ and $m_n(\omega)$ be the maximal edge length of $G_n(\omega)$. Delete the maximal edge from MST without removing the nodes of it. [i.e. if the maximal edge is $[X_i, X_j]$ then remove the set $\{x : x = \tau X_i + (1-\tau)X_j ; 0 < \tau < 1\}$ from $G_n(\omega)$.

Let the two trees be denoted by $G_{1n}(w)$ and $G_{2n}(w)$ with the respective number of points n_1 and n_2 where $n_1 + n_2 = n$.

Let $h_{1n}(w) = \left[\ell(G_{1n}(w))/n_1\right]^{1/2}$ and

$h_{2n}(w) = \left[\ell(G_{2n}(w))/n_2\right]^{1/2}$.

Let $F^*_{1n}(w) = \left\{x : d(x, G_{1n}(w)) \leq h_{1n}(w)\right\}$ and

$F^*_{2n}(w) = \left\{x : d(x, G_{2n}(w)) \leq h_{2n}(w)\right\}$ for $w \in \Omega$.

Let $\alpha^*_n(w) = F^*_{1n}(w) \cup F^*_{2n}(w)$.

Then $E_\alpha\left[\lambda(\alpha^*_n \Delta \alpha)\right] \rightarrow 0$ as $n \rightarrow \infty$.

<u>Proof</u> A and B are closed and $A \cap B = \emptyset$.

So $d(A,B) = \text{Inf}\left\{d(x,y) : x \in A,\ y \in B\right\} = \tau > 0$.

Cover $A \cup B$ by closed squares, the side of each being $\tau/20$. Only finitely many such squares are needed since $A \cup B$ is bounded. Let the squares be termed as $c_{11}, \ldots c_{1k_1}, c_{21}, \ldots c_{2k_2}$ such that

$$\bigcup_{i=1}^{k_1} c_{1i} \supseteq A, \quad \bigcup_{i=1}^{k_2} c_{2i} \supseteq B , \quad \left(\bigcup_{i=1}^{k_1} c_{1i}\right) \cap \left(\bigcup_{i=1}^{k_2} c_{2i}\right) = \emptyset ,$$

$\text{Int}(c_{1i}) \cap A \neq \emptyset$ for $i = 1, \ldots k_1$ and $\text{Int}(c_{2i}) \cap B \neq \emptyset$ for $i=1, \ldots k_2$.

$$\Pr(X_1 \notin c_{ji},\ X_2 \notin c_{ji}, \ldots) = 0 \quad \text{for } i = 1, \ldots k_j\ ;\ j = 1,2.. \quad [4.4.1]$$

Let $S_{1n}(w) = \{X_1(w), \ldots X_n(w)\} \cap A$ for $w \in \Omega$

and $S_{2n}(w) = \{X_1(w), \ldots X_n(w)\} \cap B$ for $w \in \Omega$

Let $H_{in}(w)$ = MST generated by $S_{in}(w)$ for $i = 1,2$.

Let $Y_{in}(w) = \{x: d(x, H_{in}) \leq \beta_{in}(w)\}$ where

$$\beta_{in}(w) = \left[\frac{\ell(H_{in}(w))}{\# S_{in}(w)}\right]^{1/2} \quad \text{for } i = 1,2.$$

Let $m_{in}(w)$ = Maximal edge length of $H_{in}(w)$ for $w \in \Omega$, $i=1,2$.

Then $Pr(m_{in} > (2\sqrt{2}\ \tau/20)) \dashrightarrow 0$ for $i=1,2$ [4.4.2] (from 4.4.1).

[$\sup_{x \in A_i,\ y \in A_j} d(x,y) = 2\sqrt{2}\ \tau/20$ where A_i and A_j are two adjacent squares of side $\tau/20$ (Fig.4.2.2)]

Observe that $Pr(m_n \geq \tau) \dashrightarrow 1$ as $n \dashrightarrow \infty$ since $d(A,B) = \tau$. Observe also that by taking squares of smaller side, it can be shown that $m_{in} \dashrightarrow 0$ in probability for $i = 1,2$. [From 4.4.2].

$m_{in} \dashrightarrow 0$ in probability $\Rightarrow \sqrt{m_{in}} \dashrightarrow 0$ in probability.

$$\beta_{in}(w) = \left[\frac{\ell(H_{in}(w))}{\# S_{in}(w)}\right]^{1/2} \leq \left[\frac{(\# S_{in}(w))m_{in}(w)}{\# S_{in}(w)}\right]^{1/2} = \sqrt{m_{in}(w)}$$

for $w \in \Omega$.

$\sqrt{m_{1n}} + \sqrt{m_{2n}} \dashrightarrow 0$ in probability.

Let $1 > \varepsilon_k \downarrow 0$.

$Pr(m_{in} \leq \varepsilon_k) \geq 1 - \varepsilon_k/4$ for all $n \geq N_{ik}$ for $k = 1,2,\ldots$ and $i=1,2$.

So $\Pr\left\{\left[\sqrt{m}_{1n} + \sqrt{m}_{2n}\right] \leqslant 2\sqrt{\varepsilon_k}\right\} \geq 1 - (\varepsilon_k/2)$ for all $n \geq N_k = \text{Max}\,(N_{1k}, N_{2k})$ for $k = 1,2,\ldots$

Let k_o be such that $\varepsilon_k < \tau^2/4$ for all $k > k_o$

Let $t_n = \varepsilon_1$ for $n = 1,2,\ldots N_2$

$= \varepsilon_k$ for $n = N_k + 1, \ldots N_{k+1}$ for $k = 2,3,\ldots$

Let $E_n = (\sqrt{m}_{1n} + \sqrt{m}_{2n} \leqslant 2\sqrt{t_n}) \cap (S_{1n} \neq \emptyset) \cap (S_{2n} \neq \emptyset)$

$\Pr(E_n) \to 1$ as $n \to \infty$. So $\Pr(E_n^c) \to 0$.

$w \in E_n \Rightarrow G_{1n}(w) = H_{1n}(w)$ or $H_{2n}(w)$ and

$G_{2n}(w)$ is the other for $n \geq N_{k_o} + 1$

without loss of generality let $G_{1n}(w) = H_{1n}(w)$ for $n \geq N_{k_o} + 1$

given that $w \in E_n$.

$$E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right] = \int_{E_n} \left(\int_{R^2} \left|I_{\alpha_n^*} - I_\alpha\right| d\lambda\right) d P_\alpha + \int_{E_n^c} \lambda(\alpha_n^* \Delta \alpha)\, d P_\alpha .$$

But $\int_{E_n^c} \lambda(\alpha_n^* \Delta \alpha)\, d P_\alpha \leqslant P_\alpha(E_n^c)\, \lambda\left[\alpha \oplus D(a)\right]$ where 'a' is the diameter of $A \cup B$

So $\int_{E_n^c} \lambda(\alpha_n^* \Delta \alpha)\, d P_\alpha \to 0$ as $n \to \infty$ since $P_\alpha(E_n^c) \to 0$.

$$\int_{E_n} \left(\int_{R^2} \left|I_{\alpha_n^*} - I_\alpha\right| d\lambda\right) d P_\alpha$$

$$= \int_{E_n} \left(\int_{R^2} \left|I_{F_{1n}^* \cup F_{2n}^*} - I_\alpha\right| d\lambda\right) d P_\alpha$$

$$= \int_{E_n} \left(\int_{R^2} \left|I_{Y_{1n}} - I_A\right| d\lambda\right) d P_\alpha + \int_{E_n} \left(\int_{R^2} \left|I_{Y_{2n}} - I_B\right| d\lambda\right) d P_\alpha \text{ for } n > N_{k_o} + 1$$

$$\leqslant \int_\Omega \left(\int_{R^2} \left|I_{Y_{1n}} - I_A\right| d\lambda\right) d P_\alpha + \int_\Omega \left(\int_{R^2} \left|I_{Y_{2n}} - I_B\right| d\lambda\right) d P_\alpha \text{ for } n > N_{k_o} + 1.$$

By using the theorem 4.3.2 it can be proved that the above two integrals go towards zero as $n \rightarrow \infty$. Hence the theorem.

Similarly for any number k of compact regions given that k is known, classification and estimation may be simultaneously done using MST. The procedure is stated below.

(i) Draw MST of the n observations.

(ii) Find out the maximum (k-1) edges and delete them from the MST.

(iii) Let the lengths of the k MST's be ℓ_i and the number of nodes be n_i where $\sum_1^k n_i = n$.

Let $h_i = \sqrt{\ell_i/n_i}$.

(iv) Denote the MST's by H_i, $i = 1, \ldots k$.

Let $\alpha_i^* = \{x : d(x, H_i) \leq h_i,\ x \in R^2\}$ $i = 1, \ldots k$

Let $\alpha^* = \bigcup_{i=1}^{k} \alpha_i^*$.

Then α^* is consistent estimator of the union of k compact regions.

4.5 Further remarks

It was shown that the MST based estimator α_n^* satisfies $E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right] \rightarrow 0$ as $n \rightarrow \infty$. A property of MST and α_n^* has been stated below.

Theorem 4.5.1 Let $X_1, X_2, \ldots$ be independent and identically distributed random variables defined on $(\Omega, \mathcal{A}, P_\alpha)$ to R^2 inducing $Q_\alpha \in \mathcal{G}_\alpha$ where α is a compact region. Let $G_n(w)$ be MST generated by $X_1(w), \ldots X_n(w)$, $w \in \Omega$. Let $\ell_n(w)$ be the length of $G_n(w)$. Let $\alpha^*_n(w) = \{x \in R^2 : d(x, G_n(w)) \leq h_n(w)\}$ where $h_n(w) = [\ell_n(w)/n]^{1/2}$. Then

(i) $P_\alpha \{w : D(\alpha^*_n, \alpha) > \varepsilon\} \to 0$ for any $\varepsilon > 0$ and

(ii) $P_\alpha \{w : D(G_n, \alpha) > \varepsilon\} \to 0$ for any $\varepsilon > 0$.

Proof $D(\alpha^*_n, \alpha) \leq D(G_n, \alpha) + h_n$ and $h_n \to 0$ in probability. So it suffices to show that $D(G_n, \alpha) \to 0$ in probability to prove the above theorem.

$$D(G_n, \alpha) = \text{Max} \left\{ \sup_{x \in \alpha} d(x, G_n), \sup_{x \in G_n} d(x, \alpha) \right\}$$

$\sup_{x \in G_n} d(x, \alpha)$ is defined as u_n and it was shown that $u_n \to 0$ in probability (lemma 4.2.6). Therefore, it is sufficient to show that $\sup_{x \in \alpha} d(x, G_n) \to 0$ in probability.

α is compact. Let $\varepsilon > 0$. Cover α by squares of side $\varepsilon/20$. There is a finite sub cover, say $A_1, A_2, \ldots A_k$ such that $\bigcup_{i=1}^{k} A_i \supseteq \alpha$. It is easy to show that

$$P_\alpha \{w : \text{there exists } N_i \text{ such that } X_{N_i}(w) \in A_i \text{ for } i = 1, \ldots k\} = 1.$$

So $P_\alpha \left\{w : d(x, G_n(w)) \leq \frac{\varepsilon \sqrt{2}}{20}\right\} \to 1$ as $n \to \infty$ for any $x \in \alpha$

$$P_\alpha \left\{w : \sup_{x \in \alpha} d(x, G_n(w)) \leq \frac{\varepsilon\sqrt{2}}{20}\right\} \to 1 \text{ as } n \to \infty.$$

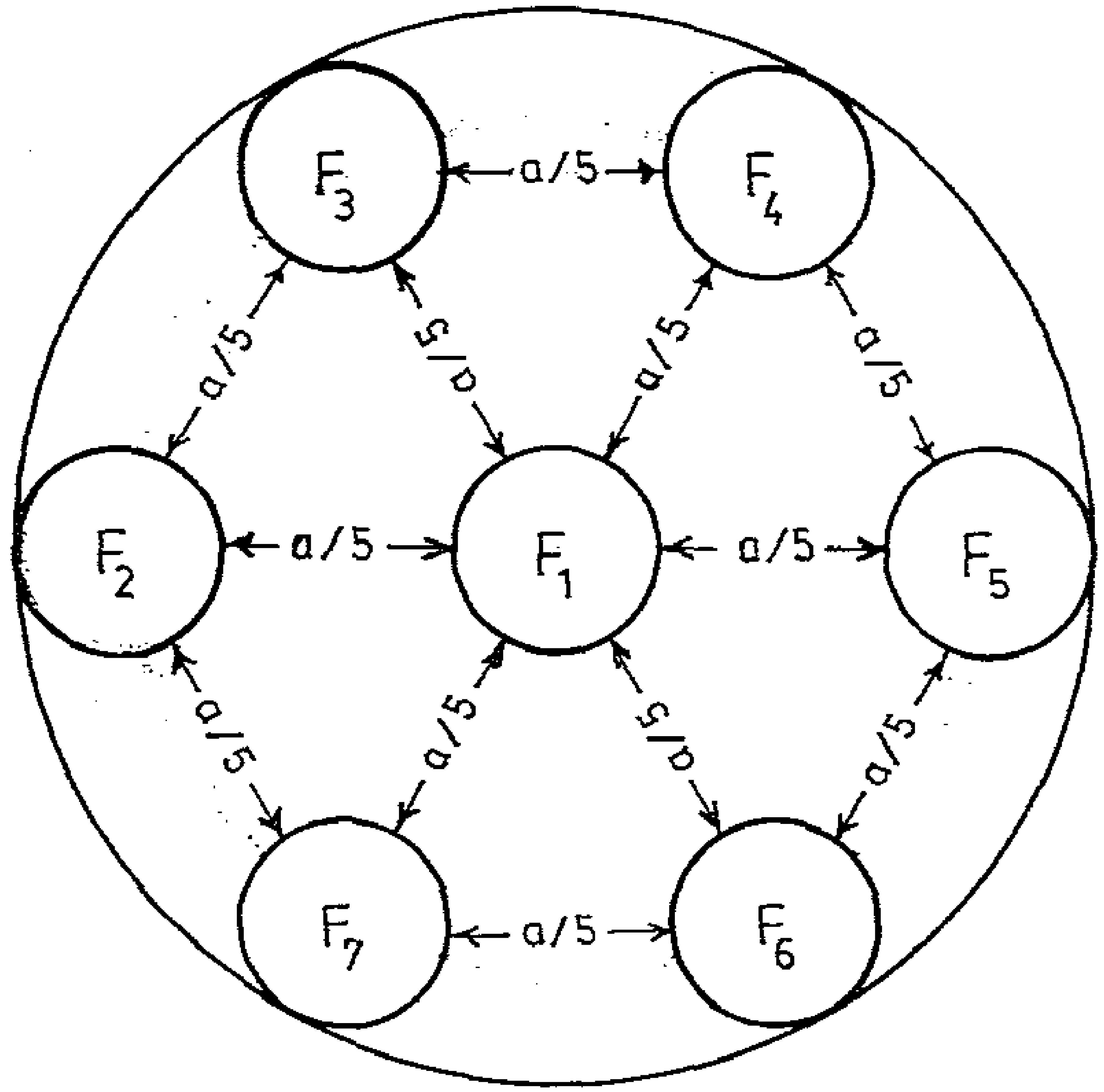
Since ε is arbitrary $\sup_{x \in \alpha} d(x, G_n(w)) \to 0$ in probability.

Hence the theorem.

In chapters V and VI, different definitions of consistency are given. The definition of consistency in higher dimensions has been discussed in Chapter VIII. Theorem 4.5.1 is helpful in that context.

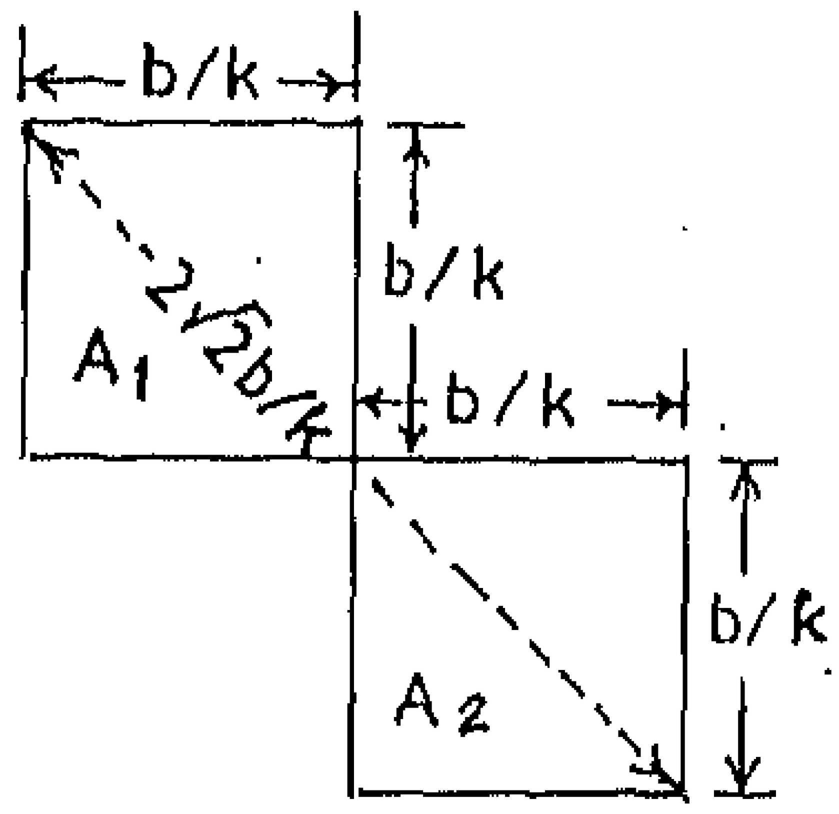
It has been shown that MST based method is useful in estimating compact regions. But in all the proofs, it is assumed that the classes are compact regions. In the next two chapters, MST based methods are developed for bounded line classes and for bounded mixture classes. In fact it will be shown that the same method can be applied in the case of bounded line classes as well as bounded mixture classes.

Experimental result using uniform distribution is given in Chapter VII. The material in this chapter is essentially taken from the papers of C.A. Murthy and D. Dutta Majumder [70,71] .



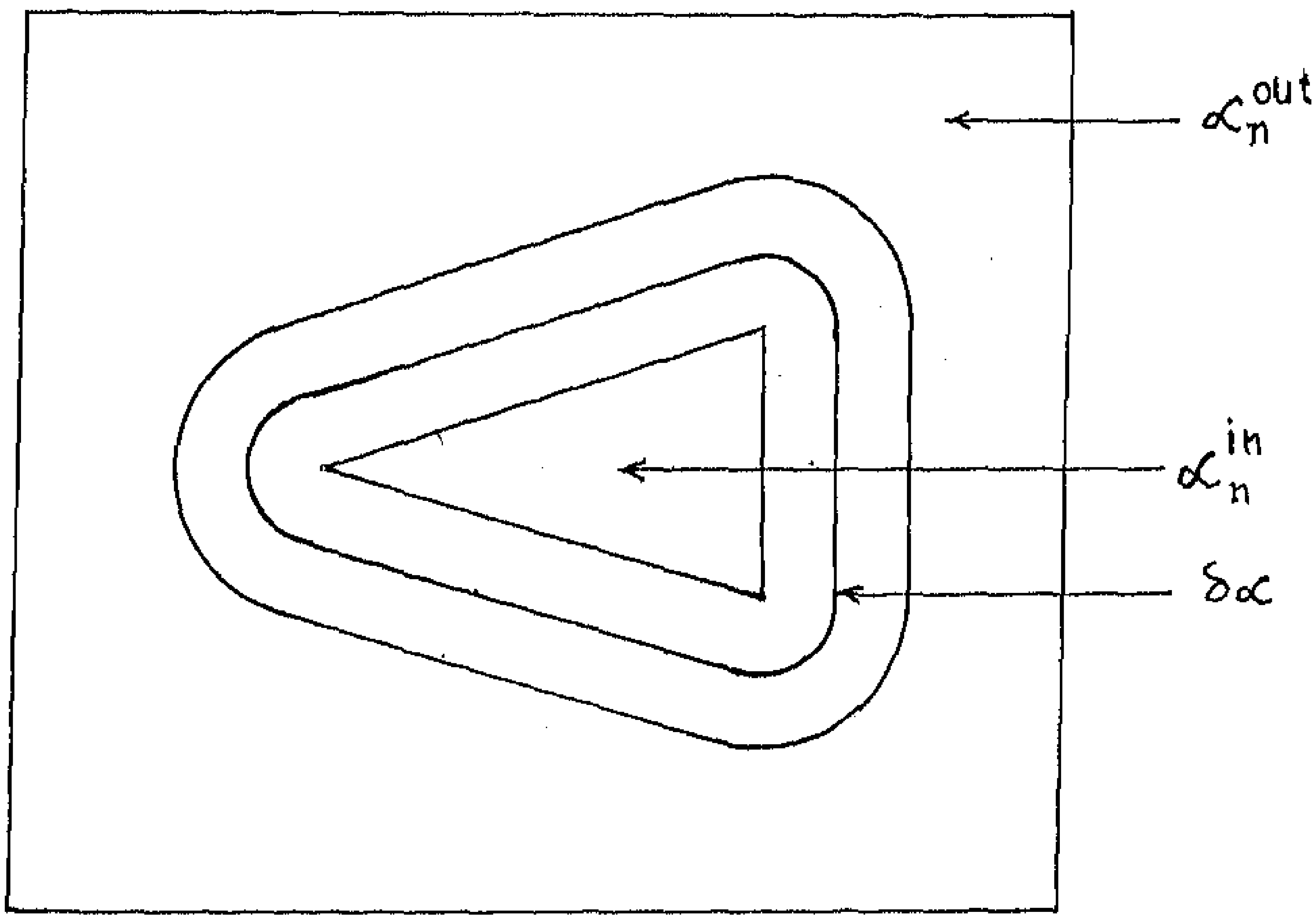
7 discs of diamater a/5 are shown to prove lemma 4.2.1.

Fig. 4.2.1



Two adjacent squares A_1, A_2 of side $\frac{b}{k}$ and shown $\underset{x \in A_1, y \in A_2}{\text{Sup}} d(x,y) = 2\sqrt{2}\, b/k$ [Lemma 4.2.2 and Lemma 4.2.6]

Fig 4.2.2



It shows α_n^{in}, α_n^{out} & $\delta\alpha$

Fig. 4.2.3

V. Consistent estimation of bounded line classes

5.1 Consistency

The definition of consistent estimator of α given in Chapter IV is the following.

$E_\alpha \left[\mu(\alpha_n^* \Delta \alpha)\right] \longrightarrow 0$ as $n \longrightarrow \infty$ where μ is a σ-finite measure and α_n^* is the estimator. μ is taken to be the lebesgue area for compact regions.

Bounded line classes have area zero. This creates a problem in the selection of the measure μ. We considered primarily two sets G_n and α_n^* where G_n is MST of the random vectors $X_1, \ldots X_n$ and $\alpha_n^* = \{x : d(G_n, x) \leq h_n\}$ where h_n is a function of $X_1, \ldots X_n$ and $h_n \longrightarrow 0$ in probability.

It is shown that

(i) $D(G_n, \alpha) \longrightarrow 0$ in probability - 5.1.1

and (ii) $E_\alpha \left[\lambda(\alpha_n^* \Delta \alpha)\right] \longrightarrow 0$ - 5.1.2.

5.1.1 Define consistency of G_n. 5.1.1 and 5.1.2 show the consistency of α_n^*. If it is known that α is a bounded line class then G_n can be used as an estimator for α. If it is not known that α is a bounded line class then α_n^* may be used as an estimator for α. It will be shown that α_n^* is same as α_{n2}^* defined in Chapter IV.

5.2 Estimation of bounded line classes

The proofs for consistency are given using the following propositions.

Proposition 5.2.1 Let $\alpha \subseteq R^2$ be a rectifiable Jordan arc with the arc length $L > 0$ and it joins y_0 and y_1. For every n, let $z_0, z_1, \ldots z_{2n} \in \alpha$ be such that the arc length of the curve joining z_i and z_{i+1}, $i=0,\ldots(2n-1)$ is $\frac{L}{2n}$ where $z_0 = y_0$, $z_{2n} = y_1$ and $z_i \neq z_j$ for $i \neq j$. Let $\mathcal{V}_{2i-1}$, $i = 1,\ldots n$ be open discs of radius $L/2n$ with centres at z_{2i-1}. Let

$$x_i \in \mathcal{V}_{2i-1} \cap \alpha, \ i=1,\ldots,n. \ \text{Let } H_n = \bigcup_{i=1}^{n-1} \left[x_i, x_{i+1}\right]. \ \text{Let } G_n \text{ be MST of } x_1,\ldots x_n.$$

Then $\lim\limits_{n\to\infty} \sup\limits_{x\in H_n} \inf\limits_{y\in\alpha} d(x,y) = 0$ and

$\lim\limits_{n\to\infty} \sup\limits_{y\in\alpha} \inf\limits_{x\in H_n} d(x,y) = 0$. Similarly $\lim\limits_{n\to\infty} D(G_n,\alpha) = 0$

<u>Proof</u>

$$\left.\begin{array}{l} \inf\limits_{x\in H_n} d(x,y) \leq \frac{L}{2n} \text{ for every } y\in\alpha . \\ \inf\limits_{y\in\alpha} d(x,y) \leq \frac{L}{2n} \text{ for every } x\in H_n . \end{array}\right\} \text{Fig. 5.2.1}$$

So $\lim\limits_{n\to\infty} \sup\limits_{y\in\alpha} \inf\limits_{x\in H_n} d(x,y) \leq \lim \frac{L}{2n} = 0$ and

$\lim\limits_{n\to\infty} \sup\limits_{x\in H_n} \inf\limits_{y\in\alpha} d(x,y) \leq \lim \frac{L}{2n} = 0$.

So $\lim\limits_{n\to\infty} D(H_n,\alpha) \to 0$ as $n \to \infty$ where D is Hausdorff metric.

Similarly $D(G_n,\alpha) \to 0$ as $n \to \infty$. Hence the proposition.

Proposition 5.2.2 Let $\alpha \subseteq R^2$ be a rectifiable Jordan curve with arc length $L > 0$. For every n take $z_0, \ldots z_{2n-1} \in \alpha$ such that $z_i \neq z_j$ for $i \neq j$ and the arc length of the curve joining z_i and z_{i+1} is $\frac{L}{2n}$ for $i = 0, \ldots (2n-1)$ where $z_{2n} = z_0$. Let ϑ_{2i-1}, $i=1, \ldots n$ be open discs of radius $\frac{L}{2n}$ around z_{2i-1}. Let $x_i \in \vartheta_{2i-1} \cap \alpha$ for $i = 1, \ldots n$.

Let $H_n = \bigcup_{i=1}^{n-1} [x_i, x_{i+1}]$. Then $D(H_n, \alpha) \longrightarrow 0$ as $n \longrightarrow \infty$.

Let G_n be MST of $x_1, \ldots x_n$.

Then $D(G_n, \alpha) \longrightarrow 0$.

Proof The proof is similar to the proposition 5.2.1.

Proposition 5.2.3 Let α be a bounded line class and is a subset of R^2. Let α also be a rectifiable Jordan arc joining y_0 and y_1 with arc length $L > 0$. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ where the induced measure be $Q_\alpha \in \mathcal{H}_\alpha$. Let $G_n(w)$ be MST generated by $X_1(w), \ldots X_n(w)$, $w \in \Omega$. Then

$$P_\alpha\{w : D(G_n(w), \alpha) > \varepsilon\} \longrightarrow 0 \text{ as } n \longrightarrow \infty \text{ for every } \varepsilon > 0 .$$

Proof $Q_\alpha \in \mathcal{H}_\alpha \Rightarrow \vartheta$ open in R^2, $\vartheta \cap \alpha \neq \emptyset \Rightarrow Q_\alpha(\vartheta \cap \alpha) \neq 0$.

Let k be a positive integer such that $\frac{L}{2k} < \varepsilon$. Find out $z_1, z_2, \ldots z_{2k+1} \in \alpha$ such that $z_i \neq z_j$ $\forall i \neq j$ and the arc length of the curve joining z_i and z_{i+1} is $L/2k$ for $i = 0, 1, \ldots (2k-1)$ where $z_0 = y_0$ and $z_{2k} = y_1$. Let ϑ_i's be open discs of radius $(L/2k)$ around z_{2i-1}, $i = 1, 2, \ldots k$. Then

$P_\alpha \{ w$: There exists $N_1, N_2, \ldots N_k$ such that

$$X_{N_i}(w) \in \mathcal{V}_i \cap \alpha \quad \text{for } i = 1, \ldots k \} = 1$$

$\Rightarrow P_\alpha \{ w : D(G_n(w), \alpha) > \frac{L}{2k} \} \longrightarrow 0$ as $n \longrightarrow \infty$ (From Proposition 5.2.1)

$\Rightarrow P_\alpha \{ w : D(G_n(w), \alpha) > \varepsilon \} \longrightarrow 0$ as $n \longrightarrow \infty$.

Proposition 5.2.4 Let α be a bounded line class and is a subset of R^2. Let α also be a rectifiable Jordan curve with arc length $L > 0$. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ to R^2 where the induced measure $Q_\alpha \in \mathcal{H}_\alpha$. Let $G_n(w)$ be MST generated by $X_1(w), \ldots X_n(w)$, $w \in \Omega$.

Then $P_\alpha \{ w : D(G_n(w), \alpha) > \varepsilon \} \longrightarrow 0$ for every $\varepsilon > 0$.

<u>Proof</u> Let k be a positive integer such that $\frac{L}{2k} < \varepsilon$. Take $z_0, z_1, \ldots z_{2k-1} \in \alpha$ such that $z_i \neq z_j$ for $i \neq j$ and the arc length of the curve joining z_i and z_{i+1} is $\frac{L}{2k}$ for $i = 0, \ldots (2k-1)$ where $z_{2k} = z_0$. Let $\mathcal{V}_i$, $i=1, \ldots k$ be open discs of radius $\frac{L}{2k}$ around z_{2i-1}.

Then $P_\alpha \{ w$: there exists $N_1, \ldots N_k$ such that

$$X_{N_i}(w) \in \mathcal{V}_i \cap \alpha \quad \text{for } i = 1, \ldots k \} = 1.$$

i.e. $P_\alpha \{ w : D(G_n(w), \alpha) > \frac{L}{2k} \} \longrightarrow 0$ as $n \longrightarrow \infty$

(From proposition 5.2.2)

$\Rightarrow P_\alpha \{ w : D(G_n(w), \alpha) > \varepsilon \} \longrightarrow 0$ as $n \longrightarrow \infty$.

Proposition 5.2.5 Let $\alpha \subseteq R^2$ be a bounded line class. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ to R^2 where the induced measure $Q_\alpha \in \mathcal{H}_\alpha$. Let $G_n(w)$ be MST generated by $X_1(w), \ldots X_n(w)$, $w \in \Omega$. Then $P_\alpha\left\{w : D(G_n(w), \alpha) > \varepsilon\right\} \to 0$ as $n \to \infty$ for any $\varepsilon > 0$.

<u>Proof</u> α is a bounded line class. So $\alpha = \bigcup_{i=1}^{k} \alpha_i$ where each α_i is a rectifiable Jordan arc or Jordan curve and α is path connected. [From definition 2.4.3] Let the arc length of each α_i be denoted as L_i and $L = \sum_{i=1}^{k} L_i$. Let M be a positive integer such that $L/2M < \varepsilon$. Get hold of discs (as described in propositions 5.2.3 and 5.2.4) $\mathcal{V}_{i1}, \mathcal{V}_{i2}, \ldots$ for each α_i, $i = 1, \ldots k$ so that the length of the curve in $\mathcal{V}_{ij} \cap \alpha$ is at least equal to $L/2M$, for $j = 1, \ldots \tau_i$ and for $i = 1, \ldots k$. Then $P_\alpha\{w$: there

exists N_{ij} such that $X_{N_{ij}}(w) \in \mathcal{V}_{ij} \cap \alpha$ for $j = 1, \ldots \tau_i$ and for $i = 1, \ldots k\} = 1$

So $P_\alpha\left\{w : D(G_n(w), \alpha) > L/2M\right\} \to 0$ as $n \to \infty$

$\Rightarrow P_\alpha\left\{w : D(G_n(w), \alpha) > \varepsilon\right\} \to 0$ for any $\varepsilon > 0$.

Proposition 5.2.6 Let α be a bounded line class and is a subset of R^2. Let α also be a rectifiable Jordan arc joining y_0 and y_1. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors on $(\Omega, \mathcal{A}, P_\alpha)$ to R^2 inducing $Q_\alpha \in \mathcal{H}_\alpha$. Let $G_n(w)$ be MST generated $X_1(w), X_2(w), \ldots X_n(w)$, $w \in \Omega$. Let $\ell_n(w)$ be length of $G_n(w)$. Let $h_n(w) = \left[\ell_n(w)/n\right]^{1/2}$

Let $\beta_n^*(w) = \left\{x : d(G_n(w), x) \leq h_n(w)\right\}$.

Then $E_\alpha\left[\lambda(\beta_n^* \Delta \alpha)\right] \to 0$ as $n \to \infty$.

<u>*Proof*</u> Let the arc length of α be L.

Then $\ell_n(w) \leq L$ for all n and w. So $h_n(w) \longrightarrow 0$ for every w as $n \longrightarrow \infty$

$\lambda(\beta_n^*) \leq \left[L + 2h_n(w)\right] 2h_n(w)$ for every n and w (fig.5.2.2)

$\leq (L + 2\sqrt{L})\ 2h_n(w)$.

Now $E_\alpha\left[\lambda(\beta_n^* \Delta\, \alpha)\right]$

$$= \int_\Omega \lambda(\beta_n^*(w)\ d\,P_\alpha(w)$$

$$\leq 2(L + 2\sqrt{L}) \int_\Omega h_n(w)\, d\,P_\alpha(w)$$

$$\leq 2(L + 2\sqrt{L}) \int_\Omega \left(\frac{L}{n}\right)^{1/2} d\,P_\alpha$$

$$= 2(L + 2\sqrt{L})\ \sqrt{L}\ /\ \sqrt{n} \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty .$$

Hence the proposition.

Proposition 5.2.7 Let $\alpha_1, \alpha_2 \subseteq R^2$ be two rectifiable Jordan arcs or curves such that $\alpha_1 \cap \alpha_2 = \{y_0\}$. Let $S_1 = \{x_{11}, \ldots x_{1n}\} \cup \{y_0\} \subseteq \alpha_1$ and $S_2 = \{x_{21}, \ldots x_{2m}\} \cup \{y_0\} \subseteq \alpha_2$. Let the arc length of α_i be L_i. Let G_i be MST of S_i, $i = 1,2$. Let G be MST of $S_1 \cup S_2$. Let the length of G_i be ℓ_i and the length of G be ℓ.

Then $\ell \leq \ell_1 + \ell_2 \leq L_1 + L_2$.

<u>*Proof*</u> Observe that $G_1 \cup G_2$ is a connected graph of $S_1 \cup S_2$. G is connected graph of $S_1 \cup S_2$ with minimum length. So $\ell \leq \ell_1 + \ell_2$.

Let $f : [0,1] \longrightarrow \alpha_1$ be such that f is one one on $[0,1)$ {If α is a Jordan arc then f is one one on $[0,1]$}, f is continuous and $f[0,1] = \alpha_1$.

Let $0 \leq a_1 < a_2 \ \ldots \ < a_n < a_{n+1} \leq 1$ be such that $f(a_i) \in S_1$ for $i = 1,..(n+1)$.

Let $H = \bigcup_{i=1}^{n} [f(a_i), f(a_{i+1})]$

$L_1 \geq$ length of H (follows from the definition of arc length)

$\geq \ell_1$ (follows from the definition of MST because H is a connected graph of S_1).

So $L_1 \geq \ell_1$. Similarly $L_2 \geq \ell_2$.

Hence the proposition.

Proposition 5.2.8 $\alpha_1, \alpha_2 \subseteq R^2$ be two rectifiable Jordan arcs or Jordan curves such that $\alpha_1 \cap \alpha_2 \neq \emptyset$. Let $S_1 = \{x_{11}, \ldots x_{1n}\} \subseteq \alpha_1$ and $S_2 = \{x_{21}, \ldots x_{2m}\} \subseteq \alpha_2$ be such that $S_1 \cap S_2 = \emptyset$. Let the arc length of α_i be L_i, $i = 1,2$. Let G_i be the MST of S_i, $i = 1,2$. Let G be the MST of $S_1 \cup S_2$ and the length of it be ℓ. Let $G_3 = G_1 \cup G_2 \cup [x_0, y_0]$ where $x_0 \in S_1$, $y_0 \in S_2$ and $d(x_0, y_0) = d(S_1, S_2)$. Let length of G_3 be ℓ_3.

Then $\ell \leq \ell_3 \leq L_1 + L_2$.

<u>Proof</u> $\ell \leq \ell_3$ is true because G_3 is a connected graph of $S_1 \cup S_2$. Let $f : [0,1] \rightarrow \alpha_1$ be continuous and one one on $[0,1)$ {If α_1 is a Jordan arc then f is one one on $[0,1]$ and $f[0,1] = \alpha$.}

Similarly define $g : [0,1] \rightarrow \alpha_2$.

Let $0 \leq a_1 < \ldots < a_n \leq 1$ be such that $f(a_i) \in S_1$ for $i = 1, \ldots n$.

Similarly $0 \leq b_1 < \ldots < b_m \leq 1$ be such that $g(b_i) \in S_2$ for $i = 1, \ldots m$.

Now there exists a path between $f(a_i)$ and $g(b_i)$ for every i and j.
Let the path with the shortest length be between $f(a_0)$ and $g(b_0)$ where $a_0 \in \{a_1,\ldots a_n\}$ and $b_0 \in \{b_1,\ldots b_m\}$.

Let $h : [0,1] \longrightarrow \alpha_1 \cup \alpha_2$ be continuous and $h(0) = f(a_0)$ and $h(1) = g(b_0)$. Let the length ofthe path be ℓ_5.

$\ell_5 \geq d(x_0,y_0)$.

length of $G_1 \leq \sum_{i=1}^{n-1} \| f(a_i) - f(a_{i+1}) \| \leq L_1 - [$lengths of the curves between 0 to a_1 and a_n to $1]$.

length of $G_2 \leq \sum_{i=1}^{m-1} \| g(b_i) - g(b_{i+1}) \| \leq L_2 - [$lengths of the curves between 0 to b_1 and b_m to $1]$.

So $\ell_3 \leq L_1 + L_2$.

Hence the proposition.

Proposition 5.2.9 Let $\alpha_1,\alpha_2,\alpha_3 \subseteq R^2$ be such that they are rectifiable Jordan arcs or Curves, $\alpha_1 \cap \alpha_2 \neq \emptyset$, $\alpha_2 \cap \alpha_3 \neq \emptyset$ but $\alpha_1 \cap \alpha_3 = \emptyset$. Let $S_i = \{x_{i1},\ldots x_{im_i}\} \subseteq \alpha_i$ for $i = 1,2,3$ such that $S_i \cap S_j = \emptyset$ for $i \neq j$.
Let $d(S_1,S_3) > d(S_1,S_2) > d(S_2,S_3)$. Let G_i be MST of S_i, $i = 1,2,3$ and ℓ_i's arethe corresponding length.

Let $G = G_1 \cup G_2 \cup G_3 \cup [x_0,y_0] \cup [x_1,y_1]$ where

$d(x_0,y_0) = d(S_1,S_3)$; $x_0 \in S_1$, $y_0 \in S_3$ and

$d(x_1,y_1) = d(S_1,S_2)$; $x_1 \in S_1$, $y_1 \in S_2$.

G is a spanning tree of $S_1 \cup S_2 \cup S_3$. Let the length of G be ℓ_4 and the arc length of α_i be L_i, $i = 1,2,3$.

Then $\ell_4 \leq \sum_{i=1}^{3} L_i$.

<u>Proof</u> Let f_i describe α_i for all i

Let $0 \leq a_1 < \dots < a_{m_1} \leq 1$ be such that $f_1(a_i) \in S_1$ for all i

$0 \leq b_1 < \dots < b_{m_2} \leq 1$ be such that $f_2(b_i) \in S_2$ for all i

and $0 \leq c_1 < \dots < c_{m_3} \leq 1$ be such that $f_3(c_i) \in S_3$ for all i.

Since $\alpha_1 \cap \alpha_2 \neq \emptyset$, there must be a path between every $f_1(a_i)$ and $f_2(b_j)$. Let the path between $f_1(a_0)$ and $f_2(b_0)$ be the shortest where $a_0 \in \{a_1, \dots a_m\}$ and $b_0 \in \{b_1, \dots b_{m_2}\}$ and the length of it be ℓ_5 .

Similarly, let the path between $f_2(b_0')$ and $f_3(c_0)$ be the shortest where $b_0' \in \{b_1, \dots b_{m_2}\}$ and $c_0 \in \{c_1, \dots c_{m_3}\}$ and the length of it be ℓ_6 .

$$e_1 = \sum_{i=1}^{m_1-1} \| f_1(a_i) - f_1(a_{i+1}) \| \geq \text{length of } G_1,$$

$$e_2 = \sum_{i=1}^{m_2-1} \| f_2(b_i) - f_2(b_{i+1}) \| \geq \text{length of } G_2,$$

$$e_3 = \sum_{i=1}^{m_3-1} \| f_3(c_i) - f_3(c_{i+1}) \| \geq \text{length of } G_3.$$

Now $\ell_6 \geq \|x_0 - y_0\|$ and $\ell_5 \geq \|x_1 - y_1\|$. [The length of the path between two points is greater than or equal to the distance between them.]

So $\ell_4 \leq e_1 + e_2 + e_3 + \ell_5 + \ell_6 \leq L_1 + L_2 + L_3$.

Hence the proposition.

The following proposition is a consequence of propositions 5.2.7, 5.2.8 and 5.2.9.

Proposition 5.2.10 Let $\alpha \subseteq R^2$ be a bounded line class. Let $\{x_1, x_2, \ldots x_n\} = S \subseteq \alpha$. Let G be MST of S. Let ℓ be the length of G. Let $\alpha = \bigcup_{i=1}^{k} \alpha_i$ where α_i's are rectifiable Jordan arcs or Jordan curves and α is path connected. Let the arc length of each α_i be L_i. Then $\ell \leq \sum_{i=1}^{k} L_i$.

<u>Proof</u> Follows from propositions 5.2.7, 5.2.8 and 5.2.9.

Theorem 5.2.1 Let $\alpha \subseteq R^2$ be a bounded line class. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ to R^2 and inducing $Q_\alpha \in \mathcal{H}_\alpha$. Let $G_n(w)$ be MST of $\{X_1(w), \ldots X_n(w)\}$, $w \in \Omega$ and $\ell_n(w)$ be its length.

Let $h_n(w) = (\ell_n(w)/n)^{1/2}$. Then

(i) $h_n(w) \to 0$ a.e.

(ii) $P_\alpha\{w : D(G_n(w), \alpha) > \varepsilon\} \to 0$ as $n \to \infty$ for any $\varepsilon > 0$.

(iii) $E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right] \to 0$ as $n \to \infty$ where

$$\alpha_n^* = \{x : d(G_n(w), x) \leq h_n\} .$$

<u>Proof</u> $\alpha = \bigcup_{i=1}^{k} \alpha_i$ where each α_i is a rectifiable Jordan arc or Jordan curve and α is path connected. Let the arc length of α_i be L_i. Let $L = \sum_{i=1}^{k} L_i$.

(i) Then $\ell_n(w) \leq L$ from proposition 5.2.10

So $h_n(w) \longrightarrow 0$ a.e.

(ii) $P_\alpha \{w : D(G_n(w),\alpha) > \varepsilon\} \longrightarrow 0$ for any $\varepsilon > 0$ has already been proved in proposition 5.2.5.

(iii) $\lambda(\alpha_n^*) \leq [L + 2kh_n(w)]\ 2h_n(w)$. [Fig.5.2.2.]

$\leq 2(L + 2k\sqrt{L})\sqrt{\frac{L}{n}}$.

So $E_\alpha [\lambda(\alpha_n^* \Delta \alpha)]$

$= E_\alpha \lambda(\alpha_n^*)$

$= \int_\Omega \lambda(\alpha_n^*)\ d\,P_\alpha \leq 2(L + 2k\sqrt{L}) \int_\Omega \sqrt{\frac{L}{n}}\ d\,P_\alpha$

$= [2(L + 2k\sqrt{L})\ \sqrt{L}]\ /\sqrt{n} \longrightarrow 0$ as $n \longrightarrow \infty$.

Hence the theorem.

5.3 Classification of bounded line classes.

In the previous section, a procedure for the estimation of bounded line classes is described. In this section, classification and estimation of bounded line classes using MST is described. A theorem is proved, the proof of which is similar to the one described in theorem 4.4.1.

Theorem 5.3.1 Let α_1, α_2 be two bounded line classes such that $d(\alpha_1,\alpha_2) = \tau > 0$. Let $Q_1 \in \mathcal{H}_{\alpha_1}$ and $Q_2 \in \mathcal{H}_{\alpha_2}$ and let $0 < q < 1$. Let $Q_\alpha = q\, Q_1 + (1-q)Q_2$ where $\alpha = \alpha_1 \cup \alpha_2$. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ inducing Q_α. Let $G_n(\omega)$ be MST of $\{X_1(\omega), \ldots X_n(\omega)\}$, $\omega \in \Omega$. Delete the edge with maximum edge from $G_n(\omega)$. Let the two resulting MST's be denoted by $G_{1n}(\omega)$ and $G_{2n}(\omega)$ with the respective number of points n_1 and n_2 where $n_1 + n_2 = n$.

Let $h_{1n}(\omega) = \left[\ell(G_{1n}(\omega))/n_1\right]^{1/2}$ and

$h_{2n}(\omega) = \left[\ell(G_{2n}(\omega))/n_2\right]^{1/2}$.

Let $F^*_{in}(\omega) = \{x : d(x, G_{in}(\omega)) \leq h_{in}(\omega)\}$ for $i = 1,2$.

Let $\alpha^*_n(\omega) = F^*_{1n}(\omega) \cup F^*_{2n}(\omega)$.

Then (i) $P_\alpha\{\omega : D((G_{1n}(\omega) \cup G_2(\omega)), \alpha) > \varepsilon\} \longrightarrow 0$

for any $\varepsilon > 0$ and

(ii) $E_\alpha\left[\lambda(\alpha^*_n \,\Delta\, \alpha)\right] \longrightarrow 0$ as $n \longrightarrow \infty$.

Proof Let $\alpha_1 = \bigcup_{i=1}^{k_1} \alpha_{1i}$ and $\alpha_2 = \bigcup_{i=1}^{k_2} \alpha_{2i}$ where each α_{1i} and α_{2i} is a rectifiable Jordan arc or Jordan curve and α_1 and α_2 are path connected. Let the arc length of each α_{ij} be a_{ij}. Let

$a_i = \sum_{j=1}^{k_i} a_{ij}$ for $i = 1,2$. Let $\varepsilon > 0$ and let M be a positive integer such that $L_1/M < \varepsilon$, $L_2/M < \varepsilon$ and $\frac{L_1 + L_2}{2M} < \tau/2$. Get hold of open discs $\mathcal{V}_{ij\xi}$ $\xi = 1, \ldots \nu_{ij}$, $j = 1, \ldots k_i$, $i = 1,2$ for every α_{ij} (as described in propositions 5.2.3 and 5.2.4) such that the length of the curve in $\mathcal{V}_{ij\xi} \cap \alpha_{ij}$ is at least equal to $L_i/2M$ for every ξ, j and i. Then $P_\alpha\{w :$ there exists positive integers $N_{ij\xi}$ such that $X_{N_{ij\xi}}(w) \in \mathcal{V}_{ij\xi} \cap \alpha$ for all i, j, $\xi\} = 1$ - [5.3.1]

Let $m_n(w)$ = maximal edge length of $G_n(w)$.

Let $S_{in}(w) = \{X_1(w), \ldots X_n(w)\} \cap \alpha_i$ for $i = 1,2$.

Let $H_{in}(w)$ = MST generated by $S_{in}(w)$ for $i = 1,2$

Let $Y_{in}(w) = \{x : d(x, H_{in}) \leq \beta_{in}(w)\}$ where

$$\beta_{in}(w) = \left[\frac{\ell(H_{in}(w))}{\# S_{in}(w)}\right]^{1/2} \quad \text{for } i = 1,2.$$

Let $m_{in}(w)$ = Maximal edge length of $H_{in}(w)$ for $w \in \Omega$.

Then $P_\alpha\left\{w : m_{in}(w) > \frac{L_i}{2M}\right\} \longrightarrow 0$ as $n \longrightarrow \infty$ for every i (follows from 5.3.1).

Observe that by taking discs of smaller radius it can be shown that $m_{in} \longrightarrow 0$ in probability.

Let $E_n = \{w : m_{1n}(w) + m_{2n}(w) \leq \tau/2\} \cap (S_{1n} \neq \phi) \cap (S_{2n} \neq \phi)$

$P(E_n) \longrightarrow 1$ because both m_{1n} and m_{2n} go towards zero in probability.

$w \in E_n \Rightarrow m_n \geq \tau$ because $d(\alpha_1, \alpha_2) = \tau$.

i.e., $w \in E_n \Rightarrow G_{in}(w) = H_{in}(w)$, $h_{in} = \beta_{in}$ and $F^*_{in} = Y_{in}$ (without loss of generality)

So $P_\alpha\{w : D((G_{1n}(w) \cup G_{2n}(w)), \alpha) > \varepsilon\}$

$= P_\alpha[\{w : D((G_{1n}(w) \cup G_{2n}(w)), \alpha) > \varepsilon\} \cap E_n] +$

$P_\alpha[\{w : D((G_{1n}(w) \cup G_{2n}(w)), \alpha) > \varepsilon\} \cap E_n^c] \longrightarrow 0 + 0 = 0$

[The first probability goes towards zero because of Theorem 5.2.1. The second probability goes towards zero because $P_\alpha(E_n^c) \longrightarrow 0$.]

Similarly

$$E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right]$$

$$= \int_\Omega \lambda(\alpha_n^*)\, d P_\alpha = \int_{E_n} \lambda(\alpha_n^*)\, d P_\alpha + \int_{E_n^c} \lambda(\alpha_n^*)\, d P_\alpha$$

$$= \int_{E_n} \lambda\left[(Y_{1n}(\omega)) \cup Y_{2n}(\omega)\right]\, d P_\alpha + \int_{E_n^c} \lambda(\alpha_n^*)\, d P_\alpha.$$

The second integral goes towards zero since $\lambda(\alpha_n^*)$ is bounded and $P_\alpha(E_n^c) \longrightarrow 0$.

$$\int_{E_n} \lambda(Y_{1n}(\omega) \cup Y_{2n}(\omega))\, d P_\alpha \leq \int_\Omega \lambda(Y_{1n}(\omega))\, d P_\alpha$$

$$+ \int_\Omega \lambda(Y_{2n}(\omega))\, d P_\alpha \longrightarrow 0 \text{ from theorem 5.2.1.}$$

Hence the theorem.

5.4 Remarks

Consistent estimator for α may be defined in another way also, as mentioned below.

Definition 5.4.1 A set α_n^* is said to be consistent to α if

(i) $P_\alpha\left\{\omega : D(\alpha_n^*, \alpha) > \varepsilon\right\} \longrightarrow 0$ for any $\varepsilon > 0$ and

(ii) $E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right] \longrightarrow 0$ as $n \longrightarrow \infty$.

It was shown in earlier sections that the MST based estimator α_n^* (as defined in earlier sections) satisfies $E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right] \longrightarrow 0$. We shall show that α_n^* also satisfies $P_\alpha\left\{\omega : D(\alpha_n^*, \alpha) > \varepsilon\right\} \longrightarrow 0$ for any $\varepsilon > 0$.

Theorem 5.4.1 Let $\alpha \subseteq R^2$ be a bounded line class. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ to R^2 and inducing $Q_\alpha \in \mathcal{H}_\alpha$. Let $G_n(w)$ be MST of $X_1(w), \ldots X_n(w)$, $w \in \Omega$ and $\ell_n(w)$ be its length. Let $h_n(w) = \left[\ell_n(w)/n\right]^{1/2}$ and

$$\alpha_n^*(w) = \left\{ x : d(G_n\ x) \leq h_n \right\};\ w \in \Omega .$$

Then for any $\varepsilon > 0$, $P_\alpha \left\{ w : D(\alpha_n^*, \alpha) > \varepsilon \right\} \to 0$.

<u>Proof</u> $D(\alpha_n^*, \alpha) \leq D(G_n, \alpha) + h_n$

$h_n \to 0$ in probability and $D(G_n, \alpha) \to 0$ in probability.

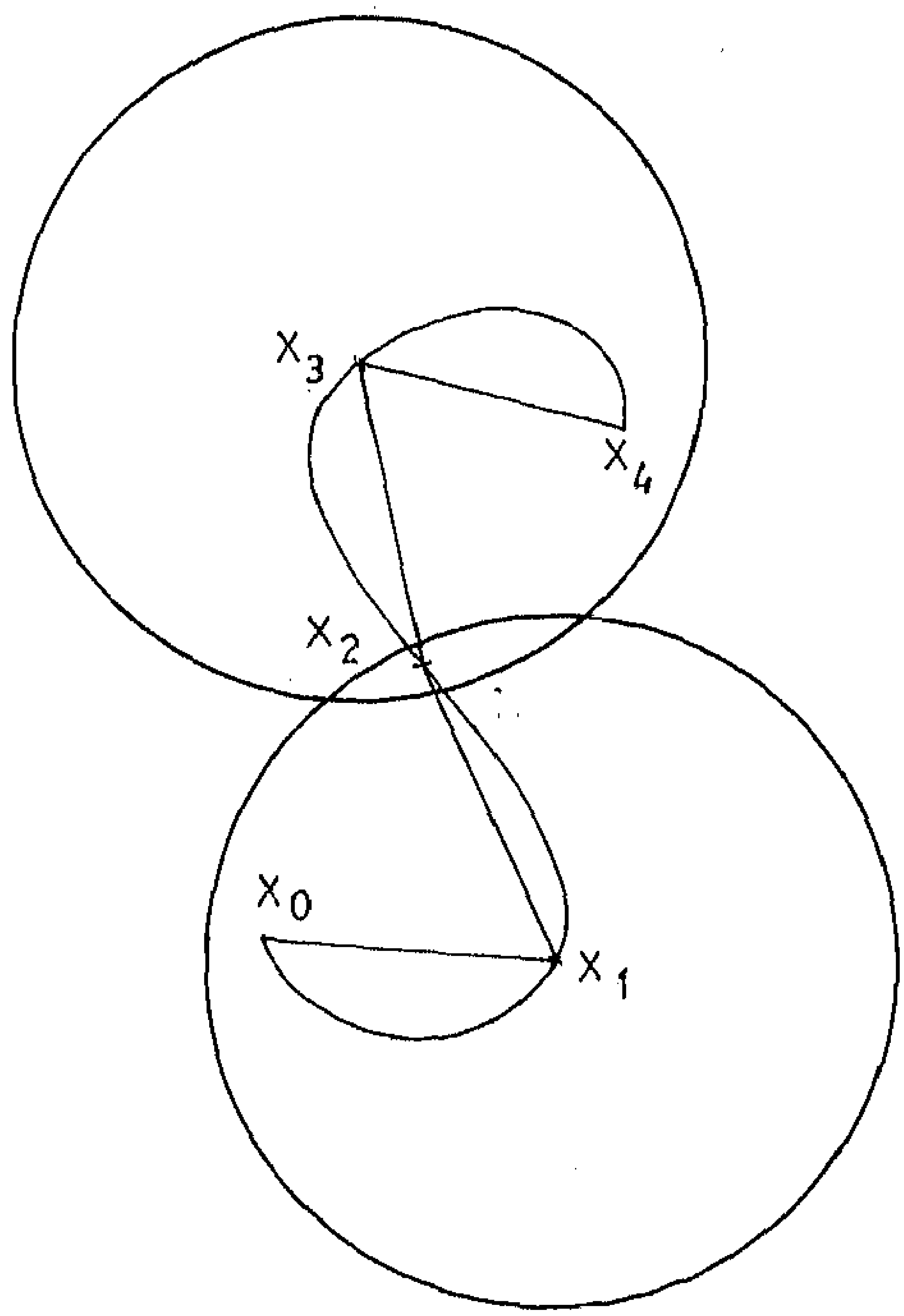
Hence the theorem.

α_n^* is consistent according to definition 5.4.1 as well as according to 5.1.1 and 5.1.2. Some properties of definition 5.4.1 are discussed in Chapter VI.

The estimators for bounded mixture classes are discussed in Chapter VI.

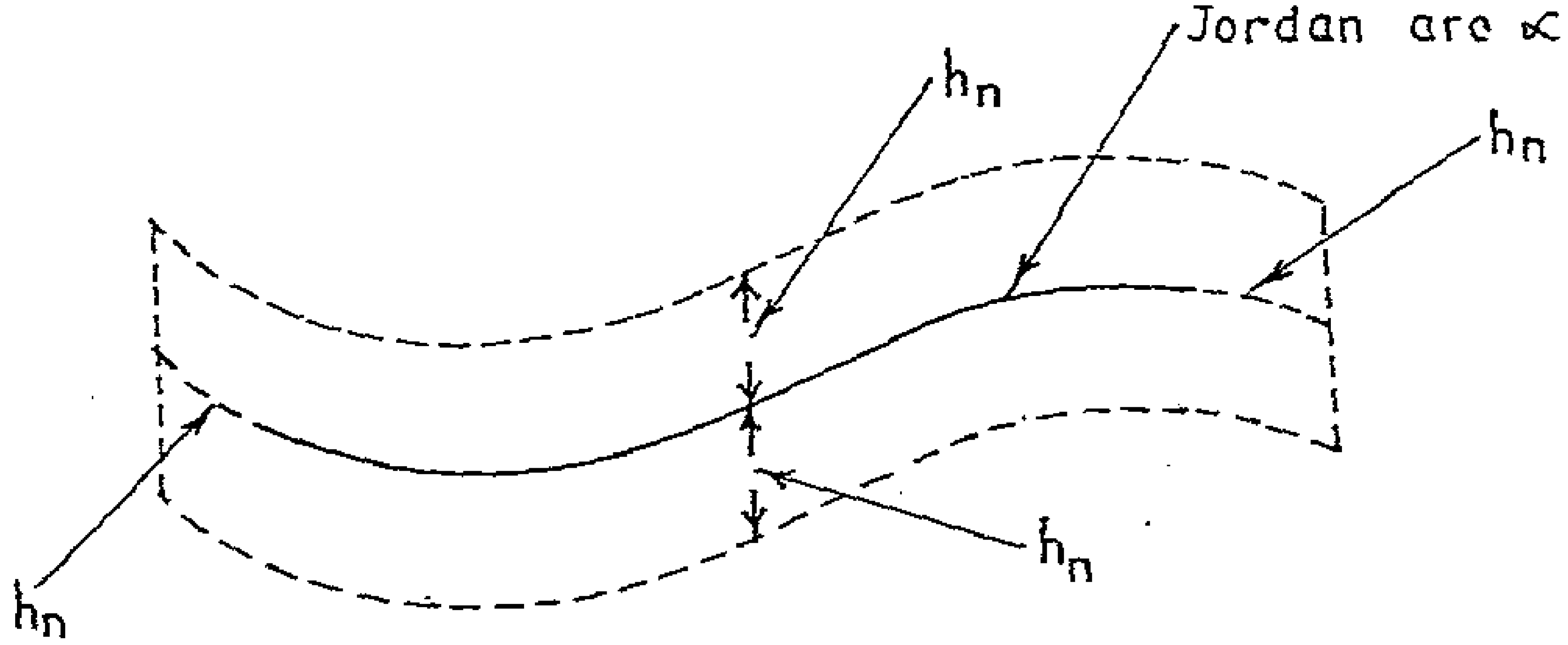
Observe that the concavity properties and C1 are not utilized in the estimation procedure. Only rectifiability is used in the estimation procedures. Indirectly, the concavity properties are used (Via C1) in the definition of the measure μ on bounded line classes. Even this condition - it will be stated in Chapter VIII - is unnecessary on the curves. Hausdorff measure can straightly be obtained on the bounded line classes.

Note that the MST itself may be taken as an estimate of bounded line class if the information is known apriori.



Two discs of radius $\frac{L}{2n}$ are shown at the points X_1 and X_3 [Proposition 5.2.1]

Fig.5.2.1



$\lambda\{x : d(\alpha, x) \leqslant h_n\} \leqslant [\text{length of } \alpha + 2h_n]\, 2h_n$

[Proposition 5.2.6]

Fig. 5.2.2

VI. Estimation of Bounded Mixture Classes

6.1 Definition of Consistency

A bounded mixture class is a union of compact regions and bounded line classes such that it is path connected [Def.2.5.1]. Probability measures on bounded mixture classes are convex combinations of probability measures on compact regions and bounded line classes.

In Chapter IV, it was defined that $\alpha_n^* \subseteq R^2$ is consistent to $\alpha \subseteq R^2$ if $E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right] \longrightarrow 0$ where λ is lebesgue area. If α is a bounded mixture class it would have a bounded line class whose area is zero. That necessitates a new definition for consistency. We shall consider two sets namely G_n, which is MST of $X_1, \ldots X_n$ and $\alpha_n^* = \left\{x : d(G_n, x) \leq h_n\right\}$ where $h_n = \left[\ell(G_n)/n\right]^{1/2}$. We shall show that

$$P_\alpha\left\{w : D(\alpha_n^*, \alpha) > \varepsilon\right\} \longrightarrow 0 \text{ as } n \longrightarrow \infty \text{ for any } \varepsilon > 0 \ldots \qquad 6.1.1$$

and $$E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right] \longrightarrow 0 \text{ as } n \longrightarrow \infty \quad \ldots \quad \ldots \qquad 6.1.2$$

i.e., any estimator α_n is said to be consistent to α if 6.1.1 and 6.1.2 are satisfied.

6.2 Consistent estimation of bounded mixture classes

A theorem was stated in Chapter IV (Theorem 4.1.1) for consistent estimation of compact sets of R^2 under the assumption of uniform distribution. Initially, the result is generalized for bounded mixture classes and for any distribution defined on it according to definition 2.5.2. Later, minimal spanning tree based estimator is proved to be consistent.

Theorem 6.2.1 Let $A \in \mathcal{E}$ and B be a rectifiable Jordan arc and also a bounded line class. Let $\alpha = A \cup B$ be a path connected set. Let Q be a probability measure on α such that $Q = qQ_1 + (1-q)Q_2$, $0 < q < 1$ where $Q_1 \in \mathcal{G}_A$ and $Q_2 \in \mathcal{H}_B$. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ and taking values in α such that P_α induces Q on α. Let $\varepsilon_n \downarrow 0$ and $n\,\varepsilon_n^2 \longrightarrow \infty$.

Let $\alpha_{n1}^*(w) = \bigcup_{i=1}^{n} \left\{ x : \|X_i(w) - x\| \leq \varepsilon_n \right\}$ for $w \in \Omega$.

Then $E_\alpha\left[\lambda(\alpha_{n1}^* \,\Delta\, \alpha)\right] \longrightarrow 0$ as $n \longrightarrow \infty$.

Proof $n\,\varepsilon_n^2 \longrightarrow 0$

$\Rightarrow n\,\varepsilon_n^2 > k$ for $n \geq N_k$ for $k = 1,2,\ldots$

Let $\theta_i = \varepsilon_1$ for $i = 1,\ldots N_2$

$= \varepsilon_k$ for $i = N_k+1,\ldots N_{k+1}$ for $k = 2,3,\ldots$

Then $n\,\varepsilon_n^2\,\theta_n \longrightarrow \infty$ as $n \longrightarrow \infty$ and $\theta_n \longrightarrow 0$.

Let $C_n = \left\{ x : f(x) \geq \theta_n \right\}$ where f is the density of Q_1 on A.

$\lambda(C_n^c \cap A) \longrightarrow 0$ as $n \longrightarrow \infty$.

Let $C_n^{in} = C_n \cap \left[C_n^c \oplus D(\varepsilon_n)\right]^c$ where $D(\varepsilon_n)$ is the closed disc of radius ε_n around origin and $\oplus$ represents Minkowski addition.

$$E_\alpha\left[\lambda(\alpha_{n1}^* \,\Delta\, \alpha)\right] = \int_{R^2} \left(\int_{\Omega} \left| I_{\alpha_{n1}^*} - I_\alpha \right| dP_\alpha \right) d\lambda.$$

Let $A_n^{in} = A \cap [A^c \oplus D(\varepsilon_n)]^c$

$B_n^{in} = B \cap [B^c \oplus D(\varepsilon_n)]^c$

$\alpha_n^{out} = [\alpha \oplus D(\varepsilon_n)]^c$

$E_\alpha [\lambda(\alpha_{n1}^* \Delta \alpha)]$

$$= \int_{A_n^{in} \cap C_n^{in}} \left(\int_\Omega \left| I_{\alpha_{n1}^*} - I_\alpha \right| d P_\alpha \right) d\lambda$$

$$+ \int_{A_n^{in} \cap (C_n^{in})^c} \left(\int_\Omega \left| I_{\alpha_{n1}^*} - I_\alpha \right| d P_\alpha \right) d\lambda$$

$$+ \int_{B_n^{in}} \left(\int_\Omega \left| I_{\alpha_{n1}^*} - I_\alpha \right| d P_\alpha \right) d\lambda$$

$$+ \int_{\alpha_n^{out}} \left(\int_\Omega \left| I_{\alpha_{n1}^*} - I_\alpha \right| d P_\alpha \right) d\lambda$$

$$+ \int_{(\alpha_n^{out} \cup A_n^{in} \cup B_n^{in})^c} \left(\int_\Omega \left| I_{\alpha_{n1}^*} - I_\alpha \right| d P_\alpha \right) d\lambda$$

$= J_{1n} + J_{2n} + J_{3n} + J_{4n} + J_{5n}$ where J_{in}'s are the corresponding integrals.

$(\alpha_n^{out} \cup A_n^{in} \cup B_n^{in})^c$

$= (\alpha_n^{out})^c \cap (A_n^{in})^c \cap (B_n^{in})^c$

$$= [\alpha \oplus D(\varepsilon_n)] \cap \{A^c \cup [A^c \oplus D(\varepsilon_n)]\} \cap \{B^c \cup [B^c \oplus D(\varepsilon_n)]\}.$$

$\longrightarrow \alpha \cap cl(A^c) \cap cl(B^c) = \delta\alpha.$ So $J_{5n} \longrightarrow 0.$

$x \in \alpha_n^{out} \Rightarrow \int_\Omega \left| I_{\alpha^*_{n1}} - I_\alpha \right| \, d P_\alpha = \int_\Omega I_{\alpha^*_n} \, d P_\alpha \longrightarrow 0$ as $n \longrightarrow \infty$

So $J_{4n} \longrightarrow 0$.

$\lambda(B_n^{in}) \longrightarrow 0$. So $J_{3n} \leqslant \lambda(B_n^{in}) \longrightarrow 0$.

$\lambda(A_n^{in} \cap (C_n^{in})^c) \leq \lambda\left[(C_n^{in})^c \cap A\right] \longrightarrow 0$ as $n \longrightarrow \infty$. So $J_{2n} \longrightarrow 0$.

$$J_{1n} = \int_{A_n^{in} \cap C_n^{in}} \left(\int_\Omega (1 - I_{\bigcup_{i=1}^{n} \{(x,\omega): \|X_i(\omega) - x\| \leq \varepsilon_n\}}) \, d P_\alpha \right) d\lambda$$

$$= \int_{A_n^{in} \cap C_n^{in}} \left(\int_\Omega I_{\bigcap_{i=1}^{n} \{(x,\omega): \|X_i(\omega) - x\| > \varepsilon_n\}} \, d P_\alpha \right) d\lambda$$

$$= \int_{A_n^{in} \cap C_n^{in}} \left[1 - P_\alpha(X_1 \in (x \oplus D(\varepsilon_n)))\right]^n d\lambda(x)$$

$$\leqslant \int_{A_n^{in} \cap C_n^{in}} \left[1 - \pi \varepsilon_n^2 \theta_n q\right]^n d\lambda(x) \text{ eventually. [From 2.3.c.4]}$$

$$(1 - \pi \varepsilon_n^2 \theta_n q)^n \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

So $J_{1n} \longrightarrow 0$. Hence the theorem.

The following theorem is a consequence of theorem 6.2.1.

Theorem 6.2.2 Let $\alpha \subseteq R^2$ be a bounded mixture class. Let Q be a probability measure on α, as defined in definition 2.5.2. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ such that P_α induces Q on α. Let $\varepsilon_n \downarrow 0$ and $n \varepsilon_n^2 \longrightarrow \infty$.

Let $\alpha^*_{n1}(w) = \bigcup_{i=1}^{n} \left\{ x : \|X_i(w) - x\| \leq \varepsilon_n \right\}$ for $w \in \Omega$.

Then $E_\alpha\left[\lambda(\alpha^*_{n1} \Delta \alpha)\right] \longrightarrow 0$ as $n \longrightarrow \infty$.

Proof It is a consequence of theorem 6.2.1.

The next theorem would show that minimal spanning tree based estimator is consistent to bounded mixture class.

Theorem 6.2.3 Let $\alpha = A \cup B$ where $A \in \mathcal{E}$, B is a bounded line class and α is path connected. Let $Q = qQ_1 + (1-q)Q_2$ where $0 < q < 1$ and $Q_1 \in \mathcal{G}_A$ and $Q_2 \in \mathcal{H}_B$. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ where P_α induces Q and α. Let G_n be MST of $X_1, \ldots X_n$.
Let $\ell_n(w)$ = length of $G_n(w)$, for $w \in \Omega$.
Let $h_n(w) = (\ell_n(w)/n)^{1/2}$.
Let $\alpha^*_n(w) = \left\{ x : d(G_n(w), x) \leq h_n \right\}$ for $w \in \Omega$.
Then (i) $E_\alpha\left[\lambda(\alpha^*_n \Delta \alpha)\right] \longrightarrow 0$ as $n \longrightarrow \infty$.

(ii) $P_\alpha\left\{ w : D(\alpha^*_n, \alpha) > \varepsilon \right\} \longrightarrow 0$ for any $\varepsilon > 0$.

Proof It will be proved using the following lemmas.

Lemma 6.2.1 (i) $h_n(w) \longrightarrow 0$ in probability.

(ii) $nh_n^2(w) \longrightarrow \infty$ in probability.

(iii) Let $u_n(w) = \sup \left\{ d(x, G_n(w)) : x \in \alpha \right\}$.

Then $u_n \longrightarrow 0$ in probability.

Proof The proofs are similar to the proofs of lemmas 4.2.1, 4.2.2, 4.2.3 and 4.2.6.

Lemma 6.2.2 $\mathcal{E}_\alpha \left[\lambda(\alpha_n^{*}\Delta\alpha)\right] \longrightarrow 0.$

Proof It is shown that $h_n(w) \longrightarrow 0$ in probability

$nh_n^2(w) \longrightarrow \infty$ in probability and

$u_n(w) \longrightarrow 0$ in probability.

Let ε_n be a sequence of numbers such that $1 > \varepsilon_n$ for all n, $\varepsilon_n \downarrow 0$.

Since $nh_n^2 \longrightarrow \infty$ in probability, there exists $N_{1k} > 0$ such that for all $n \geq N_{1k}$, $P\,(nh_n^2 \geq k) \geq 1 - \frac{\varepsilon_k}{4}$, for all $k = 1,2,3,\ldots$ Similarly since $h_n \longrightarrow 0$, there exists $N_{2k} > 0$ such that for all $n \geq N_{2k}$, $P_r(h_n \leqslant \varepsilon_k) \geq 1 - \frac{\varepsilon_k}{4}$ for all $k = 1,2,\ldots$ Similarly, there exists $N_{3k} > 0$ such that for all $n \geq N_{3k}$, $P_r(u_n \leqslant \varepsilon_k) \geq 1 - \frac{\varepsilon_k}{4}$ for all $k = 1,2,\ldots$

Let $N_k = \text{Max}\,(N_{1k}, N_{2k}, N_{3k})$, $k = 1,2,\ldots$

Let $s_i = 1$ for $i = 1,\ldots N_2$

$= k$ for $i = N_k + 1, \ldots N_{k+1}$ for $k = 2,3,\ldots$

Let $\gamma_n = (s_n/n)^{1/2}$.

Let $t_i = \varepsilon_1$ for $i = 1,\ldots N_2$

$= \varepsilon_k$ for $i = N_k + 1,\ldots N_{k+1}$ for $k = 2,3,\ldots$

It is clear that $s_n \longrightarrow \infty$ and $t_n \longrightarrow 0$.

$Pr(t_n > h_n > \gamma_n) \longrightarrow 1$ as $n \longrightarrow \infty$.

So $\gamma_n \longrightarrow 0$ since $t_n \longrightarrow 0$.

$Pr(nt_n^2 > nh_n^2 > n\gamma_n^2) \longrightarrow 1$ as $n \longrightarrow \infty$.

So $nt_n^2 \longrightarrow \infty$.

Let $D(2t_n)$ be closed disc of radius $2t_n$ around the origin.

$$\text{Let } \alpha_n^{in} = \alpha \cap \left[\alpha^c \oplus D(2t_n)\right]^c$$

$$\alpha_n^{out} = \left[\alpha \oplus D(2t_n)\right]^c$$

$$A_n^{in} = A \cap \left[A^c \oplus D(2t_n)\right]^c$$

$$B_n^{in} = B \cap \left[B^c \oplus D(2t_n)\right]^c$$

Let $I_{n2}(x,w) = 1$ if $x \in \alpha_n^*(w)$

$= 0$ otherwise

$I_1(x,w) = 1$ if $x \in \alpha$

$= 0$ otherwise.

Let $I_{(\cdot)}$ represent the indicator function of '$\cdot$'.

$$\text{Let } J_{3n} = \int_{\alpha_n^{out}} \left(\int_{\Omega} \left| I_{n2} - I_1 \right| d P_\alpha \right) d\lambda$$

$$= \int_{\alpha_n^{out}} \left(\int_{(h_n \leq t_n) \cap (u_n \leq t_n)} I_{\{(x,w):\, d(x, G_n(w)) \leq h_n\}} d P_\alpha \right) d\lambda$$

$$+ \int_{\alpha_n^{out}} \left(\int_{(h_n > t_n) \cap (u_n \leq t_n)} I_{\{(x,w):\, d(x, G_n(w)) \leq h_n\}} d P_\alpha \right) d\lambda$$

$$+ \int_{\alpha_n^{out}} \left(\int_{(h_n \le t_n) \cap (u_n > t_n)} I_{\{(x,w):\ d(x, G_n(w)) \le h_n\}} \, d P_\alpha \right) d\lambda$$

$$+ \int_{\alpha_n^{out}} \left(\int_{(h_n > t_n) \cap (u_n > t_n)} I_{\{(x,w):\ d(x, G_n(w)) \le h_n\}} \, d P_\alpha \right) d\lambda$$

$= J_{31n} + J_{32n} + J_{33n} + J_{34n}$ where J_{3in}'s are the respective integrals.

$$J_{31n} \le \int_{\alpha_n^{out}} \left(\int_{(h_n \le t_n) \cap (u_n \le t_n)} I_{\{(x,w)\,;\ d(x, G_n(w)) \le t_n\}} \, dP_\alpha \right) d\lambda$$

$$x_0 \in \alpha_n^{out} \Rightarrow I_{\{(x_0,w):\ d(x_0, G_n(w)) \le t_n\}} = 0$$

for $w \in (h_n \le t_n) \cap (u_n \le t_n)$

So $J_{31n} \to 0$ as $n \to \infty$.

$J_{32n} + J_{33n} + J_{34n}$

$$\le \int_{\alpha_n^{out}} \left(\int_{(h_n > t_n)} I_{\{(x,w)\,;\ d(x, G_n(w)) \le h_n(w)\}} \, d P_\alpha \right) d\lambda$$

$$+ \int_{\alpha_n^{out}} \left(\int_{(u_n > t_n)} I_{\{(x,w)\,:\ d(x, G_n(w)) \le h_n(w)\}} \, d P_\alpha \right) d\lambda$$

$$= \int_{[\alpha \oplus D(2a)]^c} \left(\int_{(h_n > t_n)} I_{n2} \, d P_\alpha \right) d\lambda \quad \left[\text{where } a = \sup_{x,y \in \alpha} d(x,y)\right]$$

$$+ \int_{\alpha_n^{out} \cap (\alpha \oplus D(2a))} \left(\int_{(h_n > t_n)} I_{n2} \, d P_\alpha \right) d\lambda$$

$$+ \int_{[\alpha \oplus D(2a)]^c} \left(\int_{(u_n > t_n)} I_{n2} \, dP_\alpha \right) d\lambda$$

$$+ \int_{[\alpha \oplus D(2a)] \cap \alpha_n^{out}} \left(\int_{(u_n > t_n)} I_{n2} \, dP_\alpha \right) d\lambda$$

$$= 0 + \lambda(\alpha \oplus D(2a)) \Pr(h_n > t_n) + 0 + \lambda(\alpha \oplus D(2a)) |$$

$\longrightarrow 0$ as $n \longrightarrow \infty$.

So $J_{3n} \longrightarrow 0$.

Let $$J_n = E_\alpha \left[\lambda(\alpha_n^* \Delta \alpha) \right] = \int_{\alpha_n^{in}} \left(\int_{(h_n > \gamma_n)} (1 - I_{n2}) \, dP_\alpha \right) d\lambda$$

$$= \int_{\alpha_n^{in}} \left(\int_{(h_n > \gamma_n)} (1 - I_{n2}) \, dP_\alpha \right) d\lambda + \int_{\alpha_n^{out}} \left(\int_\Omega I_{n2} \, dP_\alpha \right) d\lambda$$

$$+ \int_{(\alpha_n^{in} \cup \alpha_n^{out})^c} \left(\int_\Omega |I_1 - I_{n2}| \, dP_\alpha \right) d\lambda$$

$= J_{1n} + J_{2n} + J_{3n} + J_{4n}$ where J_{1n}, J_{2n}, J_{3n} and J_{4n} are the respective integrals.

$J_{4n} \leq \lambda(\alpha_n^{in} \cup \alpha_n^{out})^c \longrightarrow \lambda(\delta\alpha) = 0.$

$J_{3n} \longrightarrow 0$ as proved earlier.

$J_{2n} \leq \Pr(h_n > \gamma_n) \, \lambda(\alpha) \longrightarrow 0$ as $n \longrightarrow \infty$.

$$J_{1n} = \int_{\alpha_n^{in}} \left(\int_{(h_n > \gamma_n)} I_{\{(x,w) : d(x, G_n(w)) > h_n\}} \, dP_\alpha \right) d\lambda$$

$$\leq \int_{\alpha_n^{in}} \left(\int_{(h_n > \gamma_n)} I_{\{(x,w) : d(x, G_n(w)) > \gamma_n\}} \, dP_\alpha \right) d\lambda$$

$$\leq \int_{\alpha_n^{in}} \left(\int_{(h_n > \gamma_n)} I_{\bigcap_{i=1}^{n} \left\{ (x,\omega) : \| X_i(\omega) - x \| > \gamma_n \right\}} \, dP_\alpha \right) d\lambda$$

$$\leq \int_{\alpha_n^{in}} \left(\int_{\Omega} I_{\bigcap_{i=1}^{n} \left\{ (x,\omega) : \| X_i(\omega) - x \| > \gamma_n \right\}} \, dP_\alpha \right) d\lambda$$

$$= \int_{A_n^{in}} \left[\int_{\Omega} I_{\bigcap_{i=1}^{n} \left\{ (x,\omega) : \| X_i(\omega) - x \| > \gamma_n \right\}} \, dP_\alpha \right] d\lambda$$

$$+ \int_{B_n^{in}} \left[\int_{\Omega} I_{\bigcap_{i=1}^{n} \left\{ (x,\omega) : \| X_i(\omega) - x \| > \gamma_n \right\}} \, dP_\alpha \right] d\lambda$$

$$+ \int_{\alpha_n^{in} \cap (B_n^{in} \cup A_n^{in})^c} \left[\int_{\Omega} I_{\bigcap_{i=1}^{n} \left\{ (x,\omega) : \| X_i(\omega) - x \| > \gamma_n \right\}} \, dP_\alpha \right] d\lambda$$

$= J_{11n} + J_{12n} + J_{13n}$ where J_{11n} are the respective integrals.

From theorem 6.2.1 $J_{11n} \dashrightarrow 0$ and $J_{12n} \dashrightarrow 0$.

$J_{13n} \dashrightarrow 0$ because $\lambda\left[\alpha_n^{in} \cap (B_n^{in} \cup A_n^{in})^c\right] \dashrightarrow 0$.

So $J_{1n} \dashrightarrow 0$.

Lemma 6.2.3 $P_\alpha\left\{\omega : D(\alpha_n^*, \alpha) > \varepsilon\right\} \dashrightarrow 0$ for any $\varepsilon > 0$.

<u>Proof</u> Let $m_n(\omega)$ = Maximum edge length of $G_n(\omega)$.

It can be easily shown that $m_n(\omega) \dashrightarrow 0$ in probability.

(The proof is similar to lemma 4.2.2)

$d(x,\alpha) \leq h_n + \frac{m_n}{2}$ for $x \in \alpha_n^*$ [Fig.6.2.1]

i.e., $\sup_{x \in \alpha_n^*} d(x,\alpha) \leq h_n + \frac{m_n}{2}$

$\sup \{ d(x, G_n) : x \in \alpha \} = u_n$

So $\sup \{ d(x,\alpha_n^*) : x \in \alpha \} \leq u_n + h_n$

So $D(\alpha_n^*, \alpha) \leq \text{Max} \{ h_n + \frac{m_n}{2}, u_n + h_n \}$

But $h_n \to 0$ in probability, $m_n \to 0$ in probability and $u_n \to 0$ in probability.

So $P_\alpha \{ w : D(\alpha_n^*, \alpha) > \varepsilon \} \to 0$ as $n \to \infty$.

Hence the theorem.

Theorem 6.2.4 Let $\alpha \subseteq R^2$ be a bounded mixture class. Let Q be a probability measure defined on α (as defined in definition 2.5.2). Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ to R^2 such thatthey induce Q on α. Let $G_n(w)$ be MST generated by $X_1(w), \ldots X_n(w)$, $w \in \Omega$.

Let $\ell_n(w)$ be the length of $G_n(w)$.

Let $h_n(w) = [\ell_n(w)/n]^{1/2}$

Let $\alpha_n^*(w) = \{ x \in R^2 : d(x, G_n) \leq h_n \}$, $w \in \Omega$.

Then (i) $E_\alpha [\lambda(\alpha_n^* \Delta \alpha)] \to 0$ and

(ii) $P_\alpha \{ w : D(\alpha_n^*, \alpha) > \varepsilon \} \to 0$ for any $\varepsilon > 0$.

<u>Proof</u> A consequence of theorem 6.2.3.

From the results of Chapter IV and V and from the results of the theorems proved above, $\{x : d(x, G_n) \leq h_n\}$ is a consistent estimator for any bounded class, be it a compact region, bounded line class or bounded mixture class. In the next section, it will be shown that MST can be used for classification too.

6.3 Classification of bounded mixture classes.

In the earlier chapters, it was shown that deleting the edge with maximal length would provide classification in the case of two compact regions or bounded line classes. It will be shown here that the same holds good in the case of bounded mixture classes also.

Theorem 6.3.1 Let $A, B \subseteq R^2$ be two bounded mixture classes such that $d(A,B) = \tau > 0$. Let Q_1 and Q_2 be two probability measures on A and B respectively such that they follow the definition 2.5.2. Let $0 < q < 1$ and $Q = q\, Q_1 + (1-q)Q_2$. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ to R^2 such that they induce the measure Q on $\alpha = A \cup B$. Let $G_n(w)$ be MST generated by $X_1(w), \ldots X_n(w)$, $w \in \Omega$ and $m_n(w)$ be the maximal edge length of $G_n(w)$. Delete the maximal edge from MST. Let the two trees be denoted by $G_{1n}(w)$ and $G_{2n}(w)$ with the respective number of points n_1 and n_2 where $n_1 + n_2 = n$.

Let $h_{in}(w) = (\mathcal{L}(G_{in}(w))/n_i)^{1/2}$ for $i = 1,2$ and $w \in \Omega$,

$$F^*_{in}(\omega) = \{x : d(x, G_{in}(\omega)) \leq h_{in}(\omega)\}, \ \omega \in \Omega \quad \text{and}$$

$$\alpha_n^*(\omega) = F^*_{1n}(\omega) \cup F^*_{2n}(\omega).$$

Then (i) $E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right] \longrightarrow 0$ as $n \longrightarrow \infty$ and

(ii) $P_\alpha\{\omega : D(\alpha_n^*, \alpha) > \varepsilon\} \longrightarrow 0$ as $n \longrightarrow \infty$ for any $\varepsilon > 0$.

Proof The proof is similar to the proof of the theorem 4.4.1. The details are given below.

Cover $A \cup B$ by closed squares, the side of each being $\tau/20$. Only finitely many such squares are needed because $A \cup B$ is bounded. Let the squares be termed as $C_{11}, C_{12}, \ldots C_{1k_1}, C_{21}, \ldots C_{2k_2}$ such that

$$\left(\bigcup_{i=1}^{k_1} C_{1i}\right) \supseteq A, \quad \bigcup_{i=1}^{k_2} C_{2i} \supseteq B, \quad \left(\bigcup_{i=1}^{k_1} C_{1i}\right) \cap \left(\bigcup_{i=1}^{k_2} C_{2i}\right) = \emptyset,$$

$\mathrm{Int}(C_{1i}) \cap A \neq \emptyset$ for $i = 1, \ldots, k_1$ and $\mathrm{Int}(C_{2i}) \cap B \neq \emptyset$ for $i = 1, \ldots k_2$.

So $P(X_1 \notin C_{ji}, X_2 \notin C_{ji}, \ldots) = 0$ for $i = 1, \ldots k_j$, $j = 1, 2.$ – [6.3.1].

Let
$$S_{1n}(\omega) = A \cap \{X_1(\omega), \ldots X_n(\omega)\} \quad \text{and}$$
$$S_{2n}(\omega) = B \cap \{X_1(\omega), \ldots X_n(\omega)\} \quad \text{for } \omega \in \Omega.$$

Let
$$H_{in}(\omega) = \text{MST generated by } S_{in}(\omega) \quad \text{and}$$
$$Y_{in}(\omega) = \{x : d(x, H_{in}) \leq \beta_{in}(\omega)\} \quad \text{where}$$

$$\beta_{in}(\omega) = \left[\frac{\ell(H_{in}(\omega))}{\# S_{in}(\omega)}\right]^{1/2} \quad \text{for } i = 1, 2 \text{ and for } \omega \in \Omega.$$

Let $m_{in}(w)$ = Maximal edge length of $H_{in}(w)$ for $w \in \Omega$ and for $i = 1,2$.

Then $\Pr(m_{in} > (2\sqrt{2}\ \tau)/20) \longrightarrow 0$ as $n \longrightarrow \infty$ for $i = 1,2$. — [6.3.2]

(Follows from 6.3.1).

Observe that $\Pr(m_n \geq \tau) \longrightarrow 1$ as $n \longrightarrow \infty$ since $d(A,B) = \tau$.

Observe also that by taking squares of smaller side it can be shown that $m_{in} \longrightarrow 0$ in probability for $i = 1,2$. (follows from 6.3.2).

$m_{in} \longrightarrow 0$ in probability $\Rightarrow \sqrt{m_{in}} \longrightarrow 0$ in probability for $i = 1,2$.

$$\beta_{in}(w) = \left[\frac{\ell(H_{in}(w))}{\# S_{in}(w)}\right]^{1/2} \leq \left[\frac{(\# S_{in}(w))\ m_{in}(w)}{\# S_{in}(w)}\right]^{1/2} = \sqrt{m_{in}}(w)$$

for $i = 1,2$ and $w \in \Omega$.

So $\beta_{in}(w) \longrightarrow 0$ in probability for $i = 1,2$.

$\sqrt{m_{1n}} + \sqrt{m_{2n}} \longrightarrow 0$ in probability.

Let $1 > \varepsilon_k \downarrow 0$. Then $\Pr(m_{in} \leq \varepsilon_k) \geq 1 - \frac{\varepsilon_k}{4}$ for all $n \geq N_{ik}$,

$i = 1,2$ and for all k.

So $\Pr(\sqrt{m_{1n}} + \sqrt{m_{2n}} \leq 2\sqrt{\varepsilon_k}) \geq 1 - \varepsilon_k$ for

$n \geq N_k = \mathrm{Max}(N_{1k}, N_{2k})$ for $k = 1,2,\ldots$

Let K_o be such that $\varepsilon_k < \tau^2/4$ for all $k > k_o$.

Let $t_n = \varepsilon_1$ for $n = 1,2,\ldots N_2$

$= \varepsilon_k$ for $n = N_k + 1, \ldots N_{k+1}$ for $k = 2,3,\ldots$

and let $E_n = (\sqrt{m_{1n}} + \sqrt{m_{2n}} \leq 2\sqrt{t_n}) \cap (S_{1n} \neq \phi) \cap (S_{2n} \neq \phi)$

$\Pr(E_n) \longrightarrow 1$ as $n \longrightarrow \infty$. So $\Pr(E_n^c) \longrightarrow 0$ as $n \longrightarrow \infty$

$w \in E_n \Rightarrow G_{1n}(w) = H_{1n}(w)$ or $H_{2n}(w)$ and

$G_{2n}(w)$ is the other for $n \geq N_{k_o} + 1$.

without loss of generality let $G_{in}(w) = H_{in}(w)$ for $w \in E_n$.

$$E_\alpha\left[\lambda(\alpha_n^* \Delta \alpha)\right] = \int_{E_n} \left(\int_{R^2} \left| I_{\alpha_n^*} - I \right| d\lambda \right) d P_\alpha + \int_{E_n^c} \left(\int_{R^2} \left| I_{\alpha_n^*} - I \right| d\lambda \right) d P_\alpha .$$

But $\int_{E_n^c} \lambda(\alpha_n^* \Delta \alpha)\, d P_\alpha \longrightarrow 0$ since α and α_n^* are bounded and $P_\alpha(E_n^c) \longrightarrow 0$.

Now $\int_{E_n} \left(\int_{R^2} \left| I_{\alpha_n^*} - I_\alpha \right| d\lambda \right) d P_\alpha = \int_{E_n} \left(\int_{R^2} \left| I_{F_{1n}^* \cup F_{2n}^*} - I_\alpha \right| d\lambda \right) d P_\alpha$

$$= \int_{E_n} \left(\int_{R^2} \left| I_{Y_{1n}} - I_A \right| d\lambda \right) d P_\alpha + \int_{E_n} \left(\int_{R^2} \left| I_{Y_{2n}} - I_B \right| d\lambda \right) d P_\alpha$$

for $n > N_{k_o} + 1$

$$\leq \int_\Omega \left(\int_{R^2} \left| I_{Y_{1n}} - I_A \right| d\lambda \right) d P_\alpha + \int_\Omega \left(\int_{R^2} \left| I_{Y_{2n}} - I_B \right| d\lambda \right) d P_\alpha$$

for $n > N_{k_o} + 1$

From theorem 6.2.3 and subsequently from theorem 6.2.4 the above two integrals can be shown to be going towards zero.

(ii) Let $\varepsilon > 0$ and

$$\Theta_n = \left\{ w : D(\alpha_n^*, \alpha) > \varepsilon \right\}.$$

Then $P_\alpha(\theta_n) = P_\alpha(\theta_n \cap E_n) + P_\alpha(\theta_n \cap E_n^{\,c})$.

$$P_\alpha(\theta_n \cap E_n^{\,c}) \longrightarrow 0 \text{ since } P_\alpha(E_n^{\,c}) \longrightarrow 0$$

$$P_\alpha(\theta_n \cap E_n) \leq P_\alpha \left\{ w : D(Y_{1n}, A) + D(Y_{2n}, B) > \varepsilon \right\} \text{ for } n > N_{k_0} + 1$$

From theorem 6.2.3 and subsequently from theorem 6.2.4 it can be shown that $P_\alpha(\theta_n \cap E_n) \longrightarrow 0$.

Hence the theorem.

The statement for the generalization of theorem 6.3.1 to any number k of classes is stated below. The proof is not given since it follows from theorem 6.3.1.

Theorem 6.3.2 Let $\alpha = A_1 \cup \dots \cup A_k$ be a subset of R^2 such that each A_i is a bounded mixture class and $d(A_i, A_j) = \tau_{ij} > 0$ for $i \neq j$. Let Q_i be a probability measure on A_i (as in definition 2.5.2). Let $0 < q_i < 1$ be such that

$$\sum_1^k q_i = 1 \quad \text{and} \quad \text{let} \quad Q = \sum_1^k q_i Q_i \ .$$

Let $X_1, X_2, \dots$ be independent and identically distributed random variables, defined on $(\Omega, \mathcal{A}, P_\alpha)$ and inducing Q on α. Let $G_n(w)$ be MST of $X_1(w), \dots X_n(w)$. Find out the (k-1) edges with maximum length. Delete them from MST. Let the resulting trees be named $G_{in}(w)$, $i = 1, \dots k$ with the corresponding number of points n_i where $\Sigma\, n_i = n$. Let the length of $G_{in}(w)$ be $\ell_{in}(w)$, $i = 1, \dots k$.

Let $h_{in} = (\ell_{in}/n_i)^{1/2}$

$$F_{in}^{*}(w) = \{x : d(x, G_{in}) \leq h_{in}\} \quad \text{and}$$

$$\alpha_n^{*}(w) = \bigcup_{i=1}^{k} F_{in}^{*}(w)$$

Then (i) $E_{\alpha} \ \lambda(\alpha_n^{*} \Delta \alpha) \longrightarrow 0$ and

(ii) $P_{\alpha}\{w : D(\alpha_n^{*}, \alpha) > \varepsilon\} \longrightarrow 0$ for any $\varepsilon > 0$.

In the next section a few examples are stated which show the differences between

(i) $E_{\alpha}\left[\lambda(\alpha_n^{*} \Delta \alpha)\right] \longrightarrow 0$ as $n \longrightarrow \infty$ and

(ii) $P_{\alpha}\{w : D(\alpha_n^{*}, \alpha) > \varepsilon\} \longrightarrow 0$ for any $\varepsilon > 0$.

6.4 Note on the definition of consistency.

In section 6.1 it was defined that $\alpha_n^{*} \subseteq R^2$ is said to be consistent to $\alpha \subseteq R^2$ if

$$P_{\alpha}\{w : D(\alpha_n^{*}, \alpha) > \varepsilon\} \longrightarrow 0 \text{ for any } \varepsilon > 0 \quad - [6.1.1].$$

and $$E_{\alpha}\{\lambda(\alpha_n^{*} \Delta \alpha)\} \longrightarrow 0 \text{ as } n \longrightarrow \infty \quad - [6.1.2].$$

A few examples are given in this section to show that 6.1.1 ≠> 6.1.2 and 6.1.2 ≠> 6.1.1 .

Example 6.4.1 Let $\alpha = [0,1] \times [0,1]$ and the distribution on it be uniform. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables following uniform distribution on α.

Let $\alpha_{n1}^* = \{X_1(w), \ldots X_n(w)\}$

Then $P_\alpha\{w : D(\alpha_{n1}^*, \alpha) > \varepsilon\} \longrightarrow 0$ for any $\varepsilon > 0$ but

$E_\alpha[\lambda(\alpha_{n1}^* \Delta \alpha)] = \lambda(\alpha)$ for all n. Here α_n^* has area zero.

Instead let

$$\alpha_{n2}^* = \{X_1(w), \ldots X_{n-1}(w)\} \cup (X_n(w) \oplus D(\varepsilon_n)) \quad \text{where } \varepsilon_n \downarrow 0 .$$

Then $P_\alpha\{w : D(\alpha_{n2}^*, \alpha) > \varepsilon\} \longrightarrow 0$ but

$E_\alpha[\lambda(\alpha_{n2}^* \Delta \alpha)] \longrightarrow \lambda(\alpha)$ as $n \longrightarrow \infty$. α_{n2}^* will become a disconnected set as $n \longrightarrow \infty$. Even if, connectivity condition is imposed, 6.1.1 $\not\Rightarrow$ 6.1.2.

Let $\alpha_{n3}^* =$ MST of $\{X_1(w), \ldots X_n(w)\} \cup (X_n(w) \oplus D(\varepsilon_n))$ where $\varepsilon_n \downarrow 0$. Then $P_\alpha\{w : D(\alpha_{n3}^*, \alpha) > \varepsilon\} \longrightarrow 0$ but

$E_\alpha[\lambda(\alpha_{n3}^* \Delta \alpha)] \longrightarrow \lambda(\alpha)$ as $n \longrightarrow \infty$.

Example 6.4.2 Let $\alpha = [0,1] \times [0,1] \cup \{(2,y) : 0 \leq y \leq 1\}$.

Let Q_1 be uniform on $A = [0,1] \times [0,1]$ and Q_2 be uniform (according to the length) on $B = \{(2,y) : 0 \leq y \leq 1\}$ and let $Q = \frac{Q_1 + Q_2}{2}$. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors following Q on α.

Let G_n be MST of $\{X_1, \ldots X_n\}$ and ℓ_n be its length.

Let $h_n = (\ell_n/n)^{1/2}$ and $\alpha_n^* = \{x : d(x, G_n) \leq h_n\}$.

Then $E_{\alpha}\left[\lambda(\alpha_n^* \Delta \alpha)\right] \longrightarrow 0$ but

$$P_{\alpha}\left\{w : D(\alpha_n^*, \alpha) > 1/3\right\} \longrightarrow 1 \quad \text{as} \quad n \longrightarrow \infty .$$

Here α is a disconnected set. Even if α is a connected set, one can construct examples of above sort as given below.

Example 6.4.3 $\alpha = [0,1] \times [0,1]$ and $X_1, X_2, \ldots$ are independent and identically distributed following uniform distribution on α. Let G_n be MST of $X_1, \ldots X_n$ and ℓ_n be its length.

Let $h_n = (\ell_n/n)^{1/2}$ and

$$\alpha_n^* = \{(0,y) : -1 \leq y \leq 0\} \cup \{x : d(x, G_n) \leq h_n\}.$$

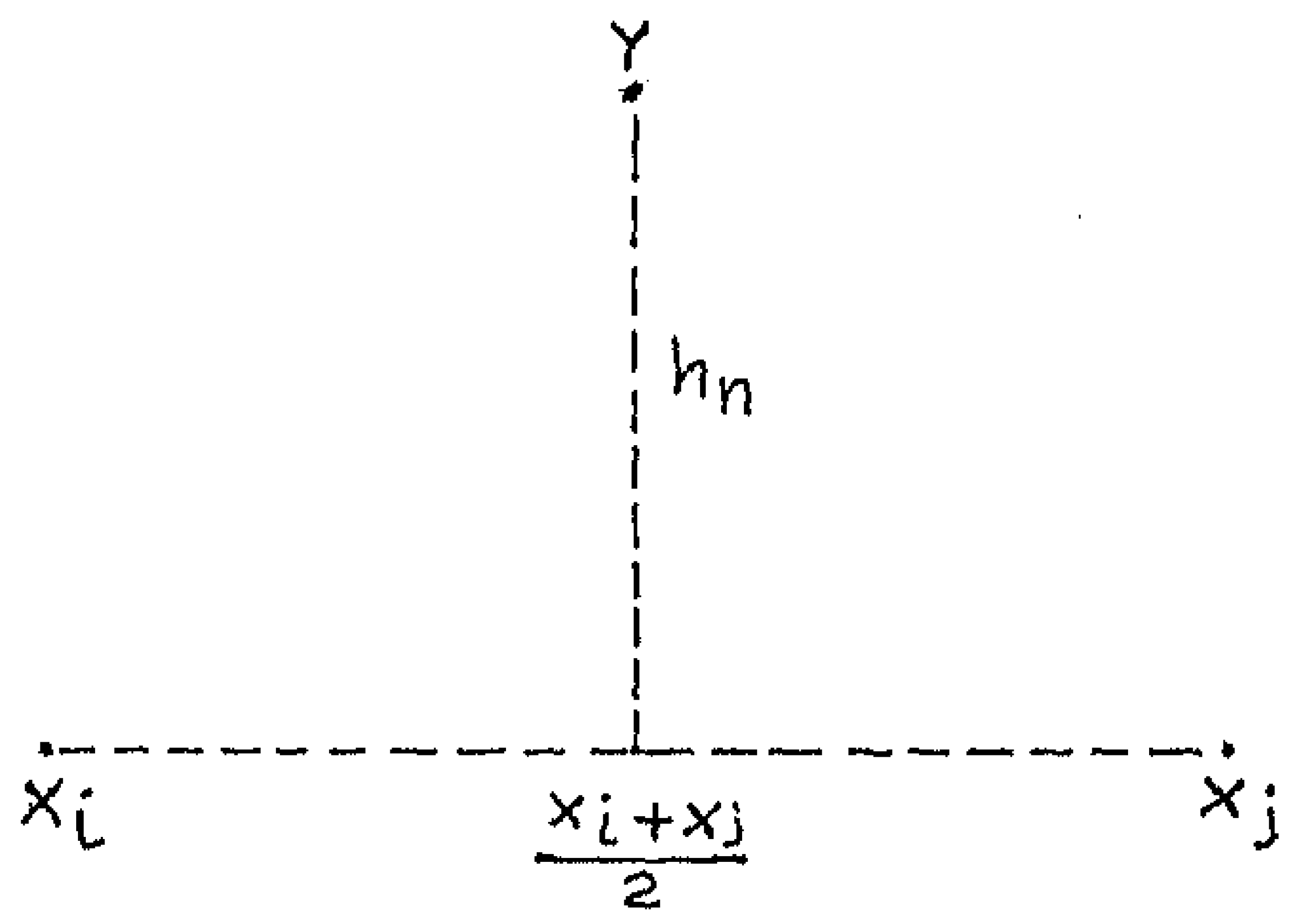
Then $E_{\alpha}\left[\lambda(\alpha_n^* \Delta \alpha)\right] \longrightarrow 0$ but

$$P_{\alpha}\left\{x : D(\alpha_n^*, \alpha) > 1/3\right\} \longrightarrow 1 \quad \text{as} \quad n \longrightarrow \infty .$$

From the above three examples, it is clear that 6.1.1 and 6.1.2 are both necessary for getting a reasonably good estimate of α.

6.5 Further remarks.

The results of Chapters IV, V and this chapter show that $\alpha_n^* = \{x : d(x, G_n) \leq h_n\}$ is consistent to any bounded class α where G_n is MST of the first n observations and $h_n = \left[\ell(G_n)/n\right]^{1/2}$. The generalization of this result to n dimensions ~~and relaxation boundedness criterion are~~ is discussed in Chapter VIII. MST has been used for classification and estimation of classes. The conditions imposed on probability measures are not stringent. But this method fails to classify the observations if the classes are overlapping.



Let $d(x_i, x_j) = m_n$. Let $d(\frac{x_i+x_j}{2}, Y) = h_n$.

Then $d(Y, x_i) \leq h_n + \frac{m_n}{2}$ [Lemma 6.2.3]

Fig. 6.2.1

VII. Computer Implementation

7.1 Introduction

In Chapters IV, V and VI it was shown that MST based estimator is consistent for the classes. In Chapter III the computational complexity in obtaining MST is discussed. The main concern of this chapter is the implementation of this method on computer. Efficient and faster implementation is not discussed here because a computer cannot draw discs around uncountably many points. The algorithm given in this chapter closely approximates the MST based estimator. The algorithm is applied on a data consisting of 30 points. It is a uniformly distributed data on [0,1] X [0,1] taken from random number tables [72]. The problems involved in obtaining the algorithm are discussed.

7.2 Problems in the implementation

The main problem in the implementation of finding the estimator using the computer is to find out a way to draw discs around every point on the spanning tree. Observe that (Fig.7.2.1) drawing discs around every point on a line segment amounts to drawing parallel lines on both sides of the line segment with semi circles at the end of the line segment. A line segment can be represented by the coordinates of the two points between which the segment is drawn. On every semi circle 3 points (A, C, E) are taken (Fig.7.2.1) such that the distance between any two consecutive points is same and A and E are on two different parallel line segments and the curve between the consecutive points is approximated by a straight line.

If two edges (AB,BC) are connected in the way shown in fig.7.2.2 then on one side of B a curve is to be drawn and on the other side legitimate points on the parallel line should not be included since those points fall within the radius of the other line . Three points on the curve DE are taken such that D and E fall on two different parallel lines for connected edges on the same side and the distance between D and Q is same as that of Q and E (Fig.7.2.2).

In order to delete points which fall within the radius of a point on another line segment, A_1, A_2, A_3 and B_1, B_2, B_3 are chosen on the line segments (Fig.7.2.2) such that $d(A,A_1) = d(A_1,A_2) = d(A_2,A_3) = d(A_3,B)$ and similarly $d(B,B_1) = d(B_1,B_2) = d(B_2,B_3) = d(B_3,C)$. Similarly on the parallel line segments for $AB(X_1X_5)$ and $BC(Y_1Y_5)$ points are taken (X_2,X_3,X_4 and Y_2,Y_3,Y_4) such that $d(X_i,X_{i+1}) = d(X_j,X_{j+1})$ for $i,j = 1,2,3,4$ and $d(Y_i,Y_{i+1}) = d(Y_j,Y_{j+1})$ for $i,j = 1,2,3,4$. Distances are calculated from every X_i to every one of A_j and B_j. Only those X_i's whose distance is $\geq$ the radius for every A_j and B_j are accepted. Similarly for Y_i's also. This process is to be repeated for every one of the points chosen on every parallel line segment and for every one of the points chosen on the MST. The choice is always the same. First and last points on the line segment and the three middle points . Naturally, the more the number of points chosen on the parallel line segments and edges of the spanning tree, the more the accuracy.

The natural question now is, how to join the points? In what order the points are to be joined? It is explained below using the figure 7.2.3. In that figure 30 points are shown from $[0,10] \times [0,10]$. MST is shown. After finding MST, a node with degree one is to be chosen (say it is 7). In order to know which parallel line is to be joined to which parallel line, the nodes are to be ordered. The ordering of the nodes for the MST is shown below.

7, 10, 24, 16, 23, 16, 27, 1, 25, 18, 3, 20, 9, 15, 13, 15, 9, 20, 3, 4, 11, 4, 3, 22, 14, 28, 14, 22, 3, 18, 25, 1, 27, 16, 24, 10, 26, 19, 2, 5, 30, 12, 21, 6, 29, 8, 29, 6, 21, 12, 30, 5, 2, 17, 2, 19, 26, 10, 7.

From the ordering itself, it can be seen how the boundary can be drawn. The ordering is done in the following way.

In the figure node 7 is connected to node 10. So the second node is naturally 10. If 10 is connected to exactly one node (say i) other than 7 then the next node is i. If not, find out the nodes which are connected to 10. In the figure the nodes are 24 and 26. After ascertaining them to be 24 and 26, make a linear transformation on 7,10,24 and 26 such that node 10 will become (0,0). i.e., subtract the x and y coordinates of node 10 from the x and y coordinates of nodes 7,10,24,26. Call the transformed nodes of 7,10,24 and 26 to be A,B,C and D respectively. Observe the quadrant in which the line segment AB is lying. Then the algorithm proceeds in the following way.

(i) AB is on the negative y axis.

Choose that transformed point which falls in 4th quadrant. If both points are in 4th quadrant then choose that one with minimum y coordinate. [y coordinates of the points cannot be the same]. If no point falls in 4th quadrant, choose that point with minimum y coordinate and which falls in 1st quadrant. If no point falls in 4th and 1st quadrant, choose that point with maximum y coordinate.

(ii) AB is in 4th quadrant.

If there exists a transformed point which is in between the straight line AB (not the line segment AB) and the positive x axis choose that one. For more than one point satisfying the above condition, choose that one with minimum y coordinate. If there exists no point in 4th quadrant satisfying that condition then choose the point in 1st quadrant with minimum y coordinate. If there exists no point in 1st quadrant too then choose the point in 2nd or 3rd quadrants with maximum y coordinate. If there exists no such point in 2nd and 3rd quadrants, choose that point in 4th quadrant whose x coordinate is minimum.

The other 6 cases can be dealt with similarly. The basic idea is to move the point A with centre at B in anti-clock wise direction and choosing that transformed point whose edge joined to B meets this revolving line first.

After choosing the node 24, the next node to be chosen is 16. Node 16 is joined to two nodes, 23 and 27. To choose one from them, 24 is to be transformed to A and 16 is to be transformed to B where A and B are as mentioned above. So node 23 will be chosen. But node 23 has degree one. The next node will be node 16 itself. Now 16 is joined to two nodes 27 and 24 and one from them is to be chosen. 23 is to be transformed to A and 16 is to be transformed to B. This is the way the ordering proceeds. It is to be stoppedwhen the initial node and the final node are the same.

While ordering the nodes, the points on the parallel line segments are to be found out simultaneously. That is, find out the five points on the parallel line segment in the transformed coordinate system of 7 and 10 (7 is A and 10 is B) and again transform them and store. The same is to be done with 10 and 24. And so on. If the angle between the line segments 7 and 10 and 10 and 24 is greater than 180° (The angle can be found outwhen node is being chosen) then one point is to be chosen on the curve. (as shown in 7.2.2). This is to be done for every edge. If there is a node with degree one some points on the semicircle (as shown in 7.2.1) are to be chosen. The first set of output is[call it F] the set of the points on the parallel line segments and curves and they are ordered.

In the next step find out points on every edge of MST as shown in Fig.7.2.2. Call the set to be H.

In the third step, reject all the points in F, whose distance from at least one point in H is less than the radius. Let the remaining set be F_1.

At step 4, take the first point in F_1 (say x_1). Let the second point be x_2. Find out x_{11}, x_{12} and x_{13} such that they fall on the line segment joining x_1 and x_2 and $d(x_1,x_{11}) = d(x_{11},x_{12}) = d(x_{12},x_{13}) = d(x_{13},x_2)$. Check whether $d(x_{1i},H) < \eta$ = (the radius x ε) for i = 1,2,3 where $1 > \varepsilon > 0$ and ε is close to one. ε is to be chosen suitably. While approximating semi circles and arcs (Figures 7.2.1 and 7.2.2), observe that the approximated line segments fall within the disc. In order to draw the boundary, the threshold value needs to be reduced. We have chosen ε to be 0.95 in fig. 7.2.3 and 0.9 in Fig. 7.3.1.

If $d(x_{1i},H) < \eta$ for at least one i then x_1 should not be joined to x_2. Go to the next point x_3 and check whether x_1 is to be joined to it or not. If x_1 is to be joined to a point x_j then check whether the next point x_{j+1} can be joined to x_j (not x_1). If all the points are exhausted, then check whether the last point x_k [say x_1 is joined to x_{j_1}, x_{j_1} to x_{j_2}, ... x_{j_m} to x_k and x_k is not joined to any other point] then check whether x_k can be joined to x_1 or not. If x_k cannot be joined to x_1 then find out $x_m = \frac{x_k + x_1}{2}$. Choose y_1,y_2,y_3,y_5,y_6 and y_7 as shown in figure 7.2.4. Choose y_1,y_2 and y_3 if $\text{Min}\{d(H,y_1), d(H,y_2), d(H,y_3)\} > \text{Min}\{d(H,y_5), d(H,y_6), d(H,y_7)\}$ otherwise choose y_5,y_6,y_7. If the minimums are same then choose (y_1,y_2,y_3) if $\text{Max}\{d(H,y_1), d(H,y_2), d(H,y_3)\} > \text{Max}\{d(H,y_5),$

$d(H,y_6)$, $d(H,y_7)\}$. Otherwise (y_5,y_6,y_7). If the maximums are also same then choose any one of the triplets, without loss of generality let the triplet be y_1,y_2,y_3. Then join x_k with y_1, y_1 to y_2, y_2 to y_3 and y_3 to x_1. This is a brute force way of joining x_1 with x_k.

At step 5, from the remaining points in the set F_1, repeat the above step until all the points in F_1 are exhausted. If there exists a point in F_1 which cannot be connected to any other point then it is to be assumed that some information was lost in the approximation. Similarly for the case of z_1 and z_2 where z_1 is joined to z_2 and they cannot be joined to any other point. Only if the number points in a chain is greater than or equal to 3, the brute force method of joining the first and last points in the chain, described in step 4 is to be applied. It is not guaranteed that the method of joining first and last points in a chain would yield good results. It is also a compromise. The output of the algorithm is a collection of sets of points, the points in each set are ordered.

The result will be more accurate if more number of points are chosen on circles and line segments.

7.3 Results

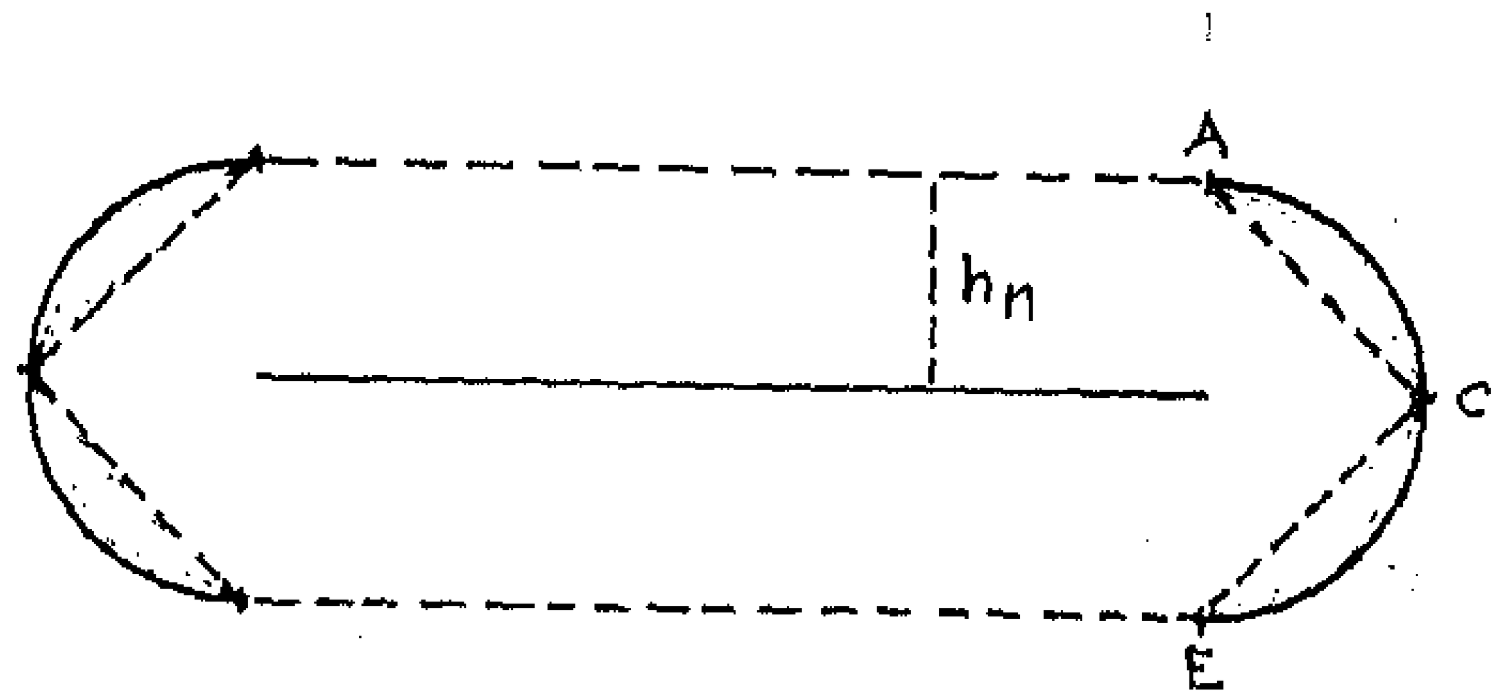
30 points from $[0,1] \times [0,1]$ are chosen randomly. In fact the data is taken from random number tables [72]. It is known here that the set is $[0,1) \times [0,1]$. Figure 7.3.1 shows the results. The MST as well

as the estimated set are shown. The algorithm, described in the previous section was applied.

In figure 7.2.3, the same points are multiplied by 10 and plotted. i.e. Now the points are from $[0,10] \times [0,10]$. The results are shown in Fig.7.2.3. Observe that if each point is multiplied by a factor $c > 0$ then the radius will be multiplied by a factor $\sqrt{c}$. Therefore if there is a dilation or shrinking, then the set will not be dilated or shrinked by the same factor.

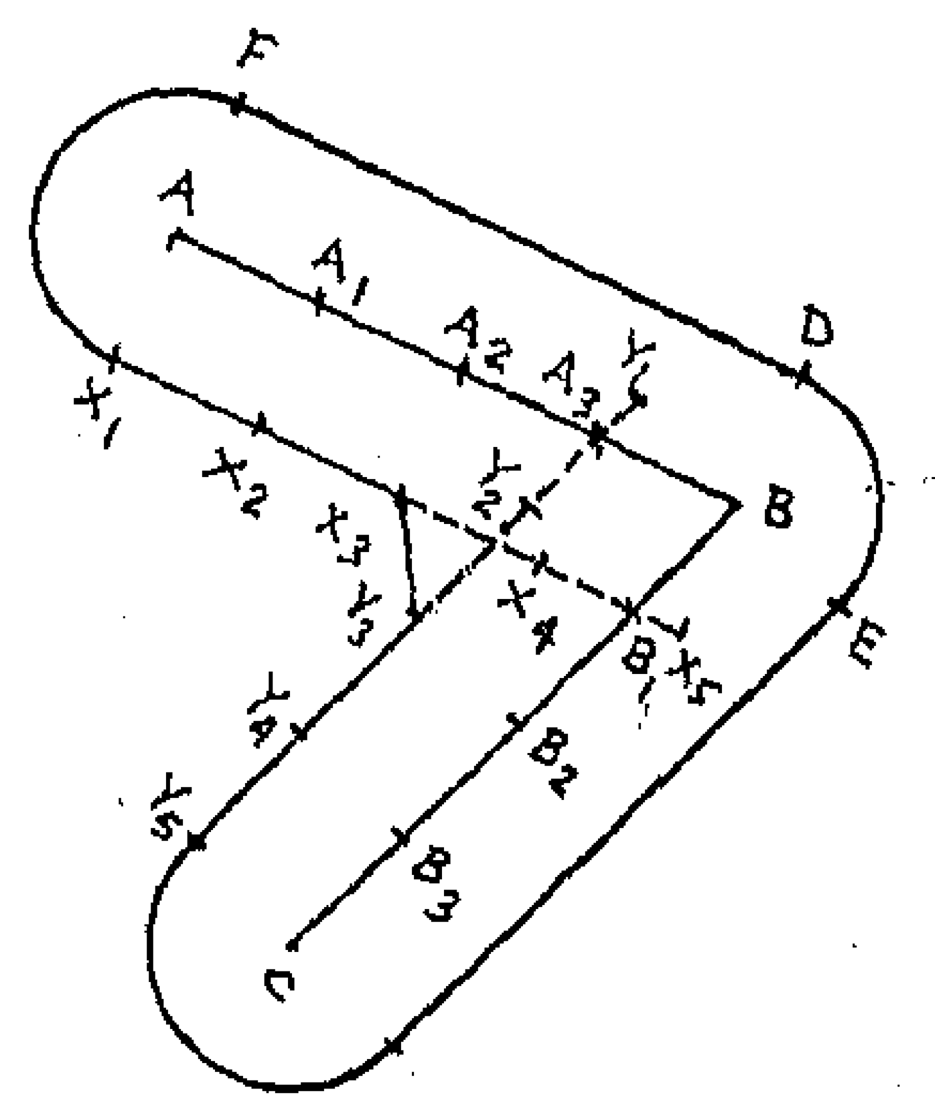
Observe that some points on the parallel line segments, which should have appeared in the set F_1, did not find any place in it. [Fig.7.2.3 and Fig.7.3.1] For example, in fig.7.2.3, more points should have been on the left side parallel line segment of the points 15 and 9. It is because of the inherent approximation in the computer while calculating the radius and the distances. Similar instances can be found out in figures 7.2.3 and 7.3.1.

The procedure described in this chapter is applicable only to R^2. It cannot be applied to R^3.



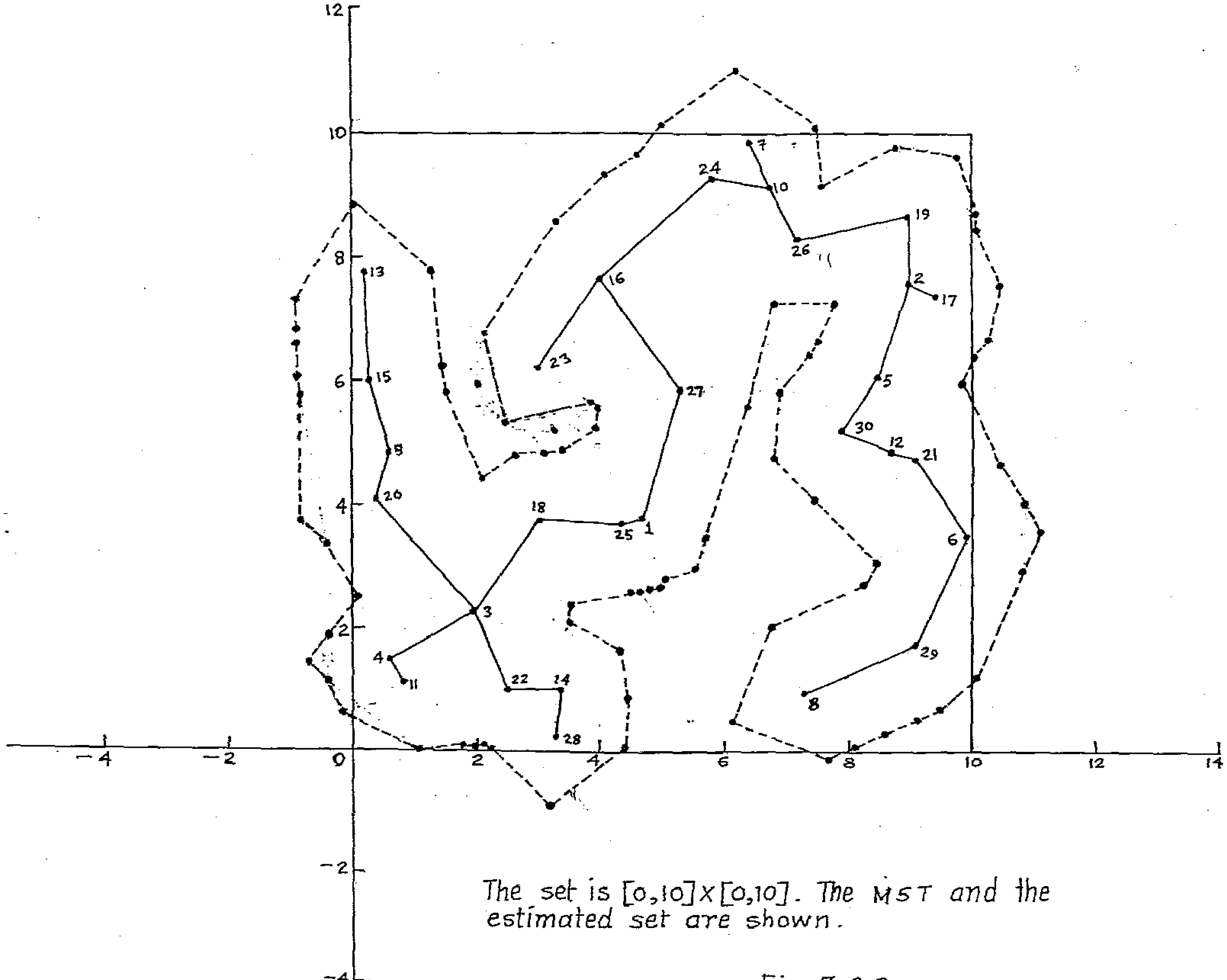
Drawing discs of radius h_n is same as drawing lines with circles at the end.

Fig. 7.2.1



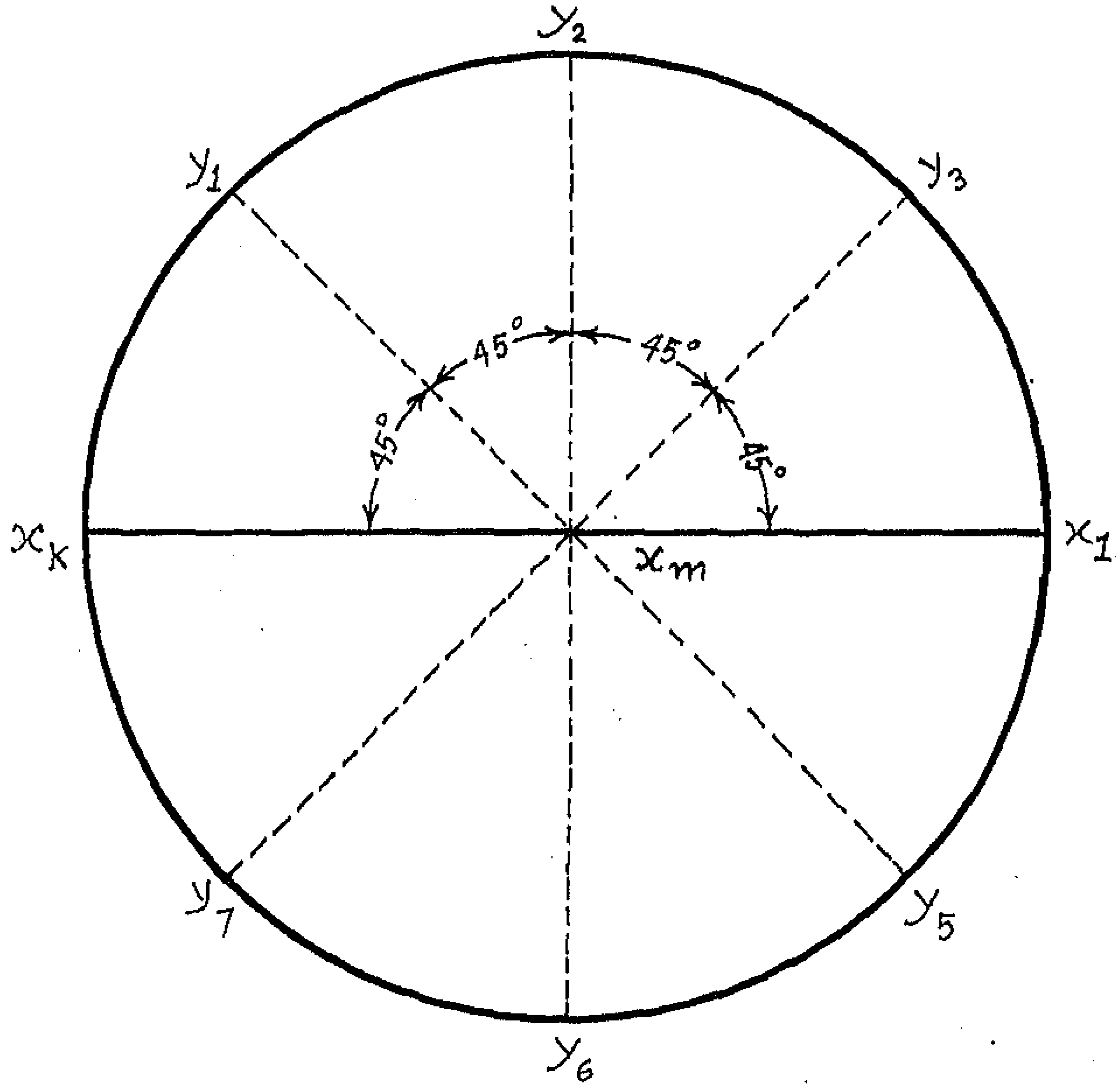
The rejected points [Y_1, Y_2, X_4, X_5] and, in the process, the difference between the actual boundary and the boundary obtained are shown.

Fig. 7.2.2



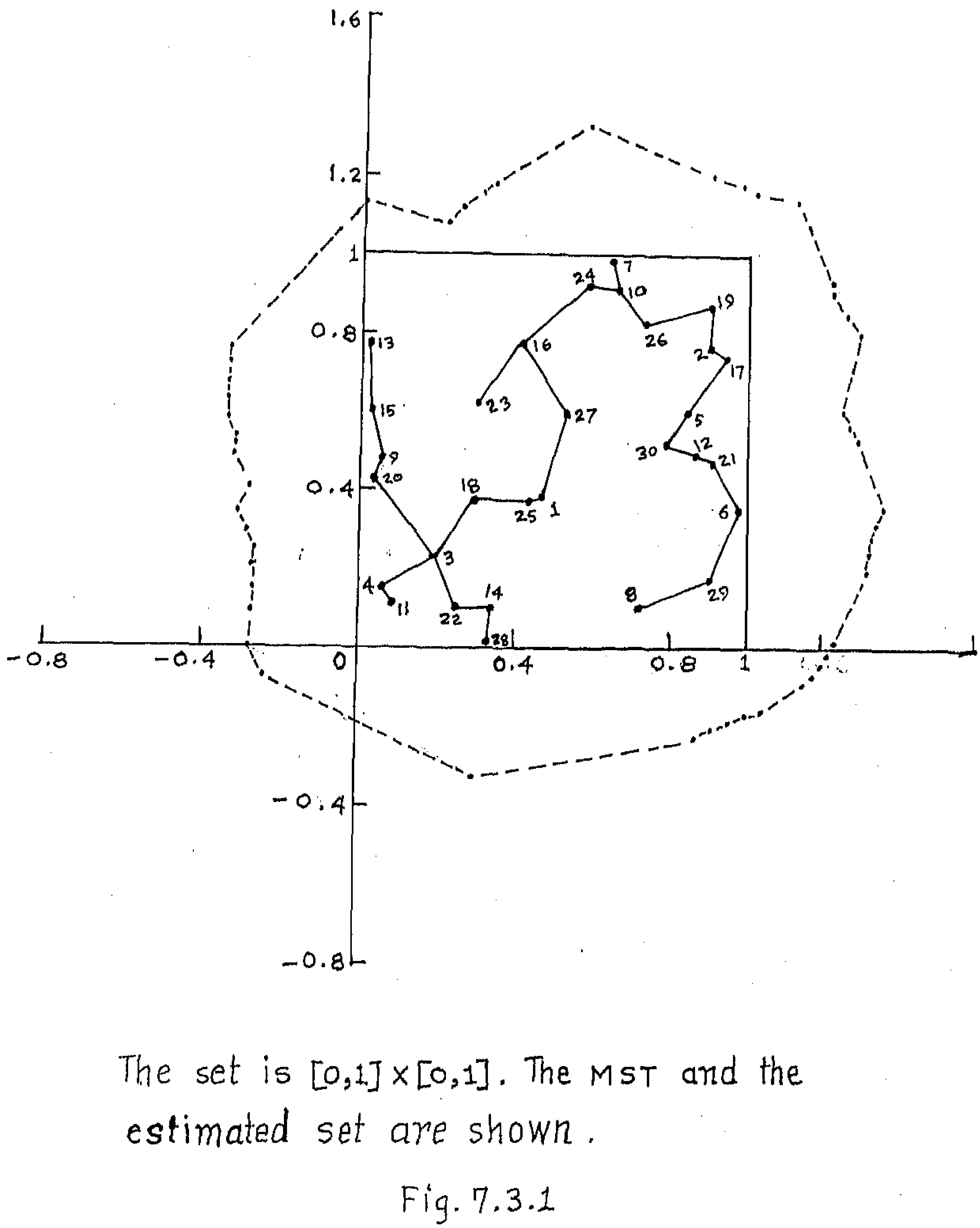
The set is [0,10]X[0,10]. The MST and the estimated set are shown.

Fig. 7.2.3



$d(y_i, y_{i+1})$ is same for $i = 0, \ldots, 7$ when $y_0 = x_K$, $y_4 = x_1$ and $y_8 = x_K$

Fig. 7.2.4



The set is [0,1] x [0,1]. The MST and the estimated set are shown.

Fig. 7.3.1

VIII. Generalization and scope for further work.

8.1 Generalization to R^m in the case of compact regions.

In the earlier chapters, consistent estimators for path connected compact subsets of R^2 are developed. In this chapter the same idea will be generalized for R^m. The results are stated without proofs.

Definition 8.1.1 $\alpha \subseteq R^m$ is said to be a compact region if α is path connected, compact, $c\ell(\text{Int}(\alpha)) = \alpha$, $\text{Int}(\alpha)$ is path connected and $\lambda_m(\delta\alpha) = 0$ where λ_m is lebesgue measure on R^m.

The difference between the above definition and the compact region in R^2 is the condition on $\delta\alpha$. $\delta\alpha$ is assumed to consist of finitely many rectifiable curves in the case of R^2. But in all the proofs, the only fact that was used about the boundary was that its lebesgue measure is zero.

Definition 8.1.1 illustrates the minimum possible conditions on α. It may be of interest to look at the structure of $\delta\alpha$ and derive conditions on the nature of $\delta\alpha$, which may be useful in some other estimation procedures. So, let us, for the present moment, accept the definition 8.1.1. Probability measures on compact regions are defined below. The definition is same, as the one given in Chapter II, excepting that, lebesgue measure is in R^m but not in R^2.

Definition 8.1.2 Q_α is a continuous distribution on a compact region $\alpha \subseteq R^m$ if it satisfies the following properties.

(i) Q_α is a probability measure with support α.

(ii) $Q_\alpha << \lambda_m$ where λ_m is lebesgue measure on R^m.

(iii) The density f is continuous on Int(α).

(iv) $\text{Int}\left\{ x: f(x) \geq \frac{1}{n} \right\}$ is path connected for sufficiently large n.

(v) $\inf_{x \in \alpha_\varepsilon^{in}} f(x) > 0$ for every $\varepsilon > 0$ where $\alpha_\varepsilon^{in} = \alpha \cap \left[\alpha^c \oplus D(\varepsilon)\right]^c$.

Let $\mathcal{G}_\alpha$ represent the set of all probability measures on α satisfying the above conditions.

A theorem was stated in Chapter IV (theorem 4.1.2) generalizing the result of Grenander [theorem 4.1.1, reference 45] to any bounded borel set in R^m. Using that theorem as well as the theorems 4.2.1, 4.2.2 and 4.3.1 the following theorem can be proved, the proof of which follows the proofs of the theorems mentioned.

Theorem 8.1.1 Let $\alpha \subseteq R^2$ be a compact region in R^m and Q be a probability measure on α, $Q \in \mathcal{G}_\alpha$. Let $X_1, X_2, \ldots$ be independent and identically distributed random vectors defined on $(\Omega, \mathcal{A}, P_\alpha)$ to R^m inducing Q on α. Let $G_n(w)$ be MST of $\{X_1, \ldots, X_n\}$. Let $\ell_n(w)$ be its length.

Let $h_n(w) = (\ell_n(w)/n)^{1/m}$.

Let $\alpha_{n1}^*(w) = \bigcup_{i=1}^{n} \left\{ x \in R^m : \| X_i(w) - x \| \leq h_n(w) \right\}$.

Let $\alpha_{n2}^*(w) = \left\{ x \in R^m : d(x, G_n(w)) \leq h_n(w) \right\}$, $w \in \Omega$.

Then (i) $E_\alpha\left[\lambda_m(\alpha_{n1}^{*} \Delta \alpha)\right] \to 0$

(ii) $E_\alpha\left[\lambda_m(\alpha_{n2}^{*} \Delta \alpha)\right] \to 0$ as $n \to \infty$,

(iii) $P_\alpha\left\{w : D(\alpha_{n2}^{*}, \alpha) > \varepsilon\right\} \to 0$ for any $\varepsilon > 0$ and

(iv) $P_\alpha\left\{w : D(G_n,\alpha) > \varepsilon\right\} \to 0$ for any $\varepsilon > 0$.

Proof: Follows from theorems 4.1.2, 4.2.1, 4.2.2, 4.3.1 and by making suitable modifications because of the dimension of the space. Another fact stated below is also to be used [60].

If V is a sphere of radius r in R^m then

$$\lambda_m(V) = \frac{2^i(2\pi)^{i-1}(i-1)!\, r^{2i-1}}{(2i-1)!} \quad \text{for } i = 1,2,\ldots \text{ if } m = 2i-1$$

$$= \frac{\pi^i r^{2i}}{i!} \quad \text{for } i = 1,2,\ldots \text{ if } m = 2i. \ [60]$$

Similarly the classification procedure for compact regions can be generalized also.

The generalization of bounded line classes to R^m and the related questions on probability measures and consistency are treated in the next section.

8.2 On generalization of bounded line classes to R^m.

Bounded line classes in R^2 are subsets of R^2 whose area is zero. In R^2, a bounded line class is a union of rectifiable curves. Every rectifiable curve is a continuous image of $[0,1]$ where $[0,1]$ is a compact region in R^1. So a bounded line class α in R^3 may be defined as $\left(\bigcup_{i=1}^{k_1} A_i\right) \cup \left(\bigcup_{i=1}^{k_2} B_i\right)$ such that α is path connected, there exist compact regions A_{i1}'s in R^2 for $i = 1,\ldots k$, such that $f_i: A_{i1} \rightarrow A_i$ is a homeomorphism for $i = 1,\ldots k_1$ and there exist $g_i : [0,1] \rightarrow B_i$ such that g_i is a homeomorphism for all $i = 1,\ldots k_2$. [A Jordan curve can be expressed as a path connected union of two Jordan arcs and a Jordan arc may be looked upon as a homeomorphism of $[0,1]$].

In general, a bounded line class in R^m may be defined in the following way.

Definition 8.2.1 A bounded line class in R^m is a path connected compact subset α of R^m such that

$$\alpha = \bigcup_{i=1}^{m-1} \left(\bigcup_{j=1}^{k_i} A_{ij} \right)$$, there exists a compact region $B_{ij} \subseteq R^i$ for every i and j, such that $g_{ij}: B_{ij} \rightarrow A_{ij}$ is a homeomorphism and $\mu_i(A_{i_1 j_1} \cap A_{i_2 j_2}) = 0$ for $i = i_1$ and $i = i_2$ [μ_i will be defined later in this section].

In order to define measures on the above classes in R^m, let us recall Chapter II. Initially, a measure μ was defined on bounded line classes and any other measure P is assumed to be absolutely continuous with respect to μ. It can be shown that the measure μ is same as the Hausdorff measure [60] on these sets. Since A_{ij}'s are borel sets and for every borel set hausdorff measure can be defined [60] and also hausdorff measures define lengths of sets in R^2, areas of sets in R^3 etc...., they serve the purpose of measure on these sets. But, in order to define hausdorff measure on curves in R^2, concavity conditions (definition 2.4.2) are unnecessary. Concavity conditions are defined because some sets were not needed to be considered as classes. μ on bounded line classes in R^2 could as well have been defined without the concavity conditions.

Let μ_i, $i \leqslant m$ be hausdorff measure on borel subsets of R^m ~~such~~ where ~~that~~ $\mu_i = \lambda_i$ for $i = m$. Define μ_i to be the measure on borel subsets of A_{ij}'s [see definition 8.2.1] for i and j. The condition on the intersection of A_{ij}'s is incorporated to make the definition of measure on borel subsets of R^m unique. Define μ on a borel set B in R^m in the following way for $\alpha = \bigcup_{i=1}^{m-1}(\bigcup_{j=1}^{k_i} A_{ij})$

$$\mu(B) = \sum_{i=1}^{m-1} \sum_{j=1}^{k_i} \mu_i(A_{ij} \cap B).$$

Definition 8.2.2 Let the set of all probability measures to be defined on a bounded line class $\alpha = \cup \cup A_{ij}$ be denoted by $\mathcal{H}_\alpha$ where any $Q \in \mathcal{H}_\alpha$ satisfies the following properties.

There exist $f_{ij}: R^m \dashrightarrow R^+$ such that

(i) $f_{ij}(x) > 0$ for $x \in A_{ij}$

(ii) $f_{ij}(x) = 0$ for $x \notin A_{ij}$

(iii) f_{ij} is continuous on A_{ij}

(iv) $Q_{ij}(B) = \int_B f_{ij}\, d\mu_i$ for all i, j where B is a borel subset of R^m.

(v) $Q_{ij}(A_{ij}) = 1$

and (vi) there exist $q_{ij} > 0$ such that $\sum_{i,j} q_{ij} = 1$ and

$$Q = \sum_{i,j} q_{ij} Q_{ij} .$$

Definition 8.2.2 is very similar to the properties given in Chapter II for bounded line classes in R^2 (2.4.e.1 to 2.4.e.5).

The definition of consistency for bounded line classes in R^2 is given in the following way [Chapter V]. The MST G_n of the observations and the set $\alpha_n^* = \{x : d(G_n, x) \leq h_n\}$ should satisfy (i) $P_\alpha\{w : D(G_n,\alpha) > \varepsilon\} \to 0$ for every $\varepsilon > 0$ and (ii) $E_\alpha[\lambda_2(\alpha_n^* \Delta \alpha)] \to 0$. Later, in section 5.4, another property for α_n^* has also been stated. $P_\alpha\{w : D(\alpha_n^*,\alpha) > \varepsilon\} \to 0$ for any $\varepsilon > 0$.

For the case of R^m, if it can be shown that

8.2.1 $E_\alpha[\lambda_m(\alpha_n^* \Delta \alpha)] \to 0$

8.2.2 $P_\alpha\{w : D(\alpha_n^*, \alpha) > \varepsilon\} \to 0$ for any $\varepsilon > 0$ and

8.2.3 $P_\alpha\{w : D(G_n,\alpha) > \varepsilon\} \to 0$ for any $\varepsilon > 0$

then α_n^* may be called as consistent estimator of α.

The main problem in defining consistent estimator is that, the estimator must possess, in a sense, the properties of the original set as well as it must be the same for all types of classes. The minimal spanning tree based estimator works well for compact regions in R^m. It can be shown to satisfy the properties 8.2.1, 8.2.2 and the MST can be shown to satisfy 8.2.3 for bounded line classes in R^m also.

8.3 Definition and estimation of bounded mixture classes in R^m.

Bounded mixture classes in R^2 are path connected subsets of R^2 and they are unions of bounded line classes and compact regions (Chapter II). Similar definition can be stated for R^m also. Probability measures on bounded mixture classes in R^2 are convex combinations of probability measures on the individual classes. Similar definition can be stated for R^m. The minimal spanning tree G_n can be shown to satisfy 8.2.3 for bounded mixture classes in R^m. The MST based estimator α_n^* can be shown to satisfy 8.2.1 and 8.2.2 for bounded mixture classes in R^m.

In effect, all the results stated in connection with R^2 concerning compact regions, bounded line classes and bounded mixture classes can be generalized to R^m. But there is scope for improvement in the definitions of these classes in R^m.

8.4 Estimation of Unbounded Classes.

In Chapter II, unbounded closed regions are defined. The MST based estimator fails to work in this case because $\lambda(\alpha_n^* \Delta \alpha)$ is infinity always.

Probably a definition for consistency may be given in the following way. A set β_n is said to be consistent to α if

$$E_\alpha\left[\lambda(\beta_n \cap \alpha^c) + Q_\alpha(\alpha \cap \beta_n^c)\right] \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty$$ where Q_α is the probability measure on α and λ is a σ finite measure. Further work needs to be done on the usefulness of the above definition in the context of α_n^*.

8.5 On the assumption of disjoint classes.

MST fails to provide good results if the classes are not disjoint. It is necessary to develop procedures for automatic classification of observations so that the classification procedure is 'good'. Discriminant analysis deals with those questions. In this dissertation, basically, estimation of sets is given priority.

If the classes are overlapping they cannot be estimated and separated simultaneously because the probability of misclassification is not zero for any number n of observations. Estimation in a 'good' way is not possible though classification procedure may be 'good'.

8.6 Conclusion

The definitions of bounded classes in R^m give only the minimum possible conditions imposed on the sets. They can be improved. Probably estimation procedures may be developed on the basis of the improvements. For unbounded sets, the estimation procedure in R^2 is not developed. The definition of consistency followed in other chapters as well as in this chapter is not a proper one in the context of unbounded sets. Estimation procedures are to be developed for these sets.

This thesis gives classification and estimation procedures when the classes are bounded and disjoint. If the classes are not disjoint simultaneous classification and estimation may not be possible in a 'good' way.

IX. Appendix

A few examples and proofs which were mentioned in earlier chapters are stated here.

A1 An example of a set $\alpha \subseteq R^2$ which is path connected, compact, $Int(\alpha) = \emptyset$ but $\lambda(\alpha) > 0$.

Let $A = [0,1] \times [0,1]$

Let $c_y = \{(x,y) : 0 < x < 1\}$ for all $y \in [0,1]$.

Let $F = \{c_y : y$ rational, $0 < y < 1\}$.

Then there exists a function f such that

$f : \{1,2,\ldots\} \rightarrow F$ which is one-one and onto.

So without loss of generality let

$$F = \{c_{y_i} : y_i \text{ rational}, \ 0 < y_i < 1, \ i = 1,2,\ldots\}.$$

Let $\varepsilon > 0$.

$$\text{Let } E = \bigcup_{i=1}^{\infty} \ \bigcup_{y \in \left(y_i - \frac{\varepsilon}{2^{i+1}},\ y_i + \frac{\varepsilon}{2^{i+1}}\right)} c_y$$

E is open. $\lambda(E) \leq \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon$

Let $\alpha = A \cap E^c$

α is path connected, compact and $Int(\alpha) = \emptyset$.

But $\lambda(\alpha) \geq 1 - \varepsilon > 0$ for $0 < \varepsilon < 1$.

A2 An example of a set $\alpha \subseteq R^2$, $\lambda(\alpha) = 0$, α is path connected, compact and α is a union of uncountably many curves.

Let C be the cantor's set [53]. Let $x_1 = (\frac{1}{2}, 1) \in R^2$.
Let $A = \{(x,0) : x \in C\}$. Let c_y be the set of all points on the line segment joining y and x_1, $y \in A$.

$$\text{Let } \alpha = \bigcup_{y \in A} c_y .$$

A3 An example of a set $\alpha \subseteq R^2$, $\lambda(\alpha) = 0$, α is path connected, compact and α is a union of countably infinite curves.

Let $B = \{\frac{1}{n} : x \text{ positive integer}\} \cup \{0\}$.
Let $x_1 = (\frac{1}{2}, 1) \in R^2$. Let $A = \{(x,0) : x \in B\}$.
Let c_y be the set of all points on the line segment joining x_1 and y, $y \in A$.

$$\text{Let } \alpha = \bigcup_{y \in A} c_y .$$

A4 An example of a curve α that is not rectifiable.
Let $f : [0,1] \longrightarrow R^2$ be defined in the following way

$$f(x) = \left(\frac{1}{n} \cos\left\{\frac{3\pi}{2}\left[2n+1 - 2n(n+1)x\right]\right\},\ \frac{1}{n} \sin\left\{\frac{3\pi}{2}\left[2n+1 - 2n(n+1)x\right]\right\} \right)$$

$$\text{for } x \in \left[\frac{2n+1}{2n(n+1)}, \frac{1}{n}\right] \text{ where } n = 1,2,3,\ldots$$

$$= \left(2x(n+1)-2,\ \frac{1 + 2n\left[1-(n+1)x\right]}{n+1}\right)$$

for $x \in \left[\frac{1}{n+1}, \frac{2n+1}{2n(n+1)}\right]$ where $n = 1,2,\ldots$

$= (0,0)$ at $x = 0$.

Let $\alpha = \{f(x) : x \in [0,1]\}$.

A5 An example of a rectifiable curve α which does not have finitely many concavities. [Similar to A4].

Let $f : [0,1] \dashrightarrow R^2$ be defined in the following way.

$$f(x) = \left(\frac{1}{n^2} \operatorname{Cos}\left\{\frac{3\pi}{2}\left[2n+1 - 2n(n+1)x\right]\right\},\ \frac{1}{n^2} \operatorname{Sin}\left\{\frac{3\pi}{2}\left[2n+1 - 2n(n+1)x\right]\right\}\right)$$

for $x \in \left[\frac{2n+1}{2n(n+1)}, \frac{1}{n}\right]$ where $n = 1,2,\ldots$

$$= \left(\frac{\left[(n+1)x - 1\right] 2}{n},\ \frac{1 - \left[(n+1)x - 1\right] 2n}{(n+1)^2}\right)$$

for $x \in \left[\frac{1}{n+1}, \frac{2n+1}{2n(n+1)}\right]$ where $n = 1,2,\ldots$

$= (0,0)$ at $x = 0$.

Let $\alpha = \{f(x) : x \in [0,1]\}$.

A6, A7 and A8 Measurability of a few functions is proved here. For R^1 and R^2 the σ-fields are assumed to be borel σ-fields. For the cross product of two spaces the σ-field is the product σ-field.

A6 Measurability of a few functions is proved here.

Set up : $\alpha \subseteq R^2$ is path connected, compact. $X_1, X_2, \ldots X_n$ are independent random vectors defined from $(\Omega, \mathcal{A}) \longrightarrow (R^2, B(R^2))$ and taking values in α. [$\mathcal{A}$ is a σ-field on Ω. $B(R^2)$ is the borel σ-field on R^2.] Let $G_n(w)$ = MST of $\{X_1(w), \ldots X_n(w)\}$, $w \in \Omega$. Then the following functions are measurable.

(i) $\ell_n(w)$ = length of $G_n(w)$

(ii) $m_n(w)$ = Maximum edge length of $G_n(w)$.

(iii) $u_n(w) = \sup_{x \in G_n} \inf_{y \in \alpha} \| x-y \|$.

(iv) $h_n(w) = \sqrt{\dfrac{\ell_n(w)}{n}}$.

<u>Proof</u> $X_i : (\Omega, \mathcal{A}) \longrightarrow (R^2, B(R^2))$ is measurable for all $i = 1,2,\ldots n$.

$Y_i : (\Omega \times R^2) \longrightarrow R^2 \times R^2$ such that

$Y_i(w,x) = (X_i(w), x)$ is measurable.

$Z : R^2 \times R^2 \longrightarrow R^+$ such that

$Z(x,y) = \| x-y \|$ is measurable.

$$a_{ij}(w) = Z(Y_i(w, X_j(w))) = Z(X_i(w), X_j(w)) = \| X_j(w) - X_i(w) \|$$ is measurable.

$\ell_{nk}(w)$ = Sum of $(n-1)$ of $a_{ij}(w)$'s for a particular spanning tree k of $X_1, \ldots X_n$ where $[X_i, X_j]$ is an edge of k.

$\ell_n(w) = \min_k \ell_{nk}(w)$ is measurable because minimum or maximum of finitely many measurable functions is measurable.

$m_n(w)$ = supromum of $a_{ij}(w)$'s where the supromum is taken on a certain proper subset of $\{a_{ij}(w) : i,j = 1,\ldots n\}$.

So $m_n(w)$ is measurable.

(iii) Linear combination of measurable functions is measurable.

So $b_{ij}(w) = \tau X_i(w) + (1 - \tau)\ X_j(w)$ is measurable.

$c_{ij}(w,x) = \|b_{ij}(w) - x\|$ is measurable for every $x \in R^2$.

$c_{ij}(w,x)$ is measurable for every $w \in \Omega$ where $x \in R^2$.

Limit of a sequence of measurable functions is measurable and observe also that supremum or infimum in a set is a limit of sequence of points in the set.

So $u_n(w)$ is measurable.

(iv) $h_n(w)$ is measurable since $\ell_n(w)$ is measurable.

A7 It will be shown here that I_{1n} is measurable where

$$I_{1n}(x,w) = 1 \text{ if } x \in \alpha^*_{n1}(w) = \bigcup_{i=1}^{n} \left\{ x : \|X_i(w) - x\| \leq h_n(w) \right\}$$

$$= 0 \text{ otherwise },$$

where $X_1, X_2, \ldots X_n$ and h_n are as defined in A6.

Observe that $g_i(x,w) = (X_i(w),\ h_n(w), x)$ is measurable for $x \in R^2$.

$\Psi_1(x_1,x_2,x_3) = \|x_1 - x_3\| - x_2$ is measurable for $x_1, x_3 \in R^2$ and $x_2 \in R^+$.

$$\left.\begin{aligned} \Psi_2(x) &= 1 \text{ if } x \geq 0 \\ &= 0 \text{ if } x < 0 \end{aligned}\right\} \Psi_2 \text{ is measurable.}$$

$\Psi_3 = \sum_{i=1}^{n} \Psi_2\, \Psi_1\, g_i$ is measurable.

Hence I_{1n} is measurable.

A8 It is to be shown that I_{n2} is measurable where

$$I_{n2}(x,w) = 1 \text{ if } x \in \alpha^*_{n2}(w) = \{x : d(x, G_n(w)) \leq h_n(w)\}$$
$$= 0 \text{ otherwise,}$$

where $X_1, \ldots X_n$, G_n and h_n are as defined in A6.

$a_{ij}(w) = \| X_i(w) - X_j(w) \|$ is measurable.

$\ell_{nk}(w)$ = Sum of $(n-1)$ of $a_{ij}(w)$'s where $[X_i, X_j]$ is an edge of a spanning tree k of $X_1, \ldots X_n$.

$\ell_{nk}(w)$ is measurable for every k.

$r_k(x,w) = d(x,k)$ is measurable.

$s_k(x,w) = (d(x,k), \ell_{nk}(w))$.

Let $\ell_n(w) = \underset{k}{\text{Min}}\ \ell_{nk}(w) = \ell_{nk_0}(w)$ (say).

Let $V_{k_1k_2}(x,w) = (d(x,k_1), \ell_{nk_1}(w))$ if $\ell_{nk_1}(w) \leq \ell_{nk_2}(w)$
$= (d(x,k_2), \ell_{nk_2}(w))$ otherwise.

$V_{k_1k_2}$ is measurable. Similarly $V_{k_1k_2k_3}$ and so on.

By using V the following function (W) can be shown to be measurable because of finitely many conditions imposed.

$W(x,w) = (d(x,k_0), \ell_{nk_0}(w))$.

So $z(x,w) = d(x,k_0) - \sqrt{\frac{\ell_{nk_0}(w)}{n}}$ is measurable.

So I_{n2} is measurable.

A9 α path connected, compact, $Int(\alpha) \neq \varphi$. It is assumed that $Int(\alpha)$ is path connected. An example of a set α is given below where $Int(\alpha)$ is not path connected.

Let $A = [0,1] \times [0,1]$ and $B = [1,2] \times [1,2]$.

Let $\alpha = A \cup B$.

Here α is path connected but the connection is weak. That is why those sets are not considered as classes.

REFERENCES

1. MR Anderberg, 'Cluster Analysis for Applications', Academic Press, New York (1973).

2. PHA Sneath and RR Sokal, 'Numerical Taxonomy', WH Freeman, San Francisco (1973).

3. A.J. Cole, ed., 'Numerical Taxonomy', Academic Press, New York (1969).

4. B. Everitt, 'Cluster Analysis', John Wiley and Sons, New York (1974).

5. J.A. Hartigan, 'Clustering Algorithms', John Wiley, New York (1975).

6. N. Jardine and R. Sibson, 'Mathematical Taxonomy', John Wiley, New York, (1971).

7. R. Dubes and A.K. Jain, 'Clustering Techniques - The User's Dilemma', Pattern Recognition, Vol.8, pp.247-260 (1976).

8. E.W. Forgy, 'Cluster analysis of multivariate data: Efficiency versus interpretability of classifications, Biometric soc meeting, Riverside, California [Abstract in Biometrics, Vol.21, No.3, pp.768] (1965).

9. Ball G.H. and Hall D.J., 'ISODATA; A novel method of data analysis and pattern classification', Technical report of Stanford Research Institute, Menlo Park, California (1965).

10. J.B. Macqueen, 'Some methods for classification and analysis of multivariate observations', Proc. of 5th symposium on mathematics, statistics and probability, Berkeley, Vol.1, pp.281-297, (1967).

11. D. Wishart, 'An algorithm for hierarchical classifications', Biometrics, Vol.22, No.1, pp.165-170.

12. C.T. Zahn, 'Graph theoretic methods for detecting and describing gestalt clusters', IEEE tr. on computers, No.20, pp.68-86 (1971).

13. R.F. Ling, 'Probability theory of cluster analysis', Journal of American Statistical Association, Vol.68, pp.159-164 (1973).

14. R.F. Ling, 'On the theory of construction of K-clusters', Comput. Journal, Vol.15, pp.326-332 (1972).

15. E. Diday, 'Optimization in non-hierarchical clustering', Pattern Recognition, Vol.6, pp.17-33 (1974).

16. R.O. Duda and P.E. Hart, 'Pattern classification and scene analysis', Wiley-Interscience, (1973).

17. K. Fukunaga, 'Introduction to statistical pattern recognition', Academic Press, New York (1972).

18. P.A. Devijver and J. Kittler, 'Pattern Recognition - A statistical approach', Prentice Hall international, London, (1982).

19. A. Wald, 'Sequential analysis', John Wiley, New York (1947).

20. Fix, E. and Hodges, J.L., 'Discriminatory analysis, non parametric discrimination', USAF School of aviation medicine, Randolph Field, Texas, Project 21-49-004, Rept.4, Feb [1951] (Reprinted in Machine Recognition of Patterns', A.K. Agrawala Ed., IEEE Press, New York [1977], pp.261-279).

21. E. Fix and J.L. Hodges, 'Discriminatory analysis, small sample performance', USAF School of Aviation Medicine, Randolph Field, Texas, Project 21-49-004, Rept.11, Aug. [1952] (Reprinted in 'Machine Recognition of Patterns', A.K. Agrawala Ed., IEEE Press, New York (1977) pp.280-322.

22. T.M. Cover and P.E. Hart, 'Nearest neighbour pattern classification', IEEE Trans. on Information theory, Vol.13, pp.21-27, Jan. (1967).

23. R.A. Fisher, 'The case of multiple measurements in taxonomic problems', Ann. Eugenics, Vol.7, part-II, pp.179-188 (1936).

24. L. Fisher and J.W. Van Ness, 'Admissible clustering procedures', Biometrika, Vol.58, No.1, pp.91-104 (1971).

25. R. Gnanadesikan, J.R. Kettenring and J.M. Landwehr, 'Interpreting and assessing the results of cluster analyses', International Statistical Institute Bulletin, pp.451-463 (1977).

26. Z. Sidak, 'Some ideas for comparisons of clustering procedures', Invited paper for the 12th European meeting of statisticians, Varna, September (1979).

27. W.M. Rand, 'Objective criteria for the evaluation of clustering methods', Journal of American Statistical Association, Vol.66, pp.846-850 (1971).

28. J. Van Ryzin (ed.), 'Classification and clustering', Academic Press, New York (1977).

29. P.B. Baker and L. Hubert, 'Measuring the power of hierarchical cluster analysis', Journal of American Statistical Association, Vol.70, pp.31-38 (1975).

30. B.S. Everitt, 'Unresolved problems in cluster analysis', Biometrics, Vol.35, pp.169-181 (1979).

31. J.A. Hartigan, 'Consistency of single linkage for high density clusters', Journal of American Statistical Association, Vol.76, pp.388-394 (1981).

32. R.M. Cormack, 'A review of classification', Journal of Royal Statistical Society, Series A, pp.321-367 (1971).

33. L. Hubert, 'Some applications of graph theory to clustering', Psychometrika, Vol.39, pp.283-309 (1974).

34. L.L. Mcquitty, 'A mutual development of some typological theories and pattern analytic methods', Educ. psychol. Measur., Vol.27, pp.21-48 (1967).

35. C.J. Van Rijsbergen, 'A clustering algorithm', Comput. J., Vol.13, pp.113-115 (1970).

36. W.H.E. Day, 'Validity of clusters formed by graph theoretical cluster methods', Mathl. Biosci., Vol.36, pp.299-317 (1977).

37. R. Dubes and A.K. Jain, 'Validity studies in clustering methodologies', Pattern Recognition, Vol.11, pp.235-254 (1979).

38. Thomas A. Bailey Jr and R. Dubes, 'Cluster validity profiles', Pattern Recognition, Vol. 15, No.2, pp.61-83 (1982).

39. R.A. Jarvis, 'Computing the shape hull of a set of points in the plane', Proc. of the IEEE Comp. Soc. Conf. on Pattern Recognition and Image Processing, pp.231-241, (1977).

40. J. Fairfield, 'Contoured shape generation forms that people see in dot patterns', in proc. IEEE Conf. on Cybernetics and Society, pp.60-64, (1979).

41. S.G. Akl and G.T. Toussaint, 'Efficient convex hull algorithms for pattern recognition applications', in Proc. 4th Intern. Joint Conf. on Pattern Recognition, Kyoto, pp.483-487, (1978).

42. G.T. Toussaint, 'Pattern Recognition and geometrical complexity', 5th Intern. Conf. on Pattern Recognition, Miami Beach, Florida, pp.1324-1347, Dec. (1980).

43. H. Edelsbrunner, D.G. Kirkpatrick and R. Seidel, 'On the shape of a set of points in the plane', IEEE Tr. on Information theory, Vol.IT-29, No.4, pp.551-559, July, (1983).

44. U. Grenander, 'Abstract Inference', John Wiley, New York (1981), pp.465-468.

45. U. Grenander, 'Abstract Inference', John Wiley, New York, pp.373-375, (1981).

46. J.P. Rasson, A. Hardy, L. Szary and P. Schindler, 'An optimal clustering method based on the Rasson's criterion and resulting from a new approach', published in Pattern Recognition Theory and Applications, edited by J. Kittler, K.S.Fu and L.F. Pair, Reidel Publishing Company, pp.63-71, (1982).

47. S.K. Parui, 'Some studies in analysis and recognition of 2-dimensional shapes', Ph.D. thesis submitted to Indian Statistical Institute, Calcutta, India, 1984.

48. C.R. Rao, 'Linear statistical inference and its applications', Wiley Eastern Ltd., New Delhi, 2nd edition (1973).

49. E.L. Lehmann, 'Testing Statistical Hypotheses', Wiley Eastern Ltd., New Delhi (1976).

50. S.K. Parui and D. Dutta Majumder, 'A new definition of shape similarity', Pattern Recognition Letters, Vol.1, pp.37-42 (1982).

51. S.K. Parui and D. Dutta Majumder, 'Shape similarity measures for open curves', Pattern Recognition Letters, Vol.1, pp.129-134 (1983).

52. S.K. Parui and D. Dutta Majumder, 'How to quantify shape distance for two dimensional regions', Proc. 7th Int. Conf. Pattern Recognition, Montreal, pp.72-74 (1984).

53. J. Dugundji, 'Topology', Allyn Bacon, Boston (1966).

54. David A. Binder, 'Approximations to Bayesian clustering rules', Biometrika, Vol.68, pp.275-285 (1981).

55. F.H.C. Marriott, 'Optimization methods of cluster analysis', Biometrika, Vol.69, pp.417-421 (1982).

56. David J. Strauss, 'A model for clustering', Biometrika, Vol.62, pp.467-475 (1975).

57. Stephen P. Smith and Anil K. Jain, 'Testing for uniformity in multidimensional data', IEEE Tr. on Pattern analysis and machine Intelligence, Vol.PAMI-6, No.1, pp.73-81, Jan. (1984).

58. T.M. Apostol, 'Mathematical Analysis', Addison Wesley (1971).

59. R.B. Ash, 'Real analysis and probability', Academic Press, New York (1972).

60. P. Billingsley, 'Probability and measure', John Wiley, New York (1979).

61. C.A. Murthy and D. Dutta Majumder, 'Definitions of classes in R^2', to be published in the proceedings of fifth Int. symposium of Data Analysis and Informatics held in Versailles, France, Sept. 29-Oct.2, (1987).

62. Jerome H. Friedman and Lawrence C. Rafsky, 'Multivariate generalizations of the Wald-Wolfowitz and Smirnov two sample tests', Ann. of Statistics, Vol.7, No.4, pp.697-717 (1979).

63. J.B. Kruskal, 'On the shortest spanning subtree of a graph and the travelling salesman problem', Proc. Amer. Math. Soc., Vol.7, pp.48-50, (1956).

64. R.C. Prim, 'Shortest connection networks and some generalizations', Bell System Tech. J., Vol.36, pp.1389-1401 (1957).

65. V.K.M. Whitney, 'Algorithm 422, minimal spanning tree', Comm. ACM, Vol.15, pp.273-274 (1972).

66. M.I. Shamos and D. Hoey, 'Closest point problems', 16th Ann.Symp. Foundations of Computer Science, IEEE, pp.151-162 (1975).

67. J.L. Bentley and J.H. Friedman, 'Fast algorithms for constructing minimal spanning trees in coordinate spaces', Stanford Linear Accelerator report (SLAC) PUB-1665 (1975).

68. F.J. Rohlf, 'Aprobabilistic minimum spanning tree algorithm', IBM research report, c6502 (1977).

69. F. Harary, 'Graph theory', Chapter 2, Addison Wesley (1969).

70. C.A. Murthy and D. Dutta Majumder, 'A method of finding boundary of a cluster', Published in Data Analysis and Informatics 4, Edited by E. Diday et al, pp.137-148 (1986).

71. C.A. Murthy and D. Dutta Majumder, 'On estimation of compact regions in R^2 in the context of cluster analysis', communicated to Advances in Applied Probability (1987).

72. A.M. Goon, M.K. Gupta and B. Das Gupta, 'Fundamentals of Statistics, Vol.II', World Press, Calcutta (1968).