Close Pairs of Galaxies in Rich Clusters:

A Statistical Analysis

by

Glen Alfred Rutledge

B.Sc., McMaster University , 1992

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Physics and Astronomy

We accept this thesis as conforming to the required standard

[redacted]

Dr. C.J. Pritchet, Co-Supervisor (Department of Physics and Astronomy)

[redacted]

Dr. R.D. McClure, Co-Supervisor (Dominion Astrophysical Observatory)

[redacted]

Dr. F.D.A. Hartwick, Departmental Member (Department of Physics and Astronomy)

[redacted]

Dr. C. Morgan, Outside Member (Department of Philosophy)

[redacted]

Dr. S. Morris External Examiner (Dominion Astrophysical Observatory)

© Glen Alfred Rutledge, 1994

University of Victoria

All rights reserved. Thesis may not be reproduced in whole or in part, by photocopying or other means, without permission of the author.

Supervisors: Dr. C.J. Pritchet and Dr. R.D. McClure

Abstract

The angular correlation function $w(\theta)$ is used to develop a statistically robust technique to search for an excess of close pairs of galaxies (linear projected separations less than 20 h^{-1} kpc) superimposed on the overall density profile of rich clusters of galaxies. This technique is applied to 10 local ($z < 0.1$) rich clusters of galaxies. Abell 262 is found to have too few galaxies to be useful for this technique. Two clusters, Abell 400 and Abell 1904 are found to have at least a 2σ excess of close pairs, while Abell 1367, Abell 2670, and Hercules are found to have a 1σ excess. The remaining four clusters, Abell 401, Coma, Abell 2199, and Abell 2634 all show no significant excess of close pairs. The clusters that show an excess of close pairs tend to be of the dynamically less evolved Rood-Sastry (RS) types I, F, and C. The more dynamically evolved RS types B and cD do not show excesses. An exception to this trend is Abell 2670, which is a cD cluster that shows an excess of close pairs, and also appears to have substructure on the scale of 100 h^{-1} kpc. This result is in agreement with the substructure studies of Bird (1994a, 1994b), which suggest that this cluster has recently undergone a line of sight merger.

Examiners:

[redacted]

Dr. C.J. Pritchet, Co-Supervisor (Department of Physics and Astronomy)

[redacted]

Dr. R.D. McClure, Co-Supervisor (Dominion Astrophysical Observatory)

[redacted]

Dr. H.D.A. Hartwick, Departmental Member (Department of Physics and Astronomy)

[redacted]

Dr. C. Morgan, Outside Member (Department of Philosophy)

[redacted]

Dr. S. Morris, External Examiner (Dominion Astrophysical Observatory)

Table of Contents

Preliminary Pages:

Chapter One:

Chapter Two:

List of Tables

List of Figures

Acknowledgement

Thanks everybody.

Chapter One

Introduction

1.1 Why Study Close Physical Pairs of Galaxies?

The idea that, what goes on *inside* a galaxy (*i.e.*, star formation/bursts, nuclear activity, morphology, etc.) is directly influenced by what goes on *outside* a galaxy (*i.e.*, tidal interactions, gaseous infall, local galaxy density, local dark matter density, etc.) dates back to the mid 1950s, when Baade and Minkowski (1954) suggested that the bright radio source, Cygnus A, may be caused by the interaction of two galaxies. Today it is accepted as a crucial parameter needed to describe a galaxy's evolution. In particular, it is becoming increasingly evident from empirical evidence that a close physical neighbor tidally interacting with a galaxy, or even merging with a galaxy, is of paramount importance in determining its observed appearance.

Morphology

Gravitational influences of a neighbouring galaxy can change the spatial distribution and velocity distribution of matter in a galaxy such that its morphology will appear different from that of an isolated galaxy, as Toomre and Toomre (1972) illustrated in their models of spiral-spiral interactions. Their models showed the same 'bridges' and 'tails' that are present in many of the galaxies in Arp's (1966) 'Atlas of Peculiar Galaxies', and Vorontsov-Velyaminov's (1977) 'Atlas of Interacting Galaxies'. It is possible, if the orbits of close pairs (separated by $\sim 20\ h^{-1}$kpc) of galaxies are approximately parabolic, that within a Gyr the galaxies will merge to form a single unit, due to dynamical friction (Toomre 1977, Barnes 1988). It has been suggested that all ellipticals and bulges of spiral galaxies were formed from a merging of spiral or irregular galaxies (Toomre 1977, see Schweizer 1990 for a review). Recent evidence for this scenario comes from the work of Hernquist (1992, 1993), who shows that the typical $r^{1/4}$ surface brightness profiles of giant elliptical systems are easily reproduced in N-body simulations of merging galaxies with a disk+bulge morphology.

Star Formation Bursts

During an interaction with another galaxy, collisions between molecular clouds (Sanders et al. 1988), or some other mechanism could help to induce a burst of star formation. It has been shown that the distribution of UBV colors in the Hubble sequence can be explained if all galaxies are approximately the same age, and have experienced monotonically decreasing star formation rates (SFRs) of different characteristic decaying time scales τ (Tinsley 1968, Searle et al. 1973, Larson and Tinsley 1974, Huchra 1977). Under this hypothesis, the 'red' ellipticals have small τ and convert most of their gas into stars at an early stage in their lives and thus today are dominated by an older redder population of stars. The 'blue' late type spirals, on the other hand, have a larger τ and thus convert their gas to stars at a much more conservative rate, such that today there is still a substantial amount of star formation occurring.

In a provocative paper by Larson and Tinsley (1978), it was shown that the distribution of UBV colors for interacting and non-interacting galaxies is quite different. The non-interacting galaxies form a rather tight sequence in the UBV color-color plane, while the interacting galaxies show substantial scatter, both to the red and blue of the 'normal' non-interacting locus. They modelled this with short bursts of star formation superimposed on a monotonically decreasing SFR, and concluded that the colors of the interacting galaxies could be well modelled if the bursts had typical durations of the order $\sim 2 \times 10^7$ yrs, involving up to 5% of the mass of the galaxy. The galaxies to the blue of the 'normal' locus represent those in which bursts occurred in the recent past, and may still be occurring, while the galaxies to the red of the 'normal' locus represent ones where bursts occurred $\sim 2 \times 10^8$ yrs ago; although this latter value, they stress, is rather uncertain. These galaxies are redder than the 'normal' locus since the hot blue stars in the burst have relatively short lifetimes (of the order $\leq 1 \times 10^8$ yrs), and thus the remaining population from the burst will be predominately red. Also, the burst consumes much of the gas that would have been used to make stars through the standard monotonically decreasing SFR, and thus this SFR is decreased. They also found that the interacting galaxies which showed signs of a relatively recent interaction, as determined from the tidal features

described by Toomre and Toomre (1972), mostly appeared blueward of the 'normal' locus, as expected from their model.

This paper instigated a large amount of research in this field, the majority of which has supported the view that galaxy interactions can induce bursts of star formation. An excellent review has been published by Kennicutt (1990), to which the interested reader is referred, since a review here is beyond the scope of this work.

Nuclear Activity

The material in a galaxy which is interacting with a close neighbor can lose much of its angular momentum in the process, and thus flow towards the nucleus of the galaxy. This radial inflow of material can induce, or at least 'feed', some kind of nuclear activity in the galaxy, as first described in detail by Gunn (1979). Since 1979, a multitude of papers have tried to correlate nuclear activity of many manifestations (ultraluminous infrared sources, QSOs, Seyferts, radio power, etc.) with the presence of a close neighbor. This field has been reviewed by Heckman (1990). In summary, there is a substantial amount of evidence which points towards nuclear activity being influenced by interactions, although the time scale between the actual interaction and the beginning or end of the activity is still rather uncertain.

1.2 Searching for Correlations

Assuming there is some observational phenomenon (color, radio power, nuclear activity, etc.) that is expected to be significantly affected by the presence of a close neighbor, there are essentially three ways to provide observational evidence for this association. (Note that 'frequency of close pairs' refers to the number of close pairs relative to the underlying number density):

(i) **Phenomenon -> Pair**

A sample of galaxies is selected that displays the observational phenomenon, along with a 'control' sample that does not display the phenomenon. A suitable

statistical test is performed to see if the distribution of neighbouring galaxies is significantly different in the two samples.

(ii) **Pair -> Phenomenon**

A sample of close pairs of galaxies is selected, along with a 'control' sample of isolated galaxies. An observing program is designed which attempts to observe the phenomenon of interest in both samples of galaxies, and a suitable statistical test is performed to see if there is a significant difference in the frequency of occurrence in the two samples.

(iii) **Statistical Trends**

A statistical description of the frequency of close pairs is developed as a function of local galaxy density, redshift, or some other parameter. The observational phenomenon of interest is observed as a function of the same parameter, and it is seen whether the same trend exists.

The last method is the least exploited of the three, even though it has the distinct advantage that the trends observed in a single statistical description of the close pairs can be compared to the trends in *several* different observational phenomenon. In addition, a statistical description of the frequency of close pairs can be used to better understand clustering on both small and large scales. It is the statistical trends of close pairs that will be of interest in this work.

Statistical Trends for Field Galaxies

In the field, the frequency of certain phenomenon such as quasars (Boyle et al. 1988), IRAS galaxies (Lonsdale et al. 1990) and the luminosity density (Eales 1993) tend to increase as a function of redshift. These trends may be correlated with an increase in the frequency of close pairs as a function of redshift. Zepf and Koo (1989) found, by comparing the number of close pairs from a faint galaxy sample with that of a low redshift sample, that there is an increase in the frequency of close pairs with redshift. Recently, this result has been corroborated by two different studies. Burkey et al. (1994) also infer

an increase in the frequency of pairs with redshift using HST data. By comparing the number of faint close pairs found in a CCD survey with what is expected from the inward extrapolation of a power law angular correlation function $w(\theta)$ (δ=0.8, $w(3'')$=1), Carlberg et al. (1994) have also found an excess of close pairs in higher z samples.

These trends are also important to our understanding of the formation and evolution of large scale clustering, since the increase in the frequency of close pairs with redshift implies an increase in the interaction/merger rate with redshift as well. In the standard hierarchical clustering models, merging of galaxies is a natural consequence (Efstathiou 1990), such that the merger rate is predicted to increase with redshift (Toomre 1977). Potentially, this merger rate can be used as a constraint on the density parameter of the universe, Ω (Carlberg 1990).

Statistical Trends in Rich Clusters

The environment within a rich cluster is much different from that of the field, and thus the formation and evolution of close pairs in clusters may be expected to differ significantly from the field. If binary galaxies form from a capture process, rather than a double condensation from the same primordial gas cloud, then the frequency of close pairs of galaxies should be proportional to the square of the galaxy number density (supported by the early work of Page et al. [1961]) assuming no pair destruction mechanisms exist. In this sense, clusters should have a much higher frequency of close pairs than the field. However, close pairs of galaxies in clusters can be destroyed both through merging, as in the field, and also through tidal disruption, which occurs rarely in the field. For these reasons, the frequency of pairs is expected to be highest in infant rich clusters, but the evolution of pairs in these clusters is expected to occur at a more dramatic rate than in the field. Also, once a rich cluster is established, the typically high velocity dispersions (~ 1000 km s^{-1}) make interactions and mergers between galaxies highly unlikely. Toomre and Toomre (1972) have shown that merging preferentially occurs at relative velocities less than ~ 200 km s^{-1}.

Several phenomena associated with rich clusters show trends that may be associated with similar trends in the frequency and evolution of their close pairs. Ellingson and Yee (1993a, 1993b and references therein) found that active galactic nuclei (AGN) have evolved approximately 5 times faster, since $z \sim 0.7$, in rich clusters of galaxies than in poor clusters. The bright AGN were preferentially found in regions of lower velocity dispersion. Yates et al. (1989) found, from a sample of 25 powerful radio galaxies in the redshift range 0.15 to 0.82, that the radio galaxies with redshifts greater than 0.3 occupy environments similar to local Abell class 0 clusters (*i.e.* 30-49 galaxies inside 1 Abell radius ($\sim 1.5\ h^{-1}$ Mpc), and within 2 magnitudes of the third brightest galaxy), while radio galaxies with redshifts smaller than 0.3 occupy regions which are a factor of 3 times less dense.

The Butcher-Oemler effect (Butcher and Oemler 1978a, 1978b, 1984a, 1984b; Butcher et al. 1980, 1983; Dressler and Gunn 1983) is essentially an excess of blue galaxies in the cores of rich clusters of galaxies at redshifts of ~ 0.3-0.4 compared to local clusters. This effect may also be a result of interactions and mergers of close pairs. There has been debate as to whether the Butcher-Oemler effect could be explained strictly through cosmic evolution (*i.e.* a general blueing of galaxies in *all* environments as a function of redshift). Work by Allington-Smith et al. (1992) seems to indicate that the dense environment of rich clusters is necessary, as the poor and intermediate richness groups associated with radio galaxies that they studied did not show the same rapid evolution of the blue population from a redshift of $z \sim 0.5$ to the present.

Lavery and Henry (1988), and Lavery (1990) find that of 19 blue cluster galaxies at $z \sim 0.2$, 25% were multiple systems. By performing a statistical nearest neighbor test on a blue and red sample of galaxies, Lavery and Henry (1988) find that the blue sample has a distribution of nearest neighbor separations that is smaller in the mean and statistically different (at the 98% confidence level for a Kolmogorov-Smirnoff test) from the distribution of the red sample. Lavery, Pierce and McClure (1992) performed high resolution imaging on 23 blue galaxies with spectroscopically confirmed membership in the clusters A370 ($z \sim 0.37$), and CL0024+1654 ($z \sim 0.39$). They find that 10 of the 23

galaxies have close companions on a scale of 1-2 ″ (~ 9 h^{-1} kpc), and several of the remaining blue galaxies have peculiar morphology typical of recent interactions (Toomre and Toomre 1972). More recently, Pierce, McClure and Lavery (1993) performed high resolution imaging of the distant cluster GHO 0020+0407 ($z \sim 0.698$). They find a dramatic excess, using the angular correlation function (3x more than would be expected for a random distribution of objects), in the number of close pairs and triplets with separations on the order of 1.5″ (~ 17 h^{-1} kpc depending q_0)

These results suggest that there is an excess of close pairs in clusters between redshifts of 0.2 and 0.7 compared to what would be expected from a random distribution. This excess seems to be correlated with an increase in the fraction of blue galaxies in these clusters, but see Dressler (1993) for cautionary remarks about assuming that mergers and interactions are the only cause of the Butcher-Oemler effect. All of these studies have focussed on the blue galaxies in the clusters. Local clusters with very few blue galaxies have not been studied sufficiently to say whether there is an actual trend in the frequency of close pairs in clusters as a function of redshift. Pierce, McClure and Lavery (1993) calculate the angular correlation function $w(\theta)$ for the local irregular Hercules cluster ($z \sim 0.04$) and find there is also an excess of close pairs with typical separations of ~ 35″ (20 h^{-1} kpc). They conclude that since Hercules is rather 'young' dynamically, interactions and mergers in clusters are more likely related to the dynamical state of the cluster rather than strictly to lookback time. However, there also existed a lot of other structure in the angular correlation function of Hercules at larger separations that did not exist in the angular correlation function of GHO 0020+0407; this was left unexplained.

1.3 Describing the Distribution of Close Pairs with $w(\theta)$

The angular correlation function $w(\theta)$ has been used extensively for the field galaxy population (Peebles 1980) to describe the large scale structure in the universe. The structure in the field is well described by a power law of slope $-\delta$, where $\delta \sim 0.8$ for relatively bright galaxies; δ may become shallower for fainter galaxy samples (Infante and Pritchet 1994). The angular correlation function is simply the projection onto the celestial

sphere of the more fundamental spatial correlation function $\xi(r)$. Therefore $w(\theta)$ is directly related to $\xi(r)$ by a fairly simple integral relation, as first demonstrated by Limber (1953, 1954). As mentioned earlier, Carlberg et al. (1994) used the inward extrapolation of $w(\theta)$ for their faint sample of galaxies in order to test for an excess of close pairs in the field.

Rich clusters of galaxies, on the other hand, have traditionally been assumed to be virialized systems (Zwicky 1937; Bahcall 1977 for a review) with azimuthally symmetric density profiles such as a King Profile (Dressler 1978), or a de Vaucouleurs Profile (Beers and Tonry 1986). Subclustering in rich clusters was first taken seriously when Geller and Beers (1982) published contour plots of several local rich clusters. They identified substructure when isolated clumps were associated with peak densities greater than three times the background noise level. West and Bothun (1990) argued, however, that much of their 'substructure' could simply be the result of rather large statistical fluctuations inherent in the characteristically small number of galaxies in a cluster. See Beers (1992) for a recent review on the current views of subclustering in clusters.

The existence of substructure in clusters is crucial to the existence of interactions and merging among close pairs, since the small velocity dispersions in the sub clumps are necessary to prevent the destruction of the physical pairs. Likewise, the presence of close pairs implies subclustering should exist. A scenario of cluster formation presented by Roos and Norman (1979), where clusters form from an agglomeration of smaller subclumps, would predict subclustering and a higher frequency of interactions/mergers. Interacting pairs of galaxies may also be expected in the scenarios described by Rivolo and Yahil (1983), and Cavaliere et al. (1986) in which subclumps of galaxies fall into a virialized cluster but are not disrupted on their first passage. Instead, the clump is elongated on its passage near the core, and then recongregates at apoapsis where mergers and interactions can again occur.

As mentioned above, Pierce, McClure and Lavery (1993) used the angular correlation function $w(\theta)$ to look for an excess of close pairs in both a high redshift cluster, and a local cluster. A more thorough analysis of the distribution of close pairs in clusters using $w(\theta)$

would be invaluable for comparing with the trends in the observational phenomena discussed in §1.2; these trends all seem to show a rapid decrease in the frequency of their occurrence in rich clusters from a redshift of ~ 0.7 to the present. This more thorough analysis would also provide insight into the different scenarios of cluster formation.

There are several advantages and disadvantages of using $w(\theta)$ to analyze the distribution of close pairs in cluster environments:

Advantages

(i) It is an excellent statistical tool since it averages the close pairs and triplets over all galaxies in the sample and thus maximizes the signal to noise ratio; this can be a problem for the small number of galaxies typically associated with a rich cluster.

(ii) It requires only positional data for the galaxies in a cluster. Therefore, it is not observationally time consuming, and can be extended to rather large redshift.

(iii) It is relatively straightforward to calculate, and thus is not computationally time consuming.

(iv) The method is quantitative, and thus easily comparable from one cluster to another.

Disadvantages

(i) It is not straightforward to interpret the results. Its relative ease of calculation has led to misuse and misinterpretation (Fall et al. 1976).

(ii) Large scale density gradients, such as the radial density fall off in rich clusters of galaxies, can induce correlations in $w(\theta)$ on all scales, so special care must be taken to understand these effects.

It should be understood that the power law correlations that are observed in the field using $w(\theta)$ are not expected to be found when analyzing a rich cluster of galaxies, since the 'fair sample' (Peebles 1980) assumption (*i.e.* the assumption that you are not looking at an atypical part of the universe) has been violated by looking directly at a rich cluster.

Also, the integral relation between $w(\theta)$ and $\xi(r)$, which was developed by Limber for the field, does not apply to the cluster situation due to the inhomogeneity of the sample. It is also not plausible to calculate $\xi(r)$ for clusters directly since peculiar velocities of galaxies in a cluster are on the order of their velocity separation due to the Hubble flow; this implies that velocity dependent distance estimators are not very useful. Any non-velocity distance determination is too time consuming, at present, to produce a large enough sample. However, radial velocities in the field of a cluster of galaxies are useful to determine cluster membership.

Contamination by background and foreground galaxies is not expected to be too severe for local clusters, but will get consistently worse for higher redshifts. This is particularly worrisome when comparing clusters at two different redshifts. If radial velocities have not been used to confirm membership in the larger z clusters (as is usually the case), then a statistical subtraction of the close pairs from contaminating galaxies must be made. This will ultimately rely on the work of Carlberg et al. (1994), Burkey et al. (1994), and Zepf and Koo (1989) to better determine the frequency of close pairs in the field.

1.4 Organization of This Thesis

In this thesis, a robust statistical tool will be developed to analyze the frequency of close pairs of galaxies in a rich cluster environment, such that the statistical uncertainty in the analysis is well understood, and the effects of large scale density gradients in the clusters is accounted for. The goal of this thesis is to analyze the scientific results from the application of this technique to 10 local clusters of galaxies. The results will be used to see if the frequency of close pairs in a cluster is dependent on the dynamical state of the cluster.

In Chapter Two, the angular correlation function will be defined, robust estimators will be developed, and the variance in the estimators will be rigorously analyzed. In Chapter Three, the effects of large scale density gradients on the angular correlation function will be analyzed, and an analytic method to determine the edge corrections necessary for the estimators will be developed. In Chapter Four, an adaptive kernel estimator will be used

to remove the effects of the density gradients in the clusters, and the overall analysis technique will be presented and applied to 10 local clusters. The scientific results will be discussed, along with possible directions for future work in this field.

Chapter Two

Practical Estimation of the Angular Correlation Function

The goal of this chapter is to precisely define the angular correlation function, and to discuss some of the practical problems that arise when trying to estimate its value from a discrete distribution of galaxies that possesses a large scale density gradient. The angular correlation function $w(\theta)$ will be defined in §2.1. Several formulae (referred to as estimators) will be developed in §2.2 to estimate its value from a discrete distribution of galaxies in some solid angle Ω. Finally the variance of these estimators will be analyzed in §2.3, both for objects distributed randomly, and for objects distributed with a large scale density gradient typical of a rich cluster of galaxies.

It is strongly suggested that the Appendix §A1 'Introduction to Probability', §A2 'Probability Distribution Functions', and §A3 'Mathematical Expectation' be read before this Chapter to familiarize oneself with the notation and terminology that will be used henceforth.

2.1 Defining $w(\theta)$

Let δP be the joint probability of finding one object in the infinitesimal element of solid angle $\delta\Omega_1$ and another object in the infinitesimal element of solid angle $\delta\Omega_2$ centered θ away from $\delta\Omega_1$. Furthermore, let the sample have a mean density of n. Then the angular correlation function $w(\theta)$ is defined as (Peebles 1980, eq. 31.4)

$$\delta P = n^2 (1 + w(\theta)) \, \delta\Omega_1 \delta\Omega_2 \quad . \qquad \text{(eq. 1)}$$

Alternatively, since the probability of finding an object in $\delta\Omega_1$ is $n\delta\Omega_1$, the angular correlation function is also defined by the conditional probability (see definition 1 and definition 7 in Appendix A) of finding an object in $\delta\Omega_2$, given that there is an object in $\delta\Omega_1$ (Peebles 1980, eq. 31.6),

$$\delta P\,(2|1) \;=\; n\,(1 + w\,(\theta)\,)\,\delta\Omega_2\,, \qquad \text{(eq. 2)}$$

which is equivalent to choosing an object at random, and estimating the probability of finding another object θ away (Peebles 1980, eq. 31.7),

$$\delta P \;=\; n\,(1 + w\,(\theta)\,)\,\delta\Omega \quad . \qquad \text{(eq. 3)}$$

These definitions also apply to cross-correlations between different types of objects. Let one sample of objects be labelled with an 'a', and another sample of objects be labelled with a 'b', where n_a, and n_b are the densities of each of the samples respectively. Then eq. 1, and eq. 3 become

$$\delta P \;=\; n_a n_b\,(1 + w_{ab}\,(\theta)\,)\,\delta\Omega_1\delta\Omega_2 \qquad \text{(eq. 4)}$$

$$\delta P \;=\; n_b\,(1 + w_{ab}\,(\theta)\,)\,\delta\Omega \qquad \text{(eq. 5)}$$

respectively, where all other symbols are defined as they were in the original equations. This defines the angular cross-correlation function $w_{ab}(\theta)$. By these definitions, $w_{ab}=w_{ba}$.

Note that for a Poisson point process (see 'Basic Terminology' in Appendix A), in which every object's position is completely independent of every other object, the probability of finding an object in an infinitesimal element of solid angle $\delta\Omega$ centered θ away from another object is simply $n\delta\Omega$; thus, $w(\theta)$ essentially measures the probability in excess of the random probability of finding an object θ away from another object.

The spatial correlation function $\xi(r)$ is defined similarly to $w(\theta)$, except that all the solid angles are replaced with volumes.

2.2 Estimating $w(\theta)$

The distribution of galaxies on the sky can either be analyzed by using the individual positions of the galaxies on the sky, or by splitting the sky up into cells and using the counts in each of the cells. If the cell approach is used (*e.g.* Lick Survey - Shane and Wirtanen 1967, APM galaxy survey - Maddox et al. 1990), then eq. 1 should be used to develop an estimator for $w(\theta)$ since counts in disjoint cells will be used. Hewett (1982) demonstrated that the best estimator for this type of analysis is an 'ensemble' estimator (employed by Peebles 1975, Peebles and Hauser 1974, Maddox et al. 1990) which Hewett showed was least biased by large scale density gradients and edge effects; in addition, it is not subject to the integral constraint (see Peebles 1980, eq. 47.8, and eq. 12 below). The problem with an analysis of this nature is that all the information on scales smaller than or equal to the size of the cells is lost. Therefore, if it is the small scale structure in $w(\theta)$ that is of interest, and if the catalogue permits, it is better to analyze the distribution of the galaxies using 'direct' estimators which use the individual galaxy positions. In this case, it is better to use eq. 3 to develop an estimator for $w(\theta)$ since the counts in annular bins around each galaxy in the sample will be used. Since in this work it is the close pairs of galaxies in rich clusters that are of interest, the latter approach will be taken, and the estimators will be derived from eq. 3.

Development of the Estimators

The infinitesimal probability in eq. 3 should be thought of as the probability of an event occurring within an ensemble of identically constructed experiments. If M identically distributed sets of galaxies were analyzed, such that a galaxy was chosen at random from the outcome of each experiment, then the number of times that a neighbouring galaxy would be found in an infinitesimal solid angle $\delta\Omega$, located θ away would be $N = M \cdot \delta P = M \cdot n\,(1 + w\,(\theta))\,\delta\Omega$ (with some statistical noise). In any one of the experiments the probability of obtaining more than one galaxy in $\delta\Omega$ is negligible, and thus the expectation value (see definition 8 in Appendix A) for the number of neighbouring galaxies N_n in $\delta\Omega$ is given by

$$E(N_n) = \sum_{N_n=0}^{\infty} N_n \cdot P(N_n) = 0 + 1 \cdot n(1 + w(\theta))\,\delta\Omega + \ldots \approx n(1 + w(\theta))\,\delta\Omega\,. \quad \text{(eq. 6)}$$

Therefore, if $w(\theta)$ is approximately constant over some finite solid angle $\Delta\Omega$, then the expectation value for the number of galaxies in $\Delta\Omega$ is simply the integral of eq. 6 over the solid angle $\Delta\Omega$ with $w(\theta)$ taken outside the integral,

$$E(N_n) \approx n(1 + w(\theta))\,\Delta\Omega\,, \quad \text{(eq. 7)}$$

where $w(\theta)$ is taken to be the average value within the solid angle. It is eq. 7 that will be used to estimate $w(\theta)$ from a discrete distribution of galaxies.

Given N galaxies distributed in a solid angle Ω, the following estimates of the quantities in eq. 7 are made. Each galaxy in turn is taken as a center and the separation to each of the other galaxies is calculated. These separations are placed in bins such that $DD(\theta)$ (referred to simply as DD henceforth, where the angular dependence is assumed) represents the number of separations between $\theta-\Delta\theta/2$, and $\theta+\Delta\theta/2$. Since the total number of distinct separations possible for N galaxies is $N(N-1)/2$ (*i.e.* $N-1$ for the first center, $N-2$ for the next center,... = $\sum_{i=1}^{N} N - i = N(N-1)/2$), an estimate for the probability of finding an object in an annular bin centered θ away from another galaxy is $2DD/[N(N-1)]$. Since there are $N-1$ possible neighbours, an estimate for the expectation value for the number of neighbours in the annular bin is given by

$$E\left(\tilde{N}_n\right) = \frac{2DD}{N}\,, \quad \text{(eq. 8)}$$

where the tilde above a variable indicates that it is an estimate of the true value.

Since one galaxy is used as a centre, an estimate for the density of the remaining galaxies is

$$\tilde{n} = \frac{N-1}{\Omega} \qquad \text{(eq. 9)}$$

The last value in eq. 7 that needs to be estimated is $\Delta\Omega$. If there were no boundaries, then edge effects would not be a problem, and $\Delta\Omega$ would simply be $2\pi\left(\cos\left(\theta - \Delta\theta/2\right) - \cos\left(\theta + \Delta\theta/2\right)\right)$, the solid angle subtended by an annular ring with inner boundary $\theta-\Delta\theta/2$, and outer boundary $\theta+\Delta\theta/2$. However, any galaxy being used as a center that is closer than $\theta+\Delta\theta/2$ to a boundary edge will have some of this solid angle lie outside the sample boundaries. A new quantity $\langle\Delta\Omega\rangle$ to take these edge effects into account is defined for each bin (its dependence on θ here is assumed - *i.e.* it should be $\langle\Delta\Omega\rangle(\theta)$). Peebles (1980, §47, p. 184) defines this quantity as,

> (i) ...the mean value of the solid angle subtended by the annulus $\theta-\Delta\theta/2$ to $\theta+\Delta\theta/2$ and within Ω for ring centers randomly placed in Ω

Alternatively, it could be defined as,

> (ii) ...the mean value of the solid angle subtended by the annulus $\theta - \Delta\theta/2$ to $\theta + \Delta\theta/2$ and within Ω for ring centers placed *at the positions of the galaxies in* Ω

This latter definition is perhaps more appropriate, since it more directly determines how the boundaries affect the counts around each galaxy. While definition (i) has been used extensively in the past (Koo and Szalay 1984, Pritchet and Infante 1986, Ling et al. 1986, Collins et al. 1988), simulations by Infante (1994) have shown that it can produce spurious 'humps' in the correlation function at large θ, and thus definition (ii) is recommended. So, in general, without regard to which definition is used here, an estimate for the solid angle $\Delta\Omega$ in eq. 7 can be written as,

$$\Delta\tilde{\Omega} = \langle\Delta\Omega\rangle\,. \qquad \text{(eq. 10)}$$

Therefore, an estimate of $w(\theta)$ can be made by substituting the estimates given in equations 8,9, and 10 into eq. 7, and rearranging to give

$$1 + \tilde{w}(\theta) = \frac{2DD}{N(N-1)} \cdot \frac{\Omega}{\langle \Delta\Omega \rangle}. \qquad \text{(eq. 11)}$$

Integral Constraint

Notice that since $\sum_{\text{all bins}} DD = N(N-1)/2$, eq. 11 can be re-arranged such that,

$$\sum_{\text{all bins}} (1 + w(\theta)) \frac{\langle \delta\Omega \rangle}{\Omega} = \frac{\sum_{\text{all bins}} DD}{N(N-1)/2} = 1 \qquad \text{(eq. 12)}$$

which is commonly referred to as the *integral constraint*. In words, it implies that the sum over all the bins of $(1+w(\theta))$ times the fractional solid angle of each bin is equal to a constant; if $w(\theta)$ is positive at small angles, as it is in practice, then it must necessarily be negative at some larger angle. Due to the weighting with the fractional solid angle of each bin, a small negative value of $w(\theta)$ for a bin that has a large fractional solid angle can compensate for several bins with large values of $w(\theta)$ and small fractional solid angles. This constraint is only important if comparisons between different data sets with different boundary conditions is desired. Since in this work, all comparisons will be made between a data set and a simulation performed with the same boundary conditions and large scale density gradients, a correction for the *integral constraint* will not be attempted.

Edge Corrections

For a given data set and sample space Ω, every term in eq. 11 is fixed except for $\langle \Delta\Omega \rangle$. Estimating the value for this quantity is often referred to as making *edge corrections;* its value for each bin is determined from the boundary conditions for the sample (including the boundaries of regions within the sample space which have been

removed due to bright star saturation, bad CCD columns etc.). There have been several different methods used in the past for making edge corrections (see Sharp [1979], Hewett [1982] for the details of the different methods employed); however the most common method used today is Monte Carlo estimation. Although it involves more computer cpu time than other methods, its relative ease to program, and its ability to take into consideration complex boundary conditions, has made it the method of choice.

In order to use the Monte Carlo method, the following steps are performed,

(i) N_r objects are distributed randomly with the identical boundary conditions as the data.

(ii) The separations between all of the objects are computed and binned in the same manner as the data. This quantity will be referred to as $RR(\theta)$ (referred to simply as RR henceforth, where the angular dependence is assumed).

(iii) The separations between the galaxies in the data set and the randomly placed objects are computed and binned in the same manner as the data. This quantity will be referred to as $DR(\theta)$ (referred to simply as DR henceforth, where the angular dependence is assumed).

(iv) Steps (i)-(iii) are performed N_t times and the mean and standard deviations of the quantities RR, and DR are calculated, represented by $<RR>$, $\sigma(RR)$, and $<DR>$, $\sigma(DR)$ respectively. The error in the mean is estimated in the standard manner as $\sigma(<RR>) = \sigma(RR)/N_t^{1/2}$, and, $\sigma(<DR>) = \sigma(DR)/N_t^{1/2}$. Therefore, N_t is chosen large enough to achieve the desired level of accuracy.

Note that to decrease the statistical noise in these two quantities, either N_t can be increased, as shown in step (iv), or N_r, the number of random objects used can be increased which reduces the statistical noise in each trial which scales approximately as $1/N_r^{1/2}$. In both cases, the error in the mean is proportional to $1/(N_tN_r)^{1/2}$. Therefore, one could obtain the same statistical error in the estimate for RR (same arguments apply for DR) with one trial (*i.e.* $N_t'=1$), having $N_r'=N_rN_t$ random points. However, since there are

$N_r'(N_r'-1)/2 \approx N_t^2(N_r(N_r-1)/2)$ comparisons needed to analyze this trial compared to $N_t(N_r(N_r-1)/2)$ comparisons needed for N_t trials with N_r objects, the latter method is more efficient by a factor of N_t (see Landy and Szalay 1993). The preferred method is to use the same number of random points as there are galaxies in the sample. In this way $\sigma(RR)$ is a rough estimate for the statistical error in DD.

Since there are a total of $N_r(N_r-1)/2$, and N_rN separations for the random-random and data-random correlations respectively, an estimate for the fractional solid angle of each bin is given as,

$$\frac{\langle \Delta\tilde{\Omega} \rangle}{\Omega} = \frac{2RR}{N_r(N_r-1)} \qquad \text{(eq. 13)}$$

if the definition for $\langle \Delta\Omega \rangle$ is taken to be (i) on page 16. Otherwise, it can be estimated as

$$\frac{\langle \Delta\tilde{\Omega} \rangle}{\Omega} = \frac{DR}{NN_r} \qquad \text{(eq. 14)}$$

if the definition for $\langle \Delta\Omega \rangle$ is taken to be (ii) on page 16, and N still defines the number of galaxies in the data set.

Defining the Estimators

Depending on whether eq. 13 or eq. 14 is substituted into eq. 11 as an estimate for $\langle \Delta\Omega \rangle$, two distinct estimators are defined, which will be labelled w_1 and w_2 respectively (the dependence on θ is assumed). A third estimator which was introduced by Landy and Szalay (1993) will also be defined here, and will be labelled w_3. It will be discussed in detail in §2.3. To simplify notation, the following terms are defined,

$$dd = \frac{2DD}{N(N-1)} \qquad dr = \frac{DR}{NN_r} \qquad rr = \frac{2RR}{N_r(N_r-1)} \qquad \text{(eq. 15)}$$

such that these terms represent the probabilities of finding a neighbouring galaxy in an annular bin for each of the three correlations. They will be referred to henceforth as the fundamental pair probabilities, or simply as the pair probabilities. The three estimators can then be expressed as follows,

$$1 + w_1 = \frac{dd}{rr} \qquad \text{(eq. 16)}$$

$$1 + w_2 = \frac{dd}{dr} \qquad \text{(eq. 17)}$$

$$1 + w_3 = \frac{dd - 2dr + 2rr}{rr} \qquad \text{(eq. 18)}$$

These estimators are simple combinations of the three fundamental pair probabilities *dd, dr,* and *rr.* (Note that *dr* could be replaced by *rd*, since it does not matter whether the data are used as centers, or the random points are used as centers, the number of separations in each bin is the same.) It is therefore the behaviour of these quantities that should be analyzed and understood. It is then easy to see why the estimators above (and any other estimator that is composed of combinations of these quantities) give different results in different circumstances and possess different levels of statistical noise (see §2.3 for details).

For example, a practice that was first suggested by Sharp (1979), expanded on by Hewett (1982), and utilized by Koo and Szalay (1984), Neuschaefer et al. (1991), and Infante and Pritchet (1994) is the use of the cross correlation w_{rd} to estimate the correlations that are induced in the data by the interaction between the large scale density gradients and the boundaries without actually modelling these gradients. This value is subtracted from the angular correlation of the data w_{dd} to give a *corrected* estimate of the angular correlation. However, different estimators were used by different authors (Koo

and Szalay [1984], and Neuschaefer et al. [1991] used w_1, while Infante and Pritchet [1994] used w_2) so it is possible to see if different corrections were made for the different estimators. It is trivial to see that the estimate for w_{rd} is the same regardless of what estimator is used. Simply replace *dd* by *dr*, *dr* by *rr*, and keep *rr* the same in equations 16 to 18, and the result is *dr/rr -1* as a correction for w_{dd} in all cases. It was also suggested by Hewett (1982) that w_{dr} be subtracted from w_{dd} as a correction, but this was not implemented by any of the authors. The estimates for w_{dr} are not the same for all three estimators, but rather are *dr/rr-1, 0, (rr-dr)/rr* for estimators w_1,w_2, and w_3 respectively. This can be seen by making the substitution of *dd* by *dr* in equations 16 to 18, and keeping the other values fixed. Since Infante and Pritchet (1994) use the w_2 estimator, making the correction for w_{dr} would have no effect but it would double the correction made by the other two authors. This simple example serves two purposes. Firstly, it shows the insight that is developed by analyzing the estimators in terms of their fundamental pair probabilities, and secondly it shows that in practice $w_{ab} \neq w_{ba}$, even though by their definitions given in equations 4 and 5, they should be equal.

For sufficiently simple boundary conditions, such as a circular boundary, the edge corrections for an estimator can be calculated analytically; this will be done in Chapter Three for *rr*, and in Appendix B for *dr*. However, due to the intuitive nature of the notation developed in this section, it will be used to describe the estimators regardless of how the edge corrections are calculated. Therefore, *rr* will always represent $\langle \Delta\Omega \rangle / \Omega$ as calculated using definition (i) on page 16, and *dr* will always represent $\langle \Delta\Omega \rangle / \Omega$ as calculated using definition (ii) on page 16.

2.3 Variance of the Estimators

The variance in the estimate for w was shown by Peebles (1980, §48) to be approximately inversely proportional to the number of independent pairs in the bin (*i.e.* Poisson errors), for the regime of weak clustering. The error estimate given by

$$\delta w = \sqrt{\frac{1+w}{DD}} \qquad \text{(eq. 19)}$$

is almost ubiquitous throughout the literature (this will be referred to as either the canonical estimate or the Poisson error throughout this chapter), and it was suggested by Sharp (1979) that this is even a slight overestimate for the errors. This conclusion, however, is not supported by the results in this section nor the results of Ling et al. (1986), Barrow et al. (1984), Bhavsar (1990), who used bootstrap resampling to estimate the error. In all cases, the estimate given by eq. 19 is a lower limit to the error determined numerically. Bootstrap resampling was also recently used by Brainerd et al. (1994), but a comparison to the expected Poisson errors was not made. Landy and Szalay (1993) showed that for a Poisson point process (see Appendix A for definition), the error in w is dependent on which estimator is used. They developed an estimator (w_3) which has approximately Poisson variance (for the weak clustering regime), and showed that the estimators w_1 and w_2 possess variances that are in excess of the Poisson variance. This result is corroborated in this section, but it is also shown that when significant large scale density gradients exist in the data, the variance is significantly in excess of the canonical estimate (*i.e.* the square of eq. 19) *for all three estimators.*

In the rest of this section, the variance of the three estimators w_1, w_2, and w_3 will be analyzed using the variance of the three fundamental pair probabilities *dd, dr,* and *rr*, as seen through their error propagation into the estimators. Monte Carlo simulations will be performed both for objects randomly distributed in a circular sample space and for objects distributed pseudo-randomly with a significant large scale density gradient in a circular sample space. The importance of studying the estimators by analyzing the fundamental quantities was first emphasized by Landy and Szalay (1993), and much of their methodology has been implemented here.

Notation and Terminology

For any given outcome from the ensemble of random and data sets created, the values of the pair probabilities can be expressed as,

$$dd = \langle dd \rangle (1 + \alpha) \qquad \text{(eq. 20)}$$

$$dr = \langle dr \rangle (1 + \beta) \qquad \text{(eq. 21)}$$

$$rr = \langle rr \rangle (1 + \gamma) \qquad \text{(eq. 22)}$$

such that α, β, and γ are the fluctuations about the mean, and <...> indicates an average over the ensemble of experiments. By definition $<\alpha> = <\beta> = <\gamma> = 0$, and thus, the variance of these quantities are simply $<\alpha^2>$, $<\beta^2>$, and $<\gamma^2>$ respectively. Since *rr* is completely determined by the geometry of the sample space, γ can be made arbitrarily small by distributing an arbitrarily large number of random objects in the sample space. It is shown in §3.2 (eq. 46 on page 53) how *rr* (equivalent to f_{S11} in §3.2) can be calculated analytically for a circular sample space, and it is this 'exact' form that will be used throughout this section. Likewise, the variance in *dr* will be due entirely to fluctuations in the data set since the number of random objects can be chosen arbitrarily large. Again, an analytic method to calculate *dr* for a given data set has been developed in Appendix B (eq. 51 on page 108), and it is this 'exact' form that will be used throughout this section. Therefore, only an ensemble of data sets need to be created, while the need for random sets has been removed. It is therefore guaranteed that all the fluctuation in any of the quantities is completely due to the fluctuations in the data. Also, an enormous reduction in computer time has been gained by using the analytic methods, since each data set would need to be cross correlated with an ensemble of random sets in order to determine the fluctuations in the quantity *dr*, and each analytic cross correlation is on the order of 500 times faster than the Monte Carlo method.

Using a 2nd order Taylor expansion, it is straightforward to show that the variance in the three estimators can be written as

$$var(w_1) = \frac{\langle dd \rangle^2}{\langle rr \rangle^2} \langle \alpha^2 \rangle \qquad \text{(eq. 23)}$$

$$var(w_2) = \frac{\langle dd \rangle^2}{\langle dr \rangle^2} \{ \langle \alpha^2 \rangle + \langle \beta^2 \rangle - 2\langle \alpha\beta \rangle \} \qquad \text{(eq. 24)}$$

$$var(w_3) = \frac{\langle dd \rangle^2}{\langle rr \rangle^2} \langle \alpha^2 \rangle + 4\frac{\langle dr \rangle^2}{\langle rr \rangle^2} \langle \beta^2 \rangle - 4\frac{\langle dd \rangle \langle dr \rangle}{\langle rr \rangle^2} \langle \alpha\beta \rangle \qquad \text{(eq. 25)}$$

where the $<\alpha\beta>$ term is the covariance between α and β. Note that these results are different from those of Landy and Szalay (1993), since these are exact to second order, whereas their variances are calculated assuming $<dr> = <rr>$, and $<dd> \sim <rr>$ (*i.e.* small correlation amplitude). Neither of these assumptions are true for our purposes, since $<dr>$ can vary substantially from $<rr>$ when large scale density gradients are present in the data; this is due to the different manner in which gradients interact with the boundaries (see §3.2 of Chapter Three). Furthermore, $<dd>$ is close to $<rr>$ only in the weak clustering regime.The variance and covariance of α and β are simply the normalized variance and covariance of the pair probabilities *dd* and *dr*,

$$var(\alpha) = \langle \alpha^2 \rangle = \frac{\langle dd^2 \rangle - \langle dd \rangle^2}{\langle dd \rangle^2} = \frac{var(dd)}{\langle dd \rangle^2} \qquad \text{(eq. 26)}$$

$$var(\beta) = \langle \beta^2 \rangle = \frac{\langle dr^2 \rangle - \langle dr \rangle^2}{\langle dr \rangle^2} = \frac{var(dr)}{\langle dr \rangle^2} \qquad \text{(eq. 27)}$$

$$cov(\alpha, \beta) = \langle \alpha\beta \rangle = \frac{\langle dd \cdot dr \rangle - \langle dd \rangle \langle dr \rangle}{\langle dd \rangle \langle dr \rangle} = \frac{cov(dd, dr)}{\langle dd \rangle \langle dr \rangle} \qquad \text{(eq. 28)}$$

The variance of the estimators becomes different when $<\alpha^2>$ is not the dominant term in equations 23 to 25. Landy and Szalay (1993) developed analytic approximations for $<\alpha^2>$, $<\beta^2>$, and $<\alpha\beta>$ under the conditions of weak or no clustering, and thus were able to judiciously choose a combination of the pair probabilities that minimizes the variance and provides an unbiased estimate of *w*. By keeping track of $<\alpha^2>$, $<\beta^2>$, and $<\alpha\beta>$ in the following simulations, it is possible to determine how these quantities contribute to the variance of the three estimators for different bins and different density gradients.

Simulations

(i) Poisson Point Process Simulations (*i.e.* no density gradient)

In these simulations, each experiment consisted of the following steps:

(i) 400 galaxies were randomly distributed in a circular sample space.

(ii) The pair separations were calculated and placed into 8 logarithmic bins similar to the binning of the cluster data in Chapter Four, thus producing *DD*. (See Table 6, §4.1.)

(iii) The quantity *dd* for each bin was calculated using eq. 15, and *dr* and *rr* were calculated using the analytic methods described earlier. The three estimators w_1, w_2, and w_3 were calculated using equations 16 to 18.

This experiment was repeated 500 times, thus producing 500 outcomes of the quantities *DD*, *dd*, *dr*, w_1, w_2, and w_3. From these quantities, $<DD>$, $<\alpha^2>$, $<\beta^2>$, $<\alpha\beta>$, $var(w_1)$, $var(w_2)$, and $var(w_3)$ were calculated. The results for $<\alpha^2>$, $<\beta^2>$, and $<\alpha\beta>$ are presented in Table 1 for each of the 8 bins. Notice that in the first few bins, $<\alpha^2>$ is the dominant term by several orders of magnitude, and in the rest of the bins, $<\alpha^2>$, $<\beta^2>$, and $<\alpha\beta>$ have comparable values. The variances calculated for w_1, w_2, and w_3, are presented in Table 2, along with the canonical estimates given by the square of eq. 19.

The variances as calculated directly were found to agree to within 1% of the estimates derived from the 2nd order Taylor expansion in equations 23 to 25. In order to better appreciate the difference between the calculated variance, and the canonical variance, see Figure 1 on page 27, which shows the percent excess for each of the 8 bins. Notice that the first three bins show very little excess since $<\alpha^2>$ is the dominant term. The rest of the bins are dramatically different, since $<\alpha^2>$, $<\beta^2>$, and $<\alpha\beta>$ are all of comparable magnitude. The estimator w_3 shows no more than a 5% excess in any of the bins as predicted by Landy and Szalay (1993), while the other two estimators show strong excess for the bins representing large separations. It is uncertain whether the dependence of the excess variance on the bin number is a result of the larger separation, or the larger bin width for these bins (due to the log binning). This could be determined by performing the same experiment with linear binning, but this was not followed up.

Table 1 : Poisson Process - Fundamental Normalized Variances

bin	$<\alpha^2>$	$<\beta^2>$	$<\alpha\beta>$
1	0.1185	0.5168E-05	0.2942E-04
2	0.2509E-01	0.1078E-04	-.2368E-04
3	0.5805E-02	0.2149E-04	0.4815E-04
4	0.1434E-02	0.4393E-04	0.9797E-04
5	0.6971E-03	0.9147E-04	0.1845E-03
6	0.7818E-03	0.1744E-03	0.3475E-03
7	0.7436E-03	0.1823E-03	0.3631E-03
8	0.8390E-03	0.2122E-03	0.4186E-03

(ii) Large Scale Density Gradient Simulations

In these simulations, the experiments were identical to those described in part (i), except that objects were distributed in the circle according to a joint probability density function which describes the large scale density gradient for the sample space. A modelling technique which is described in detail in Chapter Four was used to estimate this probability density function from the discrete distribution of the galaxies in the Hercules

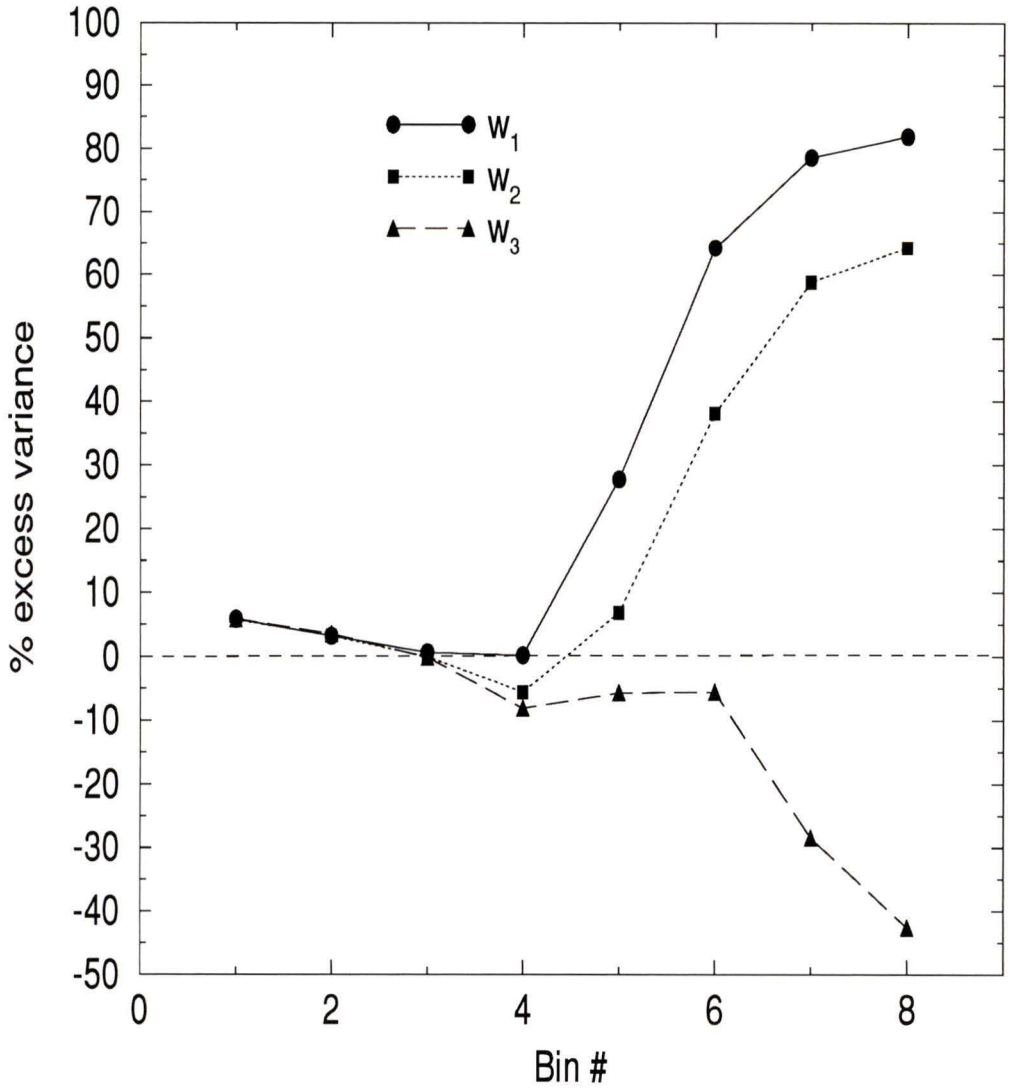


Figure 1 :Excess variance for random data sets calculated for 8 logarithmic bins. The % excess refers to the excess over the variance that is expected from the standard formulae given in eq. 19, as calculated from Monte Carlo simulations of randomly placed objects constrained within a circular boundary.

Table 2 : Poisson Process - Estimator Variances

bin	var(w_1)	(1+w_1)/DD	var(w2)	(1+w_2)/DD	var(w3)	(1+w_3)/DD
1	0.3408	0.3209	0.3404	0.3207	0.3406	0.3205
2	0.1587	0.1537	0.1587	0.1536	0.1591	0.1535
3	0.0764	0.0759	0.0758	0.0758	0.0757	0.0758
4	0.0378	0.0377	0.0357	0.0377	0.0348	0.0377
5	0.0264	0.0191	0.0205	0.0191	0.0180	0.0190
6	0.0280	0.0100	0.0162	0.0100	0.0095	0.0100
7	0.0273	0.0058	0.0142	0.0058	0.0045	0.0058
8	0.0290	0.0052	0.0146	0.0052	0.0037	0.0052

cluster (also presented in Chapter Four). The pdf is presented as a two dimensional surface in Figure 17 on page 74.

The results for $<\alpha^2>$, $<\beta^2>$, and $<\alpha\beta>$ are presented in Table 3 for each of the 8 bins. Notice that in all but the last few bins, $<\alpha^2>$ is the dominant term by several orders of magnitude, and in the last few bins, $<\beta^2>$, and $<\alpha\beta>$ are still an order of magnitude smaller. The variances calculated for w_1, w_2, and w_3, are presented in Table 4, along with the canonical estimates given by the square of eq. 19. The variances as calculated directly were again found to agree to within 1% of the estimates derived from the 2^{nd} order Taylor expansion in equations 23 to 25. The percent excess of the calculated variance over the canonical variance is presented in Figure 2 on page 30. Notice the increase in the excess variance with the bin's central separation as observed in figure 1. Also, notice the fairly dramatic increase in the excess variance, as compared to the Poisson simulations, for all but the last three bins. This is undoubtably caused by the significant increase in $<\alpha^2>$ as compared to the Poisson point process simulations, since the values for $<\beta^2>$, and $<\alpha\beta>$ only increased marginally for the two types of simulations. Also, the differences in the

variance for each of the three estimators is negligible for all but the last three bins due to the dominance by $<\alpha^2>$.

Table 3 : Large Scale Density Gradient - Fundamental Normalized Variances

bin	$<\alpha^2>$	$<\beta^2>$	$<\alpha\beta>$
1	0.6023	0.1130E-04	0.1124E-04
2	0.1490	0.2338E-04	0.9278E-05
3	0.4611E-01	0.4903E-04	0.1513E-03
4	0.1835E-01	0.1055E-03	0.2437E-03
5	0.1115E-01	0.2293E-03	0.5817E-03
6	0.5667E-02	0.4521E-03	0.7383E-03
7	0.3042E-02	0.4861E-03	0.9604E-03
8	0.4189E-02	0.5531E-03	0.1125E-02

Table 4 : Large Scale Density Gradient - Estimator Variances

bin	var(w_1)	(1+w_1)/DD	var(w_2)	(1+w_2)/DD	var(w_3)	(1+w_3)/DD
1	1.1860	0.8043	1.1809	0.8026	1.1860	0.8019
2	0.2752	0.1845	0.2728	0.1837	0.2753	0.1834
3	0.0855	0.0450	0.0838	0.0446	0.0849	0.0445
4	0.0342	0.0111	0.0328	0.0110	0.0333	0.0109
5	0.0197	0.0028	0.0175	0.0028	0.0175	0.0028
6	0.0081	0.0008	0.0064	0.0008	0.0064	0.0008
7	0.0030	0.0003	0.0016	0.0003	0.0011	0.0003
8	0.0036	0.0002	0.0022	0.0002	0.0016	0.0002

Conclusions

(i) Poisson Point Process

(1) For the first 4 bins, the dominant source of variance is from $<\alpha^2>$, and thus all three estimators have nearly identical variances.

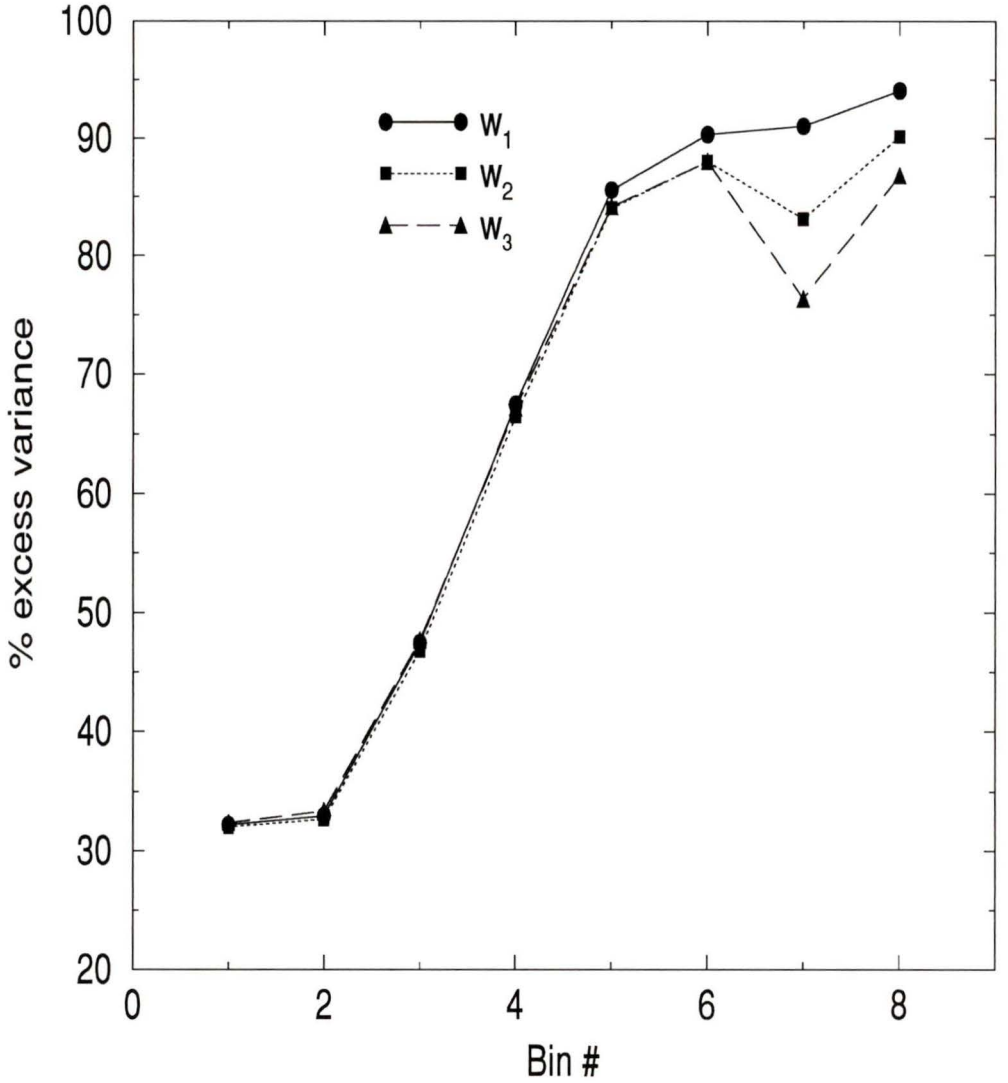


Figure 2 :Excess variance for pseudo-random data sets created using the density profile determined for the Hercules cluster using the adaptive kernel technique with *k=10*. Refer to Figure 1 on page 27 for a description.

(2) w_3 possesses significantly smaller variance than w_2 and w_1 for the bins at larger θ and essentially has Poissonian variance for all the bins; therefore w_3 is the preferred estimator if the larger separations are of interest. (A similar conclusion was found by Landy and Szalay 1993.)

(ii) Large Scale Density Gradient

(1) $<\alpha^2>$ is dominant for all but the last three bins and thus all three estimators have nearly identical variances for all but these bins.

(2) Since none of the estimators have a variance that is Poissonian, the variance must be determined numerically for all three estimators, making any of these estimators acceptable from this perspective.

Chapter Three

The Theory of Correlation Functions

Correlation functions, or variants of them, provide us with a powerful tool for analyzing the distribution of observational data, especially if the structure that exists in the data is complex. This tends to be true when studying the distribution of galaxies in space, where the auto correlation functions such as $w(\theta)$ and $\xi(\mathbf{r})$ have proven to be extremely valuable, as evidenced by the discussion in Chapter Two. It is important to develop practical methods for estimating these functions, as was done in Chapter Two. However, in order to properly interpret these correlation functions and understand their limitations, it is crucial to understand first the theory behind them, to which this chapter is dedicated. Since this investigation is primarily concerned with the angular distribution of galaxies, the focus will be on correlation functions in two dimensions, and in particular, on the auto correlation function $w(\theta)$.

A 'descriptive' statistical analysis of observational data, as presented in Chapter Two, is always the most objective since it involves the least number of assumptions about the data. However, such an analysis does not always allow for a complete understanding of how systematic effects can bias the results. Therefore, §3.1, and §3.2 are dedicated to understanding how boundaries, large scale density gradients and the interaction between them affect the distribution of pairs, as described by $w(\theta)$. In this chapter, the distribution of galaxies will be modelled by a continuous probability density function (pdf), since it offers a more intuitive understanding of how systematic effects contribute to the distribution of pairs. The angular correlation function, $w(\theta)$, then becomes the auto correlation function of the pdf. This type of modelling has proven to be excessively fruitful in the past (Limber 1953, 1954, 1957; Layzer 1956; Peebles 1980, §38). The ideas in this chapter directly related to auto correlation functions of continuous real-valued functions have been culled mainly from the work of Bracewell (1965).

3.1 Correlation Functions in One Dimension

For the purposes of studying the distribution of galaxies on the sky, one needs to use correlation functions in two dimensions. However, it is easier to first study them in one dimension to elucidate their basic properties, and to understand how they can be employed in other dimensions.

Theory of Correlation Functions

The auto correlation function of a real-valued function $f_1(x)$ is defined as

$$\int_{-\infty}^{\infty} f_1(u) f_1(u-r)\, du \qquad \text{(eq. 29)}$$

which will be represented symbolically as $f_1 \bullet f_1(r)$ throughout this paper. In words, one can think of this function as follows:

i) Create two identical copies of the function $f_1(x)$.

ii) Shift one of the functions by r in the positive direction.

iii) At every position, multiply these two functions together and add up the results for every position along the axis.

iv) the sum found in iii) represents the value of the autocorrelation function at r.

It is graphically described in figure 3.

Two simple properties of the autocorrelation function are given without proof, but are intuitively obvious from an analysis of figure 3.

i) $f_1 \bullet f_1(r)$ is an even function. That is, if the function is shifted in the positive or negative direction by r, the same answer is obtained, which can be seen by making the change of variables $w=u-r$ in eq. 29. This can also be seen by observing that, in figure 3, the shaded area remains the same regardless of whether: (a) the dashed curve remains

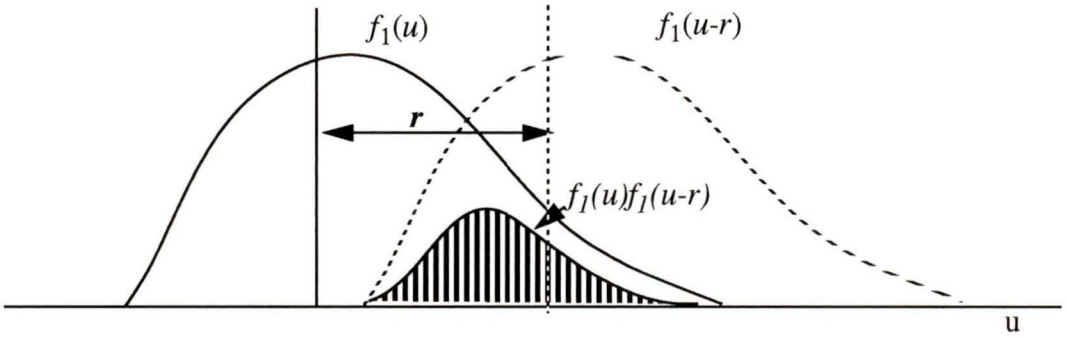


Figure 3 :The auto correlation function of f at a position r, represented by the shaded area under the curve labelled $f_1(u)f_1(u-r)$.

adapted from Bracewell (1965)

stationary, and the solid curve is shifted by $-r$, or, (b) the solid curve remains stationary, and the dashed curve is shifted by $+r$.

ii) $f_1 \bullet f_1(r)$ reaches a maximum at the origin *(i.e.* $r = 0$). That is, when $r = 0$, the auto correlation function is simply equal to the area under the function $f_1(x)$; when $r \neq 0$, an identical copy is shifted, and the auto correlation, represented by the shaded area in figure 3, must be less than or equal to the area under the function $f_1(x)$. (See Bracewell 1965, pg 48 for a proof.) Note that the auto correlation of a function at a non-zero shift r, can be equal to the auto correlation at $r = 0$ only if there is perfect periodicity in the function, as is the case for a sinusoidal curve.

The auto correlation function can be used to analyze many different types of observational data, but in this section it will be used to analyze how objects are distributed in one dimension, and in §3.2 it will be used to analyze how objects are distributed in two dimensions. Imagine that objects are distributed along the x axis, such that the position of some object is a random variable X with a probability density function (pdf) $f_1(x)$. Then the probability that X will have a value in the interval between $x=a$, and $x=b$, where $a \leq b$, $\{a, b\} \in \Re$, is given by

$$P(a \le X \le b) = \int_a^b f_1(x)\,dx. \qquad \text{(eq. 30)}$$

Now create a new random variable R, defined as $r = x_i - x_j$ (*i.e.* the distance between two objects), then the probability density function for this random variable is proportional to the auto correlation function of $f_1(x)$. Therefore, the probability that R, the separation between two objects, will have a value between $r=a$, and $r=b$, where $a \le b$, $\{a, b\} \in \Re$, is given by

$$P(a \le R \le b) \propto \int_a^b f_1 \bullet f_1(r)\,dr\,. \qquad \text{(eq. 31)}$$

This can be understood by referring to figure 3 on page 34, and drawing a thin vertical slice of infinitesimal thickness du anywhere along the u axis. Since the one probability density function is shifted by r, the probability that there is an object in the interval du at $x=u$, and also another object in the interval du centered a distance r away at x=u-r, is simply $f_1(u)f_1(u-r)\,du$. By adding up these probabilities for all values of u, a measure of the probability that any two objects are separated by r is obtained. Suppose that this probability density function is normalized by the probability that any two objects will be separated by some value of r between 0 and ∞ (which of course must be the case if only values $r \ge 0$ are considered). Then the normalized probability density function $f_{11}(r)$ can be expressed as

$$f_{11}(r) = \frac{f_1 \bullet f_1(r)}{\int_0^\infty f_1 \bullet f_1(r)\,dr} = \frac{\int_{-\infty}^{\infty} f_1(u)f_1(u-r)\,du}{\int_{-\infty}^{\infty}\int_0^\infty f_1(u)f_1(u-r)\,du\,dr}, \qquad \text{(eq. 32)}$$

such that eq. 31 becomes

$$P(a \le R \le b) = \int_a^b f_{11}(r)\, dr. \qquad \text{(eq. 33)}$$

In practice, however, one does not deal with the entire one dimensional space (or in general, the n-dimensional space), but rather some subsection of it which will be denoted as S, commonly referred to as the sample space. For instance, in one dimension one may be dealing with a line segment with lower and upper bounds defined by x_l and x_u respectively, such that the function suddenly drops to zero outside this segment. A function of this nature will be denoted as $f_{S1}(x)$, and the autocorrelation function as $f_{S11}(r)$, as shown graphically in figure 4 on page 37, and defined by

$$f_{S11}(r) = \frac{f_{S1} \bullet f_{S1}(r)}{\int_0^\infty f_{S1} \bullet f_{S1}(r)\, dr} = \frac{\int_{-\infty}^{\infty} f_{S1}(u) f_{S1}(u-r)\, du}{\int_0^\infty \int_{-\infty}^{\infty} f_{S1}(u) f_{S1}(u-r)\, du dr}$$

$$= \frac{\int_{(x_l+r)}^{x_u} f_1(u) f_1(u-r)\, du}{(x_u - x_l) \int_0^{x_u} \int_{(x_l+r)}^{x_u} f_1(u) f_1(u-r)\, du dr} \qquad \text{(eq. 34)}$$

where the limits of the integration were determined from an inspection of figure 4.

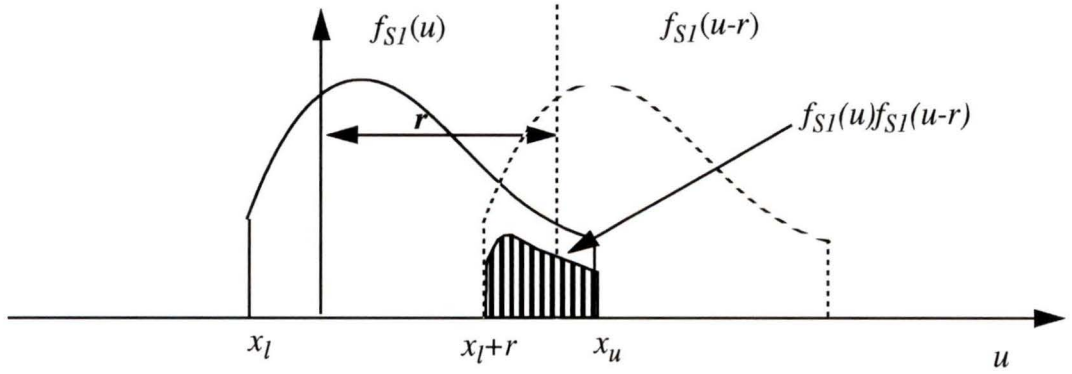


Figure 4 :The auto correlation function of f_1, constrained to some region S defined by the lower and upper bounds x_l and x_u respectively. The limits of the integration can be determined from this diagram.

So, when dealing with a finite sample space, as is always the case in practice, the same pdf is used as for the case with an infinite sample space, but the limits of the integration are changed. This is a crucial point to comprehend, as it implies that the correlation function that we observe for a finite sample space is partially determined by the distribution function $f_1(x)$, and partially determined by the geometry of the sample space. It is therefore necessary to remove the contribution of the geometry in order to see what the intrinsic correlation function is due to $f_1(x)$ alone. More details on how this can be done will be presented in §3.2

The cross-correlation function of two real-valued functions $f_1(x)$ and $f_2(x)$ is likewise defined as

$$f_1 \bullet f_2(r) = \int_{-\infty}^{\infty} f_1(u) f_2(u-r)\, du. \qquad \text{(eq. 35)}$$

This can be thought of graphically in the same way as figure 3 on page 34; however, the two functions are different, and it is the second function that is shifted before

multiplication. The two simple properties defined for the auto correlation function on page 33 do not hold for the cross correlation function,

i) It can be seen that $f_1 \bullet f_2(-r) = f_2 \bullet f_1(r) \neq f_1 \bullet f_2(r)$ by making the substitution $w=u-r$ in eq. 35. In general, $f_1 \bullet f_2(r)$ is not an even function unless both f_1 and f_2 are even functions.

ii) In general $f_1 \bullet f_2(r)$ does not reach a maximum at the origin.

The probability density function $f_{12}(r)$ is likewise defined such that $f_{12}(r)dr$ defines the probability that the separation between an object distributed according to $f_1(x)$ and an object distributed according to $f_2(x)$ will be in the interval dr centered on r,

$$f_{12}(r) = \frac{f_1 \bullet f_2(r)}{\int_{-\infty}^{\infty} f_1 \bullet f_2(r)\, dr} = \frac{\int_{-\infty}^{\infty} f_1(u) f_2(u-r)\, du}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(u) f_2(u-r)\, du dr} \qquad \text{(eq. 36)}$$

Following this notation, a general cross correlation function $f_{ij}(r)$ can be defined such that if $i=1$, $j=1$, then f_{ij} is the auto correlation function; if $i=1$, $j=2$, then f_{ij} is the cross correlation function where f_2 is shifted; and if $i=2$, $j=1$, then f_{ij} is the cross correlation function where f_1 is shifted. Here $f_{ij}(r)$ is defined as

$$f_{ij}(r) = \frac{f_i \bullet f_j(r)}{\int_{-\infty}^{\infty} f_i \bullet f_j(r)\, dr} = \frac{\int_{-\infty}^{\infty} f_i(u) f_j(u-r)\, du}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_i(u) f_j(u-r)\, du dr}. \qquad \text{(eq. 37)}$$

3.2 Correlation Functions in Two Dimensions

In this section, the focus will be on the angular distribution of galaxies on the sky constrained to some region of the sky, which will be referred to as the sample space S. Correlation functions in two dimensions are identical to correlation functions in one dimension, except that now the integral is over a plane rather than a line. The distribution of galaxies in the sample space must therefore be described by a two dimensional joint probability density function $f_{SI}(x,y)$, and the auto correlation of this function will also be a function of two variables (*e.g.* r and ϕ). However, since it is generally the separation of galaxies that is of interest, and not the position angle, the azimuthal component is averaged out. Recall from Chapter Two, that for the discrete case, the angular correlation function was calculated from the distribution of galaxies in a sample space simply by calculating the separations between all galaxies without regard to the position angles between them. Therefore, the angular correlation function, for the discrete case, tacitly assumed azimuthal symmetry; likewise, it will be assumed that all references to correlation functions in the continuous regime refer to azimuthally averaged correlation functions.

Therefore, the auto correlation function in two dimensions should be thought of as follows:

(i) Make two identical copies of the pdf $f_{SI}(x,y)$.

(ii) Shift one of the functions by r in any direction.

(iii) At every position in the plane, multiply these two functions together, and add up the results.

(iv) Rotate the copied function by $d\phi$, and perform step (iii).

(v) Repeat step (iv) until the original position angle is obtained. The sum of all these results represents the azimuthally averaged auto correlation function of $f_{SI}(x,y)$, which is a function of only one variable r, denoted as $f_{SII}(r)$.

Assume $f_{S2}(x,y)$ is the joint pdf describing the distribution of galaxies in the sample space, and $f_{S1}(x,y)$ is a flat joint pdf (*i.e.* constant probability over the whole sample space). Then the auto and cross-correlation functions $f_{S22}(\theta)$, $f_{S21}(\theta)$, and $f_{S11}(\theta)$ are connected to the pair probabilities *dd*, *dr* and *rr* (defined on page 19 by eq. 15 in Chapter Two) respectively by simple integrals over the same bins for which the discrete quantities were defined. Notice that the variable r was replaced by the variable θ in order to emphasize that the separations between galaxies on the sky are angular. Therefore, the three estimators defined in Chapter Two (equations 16, 17, and 18) can be expressed for the continuous case as,

$$1 + w_1(\theta) = \frac{f_{S22}(\theta)}{f_{S11}(\theta)} \qquad \text{(eq. 38)}$$

$$1 + w_2(\theta) = \frac{f_{S22}(\theta)}{f_{S21}(\theta)} \qquad \text{(eq. 39)}$$

$$1 + w_3(\theta) = \frac{f_{S22}(\theta) - 2f_{S21}(\theta) + 2f_{S11}(\theta)}{f_{S11}(\theta)} \qquad \text{(eq. 40)}$$

where the integral of the pdfs over the bin is assumed.

In order to understand and interpret the angular correlation function properly, it is important to recognize the factors that influence it. There are essentially three quantities that affect the distribution of pairs in a finite sample space: namely, boundaries, large scale density gradients (*i.e.* gradients that are on the same scale as the dimension of the sample space), and small scale density gradients (*i.e.* structure on scales that are small compared to the scale of the sample space)

(i) Boundaries

If galaxies are randomly distributed in a circular sample space with diameter D, then it is obvious that the largest separation that one can observe in the data is D. Furthermore, as one approaches smaller and smaller separations, one should see bins with more and more pairs (assuming that the annular area of the bin remains the same), since the smaller separations will be less affected by the boundary. See figure 5.

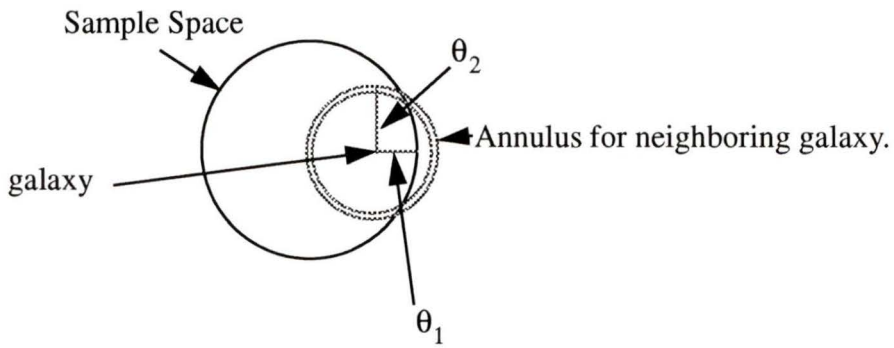


Figure 5 :Boundary effects on the distribution of pairs. Neighboring galaxies that lie within the annulus are defined to have a separation of $\theta_2 \pm \Delta\theta/2$, and thus if part of the annulus lies outside of the boundary (*i.e.* $\theta_2 + \Delta\theta/2 > \theta_1$, where θ_1 is the distance that the central galaxy is away from the edge), then the area available, and thus the probability for a neighboring galaxy to exist at $\theta_2 \pm \Delta\theta/2$, is decreased.

(ii) Large Scale Density Gradients

Large scale density gradients refer to smoothly varying gradients which are of the order of the size of the sample space. These gradients play a significant role in determining the distribution of pairs on all scales up to the size of the sample space. For example, the radial density fall off in a cluster of galaxies produces more pairs at small separations than the same number of galaxies would produce if they were randomly distributed simply because more galaxies are packed into the central region.

Since large scale density gradients are easily modelled by low order functions, these functions can be treated as pdfs and the autocorrelation of these functions used to predict the probability of having a pair of galaxies separated by a given angle. This will be demonstrated for circularly symmetric functions such as King model profiles, and de Vaucouleur profiles confined to a circular sample space. (These are analytic expressions used to model the radial density fall-off of clusters - Dressler [1978], and Beers and Tonry [1986] for King and de Vaucouleurs profiles respectively.)

(iii) Small Scale Density Gradients

Small scale density gradients refer to clustering in which the structures are small compared to the size of the sample space. For example, a small clump of galaxies whose dimension is much smaller than the dimension of the sample space will contribute to the distribution of pairs both through the autocorrelation of its own distribution of galaxies, and through the cross correlation of the clump galaxies with other galaxies in the sample space. However, if there is no structure on scales larger than the scale of the clump (*i.e.* there are several clumps of galaxies randomly distributed in the sample space), then the distribution of pairs on these larger scales will appear random.

Although these three different factors have been listed separately, they are intimately connected, and it is the interaction between these factors that presents the most problems when interpreting the data (*e.g.* the interaction between large scale gradients and boundaries is widely known to cause serious bias in the estimate of $w(\theta)$ - Sharp 1979, Infante and Pritchet 1994). It is the goal of this work to disentangle the relative contributions of these factors and to try and isolate the effect of small scale structure.

There is always an ambiguity when one defines densities for a discrete data set, in that a scale must always be defined, rather arbitrarily, over which objects are counted. If the chosen scale is too large, then structure is missed on the smaller scales. If the chosen scale is too small, then one approaches the discrete limit in which every cell has either 1 or 0 objects; in this case, Poisson noise dominates the signal rendering the density essentially

meaningless. Large scale density gradients can be defined sufficiently well by counting objects in relatively large regions (compared to the size of the sample space), but small scale density gradients must be defined by counting objects in relatively small regions. Since clusters of galaxies have at most 300-400 galaxies in the sample space, only the large scale density gradients will be sufficiently defined, whereas the small scale density gradients will be dominated by noise. It is therefore difficult to describe structure on small scales by a density. This is one of the reasons (in addition to the obvious complexity of the distribution of galaxies on small scales which is apparent from a visual inspection) that the statistic $w(\theta)$ was developed. This statistic is superior to a simple density for small scales since the structure on these scales is averaged over the whole sample space, thus increasing the signal to noise ratio (S/N). The statistic $w(\theta)$ is able to pick out clustering on small scales (if it is common throughout the sample space) which would otherwise pass as Poisson noise if analyzed on an object-by-object basis.

Modelling the Radial Density Fall-off in Clusters

In this section, the interaction between large scale density gradients and boundaries will be analyzed in order determine their effect on the distribution of pairs. This will be done by calculating auto and cross correlation functions of joint pdfs constrained to a finite sample space by using the methodology developed in §3.1. For simplicity, the joint pdfs are assumed to be circularly symmetric, and the sample space is assumed to have a circular boundary centered on the pdf. The formulae given in equations 38, 39, and 40 will then be used to see how the different estimators of $w(\theta)$ respond to these different conditions.

The position of an object in the plane of the sky is described by two random variables labelled Θ and Φ (using polar coordinates, where Θ describes the radial distribution, and Φ describes the azimuthal distribution), with values θ and ϕ distributed according to the joint pdf $f_S(\theta, \phi)$. The formulae of spherical trigonometry can be replaced by the formulae of Cartesian trigonometry when the arcs on the sphere are small enough that the cubes of their values can be neglected; this approximation produces a maximum error of $1''$ for the largest separations analyzed. Therefore, Cartesian formulae are used, but, the

radial distance is still referred to as an angle θ, and the azimuthal angle as ϕ. The boundary of the sample space is defined by a circle centered on the circularly symmetric pdf. The general two dimensional correlation function in polar coordinates for a pdf constrained to a sample space S is defined as

$$f_{Sij}(\theta, \phi) = \frac{f_{Si} \bullet f_{Sj}}{\int_0^{2\pi} \int_0^{\infty} f_{Si} \bullet f_{Sj} \theta d\theta d\phi} \qquad \text{(eq. 41)}$$

$$= \frac{\int_0^{2\pi} \int_0^{\infty} f_{Si}(\vartheta, \varphi) f_{Sj}(\vartheta_j, \varphi_j)\, \vartheta d\vartheta d\varphi}{\int_0^{2\pi} \int_0^{\infty} \int_0^{2\pi} \int_0^{\infty} f_{Si}(\vartheta, \varphi) f_{Sj}(\vartheta_j, \varphi_j)\, \vartheta d\vartheta d\varphi \theta d\theta d\phi}, \qquad \text{(eq. 42)}$$

where ϑ_j and φ_j must be determined as a function of the shift (θ, ϕ), and the position of the integrating variables (ϑ, φ). Since the boundary is circular, say of radius θ_b, and centered on the circularly symmetric pdf, f_{Si} can be defined as

$$f_{Si}(\theta, \phi) = \begin{cases} f_i(\theta, \phi) & \theta \le \theta_b \\ 0 & \theta > \theta_b \end{cases}, \qquad \text{(eq. 43)}$$

where $f_i(\theta, \phi) = f_i(\theta, \phi + \Delta\phi) = f_i(\theta)$. The integral of f_{Si} in eq. 41 can be converted to an integral of f_i by changing the limits of the integration, as demonstrated for one dimension in eq. 34, where the limits of the integration are determined from the geometry of the sample space. In 1-d, the limits of the integration are quite easy to derive, since any sample space is defined simply by a lower and upper bound. In 2-d, however, the geometries can be much more complex, and thus the limits of the integration equally complex. Fortunately, with a circular sample space, the limits can be determined by

straightforward geometrical arguments. To determine the limits, the overlapping region for two identical circles of radius θ_b must be analyzed for various shifts in θ and ϕ. To further simplify the problem, only shifts in θ need to be considered, keeping $\phi = 0$, since the joint pdfs that are used are circularly symmetric. The geometry of the problem is presented in figure 6, for a radial shift of θ, and an azimuthal shift of 0.

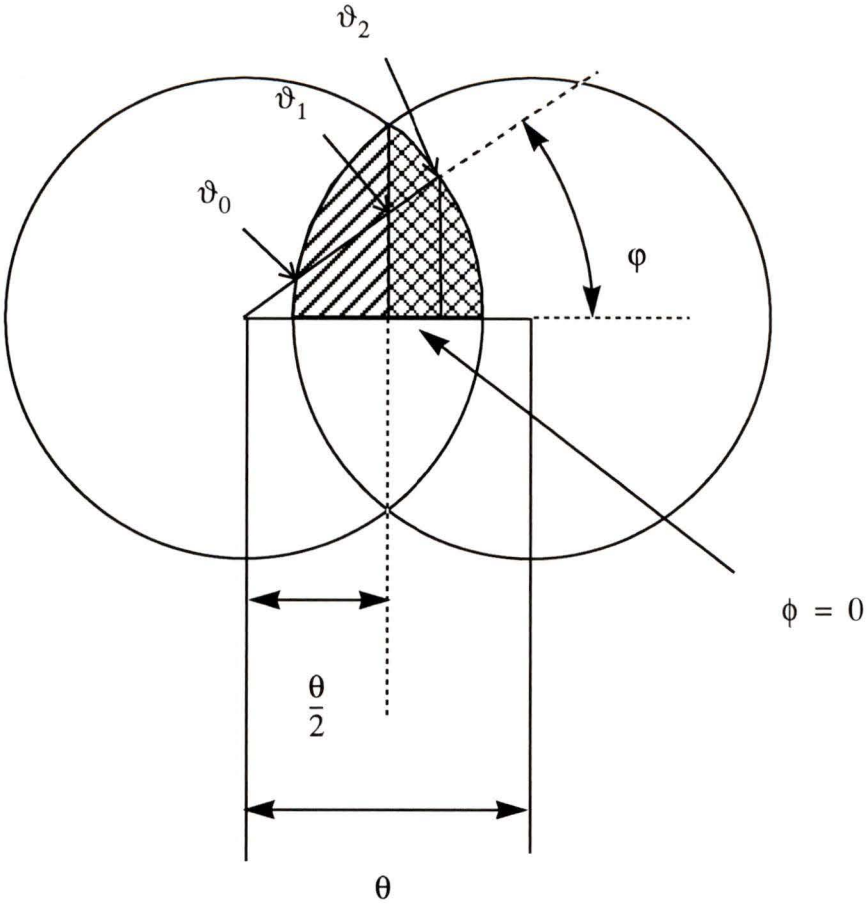


Figure 6 :Geometry of overlapping circular sample spaces. The one circle is shifted to the right by θ, with no azimuthal shift. The integral must be calculated for the overlapping region, but due to the symmetry, only the upper half needs to be integrated.An arbitrary line of integration is drawn for the dummy variable φ, which shows the key values of ϑ that must be known as a function of φ.

The overlapping region in figure 6 is the region that the integral must be calculated over, but since the joint pdfs are circularly symmetric, the integral needs only to be calculated over the upper half (which is hatched) and then multiplied by 2. If auto

correlation functions alone were being calculated, then the integral would need only to be extended over one quarter of the region, which is double hatched. Since cross-correlation functions must also be calculated, the pdfs in each of the circles will not be the same, and the integral must be computed over the entire upper half. As it turns out, given two pdfs, f_{Si}, and f_{Sj}, it is easier to calculate the cross correlation function by first shifting f_{Si} to the right and integrating over the upper right quadrant, and then shifting f_{Sj} to the right and integrating over the upper right quadrant, and finally adding the two results together. It is therefore necessary to determine the limits of φ for the upper right quadrant, and the values of ϑ_1 and ϑ_2 as a function of φ. This can easily be done by referring to figure 7, which is a magnification of the triangle in figure 6.

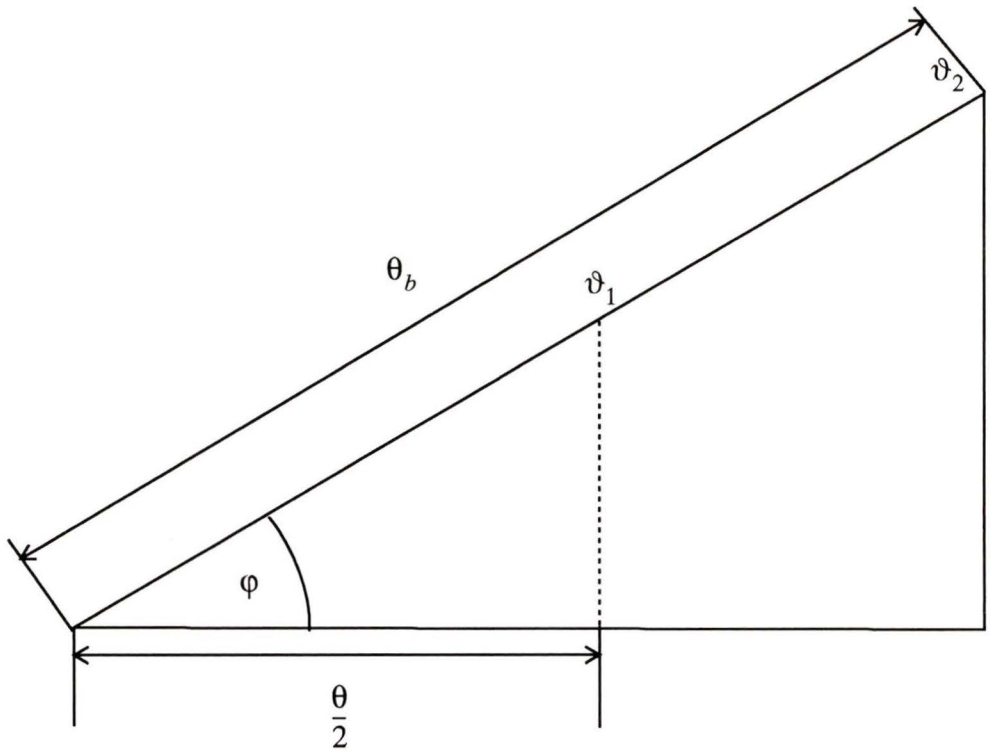


Figure 7 :Limits of 2-d integration. One can see that $\vartheta_2 = \theta_b$ for all values of φ, and that $\vartheta_1 = \frac{\theta}{2} \cdot \frac{1}{\cos\varphi}$.

Figure 7 illustrates how the limits of the integration in ϑ are related to φ, and thus the limits of φ itself need to be determined. From figure 6, the lower limit is 0, and thus $\varphi_1 = 0$, and the upper limit occurs at the upper intersection of the two circles, which produces the same triangle as in figure 7, with the exception that ϑ_1 is replaced by θ_b, and thus the relation $\varphi_2 = \text{acos}\left(\frac{\theta}{2\theta_b}\right)$ is obtained. Now that the limits of the integration have been obtained, it must be determined how (ϑ_j, φ_j) are related to the other variables (*i.e.* the arguments to use for the shifted pdf). Since the functions are circularly symmetric, only ϑ_j needs to be determined. From figure 8, it can see that

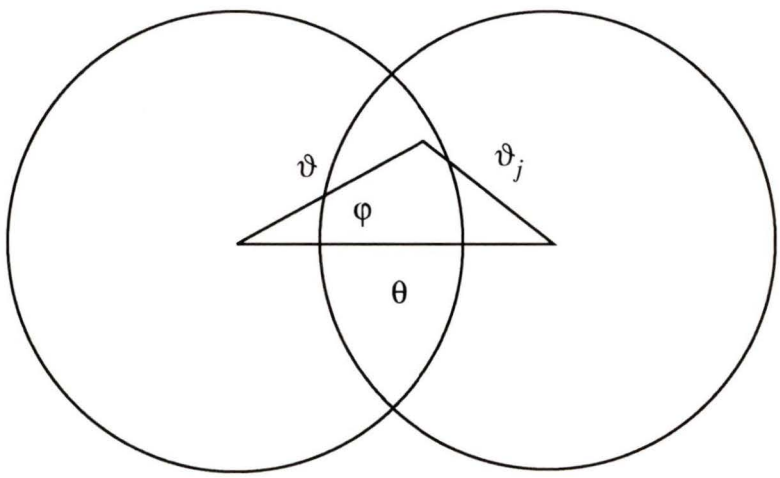


Figure 8 :Determining the argument of the shifted pdf. From the cosine law,

$$\vartheta_j^2 = \theta^2 + \vartheta^2 - 2\vartheta\theta\cos(\varphi)$$

$\vartheta_j^2 = \theta^2 + \vartheta^2 - 2\vartheta\theta\cos(\varphi)$, through the cosine law. Therefore, eq. 41 reduces to

$$f_{Sij}(\theta, \phi = 0) = \frac{\int_0^{\operatorname{acos}\left(\frac{\theta}{2\theta_b}\right)} \int_{\frac{\theta}{2\cos(\varphi)}}^{\theta_b} f_i(\vartheta) f_j(\vartheta_j)\, \vartheta d\vartheta d\varphi}{\int_0^{2\theta_b} \int_0^{\operatorname{acos}\left(\frac{\theta}{2\theta_b}\right)} \int_{\frac{\theta}{2\cos(\varphi)}}^{\theta_b} f_i(\vartheta) f_j(\vartheta_j)\, \vartheta d\vartheta d\varphi \theta d\theta}$$

$$+ \frac{\int_0^{\operatorname{acos}\left(\frac{\theta}{2\theta_b}\right)} \int_{\frac{\theta}{2\cos(\varphi)}}^{\theta_b} f_j(\vartheta) f_i(\vartheta_j)\, \vartheta d\vartheta d\varphi}{\int_0^{2\theta_b} \int_0^{\operatorname{acos}\left(\frac{\theta}{2\theta_b}\right)} \int_{\frac{\theta}{2\cos(\varphi)}}^{\theta_b} f_j(\vartheta) f_i(\vartheta_j)\, \vartheta d\vartheta d\varphi \theta d\theta}, \qquad \text{(eq. 44)}$$

where $\vartheta_j^2 = \theta^2 + \vartheta^2 - 2\vartheta\theta\cos(\varphi)$. Now one can simply substitute circularly symmetric functions for the pdfs in eq. 44 to calculate their auto and cross correlation functions in a circular sample space of radius θ_b.

In the past, cluster profiles have been primarily modelled by King model profiles (Dressler 1978), and de Vaucouleurs' $r^{1/4}$ profiles (Beers and Tonry 1986). These will be investigated along with a flat profile, which would simply show the effects of the boundaries on the distribution of pairs (*i.e.* the objects are randomly distributed in the sample space). These profiles are represented analytically as

$$f_i(\theta) = \begin{cases} 1 & \text{flat} \\ \left(1+\left(\frac{\theta}{a}\right)^2\right)^{-1} & \text{King} \\ exp\left[-7.67 \cdot \left(\frac{\theta}{a}\right)^{1/4}\right] & \text{de Vaucouleurs} \end{cases} \qquad \text{(eq. 45)}$$

where the scaling parameter a for the King and de Vaucouleurs model must be determined, as well as θ_b, the radius of the circular sample space. The fundamental quantity is the ratio of a to θ_b, denoted as α, which determines the shape of the correlation functions, while their absolute values determine the scale on which the separations are measured. Since in Chapter Four, the real cluster data is bounded by a circle with a linear projected radius of 0.75 h^{-1} Mpc, this will be used for θ_b. Dressler (1978) derives a mean core radius of 0.25 h^{-1} Mpc for 12 rich clusters, and thus this value of a will be used for the King model profile, which reduces to $\alpha = 1/3$ for the unitless argument θ/θ_b. Beers and Tonry (1986) observed the cores of rich clusters, and found that $a \approx 3\ h^{-1}$ Mpc provided a good fit for the de Vaucouleurs profile, and thus $\alpha = 4$ will be used for this profile. It should be stressed that neither of these values provide a good fit to *all* clusters, but they are adequate for the illustrative purposes of this analysis.

These three profiles are presented in figure 9 as a function of the unitless argument θ/θ_b. They have been normalized to have the same values at the edge of the sample space (*i.e.* where $\theta/\theta_b = 1.0$). Notice that de Vaucouleurs' profile predicts a cusp near the center of the cluster, whereas the King profile predicts a somewhat rounded flat core. A maximum value of 2 has been plotted in this figure but the de Vaucouleurs profile rises up to a density of 22.6 before θ/θ_b reaches 0. Some clusters seem to show this cusp (*e.g.* Abell 2029, and Abell 154 - Dressler 1978), whereas others do not show it as clearly; the autocorrelation of both of these functions will therefore be analyzed.

Eq.44 was numerically integrated across 10 mutually exclusive bins using the pdfs given in eq. 45. The results of this integration represent the probability of finding a pair of

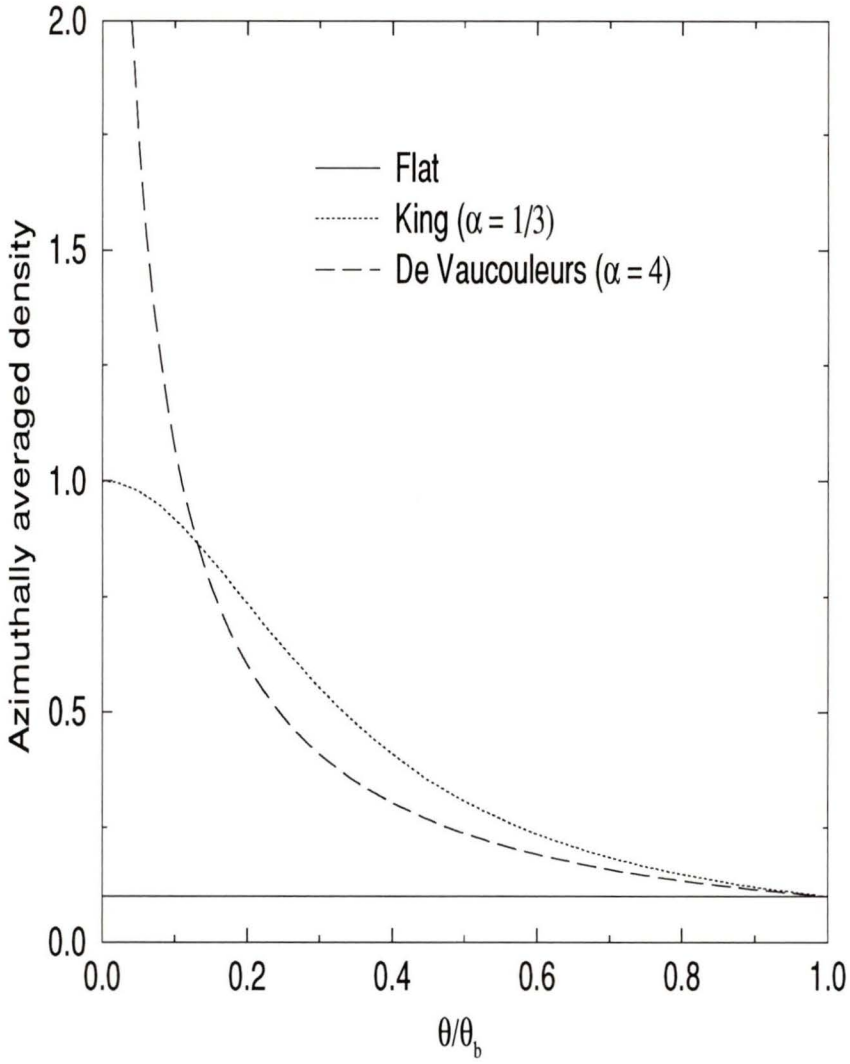


Figure 9 :pdfs used to model the density gradient produced by the cluster profiles.

galaxies with a separation in the bin. However, since the bins are not of equal area, the probability for each bin has been divided by the area of the bin and renormalized. The final result represents the probability per unit area for each bin, and is displayed graphically in figure 10. To show that these values do actually represent the probability of finding a pair of galaxies in the bin, three simple numerical simulations were performed. In these simulations, 400 galaxies were distributed in a circle, and the separations between all galaxies were calculated, totalling $\frac{400 \times 399}{2} = 79800$ separations.

In the first simulation, the objects were distributed randomly; in the second simulation the objects were distributed with a King profile; and in the last simulation, the objects were distributed with a de Vaucouleurs profile. The scaling parameter α was chosen to be identical to that of the profiles used in the numerical integration. The separations were placed in bins of the same dimension as the numerical integration. The experiment was performed 20 times and the mean and standard deviation were calculated for each bin. These values are also plotted in figure 10 and show excellent agreement with the numerical integrations. The error bars could be made arbitrarily small by simply using more galaxies or performing more experiments. The decrease in the probability per unit area with increasing separation for the flat profile is simply a result of the circular boundary, but for the other two profiles, the distribution of pairs is also affected by the density gradient. The King and de Vaucouleurs profiles show an almost identical distribution of pairs at large separations, but deviate at smaller separations, where the de Vaucouleurs profile predicts more pairs in the bin of smallest separations, as one would expect from the central cusp in the profile. The advantage of estimating the number of pairs analytically is that the results are essentially error free (in a statistical sense), and the results are obtained in a few seconds; on the other hand, the simulations described above require about 20 minutes on the same machine, and they possess statistical noise.

As mentioned on page 40, the integral of $f_{SII}(\theta)$ over a bin defined by some interval in θ is equivalent to the quantity *rr,* defined in Chapter Two (eq. 15 on page 19), evaluated over the same bin. These quantities essentially define the estimation for the edge correction $\langle\Omega\rangle/\Omega$ defined according to definition (i) on page 16. The term *rr* is the

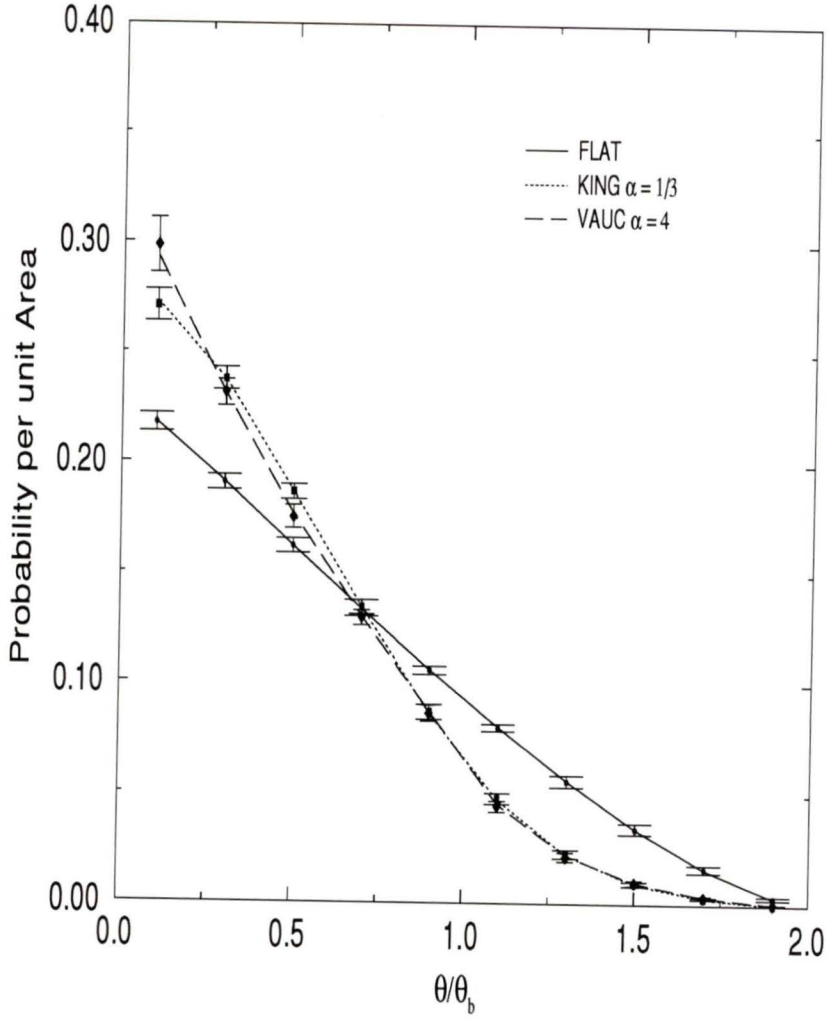


Figure 10 :Auto correlation of the pdfs integrated over 10 bins, and divided by the area of the bin. This represents the probability per unit area of finding a pair of galaxies with a separation in the bin.

Monte Carlo estimate defined in eq. 13 on page 19, and thus a very large number of random points (and a lot of cpu time) must be used to minimize the statistical noise, whereas f_{S11} is essentially error free. Therefore f_{S11} will be used to make edge corrections whenever the angular correlation function is calculated according to w_1 or w_3 (w_2 does not use *rr*). By integrating eq. 44 analytically, a simplified form of $f_{S11}(\theta)$ (eq. 46) is obtained which simply needs to be numerically integrated across the bin to give an exact value for *rr*. Notice that the normalization has been changed from an integral starting at 0 to an integral starting at θ_{res}; this is because, in general, the data will have some resolution limit below which it will not be possible to distinguish two objects as separate entities. Therefore, θ_{res} would also be used as the lower limit of integration for a bin if its value is larger than the lower limit of that bin.

$$f_{S11}(\theta) = \frac{\frac{\theta_b^2}{2}\cdot \operatorname{acos}\left(\frac{\theta}{2\theta_b}\right) - \left(\frac{\theta}{4}\cdot\left(\theta_b^2 - \frac{\theta^2}{4}\right)\right)}{\int_{\theta_{res}}^{2\theta_b} \frac{\theta_b^2}{2}\cdot \operatorname{acos}\left(\frac{\theta}{2\theta_b}\right) - \left(\frac{\theta}{4}\cdot\left(\theta_b^2 - \frac{\theta^2}{4}\right)\right)\theta d\theta}, \qquad \text{(eq. 46)}$$

where θ_{res} is the resolution limit for the data.

The quantity *dr* which is used to estimate $\langle\Omega\rangle/\Omega$ (defined according to definition (ii) on page 16) cannot be estimated using these techniques, since it requires the discrete distribution of the data galaxies to estimate its value. It is however estimated using an analytic technique described in Appendix B (eq. 51 on page 108).

Estimating *w*(θ)

Numerical integration of $f_{S11}(\theta)$, $f_{S21}(\theta)$, and $f_{S22}(\theta)$ was performed for the King and de Vaucouleurs profiles discussed earlier over 20 evenly spaced bins between 0 and θ_{max} for three different boundary conditions. $w_1(\theta)$, $w_2(\theta)$, and $w_3(\theta)$ were then calculated using eq. 38, eq. 39, and eq. 40 respectively. In all three scenarios, it was

assumed that the linear scaling parameter a remained constant, such that the physical dimension of the cluster profiles remained constant. Only the radius of the boundary circle changed: in the first scenario, it had a radius of 750 kpc, in the second, it had a radius of 1500 kpc, and in the final scenario, a radius of 3000 kpc. In essence, this inversely affects only the unitless scaling parameter $\alpha = a/\theta_b$ for the calculations; however, as θ_b was increased, the upper limit of the integration in θ (see eq. 44) was decreased such that the integration was done over the same linear distance. θ_{max} was chosen such that the linear distance for the integration was always 1500 kpc, which was the maximum separation for any two objects in scenario 1. In this way, scenario 1 represents the case where the density gradient is of the order of the size of sample space (i.e a large scale density gradient), and scenarios 2 and 3 represent cases where the density gradient is progressively on smaller scales. It is interesting to see how the three different estimators respond to density gradients on these three different scales. The results of these calculations are shown in figure 11 on page 55, and figure 12 on page 56, for a King profile and a de Vaucouleurs profile respectively. In both cases, the left graphs a, c and e represent the results of the numerical integration of eq. 44 for $f_{S11}(\theta)$, $f_{S21}(\theta)$, and $f_{S22}(\theta)$. The right graphs b, d, and f represent the calculations of the three different estimators (see eqs. 38 - 40). The top graphs are the results for the largest scale density gradient, and the middle and bottom graphs are the results for the smaller scale density gradients.

The results are similar for the King profile and de Vaucouleurs profile, except for the slight excess of very close pairs predicted for the de Vaucouleurs profile due to its central cusp. The focus of the following discussion will therefore concentrate the results for the King profile. There are several interesting aspects to point out.

(i)There is a direct relationship between the values of the auto/cross-correlation function and the estimates of the angular correlation function. Recall that the only difference between the estimators is the way in which $f_{S11}(\theta)$ and $f_{S21}(\theta)$ are used to normalize $f_{S22}(\theta)$. These results are obvious from an inspection of equations 38, 39 and

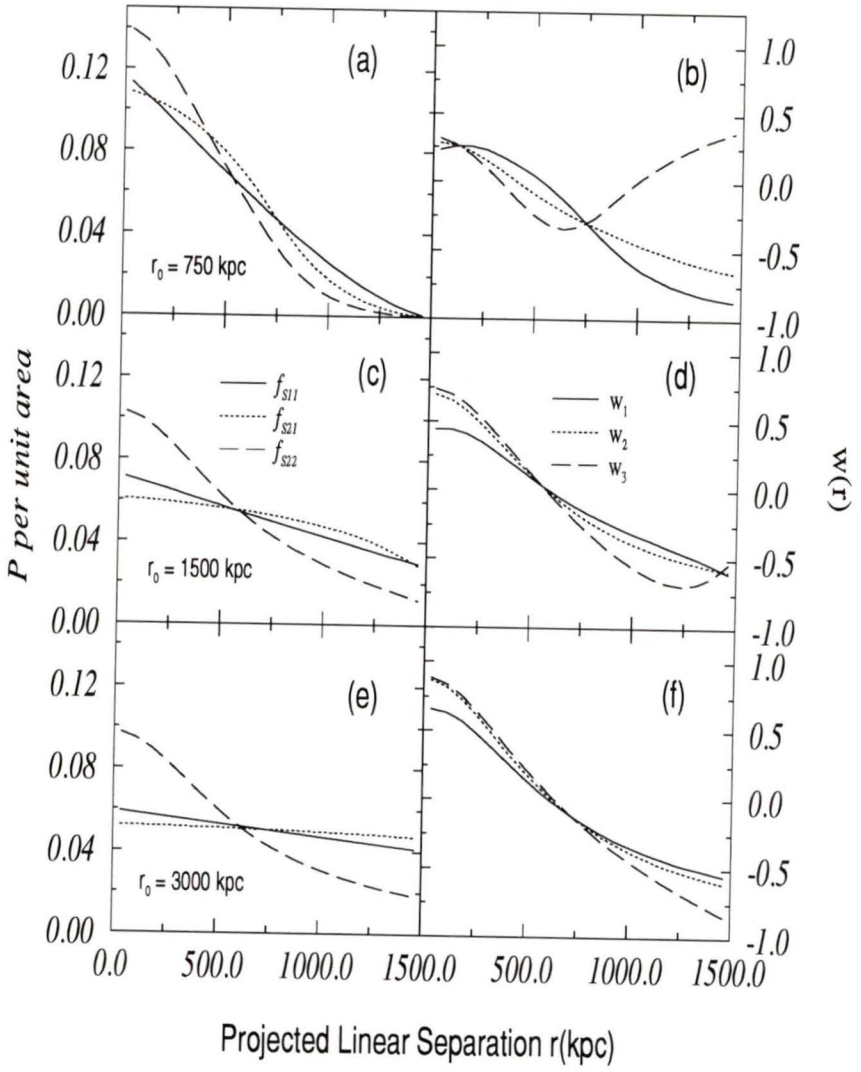


Figure 11 :Angular correlation induced by a King profile with a core radius a=250 kpc constrained to a circular sample space of radius r_0.

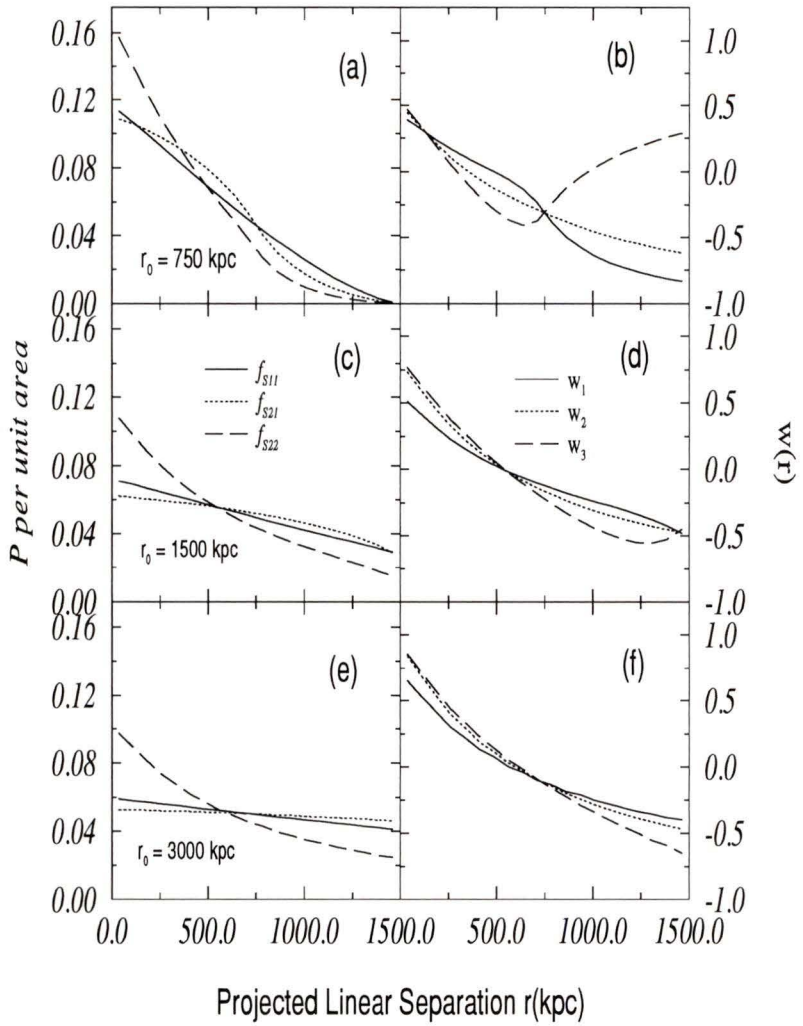


Figure 12 :Angular correlation induced by a de Vaucouleur profile with scale length a=3000 kpc constrained to a circular sample space of radius r_0.

40., if one substitutes $f_{S21}(\theta) = f_{S11}(\theta) \pm \varepsilon$ for cases a) and c) respectively, and $f_{S21}(\theta) = f_{S11}(\theta)$ for case b).

(a) When $f_{S21}(\theta) > f_{S11}(\theta)$, then $w_1 > w_2 > w_3$

(b) When $f_{S21}(\theta) = f_{S11}(\theta)$, then $w_1 = w_2 = w_3$

(c) When $f_{S21}(\theta) < f_{S11}(\theta)$, then $w_1 < w_2 < w_3$

(ii) The scatter between the three different estimators becomes progressively smaller as the density gradient scale decreases. In fact, when the radius of the boundary circle is increased such that it is several orders of magnitude larger than the scaling length, the estimators give nearly identical results.

(iii) For all scales of the density gradient, the cross correlation $f_{S21}(\theta)$, which is used to normalize $w_2(\theta)$, always flattens out as the separation approaches 0. The same behaviour is seen for $f_{S22}(\theta)$, but $f_{S11}(\theta)$ continues to rise. This is due to the fact that $f_{S11}(\theta)$ is affected by the circular boundary similarly on all scales due to the even distribution of objects within the circle; in contrast, $f_{S21}(\theta)$, and $f_{S22}(\theta)$ are less affected by the boundaries on small scales due to the concentration of objects near the center of the circle. For this reason, it would seem that $f_{S21}(\theta)$ is the more appropriate function for normalizing $f_{S22}(\theta)$, since it is affected by the boundary conditions in the same way as $f_{S22}(\theta)$.

(iv) At small separations, the values of $w_2(\theta)$, and $w_3(\theta)$ are very similar, but diverge as the size of the separation increases. This effect is most dramatic for the larger density gradients, and implies that $w_3(\theta)$ is not an appropriate estimator when large scale density gradients exist due to the spurious correlation induced at large separation.

The results of this analysis seem to indicate that it may be possible to distinguish between correlations induced by large scale density gradients and those induced by small scale density gradients by observing the scatter among the different estimators. It is evident that large scale density gradients in a finite sample space affect the distribution of pairs on all scales, and thus if the distribution of close pairs in clusters is to be analyzed,

then these contributions must be removed. A technique to do this will be developed and utilized in Chapter Four to study the distribution of close pairs in ten local rich clusters.

Chapter Four

Statistical Analysis of Close Pairs in Rich Clusters

In this chapter, the excess number of close pairs (defined to have separations ≤ 20 h^{-1} kpc) will be calculated for 10 nearby clusters; the angular correlation function will be used to calculate the excess number of pairs by first calculating the 'real' number of close pairs for the cluster, and then calculating the 'expected' number of close pairs from the cluster density profile. The cluster density profiles will be determined using an adaptive kernel estimator, which creates a continuous pdf from the discrete distribution of the cluster's galaxies. The scientific results of this analysis will be discussed, and future work will be described.

The estimator w_2 will be used for two reasons. Firstly, as shown in Chapter Two, all of the estimators have nearly identical variance for small angular separations in a cluster environment. Therefore, due to the relative complexity of w_3 compared to w_1 and w_2, it is not the preferred estimator; this is in contrast to the situation for weakly clustered data in which the large angular separations are of interest. In addition, w_3 produces spurious power in the correlation function at large θ when a large scale density gradient exist (see figures 11 and 12). Secondly, the estimator w_2 is preferred over w_1, since it correctly determines what the edge corrections should be when a large scale density gradient exists within data, as is the case in a cluster environment. In a cluster environment, where there is a higher density of galaxies in the center of the sample space, w_1 will over-estimate the edge correction for small separations, and thus under-estimate the correlation at small separations as was shown in figure 11 on page 55, and figure 12 on page 56.

The ambiguity that exists when using the angular correlation function to study the distribution of close pairs in a cluster environment is two-fold. Firstly, as shown in Chapter Two, there is not an analytic expression for the statistical uncertainty in each of the bins of the correlation function, and thus the errors must be determined numerically. Secondly, as shown in Chapter Three, the large scale density gradients present in a cluster environment affect the distribution of pairs on all scales, and this effect must be removed

before determining if there is an excess of close pairs from the small scale clustering alone. This can be thought of as looking for sub-clustering superimposed on a large scale density gradient.

The method that is developed in this chapter simultaneously removes both of these ambiguities and can be summarized as follows:

(i) The angular correlation function is estimated for the galaxies in the cluster using the w_2 estimator as described in Chapter Two.

(ii) The density profile of the cluster is estimated using an adaptive kernel estimator.

(iii) An ensemble of galaxy distributions is produced using the density estimated in step (ii) as a joint pdf. The angular correlation is calculated for each outcome of the ensemble using the w_2 estimator, and the mean and standard deviation ($<w_2>$ and $\sigma(w_2)$) are calculated for the ensemble.

Each step is described in detail below.

By overplotting w_2 for the real cluster data and $\langle w_2 \rangle \pm \sigma(w_2)$ from the simulations, it is possible to assess whether w_2 from the real data can simply be regarded as a random fluctuation of the overall density profile, or whether there is a real excess of close pairs superimposed on the overall density profile.

4.1 Estimating the Angular Correlation Function for the Clusters

In order to study local clusters of varying stages of evolution, and to test the method developed in this chapter, ten rich local clusters of galaxies have been analyzed. The positional data for the galaxies in these clusters was taken from Butcher and Oemler (1985). All the clusters are in the redshift range $0 < z < 0.1$ (redshifts taken from Sarazin, Rood, and Struble 1982). They have been imaged, by Butcher and Oemler (1985), down to a limiting magnitude J_{lim} (IIIa-J emulsion behind a Wratten 4 filter) between 16 and 19 (depending on the redshift), and out to a linear projected radius of approximately 0.75 h^{-1} Mpc. The resolution limit for separating two galaxy images was approximately 10

arcseconds, which corresponds to a linear projected resolution limit between 2.4 and 10.9 h^{-1} kpc. The clusters Abell 262, Abell 1367, and Hercules are irregular clusters with large populations of spiral galaxies, while the remaining clusters are compact, with few spirals in their core (see Appendix C for a short description of different cluster classification schemes). The above information is presented in table 5, along with the galactic latitude, and a brief description of each cluster.

Table 5 : Cluster Properties

cluster	# galaxies	z	b	Sample Dimension (h^{-1}Mpc)	resolution limit (h^{-1}kpc)	J_{lim}	Description[a]
A262	38	0.0164	-25°	R < 0.532 = 37′	2.4	16.0	BMIIISRI
A400	82	0.0234	-45°	R < 0.71 = 35′	3.4	17.0	BMIISPC
A401	128	0.0748	-39°	R < 0.76 = 11′.5	10.9	18.75	BMISPC
A1367	143	0.0213	74°	R < 0.69 = 37′	3.09	17.0	BMIIISRI
Coma	256	0.0235	87°	R < 0.62 = 30′	3.41	17.5	BMIISPC
A1904	102	0.0714	62°	R < 0.73 = 11′.8	10.4	19.0	BMIISPC
Hercules	144	0.0371	44°	R < 0.78 = 24′.2	5.4	18.0	BMIIISRI
A2199	161	0.0305	43°	R < 0.80 = 30′	4.4	17.5	BMISPC
A2634	109	0.0322	-33°	R < 0.96 = 34′	4.7	17.5	BMISPC
A2670	126	0.0749	-69°	R < 0.87 = 13′.2	10.8	18.5	BMISPC

a. BM = Bautz-Morgan type, SR/P = spiral rich/poor, C/I = compact/irregular

The centers of the clusters were chosen by eye; the circular boundary defining the sample space was chosen to maximize the number of galaxies included in the sample, while at the same time to minimize the amount of 'dead' space (where no galaxies exist due to the area not being imaged). The linear projected distributions of the galaxy positions as well as the boundary used for each sample (as defined above) are presented in figure 13, and figure 14. The galaxies that have a close neighbor within 20 h^{-1} kpc have been flagged and are presented as filled circles, whereas all other galaxies are presented as open circles. Due to the low redshift of these clusters, the contamination by foreground and background galaxies is expected to be small (Butcher and Oemler 1985), and thus no correction has been made.

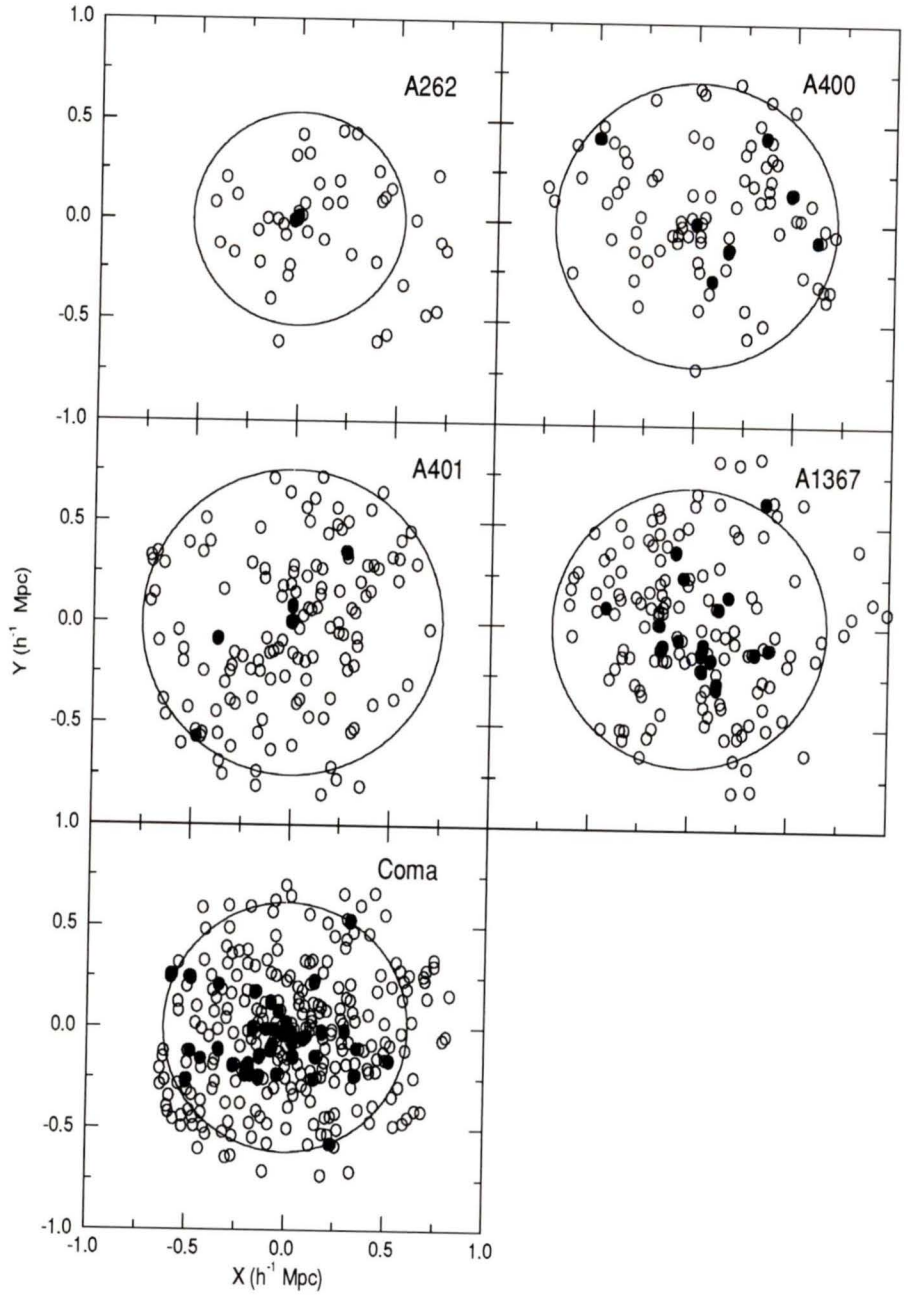


Figure 13 :Relative projected linear coordinates of galaxies converted using $H_0 = 100h$ km s^{-1} Mpc^{-1}. The open circles represent galaxies that do not have a neighbor within 20 h^{-1} kpc, and the filled circles represent galaxies that do. The large overlaid circle represents the boundaries that were used for each cluster.

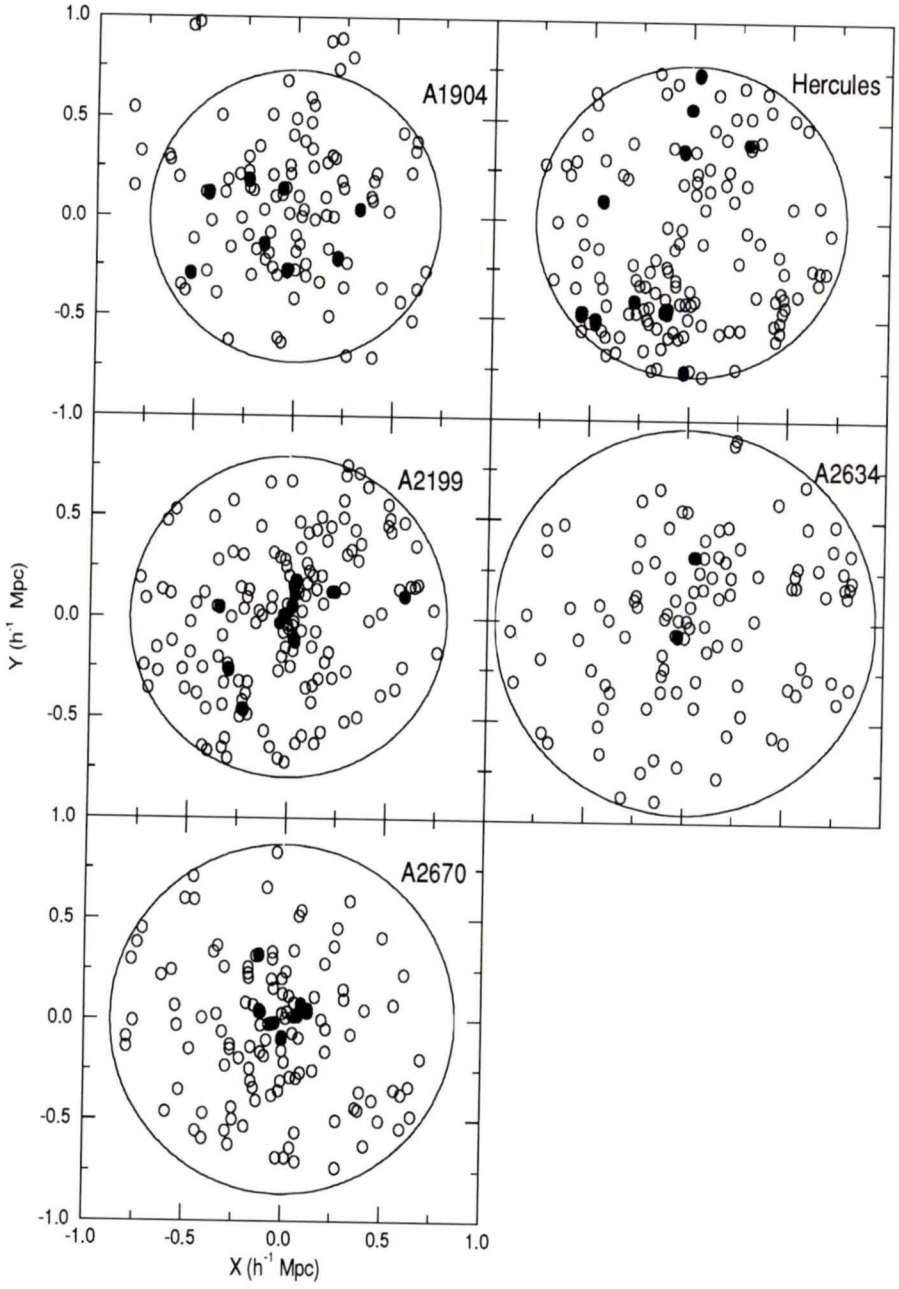


Figure 14 :Relative projected linear coordinates of cluster galaxies. As in figure 13 on page 62.

The binning system used for the angular separations consists of 8 logarithmic bins of constant width between 5 h^{-1} kpc and 1500 h^{-1} kpc. In this way, the first bin represents all separations between 5 and 10 h^{-1} kpc, and the second bin represents all separations between 10 and 20 h^{-1} kpc. The first two bins will be used to study the close pairs in the cluster, while the remaining bins will be used to study the correlations from the radial density fall off in the clusters. (These latter bins cover all separations out to the maximum separation, which is typically 2 times the radius of the sample space.) In table 6 on page 64, the bin number and the separations that they include is presented. Notice that the resolution limit for some of the clusters, as presented in table 5, is larger than the lower limit of the first (and sometimes second) bin. This implies that the probability of finding a pair of galaxies in one of these bins is reduced.This does not produce a bias in the estimation of w_2 since the quantities *dd* and *dr* are affected by the resolution limit in the same manner (see Appendix B).

Table 6 : Bin Separations

bin	log(h^{-1}kpc)		linear(h^{-1}kpc)	
	lower	upper	lower	upper
1	0.699	1.009	5	10.2
2	1.009	1.318	10.2	20.8
3	1.318	1.628	20.8	42.5
4	1.628	1.938	42.5	86.7
5	1.938	2.247	86.7	176.6
6	2.247	2.557	176.6	360.6
7	2.557	2.866	360.6	734.5
8	2.866	3.176	734.5	1500

The angular correlation for each of the clusters was calculated using the w_2 estimator defined on page 20, where *dd* was derived from the data and *dr* was calculated using eq. 51 on page 108 of Appendix B. These correlations are presented in figure 18 on page 76 and figure 19 on page 77 as filled black circles.

4.2 Estimating the Density Profile of the Clusters

The process of estimating a density from a discrete distribution of points is a crucial link between the world of observation and the world of theory. All observations of phenomena consist of measurements averaged over finite intervals in time or space and are presented as discrete values, while theory generally describes phenomenon through continuously varying functions. An excellent review of the different methods used for this type of estimation is given by Silverman (1986) in his book "Density Estimation for Statistics and Data Analysis".

The viewpoint taken for an analysis of this sort is that the data sample is a set of independent identically distributed data points. That is, they are drawn independently of one another from a common pdf (see §A1 in Appendix A). The challenge is to estimate the underlying pdf from a discrete set of realizations. There are two distinct approaches that can be taken, namely the parametric and non-parametric approaches. In the parametric approach, an analytic form is assumed for the pdf and a suitable statistical method (*i.e.* least squares, maximum likelihood,...) is used to estimate the parameters in the analytic model (*e.g.* μ and σ for a Gaussian distribution). In the non-parametric approach, no assumptions are made about the form of the pdf (except that the distribution has a pdf), and quite possibly, the pdf may not be analytically representable.

The non-parametric approach will be used here since there is no a priori reason to choose a particular parametric family. The King profile and the de Vaucouleurs profile have been used in the past by Dressler (1978) and Beers and Tonry (1986) respectively, but their results (especially the azimuthal symmetry inherent in these two profiles) indicate that neither form is appropriate for all clusters.

Non- parametric Density Estimators

Assume that there are N objects distributed in a two dimensional plane and let x_i be a two dimensional vector which describes the position of object i in the plane. The following approaches can be used to estimate the underlying pdf of these objects (for other methods, see Silverman 1986).

(i) Histogram Estimator

The most direct method to estimate the pdf is simply to bin the data into square cells with sides of length h and area h^2. An estimate for the pdf is then given by the histogram

$$\tilde{f(x)} = \frac{1}{N} \cdot \frac{\#\text{ of } x_i \text{ in the same bin as } x}{h^2}, \qquad \text{(eq. 47)}$$

where the parameter h determines how much the data will be smoothed out (often called the smoothing parameter).

Advantages

• It is trivial to calculate and thus it is an excellent estimator for simple applications such as just 'looking' at the data.

Disadvantages

• The data is binned and thus there are discontinuities at the edges of each bin, which can be a problem if derivatives of the pdf are required.

• More importantly, the histogram can look significantly different depending on where the edges of the bins fall.

• Since the smoothing parameter h is fixed, the statistical noise in each bin can vary significantly if there are long tails in the pdf such that certain regions are quite dense, while other regions are quite sparse. If h is chosen too small, then the dense regions may be well defined, but the sparse regions will be too noisy. If h is chosen too large, then the sparse regions may be well defined, but the dense regions will be oversmoothed, and potential structure in the pdf will be lost.

(ii) Kernel Estimator

A more powerful method for estimating the pdf is to use the kernel estimator which is defined as

$$\tilde{f(x)} = \frac{1}{Nh^2} \sum_{i=1}^{N} K\{\left(\frac{x - x_i}{h}\right)\}, \qquad \text{(eq. 48)}$$

where the kernel $K(x)$ has the property

$$\int_{R^2} K(x) = 1 \qquad R^2 \equiv \text{2-d space} \qquad \text{(eq. 49)}$$

and the smoothing parameter h defines the scale on which the data is smoothed.

In words, the density at some position x is defined by summing up the contributions from all the objects, where the contribution from object i is defined by the value of some normalized function centered at x_i and evaluated at x. If $K(x)$ is a symmetric bivariate Gaussian given by $K(x) = \exp\left(-x^2/2\right)/2\pi$, then the parameter h is simply the standard deviation of the Gaussian. To get an intuitive feel for the appearance of $\tilde{f(x)}$, one can imagine a Gaussian with $\sigma=h$ centered at the position of each galaxy, where overlapping areas of different Gaussians are summed.

From a practical point of view, a fine grid system (approximating a continuous surface) is constructed; for each cell in the grid, the distance to each galaxy is evaluated and passed to the kernel function, which in turn passes back the contribution to that cell from each galaxy. The sum of the contributions from all the galaxies is the density for that cell.

Advantages

• There is no binning involved, so that for a given h, the density does not vary depending on where the bin edges are located.

• For most functional forms of the kernel, the resulting pdf can be differentiated.

Disadvantages

• Again, since the smoothing parameter h is fixed, it may not be possible to define an acceptable density in all regions.

(iii) Adaptive Kernel Estimator

This method is identical to the kernel estimator except that the smoothing parameter h is allowed to vary from object to object. Therefore, there are two stages to this estimation. The first stage involves a 'rough' density estimate, such that the local density of each object is estimated. The next stage involves using the kernel estimator where h is determined as a function of the 'rough' density estimate for each object, such that h is large is low density regions, and small in high density regions.

The initial density estimate is usually determined using the kernel estimator where the smoothing parameter is set by the total number of objects in the sample space. It is also possible to determine the local density of each object by using the distance to the k^{th} closest neighbor. However, it has been found from personal experience, and from the work of Breiman et al. (1977) and Beers (1992) that the final density estimate is not strongly dependent (within reason) on the method used to determine the 'rough' estimate.

Advantages

• This technique has all the advantages listed for the kernel estimator, as well as the advantage that data can be smoothed on different scales for different density regions of the data; thus the disadvantages of having a fixed smoothing scale listed for the other two estimators are alleviated.

Disadvantages

- The relationship between h and the 'rough' density estimate must be chosen rather arbitrarily.

Bias and Variance in Density Estimation

When estimating a density, two important concepts to keep in mind are the bias and variance in the estimate. For the kernel estimator, the bias is a result of the fact that by representing each object by a kernel, it is no longer the underlying pdf that is being estimated, but rather the underlying pdf convolved with the kernel. A well known example would be trying to estimate the density for a set of objects that are distributed according to a Gaussian with standard deviation σ by using a Gaussian kernel estimator. The resulting estimate would produce a Gaussian density with a variance of $\sigma^2 + h^2$. Therefore, choosing the smoothing parameter h too large significantly biases the density estimate. The variance in the estimator is a result of the statistical noise inherent in trying to define a continuous function by a discrete set of points. By decreasing the smoothing parameter h, the density at each point in the plane is defined by fewer objects, and thus the statistical noise is increased.

Density Estimation for Clusters

The problems that arise when estimating densities for clusters of galaxies have been summarized by Beers (1992), who also elucidated the advantages of using the adaptive kernel estimator. Likewise for this work, the adaptive kernel estimator with a Gaussian kernel was used to estimate the density of each cluster; the 'rough' density estimate was determined by d_k, the distance to the k^{th} closest neighbor, such that the smoothing parameter is defined as $h = d_k$. In this way the density at each point in the plane is mainly determined from the closest $\sim k$ objects, such that the variance at each point in the plane is approximately constant. The 'best' value for k will be determined in the next section. Since the resulting pdf will in turn be used to produce an ensemble of discrete sets of galaxies, a very efficient algorithm suggested by Silverman (1986, pg 143) will be used to produce each data set without actually producing the pdf. Given that there are N galaxies

distributed in a circular sample space with positions defined by the coordinates (x_i,y_i), the following steps are performed N times to produce a pseudo-random data set of N objects.

(i) The k^{th} closest neighbor is determined for each galaxy, such that d_{ik} is the distance to the k^{th} closest neighbor of the i^{th} galaxy.

(ii) A value i is randomly chosen with replacement from $\{1,..., N\}$.

(iii) A value ε is generated as a Gaussian deviate with $\sigma = d_{ik}$.

(iv) An azimuthal angle θ is randomly chosen between 0 and 2π.

(v) A new galaxy position is defined as $w=x_i+cos(\varepsilon)$, $z=y_i+sin(\varepsilon)$.

(vi) If the position of the new galaxy is not within the sample space, it is disregarded, and steps (ii) to (vi) are repeated.

This technique is essentially a variation of the bootstrap resampling technique which produces pseudo sets by choosing N galaxies with replacement from the original distribution. There are several advantages to using this technique. Firstly, it alleviates the problem of having to define a grid system on which the pdf would be defined. This pdf would have to be defined on a relatively fine grid such that its discreteness does not affect the distribution of galaxies produced from the pdf; this process is therefore very time consuming. Secondly, due to step (vi), no edge correction needs to be made in step (i). Lastly, it is possible to label each galaxy in the pseudo-random distributions with the galaxy index from which it was drawn, a fact that will prove to be extremely important, as shown in the next section.

4.3 Calculating w_2 from the Simulations

There is a subtle point that must be considered when estimating the angular correlation function for pseudo-random sets produced from a pdf determined with a kernel estimator. When the angular correlation function is estimated, there are contributions from two sources. There is a contribution from the cross correlation of all the different kernels (which essentially defines the contribution from the large scale density gradient) and there

is a contribution from the auto correlation of the kernel itself (which is spurious). When w_2 is estimated for the 'real' galaxy positions, the density field around each galaxy is defined by the *N–1* other galaxies in the sample, and not the galaxy that is used as the center. Therefore, the auto correlation of the kernel must be removed from the estimate of w_2 for the simulated data sets. This may at first seem like an imposing task, but in fact, it is handled quite easily using the technique described in the previous section. Recall that it is possible to label each galaxy in the simulated set with an index that indicates from which galaxy in the 'real' data set it was created. Therefore, a check can be made when calculating *DD* so that the separation between any two galaxies that are a realization from the same 'real' galaxy center can be ignored. In this way, the number of pairs in any simulated set must be counted so that dd is defined as *DD*/(total number of independent pairs kept) rather than *2DD/(N(N–1))*.

Determining a value for *k*

In order to choose a value for *k*, the Hercules cluster was analyzed in detail using 4 different values of *k*, namely *k=5*, *k=10*, *k=20*, and *k=50*. The distribution of the Hercules galaxies is presented in figure 14 on page 63. This distribution was used to estimate a pdf using the 4 different values of *k*. Contour plots of these pdfs can be found in figure 15. Notice the different levels of subclustering that exist for the different values of *k*. An ensemble of 1000 pseudo-random distributions were created for each of the 4 values of *k*, thus producing $<w_2>$ and $\sigma(w_2)$. These values are overplotted, as solid curves, on top of w_2 calculated for the real Hercules galaxy distribution (filled circles) in figure 16 on page 73. It is reassuring how insensitive these results seem to be to the value of *k*. For *k=5*, the first bin has about a *0.9* σ excess of close pairs, while for the *k=50*, the first bin has about a *1.2* σ excess. A value of *k=10* was chosen for the rest of the clusters more because its contour is representative of the large scale density gradient that exists in the Hercules cluster (*i.e.* from a subjective viewpoint), than the results presented in figure 16. This pdf is presented as a surface in figure 17 on page 74 to give the reader a better feel as to how the galaxies in the pseudo-random sets are distributed.

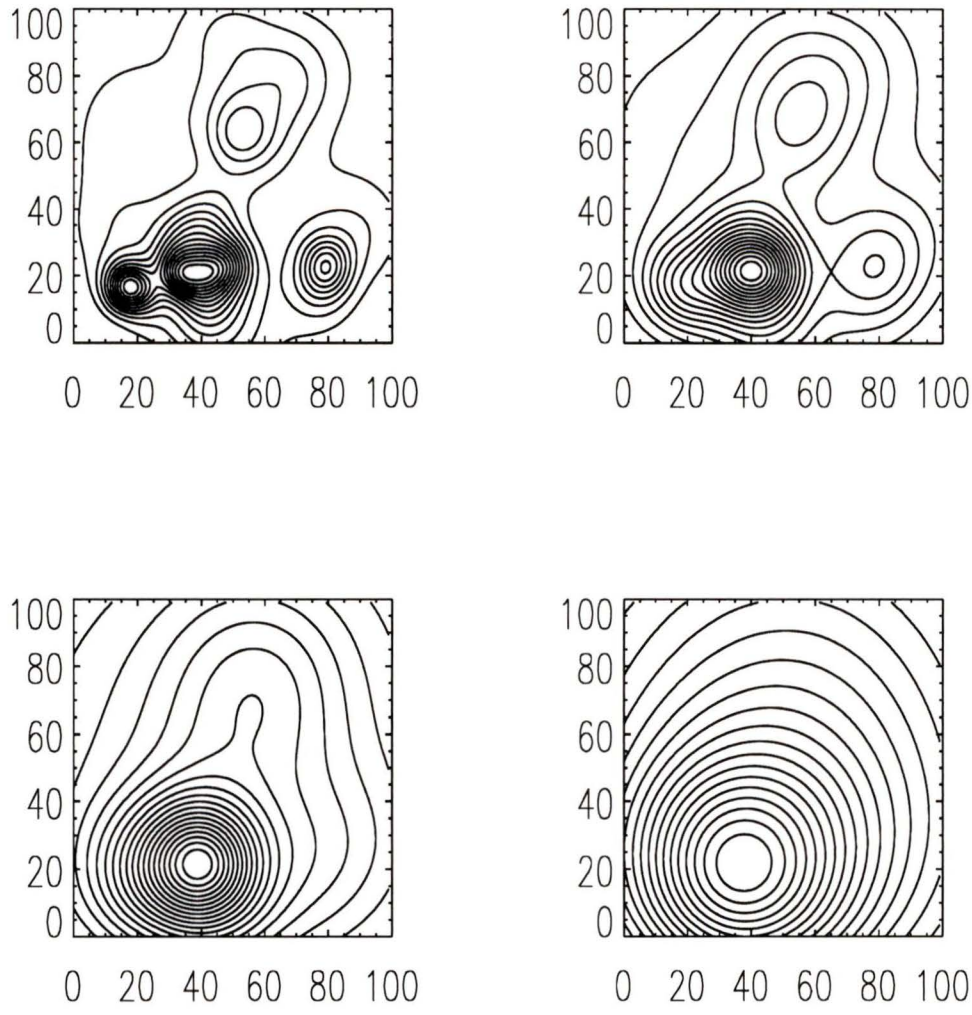


Figure 15 :Density estimates for the Hercules cluster using the variable kernel technique with *k=5, k=10, k=20,* and *k=50* for the top left, top right, bottom left, and bottom right respectively. The actual positions of the Hercules galaxies can be found in figure 14 on page 63.There are 20 contours which are evenly spaced between the minimum and maximum of each plot. The tick labels simply represent the grid system on which the densities were digitized and have no physical significance for the plots.

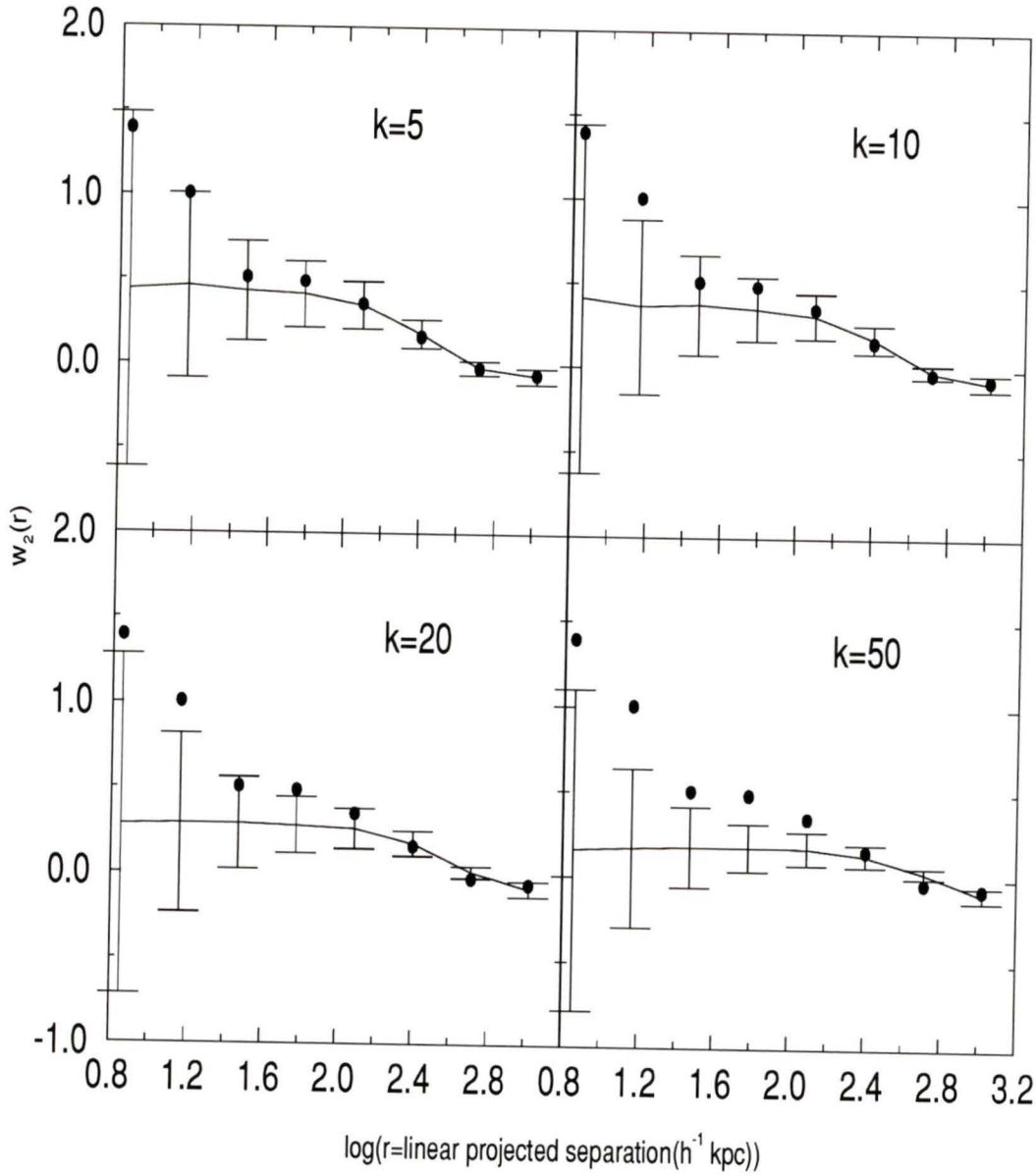


Figure 16 :Angular Correlation functions for 4 simulations of the Hercules density gradient using different values of *k* in the adaptive kernel density estimator.The filled circles represents the correlations from the 'real' galaxy positions. The solid line represents the mean correlations (with 1σ error bars) from the galaxies in 1000 simulated clusters.

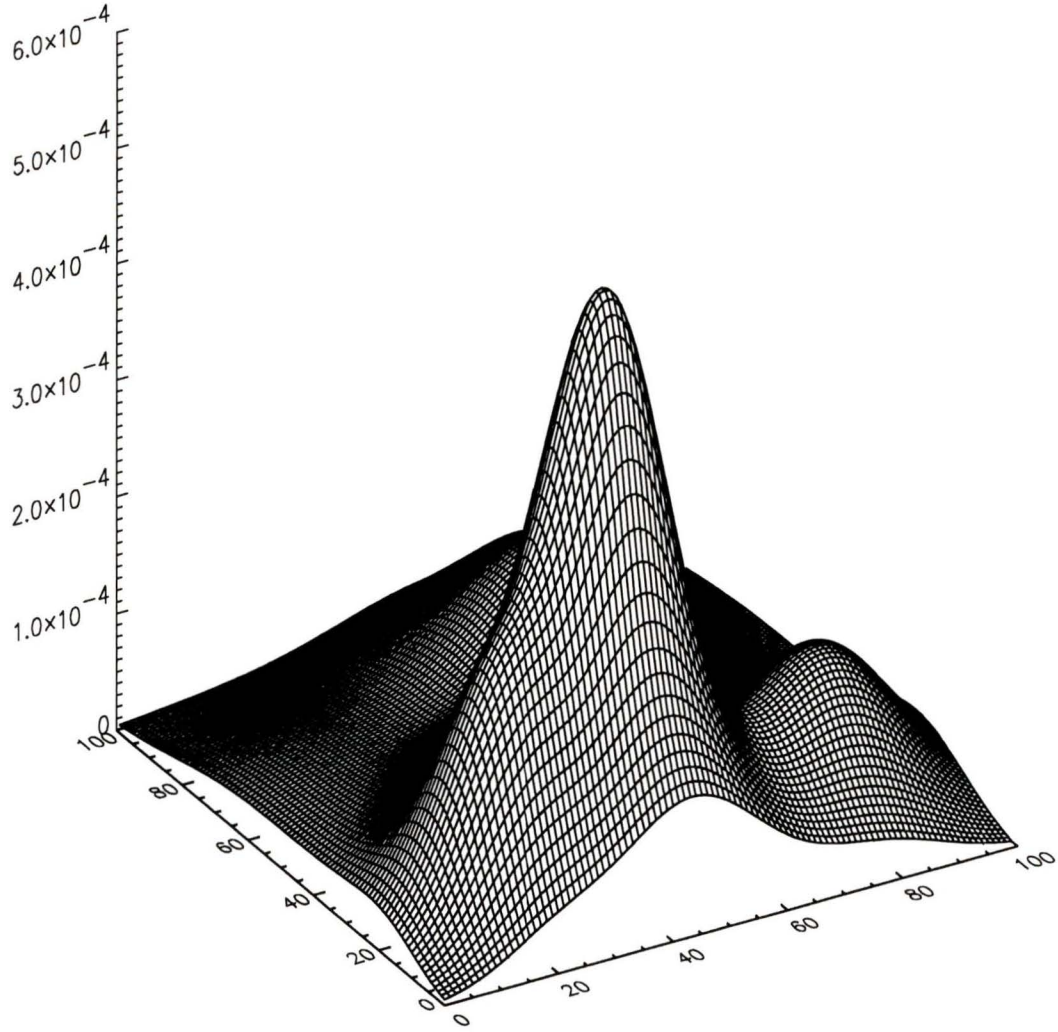


Figure 17 :Probability density function used for the Hercules cluster in order to create pseudo-random data sets from which the angular correlation function was calculated. This profile was determined using the adaptive kernel technique with *k=10*.

4.4 Results for Local Clusters

The density of each cluster was estimated using the adaptive kernel technique with a Gaussian kernel; for all clusters, h was set to be the distance to the k^{th} closest neighbor, and k was set to 10. The steps presented on page 70 were followed to produce an ensemble of 1000 pseudo-random galaxy distributions for each cluster. The angular correlation for each realization was estimated using the w_2 estimator, where the correction discussed on page 70 for the auto correlation of the kernel was made. From each ensemble, $<w_2>$ and $\sigma(w_2)$ were calculated. These values are overplotted on the estimate of w_2 for the 'real' galaxy distributions in figures 18 and 19. Note that if the first bin of the correlation function is entirely below the resolution limit of the cluster, (as is the case for Abell 401, Abell 1904, and Abell 2670) then both the simulation and the "real" correlation functions will have a zero result for this bin. In addition, there is no error bar attached to this bin. Since it is the first two bins that contain the information about the close pairs in the clusters, two tables have been compiled; Table 7 and Table 8 contain the fundamental quantities for these bins, as well as the estimate of w_2 for both the 'real' data, and the simulations.

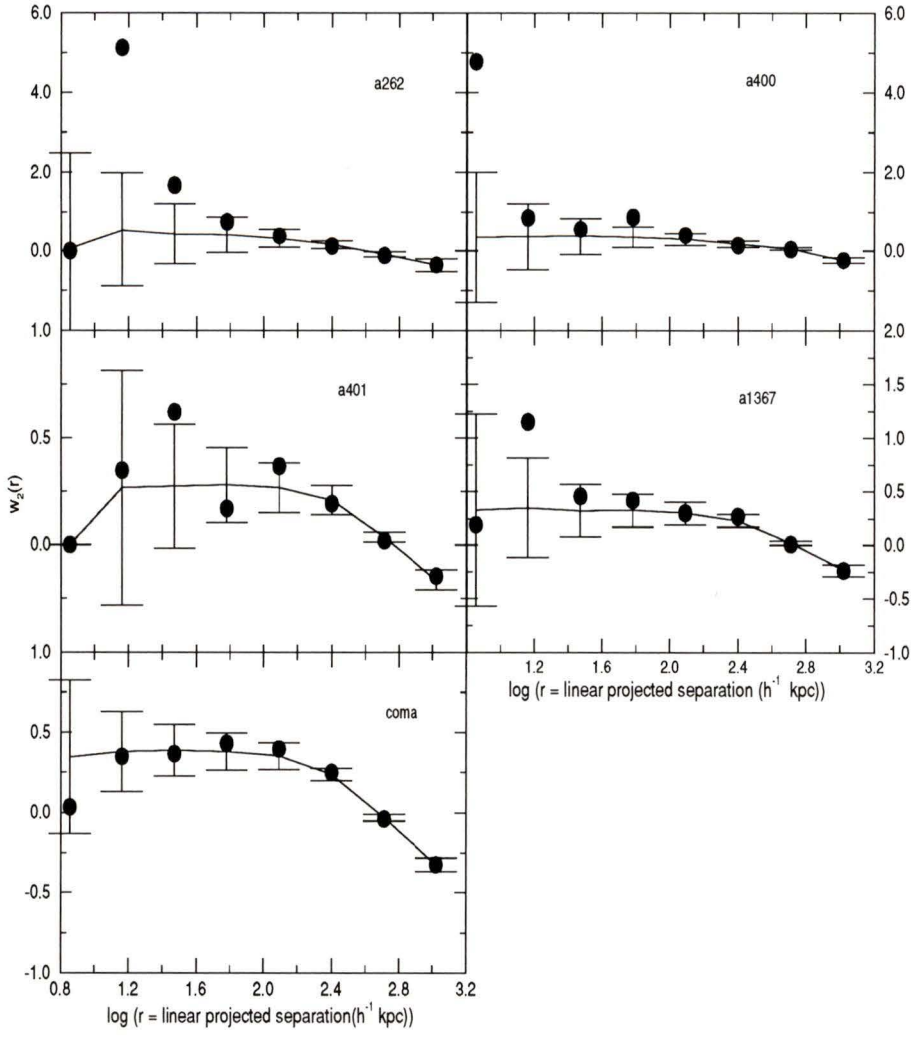


Figure 18 :Angular correlation functions. Angles have been converted to projected linear separations using H_0=100h km s^{-1} Mpc^{-1}. The filled circles represent the values calculated from the real galaxy positions. The solid line connects points with no symbols which represent the correlations calculated from the simulations; the error bars are 1 sigma.

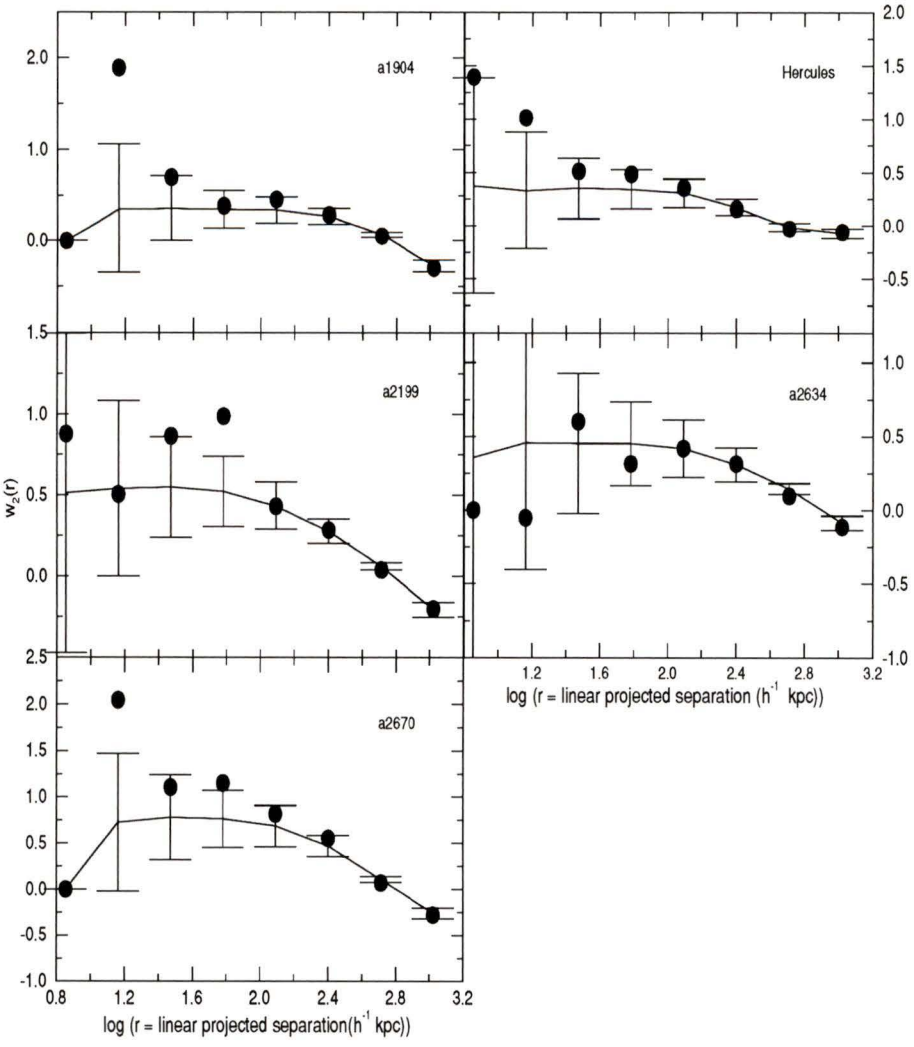


Figure 19 :Angular correlation functions. As in figure 18 on page 76.

Table 7 : Bin #1 (5-10 h^{-1} kpc)

	Data				Simulations			
cluster	*DR*	*2RR*	*2DD*	w_2	*<2DD>*	*σ(2DD)*	*<w2>*	*σ(w2)*
A262	.448	.356	0	0	.470	1.02	.238	2.687
A400	1.15	.912	6	4.77	1.584	1.77	.544	1.729
A401	0	0	0	0	0.	0.	0.	0.
A1367	3.53	3.51	4	.142	4.37	2.98	.311	.895
Coma	13.8	13.7	14	.022	18.85	6.31	.391	.466
A1904	0	0	0	0	0.	0.	0.	0.
Hercules	2.48	2.39	6	1.43	3.50	2.57	.407	1.033
A2199	3.21	3.11	6	.880	4.86	3.25	.535	1.027
A2634	.992	1.02	0	0	1.38	1.75	.417	1.804
A2670	0	0	0	0	0.	0.	0.	0.

Table 8 : Bin #2 (10-20 h^{-1} kpc)

	Data				Simulations			
cluster	*DR*	*2RR*	*2DD*	w_2	*<2DD>*	*σ(2DD)*	*<w2>*	*σ(w2)*
A262	1.79	1.62	10	4.74	2.55	2.24	.618	1.422
A400	4.48	4.36	8	.807	5.95	3.53	.399	.829
A401	8.78	8.71	12	.378	11.47	4.93	.298	.558
A1367	13.9	13.9	30	1.18	18.14	6.18	.313	.447
Coma	57.5	55.8	76	.328	77.64	14.11	.381	.251
A1904	6.5	6.03	18	1.80	8.24	4.25	.339	.691
Hercules	10.8	11.0	22	1.04	14.84	5.64	.364	.518
A2199	13.4	13.1	20	.499	20.38	7.04	.550	.535
A2634	4.11	4.03	4	-.18	5.65	3.41	.398	.845
A2670	6.5	6.36	20	2.10	11.21	4.89	.732	.754

Discussion

Notice that for large angular separation, the correlation functions give essentially identical results for both the real data and the simulated data; this is a result of the data sets having the same large scale density gradients. Therefore, the correlations at small

separations, that are induced by the large scale density profile, still exist in the simulated angular correlation function. For this reason, the correlations for small separations in the real data should be viewed as a correlation superimposed on the simulated correlation. The 1σ error bars for the simulated data plotted in figures 18 and 19 give an indication of how probable it is that the close pairs in the clusters are simply random fluctuations of the underlying density profile.

It is very illuminating to compare the correlation functions in figures 18 and 19 with the projected linear coordinates given in figures 13 and 14. This gives an impression of the spatial distribution of close pairs in the cluster, and at the same time indicates if there are more pairs than one would expect. There are several interesting results from this analysis, although the small sample size makes it difficult to draw any definite conclusions.

The highest excess of close pairs seems to be in the clusters Abell 262, Abell 400, and Abell 1904, all of which show at least a 2σ excess. The excess in Abell 262 should be viewed with scepticism since the rather meager 38 galaxies in this cluster make the density profile very hard to estimate. However, it is still worthwhile noting that there are 8 galaxies all contained within the central 30 h^{-1} kpc, which would make this region very interesting in terms of galaxy interactions and dynamics. The excess in Abell 400 and Abell 1904 seem to be legitimate, since the close pairs are fairly evenly distributed within the cluster. Contrary to the ideas of Pierce et al. (1993), who suggest that the excess in close pairs should be strongest in clusters which are in a dynamically young state of their evolution, both Abell 400 and Abell 1904 are centrally concentrated spiral poor clusters and are generally regarded as being fairly relaxed.

A smaller but marginally significant excess ($\sim 1\sigma$) in close pairs exist in Abell 1367, Hercules, and Abell 2670. It should be noted that, since both the first and second bin of Hercules show a 1σ excess, the net excess of close pairs is $\sim 1.4\sigma$. Both Abell 1367 and Hercules are irregular spiral rich clusters, and thus their excess is understandable if it is a function of the state of virialization. Abell 2670, on the other hand, is a centrally concentrated spiral poor cluster. The rest of the clusters are centrally concentrated spiral

poor clusters, and none of them show a significant excess of close pairs, whereas Coma and Abell 2634 show a deficit of close pairs if anything. Abell 401 shows a marginally significant excess in the 3^{rd} bin (20 - 40 h^{-1} kpc), and Abell 2199 shows a marginally significant excess in the 4^{th} bin (40-80 h^{-1} kpc).

These results seem to indicate that an excess of close pairs in a cluster is relatively uncorrelated with the cluster's morphological classification. Even though the irregular clusters Abell 1367 and Hercules (Abell 262 will no longer be discussed due to the small number of galaxies in this cluster) show a marginal excess of close pairs, there are three regular clusters (Abell 400, Abell 1904, and Abell 2670) that show similar excesses. Furthermore, an excess of close pairs is found in at least one of each Bautz-Morgan type. This seemingly contradicts the suggestions of Pierce et al. (1993), who propose that an excess of close pairs in a cluster may be correlated with the dynamical state of the cluster. This is only true, however, if the morphological classifications presented in Table 5 are representative of the dynamical state of each cluster. By referring to the projected linear coordinates of the clusters in figures 13 and 14, there are two subjective observations which allows one to question this assumption:

(i) The distribution of galaxies in Abell 1904 is not as centrally concentrated as the other regular clusters.

(ii) Although Abell 400 has a compact core, the outer halo is quite clumpy.

These observations are, of course, qualitative, but they do suggest that the simple classification of clusters into regular or irregular categories may be too coarse. A similar observation was made by Rood and Sastry (1971), who found that the Abell clusters are categorized more naturally into 6 classes (now commonly referred to as RS types). This classification system was revised slightly and applied to all 2712 Abell clusters by Struble and Rood (1982, 1984, 1987a); they eventually produced a complete catalog (Struble and Rood 1987b). A short description of this system is found in Appendix C.

There is observational evidence that the RS types are correlated with the dynamical state of a cluster. Abramopolous and Ku (1983) found that the X-ray luminosity of

clusters is correlated with their RS type (see also Struble and Rood 1984, and more recently, Nesci 1991). There is also a relationship between the RS type of a cluster and the presence of narrow angle tail (NAT) radio sources (O'Dea and Owen 1985). In addition, Struble and Rood (1987c) emphasize that there is an increase in the fraction of clumpy irregular (RS type I_c) clusters between the distance class $D \leq 5$ ($z \sim 0.14$), and $D = 6$ ($z \sim 0.18$) clusters. The evolutionary sequence of RS types was suggested by Struble and Rood (1982, 1987c) to be as follows: "I and F clusters are dynamically young, C and L clusters, along with unsubscripted B clusters, are intermediate in dynamical age, and B_b and cD clusters are dynamically the oldest".

Table 9 summarizes the RS types (from Struble and Rood 1987b) for each of the

Table 9 : RS Type Correlations

Cluster	RS type	σ (bin 1)	σ (bin 2)
A 262	**C**	**0**	**2.5**
A 400	**I_c**	**2.5**	**0.6**
A 401	cD	R	0
A 1367	**F**	**0**	**1.5**
Coma	B	-0.8	0
A 1904	**C**	**R**	**2**
Hercules	**F**	**1**	**1**
A 2199	cD	0.3	0
A 2634	cD_s	0	-0.6
A 2670	**cD**	**R**	**1.5**

clusters studied, along with the excess number of pairs determined from this study; this excess is presented as a sigma result for the first two bins of the correlation function (see figures 18 and 19). The clusters that show an excess of more than 1 σ are in bold letters. An 'R' indicates that the bin is below the resolution limit of 10″, and thus not usable.

Notice that Abell 400 and Abell 1904, which were both classified as regular clusters in the Butcher and Oemler (1985) scheme (see Table 5), are now classified as I_c, and C

respectively. With respect to dynamical evolution, the Ic class is in the least evolved category, while the C class is in the intermediate category. This makes the excess of close pairs found for these two clusters more understandable. From an inspection of Table 9, one can see that *an excess of close pairs is found preferentially in clusters of RS type I, F, or C, while no excess is found in the more dynamically evolved RS types B, or cD.* (Note that Coma appears to be more dynamically evolved than most RS type B clusters.) An exception to this rule is Abell 2670, which is a cD cluster, and still seems to show an excess of close pairs.

From an inspection of the angular distribution of galaxies in Abell 2670 (see figure 14), it can be seen that 6 of the 10 pairs of galaxies that contribute to the second bin of the correlation function (*i.e.* the bin with the excess) are found in a somewhat compact group. There is also an interesting feature in the correlation function of Abell 2670 (see figure 19); the third, fourth and fifth bins all show approximately a 1 sigma excess. Since the fifth bin contains separations on the order of $100h^{-1}$ kpc, this implies that there is substructure on these scales. Although each bin is only a 1 sigma result, the fact that all 3 bins show this excess makes the result very significant. This is in agreement with the recent work of Bird (1994a, 1994b), who detects substructure in this cluster and suggest that the cluster has experienced a line of sight merger. This may indeed be the reason that this cluster shows an excess of close pairs.

A few words of precaution should be included here about using the angular correlation function to detect substructure in clusters of galaxies. Since this analysis is always comparing the angular correlation of the "real" galaxies in the cluster to the correlation of galaxies in simulated clusters, it is important to be aware of how the simulated clusters were created. Recall that the 10th closest neighbor was used to define a rough density estimate for each galaxy in the cluster; from this, the smoothing scale was determined for that galaxy, and this was used in the kernel estimation of the cluster pdf. This implies that structures in the galaxy distribution that are defined by less than ~ 10 galaxies will not be modelled well by the pdf, and will turn up as an excess in the angular correlation function. Since there are a different number of galaxies in each of the clusters, the 10th closest

neighbor defines a different scale for each clusters, and therefore, the pdf models different scale structures. For example, since Coma has 256 galaxies, its pdf models relatively fine features in the galaxy distribution, while Abell 262, having only 38 galaxies, has a pdf which models only the coarser features. Therefore, if this technique were to be used to search for substructure, it would be better to choose *k,* in the adaptive kernel estimation of the pdf, as a function of the number of galaxies in the cluster. Alternatively, the kernel estimation could be used with a constant scale length.

In summary, of the 10 clusters studied, none of the most dynamically evolved clusters (RS type B [Coma], and cD clusters) have an excess of close pairs except for Abell 2670; this cluster, however, may be undergoing a line of sight merger, which would make this excess understandable. The rest of the clusters are less dynamically evolved (RS type I_c, F, and C), and also show an excess in the number of close pairs. Although these excesses are only $1 - 2.5\ \sigma$ results, the fact that all the dynamically less evolved clusters have this excess makes the correlation significant. This is in agreement with the predictions made by Pierce et al. (1993).

Although these clusters have been analyzed on an individual basis, this technique is probably more effective when applied to large samples of clusters, such that comparisons are made between different subsets of the sample. In this regard, the next step in this project will be to apply this technique to a much larger sample of local clusters. This will be possible in the near future as the 'Astrophysics Data Systems - Automated Plate Scanner' (ADS-APS) group is presently producing catalogues from the Palomar Observatory Sky Survey using a neural net for galaxy-star classification. It would be very desirable to extend this work to higher redshifts as well, but it should be stated that much more care must be taken to account for contamination from foreground and background galaxies to make this technique useful. Since complete redshift data for galaxies in clusters is rare, this correction will inevitably have to be statistical, and will depend on the work of Zepf and Koo (1989), Carlberg et al. (1994), Burkey et al. (1994), and hopefully others, to better define the frequency of pairs in the field.

References

Abell, G.O. 1958, Astrophysical Journal Supplement Series, **3**, 211.

Abell, G.O. 1975, in Galaxies and the Universe, volume #9 of Stars and Stellar Systems. eds. A. Sandage, M. Sandage, and J. Kristian (University of Chicago Press).

Abramopoulos, F. and Ku, W. 1983, Astrophysical Journal, **271**, 446.

Allington-Smith, J.R., Ellis, R., Zirbel, E.L., and Oemler, A.J. 1993, Astrophysical Journal, **404**, 521.

Arp, H. 1966, *Atlas of Peculiar Galaxies*, Astrophysical Journal Supplement Series, **14**, 3.

Baade, W., Minkowski, R. 1954, Astrophysical Journal, **119**, 206

Bahcall, N.A. 1977, Annual Review of Astronomy and Astrophysics, **15**, 505.

Barnes, J.E. 1988, Astrophysical Journal, **331**, 699.

Barrow, J.D., Bhavsar, S.P., and Sonoda, D.H. 1984, Monthly Notices of the Royal Astronomical Society, **210**, 19p

Bautz, L.P. 1972, Astronomical Journal, **77**,1.

Bautz, L.P., and Morgan, W.M. 1970, Astrophysical Journal Letters, **162**, L149.

Beers, T.C. 1992, in *Statistical Challeges in Modern Astronomy*, eds. Eric D. Feigelson and G. Jogesh Babu (Springer-Verlag), pg111

Beers, T.C., and Tonry J.L. 1986, Astrophysical Journal, **300**, 557.

Bhavsar, S.P. 1990, in *Errors, Bias and Uncertainties in Astronomy* (Cambridge Univ. Press), eds. C. Jasheck, and F. Murtagh, p. 107.

Bird, C.M. 1994a, Astronomical Journal, **107**, 1637.

Bird, C.M. 1994b, Astrophysical Journal, **422**, 480.

Boyle, B.J., Shanks, T., and Peterson, B.A. 1988, Monthly Notices of the Royal Astronomical Society, **235**, 935.

Bracewell, R., 1965, *The Fourier Transform and Its Applications* (McGraw Hill, Inc.)

Breiman, L., Meisel, W., and Purcell, E. 1977, Variable Kernel estimators of multivariate densities. Technometrics, **19**, 135.

Burkey, J.M., Keel, W.C., and Windhorst, R.A. 1994, preprint

Butcher, H., and Oemler, A., Jr. 1978a, Astrophysical Journal, **219**, 18.

Butcher, H., and Oemler, A., Jr. 1978b, Astrophysical Journal, **226**, 559.

Butcher, H., and Oemler, A., Jr. 1984a, Astrophysical Journal, **285**, 426.

Butcher, H., and Oemler, A., Jr. 1984b, Nature, **310**, 31.

Butcher, H., and Oemler, A., Jr. 1985, Astrophysical Journal Supplement Series, **57**, 665.

Butcher, H., Oemler, A., Jr., and Wells, D.C. 1980, in I.A.U. Symposium 92, *Objects at High Redshift*, eds. G. O. Abell, and P.J.E. Peebles (Dordecht: Reidel) p 49.

Butcher, H., Oemler, A., Jr., and Wells, D.C. 1983, Astrophysical Journal Supplement Series, **52**, 183.

Brainerd,T.G., Smail,I., Mould, J., 1994, preprint

Carlberg, R.G. 1990, Astrophysical Journal Letters, **359**, L1

Carlberg, R.G., Pritchet, C.J., and Infante, L. 1994, Astrophysical Journal, in press

Cavaliere, A., Santangelo, P., Tarquini, G., and Vittorio, N. 1986, Astrophysical Journal, **305**, 651.

Collins, C.A., Heydon-Dumbleton, N.H., MacGillivray, H.T., 1988, Monthly Notices of the Royal Astronomical Society, **236**, 7p.

Dressler, A. 1978, Astrophysical Journal, **223**, 765.

Dressler, A. 1993, in *Observational Cosmology*, Astronomical Society of the Pacific Conference Series Volume 51, eds. G Chincarini, A. Iovino, T. Maccacaro, and D Maccagni.

Dressler, A., and Gunn, J. E., 1983, Astrophysical Journal, **270**, 7.

Eales, S. 1993, Astrophysical Journal, **404**, 51.

Efstathiou, G. 1990, in *Dynamics and Interactions of Galaxies*, ed. R. Wielen, Berlin: Springer, p2.

Ellingson, E., and Yee, H.K.C. 1993a in *The Evolution of Galaxies and Their Environment* (NASA. Ames Research Center) p 309.

Ellingson, E., and Yee, H.K.C. 1993b, Astrophysical Journal, 411, 43.

Fall, S.M., Geller, M.J., Jones, B.J.T., and White, S.D.M. 1976, Astrophysical Journal Letters, **205**, L121.

Freund, J.E., 1992, *Mathematical Statistics* 5th edition (Prentice Hall Inc.)

Geller, M.J., Beers, T.C. 1982, Publications of the Astronomical Society of the Pacific, **94**, 421.

Gunn, J.E. 1979, in Active Galactic Nuclei (Cambridge Univ. Press), eds C. Hazard, S. Mitton., p. 213.

Heckman, T. 1990, in *Paired and Interacting Galaxies,* IAU Coll. No 124 (NASA Conference Publication 3098), eds J.W. Sulentic, W.C. Keel, and C.M. Telesco., p. 359.

Henry, J.P., and Lavery, R.J. 1987, Astrophysical Journal, **323**, 473.

Hernquist, L., 1992, Astrophysical Journal, **400**, 460.

Hernquist, L., 1993, Astrophysical Journal, **409**, 548

Hewett, P.C., 1982, Monthly Notices of the Royal Astronomical Society, **201**, 867.

Hubble, E. 1936, Mount Wilson Contributions, No. 548; Astrophysical Journal, **84**, 158.

Huchra, J.P., 1977, Astrophysical Journal, **217**, 928.

Infante, L., 1994, Astronomy & Astrophysics , **282**,353.

Infante, L., and Pritchet,C.J., 1994, Astrophysical Journal, in press

Kennicutt, R.C. Jr. 1990, in *Paired and Interacting Galaxies*, IAU Coll. No 124 (NASA Conference Publication 3098), eds J.W. Sulentic, W.C. keel, and C.M. Telesco., p. 269.

Koo, D.C. 1986, Astrophysical Journal, **311**, 651.

Koo,D.C, and Szalay, A.S., 1984, Astrophysical Journal, **282**,390.

Landy,S.D., and Szalay, A.S., 1993, Astrophysical Journal, **412**, 64L

Larson, R.B., and Tinsley, B. M. 1974, Astrophysical Journal, **192**, 293.

Larson, R.B., and Tinsley, B. M. 1978, Astrophysical Journal, **219**, 46.

Larson, R. B., Tinsley, B.M., and Caldwell, C.N. 1980, Astrophysical Journal, **237**, 692.

Lavery, R.J. 1990, in *Dynamics and Interactions of Galaxies*, ed. R. Wielen, Berlin: Springer, p30..

Lavery, R. J., and Henry, J. P. 1986, Astrophysical Journal Letters, **304**, L5.

Lavery, R. J., and Henry, J. P. 1988, Astrophysical Journal, **330**, 596.

Lavery, R.J., Pierce, M.J., McClure, R.D. 1992, Astronomical Journal **104,** 2067.

Layzer, D., 1956, Astronomical Journal, 61, 383.

Leir, A., and van den Bergh, S. 1977, Astrophysical Journal Supplement Series, **34**, 381.

Limber, D.N., 1953, Astrophysical Journal, **117**,134.

Limber, D.N., 1954, Astrophysical Journal, **119**,655.

Limber, D.N., 1957, Astrophysical Journal, **125**,9.

Ling, E.N., Frenk, C.S., and Barrow, J.D., 1986, Monthly Notices of the Royal Astronomical Society, **223**, Short Communication,21p.

Lonsdale, C.J., Hacking, P.B., Conrow, T.P., and Rowan-Robinson, M. 1990, Astrophysical Journal, **358**, 60.

Maddox, S.J., Efstathiou, G., Sutherland, W.J., and Loveday, J., 1990, Monthly Notices of the Royal Astronomical Society, **242**, 43p

Matthews, T.A., Morgan, W.M., and Schmidt, M. 1964, Astrophysical Journal, **140**, 35.

Morgan, W.M. 1961, Proceedings of the National Acadamy of Science U.S., **47**, 905.

Nesci, R. 1991, Astronomy and Astrophysics, **252**, 13.

Neuschaefer, L.W., Windhorst, R.A., Dressler, A., 1991, Astrophysical Journal, **382**, 32.

O'Dea, L.P., and Owen, F.N. 1985, Astronomical Journal, **90**, 954.

Oemler, A. Jr. 1974, Astrophysical Journal, **194**, 1.

Page, T., Dahn, C.C., and Morrison, F.F. 1961, Astronomical Journal, **66**, 614.

Peebles,P.J.E., 1975, Astrophysical Journal, **196**, 647.

Peebles, P.J.E., 1980, in *The Large-Scale Structure of the Universe.*, Princeton Series in Physics, eds. Arthur S. Wightman, and Philip W. Anderson.

Peebles, P.J.E., and Hauser, M.G., 1974, Astrophysical Journal Supplement Series, **28**, 19.

Pierce, M.J.,McClure, R.D., and Lavery, R.J. 1993, Second Semester, CFHT Bulletin No. 29.

Pritchet, C.J., and Infante, L. 1986, Astronomical Journal, **91**, 217.

Rivolo, A. R. and Yahil, A. 1983, Astrophysical Journal, **274**, 474.

Rood, H.J., and Sastry, G. 1971, Publications of the Astronomical Society of the Pacific, **83**, 313.

Roos, N., and Norman, C.A. 1979, Astronomy & Astrophysics , **76**, 75.

Sanders, D.B., Soifer,B.J., Elias, J.H., Madore, B.F., Matthews, K., Neugebauer, G. and Scoville, N.Z., 1988, Astrophysical Journal, **325**, 74.

Sarazin, C.L., Rood, H.J, and Struble, M.F., 1982, Astronomy & Astrophysics , **108**, L7.

Schweizer, F. 1990, in *Dynamics and Interactions of Galaxies*, ed. R. Wielen, Berlin: Springer.

Searle, L., Sargent, W.L.W., Bagnuola, W.G., 1973, Astrophysical Journal, **179**, 427.

Shane, C. D., and Wirtanen, C. A. 1967, Publications of the Lick Observatory, **22**, 1.

Sharp,N.A., 1979, Astronomy & Astrophysics , 74, 308.

Silverman, B.W. 1986, *Density Estimation for Statistics and Data Analysis (Monographs on Statistics and Applied Probability #26)*, Chapman and Hill.

Struble,F.S., and Rood, H.J. 1982, Astronomical Journal, **87**, 7.

Struble,F.S., and Rood, H.J. 1984, Astronomical Journal, **89**, 1487.

Struble,F.S., and Rood, H.J. 1987a, Astronomical Journal, **93**, 1035.

Struble,F.S., and Rood, H.J. 1987b, Astrophysical Journal Supplement Series, **63,** 555.

Struble,F.S., and Rood, H.J. 1987c, Astrophysical Journal, **323**, 468.

Struble,F.S., and Rood, H.J. 1988, Sky and Telescope, volume **75**, January, pg 16.

Thompson, L. A. 1986, Astrophysical Journal, **306**, 384.

Tinsley, B.M., 1968, Astrophysical Journal, **151**, 547

Toomre, A. 1977 in *The Evolution of Galaxies and Stellar Populations*, eds. B.M. Tinsley and R.B. Larson, (Yale Observatory: New Haven), p. 401.

Toomre, A., and Toomre, J. 1972, Astrophysical Journal, **178**, 623.

Vorontsov-Velyaminov, B.A., 1977, *Atlas of Interacting Galaxies*, Astronomy & Astrophysics Supplement Series, **28**, 1.

West, M.J. and Bothun, G.D. 1990, Astrophysical Journal, **350**, 36.

Yates, M.G., Miller, L., and Peacock, J.A. 1989, Monthly Notices of the Royal Astronomical Society, **240**, 129

Zwickey, F. 1937, Astrophysical Journal, **86,** 217.

Zepf, S.E., and Koo, D.C. 1989, Astrophysical Journal, **337**, 34.

Appendix A

Mathematical Statistics

Since this thesis deals with mathematical statistics to a large extent, some of the terminology, notation, and theorems that are used throughout the chapters will be presented here. In order to use the concept of 'probability' in science effectively, a set of axioms, or rules, must be defined which allow probabilities to be employed and interpreted in a consistent manner. The terminology and notation will be discussed in paragraph format, while the theorems, definitions, and postulates will be numbered consecutively so that they may easily be referred to throughout the chapters in this thesis. The notation and formalism presented in John E. Freund's book entitled Mathematical Statistics is followed closely, and it is here that the reader is referred for more details. Most of the theorems, definitions, and postulates are taken verbatim from this text, and proofs for the theorems may be found there; they are not presented here for the sake of brevity.

A1 Introduction to Probability

Basic Terminology

experiment - any process of observation or measurement. (*e.g.* taking an image of a cluster of galaxies, and calculating the angular separation between two galaxies would be considered an experiment.)

outcomes (also called realizations)- the results that are obtained from an *experiment*. (*e.g.* the value that we obtain for the separation between the two galaxies in the *experiment* described above would be considered the *outcome o*f the experiment.)

sample space - the set of all possible outcomes of an experiment, usually denoted by the label S, where each possible outcome is called an **element** of the sample space. They can be conveniently described using set notation,

$$S = \{\text{set of all elements}\}$$

discrete -the sample space consists of a *finite number* of elements, such as the possible outcomes from the roll of a die, which would be represented by $S = \{1, 2, 3, 4, 5, 6\}$, or a *countably infinite* number of elements, such as the number of rolls of a die that are required before a 1 is rolled, which could occur on the first roll, second roll, ... etc., and would be represented by $S = \{1, 2, 3, \ldots\}$.

continuous - the sample space consists of a continuum of elements, such as all the points on a real number line segment. These types of sample spaces arise whenever the outcomes of experiments are the measurements of physical quantities such as time, length, temperature, etc. Following our example of measuring the angular separation between two galaxies, with the further stipulation that the angular resolution of the galaxies is $10''$, and the boundary for the cluster is a circle of diameter $3600''$, then the sample space for this experiment can be described as $S = \{x | 10'' \leq x \leq 3600'', x \in \Re\}$; in this case, x represents the outcome of an experiment which measures the angular separation between two galaxies. This notation should be read as, "The sample space consists of values x, where x lies between $10''$ and $3600''$, and x is a real number".

event - a subset of the sample space consisting of 1 or more elements considered collectively. For example, in the above example, we could define an event A as the collection of all separations that lie between $100''$ and $200''$, which would be denoted in set notation as, $A = \{x | 100'' \leq x \leq 200'', x \in \Re\}$

ensemble - a collection of identically constructed experiments from which the outcomes can be used to predict probabilities for different events.

independent identically distributed values - a series of values which are drawn independently from the same probability distribution function.

Poisson point process - a process which defines how a discrete set of values are distributed

in either space or time. For example, let N objects be distributed in some solid angle Ω. They are distributed according to a Poisson point process if the following 3 criteria are true:

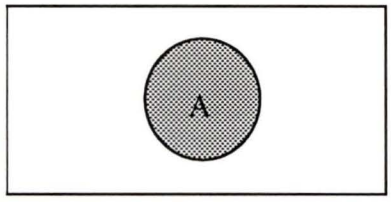
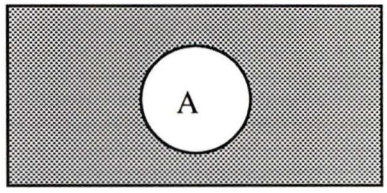
(i) The probability of finding an object in a small solid angle element $\Delta\Omega$ is $\alpha \cdot \Delta\Omega = \frac{N}{\Omega} \cdot \Delta\Omega$.

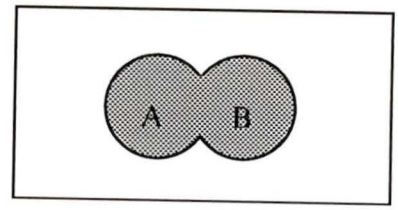
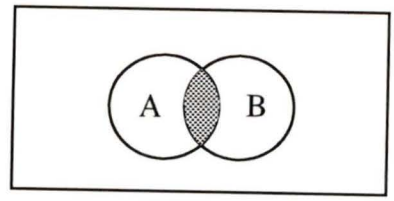
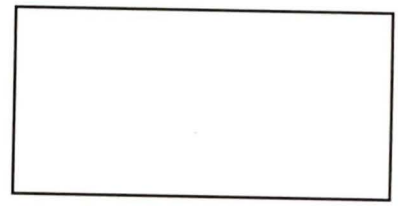
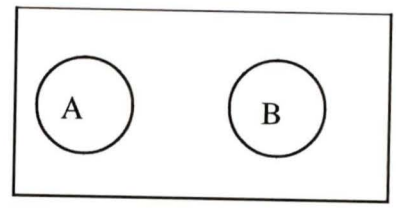
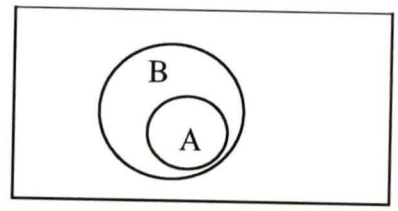
(ii) The probability of finding more than one object in the solid angle element $\Delta\Omega$ is negligible.

(iii) The probability of finding an object in the solid angle element $\Delta\Omega$ does not depend on what happens in any other element of solid angle.

Basic Manipulations

One is often interested in events which consist of combinations of two or more events (say event A and event B), or manipulations of one event. These are most easily understood using Venn diagrams, where the sample space is represented by a rectangle, and events are represented by regions inside the rectangle. The basic manipulations and combinations will be presented by writing in the first column its name, in the next column its symbolic representation, and in the last column its representation in a Venn diagram; the shaded region represents the area described in the first two columns.

Name	Symbolic Representation	Venn Diagram
event	A (or some other label)	
complement	A′	

union	$A \cup B$	
intersection	$A \cap B$	
empty set	$\varnothing$	
mutually exclusive	$A \cap B = \varnothing$	
A contained in B	$A \subset B$	

Postulates, Definitions and Theorems

Given some sample space S, which consists of all possible outcomes from an experiment, the probability that event A will occur will be written as P(A), and the probability that event B will occur will be written as P(B), etc., and probabilities will be assigned within the limits of the following three postulates,

Postulate 1 The probability of an event is a nonnegative real number; that is, $P(A) \geq 0$ for any subset A of S.

Postulate 2 $P(S) = 1$.

Postulate 3 If $A_1, A_2, A_3,...$, is a finite or infinite sequence of mutually exclusive events of S, then $P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$

N.B. For continuous sample spaces, one can only assign probabilities to subsets of the sample space which are formed from unions or intersections of finitely many, or countably infinitely many intervals (or their complements). This is simply a fancy way of saying that we can only assign probabilities to intervals in the sample space, rather than points, since by definition, a point is an infinitely small interval, and thus has zero probability.

For reference a few of the basic theorems of probability will be given. Proofs of these theorems may be found in Freund (1992).

Theorem 1 If A, and A' are complementary events in a sample space S, then $P(A') = 1 - P(A)$

Theorem 2 $P(\varnothing) = 0$ for any sample space S

Theorem 3 If A and B are events in a sample space S, and $A \subset B$, then $P(A) \leq P(B)$.

Theorem 4 $0 \leq P(A) \leq 1$ for any event A.

Theorem 5 If A and B are any two events in a sample space S, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Theorem 6 If A, B, and C are any three events in a sample space S, then

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Conditional Probability

In all of the above cases, only the probability of combinations of events have been considered, without regard to which one happened first. For example, asking what the probability is of both event A and event B occurring is not necessarily the same as asking

what the probability is of event B occurring given that event A has already occurred. In the latter case, the sample space has been changed by stating that event A has occurred. Now any other event that occurs must consist of elements that are also contained in event A and the probability is normalized by all the elements in event A rather than all the elements in S. The conditional probability of event B given that event A has occurred is written as $P(B|A)$, and it is defined as follows,

Definition 1 If A and B are any two events in a sample space S and $P(A) \neq 0$, the conditional probability of B given A is

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

By simple manipulation of Definition 1, the **multiplication rules** for probability can be derived,

Theorem 7 If A and B are any two events in a sample space S, and $P(A) \neq 0$, then

$$P(A \cap B) = P(A)P(B|A)$$

Theorem 8 If A, B, and C are any three events in a sample space S such that $P(A \cap B) \neq 0$, then

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

Independent Events

In words, two events are independent if the occurrence or non-occurrence of one does not affect the probability of the occurrence of the other. This can be written symbolically as $P(B|A) = P(B)$, which can be shown to imply that $P(A|B) = P(A)$, provided that $P(A) \neq 0, P(B) \neq 0$. By using this expressions in Theorem 7, the formal definition for independence is obtained,

Definition 2 Two events, A and B, are independent if and only if

$$P(A \cap B) = P(A)P(B)$$

which in words can be stated as, the probability of both A and B occurring is equal to the product of their individual probabilities if A and B are independent. A useful theorem follows this definition,

Theorem 9 If A and B are independent, then A and B′ are also independent

A2 Probability Distribution Functions

In most cases, a sample space consists of some region of n-dimensional space, such as a line segment in one dimension (1-d), a square or a circle in two dimensions (2-d) etc. For the sake of clarity probability distribution functions will be defined in 1-d, and later their properties will be discussed in 2-d. The elements of the sample space may be discrete (*e.g.* integer values along a line segment) or continuous (*e.g.* an infinitesimal line segment dx), but they need not occur with equal probability. Probabilities can be assigned to the elements as one pleases, so long as the postulates given on page 92 are adhered to, and in general are assigned using a **probability measure** which consists of a real valued function $f(x)$, which describes the probability that an element with a position x in the sample space will occur with respect to all other elements in the sample space. In this case, the position in the sample space would be a random variable with a label X. However, a new random variable could be defined which is a function of the position in the sample space, such as $y=q(x)$, which would have a label Y. In summary then, a sample space S which consists of either a series of values along the x axis in the discrete case, or a line interval in the continuous case is defined. A probability measure for the sample space is produced according to the postulates on page 92 and is defined by the probability distribution function $f(x)$ which is a real valued function that is only non-zero for elements contained within the boundaries of the sample space and describes the probability of any element occurring. Finally, a random variable Y is defined over the sample space which is a real valued function $q(x)$, which for the simplest case is $y=x$, and in this special case

simply describes the probability of each element occurring. The formal definition for a random variable is as follows,

Definition 3 If S is a sample space with a probability measure and Y is a real-valued function defined over the elements of S, then Y is called a **random variable.**

Random variables will be labelled with capital letters, their values with small letters, and probability distribution functions with a small f. So if the elements are discrete, and one wants to know what the probability is that $X=x$, one can write $P(X = x) = f(x)$, and here $f(x)$ is known as a **probability mass function (pmf).** It is only non-zero for a finite, or countably infinite number of x values. This of course won't work if the elements are continuous, since the probability that $X=x$ is zero for all values of x. That is, as was stated earlier, one must consider intervals. One could ask what the probability is that X lies in the interval bounded by the real numbers a and b, where $a \le b$, and this could be expressed as $P(a \le X \le b) = \int_a^b f(x)\,dx$, where here, $f(x)$ is known as a **probability density function (pdf)** and is defined for an infinite number of x values for any line segment contained within the sample space. If it is not important to distinguish between the elements being discrete or continuous, then the generic name used to describe the distribution of probabilities across the elements is **probability distribution function.** It is convenient to always label probability distribution functions with a small f so that there is no ambiguity as to what the function means. Later the **cumulative distribution function** will be described which will always be labelled with a capital F, and is defined as, $P(X \le x) = F(x)$. In all cases, subscripts will be used to distinguish between distribution functions which describe different distributions over the same sample space.

The following theorems and definitions can be divided between those for discrete elements and continuous elements, where the only difference between most of them is the replacement of a summation sign with an integral sign.

Discrete

Theorem 10 A function can serve as the probability mass function of a discrete random variable X if and only if its values, $f(x)$, satisfy the conditions

1. $f(x) \geq 0$ for each value within its domain;

2. $\sum_{x} f(x) = 1$, where the summation extends over all values within its domain.

Definition 4 if X is a discrete random variable, the function given by,

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \text{ for } -\infty < x < \infty$$

where *f(t)* is the value of the probability mass function of X at t, is called the **cumulative distribution function** of X.

Theorem 11 The values *F(x)* of the mass function of a discrete random variable X satisfy the conditions

1. $F(-\infty) = 0$ and $F(\infty) = 1$;

2. if $a < b$, then $F(a) \leq F(b)$ for any real numbers a and b.

Theorem 12 If the range of a random variable X consists of the values $x_1 < x_2 < x_3 < \ldots < x_n$, then $f(x_1) = F(x_1)$ and

$$f(x_i) = F(x_i) - F(x_{i-1}) \quad \text{for } i = 2, 3, \ldots, n$$

Continuous

Theorem 13 A function can serve as a probability density function of a continuous random variable X if its values, *f(x)*, satisfy the conditions,

1. $f(x) \geq 0$ for $-\infty < x < \infty$;

2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Definition 5 if X is a continuous random variable, the function given by,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt \text{ for } -\infty < x < \infty$$

where *f(t)* is the value of the probability density function of X at t, is called the **cumulative distribution function** of X.

Theorem 14 The values *F(x)* of the density function of a continuous random variable X satisfy the conditions

1. $F(-\infty) = 0$ and $F(\infty) = 1$;

2. if $a < b$, then $F(a) \leq F(b)$ for any real numbers a and b.

Theorem 15 If *f(x)* and *F(x)* are the values of the probability density and cumulative distribution function of X at x, then

$$P(a \leq x \leq b) = F(b) - F(a)$$

for any real constants a and b with $a \leq b$, and

$$f(x) = \frac{dF(x)}{dx}$$

where the derivative exists.

Multivariate Distribution functions

The definitions and theorems defined above for univariate distribution functions (*i.e.* one variable) apply equally well to bivariate, and in general multivariate distribution functions. In particular, for bivariate distributions the theorems are identical except that now one either adds or integrates over elements which are distributed in a two dimensional plane rather than a line, and thus single sums and integrals become double sums and integrals. These distribution functions defined as *f(x,y)* are called **joint probability distribution functions.** If one integrates (or sums) one of the variables out, then a distribution function of one variable is obtained which will be called the **marginal**

distribution function. A formal definition for this function follows, which applies to the discrete element case as well if the integral symbols are replaced by summation symbols over the domain of the sample space, and the word *density* is replaced by *mass*.

Definition 6 If X and Y are continuous random variables and $f(x,y)$ is the value of their joint probability density at (x,y), the function given by

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad \text{for } -\infty < x < \infty$$

is called the **marginal density** of X. Correspondingly, the function given by

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \quad \text{for } -\infty < y < \infty$$

is called the **marginal density** of Y.

In practice, the value of one of the two random variables in a bivariate distribution may already be known, and thus it is useful to know what the probability distribution is for the second variable given the value of the first variable. This is nothing more than the conditional probability discussed in Definition 1, and can be defined as follows. As before, the definition for the discrete case is identical except the range of the variables is restricted to the elements contained in the sample space, and the word *density* is replaced by *mass*. Notice that that $f(x,y)$ is merely renormalized, so that the new distribution function of only one variable still adheres to Postulate 2 on page 93.

Definition 7 If $f(x,y)$ is the value of the joint density of the continuous random variables X and Y at (x,y), and $h(y)$ is the value of the marginal density of Y at y, the function given by

$$f(x|y) = \frac{f(x, y)}{h(y)} \qquad h(y) \neq 0$$

for $-\infty < x < \infty$, is called the **conditional density** of X given $Y=y$.

Correspondingly, if $g(x)$ is the value of the marginal density of X at x, the function given by

$$w(y|x) = \frac{f(x,y)}{g(x)} \qquad g(x) \neq 0$$

for $-\infty < y < \infty$, is called the **conditional density** of Y given $X=x$.

A3 Mathematical Expectation

For the sake of brevity, only sample spaces that consist of continuous elements will be considered. However, the majority of theorems etc. are identical if you consider sums in place of the integrals. For more on discrete sample spaces, see Freund (1992).

One can define the expectation value of the random variable X as

Definition 8 If X is a continuous random variable and $f(x)$ is the value of its probability density at x, the expected value of X is

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Since $f(x)$ is only non-zero in the interval within the sample space S, the limits of the integration are defined by the limits of the sample space. For instance, if S was defined as the segment of the x axis between the two real values a and b, where $a \leq b$, then the expected value of X would reduce to $\int_a^b x \cdot f(x)\, dx$. Therefore, for a continuous sample space, the boundaries of the sample space define the limits of the integration for the expectation value.

It is of general interest to know what the expected value of a random variable $Y=q(X)$ is over a sample space S, with a probability measure given by a probability density function $f(x)$; thus, for reference, the general theorem for this is presented below along with some examples for specific forms of $q(x)$.

Theorem 16 If X is a continuous random variable and $f(x)$ is the value of its probability density function at x, the expected value of $q(X)$ is given by

$$E(q(X)) = \int_{-\infty}^{\infty} q(x) \cdot f(x)\, dx$$

Theorem 17 If a and b are constants, then

$$E(aX + b) = aE(X) + b$$

Corollary 1 If a is a constant, then

$$E(aX) = aE(X)$$

Corollary 2 If b is a constant, then

$$E(b) = b$$

Theorem 18 If c_1, c_2,..., and c_n are constants, then

$$E\left[\sum_{i=1}^{n} c_i q_i(X)\right] = \sum_{i=1}^{n} c_i E[q_i(X)]$$

These theorems can easily be extended to two dimensional sample spaces, where the joint probability density $f(x,y)$ defines the probability that any element in the 2-d sample space will occur, and $z=q(x,y)$ defines a random variable Z over the sample space. It can be shown that

Theorem 19 If X and Y are continuous random variables and $f(x,y)$ is the value of their joint probability density at (x,y), then the expected value of $Z = q(X,Y)$ is

$$E[q(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} q(x,y) f(x,y)\, dxdy$$

where again, the boundaries of the sample space define the limits of the integration, since $f(x,y)$ is zero for any values of (x,y) that are not in the sample space.

Moments

It is often convenient to describe a probability density function by a series of numbers known as its "moments", which describe characteristics of the distribution (*e.g.* mean, standard deviation, skewness, kurtosis, etc.). In general, a moment is nothing more than the expected value of some random variable, but it has been found that certain moments are more useful in practice than others; these will be briefly summarized below.

Definition 9 The ***r*th moment about the origin** of a random variable X, denoted by μ'_r, is the expected value of X^r, and can be written symbolically as

$$\mu'_r = E(X^r) = \int_{-\infty}^{\infty} x^r \cdot f(x)\, dx,$$

but in practice, the first moment about the origin is the most useful; this is usually referred to as the mean of the distribution, and is simply denoted as μ. For a discrete set of independent identically distributed values, the sample mean is an unbiased estimate of the mean of the distribution, and is defined by,

$$\langle x \rangle = \frac{\sum_{i=1}^{N} x_i}{N}$$

Definition 10 The ***r*th moment about the mean** of a random variable X, denoted by μ_r, is the expected value of $(X-\mu)^r$, symbolically,

$$\mu_r = E\left[(X-\mu)^r\right] = \int_{-\infty}^{\infty} (x-\mu)^r \cdot f(x)\, dx$$

where μ_2 is generally referred to as the variance of the distribution, and is more commonly denoted by σ^2, var (X), or V (X), and the positive square root of this value is referred to as the standard deviation σ.

It is often convenient to calculate the variance of a distribution by calculating the first and second moments about the origin, and then using the following theorem.

Theorem 20 $$\sigma^2 = \mu'_2 - \mu^2$$

For a discrete set of independent identically distributed values, the sample variance is an unbiased estimate of the variance of the distribution function, and is defined as,

$$\sigma^2 \approx s^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \langle x \rangle)^2$$

Product Moments

We can now consider how to calculate moments for a two dimensional sample space defined by two random variables X and Y with a joint probability function given by f(x,y),

Definition 11 The ***r*th and *s*th product moment about the origin** of the random variables X and Y, denoted by $\mu'_{r,s}$, is the expected value of $X^r Y^s$; symbolically

$$\mu'_{r,s} = E\left(X^r Y^s\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s \cdot f(x,y)\, dxdy$$

where $\mu'_{1,0} = E(X) = \mu_X$, and $\mu'_{0,1} = E(Y) = \mu_Y$

Definition 12 The ***r*th and *s*th product moment about the mean** of the random variables X and Y, denoted by $\mu_{r,s}$, is the expected value of $(X-\mu_X)^r (Y-\mu_Y)^s$; symbolically

$$\mu_{r,s} = E\left[(X-\mu_X)^r (Y-\mu_Y)^s\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_X)^r (y-\mu_Y)^s \cdot f(x,y)\, dxdy,$$

where the quantity $\mu_{1,1}$ is of special importance, since it indicates if there is any relationship between the two random variables X, and Y. From this,

Definition 13 $\mu_{1,1}$ is called the covariance of X and Y, and it is denoted by $\sigma_{X,Y}$, $\mathrm{cov}(X,Y)$, or $C(X,Y)$.

Theorem 21 $\sigma_{XY} = \mu'_{1,1} - \mu_X \mu_Y$

Notice that if large values of X are associated with large values of Y and small values of X are associated with small values of Y then the covariance of these two random variables will be positive, but if large values of X are associated with small values of Y, and vice versa, then the covariance between X and Y will be negative, and if the value of X does not affect the value of Y, then,

Theorem 22 If X and Y are independent, then $E(X, Y) = E(X) \cdot E(Y)$ and $\sigma_{XY} = 0$.

For a discrete set of independent identically distributed pairs of values x_i and y_i, the sample covariance is an unbiased estimate of the covariance of the joint probability distribution function, and is defined as,

$$s_{xy}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \langle x \rangle) \cdot (y_i - \langle y \rangle)$$

Moments of Linear Combinations of Random Variables

In many cases, new random variables are defined over the sample space which are linear combinations of pre-existing random variables, and thus the following theorems are useful,

Theorem 23 If $X_1, X_2, \ldots, X_n$ are random variables and

$$Y = \sum_{i=1}^{n} a_i X_i$$

where $a_1, a_2, \ldots, a_n$ are constants, then

$$E(Y) = \sum_{i=1}^{n} a_i E(X_i)$$

and

$$\text{var}(Y) = \sum_{i=1}^{n} a_i \cdot \text{var}(X_i) + 2\sum\sum_{i<j} a_i a_j \cdot \text{cov}(X_i, X_j)$$

where the summation extends over all values of i and j, from 1 to n, for which i is less than j.

Corollary If the random variables $X_1, X_2, \ldots, X_n$ are independent and $Y = \sum_{i=1}^{n} a_i X_i$, then

$$\text{var}(Y) = \sum_{i=1}^{n} a_i^2 \cdot \text{var}(X_i)$$

Theorem 24 If $X_1, X_2, \ldots, X_n$ are random variables and

$$Y_1 = \sum_{i=1}^{n} a_i X_i \text{ and } Y_2 = \sum_{i=1}^{n} b_i X_i$$

where $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ are constants, then

$$\text{cov}(Y_1, Y_2) = \sum_{i=1}^{n} a_i b_i \cdot \text{var}(X_i) + \sum\sum_{i<j} (a_i b_j + b_i a_j) \cdot \text{cov}(X_i, X_j)$$

where the summation extends over all values of i and j, from 1 to n, for which i is less than j.

Corollary If the random variables $X_1, X_2, \ldots, X_n$ are independent and $Y_1 = \sum_{i=1}^{n} a_i X_i$, and $Y_2 = \sum_{i=1}^{n} b_i X_i$ then

$$\text{cov}(Y_1, Y_2) = \sum_{i=1}^{n} a_i b_i \cdot \text{var}(X_i)$$

Appendix B

Calculations

B1 Analytic Calculation of *dr*

The quantity *dr(θ)* (referred to as *dr*, where the angular dependence is assumed) represents (see definition (ii) on page 16)

> "(ii) ...the mean value of the solid angle subtended by the annulus $\theta - \Delta\theta/2$ to $\theta + \Delta\theta/2$ and within Ω for ring centers placed *at the positions of the galaxies in* Ω"

This quantity can be estimated from a discrete distribution of N galaxies in a sample space S by using the Monte Carlo Method. With this method, there are N_r objects distributed in the same sample space as the galaxies. The separations from each of the N galaxies to each of the N_r objects (a total of $N \cdot N_r$ separations) are calculated, and the number of separations that fall in the interval θ–$\Delta\theta$ and θ+$\Delta\theta$ (which is labelled *DR*, where the angular dependence is assumed) are counted. The fractional number of separations that lie in the bin is then an estimate for the quantity *dr* given by (see also eq. 17 on page 19),

$$dr = \frac{DR}{N \cdot N_r} \qquad \text{(eq. 50)}$$

This estimate will have a statistical uncertainty of $\sqrt{DR}/(N \cdot N_r)$ if only 1 random set is used, and $\sqrt{DR}/\left(\sqrt{N_t} \cdot N \cdot N_r\right)$ if the average of N_t random sets are used.

This quantity is only dependent on the distribution of galaxies in the sample space and the geometry of the sample space, so if the geometry is sufficiently simple, it is possible to calculate *dr* analytically. If the sample space is circular, then for any bin defined as a circular annulus centered on a galaxy with an inner boundary θ_1, and an outer boundary θ_2, then it is possible to calculate the fraction of the bin contained within the sample space

given that the galaxy is a distance r from the center. This value averaged over all the galaxies in the sample gives an error free (in the statistical sense) estimate of dr.

The geometry of the problem is presented in figure 20. Due to the small angular size

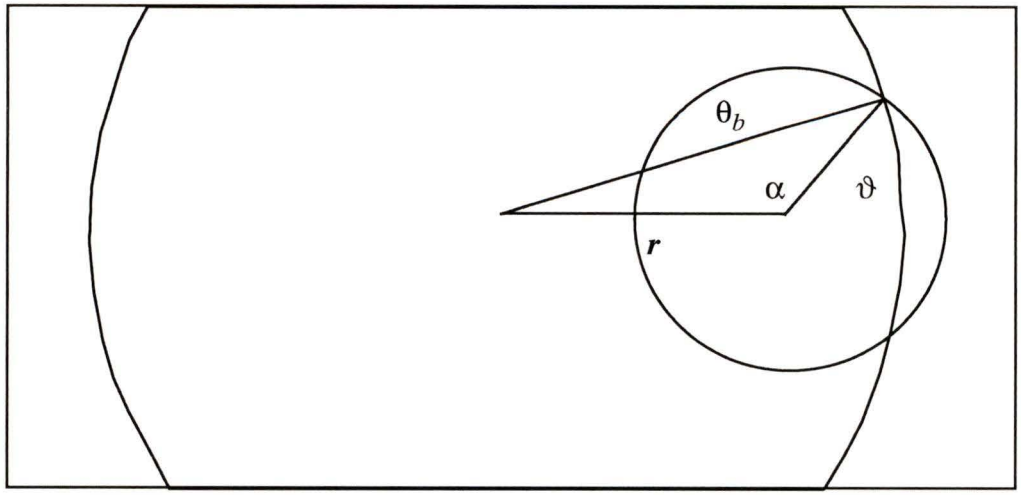


Figure 20 :Geometry for analytical calculation of *dr*. The large circle that is only partially shown represents the circular sample space. The center of the smaller circle is the position of a galaxy in the sample space. The smaller circle represents an infinitesimal integrating annulus.

on the sky of the circular sample spaces in cluster analysis, these calculations will be done assuming a flat two dimensional plane. In this problem, θ_b is the radius of the circular sample space, r is the distance of a given galaxy from the center of the sample space, θ_1 and θ_2 are the inner and outer boundaries for the annular bin of interest, ϑ is the integrating variable for the bin, and α is the angle from the direction pointing to the center of the sample space to the point on the boundary where the integrating annulus intersects. From figure 20, it can be seen that for any infinitesimal circle of radius ϑ centered on the galaxy, α/π represents the fraction of this circle that lies within the boundary. Therefore integrating α across a bin and dividing by the area of the bin (assuming there were no boundaries) will give the fraction of the bin that is contained within the sample space. Since α is given from the cosine law as seen in figure 20, an estimate for dr is,

$$dr = \frac{\langle 2\int_{\theta_1}^{\theta_2} \alpha\vartheta d\vartheta \rangle}{\pi\left(\theta_2^2 - \theta_1^2\right)} = \frac{\langle 2\int_{\theta_1}^{\theta_2} \mathrm{acos}\left(\frac{r^2 + \vartheta^2 - \theta_b{}^2}{2r\vartheta}\right)\vartheta d\vartheta \rangle}{\pi\left(\theta_2^2 - \theta_1^2\right)} \qquad \text{(eq. 51)}$$

where the integral is performed numerically for each galaxy in the sample, and < ... > implies an average over all galaxies. If the resolution limit for the cluster being studied is smaller than θ_1, then this limit is used for the lower limit of the integration, rather than θ_1.

In order to make sure that both methods give the same result, *dr* was calculated for the Coma cluster (presented in Chapter Four) using both the Monte Carlo method, and also the analytic method. For the Monte Carlo method, the same number of objects as there are galaxies in Coma sample were distributed randomly in a circular sample space and the separations between galaxies and random objects were placed into 8 logarithmic bins (identical to the binning in Chapter Four). This process was repeated 500 times to produce <*DR*> and σ*(DR)*. By dividing <*DR*> by N^2(assuming there are *N* galaxies in Coma), and the area (assuming no boundaries), the probability per unit area was obtained. *dr* was calculated analytically for these 8 bins using eq. 51, and was divided by the area of the bin to produce the probability per unit area. These results are presented in figure 21 on page 109. The Monte Carlo method took 26 minutes of cpu time on a Sparc 10, while the analytic method took 3 seconds of cpu time on the same machine with approximately the same load conditions! Notice that, even with 500 repetitions, the error in the mean for the first bin is still 1.2%. The advantages of the analytic method are obvious.

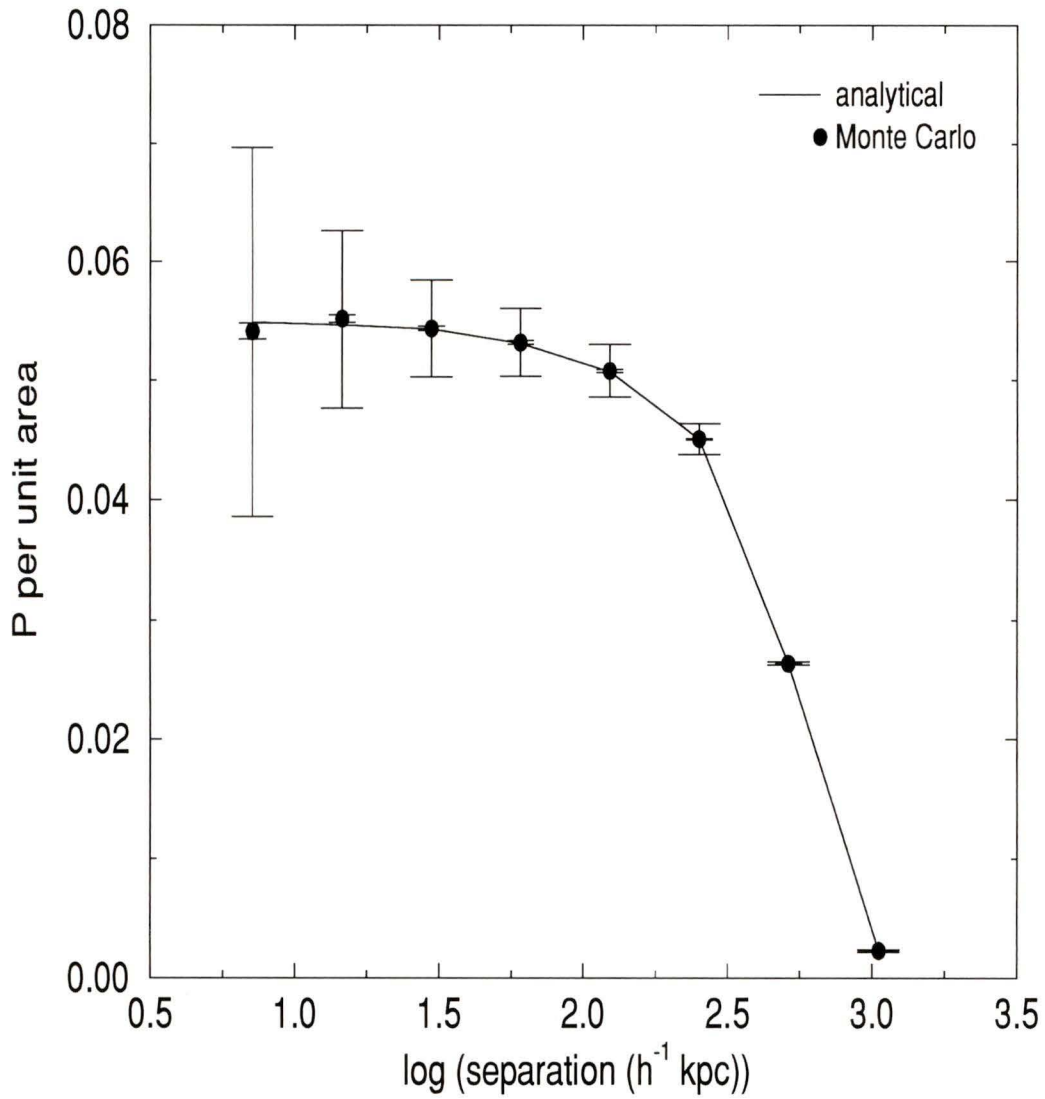


Figure 21 :Distribution of pair probability *dr*. The large error bars are the 1 σ standard deviation for each bin, while the small error bars are the 1 σ standard deviation in the mean for 500 repetitions (*i.e.* $1/\sqrt{500}$ times the error in the variable)

Appendix C

Cluster Morphological Classification Systems

An excellent summary of the different morphological classification schemes for clusters of galaxies has been presented by Struble and Rood (1988). A more detailed description of the earlier ones (§§ I-III below) can be found in Abell's (1975) summary of clusters of galaxies. The following is a compilation of the essential criteria for the most commonly used classification schemes.

I Spiral Rich / Spiral Poor

Morgan (1961) studied the 20 nearest Abell clusters (Abell 1958), and based on the morphology of their brightest members, found that there were two distinct types of clusters. The type i clusters contained a significant number late spirals and irregular galaxies, while the type ii contained very few of these late type galaxies.

II Bautz-Morgan Type

Matthews, Morgan, and Schmidt (1964) showed that more than 30 clusters contained superluminous galaxies with extended halos as their brightest member; they labelled these galaxies cD. Bautz and Morgan (1970) and Bautz (1972) classified clusters according to the type and distribution of their superluminous cluster members. A **Bautz-Morgan type I** cluster contains a cD, or giant binary galaxy, which dominates the cluster visually. A **Bautz-Morgan type II** cluster contains two or three galaxies which visually dominate the rest of the cluster galaxies; Coma is the prototype of this class. Finally, a **Bautz-Morgan type III** cluster has no particularly dominant galaxies; Virgo is a typical example. There are also type I-II and II-III, which provide more continuity to this classification scheme. Leir and van den Bergh (1977), from a visual inspection of the Palomar Observatory Sky Survey (POSS), have assigned Bautz-Morgan types to almost all of the Abell clusters.

III Regular / Irregular

Abell (1975) found that the major features of a galaxy cluster could be divided into two simple classes; he termed these classes regular and irregular. A description of these two classes is given by Abell (1975, pg. 604):

> "***Regular*** clusters are all rich, having populations of the order of 10^3 or more in the interval of the brightest 6 magnitudes. They show high central concentration and marked spherical symmetry. Their memberships consist entirely, or almost entirely, of galaxies without conspicuous dust –E and S0 galaxies. Examples are the famous clusters in Coma and Corona Borealis"
>
> "***Irregular*** clusters range from poor groups, like the Local group, to relatively rich aggregates like the Virgo cluster or the Hercules cluster. They contain little or no spherical symmetry, and no marked central concentration, although multiple condensations are often present. These clusters usually contain galaxies of all types, including appreciable numbers of late-type spirals and irregulars."

IV Revised Rood-Sastry

A more detailed classification scheme was developed by Rood and Sastry (1971), which categorizes clusters into 6 classes, some of which are further divided into various subtypes. These classes are based on the type and distribution of both the superluminous galaxies of the cluster, and the remaining 10 - 20 brightest members. This scheme has been revised slightly by Struble and Rood (1982), and is now commonly referred to as the revised Rood-Sastry type, or more simply as the RS type. Classifications of Abell clusters have been completed for distance class $D \leq 4$ ($z \leq 0.09$) (Struble and Rood 1982), $D = 5$ ($z \sim 0.14$) (Struble and Rood 1984), and $D = 6$ ($z \sim 0.18$) (Struble and Rood 1987a). A final catalog of all 2712 Abell clusters can be found in Struble and Rood (1987b), while a description of the most significant correlations can be found in Struble

and Rood (1987c). The following description of each of the RS types has been taken verbatim from Struble and Rood (1988).

(i) cD. – A cluster dominated by a single large galaxy – cD, a smaller and less luminous D galaxy, or a giant elliptical – that is at least three times bigger than any other member galaxy. Most of a cD cluster's constituents are E's and S0's.

This class has three subtypes that describe special properties of the dominant galaxy:

- **cD_s** means a satellite object is seen within its faint envelope of stars.
- **cD_p** means it displays a peculiarity like an asymmetric nucleus or envelope, a jet, or a tail.
- **cD_n** means its core has a nest of multiple nuclei.

(ii) B. – A cluster dominated by a "binary," two galaxies each at least 1 magnitude brighter than any other and separated by less than 10 times the diameter of either (Coma is an example). As with cD groups, B clusters contain mostly E and S0 galaxies.

- **B_b**. Sometimes the two bright galaxies are joined by a luminous bridge or share a common envelope of light; these are the dumbbells mentioned earlier. Clusters containing these especially close pairs are subscripted with a 'b'.

(iii) C. – A Core-halo cluster with several bright galaxies of any type concentrated near its center (the core) and many fainter galaxies spread throughout a larger surrounding area (the halo).

(iv) L. – A cluster dominated by several bright galaxies, again of any type, **arranged roughly in a line across the sky**. In some L systems the fainter galaxies also lie along this line; in others they are spread over a broader area of the sky (the Perseus cluster is an example).

- **L_a** denotes a cluster in which an arc of bright galaxies is observed rather than a nearly

straight line.

(v) **F. – A cluster appearing flattened** or elongated on the sky but not dominated by any bright galaxies (the Virgo and Hercules clusters are examples). Their members are mainly spirals and S0's. Most F clusters are actually composed of two or three subclusters that could each be classified independently.

(vi) **I. – An irregularly shaped cluster** not dominated by any bright galaxies and with only a mild concentration toward a center. Nearly all the galaxies in an I cluster are spirals.

- $\mathbf{I_c}$ denotes a clumpy system composed of several subclusters.
- $\mathbf{I_s}$ describes a smoother distribution with no noticeable subclustering at all.

Vita

Surname: **Rutledge** Given Names: **Glen Alfred**

Place of Birth: **Toronto, Ontario** Date of Birth: **August 28, 1969**

Educational Institutions Attended:

University of Victoria, Victoria , British Columbia **1992 to 1994**

McMaster University, Hamilton, Ontario **1988 to 1992**

Degrees Awarded:

B.Sc. in Physics (Honours)
McMaster University **1992**

Honours and Awards:

NSERC PGS B (National Science and Engineering Research Council Post Graduate Scholarship) -awarded for Ph.D. - September 1995

NSERC PGS A - September 1992 - August 1994

President's Research Scholarship, University of Victoria - 1992/1993, 1993/1994, 1994/1995

Dean's Honors List, McMaster University - 1989,1990,1991,1992

Senate Scholarship received for Academic excellence, McMaster University - 1990

NRC (National Research Council) Undergraduate Summer Research Award, Dominion Astrophysical Observatory - 1991

NSERC Undergraduate Summer Research Award, McMaster University - 1990

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis: Close Pairs of Galaxies in Rich Clusters: A Statistical Analysis

Author: [redacted]

Glen Alfred Rutledge

August 31, 1994