Chapter 3

Point processes and galaxy clustering II: the use and misuse of stochastic models

1 Introduction

In the study of structure formation in the Universe, the greatest difficulties are encountered in the so-called non-linear regime. This regime, where the relative density fluctuations become of order one and greater, poses problems not only for understanding the dynamical aspects of the clustering process in time, but also for ‘merely’ describing the distribution of matter throughout space at a fixed point of time. In the early stages of its evolution, the matter content of the universe can be described by a continuous density field which deviates only slightly from homogeneity. Such ‘linear’ deviations can be quantitatively described using the same statistical measures that allow the equations of motion to be solved. At more evolved stages, this continuum model is no longer applicable. Density contrasts grow to such amplitudes that the linear dynamics is no longer valid and isolated objects develop for which this description in terms of uncorrelated Fourier modes is no longer manageable nor useful. Instead, a discrete description in terms of the distribution of these isolated objects becomes more appropriate. The mathematical formalism for describing such discrete point distributions is provided by the theory of stochastic point processes.

The theory of point processes, which initially received a great impetus from astronomy through the work of Neyman & Scott (1953), has grown into a mature branch of mathematics, finding applications in such diverse fields as economy, biology and telecommunications (for a short introduction to the theory of random point processes, see the book by Cox & Isham, 1980; for more comprehensive treatments of the subject, see the books by Daley & Vere-Jones, 1988; Matthes et al., 1978; Stoyan & Stoyan, 1994; Karr, 1986). Part of the theory is devoted to finding optimal estimators for determining the statistical characteristics of point processes encountered in practical situations (e.g. Diggle (1983) and Ripley (1988) for the mathematical and Colombi et al. (1995) for the cosmological viewpoint). An example of this approach has been discussed in the first chapter of this thesis. The
other part is devoted to deriving rigorous results on specific types of point processes and to
develop techniques to specify the statistical characteristics of point processes completely.

There are several ways in which a point process may be completely specified, apart
from providing the positions of all the points explicitly. Within the cosmological literature
point processes are usually described using the correlation function approach. Here one
describes the point process in terms of an infinite hierarchy of local functions, which give the
probabilities of all possible point configurations (see chapter 3). Within the mathematical
literature a point process is generally defined by the whole set of counting probability
distributions on the Borel sets of the space under consideration (Daley & Vere-Jones, 1988;
DV-J88). These give the the probability for finding any number of points in any number
of finite volumes.

Both these methods should in principle describe a point process completely and one
ought to be able to translate results from one description in terms of the other. In an
influential paper, White (1979; W79) showed that such a relation does indeed exist and can
be cast in an elegant form, expressing the counting probabilities for single volumes in terms
of all the correlation functions. The main motivation of that work was to develop statistical
clustering measures that might provide information complementary to the two and three-
point functions, the only ones which can be determined with any accuracy. The necessity of
finding alternatives to the usual measures for the galaxy clustering process has been shown
very clearly by the deep galaxy redshift surveys (e.g. de Lapparent et al., 1986; Haynes
& Giovanelli, 1988; Shectman et al., 1992). These have revealed a wealth of structure
that cannot be described in sufficient detail by the lower order correlation functions. An
important aim of present-day cosmology is to understand the emergence of these large scale
structures in the galaxy distribution from the supposed initial conditions and subsequent
dynamics of gravitational collapse. Regardless of the problems introduced by the incomplete
knowledge of possible non-gravitational processes, the complications introduced by the non-
linear regime of gravitational collapse alone preclude an exact description of the resulting
point process to be extracted from the dynamics. Much energy has therefore been devoted
to finding approximate stochastic models aimed at describing the characteristic of the true
clustering process. Models have been proposed to mimic both the dynamical aspect of the
evolutionary process and the structural properties of the resulting galaxy distribution. In
this chapter we concentrate on this last aspect, by investigating the merits of some popular
stochastic models for describing the galaxy clustering point process.

These models can be formulated using the three descriptive methods mentioned above.
The most straightforward of these is by giving an algorithm for explicitly constructing a
random point process. Most of these constructive models are introduced to mimic certain
aspects of the observed clustering pattern or of the dynamical processes leading to the
present state. One of the earliest of these models was Fournier d’Albe’s self-similar clus-
tering hierarchy of clusters within clusters ad infinitum (see Mandelbrot, 1977). Modern
equivalents of this model are the random fractal models like the Soneira-Peebles model
(Soneira & Peebles, 1978) and the Lévy flight models (Mandelbrot, 1977; Peebles, 1980;
Lemson & Sanders, 1991; see also Coleman & Pietronero, 1992). These models were mo-
tivated both by the hierarchical and scaling behaviour of the correlation functions and by
the fact that deeper and deeper surveys showed structures on ever increasing scales.
A historically important constructive model is the cluster model introduced by Neyman & Scott to describe the Shane-Wirtanen galaxy counts (Neyman & Scott, 1953; Neyman, Scott & Shane, 1954). In this model, clusters with a certain spatial form are distributed uniformly through space and the various free parameters, such as the mass and size of the clusters, are fitted to the observations. While this specific model is no longer thought to be acceptable for the description of galaxy clustering, the rigorous mathematical analysis of it by Neyman & Scott has been of great importance for the development of the study of point processes in general.

Another model, motivated by the frothy appearance of large scale structure, where galaxies are positioned on walls and filaments surrounding large regions almost devoid of galaxies, is the Voronoi tessellation. In a cosmological context this model has been advocated and analyzed mainly by Icke and van de Weygaert (1987, 1989). A nice aspect of this model is that some results can be obtained analytically, but the physical motivation for their appearance in a cosmological context is not very strong. This however is true for all the models mentioned. None of them follow directly from the dynamical equations of a perturbed Friedman universe.

There is a class of models for which this dynamical basis does exist, namely those following from numerical N-body simulations. For these models complete statistical information on the (linear) initial conditions is available, but the complicated nature of the non-linear interactions during the simulation destroys this information, and just as for the other constructive algorithms mentioned above, the statistical properties of the resulting point process can only be estimated \textit{a posteriori}.

The second method for modeling the galaxy distribution consists of specifying some specific form for the correlation functions. The most popular of these is the hierarchical model, where the N-point functions are expressed as products of the two-point function, e.g.

\[ \xi_N(\lambda x_1, ..., \lambda x_N) = \lambda^{-\gamma(N-1)} \xi_N(x_1, ..., x_N) \]  

(3.1)

(e.g. W79). This form is consistent with observations on the three-point function and supposedly also with observations for higher orders, but is not established beyond doubt (Matarrese et al., 1986). Attempts have been made to derive this form and the values of the coefficients from solutions of the BBGKY hierarchy (e.g. Peebles, 1980, § 68), but also there a consensus has not been reached (Fry, 1984; Schaeffer, 1985; Hamilton, 1985). A similar model that has been thoroughly investigated is defined by the scaling ansatz

\[ \xi_N(\lambda x_1, ..., \lambda x_N) = \lambda^{-\gamma(N-1)} \xi_N(x_1, ..., x_N) \]  

(3.2)

(Balian & Schaeffer, 1989). This model is suggested by the power-law shape of the two-point correlation function (but see Lemson & Sanders, 1991), together with the above mentioned hierarchical form of the three-point function. Both these models have been used to derive the characteristics of the corresponding counting probabilities. The most thorough analysis along these lines is the one by Balian & Schaeffer (1989). But even they need to make strong assumptions to obtain even only approximate results. The strongest result is that for the scaling model, all quantities should be only dependent on the density, the volume and the two-point correlation function in the combination $nV \xi_2(V)$. This behaviour has indeed been observed to some degree, both in observed galaxy catalogues and in the results
of N-body calculations. The weak point of all this work is that it is very difficult to assess the importance of the assumptions that are made in the course of the derivations and thus to determine how unique the predictions are.

The third way of introducing point process models is by directly proposing a form for the finite volume counting probabilities. Hubble (1936) probably has been the first to propose such a model, when showing that a log-normal model fitted the distribution of projected galaxy counts better than the uniform Poisson model. Lately this model has become popular again mainly due to Coles and Jones (1991), although their work was aimed at continuous density distributions. More recent proposals for the counting probability distribution are the negative-binomial distribution (Fry, 1986; Elizalde & Gaztañaga, 1992) and the generalized Poisson distribution, the last of which was derived from a quasi-equilibrium thermodynamic theory of galaxy clustering by Saslaw & Hamilton (1984, SH84).

All these models for the counting probabilities are essentially modified versions of well known discrete probability distributions, which appear in problems unrelated to spatial statistics. Except for the ‘gravothermal distribution’ (GTD) from SH84, none of them have a physical motivation. The most useful feature is that these distributions are over-dispersed with respect to the Poisson distribution, which allows better fits to the clustered galaxy distribution. In Fig. 1 we show the Poisson distribution, the GTD and the NBD, all with mean value $\bar{N} = 10$, while the variance of the GTD and the NBD is $<(N - \bar{N})^2 >= 100$. The greater dispersion translates itself in excess probabilities, both for finding less and more points than the mean. For counting probabilities this implies for instance that there is a greater probability for finding no points in the sampling volume, just what one would expect for clustered distributions. However, for none of these models has a physically reasonable point process been found with the particular distribution for its counting statistics. In this Chapter I will show that the demand, that the model for the counting probabilities can be realized by a physically reasonable point process, constrains the choice of suitable models.
Correlation functions and counting probabilities

3.2 Correlation functions and counting probabilities

So much so, that the three popular models mentioned above may be excluded as possibly giving the correct description of a point process possibly relevant for galaxy clustering.

An important ingredient of the whole reasoning will depend on the above mentioned results by White (1979). In the next section I will discuss that work and its implications for the present problem. In section 3 I will show that the thermodynamic theory of gravitational clustering must be flawed, since it leads to a probability distribution that is inconsistent with the assumptions underlying its derivation. Generalizations of this distribution are also shown to be inconsistent with the general results by White (1979), and an interpretation of this inconsistency is provided. I will then show some interpretations of the negative-binomial distribution (NBD) as a point process. An argument will be given why also this process is probably inconsistent with the characteristics that a cosmologically relevant point process should possess. In section 5 I will discuss a variety of models in the light of the previous results and discuss their possible relevance for the problem of structure formation theories.

2 Correlation functions and counting probabilities

Many point processes can be defined by their $N$-point probability densities,

$$P(x_1, ..., x_N) dx_1 ... dx_N,$$

(3.3)

which give the probability of finding $N$ particles at the points $x_1$ to $x_N$ within the infinitesimal volumes $dx_1$ to $dx_N$. For these point processes, reduced correlation functions $\xi_N(x_1, ..., x_N)$ are defined in the standard way

$$P(x_1) dx_1 = n \xi_1(x_1) dx_1$$

$$P(x_1, x_2) dx_1 dx_2 = n^2 (\xi_1(x_1)\xi_1(x_2) + \xi_2(x_1, x_2)) dx_1 dx_2$$

$$P(x_1, x_2, x_3) dx_1 dx_2 dx_3 = n^3 (\xi_1(x_1)\xi_1(x_2)\xi_1(x_3) + \xi_1(x_1)\xi_2(x_2, x_3) + \text{cycl.})$$

(3.4)

$$+ \xi_3(x_1, x_2, x_3)) dx_1 dx_2 dx_3$$

etc.

with $n$ the number density and $\xi_1 \equiv 1$ (see Peebles (1980) and Chapter 2 for more details).

This description in terms of probability densities is not applicable to all point processes. The most general description is in terms of the whole family of finite dimensional probability distributions, defined on the set of finite sub-volumes

$$P(N(V_1) = N_1, N(V_2) = N_2, ..., N(V_m) = N_m) = \text{Prob}(V_i \text{ contains } \{N_i\} \text{ points, } i = 1, m)$$

$$\forall m \in \mathbb{N}, \forall V_i \subset \mathbb{R}^3,$$

(3.5)

defined in the Introduction. These distributions are essentially integer valued random measures on the space under consideration, and such measures do not necessarily allow a description in terms of a local density function, which is the basic assumption for defining the correlation functions. An example where a probability density in the usual sense cannot be defined is a lattice process. Here the origin and orientation of a lattice are chosen...
randomly, after which all the points of the process are placed on the nodes of the lattice (e.g. Stoyan & Stoyan, 1994). One may in such cases define a density but only within the theory of generalized functions. We will encounter another example later.

In this Chapter we will only concern ourselves with the single volume probability distributions

\[ P_N(V) = \text{the probability of finding exactly N points in volume } V. \]  

(3.6)

In W79, a relation was derived, expressing \( P_N(V) \) in terms of the correlation functions. The void probability function, \( P_0(V) \), the probability of finding no points in a certain volume, is given by

\[
P_0(V) = \exp \left( \sum_{i=1}^{\infty} \frac{(-n)^i}{i!} \int_{V_i} \xi_i(x_1, ..., x_i) dx_1 ... dx_i \right)
\]

\[
= \exp \left( \sum_{i=1}^{\infty} \frac{(-nV)^i}{i!} \xi_i(V) \right). \tag{3.7}
\]

For the functions \( P_N(V), \ N \geq 1 \), White derived analogous expressions, which he showed could be generated by the void probability function as follows:

\[
P_N(V) = \frac{(-n)^N}{N!} \frac{\partial^N P_0(V)}{\partial n^N}. \tag{3.8}
\]

In Chapter 2 I have shown how this generating relation may be generalized to multiple volume counting probabilities. This relation (3.8) between the void probability function and the rest of the probability distribution is a very strong result and will be an important tool in testing probability distributions for their consistency with these results. An easy example is provided by the Poisson distribution

\[
P_N^{\text{Poisson}}(V) = \frac{(nV)^N}{N!} e^{-nV} = \frac{(-n)^N}{N!} \frac{\partial^N e^{-nV}}{\partial n^N}. \tag{3.9}
\]

In the mathematical literature this result seems to be unknown. In DV-J88 it is shown that so called simple point processes\(^1\) are completely specified when the void probabilities \( P_0(V_1, V_2, ..., V_M) \) are known for a large enough family of sub-volumes \( V \) and for all \( M \). However, no 'constructive' algorithm is presented to derive the other counting probabilities from knowledge of the void probabilities. In the form presented by White, relation (3.8) is first of all an economical way of writing down the whole probability distribution from the expression for the void probability in Eq. 3.7. Even so, the further analysis of these results has proven very complicated (see for instance Fry, 1984, 1985, 1986; Schaeffer, 1985; Balian & Schaeffer, 1989). Especially since the higher order correlation functions \( \xi_j \) increase strongly with \( j \), one cannot simply truncate the sums at low orders. Even

\(^1\)In a simple point process, the probability that two point occupy the same spot, vanishes with probability 1. Non-simple processes may be generated by placing point clusters on the points defined by a simple point process.
the simple constraints \( 0 \leq P_j \leq 1 \), for all \( j \) are hard to test for all but the most trivial assumptions for the form of the correlation hierarchy (W79).

On the other hand, one may view equation (3.8) as a constraint rather than as a generating relation. Just as the relation holds for the Poisson distribution, it should hold for all model distributions aimed at describing the galaxy distribution, insofar the underlying assumptions for the model agree with those underlying White’s derivation. Since it is difficult to extract closed expressions for the counting probabilities both from the correlation function approach and from constructive algorithms for the point process themselves, various models have been proposed without such motivation. These are the main object of study in the rest of this Chapter.

First we will discuss one more interpretation of relation (3.8), provided by Bertschinger (1992). Bertschinger interprets the derivatives \( \partial^N / \partial s^N \) using the process of random thinning. Thinning is an operation on a point process that lowers the average density, but leaves the correlation functions intact. It is defined as follows: from a point process with a non-vanishing density \( n \), points are removed with probability \( q = 1 - p \), \( 0 < p \leq 1 \). This leads to a point process with a reduced density \( pn \). The fact that this operation leaves the correlation functions unchanged is best appreciated from its effect on the probability generating function. This generating function is defined by

\[
g(s; V) = \sum_{N=0}^{\infty} s^N P_N(V). \tag{3.10}
\]

When sufficient conditions are met, the generating function contains all the information of the probability distribution, which is generated by

\[
P_N(V) = \left. \frac{1}{N!} \frac{\partial^N g(s)}{\partial s^N} \right|_{s=0}. \tag{3.11}
\]

For the general form of the probability distribution defined by (3.7) and (3.8) the generating function is (Schaeffer, 1985; also Chapter 2))

\[
g(s) = \exp \left( \sum_{i=1}^{\infty} \frac{(nV(s-1))^i}{i!} \xi_i(V) \right). \tag{3.12}
\]

Suppose we have a ‘parent’ process \( P^p_N(V) \), which is thinned as described above to density \( pn \). If we label the thinned ‘offspring’ distribution with \( o \) then the counting probability for this process is given by

\[
P^o_N(V) = \sum_{M=N}^{\infty} \binom{M}{N} p^N q^{M-N} P^p_M(V) \tag{3.13}
\]

where \( q = 1 - p \). This gives the following simple relation between the generating functions of the parent and the offspring processes

\[
g^o(s; V) = \sum_{N=0}^{\infty} s^N \sum_{M=N}^{\infty} \binom{M}{N} p^N q^{M-N} P^p_M(V)
\]
Thus $g^o$ is the generating function for a point process with the same correlation functions as the parent process, but with a density reduced by a factor $p$.

Bertschinger now defines a derivative operator $\Delta/\Delta n$, working on empirically determined probability distributions, by

$$\frac{\Delta P_0(n)}{\Delta n} \equiv \lim_{p \to 1} \frac{P_0^p(n) - P_0^o(pn)}{(1 - p)n}. \quad (3.16)$$

Here $P_0^p(n)$ is the void probability distribution as estimated for a particular realization of a point process, while $P_0^o(pn)$ is the corresponding function for the same realization, thinned to $pn$, $0 < p < 1$. Bertschinger concludes that using this definition for obtaining the counting probabilities $P_N$ for $N \geq 1$, is inferior to the simply estimating these from volume sampling. The reason I discuss this interpretation here is because it will reappear later in a somewhat different form. Saslaw & Sheth (1993) adopt this interpretation to resolve an inconsistency between the work of White (1979) and the results of the gravothermal theory. As I will show in the now following section, that resolution is wrong.

3 The thermodynamic theory of gravitational clustering

In an attempt to rigorously treat the dynamics in the non-linear regime of structure formation, Saslaw & Hamilton (1984, SH84) developed a thermodynamical theory of gravitational clustering. Using standard results from statistical mechanics and thermodynamics plus some motivated guesses, they were able to derive an expression for the finite volume counting probability distribution

$$P_N^b(V) = \frac{nV(1 - b)}{N!} (nV(1 - b) + bN)^{N-1} \exp(-nV(1 - b) - bN) \quad (3.17)$$

(we will from now on refer to this distribution as the gravothermal distribution (GTD))

The parameter $b$, which in the gravothermal theory is a function both of density and of temperature $T$, is defined by

$$b = \int_0^\infty r\xi_2(r,..)dr. \quad (3.18)$$

An important assumption in SH84 is that $b$ acquires the form

$$b = \frac{b_0 nT^{-3}}{1 + b_0 nT^{-3}}. \quad (3.19)$$
As defined in Eq. (3.18), $b$ is independent of volume, and the specific ansatz in Eq. (3.19) was made to be able to complete the calculation of the probability distribution. In later work, Saslaw et al. (1990) used more general arguments to support the ansatz (3.19). Sheth (1995) has shown that with different ansatzes the equations may also be solved, which allowed him for instance to derive the negative binomial distribution.

Some statistical properties of the GTD are derived in Saslaw (1989), while a discrete probability distribution of the same form has been studied extensively under the name ‘Generalized Poisson Distribution’ by Consul (1989). Saslaw determines the generating function for the GTD:

$$g_{sh}(s) = \sum_{N=0}^{\infty} s^N P_{N}^{sh} = \exp \left\{ -N + N \left[ b + (1 - b) \sum_{N=1}^{\infty} N^{N-1} \frac{N^{-1}}{N!} b^{N-1} e^{-Nb} s^N \right] \right\}. \quad (3.20)$$

From the form of this generating function it follows that the GTD with $b$ constant is infinitely divisible. According to Saslaw, this implies that non-overlapping volumes are stochastically independent and this would imply that the distribution of counting probabilities for cells projected on the sky would have the generalized Poisson form if this were true for the underlying spatial distribution. This is however not a general property of infinitely divisible point processes. An example of an infinitely divisible point process is provided by the general Poisson cluster process with clusters of finite spatial extent. Non-overlapping volumes at a separation of the order of a typical cluster size will clearly not be independent. Under certain conditions the class of Poisson cluster processes and the class of infinitely divisible processes are in fact identical (DV-J88, Matthes et al., 1978). From this identification it can be understood why the projection of such a process may be identical to the process itself. This is however a consequence of the Poisson cluster nature of the process, not of the property that non-overlapping volumes would be independent.

Saslaw (1989) furthermore presents two point processes which have the GTD for their counting probability distribution. The first is a ‘decomposition in Poisson cluster centers’, where the clusters are Poisson distributed and contain a random number of galaxies, the value of which is drawn from the Borel distribution:

$$p_0 = 0, \quad p_j = \frac{(jb)^{j-1}}{j!} e^{-jb}, \quad j > 0. \quad (3.21)$$

The second point process that Saslaw proposes is essentially the same as the first, but with the clusters each concentrated on points, i.e. the spatial structure of the clusters is trivial. In fact, the Poisson-cluster decomposition only gives the required probability distribution when the spatial size of the clusters vanishes, so in fact Saslaw only provides one example. For this point-cluster process, non-overlapping volumes are indeed independent.

---

2 A discrete probability distribution is called infinitely divisible if, for any value of $k$, it can be obtained as the distribution of the sum of $k$ independently and identically distributed random variables, see for example Grimmett & Stirzaker, (1992). This translates itself in the following identity for the generating functions: $g(n, s) = [g(n/k, s)]^k$. An infinitely divisible point process can, for each $k$, be obtained from $k$ i.i.d. distributed point processes, its counting probabilities must than be infinitely divisible (see e.g. DV-J88, Matthes et al., 1978)
### 3.1 The GTD as a point process

I will now show that there are problems with the interpretation of the gravothermal distribution as a cosmologically interesting point process. We will approach the GTD at increasing levels of generality, depending on the freedom we will allow for the form of the parameter/function \( b \). First we assume that \( b \) is constant, independent of both density and volume. It is trivial to show that in this form the GTD does not fulfill relation (3.8).

The cause of this is related to the interpretation of this GTD as a point process. As Saslaw (1989) showed, a point process with Poisson-distributed point clusters provides a realization of this probability distribution, and in fact it can be shown that this is the only such point process. To see this we first determine the probability that a volume contains \( N \) points, under the condition that it contains at least 1.

\[
P(N|N \geq 1) = \frac{P(N)}{P(N \geq 1)} = \frac{P(N)}{1 - P(0)} = \frac{nV(1 - b)(nV(1 - b) + bN)^{N-1}e^{-nV(1-b)-Nb}}{(1 - e^{-nV(1-b)}) N!}.
\]  

(3.22)

In the limit of vanishing volume, \( V \to 0 \), this conditional probability becomes

\[
\lim_{V \to 0} P(N|N \geq 1) = \frac{(bN)^{N-1}}{N!} e^{-Nb}.
\]

(3.23)

This result implies that the probability of more than one point occupying the same spot is non-vanishing and is given by the Borel distribution (3.21).

This result gives no clue about the distribution of these clusters themselves, but we can show that they must be distributed uniformly. This follows from a generalization of the results of White to point processes consisting of randomly distributed point clusters. The distribution of these clusters is assumed to be described by the hierarchy of \( N \)-point correlation functions as in (3.5) and the generating function for the cluster counting probabilities is therefore

\[
g^*(s) = \exp \left( \sum_{N=1}^{\infty} \frac{(-sV)^N}{N!} (1 - s)^N \xi_N \right).
\]

(3.24)

The quantities pertaining to this cluster support process are denoted by \( a^* \). The probability that in the full point process the volume \( V \) contains \( N \) points is given by the following expression

\[
P_N(V) = \sum_{k=0}^{\infty} P_k(V) p_N^{*k}.
\]

(3.25)

Here \( P_k(V) \) is the probability that there are \( k \) clusters in \( V \), and \( p_N^{*k} \) is the \( k \)-times convolution of the sequence \( p_t \), which gives the probability that the sum of \( k \) random variables, each drawn from the distribution \( p_t \), is exactly \( N \). In terms of the generating functions this result simplifies considerably

\[
g(s) = g^*(G(s)).
\]

(3.26)
where \( G(s) \) is the generating function of the cluster size probability function \( p_j \):

\[
G(s) \equiv \sum_{j=1}^{\infty} s^j p_j .
\]

The average density of the resulting process \( n = \lambda \), where \( \lambda = \sum j p_j \). Using this we obtain for the generating function of a general non-simple point process :

\[
g(s) = \exp \left( \sum_{N=1}^{\infty} \frac{(-n V/\lambda)^N}{N!} (1 - G(s))^{N \xi_N} \right) .
\]

When substituting \( \lambda = (1 - b)^{-1} \), the mean of the Borel distribution, and choosing \( \xi_N \equiv 0 \), \( N \geq 2 \), one obtains the expression for the generating function (3.20) derived by Saslaw (1989), with \( G(s) \) the generating function for the Borel distribution (3.21). The vanishing of all correlation functions of order \( N \geq 2 \) implies that the clusters are distributed Poisson like.

This point cluster interpretation may be evaded when \( b \) depends explicitly on the volume, and we will also allow a dependence on density. We will not assume an explicit form for \( b(n, V) \) and an argument using the conditional probabilities as in (3.22) can not be used. We will therefore investigate the consistency with the constraint relation, (3.8). It is easily seen that forcing this relation to hold for all values of \( N \), will lead to an infinite number of differential equations in \( b \) : equality for the relation for \( P_N \) gives rise to an \( N \)th order differential equation, \( d^N b / d n^N = f_N(n, b, \beta, ..., b^{(N-1)}) \). In Appendix A the resulting differential equations are given for the first four levels of the hierarchy. For the GTD to be consistent with (3.8), all these differential equations should be solved by the same function \( b \). In figure (1) the results of numerically integrating the first four equations are presented. Only the initial condition for the equation for \( \beta \) could be freely chosen, all the others follow from this one. By choosing \( b(0.001) = 0.001/(1 + 0.001) \) we may compare the numerical solutions for \( b(n) \) to the ansatz (3.19) from SH84, with \( b_0 T^{-3} = 1 \). This ansatz is also shown in the figure. Clearly the four solutions are all different, and they are also different from the functional form used in the derivation of the GTD. This shows that also when \( b \) is a general function of \( n \) and \( V \), the results of SH84 and W79 are inconsistent.

This can also be shown without numerically integrating the individual differential equations. We will use the following relation which can easily be derived from Eq. 3.8,

\[
P_{N+1} = \frac{N}{N + 1} - \frac{n}{N + 1} \frac{dP_N}{dn} .
\]

Substituting the GTD in this relation leads to a first order differential equation for all values of \( N \). These should all be equivalent which can most easily be tested by reducing two of them to an algebraic equation upon eliminating \( \beta \). The result for \( N = 0 \) and \( N = 1 \) can be reduced to

\[
b \equiv 1 \sqrt{b} = 2 - \frac{4}{1 + e^b} .
\]

The second equation is only solved by \( b \equiv 0 \), which gives the Poisson distribution. The \( b = 1 \) solution is singular: it leads to \( P_0 = 1, P_{N>0} = 0 \), i.e. a typical volume almost surely\(^3\)
contains no points. This would seem to imply \( n \equiv 0 \), but in fact it can be realized with a finite number of infinitely massive clusters distributed throughout an infinite space. The interpretation of this inconsistency is not so clear as for the case \( b = \text{constant} \), and we will devote the next section to this.

### 3.2 Interpretation and discussion

We have thus shown that the only point process consistent with the result from SH84 is a Poisson point-cluster process. For such a point process the derivation of Eq. 3.8 in W79 is not applicable, and the inconsistency should come as no surprise. In the derivation it is explicitly assumed that one may choose a subdivision of space in cells that are small enough that the probability of more than one point occupying the same cell vanishes (see also Chapter 2). This is clearly not true for point clusters.

From the first part of the previous paragraph, it is clear that the original derivation of the GTD, applying thermodynamics to gravitational clustering, is flawed. In that derivation (SH84), it is explicitly assumed that ‘.. infintesimally close binaries must be excluded’ since they would make infinite contributions to the grand partition function. This is at variance with the resulting probability distribution which can only be obtained using point clusters. Also, such a process is not of interest as a viable model for the galaxy distribution. This Poisson point-cluster process does explain the various properties of the GTD obtained by Saslaw (1989). The infinite divisibility directly follows from the equivalence of Poisson cluster processes and infinitely divisible processes (see footnote 2). The stochastic independence of the non-overlapping volumes and the invariance of the distribution under projections follow from the point structure of the clusters.

All these characteristics are no longer generally true when the basic parameter of the GTD, \( b \), is allowed to be a general function of density and volume. From Eq. 3.23 we see that when \( b \to 0 \) as \( V \to 0 \), the point cluster interpretation may be evaded. Still, the discrepancy with White’s relation (3.8) remains, even for \( b = b(n, V) \), as was shown by the fact that the differential equations of Appendix A could only be solved by \( b \equiv 0 \) or
b ≡ 1. The interpretation of this discrepancy is more complicated. Saslaw & Sheth (1993) also observed that for constant b, the GTD did not conform to White’s relation. Their interpretation was that this signified that the ‘naive’ derivatives w.r.t. n, when applied to $P_0^{SH}(n)$ did not explicitly leave the correlation functions constant, as is implied by Eqs. 3.7 and 3.8. They asserted that one needs to interpret the derivatives using random thinning.

From the definition (3.10) of the generating function $g(s; n, V)$, one sees that

$$g^{th}(s; n, V) = \frac{\partial P_0^{th}(V)}{\partial n} = \left. \frac{\partial g^{th,p}(s; n, V)}{\partial n} \right|_{s=0}$$

$$= \lim_{\delta n \to 0} \frac{g^{th,p}(0; n, V) - g^{th,0}(0; n - \delta n, V)}{\delta n}$$

$$= \lim_{q \to 0} \frac{g^{th,p}(0; n, V) - g^{th,0}(0; n(1 - q), V)}{qn}$$

$$= \lim_{q \to 0} \frac{g^{th,p}(0; n, V) - g^{th,p}(q; n, V)}{qn}$$

$$= \frac{-1}{n} \frac{dg^{th,p}(s)}{ds} \bigg|_{s=0} = \frac{-1}{n} P_1(V). \quad (3.31)$$

In the second line, for the generating function at $n - \delta n$, they substitute the generating function of the thinned offspring distribution, $g^{th,0}$. The equality in the last line follows from the definition of the generating function, and it is easy to see that in this interpretation, White’s relation is valid for any distribution. However, while it is true that $P_0^{th}(n, V) = g^{th,p}(0; n, V)$, for the GTD it is not true that $P_0^{th}(n - \delta n, V) = g^{th,0}(0, n - \delta n, V)$. The second line is therefore wrong. This is proven in Appendix B, where it is shown that forcing $P_0^{th}(n - \delta n, V) = g^{th,0}(0, n - \delta n, V)$ leads to the same set of differential equations for b as derived in Appendix A, which we know to be inconsistent unless b ≡ 0 or b ≡ 1.

The basic idea of the gravothermal theory is that, as the universe expands, the correlation functions evolve along with the density. If this is a viable theory for galaxy clustering, it may indeed happen that at each time the counting probabilities are given by the GTD, even though thinning of the distribution does not preserve this form. However, one must then still be able to find a set of correlation functions that reproduces the GTD from the expressions for the probability distributions as derived by White. In the next section we provide an argument why this will not be possible, but even if this were possible, it would introduce complications of a different kind, for we would never be able to determine that the galaxies actually were distributed according to the GTD. For example, volume limited galaxy catalogues extracted from apparent magnitude limited catalogues are necessarily limited by an absolute magnitude cut-off (see Chapter 1). This implies that in practice we will never observe the whole distribution, but thinned versions of it. If luminosity is uncorrelated with the spatial distribution of the galaxies, the thinning introduced by an absolute magnitude cut-off preserves the form of the correlation functions. This again implies, that these limited catalogues will not be distributed according to the GTD and we will never be able to test the hypothesis.

Even so, the GTD does give a good description of the counting probabilities in some
instances. An example is shown in Fig. 3, where for various volumes the counting
probabilities are shown for two large scale cosmological simulations that are described in detail
elsewhere in this thesis. Also shown in that figure are the corresponding GTD’s, with pa-
parameter \( b \) determined from the variance, \( \langle (N - \overline{N})^2 \rangle = \overline{N} / (1 - b)^2 \). Especially for the
\( n=0 \) simulation, the GTD fits the observations remarkably well, with a value for \( b \) which
is fairly constant for this range of volumes. We know that a Poisson point-cluster process
with Borel distributed cluster sizes is exactly described by the GTD with constant \( b \). One
may then expect that, when the clusters are given finite spatial sizes, and when the cluster
masses are given by the Borel distribution, sampling with volumes that are large w.r.t. the
typical cluster size, may be well approximated by the GTD. The question is why the cluster
mass function would follow a Borel distribution. A possible cause of this can be obtained
from results derived in Appendix D, which are further discussed in the concluding section.

4 Negative binomial point processes

The second model for the counting probabilities that we will investigate is the negative
binomial distribution (NBD). This distribution is defined by (DV-J88)

\[
p_N = \frac{(\alpha - 1 + N)!}{(\alpha - 1)!N!} \left( \frac{\mu}{1 + \mu} \right)^\alpha \left( \frac{1}{1 + \mu} \right)^N.
\]

(3.32)

The mean \( \overline{N} = \alpha / \mu \) and the variance \( \langle (N - \overline{N})^2 \rangle = \overline{N}(1 + 1/\mu) \). An example of its
appearance as a discrete probability distribution is provided by the probability \( p_i \) that the
\( \alpha’ \)th success (\( \alpha \) integer) occurs at the \( (i + \alpha)’ \)th try, when drawing with success probability
\( p = 1/(1 + \mu) \). The increased variance given by the parameter \( 1/\mu \) allows one to describe
models that are over dispersed w.r.t. the Poisson distribution, i.e. \( \sigma > \sqrt{\overline{N}} \). The NBD has
therefore also been popular in the description of galaxy counts.

Carruthers & Minh (1983) investigated the probability that Zwicky clusters contained
\( N \) points, \( P_N \), and proposed the NBD with \( \alpha = 6 \) for this multiplicity function. They
explained the appearance of the NBD here and in other circumstances, such as multi-hadron
final states, as signifying a ‘universal mechanism ... involving the production of objects by
means of a suitably chaotic process.’ Carruthers (1991) investigated the moment structure
of the hierarchical model and noted the correspondence with the moment structure of the
NBD. Fry et al. (1989) determined the void probabilities for the Pisces-Penseus redshift
compilation of Giovanelli & Haynes and found that the negative binomial model fitted the
data very well. Elizalde & Gaztañaga (1992) investigated the NBD in the context of void
probabilities for the CfA-catalogue (Huchra et al., 1983). These authors also propose a
stochastic model which should provide realizations of a negative-binomial point process.
Finally, Fry (1986), Szapudi & Szalay (1993) and Borgani (1993) have investigated the
NBD using the relation between the counts-in-cells statistics and correlation functions as
derived in W79.

In spite of this great number of studies, there are only two examples known of point
processes which have the NBD for their counting probabilities (DV-J88; Gregoire, 1983).
The first example is a so called homogeneous mixed Poisson process. Each realization of
Figure 3. Counting probabilities for two cosmological simulations with power-law initial fluctuation spectrum, $P(k) \propto k^n$. Upper diagram shows results for four different sampling volumes (indicated by the numbers in the frames, 25 corresponds to a cubic grid of $25^3$ cells etc.) from an $n=-1$ simulation, the lower diagram shows the results for $n=0$. The dashed line gives the negative binomial model with the same variance as the data. The full line the corresponding gravothermal distribution. The last is seen to provides by far the better fit, especially for the $n=0$ simulation.
such a process is a Poisson process, but the density is randomly drawn from a continuous random distribution (DV-J88). To obtain the NBD, the density \( \lambda \) must be drawn from the gamma distribution.

\[
p(\lambda)d\lambda = \frac{ip \lambda^{a-1}e^{-b\lambda}}{\Gamma(a)}d\lambda .
\]

For points distributed uniformly with a density drawn from this distribution, the counting probability for a single volume is given by

\[
P_{\text{mixed}}^N \equiv \int_0^{\infty} d\lambda p(\lambda) \frac{(\lambda V)^N}{N!} e^{-\lambda V}
\]

\[
= \frac{\Gamma(a + N)}{\Gamma(a)N!} \left( \frac{b}{b + V} \right)^a \left( \frac{V}{V + b} \right)^N
\]

which equals the expression in (3.32) when we identify \( a = \alpha \) and \( V/b = 1/\mu \equiv <N>/\alpha = nV/\alpha \). The last expression defines the average density of the process, \( n = \alpha/b \).

It is an easy exercise to show that using this definition of \( n \) and for \( \alpha \) constant, the resulting expression fulfills the constraint equation (3.8), From this it immediately follows that \( \alpha \) may not be generalized to a function of the number density \( n \). Allowing \( \alpha = \alpha(n) \) and forcing consistency with (3.8) leads to differential equations in \( \alpha(n) \), just as in the case of the GTD. We already know these differential equations to be solved by the constant function; the uniqueness of solutions of differential equations then shows this to be the only solution. This point process is, according to theorem 7.3.II of DV-J88, the unique simple point process with this counting probability distribution. It is however of no interest for the study of galaxy clustering. The main problem is that mixed Poisson processes are not ergodic. The statistical properties of the point process can not be extracted from spatial averages of a single realization, since each realization is a Poisson process and therefore has trivial spatial statistical properties. If a non-trivial ergodic realization of the NBD point process should exist, the parameter \( \alpha \) necessarily has to be a non-trivial function of the sampling volume. Whether or not such a realization exists is a difficult question, to which I have found a tentative, negative answer. First however we will investigate some limits of the supposed functional form of \( \alpha(V) \), which will lead us to the second realization of the NBD as a point process.

One may constrain the possible asymptotic form of \( \alpha(V) \) using the limits \( V \to 0 \) of the conditional probabilities \( P(N|N \geq 1) \). If we assume that \( \alpha(V) \sim \alpha_0 V^{\beta} \) for \( V \to 0 \) then \( P(N|N \geq 1) \) has the limits 1 for \( N = 1 \) and 0 for \( N \geq 2 \) when \( 0 \leq \beta < 1 \). This is as one expects for a simple point process. For \( \beta = 1 \) however,

\[
\lim_{V \to 0} P(N|N \geq 1) = \frac{1}{N} \left( \frac{n}{n + \alpha_0} \right)^N \frac{1}{\ln(1 + n/\alpha_0)} .
\]

That is, for \( \alpha \sim \alpha_0 V \) as \( V \to 0 \), any possible realization will contain point clusters, with masses determined by the logarithmic distribution (3.35).

This result may also be obtained using an alternative argument. Recall the general
expression for the void probability function of a simple point process :

\[ P_0(n, V) = \exp \left( \sum_{j=1}^{\infty} \frac{(-nV)^j}{j!} \bar{\xi}_j \right). \]  

(3.36)

The void probability function of the NBD can be rewritten in a similar form :

\[ P_0^{NBD}(V) = \left( \frac{\alpha}{nV + \alpha} \right)^\alpha \]

\[ = \exp \left( -\alpha \log(1 + nV/\alpha) \right) \]

\[ = \exp \left( -\alpha \sum_{i=1}^{\infty} \frac{1}{i} \left( \frac{-nV}{\alpha} \right)^i \right). \]  

(3.37)

Comparison with the general expression, (3.36), shows that

\[ \bar{\xi}_j = (j - 1)! \alpha^{1-j} \]  

(3.38)

with special case \( \alpha = 1/\bar{\xi}_2 \). The two-point correlation function for a Poisson distribution of point clusters with integer masses will be proportional to a Dirac delta function with integral

\[ \bar{\xi}_2 \equiv \frac{1}{V^2} \int_{V^2} \xi_2(x_1, x_2) dx_1 dx_2 \propto V^{-1} \]

and thus \( \alpha \propto V \). (3.39)

With \( \alpha \propto V \) and independent of \( n \), the corresponding point process must contain point clusters, and indeed, a Poisson cluster process with clusters masses randomly drawn from the logarithmic distribution

\[ p_N = \frac{1}{N} \left( \frac{1}{1 + \mu} \right)^N \frac{1}{\log(1 + 1/\mu)} \]  

has the NBD for its counting probabilities (DV-J88). The final form is obtained from the defining expression (3.32), by substituting \( \alpha \equiv \mu V/n \):

\[ P_N = \frac{(\mu V + N - 1)!}{(\mu V - 1)!N!} \left( \frac{\mu}{1 + \mu} \right)^{\mu V} \left( \frac{1}{1 + \mu} \right)^N. \]  

(3.41)

For \( \mu = \alpha_0/n \), this reduces to the asymptotic form investigated above, and then also the constraint equation is solved. Thus in this case, in contrast to the GTD, a point process containing point clusters does follow the relation (3.8), even though that relation was derived explicitly excluding the possibility of point clusters.

Since however these point-cluster realizations are just as uninteresting for describing the galaxy distribution as they were in the case of the GTD, we now turn to the question whether or not a simple point process can be constructed that gives negative binomially distributed counting probabilities through spatial clustering of the points within each realization. Elizalde & Gaztañaga (1992) have proposed a model in the Appendix to their paper that should accomplish this. They subdivide space in a grid of cells and distribute
points on the grid, where the probability of a point ending up in a particular grid cell is proportional to the number of points already occupying that cell. They show that the ratio of the number of cells containing \( N \) points over the total number of cells approaches

\[
p_N = \frac{(\alpha - 1 + N)!}{(\alpha - 1)!N!} \left( \frac{\alpha}{nV + \alpha} \right)^N, \tag{3.42}
\]

when the number of cells and the total number of points approaches infinity. The parameter \( \alpha \) in this distribution is related to the proportionality constant which quantifies the strength of the ‘attraction’. However, this process is clearly only an approximation. Only for sampling with the grid cells will this process give the required form for the counting probabilities. Randomly placed volumes of different sizes will not reproduce this result, i.e., this algorithm does not provide the required point process realization.

The now following argument suggests that it is in fact impossible to construct such a point process. This conclusion follows from the relations implied by Eq. 3.38

\[
\xi_j(V) = (j - 1)! \xi_2(V)j^{-1} \quad \forall \; j \geq 3, \; \forall \; V \subset \mathbb{R}^3. \tag{3.43}
\]

The assumption that these relations should be valid for all subvolumes \( V \subset \mathbb{R}^3 \) puts strong constraints on the possible functional forms of the \( \xi_j \). To show that the NBD can be realized as a possibly relevant point process, one should show that these constraints are flexible enough to allow non-trivial, regular solutions, i.e., different from the constant function and not containing delta functions.

In its integral form, the first of this family of equations may be written as

\[
\frac{1}{V^3} \int_V dx \int_V dy \int_V dz \; \xi_3(x, y, z) = 2 \left( \frac{1}{V^3} \int_V dx \int_V dy \; \xi_2(x, y) \right)^2 = \frac{2}{V^4} \int_V dx \int_V dy \int_V dz \; \int_V dw \; \xi_2(x, y) \; \xi_2(z, w). \tag{3.44}
\]

By making specific choices for the volumes over which the integrals should be performed, we can reduce this integral relation to equalities in terms of the functions. We will demand that the equality in (3.43) holds for disconnected volumes and in the limit of vanishing size. To be able to take these limits, we will from now on explicitly discard the possibility that the correlation functions may contain Dirac-delta functions. Using this prescription, in Appendix C the following local equations are derived:

\[
\xi_3(x, x, x) = 2 \xi_2(x, x)^2 \tag{3.45}
\]

\[
2 \left( 2 \xi_3(x, x, x) + 6 \xi_3(x, x, y) \right) = 8 \xi_2(x, x)^2 + 16 \xi_2(x, x) \xi_2(x, y) + 8 \xi_2(x, y)^2 \tag{3.46}
\]

\[
3 \left( 3 \xi_3(x, x, x) + 6 \xi_3(x, x, y) + 6 \xi_3(x, x, z) + 6 \xi_3(y, x, z) + 6 \xi_3(x, y, z) \right) = 2 \left( 3 \xi_2(x, x) + 2 \xi_2(x, y) + 2 \xi_2(x, z) + 2 \xi_2(y, z) \right)^2. \tag{3.47}
\]

Here appropriate limits are implied where equality of two or more arguments might imply the possible divergence of the function values, for instance in the case \( \lim_{x \to y} \xi_2(x, y) \),
when $\xi_2(x, y) \propto |x - y|^{-\gamma}$, $\gamma > 0$. These equations were derived using one, two and three infinitesimal volumes respectively, and from them we can derive a local expression for the three-point correlation function

$$\xi_3(x, y, z) = \frac{8}{9} (\xi_2(x, y) \xi_2(x, z) + \xi_2(x, y) \xi_2(y, z) + \xi_2(x, z) \xi_2(y, z))$$

$$- \frac{2}{9} (\xi_2^2(x, y) + \xi_2^2(x, z) + \xi_2^2(y, z)) \quad (3.48)$$

Using four volumes one can derive another equation expressing a combination of three-point functions in terms of two-point functions (see Appendix C).

$$4 (4 \xi_3(x_1, x_1, x_1) + 6 \xi_3(x_1, x_1, x_3) + 6 \xi_3(x_1, x_1, x_4) + 6 \xi_3(x_2, x_2, x_3) + 6 \xi_3(x_2, x_2, x_4) + 6 \xi_3(x_3, x_3, x_4) + 6 \xi_3(x_1, x_2, x_3) + 6 \xi_3(x_1, x_2, x_4) + 6 \xi_3(x_1, x_3, x_4) + 6 \xi_3(x_2, x_3, x_4) = 2 (4 \xi_2(x_1, x_1) + 2 \xi_2(x_1, x_3) + 2 \xi_2(x_1, x_4) + 2 \xi_2(x_2, x_3) + 2 \xi_2(x_2, x_4) + 2 \xi_2(x_3, x_4))^2 \quad (3.49)$$

Substituting the expression for $\xi_3(x, y, x)$ found before, Eq. 3.48, we obtain an equality in terms of $\xi_2$ alone. By taking the derivatives once with respect to all the four independent coordinates, we obtain the following simple expression,

$$w_2(x_1, x_2) w_2(x_3, x_4) + w_2(x_1, x_3) w_2(x_1, x_4) + w_2(x_1, x_4) w_2(x_2, x_3) \equiv 0 \quad (3.50)$$

in terms of the function

$$w_2(x, y) \equiv \frac{\partial^2 \xi_2(x, y)}{\partial x \partial y} \quad (3.51)$$

Since one can always choose a configuration of points such that all terms in this expression have the same sign, for instance by choosing the positions defining the four vertices of a regular tetrahedron, we see that this equation can only be solved by $w_2(x, y) \equiv 0$. This can only be solved by $\xi_2 \equiv constant$. In Appendix C, apart from deriving the above equations, I give a slightly different version of this argument which does not need the limit of infinitesimal volumes.

These same arguments can also be applied to the case of the GTD. From the expressions for the variance of the counts in cells (SH84)

$$<(N - \overline{N})^2> = \frac{\overline{N}}{(1 - b)^2} = \overline{N} + \overline{N}^2 \xi_2 \quad (3.52)$$

one obtains

$$1 - b(n, V) = \frac{1}{\sqrt{1 + \overline{N}^2 \xi_2(V)}}$$

$$= 1 + \sum_{j=1}^{\infty} \frac{(-\overline{N} \xi_2(V))^j}{j!} \frac{(2j - 1)!!}{2^j} \quad (3.53)$$

\(^4k!! = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot k \text{ for } k \text{ odd}, k!! = 2 \cdot 4 \cdot 6 \cdot \ldots \cdot k \text{ for } k \text{ even.}\)
Alternatively, from the equations for the void probability functions one may deduce

\[
1 - b(n, V) = -\frac{1}{N} \sum_{j=1}^{\infty} \frac{(\bar{N})^j}{j!} \xi_j(V).
\]

By equalizing coefficients of the same powers of \( \bar{N} \) we obtain expressions analogous to Eq. 3.43. The first of these is

\[
\bar{\xi}_3(V) = \frac{9}{4N^2}.
\]

This equation is essentially the same as Eq. 3.45 for the NBD and therefore has no interesting solutions, valid \( \forall V \).

One may object to the freedom of choosing any volume in equation (3.43), since in practice sampling is only performed using macroscopic, simply connected volumes with a limited range of shapes (e.g. Elizalde & Gaztañaga, 1992). When aiming at providing a complete description of a point process, one does however need to specify the counting probabilities for all volumes. It is then clear that a probability distribution for which an exact form can only be given for a limited set of volumes, does not provide such a complete description for the system.

We can say more about the use of simply connected, macroscopic, convex sampling volumes only, for the case of the scaling model as defined by Balian & Schaeffer (1989). For the volume \( V \) with \( |V| \equiv l^3 \) they derive

\[
\bar{\xi}_2(V) = \left( \frac{l_0}{l} \right)^\gamma
\]

and

\[
\bar{\xi}_N = S_N \bar{\xi}_2^{N-1}
\]

where the parameter

\[
S_N \equiv l_0^{-(N-1)\gamma} \int \ldots \int d\mathbf{x}_1 \ldots d\mathbf{x}_N \xi_N(\mathbf{x}_1, \ldots, \mathbf{x}_N)
\]

is an integral over the unit volume \( V/|V| \). Under the extra assumption that the higher order correlation functions are related to the two-point function through some hierarchical functional form (see Eq. 3.1, Balian & Schaeffer derive the approximate result

\[
S_N = N^{-N-2} Q_N \times C_{\text{shape}}.
\]

The shape parameter \( C_{\text{shape}} \) is estimated to lie between \( 0.5^N \) and \( 1.5^N \). Important for the present discussion is the clear indication that \( C_{\text{shape}} \) indeed does depend on the geometry of the volume. In particular, Balian & Schaeffer find different results for cubes and spheres respectively. This clearly is at variance with the simple relations (3.43) and (3.55) for the NBD and the GTD respectively. This implies that the corresponding counting probabilities can not be reproduced by a hierarchical model with the scaling ansatz: the hierarchical relations between the integrated quantities \( \bar{\xi}_N \) and \( \bar{\xi}_2 \) can not be translated to hierarchical relations for the underlying functions themselves.
This has severe implications for the multi-fractal description of hierarchical models, as developed by Borgani (1993). Borgani calculates multi-fractal spectra for various hierarchical models, two of which are the GTD and the NBD. An essential aspect of his argument is that the two-point correlation function of the underlying point process is a power law, $\xi(r) \propto r^{-\gamma}$. From this he derives that hierarchical models are pure fractals with fractal dimension $D = 3 - \gamma$. For the present two models these assumptions are clearly invalid. A more promising relation between fractals and hierarchical models was investigated by Hamilton (1985) and Hamilton & Gott (1988), and earlier already by Mandelbrot (1977) and Peebles (1980). This relation will be further discussed in the concluding section of this Chapter.

5 Summary and discussion

In this Chapter I have investigated stochastic models for the galaxy distribution. At present, the non-linear regime of gravitational clustering is not sufficiently well understood to allow predictions of the statistical characteristics of the galaxy distribution to be made. Stochastic models may then help to elucidate certain characteristic aspects, both of the dynamics of the clustering process and of the final structural properties of the distribution.

As described in the text, these stochastic models can be subdivided in three groups. Models of the first kind propose a definite constructive stochastic algorithm for distributing points in space. Examples of this approach are fractal and multi-fractal algorithms such as the Rayleigh-Lévy random walk (Mandelbrot, 1977; Peebles, 1980, §62; Lemson & Sanders, 1991, Chapter 1; Coleman & Pietronero, 1992) or the Soneira-Peebles fractal (Soneira & Peebles, 1978), Poisson cluster processes (Neyman & Scott, 1953) or Voronoi tessellations (Icke & van de Weygaert, 1987; van de Weygaert & Icke, 1989). N-body simulations may also be included in this group.

Models of the second type are defined by proposing a particular form for the correlation functions. The prime examples of these are the hierarchical and scaling models. These are motivated by the low order behaviour of the observed galaxy correlation functions. These models are used to derive statistical measures that give more information on the higher order clustering properties and are more easily estimated (Balian & Schaeffer, 1989; Fry, 1984, 1985, 1986; Schaeffer, 1985; Hamilton, 1985). In models of the third kind, forms for these higher order statistics, namely the void and counting probabilities, are proposed directly.

The common characteristic of these three methods is that, in principle, they describe the point process completely. This is obviously true for the first method, in which the positions of all individual particles in the realizations are specified. However, for this method it proves difficult to extract information on the more fundamental statistical parameters characterizing the distribution. These can only be estimated from realizations of the process a posteriori.

This is obviously not a problem for the other two methods, which are defined directly in terms of the relevant statistics. There however, the relation with the underlying point process is difficult to extract. For certain hierarchical models, Hamilton (1985) and Hamilton & Gott (1988) assert that a homogenized Rayleigh-Lévy random walk would provide...
Chapter 3  Point processes II: stochastic models

the correct realization (see also Mandelbrot (1977), Peebles (1980) and references therein), but for most of these models this relation remains obscure.

In this Chapter I have investigated two models in detail which are defined by the counting probabilities on finite volumes, namely the gravothermal distribution (GTD)

\[ P_N(n,V) = \frac{nV(1-b)}{N!} (nV(1-b) + bN)^{N-1} \exp \left[ -nV(1-b) - bN \right] \]  

(3.60)

and the negative binomial distribution (NBD)

\[ p_N = \frac{(\alpha - 1 + N)!}{(\alpha - 1)!N!} \left( \frac{\mu}{1 + \mu} \right)^\alpha \left( \frac{1}{1 + \mu} \right)^N. \]  

(3.61)

The GTD is special, in the sense that it was derived in a cosmological context from a physical model for non-linear gravitational clustering by Saslaw & Hamilton (1984, SH84). This physical basis is not well established and as has recently been shown not to be unique: using a different ansatz for the form of the thermodynamic variable \( b \), also the NBD can be derived in this formalism (Sheth, 1995). However, the fact that this theory leads to a definite prediction, which moreover gave very satisfying results when compared to observations and N-body simulations, has provided much interest in the GTD (Barrow, 1992; Bhavsar, 1992).

The present investigations have exposed problems with the interpretation of the GTD as a point process. First it was shown that, as originally formulated, the GTD is inconsistent with the constraint relation

\[ P_N(V) = \frac{(-n)^N \partial^N P_0(V)}{N!} \]  

(3.62)

which expresses the whole probability distribution in terms of the void probability function. This is probably related to the fact that the only point process that leads to the GTD in its original form is a Poisson point-cluster process. These processes are not interesting for cosmology and are inconsistent with the derivation in SH84. It does however explain all the statistical characteristics of the GTD as derived by Saslaw (1989), such as its infinite divisibility, the stochastic independence of non-overlapping volumes and its invariance under projections.

These characteristics are no longer true, nor is there any longer a physical background to the model, when the basic parameter of the GTD, \( b \), is allowed to be a function of density and volume. But also in that case does the GTD not satisfy the fundamental requirement that the void probability be the generating function for the whole probability distribution. This was shown to be equivalent to the fact that the GTD is not invariant under thinning. This spells problems for any supposed hypothesis concerning the galaxy distribution. If galaxy counting probabilities were described by the GTD, any observational test which included limited samples, e.g. by including only galaxies brighter than a certain luminosity limit, would not be described by the GTD. The hypothesis is therefore not testable.

Nevertheless, the GTD, with \( b = b(V) \), seems to fit observations and results from N-body simulations very decently (see e.g. Fig. 2). A tentative explanation for this can be extracted from an interesting result concerning the point cluster interpretation of the GTD. In Appendix D it is shown how a result by Epstein (1983), leading to a discrete
analogue of the Press-Schechter theory for Poisson processes, could be extended to the Poisson point-cluster process corresponding to the GTD. Epstein derivation has its direct analogue in the excursion set derivation of the Press-Schechter formula by Bond et al. (1991). It leads to a multiplicity function for point clusters which is given by the Borel distribution. In Appendix D it is shown that the Borel distribution is invariant under this operation, i.e. at the next level in the clustering hierarchy, where Borel clusters are merged, the multiplicity function remains of the Borel form. In the limit $N \to \infty$, the Borel distribution approaches the form $p_N \sim N^{-3/2} \exp(-cN)$, which is the Press-Schechter mass multiplicity function for an $n=0$ perturbation spectrum, which corresponds to the discrete Poisson distribution. It is also for the $n=0$ simulation that the GTD is seen to fit the counts-in-cells best (Fig. 3). When sampling with volumes large enough for clusters to be unresolved, and when these clusters have the correct multiplicity-function, the GTD may actually be expected to provide a decent fit to the counts in cells, just as is observed for the $n=0$ simulation.

This same argument may explain why the NBD does more poorly in the fits. This distribution may also be obtained from a Poisson point-cluster process, but with a a logarithmic distribution for the cluster size. This distribution does not correspond to a Press-Schechter multiplicity function for any choice of the initial power spectrum. In contrast to the GTD, the NBD does have an interpretation which solves the constraint relation (3.62). The corresponding point process however is again not interesting from a cosmological point of view, since it is a mixed Poisson process and therefore not ergodic: each realization is a homogeneous Poisson process.

It was then shown that a point process can not be constructed for which the counting probability distribution for all volumes, whether connected or not, is given by the NBD. The argument explores the relations between integrals of the general N-point correlation functions and the integrals of the two-point function, that can be derived using the form of the probability distribution and the expressions derived in W79. This argument relies on the assumption that the form for the probability distribution, be it GTD or NBD, is valid for all volumes, connected or not, also in the limit of vanishing sizes. This allows one to extract local relations connecting the correlations functions from the integrated forms. These integrated relations are very similar to the ansatizes made for the hierarchical models, and indeed, one of the intermediate results is a hierarchical expression for the three-point correlation function in terms of the two-point function (Eq. 3.48). However, in contrast to the hierarchical ansatz, which is defined using the local correlation functions, here the relations are phrased in terms of integrals over general volumes. This puts severe constraints on the constituent functions, which in the cases at hand only allow trivial solutions, such as the constant function or Dirac-delta functions.

To summarize, even though the approach using counting probabilities to describe the clustering properties of point sets, allows information to be extracted complementary to the low order correlation functions, the great complications induced by the non-linearities preclude simple models, such as the GTD or the NBD, to describe these sets exactly. Such simple models are related to simple point processes, such as Poisson point-cluster processes or mixed Poisson processes. The fact that such distributions often provide decent fits to the observations can be attributed to various circumstances. First is the increased flexibility
that an extra parameter in the model offers to describe probability distributions that are over dispersed w.r.t. the Poisson distribution. Second is the incomplete sampling of the space of volumes that should be used in the determination of the probability distribution. For instance, the void probability function in principle describes a simple point process completely, when this function is known for all elements of the Borel set of sub-volumes of the space under consideration. This set is far greater than the set of spheres of a limited range of sizes, or ellipsoids of a limited range of shapes, which are most often used in practical estimates of the counting probabilities (e.g. Elizalde & Gaztañaga, 1992). Third, the point-cluster interpretation of the GTD and the NBD can be approximated by finite sized clusters, when the sampling is performed using sufficiently large volumes. If then the mass of the clusters is governed by the appropriate multiplicity function, as was shown to be the case for the GTD and the n=0 simulation, it is no real surprise that this model performs as well as it does, even though it will not be the exact description of the point process.

A final word on the thermodynamic interpretation of the GTD as viewed by Saslaw and coworkers (SH84, Saslaw et al., 1990). In that picture, \( b \) evolves from 0 to 1, where the value \( b = 1 \) corresponds to a state of complete, relaxed clustering. In the point cluster model, this evolution corresponds to the average mass of the clusters approaching higher and higher values, whilst the clusters are at greater and greater separations. This picture is very similar to one that Newton sketched in a letter to Bentley (Newton, 1692). Newton speculated on the evolution of an infinite space, uniformly filled with point masses. He concluded that the masses would approach each other and would coalesce into greater mass concentrations, until finally the Universe would be filled with ‘... great masses scattered at great distances from one to another throughout all [that] infinite space.’ At least for uniform initial conditions, the most obvious a priori assumption one would make for the distribution of mass in the Universe, the results of the present Chapter quantitatively support Newton’s intuitive beliefs.

Acknowledgments I wish to thank Ravi Sheth for the extensive discussions about all the matters presented here and many other related subjects. The work presented here in Appendix D was performed together with him. I also wish to thank Bill Saslaw for his remarks and comments, and the Institute of Astronomy, Cambridge, for its hospitality, which made those discussions possible. I wish to thank A. J. Stam for illuminating discussions and for pointing out Abel’s generalization of the binomial formula.
Appendix A

Here we derive the first four differential equations for the function \( b(n) \), that follow from substituting the gravothermal probability distribution (3.17) in the constraint relation (3.8)

\[
P_N(n, V) = \frac{(-n)^N}{N!} \frac{\partial^N P_0(n, V)}{\partial n^N}.
\] (3.63)

This gives rise to four differential equations, the order of which is \( N \) for the equation arising from substituting \( P_N \).

\[
N = 1
(1 - b(n)) e^{-b(n)} = 1 - b(n) - n \bar{b}(n)
\] (3.64)

\[
N = 2
\frac{(1 - b(n)) n (1 - b(n) + 2 b(n))}{n} e^{-2b(n)} = 2 \bar{b}'(n) + (-1 + b(n) + n \bar{b}'(n))^3 + n \bar{b}''(n)
\] (3.65)

\[
N = 3
\frac{(1 - b(n)) (n + 3 b(n) - n b(n))^2}{n^2} e^{-3b(n)} =
\frac{(1 - b(n) - n \bar{b}'(n))^3 - 3 \bar{b}''(n) - 3 (-1 + b(n) + n \bar{b}'(n)) (2 \bar{b}'(n) + n \bar{b}''(n))}{-n \bar{b}^{[3]}(n)}
\] (3.66)

\[
N = 4
\frac{(1 - b(n)) (n + 4 b(n) - n b(n))^3}{e^{4b(n)} n^3} =
\frac{(-1 + b(n) + n \bar{b}'(n))^4 + 6 (-1 + b(n) + n \bar{b}'(n))^2 (2 \bar{b}'(n) + n \bar{b}''(n))}{+ 3 (2 \bar{b}'(n) + n \bar{b}''(n))^2 + 4 \bar{b}^{[3]}(n)} + 4 (-1 + b(n) + n \bar{b}'(n)) (3 \bar{b}''(n) + n \bar{b}^{[3]}(n))
\]

\[+ n \bar{b}^{[4]}(n)
\] (3.67)

Initial conditions may only be freely chosen once, for \( b(n_{in}) \) in the first of these equations. The initial conditions for the equation of order \( N \) follow from this one and the lower order equations.

Appendix B

In this appendix it is shown that demanding that thinning a GTD point process preserves the form of the probability distribution, leads to the same set of differential equations in \( b(n) \) as was derived in Appendix A. This shows the equivalence of taking the derivative w.r.t. the density and the thinning operation.
For the GTD ‘parent’ process, the generating function is given by Eq. 3.20 (Saslaw, 1989). According to equation (3.15), the generating function for the thinned ‘offspring’ distribution is given by

\[ g_{GTD}^\circ (s) = \exp \left\{ -\bar{N} + \bar{N} \left[ b + (1 - b) \sum_{N=1}^\infty \frac{N^{N-1}}{N!} b^{N-1} e^{-N\bar{b}} (q + ps)^N \right] \right\} \]  

(3.68)

If the thinned distribution should again be a GTD, this \( g_{GTD}^\circ (s) \) should be of the form (3.20), where we must interpret \( b \) as a function of \( q \). For clarity I will write this function as \( \tilde{b}(q) \), where it is understood that \( \tilde{b}(0) = b \). We get

\[ g_{GTD}^\circ (s) = \exp \left\{ -pN + pN \left[ \tilde{b} + (1 - \tilde{b}) \sum_{N=1}^\infty \frac{N^{N-1}}{N!} \tilde{b}^{N-1} e^{-N\tilde{b}} s^N \right] \right\} \]  

(3.69)

The resulting expressions should be equal for all values of \( s \), which is equivalent to demanding equality of the corresponding coefficients. The first of the two expressions can be rewritten to

\[ g_{GTD}^\circ (s) = \exp \left\{ -\bar{N} + \bar{N} \left[ b + (1 - b) \sum_{M=0}^\infty \frac{(ps)^M}{M!} \sum_{N=1}^{\infty} \frac{N^{N-1}}{(N-M)!} \tilde{b}^{N-1} e^{-N\tilde{b}} q^{N-M} \right] \right\} \]  

(3.70)

Equating the coefficients for \( M = 0, 1 \) and 2 gives three equations for \( \tilde{b} \):

\[ M = 0 : \quad -p\bar{N} + p\bar{N}b \equiv -\bar{N} + \bar{N}b + (1 - b) \sum_{N=1}^\infty \frac{N^{N-1}}{N!} b^{N-1} e^{-N\tilde{b}} q^N \]  

(3.71)

\[ M = 1 : \quad p\bar{N}(1 - \tilde{b}) e^{-\tilde{b}} \equiv p\bar{N}(1 - b) \sum_{N=1}^\infty \frac{N^{N-1}}{(N-1)!} b^{N-1} e^{-N\tilde{b}} q^{N-1} \]  

(3.72)

\[ M = 2 : \quad p\bar{N}(1 - \tilde{b}) \tilde{b} e^{-2\tilde{b}} \equiv \frac{p^2}{2} \bar{N}(1 - b) \sum_{N=2}^\infty \frac{N^{N-1}}{(N-2)!} b^{N-1} e^{-N\tilde{b}} q^{N-2} \]  

(3.73)

Using the first of these equations we can rewrite the right-hand sides of the other two, which leads to two differential equations in \( \tilde{b} \).

\[ M = 1 : \quad (1 - \tilde{b}) e^{-\tilde{b}} \equiv -\frac{\partial}{\partial q} \left[ (1 - q)(1 - \tilde{b}) \right] \]

\[ = 1 - \tilde{b} + (1 - q) \frac{\partial \tilde{b}}{\partial q} \]

\[ = 1 - \tilde{b} - \frac{\partial \tilde{b}}{\partial p} \]  

(3.74)

\[ M = 2 : \quad (1 - \tilde{b}) \tilde{b} e^{-2\tilde{b}} \equiv \frac{p}{2} \frac{\partial^2}{\partial q^2} \left[ (1 - q)(1 - \tilde{b}) \right] \]

\[ = -(1 - q) \frac{\partial \tilde{b}}{\partial q} + \frac{(1 - q)^2}{2} \frac{\partial^2 \tilde{b}}{\partial q^2} \]

\[ = \frac{p}{p^2} \frac{\partial \tilde{b}}{\partial p} + \frac{p^2}{2} \frac{\partial^2 \tilde{b}}{\partial p^2} \]  

(3.75)
Appendix C

In this appendix I will derive equalities connecting the three and two-point correlation functions. These equalities will be obtained by making suitable choices for the volume $V$ in the integral equation derived in section 4:

$$V \int_V dx \int_V dy \int_V dz \xi_3(x, y, z) = 2 \left( \int_V dx \int_V dy \xi_2(x, y) \right)^2. \quad (3.76)$$

By choosing the volume $V$ to consist of different, disconnected components, one may obtain expressions for integrals of the form

$$\int_{V_1} dx \int_{V_2} dy \int_{V_3} dz \xi_3(x, y, z) \quad (3.77)$$

in terms of integrals of the form

$$\int_{V_1} dx \int_{V_2} dy \xi_2(x, y). \quad (3.78)$$

By taking limits $|V_i| \to 0$, these expressions can be transformed into local relations for the correlation functions themselves.

First we will derive the relations in their integral form. To this end it is useful to introduce the notation

$$I_2(U, V) \equiv \int_U dx \int_V dy \xi_2(x, y) \quad (3.79)$$

and

$$I_3(U, V, W) \equiv \int_U dx \int_V dy \int_W dz \xi_3(x, y, z). \quad (3.80)$$

These set functions are totally symmetric and linear in all their arguments, i.e.

$$I_2(U, V) = I_2(V, U), \quad (3.81)$$

$$I_3(U, V, W) = I_3(V, U, W) = I_3(U, W, V), \quad (3.82)$$

$$I_2(U_1 \cup U_2, V) = I_2(U_1, V) + I_2(U_2, V) \quad (3.83)$$

and

$$I_3(U_1 \cup U_2, V, W) = I_3(U_1, V, W) + I_3(U_2, V, W). \quad (3.84)$$

For congruent volumes, $U \cong V$, i.e. volumes of equal size and shape,

$$I_3(V, V, V) = I_3(U, U, U) \text{ and } I_3(U, V, V) = I_3(U, U, V) \quad (3.85)$$
with analogous equations holding for $I_2$. With this notation, the constraint relation (3.76) becomes

$$|V|I_3(V, V, V) = 2(I_2(V, V))^2.$$  \hspace{1cm} (3.86)

Using the characteristics of the set functions $I_2$ and $I_3$ it is straightforward to derive the following relations, corresponding to the indicated choices for the volume $V$:

$$V = V_1 \cup V_2, V_1 \cap V_2 = \emptyset \land V_1 \aeq V_2 : \quad 2|V_1|(2I_3(V_1, V_1, V_1) + 6I_3(V_1, V_1, V_2)) = 2(2I_2(V_1, V_1) + 2I_2(V_1, V_2))^2 \quad (3.87)$$

$$V = V_1 \cup V_2 \cup V_3, V_i \cap V_j = \emptyset, i \neq j \land V_1 \aeq V_2 \aeq V_3 : \quad 3|V_1|(3I_3(V_1, V_1, V_1) + 6I_3(V_1, V_1, V_2) + 6I_3(V_1, V_1, V_3) + 6I_3(V_2, V_2, V_3) + 6I_3(V_3, V_3, V_3))$$

$$= 2(3I_2(V_1, V_1) + 2I_2(V_1, V_2) + 2I_2(V_1, V_3) + 2I_2(V_2, V_3))^2.$$

Using these equalities we can derive an expression for $I_3(V_1, V_2, V_3)$ in terms of the $I_2$:

$$|V_1|I_3(V_1, V_2, V_3) = \frac{8}{9} (I_2(V_1, V_2)I_2(V_1, V_3) + I_2(V_1, V_2)I_2(V_2, V_3) + I_2(V_1, V_3)I_2(V_2, V_3))$$

$$- \frac{2}{9} (I_2^2(V_1, V_2) + I_2^2(V_1, V_3) + I_2^2(V_2, V_3)). \quad (3.88)$$

In the limit $|V_1| \to 0, i = 1, 2, 3$, the integral quantities $I_2$ and $I_3$ approach

$$\frac{1}{|V_1||V_2|}I_2(V_1, V_2) \to \xi_2(x_1, x_1)$$

$$\frac{1}{|V_1||V_2||V_3|}I_3(V_1, V_2, V_3) \to \xi_3(x_1, x_2, x_3). \quad (3.89)$$

Here it is assumed that the volume $V_1$ is ‘located’ at position $x_1$, etc. If we now define $\xi_2(x_1, x_1)$ as the appropriate limit for $I_2(V_1, V_1)$ etc, the equations derived above go over into their local counterparts, Eq. 3.45 to 3.47 in section NBD.

Analogously we can derive an equality from choosing $V$ consisting of four congruent subvolumes.

$$4|V_1|(4I_3(V_1, V_1, V_1) + 6I_3(V_1, V_1, V_2) + 6I_3(V_1, V_1, V_3) + 6I_3(V_1, V_1, V_4) + 6I_3(V_2, V_2, V_3) + 6I_3(V_2, V_2, V_4) + 6I_3(V_3, V_3, V_4) + 6I_3(V_3, V_3, V_4) + 6I_3(V_4, V_4, V_4))$$

$$= 2(4I_2(V_1, V_1) + 2I_2(V_1, V_2) + 2I_2(V_1, V_3) + 2I_2(V_1, V_4) + 2I_2(V_2, V_3) + 2I_2(V_2, V_4) + 2I_2(V_3, V_4))^2. \quad (3.90)$$

In the limit of vanishing volumes this expression leads to Eq. 3.49. If we substitute the expression for $I_3(U, V, W)$ from Eq. 3.88 in this equality, we obtain an expression purely in terms of the functions $I_2$. We will only examine the special case where the four volumes are organized in a tetrahedron. In that case all the functions $I_2(V_i, V_j), i \neq j$, reduce to $I_2(V_i, V_2)$, which results in

$$17I_2(V_1, V_2)^2 + 14I_2(V_1, V_1)I_2(V_1, V_2) - I_2(V_1, V_1)^2 = 0. \quad (3.91)$$
This implies that $I_2(V_i, V_j)$ would be independent of the separation between the volumes, which clearly can only be the case for a trivial point process such as a Poisson or a Poisson point-cluster process. This argument relies on the fact that the counting probabilities must follow the NBD for all volumes, connected or not, but does not need the limit of vanishing volumes, as is required by the argument in the main text.

Appendix D

Here we derive an interesting aspect of the gravo-thermal distribution in its interpretation as a Poisson point cluster process. We generalize an approach by Epstein (1983), who determined the mass spectrum of clusters starting from a discrete Poisson distribution of points. The work of Epstein is the discrete analogue of the Press-Schechter theory (Press & Schechter, 1974) and inspired the excursion set approach to this theory by Bond et al. (1991). Here we generalize this approach to include the merging of point clusters, just as Bond et al. were able to treat the merging of cluster in their continuum approach. It will appear that the Borel distribution, and through this the GTD, plays a fundamental role, which may explain its nice fitting properties for certain point processes.

Points are distributed throughout space according to a Poisson process with density $\nu$. Evolution with time is approximated, as in the usual Press-Schechter theory, by calculating the probability that a typical point will belong to a cluster of (integer) mass $k$ at time $T$. This probability is identified with the probability that there is a sphere around the point, which contains exactly $k$ points and has an over density $\delta$ greater than some critical over density $\delta_c(T)$, while there is no larger sphere, containing more than $k$ points, and also at an over density greater than $\delta_c(T)$. This condition finds its direct analogue in the continuous case, where, for a Gaussian random field, Bond et al. (1991) calculate the probability that a point has an over density $\delta(R) \geq \delta_c(z)$ when the field is smoothed at smoothing radius $R$, while for all larger radii the over density is lower than the critical value. The assumption is that at time $T$ this region collapses to form a new (point-)cluster of mass $k$. The dynamical evolution arises from the dependence of the critical over density, $\delta_c(T)$ on time.

Epstein (1983) finds, that the probability that a point lies at the center of a spherical region of mass $k$ and at an over density $\delta_c$ is

$$f_c(k, \delta_c) = \frac{\delta_c}{1 + \delta_c (k - 1)!} e^{-\nu_k} ,$$

where $\nu_k = k/(1 + \delta_c)$. This probability can be transformed into a multiplicity function for the resulting point-cluster mass distribution by dividing this expression by the mass, $k$, and normalizing. This results in the Borel distribution,

$$P(\text{cluster mass} = k) = \frac{(kb)^{k-1}}{k!} e^{-\nu_k} ,$$

with

$$b = \frac{1}{1 + \delta_c} .$$
We have seen in § 3, that a Poisson point-cluster process, where the multiplicity function of the clusters is given by the Borel distribution, gives the GTD for its finite volume counting probabilities. This suggests a stochastic process, leading from the Poisson process to a point-cluster process described by the GTD, with a value of \( b \) evolving from 0 to 1, just as was envisaged in the gravo-thermal theory (SH84; Saslaw et al., 1990). We will now consider whether or not we can generalize this stochastic process to apply it to a distribution of point clusters of a given mass function. For the Borel mass distribution, this turns out to be possible, and the interesting aspect is that the Borel distribution is preserved in the process, i.e. the merging of Borel clusters leads again to Borel clusters, albeit with a different parameter.

We will apply the same algorithm as used by Epstein, but now to a Poisson point-cluster process. The point-clusters are distributed with density \( \lambda_c \), and their mass is determined by the Borel distribution. Following Epstein, it will prove useful to define volumes \( V_k \), defined by the demand that when they contain \( k \) points, their over-density is exactly \( \delta_c \),

\[
V_k = \frac{k}{n(1 + \delta_c)} .
\]  

(3.95)

The probability that at the next step, corresponding to this value of \( \delta_c = \delta_c(T) \), the point at the origin will belong to a cluster of size \( k \), consists of two terms.

\[
f_c(k, \delta_c) = f_c^I(k, \delta) \times f_c^E(k, \delta) .
\]  

(3.96)

First an ‘internal’ term, \( f_c^I \), giving the probability that \( V_k \) contains exactly \( k \) points, under the condition that there is one at the center. This says that the point at the origin is in a cluster of mass \( \geq k \). Second an ‘external’ term, that gives the probability that no larger sphere contains more than \( k \) points and is at an over density \( \delta \geq \delta_c \). This prevents the point to be in a cluster of mass larger than \( k \). This second term is a conditional probability, the probability that \( V_l, l > k \) contains less than \( l \) points while \( V_k \) contains exactly \( k \) points. It is in fact equal to the probability that the shell \( V_l - V_k = V_{l-k} \) contains less than \( l - k \) points, for all \( l > k \), still under the condition that \( V_k \) contains \( k \) points. This equality holds due to the Poisson nature of the cluster distribution: disconnected volumes are stochastically independent. Furthermore, since probabilities are only dependent on the size of the volume, not on the shape, the probability that \( S_j \) contains \( j \) points is equal to the probability that \( V_j \) contains \( j \) points. Following Epstein, we consider the quantity

\[
1 - f_c^E(k, \delta_c) = \text{Prob}(\exists l > 0 \text{ such that } V_l \text{ contains } l \text{ points}) .
\]  

(3.97)

From this equation it follows that in fact \( f_c^E(k, \delta_c) \) does not depend on \( k \). This relation may be written as

\[
1 - f_c^E(\delta_c) = \sum_{l=1}^{\infty} \text{Prob}(V_l \text{ contains } l \text{ points}) \times \text{Prob( no } V_j > V_l \text{ contains } j \text{ points)}
\]

\[
= f_c^E(\delta_c) \sum_{l=1}^{\infty} P(l, V_l)
\]  

(3.98)
where $P(l, V_i)$ is the probability that $V_i$ contains $l$ points. This results in

$$f_c(k, \delta_c) = \frac{f_c^I(k, \delta)}{1 + \sum_{j=1}^\infty P(j, V_j)}.$$  \hfill (3.99)

Epstein uses a short-cut to calculate the value of the denominator, but we will simply perform the sum explicitly. For a Poisson distribution of Borel point-clusters, the term in the summand is the gravothermal probability for the volume $V_j$

$$P(j, V_j) = \frac{nV_j(1 - b)}{j!}(nV_j(1 - b) + j b)^{j-1}e^{-nV_j(1 - b)} + j b = (a - b)\frac{j}{j!}(a j)^{j-1}e^{-a j},$$  \hfill (3.100)

where I have defined

$$a \equiv \left(\frac{1 - b}{1 + \delta_c}\right) + b.$$  \hfill (3.101)

Summing this term gives $(a - b)/(1 - a)$ and

$$f_c^E(\delta_c) = \frac{1 - a}{1 - b} = \frac{\delta_c}{1 + \delta_c},$$  \hfill (3.102)

which is the same result as Epstein finds for the pure Poisson distribution.

The internal probability is a convolution of the probability that the point belongs to a cluster of mass $j$ and the probability that in the rest of the volume $V_k$ there are $k - j$ points.

$$f_c^I(k, \delta_c) = \sum_{j=1}^k \text{Prob}(\text{point in cluster of mass } j) \times \text{Prob}(k - j \text{ other points in } V_k).$$  \hfill (3.103)

The first term equals $(j/\langle j\rangle)p_j$, where $p_j$ is the Borel distribution of having a cluster of size $j$, and the term $j/\langle j\rangle$ comes from the fact that from a cluster of that size, one has $j$ points to choose from, the mean size $<j>$ being needed for the normalization. The second term is the counting probability for this point process, which as we have seen above is simply the gravothermal distribution. This gives

$$f_c^I(k, \delta_c) = \sum_{j=1}^k \frac{j}{j!}(-e^{-j b}nV_k(1 - b)(nV_k(1 - b) + (k - j)b)^{k-j-1}e^{-nV_k(1 - b) - (k-j)b}}{(k - j)!}.$$  \hfill (3.104)

In this expression, $n$ is the density of the point process, which equals the product of the density of the cluster process and the average mass per cluster, $n = n_c <j>$. Now we can reduce the above expression to

$$f_c^I(k, \delta_c) = \frac{(1 - b)^2}{1 + \delta_c} \sum_{j=1}^k \frac{k(k(1 - b)/(1 + \delta_c) + (k - j)b)^{k-j-1}(j b)^{j-1}e^{-(k(1 - b)/(1 + \delta_c) - kb}}{(j - 1)!(k - j)!}.$$  \hfill (3.105)
with the definition (3.101) this becomes

\[
f_c^I(k, \delta_c) = (a - b)(1 - b) \sum_{j=1}^{k} \frac{k(ja - jb)^{j-1}(jb)^{j-1}}{(j-1)!(k-j)!} e^{-ak} \]

\[
= \frac{(1 - b)(a - b)}{b} \left[ \sum_{j=1}^{k} \binom{k}{j} \left( \frac{jb}{ka} \right)^j \left( 1 - \frac{jb}{ka} \right)^{k-j} \right] \frac{k}{k!} e^{-ak} . \tag{3.106}
\]

The term in square brackets can be reduced using Abel’s generalization of the binomial formula (e.g. Riordan, 1968), giving the surprisingly simple result \( b/(a - b) \). Together with the result for the external probability in Eq. 3.102 we obtain for the total probability that a point finds itself in a cluster of size \( k \) at over density \( \delta_c \):

\[
f_c(k, \delta_c) = (1 - a)k\left( \frac{ka}{k} \right)^{k-1} \frac{k}{k!} e^{-ak} . \tag{3.107}
\]

This is of the same form as the original distribution of cluster masses, namely the Borel distribution, with only the value of the parameter changed:

\[
b \rightarrow b + \left( \frac{1 - b}{1 + \delta_c} \right) . \tag{3.108}
\]

The Borel distribution is therefore a stable functional form under this discrete version of the Press-Schechter model. Since the value of the over density \( \delta_c \) decreases with increasing time, this process will lead to multiplicity functions with the parameter \( b \) increasing from 0 to 1, just as in the thermodynamical theory of gravitational clustering by Saslaw and coworkers (SH84; Saslaw et al., 1990). Their theory may therefore be interpreted in this stochastic manner.

Within this same formalism one may also derive discrete analogies for the merging probabilities, as calculated by Bond et al. (1991). The question to be answered would now be, what is the probability that a point belongs to a cluster of size/mass \( k_1 \) at time \( t_1 \) and of mass \( k_2 \geq k_1 \) at \( t_2 > t_1 \).

References

REFERENCES

Consul, P.C., 1989, Generalized Poisson Distributions, Marcel Dekker, Inc
Daley, D.J., Vere-Jones, D., 1988, An Introduction to the Theory of Point Processes,
Springer Verlag, New York (DV-J88)
Londi
Clarendon Press, Oxford
Haynes, M.P., Giovannelli, R., 1988, in Rubin, V.J., Coyne, G.V., eds., Large Scale motions
in the Universe, Princeton University Press, Princeton
Hubble, E., 1936, The Realm of the Nebulae, Oxford University Press, London : Humphrey
Milford
Mandelbrot, B.B., 1977, Fractals, form, chance and dimension, W.H. Freeman and
Company, San Francisco
& Sons
Newton, I., 1692, The Correspondence of Isaac Newton, Vol. III, 1688-1694, ed. H.W.Turnbull,
Peebles, P.J.E., 1980, The Large Scale Structure of the Universe, Princeton University
Press, Princeton
Riordan, J., 1968, Combinatorial Identities, Wiley, New York
Ripley, B.D., 1988, Statistical Inference for Spatial Processes,
Cambridge University Press, Cambridge
Schaeffer, R., 1985, A&A, 144, L1
Stoyan, D., Stoyan, H., 1994 Fractals, Random Shapes and Point Fields, John Wiley & Sons