Clusters in point distributions

J. Betancort-Rijo
Instituto de Astrofísica de Canarias, E-38200 La Laguna, Tenerife, Spain
(Received 7 March 1994)

We present a procedure for obtaining the mean density of clusters and their associated structures in a very general type of distribution. We first consider the one-dimensional case, and then use it to develop the procedure according to the d-dimensional case. These procedures require the probability that a randomly placed body should contain no points. We show how this quantity may be evaluated for different types of distributions. The particular case of spherical clusters in three dimensions is treated in detail.

PACS number(s): 02.50.−r, 05.40.+j

I. INTRODUCTION

The results contained in this work are relevant to a large variety of practical problems. However, to fix ideas, we shall consider the following problem in greater detail: given a distribution of points, we may look for those distinct, nonoverlapping spots where a sphere of radius r may be placed so as to contain at least n points. Any such spot will be termed a cluster of radius r at level n [richness type $f(n)$, for a certain f, in the astronomer's jargon]. Our problem is to express the value of the mean density of the clusters in terms of the probability distribution, $P_n(r)$, for the number of points within a randomly placed sphere. Given the process generating the distribution, this last quantity may be obtained in a straightforward manner. However, the solution of our problem poses several, rather complicated difficulties, some of which have been treated in a series of preliminary works. Only now can the full solution be presented.

The proper treatment of the problem of spherical clusters in three dimensions, as formulated above, is one of the most useful consequences of the general formalism presented here. The number densities of galaxy clusters of various richnesses are readily available cosmological observables. Assume we have a model for the clustering properties of galaxies, obtained either from a theory for the formation of the large scale structure of the Universe or from a plausible extrapolation of the low order galaxy-galaxy correlation function. If we want to check the model's ability to reproduce the observed cluster densities, we must use the results presented here for the spherical clusters. More generally, whenever we consider systems where the rate for some processes, such as condensations, is a strongly nonlinear function of the local density of particles, we may use the mentioned results to obtain the mass distribution and number density of distinct condensed objects, as we point out at the end of this work.

We shall briefly review the questions involved in our problem: the first step towards its solution is to define a cluster of n points as any collection of n points contained within the prescribed geometrical body. The density of these clusters in a Poissonian distribution was first derived by Politzer and Presskill [1], and Otto *et al.* [2], following a slightly different approach, generalized Presskill *et al.*'s result to any distribution. But, given one of these clusters of n points we may find (by slightly moving the body defining it, so as to discard a few points while sweeping in as many others) several clusters of n points sharing most of their points with the first. It is then clear that the objects perceived as distinct individual clusters are not the clusters thus defined, but rather groups of clusters.

The discussion of these issues and the procedure for deriving the probability density of the relevant objects in one-dimensional Poissonian distributions was first presented by the author of this paper [3]. In another work [4], the treatment corresponding to d-dimensional Poissonian distributions was developed. The notation and definitions introduced in this latter work are the ones adopted here. The clusters of n points defined so that any collection of n points is taken as distinct are termed type-1 clusters of n points, and are represented by C_n^1. The clusters of type-1 clusters formed by all the C_n^1's that share points are termed type-2 clusters of n points and are represented by C_n^2. These latter clusters are in a one to one correspondence with the really distinct objects termed "structures." These are the underlying concentrations of points out of which the C_n^1's and C_n^2's may be formed. From a formal point of view, note that the C_n^1's and the structures (of n points) are point sets, while the C_n^2's are sets of point sets. The structures of n points and specified geometrical shape (determined by the body defining the C_n^1's) may be defined in a precise and unambiguous manner as the set of all points belonging to any of the C_k^1's, where $k \geq n$, that share points. However, the concept of structure in point distributions is an essentially fuzzy one. The question of whether a particular point belongs to a given structure has an unambiguous answer for some points, but becomes increasingly ambiguous as we consider points that lie at an increasing distance from the center of the structure. Under these circumstances any precise definition of the term structure will be arbitrary as well as unambiguous. In some contexts it may be expedient to use a particular, precise definition of structure, but as long as our sole interest is centered in the mean density and correlations of the spots at which structures occur, the exact definition of structure is totally irrelevant. All we need to know for this purpose is that a structure of n points, whichever its exact definition, contains at least a C_n^1, and a single C_n^2. So, from a practical point of view, we may identify structures with C_n^2's. However, it must be remembered that the structures are the really relevant physical objects, while the C_n^2's are merely instrumental concepts.

It is interesting to note that the structures of more than n points are all included within the structures of n points. It would then seem that characterizing the structures by the maximum value of k over all the C_k^1's included in them might be more expedient. However, for a number of reasons, it turns out that the former definition is more convenient than the latter one. The properties of the structures derived using this latter definition may be obtained immediately from those pertaining to the structures as defined at present (which are cumulative with respect to the former). The reasons for maintaining the present definitions of structure become particularly clear when the concept is used in the context of continuous random fields [5]. This reference includes a detailed discussion of the concept of structure.

Reference [4] showed how to obtain the probability density (this term is more appropriate than "mean density"), D_n^2, of C_n^2's, which is equal to the probability density of structures of n points, from the probability density, D_n^1, of C_n^1's. All that is required is to divide the latter by the mean number $\bar{N}_d$, of C_n^1's in a C_n^2 (or in a structure of n points). We also showed how $\bar{N}_d$, which corresponds to a d-dimensional problem, may be expressed in terms of $\bar{N}$, which corresponds to a one-dimensional problem. The expression given there for $\bar{N}$, which was first derived in Ref. [3], is an approximate one. The exact expression is given in a latter work (Ref. [6]), which approaches the problem from a different direction than that taken in [3], which allowed the rigorous treatment not only of the relational properties (i.e., densities and correlations) of structures in one-dimensional Poissonian distributions, but that of their inner configuration as well. This approach has the further advantage that it may easily be extended to any point distribution. To show how this may be done is the main goal of this paper, see Secs. II and III. Section IV describes the general procedure for obtaining the probability for a randomly placed body to contain n points, P_n, in terms of the properties of the physical processes generating the distribution. In this last section two examples are considered in detail: the case of spherical clusters, and a case concerning the simultaneous occurrence of events, which illustrates how the present work may be applied outside its initial context. The result from spherical clusters is in itself quite useful. So, we present it in expressions (36) and (37) in a self-contained manner so as to make it available to readers not interested in the general content of this work.

II. GENERAL DERIVATION OF D_n^2: ONE-DIMENSIONAL CASE

The general expression for the probability density, D_n^1, of C_n^1's has been derived by the authors of Ref. [2]. So, to obtain the probability density of C_n^2's we only need to show how to compute the mean number, $\bar{N}_d'$, of C_n^1's in a C_n^2 (when an explicit distinction between symbols corresponding to general and Poissonian distributions is expedient, the primed symbols will be associated with the former case).

We shall first develop a treatment for the one-dimensional distributions; we shall then see, in the following section, how the treatment of the d-dimensional distribution may be developed by means of the former. In the one-dimensional case, the only problem that really needs solving is the computation of $\bar{N}'$. In Ref. [6] the following expression for $\bar{N}$ (Poissonian case) is derived:

$$D-U=(1/\bar{N})[1-2(1-L)]\ , \tag{1}$$

where D and U stand, respectively, for the probability that the next C_k^1 to the right of a randomly chosen C_n^1 will be a C_{n-1}^1, in which case we say that it is a down-going C_n^1, and the probability for it to be a C_{n+1}^1, in which case we say that it is an up-going C_n^1. $1-L$ stands for the probability that the next C_k^1 to the right of the last C_n^1 in a C_n^2 will result in a C_{n+1}^1.

This last probability is usually negligible; it only bears some relevance when dealing with rather common structures. To obtain expression (1) we used the fact that, in a Poissonian distribution, over the ensemble of all the C_k^1's lying between the first and the last C_n^1's in a C_n^2 the values of k are symmetrically distributed around n. It is obvious, however, that this result does not hold in general. If we consider a distribution resulting from a nonuniform random Poissonian process, it is clear that, in the high number density limit, the evolution of the values of k as we move from the first to the last C_n^1 in a C_n^2 will approach the profile of the peaks of the underlying probability density field smoothed in a scale Δl (length of the interval defining the clusters) above a certain threshold (related to n). So, the asymmetry of the distribution of the values of k with respect to n is evident. However, the mentioned symmetry is not, in fact, necessary for obtaining expression (1). What is really necessary to derive (1) is the fact that, over the ensemble of all C_n^1's but the last in a C_n^2, the probability that a C_n^1 be an up-going one is equal to the probability for it to be a down-going one, hence equal, to a half. For those C_n^1's that come last in a C_n^2, the probabilities that they might be up-going or down-going are (by definition) $1-L$ or L, respectively. We could then write

$$U=(1/\bar{N})[\bar{N}-1+(1-L)];D=(1/\bar{N})(\bar{N}-1+L)\ . \tag{2}$$

This expression leads immediately to expression (1). So, for (1) to hold for any point distribution, the sole requirement is that over the ensemble of all but the last C_n^1's in a C_n^2, the probabilities that a C_n^1 be up-going or down-going be equal.

This is actually the case for a large variety of distributions termed s distributions (which includes most interesting cases). But, in a general s distribution, unlike the Poissonian case, these probabilities (U,D) are not the same over all the subensembles of the ensemble of all C_n^1's but the last in a C_n^2. In particular, these probabilities may depend on the order number of the C_n^1's within the

C_n^2's. This dependence is the more pronounced the rarer the structure. The probability of up-going C_n^1's is larger that $\frac{1}{2}$ for the first half of the C_n^1's in a C_n^2, and smaller than $\frac{1}{2}$ for the second half. But the isotropy of the distribution is enough to grant the cancellation of these differences with respect to $\frac{1}{2}$ of the probabilities that they be up-going C_n^1's. We may then write for $\bar{N}'$,

$$D-U=(1/\bar{N}')[1-2(1-L)] . \quad (3)$$

We then see that the usefulness of expression (1) is not limited to the particular context in which it was first derived. In fact, as advanced in Ref. [6], this expression happens to provide a simple solution for problems whose alternative treatments may be rather complex. It may even be used to deal with the problem of structures in differentiable random fields, as we intend to show in a future work. The huge simplification in the computation of $\bar{N}'$ implied by (3) is due to the fact that the probabilities D,U may generally be expressed directly in terms of the properties of the underlying physical processes. In this respect, the somewhat complex computation of $\bar{N}$ given in Ref. [3], which is not even exact, may be compared with that given in Ref. [6], in which its exact value is obtained immediately by means of (1).

We shall now describe the general procedure for computing U,D in a one-dimensional s distribution. In these cases the body defining the clusters is simply an interval of length Δl, and the relevant probability distribution is that of having n points within a randomly distributed interval of length Δl, $P_n(\Delta l)$.

It is easy to realize that the ratio D/U must be given by

$$r(n)\equiv\frac{D(n)}{U(n)}=\frac{P_{n-1}(\Delta l)\rho(n-1)}{P_n(\Delta l)\rho(n)} , \quad (4)$$

where $\rho(n)$ is the probability density at the extremes (equal at both extremes since isotropy is assumed) of a randomly placed interval of length Δl containing n points (the explicit dependence of Δl is dropped in most symbols). To obtain $\rho(n)$ we present a derivation which, in principle, is valid strictly for those distributions of points that are generated by random Poissonian processes. In these cases, the probability distribution $P(M/n)$ for the integral, M, of the probability density over the interval Δl, when it constrains n points, is given by

$$P(M/n)=\frac{P(M)P(n/M)}{\int_0^\infty P(M)P(n/M)dM} ;$$
$$P(n/M)=\frac{M^n}{n!}e^{-M} , \quad (5)$$

where Bayes's rule has been used, and $P(M)$ is the probability distribution for the value of M within a randomly placed interval. $P(n/M)$ is the probability that the interval contains n points when the integrated probability density within it takes the value M, which is clearly given by the above expression. We have, on the other hand, that

$$P_n=\int_0^\infty P(M)P(n/M)dM . \quad (6)$$

So, the mean value, $\bar{M}$, of M within an interval containing n points may be expressed in the form

$$\begin{aligned}\bar{M}(n)&=\int_0^\infty MP(M/n)dM\\&=P_n^{-1}\int_0^\infty P(M)MP(n/M)dM\\&=P_n^{-1}\int_0^\infty P(M)(n+1)P(n+1/M)dM\\&=(n+1)P_{n+1}P_n^{-1} .\end{aligned} \quad (7)$$

We also have

$$\bar{M}(n,\Delta l+x)=[1-\rho(n)x][\bar{M}(n,\Delta l)+x\rho(n)]+\bar{M}(n,\Delta l+x)x\rho(n) , \quad (8)$$

where x is an arbitrarily small quantity. The first term on the right of this expression is the probability that a randomly placed interval of length $\Delta l+x$ containing n points should contain no point within a distance x from the right extreme, multiplied by the mean value of $\bar{M}(n,\Delta l+x)$ when this event takes place. The second term is the probability of the complementary event multiplied by the corresponding mean value of $\bar{M}(n,\Delta l+x)$. We have for $\rho(n)$

$$\rho(n)=\frac{dM}{d\Delta l}(n,\Delta l) . \quad (9)$$

Bearing in mind that $U+D=1$, we find that

$$\bar{N}'(n)=\frac{r(n)+1}{r(n)-1}[1-2(1-L)] . \quad (10)$$

In most interesting cases $1-L$ may be set to zero. This quantity equals the probability that an interval displaced by an amount Δl with respect to an interval containing n points contains more than n points, plus one half of the probability that it contains n points. For most purposes this quantity may be approximated by

$$1-L\simeq\frac{P_{2n}(2\Delta l)}{P_n(\Delta l)}\frac{(2i)!4^i}{(i!)^2}-\frac{1}{2}P_n(\Delta l) . \quad (11)$$

It must be noted that the definition given here for $1-L$ corresponds to defining the C_n^2's as the sets of all C_n^1's whose corresponding intervals overlap (at some position or other); while, in other parts of this and other papers, we define it as the set of all the C_n^1 that share points. In principle, these definitions are not identical, but the difference is so irrelevant that it needn't be considered. One or the other are used depending on whether in the particular context under consideration it is more convenient to characterize the C_n^1's by their points or by the position (or rather, range of possible positions) of their corresponding intervals (or, in general, body).

From Ref. [1] it is clear that the mean number D_n^1 of C_n^1's per unit of length is given by

$$\begin{aligned}D_n^1&=\rho(n-1)P_{n-1}+\rho(n)P_n\\&=[1+r(n)]\rho(n)P_n=\frac{P_n}{S(n)} ,\end{aligned} \quad (12)$$

where $S(n)$ is the mean length within which the interval defining the cluster may be wiggled while it contains the same n points. We shall find that this quantity turns out to be a good tool for obtaining the solution to the d-dimensional problem.

Finally for D_n^2 we have

$$D_n^2=\frac{D_n^1}{\bar{N}'}=[\rho(n-1)P_{n-1}-\rho(n)P_n][1-2(1-L)]^{-1} . \tag{13}$$

Such as it stands, this expression is exact for any s distribution. Only when expression (9) is used to obtain $\rho(n)$ is it necessary to make a further assumption on the character of the distribution, namely, that it is the result of a Poissonian process. However, although we do not know how to express $\rho(n)$ in terms of $P_n(\Delta l)$ and its derivatives (the expression in terms of the correlations is valid for any distribution, but is rarely useful) without making that assumption, we may use expression (9) for any distribution, provided only that $\rho(n)$ [as given by (9)] is defined for any value of n, and that the second central moment is not less than $\langle n \rangle$. The reason for this is that those distributions for which this condition holds—which is true for most distributions arising from natural processes—conform to a Poissonion model.

The probability density, M_j^2, of those structures (j structures) such that the maximum value of k over all the C_k^1's that they contain is j is immediately given by

$$M_j^2=D_j^2-D_{j+1}^2 . \tag{14}$$

The mean number of C_j^1's in a j structure, $\bar{N}_c'$, may be obtained in the same manner as expression (35) in Ref. [6]; it is only necessary to change k^{-1} $(=j/\bar{n}\Delta l)$ by r [see expression (4)].

$$\bar{N}_c'(j)=1+r^{-1}(j) . \tag{15}$$

The probability distribution for N_c' is given by

$$P(N_c')=\frac{1}{\bar{N}_c'}\left[1-\frac{1}{\bar{N}_c'}\right]^{N_c'} . \tag{16}$$

III. D-DIMENSIONAL CASE

For structures in d dimensions, D_n^1 is obtained by dividing P_n by the mean d-dimensional volume, ΔV, in which the body defining the cluster may be moved while it still contains the same set of n points. This volume may be expressed in terms of the quantities S_i [see expression (12)], with i from 1 to d, corresponding to the ith one-dimensional problem obtained when only displacements (of the body defining the clusters) along the ith degree of freedom are considered (similar considerations may be found in Refs. [4] and [5]). Note that, in general, not all the degrees of freedom are translational. There are usually at most three translational ones; the others are related to the possible shapes and orientations of the body defining the clusters. The S_i's are given by

$$S_i=[\rho_i(n-1)\frac{P_{n-1}}{P_n}+\rho_i(n)]^{-1} ,$$

$$\rho_i(n)=(n+1)\frac{d}{dx}\left[\frac{P_{n+1}[V+\delta V_i(x)]}{P_n[V+\delta V_i(x)]}\right]\Bigg|_{x=0} , \tag{17}$$

where $P_n(V)$ stands for the probability for the body in question (with volume V) to contain n points when placed at random, while $P_n(V+\delta V_i)$ corresponds to the body generated by the body defining the clusters when displaced along the ith degree of freedom by an amount x. We then have for D_n^1

$$D_n^1=\alpha\left[\prod_{i=1}^{d}S_i\right]^{-1}\left[1+\sum_{j=1}^{d-1}\gamma_j\left[\frac{\prod_{i=1}^{d}S_i}{V}\right]^{j/d}\right]P_n(V) , \tag{18}$$

α, γ_j are certain geometrical coefficients. α is directly related (for regular distributions) to the coefficients represented by the same symbol in Ref. [4] [expression (3)]. In this latter work we showed how to compute this coefficient. A general procedure for obtaining the γ_i's is not available: only in some cases is it possible to conduct simple computation (see Ref. [1]). But, these coefficients are rarely relevant. For N_d' we have

$$\begin{aligned}\bar{N}_d'&=\left[\prod_{i=1}^{d}[1+B_i(2I_i')]\right]\left[\frac{1}{d}\sum_{i=1}^{d}\frac{\bar{N}_i'}{1+2I_i'}\right]\left[1+\sum_{j=1}^{d-1}\Omega_j\left[\frac{2I_1'}{n}\right]^j\right]\\ &\simeq B\left[\frac{1}{d}\sum_{i=1}^{d}\frac{\bar{N}_i'}{1+2I_i'}\right]\left[\prod_{i=1}^{d}(1+2I_i')\right]\left[1+\frac{B^{-1}-1}{[\prod(1+2I_i')]^{1/d}}\right]\left[1+\sum_{j=1}^{d-1}\Omega_j\left[\frac{2I_1'}{n}\right]^j\right]\\ 1+2I_i'&=\bar{N}_i'\left[\frac{S_i(n)}{P_n(V)}\left[\sum_{k=n}^{\infty}\frac{P_k(V)}{S_i(k)}\right]-\frac{1}{2}+\frac{\bar{N}_i'}{2}\right] .\end{aligned} \tag{19}$$

$\bar{N}_i'$ is the $\bar{N}'$ associated with the one-dimensional problem corresponding to the ith degree of freedom, which is given by (4) and (10) with ρ_i [see (17)] in the place of ρ. This expression is the generalization of expression (15) in Ref. [4]. The expression for $1+2I_i'$ was obtained using the identity

$$\frac{P_n(\Delta l)}{\bar{N}_i'S_i(n)}=D_n^2=\left[\sum_{k=n}^{\infty}D_k^1\right]\left[1+2I_i'-\frac{\bar{N}_i'(\bar{N}_i'-1)}{2}\right]^{-1} , \tag{20}$$

which expresses the fact that D_n^2 may be obtained either by following the procedure implied by expression (13), or by dividing the probability density of C_k^1's, with $n \le k$, by the mean number of these C_k^1's, in a structure. This latter quantity is given by $[1+2I'-\bar{N}'(\bar{N}'-1)/2]$, where $1+2I'(n)$ has been taken—to keep the notation used in the Poissonian case—to represent the mean number of C_k^1's (for any value of k) that may be found between the first and the last C_n^1 (both included) in a C_n^2 in a one-dimensional problem. To obtain this expression [in the right-hand parentheses in (20)] we need only note that the quantity it represents is equal to $1+2I'(n)$, minus the mean number of all the C_k^1's, with $k<n$, that may be found between the first and the last C_n^1 in a C_n^2. This last quantity, in turn, depends on $\bar{N}'$ in the same manner as $(1+2I-\bar{N})/2$ depends on $\bar{N}$ in the one-dimensional Poissonian case. The approximation in which $I=(\bar{N}^2-1)/2$ has been used; when required, more accurate expressions may be used (see Ref. [6]). To compute the coefficients in (19) it is assumed that the correlations of the points within the cluster have scales of variation which are not much smaller than the mean distance between them. Typically, these scales are of order of the size of the cluster, and this assumption is fully granted. In a somewhat artificial distribution in which the points within the cluster are ordered in a fixed array, the fact of having a point at a certain position may actually determine the exact position of the points on the other side of the cluster. In this case, the assumption in question does not hold. We shall find, however, that even in cases as extreme as this, this assumption yields reasonably good results. But our attention will be centered on the distributions for which it holds. In these cases the Ω_i's and the B_i's are equal to those corresponding to a Poissonian distribution. The Ω_i's may be obtained by taking the high n limit of $D_n^2(n)$ for a Poissonian distribution, and by comparing it with the expression for the probability density of structures in a Gaussian random field with a white noise spectrum (see Ref. [5]). There is no simple procedure to compute the B_i's; while the procedure for computing B was shown in Ref. [4] [the B's given there must be multiplied by $(13/16)^{d-1}$, since this factor was missing in that work]. B is the most relevant coefficient and, in most cases, the only one that is really needed. The general procedure described in this section looks rather cumbersome, but we shall see that when applied to individual cases, it looks much simpler.

IV. COMPUTATION OF $P_0(V)$

The preceding sections showed that all the quantities relevant to our problem may be obtained from $P_0(V)$. In principle, to compute this quantity is a straightforward question. However, to connect the standard procedure with the processes generating the distribution is rather awkward. In Refs. [2] and [7], $P_0(V)$ is expressed in terms of integrals over correlation functions of all orders. So, given the processes in question we shall first need to obtain all the correlation functions in order to obtain $P_0(V)$. But to express the correlations in terms of the properties of the underlying processes is by no means a simple matter. In fact, as we shall see, it is $P_0(V)$ which admits a relatively direct expression in terms of the processes mentioned; while the expression for $P_0(V)$ in terms of the correlations may be used to derive certain properties of the latter. We shall consider distributions generated by two different types of processes. When those processes are such that the probability density for a point to end up at certain position is independent of the positions of other points—depending only on the value at the point of a certain underlying field—we say that the distribution results from a Poissonian process. There are, on the other hand, processes whereby the probability for a point to end up at certain position depends on the positions of other points. These processes are termed "non-Poissonian processes."

In the Poissonian processes all the properties of the resulting distribution are derived from those of the underlying probability density field, $\rho(x)$. In these cases, the relevant physical problem is that of obtaining the field, the procedure depending on the detailed nature of the field in question (an example may be found in Ref. [8]). What is presented here is how to derive $P_0(V)$ from a given field.

$$
\begin{aligned}
&P_0(V)=\int_0^{\infty} e^{-w\bar{n}}P(w)dw \ , \\
&w=(1/\bar{n})\int_V \rho(\mathbf{x})d\mathbf{x} \ , \\
&P_n(V)=\int_0^{\infty}\frac{(\bar{n}w)^n}{n!}e^{-\bar{n}w}P(w)dw \ .
\end{aligned} \tag{21}
$$

The integral in the definition of w is over the body defining the cluster. This definition has been chosen so that w does not depend on the mean number density, $\bar{n}$: it depends only on the dimensionless correlations, which are determined by the underlying field. The expression for $P_n(V)$ in terms of the derivatives of $P_0(V)$ with respect to $\bar{n}$, which is obvious in the present context, may be shown to be valid for any distribution (see Ref. [7]), provided that the derivatives are carried out with the dimensionless correlations held fixed. Using this result the following useful identity may easily be shown:

$$
G(t)=P_0[V,\bar{n}(1-e^t)];\ \langle n^k\rangle=\left.\frac{d^kG(t)}{dt^k}\right|_{t=0}, \tag{22}
$$

where $G(t)$ is the generating function of the distribution $P_n(V)$, $P_0(V,\bar{n})$ is $P_0(V)$ as a function of $\bar{n}$ at fixed correlations, and $\langle n^k\rangle$ is the kth moment of the distribution.

The essential problem that needs solving to obtain $P_0(V)$ is computing the probability distribution for the values of w at a randomly chosen point, $P(w)$. This distribution corresponds to the values of the smoothed field (using a weighting function whose value is one, inside the body in question, and zero on the outside), which may be obtained immediately in some cases—e.g., when the field is a Gaussian random one—may be quite involved in others.

We shall now consider non-Poissonian processes. To obtain the properties of the resulting distributions it is expedient to substitute the real processes, whereby the position of all the points will, in general, evolve simultaneously towards their final configuration, by a process whereby the points are placed sequentially; and where the probability density for a point to be placed at a given position depends only on the position of the previously placed ones. This substitution is merely a useful mathematical trick. The contribution to the probability density at $\mathbf{x}_j$ owed to the presence of a point at $\mathbf{x}_i$ is denoted $p(r)(r=|\mathbf{x}_i-\mathbf{x}_j|)$. We shall assume that

$$p=\frac{4\pi}{\bar{n}}\int_0^\infty p(r)r^2dr\ ,$$
$$(1/p)[p-(1/V)\int_V\int_V p(|\mathbf{x}_1-\mathbf{x}_2|)d\mathbf{x}_1d\mathbf{x}_2]\ll 1\ . \tag{23}$$

That is, the integral to all space of the fractional excess probability density (the mean is $\bar{n}$) due to the presence of a point converges over scales much smaller than that of the body in question. In this case, if the points are all placed within a region of volume Ω, we have for $P_0(V)$:

$$P_0(V)=\prod_{i=1}^{N}\left[1-\frac{V}{\Omega+pi}\right]\ . \tag{24}$$

The ith factor in this product is the probability that the $(i+1)$th point is placed outside V. Border effects have been neglected here. Accounting for these effects is relatively simple: for example, at first order, we need only add in (24) $i\Delta$ to V, where Δ is the mean value of the integral of $p(r)/\bar{n}$ within V for a randomly placed point outside V. However, we shall not include these refinements here. N is the total number of points in Ω. In the large Ω/V limit we have

$$P_0(V)=\exp\left[-V\sum_{i=0}^{N}\frac{1}{\Omega+ip}\right]\sim(1+p\bar{n})^{-V/p}\ ,$$
$$\bar{n}=N/\Omega\ . \tag{25}$$

When p falls to zero (25) takes its Poissonian value, $e^{-V\bar{n}}$. We may now express $P_n(V)$ in the form

$$P_n(V)=\frac{(-\bar{n})^n}{n!}\frac{d^n}{d\bar{n}^n}P_0(V,\bar{n})$$
$$=\frac{(\bar{n})^n}{n!}\left[\prod_{i=0}^{n-1}(V+pi)\right](1+p\bar{n})^{-(V/p+n)}\ , \tag{26}$$

or, in a more compact form,

$$P_n(V)=\frac{\Gamma(V/p+n)}{\Gamma(V/p+1)}\frac{(\bar{n}p)^n}{n!}(1+p\bar{n})^{-(V/p+n)}\ . \tag{27}$$

Using expression (22) we find that

$$\langle(n-\langle n\rangle)^2\rangle=\bar{n}V(1+\bar{n}p)\ . \tag{28}$$

Comparing this result with the expression for the second order central moment in terms of the two point correlation function ε, we find

$$p=(1/V)\int_V\int_V\varepsilon(|\mathbf{x}_1-\mathbf{x}_2|)d\mathbf{x}_1d\mathbf{x}_2\ . \tag{29}$$

It must be noted that the difference between Poissonian and non-Poissonian processes concerns the processes generating the distributions and not the distributions themselves. A distribution generated by a non-Poissonian process may conform to a Poissonian model. This is generally the case when the excess probability density at the places where there are points is typically the result of the contribution of several points, rather than being dominated by the next neighbor.

We shall now consider a clustering model which is particularly relevant in problems related to the simultaneous occurrence of events (see Ref. [6]). In this model, the distribution is formed by uncorrelated clusters with negligible size, whose number of points, N, follows a given probability distribution $P(N)$. In this case, the probability of having n points within a randomly placed body of volume V, $P_n(V)$, may be obtained by means of the generating function of $P_n(V)$, which may easily be expressed in terms of the generating function of $P(N)$,

$$G(P_n(V))=\exp\{-[MV-MVG(P(N))]\}\ ,$$
$$P_n(V)=\frac{1}{n!}\frac{d^n}{d(e^t)^n}G(P_n(V))\Big|_{t=-\infty}\ , \tag{30}$$

where $G(f)$ stands for the generating function of f and M is the number density of clusters.

In deriving expression(4), we assume that the correlations between the points are not singular. So, the next C_k^1 to a C_n^1 must either be a C_{n-1}^1 or a C_{n+1}^1. When the points are bound in arbitrarily tight clusters, as in the model which has just been described, expression (4) must be changed by

$$r(n)=\sum_{i=0}^{n}\frac{P_i(V)\rho(i)P(n-i)}{P_n(V)\rho(n)}\ , \tag{31}$$

where $\rho(i)$ is now the probability density for clusters at the extremes of the interval. In the models just considered $\rho(i)$ may be dropped, since it does not depend on i.

As an example of the possible applications of the main results of this work to the above model, the following case will be considered: assume that users come to a facility at randomly chosen times. The users come in groups whose size, N, follows the distribution (which is a relevant distribution in the theory of avalanches):

$$P(N)=\frac{1}{\ln(1-A)}\frac{1}{N}A^N\ . \tag{32}$$

The mean arrival rate of the groups is M, and their mean facility use time is T. The facility may serve $n-1$ users simultaneously. The question is what is the mean time, $\langle t\rangle$, between saturations (situations during which there is at least one user at the queue). The answer to this question is clearly

$$\langle t\rangle^{-1}=D_n^2=M[r(n)-1]P_n(T)[1-2(1-L)]^{-1}\ , \tag{33}$$

where $r(n)$ is given by expression (31). To obtain the probability, $P_n(T)$, of finding n users within a randomly chosen interval of length T we use (30) and find

$$G(P_n(T)) = \exp\left\{-MT\left[1-\frac{\ln(1-Ae^t)}{\ln(1-A)}\right]\right\},$$

$$P_n(T) = \frac{(-1)^n}{n!} \times \frac{d^n}{du^n}\exp\left\{-MT\left[1-\frac{\ln(1-A+Au)}{\ln(1-A)}\right]\right\}\Bigg|_{u=0}$$

$$= \frac{\Gamma(T/p+n)}{\Gamma(T/p+1)}\frac{1}{n!}\left[\frac{A}{1-A}\right]^n (1-A)^{T/p+n},$$

$$p = -\frac{\ln(1-A)}{M}. \quad (34)$$

$(1-L)$ is given by

$$1-L = \sum_{i=n+1}^{\infty} P_i(T) + 1/2P_n(T). \quad (35)$$

The above example illustrates the one-dimensional problem. As an example of the d-dimensional one, we shall now consider the case of spherical clusters in three dimensions, which is the issue that originally motivated the present work. Given a three-dimensional distribution of points, and the probability distribution, $P_n(r)$, for the number of points within a randomly placed sphere of radius r, we have to obtain the mean density of spherical clusters of radius r. These clusters are defined as those nonoverlapping spots within the distribution where a sphere of radius r may be placed so as to contain at least n points. It is then clear that the mean density of these clusters is given by the mean density, $D_n^2(r)$, of spherical type 2 clusters of radius r.

To obtain $D_n^2(r)$ we must use the general procedure described in the preceding section. However, due to the isotropy of the present problem, we reduce it to three equal one-dimensional problems. That is, S_i is independent of i and is given by

$$S(n) = \left[\rho(n-1)\frac{P_{n-1}}{P_n} + \rho(n)\right]^{-1},$$

$$\rho(n) = \pi r^2 \frac{(n+1)}{4\pi r^2}\frac{d}{dr}\frac{P_{n+1}(r)}{P_n(r)}. \quad (36)$$

Using the coefficients given in Ref. [1], we find

$$D_n^1(r) = \frac{3\pi^2}{32(\frac{4}{3}\pi r^3)}\left[L^{-3}(n) + 3\left[1-\frac{3\pi^2}{32}\right]L^{-2}(n) - 2\left[1-\frac{3\pi^2}{32}\right]L^{-1}(n)\right]P_n(r),$$

$$D_n^2(r) = \frac{D_n^1(r)}{\bar{N}_3'}, \quad L(n) \equiv \frac{3}{4}\frac{S(n)}{r},$$

$$\bar{N}_3' = \frac{5\pi^2}{48}\left[\frac{16}{13}\right]^2 (1+2I')^2 \bar{N}'\left[1-\frac{0.47}{1+2I'}\right] \times \left[1+\Omega_1\left[\frac{2I}{n}\right]+\Omega_2\left[\frac{2I}{n}\right]^2\right], \quad (37)$$

$$\bar{N}' = \frac{r(n)+1}{r(n)-1}[1-2(1-L)]; \quad r(n) = \frac{P_{n-1}(r)\rho(n-1)}{P_n(r)\rho(n)},$$

$$1+2I' = \bar{N}'\left[\frac{S(n)}{P_n(r)}\left[\sum_{i=n}^{\infty}\frac{P_i(r)}{S(i)}\right] + \frac{\bar{N}'-1}{2}\right].$$

S, ρ are given by (36), and N', I' are given by (10) [with $r(n)$ given by (4)] and (19). The coefficients have been taken from Ref. [4] [$B = 10\pi^2/96\,(13/16)^2$]. A procedure for computing the Ω's is also given there. When the mean distance between clusters is much larger than r, $(1-L)$ is negligible. For a Poissonian distribution with mean density of points $\bar{n}$, (37) reduces to

$$D_n^2(r) = \frac{(\bar{n}V+n)^3 + 3(1-3\pi^2/32)(\bar{n}V+n)^2 - 2(1-3\pi^2/32)(\bar{n}V+n)}{(1+2I)^2\bar{N}\left[1-\dfrac{0.47}{1+2I}\right]}\frac{9}{10}\left[\frac{13}{16}\right]^2\frac{P_n(r)}{V},$$

$$N = \left[\frac{1+K}{1-K}\right][1-w],$$

$$I = \frac{2K}{(1-K)^2}[1-w]; w = [1-1/(8n) - O(n^{-2})]\,\mathrm{erfc}\left[\left[\frac{1-K}{1+K}\right]\sqrt{n-1}\right],$$

$$K = \bar{n}V/n; V = \tfrac{4}{3}\pi r^3. \quad (38)$$

Expression (37) provides a powerful tool for statistical mechanics. When nonlinear processes (i.e., condensations) take place at some high density spots within a system, we may use (37) to obtain the mean density, M, of those spots at any given time and the probability distribution of their masses. To this end we must first obtain the probability, $F(n,r)$, that a condensation of n or more particles occurs when the n particles may be found within a sphere of radius r (determined by the threshold local density for the onset of the processes). We then have

$$M = \sum F(n,r)[D_{n-1}^2(r) - D_n^2(r)]. \quad (39)$$

$F(n,r)$ may readily be obtained from the underlying statistical mechanics, but if we want to obtain M we cannot bypass the use of (37).

[1] H. Politzer and J. Presskill, Phys. Rev. Lett. **56**, 99 (1986).
[2] S. Otto, H. Politzer, J. Presskill, and M. Wise, Astrophys. J. **62**, 304 (1986).
[3] J. Betancort-Rijo, Mon. Not. R. Astron. Soc. **237**, 431 (1989).
[4] J. Betancort-Rijo, Phys. Rev. A **43**, 2694 (1991).
[5] J. Betancort-Rijo, Phys. Rev. A **45**, 3647 (1992).
[6] J. Betancort-Rijo, Phys. Rev. A **46**, 4549 (1992).
[7] S. D. M. White, Mon. Not. R. Astron. Soc. **186**, 145 (1979).
[8] J.Betancort-Rijo, Mon. Not. R. Astron. Soc. **246**, 608 (1990).