PHYSICAL REVIEW LETTERS

VOLUME 56 13 JANUARY 1986 NUMBER 2

Poisson Clusters and Poisson Voids

H. David Politzer and John P. Preskill

California Institute of Technology, Pasadena, California 91125

(Received 15 August 1985)

Expressions are derived for the expected abundance of clusters and voids in a sample of randomly distributed objects.

PACS numbers: 02.50.+s, 05.40.+j

The interpretation of an observed distribution of pointlike objects often involves assessing the significance of clusters of voids. For example, one may observe with a neutrino telescope[1] a cluster of several events within a small solid angle, or one may observe a large void in the distribution of rich clusters of galaxies.[2] One then wishes to determine the likelihood of the observed cluster or void arising as a statistical fluctuation, under the assumption that the objects are actually randomly distributed. Unless this probability is low, it is not reasonable to regard the signal detected by the telescope as evidence for an astrophysical point source of neutrinos, nor to regard the rich-cluster void as dramatic evidence for very large scale structure in the universe.

One way to test the hypothesis that a set of points is randomly distributed is to check the validity of the Poisson formula,

$$P_k(nV) = (1/k!)(nV)^k e^{-nV}, \qquad (1)$$

for the probability that a randomly selected region of volume V contains k points, if n is the mean density of the points. However, for clusters so dense or voids so dilute that the corresponding Poisson probability is extremely small, it is far more useful to know the expected number of clusters or voids located *anywhere* in the entire observed region. In principle, the expected abundance of such dense clusters or dilute voids could be determined by Monte Carlo simulations. However, rare occurrences are not so easily simulated, and at any rate it is evidently desirable to be able to express the answer in an analytic form. Thus, in this Letter, we derive such expressions for the expected abundance of clusters and voids in a sample of randomly distributed objects, in the limit in which the clusters or voids are rare. These formulas can be generalized to the case of correlated distributions. To illustrate their use, we apply our results to the distribution of rich clusters of galaxies in the sky.

The Poisson probability P_k is independent of the shape of the region in question. In contrast, to define what constitutes a cluster or void requires some specification of an acceptable shape, and the expected abundance of such configurations will, in general, depend on that specification. For example, we may decide that k objects distributed in d dimensions constitute a cluster if there is a cube of volume V and *a priori specified orientation* which contains these k objects and no others. Then, as we will show, the expected number of such clusters per unit volume is

$$D_k(n,V) = V^{-1} P_k(nV) k^d [1 + O(nV/k)], \qquad (2)$$

where n is the mean density of the objects, and k is large compared to nV. For k comparable to nV, the mean number of objects in the volume V, there are typically many overlapping clusters, and the abundance of clusters is not of great interest; there is considerable arbitrariness in counting the clusters when they overlap. The $O(nV/k)$ correction in Eq. (2) depends on how the overlapping clusters are counted.

A void in a d-dimensional distribution of objects may be defined as a cluster of k objects where $k \ll nV$. (For $k = 0$, the void is the region in which our cube of *a priori* specified orientation can be continuously transported without encountering an object.) We will show that the expected number of such voids per unit volume is

$$D_k(n,V) = V^{-1}P_k(nV)(nV)^d[1+O(k/nV)+O(1/nV)^{d-1}]. \tag{3}$$

Equations (2) and (3) apply to clusters defined not only by cubes, but also by rectangular solids of specified shape and orientation, and, in two dimensions, to circles and ellipses. The corresponding formula for three-dimensional spheres is also given below. The generalizations to shapes with arbitrary orientation and correlated distributions of objects are briefly discussed.

The reader should note that our Eqs. (2) and (3) cannot be reproduced by multiplication of the Poisson probability by the density of nonoverlapping volumes of size V. This naive procedure fails because the actual clusters or voids will not typically coincide with any member of a set of *a priori* selected nonoverlapping volumes. The naive procedure therefore underestimates the abundance of clusters or voids by the factor k^d or $(nV)^d$ respectively, which in practice can amount to orders of magnitude.

It is simplest to begin by considering clusters among randomly distributed objects in two dimensions, where k objects are said to constitute a cluster if there exists a square of side L, with sides oriented parallel to the x and y axes, which contains those k objects and no others. To search systematically for such clusters, we may introduce a square lattice with spacing ϵ, and center at each site a trial square of appropriate size and orientation. Each trial square containing exactly k objects locates a cluster, and, in the limit $\epsilon \to 0$, every cluster is found by some square. However, each given cluster is found by many different trial squares. To avoid overcounting the clusters, we order the trial squares by sweeping through the lattice; we sweep through each row from left to right, and order the rows from top to bottom. Now there is a unique trial square in this ordered sequence which has the *first encounter* with each cluster of k objects. To compute the density of clusters of k objects, we may equivalently compute the density of these first encounters.

The fraction of trial squares containing k objects which are also first encounters is given by the sum of two generic possibilities: The trial square immediately preceding the first encounter of a particular set of k objects contains either $k+1$ or $k-1$ objects. (Since we will ultimately take $\epsilon \to 0$, it suffices to consider objects leaving or entering the trial square one at a time.) The former case may be neglected in the limit $k >> nV$. Therefore, in this limit, the first encounter square always contains one object just inside its leading edge, in a strip of width ϵ and length L. Furthermore, since no trial square in the previous row contained these k objects, the first encounter square also contains one object just inside its bottom edge,[3] in a strip of width ϵ and length L. Thus, we may distinguish between two types of first encounter squares. Either there is one object along the leading edge, one object along the bottom edge, and $k-2$ objects in the interior of the square, or there is one object in the intersection of the two strips, an $\epsilon\times\epsilon$ region in the bottom right corner of the square, and $k-1$ objects in the interior of the square. Adding together the two types, we find that the fraction F of all squares containing k objects which are also first encounters is[4]

$$F = (n\epsilon L)^2\frac{P_{k-2}(nL^2)}{P_k(nL^2)} + (n\epsilon^2)\frac{P_{k-1}(nL^2)}{P_k(nL^2)} = \frac{\epsilon^2}{L^2}k^2, \tag{4}$$

up to corrections of order (nV/k). Multiplying by $P_k(nL^2)/\epsilon^2$, the expected number per unit volume of squares containing k objects, we obtain Eq. (2), with $d=2$ and $V=L^2$.

To compute the abundance of voids in the limit $k << nV$, we need consider only the case in which the trial square immediately preceding the first encounter square contains $k+1$ objects. Therefore, there is one object just outside the trailing edge of, and one just above, the first encounter square, each in a strip of width ϵ and length L. The fraction of trial squares containing k objects which are also first encounters is thus

$$F = (n\epsilon L)^2 = (\epsilon^2/L^2)(nL^2), \tag{5}$$

in the limit $k << nV$. Multiplying by $P_k(nL^2)/\epsilon^2$, we obtain Eq. (3).

This derivation may be repeated for d-dimensional cubes with specified orientation, yielding Eqs. (2) and (3). For example, in three dimensions, we may distinguish three types of first encounters which are relevant in the limit $k >> nV$. The first encounter cube may contain one object along each of three faces, and $k-3$ objects in the interior; it may contain one object along a face, one along an edge, and $k-2$ in the interior; or it may contain one object in the corner and $k-1$ in the interior. Adding together the abundances of the three types of first encounter, we obtain Eq. (2), with $d=3$. The derivation may also be trivially generalized to clusters and voids defined by d-dimensional rectangular solids of *a priori* specified shape and orientation.

It is evident from the above derivation that the abundances of clusters and voids generally depend on the shape of our trial volume. The definitions of clusters and voids used above have the disadvantage that an orientation for our trial cubes must be arbitrarily specified. It is preferable to say that k objects constitute a cluster if there is a sphere of volume V which contains these k objects and no others, since the sphere has no orientation. We will therefore derive expressions for the expected abundance of clusters and voids defined by circles in two dimensions and spheres in three dimensions.

To search for clusters defined by circles, we introduce a square lattice with spacing ϵ, and center at each site a trial circle of radius R, $\pi R^2 = V$. As before, we compute the density of clusters of k objects by finding the density of first encounters. In the limit $k \gg nV$, we may distinguish two types of relevant first encounters. Either the first encounter circle contains two objects along its circumference, and $k-2$ objects in its interior, or it contains one object at its very bottom, and $k-1$ objects in its interior. The expected density of clusters must therefore take the form

$$D_k(n,V) = V^{-1}P_k(nV)[\alpha k(k-1) + \beta k]. \quad (6)$$

To determine α and β consider the limit $nV \to 0$. Obviously, in this limit $D_1(n,V) = n$; thus $\beta = 1$. Furthermore, two objects form a cluster if their separation is less than $2R$. For each object, the probability that there is a second object within a distance $2R$ is $n\pi \times (2R)^2 = 4nV$. Dividing by two to avoid overcounting the two-object clusters, we have $D_2(nV) = 2n^2V$, and $\alpha = 1$.

In searching for voids with $k \ll nV$, we need consider only one type of first encounter. The geometry which determines the abundance of these first encounters is identical to that which determines the abundance of the first type of first encounter in the limit $k \gg nV$. Therefore,

$$D_k(n,V) = V^{-1}P_k(nV)\alpha(nV)^2 \quad (7)$$

for $k \ll nV$, and we conclude that Eqs. (2) and (3), with $d=2$, apply to clusters and voids defined by circles. (They also apply to ellipses of fixed orientation and eccentricity.)

$$D_k(n,V) = V^{-1}P_k(nV)[\alpha k(k-1)(k-2) + \beta k(k-1) + \gamma k]. \quad (8)$$

Again, we can determine α, β, and γ by consideration of the limit $nV \to 0$. From $D_1(n,V) = n$, $D_2(n,V) = 4n^2V$, we obtain $\gamma = 1$, $\beta = 3$. To determine α, we must find $D_3(n,V)$.

To calculate $D_3(n,V)$ for $nV \ll 1$, we observe that, given two points with separation $z < 2R$, the allowed positions of a third point such that all three are contained inside some sphere of radius R fill a solid with volume

$$W(z) = 2\pi\{\tfrac{16}{3}R^3 - \tfrac{5}{2}zR^2 + \tfrac{1}{24}z^3 + R^2[R^2 - (z/2)^2]^{1/2}\sin^{-1}(z/2R)\}. \quad (9)$$

(The boundary of this solid is a figure known to draftsmen as a "four-centered approximate ellipse," rotated about its minor axis. The "approximate ellipse" consists of four circular arcs, two of radius R, and two of radius $2R$.) And given one object, the probability that there is a second object with separation from the first between z and $z+dz$ is $n^2 4\pi z^2\,dz$. Therefore, the density of three-object clusters is

$$D_3(n,V) = \tfrac{1}{6}n^3\int_0^{2R} 4\pi z^2\,dz\, W(z)$$
$$= V^{-1}P_3(nV)[\tfrac{9}{16}\pi^2 + 21], \quad (10)$$

and we find $\alpha = 3\pi^2/32$, $\beta = 3$, and $\gamma = 1$. Having determined α, we also know that the density of voids is

$$D_k(n,V) = V^{-1}P_k(nV)(3\pi^2/32)(nV)^3, \quad (11)$$

in the limit $k \ll nV$.

If our trial volume is not a sphere, and is permitted to have an arbitrary orientation, then the likelihood of finding a cluster of k objects evidently increases. Consider, for example, rectangles in two dimensions with sides of fixed length but arbitrary orientation. In an ordered search for trial rectangles containing k objects, the first encounter rectangle for a given cluster of k objects will typically have three objects distributed along its boundary, and $k-3$ in its interior. The general rule is that the first encounter trial volume typically has as many objects distributed along its boundary as the number of degrees of freedom inherent in the search; e.g., for ellipsoids of fixed volume with arbitrary orientation and eccentricities, this number is eight. If the trial volume has d degrees of freedom, then the formula analogous to Eqs. (6) and (8) for the expected density of clusters is

$$D_k(n,V) = V^{-1}P_k(nV)\alpha[k^d + O(k^{d-1})], \quad (12)$$

and the formula analogous to Eq. (7) for the density of voids is

$$D_k(n,V) = V^{-1}P_k(nV)\alpha(nV)^d. \quad (13)$$

For rectangles or ellipsoids of arbitrary orientation, the coefficient α and the coefficients of the nonleading powers of k in Eq. (12) are difficult to determine analytically, but should be amenable to a Monte Carlo evaluation.

For objects with nontrivial correlations, a generalization of the Poisson formula can be derived that expresses the probability that a randomly selected volume contains k objects in terms of integrals of correlation functions.[5] A corresponding generalization of our formulas for $D_k(n,V)$ can also be derived. The derivation is particularly simple for voids on scales much larger than all correlation lengths,[6] or clusters much smaller than all correlation lengths.

As an application of our formalism, we consider the (two dimensional) distribution of rich clusters of galaxies in the sky. A large region of the sky has been found[2] to contain no nearby Abell clusters, of "richness class" $R \geqslant 1$ and at a distance less than about $250h^{-1}$ Mpc, where $H_0 = 100h$ km s^{-1} Mpc^{-1} is the Hubble constant. In order to assess the significance of this void, we should determine whether it could plausibly have arisen as a statistical fluctuation. Although the positions of the rich clusters are known to be correlated,[7] we will ignore the correlations, and estimate the likelihood of a large void assuming that the rich clusters are randomly distributed.

In a sample of 71 rich clusters projected on the sky, the largest observed empty *circular* region is one in which the expected number of clusters is seven. The expected density of circular voids of this size or larger is given by Eq. (3), with $d = 2$ and $nV = 7$, and we find $V_{\mathrm{total}} D_0(n, V) = (71/7)(7^2)e^{-7} = 0.45$. After making a "fiducial-volume" correction—circles whose centers are too close to the boundary of the sample are not allowed—we conclude that the probability of finding a circular void with $nV \geqslant 7$ is about 20%. Thus, the discovery of such a void would be unremarkable even if the rich clusters were randomly distributed.

In the same sample of 71 rich clusters, a roughly rectangular empty region is observed in which ten clusters were expected. A search in two dimensions for a rectangular void of a given area but with arbitrary center, orientation, and eccentricity has four degrees of freedom; the expected density of voids is therefore given by Eq. (13), with $d = 4$. The coefficient in Eq. (13) has been determined by a Monte Carlo calculation reported in Ref. 6 to be $\alpha \sim 0.1$. The density of empty rectangular regions of volume V or greater is therefore

$$D_0(n, V) = (0.1)\, V^{-1} (nV)^4 e^{-nV}, \tag{14}$$

and the probability of finding a void with $nV \geqslant 10$ in a total sample of 71 objects is roughly 30%.[6] (Because the determination of α is rather crude, we have not attempted to include a fiducial-volume correction.) Again, the observation of such a void would not be surprising even if the rich clusters were randomly distributed.

Since the rich clusters are not really randomly distributed,[7,8] voids inconsistent with the hypothesis of randomly distributed clusters should be expected to appear in a sufficiently large sample. For such a sample, the abundance of voids might well be used to test more realistic hypothesis concerning the distribution of clusters. Techniques for carrying out such tests have been described in Ref. 6.

This work was supported in part by the U.S. Department of Energy under Contract No. DEAC-03-81-ER40050. One of us (J.P.) is the recipent of an Alfred P. Sloan Foundation fellowship and a National Science Foundation Presidential Young Investigator Award. Stimulating conversations with John LoSecco, Steve Otto, Ben White, and Mark Wise are also gratefully acknowledged.

[1]J. M. LoSecco *et al.*, "A Study of Atmospheric Neutrinos with the IMB Detector," in Proceedings of the Nineteenth International Cosmic Ray Conference, La Jolla, California, 1985 (to be published).

[2]N. Bahcall and R. Soniera, Astrophys. J. **262**, 419 (1982).

[3]The possibility that there is one object just above the first encounter square can be neglected for $k >> nV$.

[4]The procedure described sometimes counts a given cluster more than once, but makes a negligible error in the limit $k >> nV$.

[5]S. D. M. White, Mon. Not. Roy. Astron. Soc. **186**, 145 (1979).

[6]S. Otto, H. D. Politzer, J. Preskill, and M. B. Wise, California Institute of Technology Report No. CALT-68-1254, 1985 (to be published). This paper used an alternative method for computing the density of voids: With trial volumes centered at each site of a fine lattice, the number of voids was identified as the total number of empty trial volumes divided by the average number of empty trial volumes per single void.

[7]N. Bahcall and R. Soniera, Astrophys. J. **270**, 20 (1983).

[8]D. J. Batuski and J. O. Burns. Astron. J. **90**, 1413 (1985).