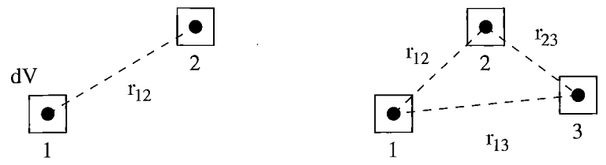

Figure 16.1. The N-point correlation functions of a density field consisting of a set of particles are calculated by looking at a set of cells of volume dV (so small that they effectively only ever contain 0 or 1 particles). The Poisson probability that two cells at separation r_{12} are both occupied is $\rho_0^2\, dV_1\, dV_2$; with clustering, this is modified by a factor $1+\xi(r_{12})$, where ξ is the two-point correlation function. Similarly, the probability of finding a triplet of occupied cells is a factor $1+\xi(r_{12},r_{13},r_{23})$ times the random probability; this defines the three-point correlation function.

measures the average galaxy density profile around clusters (at least out to radii where clusters overlap).

There is a straightforward extension of two-point statistics to larger numbers of points: the probability of finding an n-tuplet of galaxies in n specified volumes is

$$dP = \rho_0^n[1+\xi^{(n)}]\, dV_1 \cdots dV_n. \tag{16.31}$$

As with the two-point function, the probability is proportional to the product of the density field at the n points, and so

$$1+\xi^{(n)} = \left\langle \prod_i (1+\delta_i) \right\rangle. \tag{16.32}$$

Expanding the product gives a sequence of terms. For $n=3$, for example,

$$\xi^{(n)} = \xi(r_{12}) + \xi(r_{23}) + \xi(r_{31}) + \zeta(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3). \tag{16.33}$$

The term $\zeta = \langle \delta_1\delta_2\delta_3 \rangle$, which represents any excess correlation over that described by the two-point contributions we already know about, is called the **reduced three-point correlation function**. Similar expressions exist for larger numbers of points; the reduced n-point correlation function is $\langle \prod \delta_i \rangle$.

Observationally, the three-point function is non-zero. It has been suggested (Groth & Peebles 1977) that the results fit the so-called hierarchical form

$$\zeta = Q(\xi_{12}\xi_{23} + \xi_{23}\xi_{31} + \xi_{31}\xi_{12}), \tag{16.34}$$

with $Q \simeq 1$. This has been generalized to the **hierarchical ansatz**, which assumes that the reduced correlations can all be expressed in terms of two-point functions in this way (e.g. Balian & Schaeffer 1989). However, no convincing derivation of such a situation starting from linear initial conditions has ever been given.

The reduced functions are sometimes called the **connected correlation functions** by analogy with Green functions in particle physics. The analogy allows diagrammatic techniques borrowed from that subject to be used to aid calculation (Grinstein & Wise 1986; Matsubara 1995). To describe all the statistical properties of the density field requires the whole hierarchy of correlation functions (White 1979), which is a problem since only the first few have been measured. However, this is not a problem for Gaussian fields (see below), where all reduced correlations above the two-point level either vanish, if odd, or are expressible in terms of two-point functions, if even. It is because of this relation to the two-point function, and hence to the power spectrum, that the latter is often described as the 'holy grail' of cosmology: it tells us all that there is to know about the statistical properties of the density field.

16.3 Gaussian density fields

Apart from statistical isotropy of the fluctuation field, there is another reasonable assumption we might make: that the phases of the different Fourier modes δ_k are uncorrelated and random. This corresponds to treating the initial disturbances as some form of random noise, analogous to Johnson noise in electrical circuits; indeed, many mathematical tools that have become invaluable in cosmology were first established with applications to communication circuits in mind (e.g. Rice 1954). The random-phase approximation has a powerful consequence, which derives from the **central limit theorem**: loosely, the sum of a large number of independent random variables will tend to be normally distributed. This will be true not just for the field δ; all quantities that are derived from linear sums over waves (such as field derivatives) will have a joint normal distribution. The result is a **Gaussian random field**, whose properties are characterized entirely by its power spectrum.

The central limit theorem is a rather general result, whose full proof is involved (e.g. Kendall & Stuart 1958). However, the proof is simple for a restricted class of probability distributions. Suppose we make n measurements of a variable x, whose probability distribution is $f(x)$; the critical simplification will be to assume that all the moments $\int f(x)x^m\, dx$ are finite. Now introduce the **complementary function**, which is the Fourier transform of f: $C(k) = \int f(x)\, \exp(-ikx)\, dx$. This function obeys the key result that the complementary function for the distribution of the sum of two random variables x and y, with distribution functions f and g and complementary functions C_1 and C_2, is the product of the individual complementary functions:

$$C_+(k) = \int f(x)g(y)\, \exp[-ik(x+y)]\, dx\, dy = C_1(k)C_2(k). \tag{16.35}$$

For the sum of n drawings from the same distribution, $C_+(k) = C(k)^n$. Now consider a Taylor expansion of C, which involves the mean $\mu = \langle x \rangle$ and variance $\sigma^2 = \langle x^2 \rangle - \mu^2$:

$$C(k) = 1 - i\mu k - (\sigma^2+\mu^2)k^2/2 + O(k^3) = \exp\left[-i\mu k - \sigma^2k^2/2 + O(k^3)\right]. \tag{16.36}$$

Raising this to the nth power and defining $k' = k/n$, so that we deal with the distribution for the mean of the n observations, rather than the sum, gives

$$C_{\rm mean}(k') = \exp\left[-i\mu k' - (\sigma^2/n)k'^2/2 + O(k'^3/n^2)\right]. \tag{16.37}$$

The higher powers of $1/n$ in the terms beyond k'^2 mean that these terms can be neglected for large enough n, so long as the higher-order moments are finite. What remains is the complementary function of a Gaussian, as required. A fuller proof shows that the theorem in fact holds even when these moments diverge; all that is needed is a finite second moment. The simplest counterexample to the theorem is thus the **Cauchy distribution**, $f \propto 1/(1+x^2)$.

REALIZATIONS An immediate example of a Gaussian variable related to a Gaussian field arises when we consider one realization of such a field: a sample within a finite volume. The Fourier coefficients for the field within the box are linear sums over the field, and hence will themselves be (two-dimensional, complex) Gaussian variables. Call these a_k, where $\delta = \sum a_k \exp(-i\mathbf{k}\cdot\mathbf{x})$. If we call the realization volume V_R to distinguish it from the arbitrary normalization volume, the expectation value of the realization power follows from fixing the power in a given region of k-space in the continuum limit:

$$\langle |a_k|^2 \rangle = \frac{V}{V_R} P(k). \tag{16.38}$$

The power in a given mode will not take this expectation value, but will be exponentially distributed about it (a **Rayleigh distribution** for the mode amplitude; the real and imaginary parts of a_k will be Gaussian):

$$P\left(|a_k|^2 > X\right) = \exp\left(-X^2/\langle |a_k|^2 \rangle\right). \tag{16.39}$$

We now have the prescription needed to set up the conditions in k-space for creating a Gaussian realization with a desired power spectrum: assign Fourier amplitudes randomly with the above distribution and assign phases randomly between 0 and 2π. The real-space counterpart may then be found efficiently via the fast Fourier transform (FFT) algorithm. This is how the set of mode amplitudes and phases may be chosen for N-body initial conditions (see chapter 15). In fact, the central limit theorem means that, even if the dispersion in amplitudes is neglected, the field will still be closely Gaussian if many modes are used. However, its statistical properties may be suspect on scales approaching the box size, where there are few modes.

An alternative, simpler, method is to create a spatial array of white noise: Gaussian density variations in which each pixel is independent and has unit variance. This field can then be given the desired statistical properties by convolving it with a function f that is the 'square root' of the desired correlation function, i.e. $f * f = \xi(r)$. The squared Fourier transform of f is thus just the desired power spectrum; by the convolution theorem, this is also the power spectrum of the convolved white-noise field. From the computational point of view, this is not competitive with the FFT approach in terms of speed, although it may be simpler to code. However, it is often a useful concept to imagine random fields being generated in this two-step manner.

DENSITY MAXIMA To see the power of the Gaussian-field approach, consider the question of **density peaks**: regions of local maxima in $\delta(\mathbf{x})$. These are the points that will be the first to reach nonlinear density contrasts as the perturbation field develops. If we adopt the 'collapse at $\delta = \delta_c$' prescription from chapter 15, then the distribution of collapse redshifts can be deduced once we know the probability distribution for the value of $\delta(\mathbf{x})$ at the location of density maxima. This can be found as follows; we shall give only the 1D case in detail, because larger numbers of dimensions become messier to handle. See Peacock & Heavens (1985) and Bardeen *et al.* (1986; BBKS) for details of the 3D case. The field and its derivatives are jointly Gaussian

$$\boxed{p(\delta, \delta', \delta'', \ldots) = \frac{|\mathbf{M}|^{1/2}}{(2\pi)^{m/2}} \exp(-\tfrac{1}{2}\mathbf{V}^{\mathrm{T}}\mathbf{M}\mathbf{V}),} \tag{16.40}$$

for m variables in a vector $\mathbf{V}$; $\mathbf{M} \equiv \mathbf{C}^{-1}$ is the inverse of the **covariance matrix**. The elements of the covariance matrix depend only on moments over the power spectrum; in the particular case where our vector is $\mathbf{V}^{\mathrm{T}} = (\delta\ \delta''\ \delta')$, the covariance matrix is

$$\mathbf{C} = \begin{pmatrix} \sigma_0^2 & -\sigma_1^2 & \\ -\sigma_1^2 & \sigma_2^2 & \\ & & \sigma_1^2 \end{pmatrix}, \tag{16.41}$$

where

$$\sigma_m^2 \equiv \left(\frac{L}{2\pi}\right)^n \int |\delta_k|^2 k^{2m}\, d^n k. \tag{16.42}$$

This says that the field derivatives are independent of the values of the field and its second derivatives, but that the latter two quantities are correlated. We will be interested in points with $\delta' = 0$. At first sight, these may appear not to exist, since the probability that δ' is exactly zero vanishes. The way out of this paradox is to take the 'volume element' $d\delta\, d\delta''\, d\delta'$ and use the Jacobian to replace δ' by position, giving $d\delta\, d\delta''\, |\delta''|\, d^n x$; in n dimensions, $|\delta''|$ denotes the Jacobian determinant of the **Hessian matrix**, $\partial^2\delta/\partial x_i \partial x_j$. We then have only to integrate over the region of δ'' space corresponding to peaks, and we are done. To express the answers neatly, Bardeen *et al.* (1986) defined some (mainly dimensionless) parameters as follows:

$$\boxed{\nu \equiv \frac{\delta}{\sigma_0}, \quad \gamma \equiv \frac{\sigma_1^2}{\sigma_0 \sigma_2}, \quad R_* \equiv \sqrt{n}\,\frac{\sigma_1}{\sigma_2}.} \tag{16.43}$$

The meaning of these variables is as follows: ν is the 'height' of the field, in units of the rms; R_* is a measure of the coherence scale in the field; γ is a measure of the width of the power spectrum, with $\gamma = 1$ corresponding to a shell in k-space.

If we also put $x \equiv -\delta''/\sigma_2$, then the number density of stationary points in one dimension can be deduced immediately from the Gaussian distribution,

$$dN = \frac{e^{-Q/2}}{(2\pi)^{3/2}(1-\gamma^2)^{1/2} R_*}\, |x|\, dx\, d\nu, \qquad Q = \frac{(\nu - \gamma x)^2}{1-\gamma^2} + x^2, \tag{16.44}$$

with peaks corresponding to $x > 0$. Doing the ν integral first, followed by that over x, gives the total peak density as

$$N_{\mathrm{pk}} = \frac{1}{2\pi R_*}. \tag{16.45}$$

Doing the x-integral first is a tedious operation; nevertheless, the result can in fact be integrated again to yield the integral peak density (Cartwright & Longuet-Higgins 1956):

$$P(>\nu) = \frac{1}{2}\left(\mathrm{erfc}\left[\frac{\nu}{\sqrt{2(1-\gamma^2)}}\right] + \gamma e^{-\nu^2/2}\left\{1 + \mathrm{erf}\left[\frac{\gamma\nu}{\sqrt{2(1-\gamma^2)}}\right]\right\}\right). \tag{16.46}$$

The corresponding probability density for higher numbers of dimensions is quite easily deduced: the correlation between δ and δ'' means that, for very high peaks, all principal second derivatives tend to become equal and scale $\propto \nu$. It is then improbable that a stationary point is anything other than a peak, so the integration over second

derivatives is unrestricted, leading to

$$N_{\rm pk}(>\nu) = \frac{1}{\sqrt{2\pi}} \left(\frac{\gamma}{\sqrt{2\pi}R_*}\right)^n \nu^{n-1} e^{-\nu^2/2}. \tag{16.47}$$

In general, allowing for the constraint that all second derivatives must be negative is quite involved; details are given by Bardeen *et al.* (1986) and Bond & Efstathiou (1987). The final result can be expressed as

$$\frac{dN_{\rm pk}}{d\nu} = \frac{1}{(2\pi)^{(n+1)/2}} R_*^n e^{-\nu^2/2} G(\gamma, \gamma\nu), \tag{16.48}$$

where

$$G(\gamma, \gamma\nu) = \int_0^\infty F(x) \frac{\exp\left[-\frac{1}{2}(x-\gamma\nu)^2/(1-\gamma^2)\right]}{[2\pi(1-\gamma^2)]^{1/2}}\, dx. \tag{16.49}$$

In one dimension, $F(x) = x$ (Rice 1954), whereas $F(x) = x^2 + \exp(-x^2) - 1$ in two dimensions (Bond & Efstathiou 1987). $F(x)$ is given for $n = 3$ by equations (A1.9) and (4.5) of Bardeen *et al.* (1986). The integral densities in higher dimensions come out to be

$$\begin{aligned} N_{\rm pk} &= \frac{1}{4\pi\sqrt{3}R_*^2} \quad \text{(2D)} \\ N_{\rm pk} &= \frac{29 - 6\sqrt{6}}{8\pi^2 5^{3/2} R_*^3} \simeq 0.016 R_*^{-3} \quad \text{(3D)}. \end{aligned} \tag{16.50}$$

The above results are illustrated in figure 16.2. We see that the height distribution for peaks is much narrower than the unconstrained positive tail of a Gaussian. Since we may use the spherical model to justify taking $\delta \simeq 1$ to mark the creation of a bound object, and since δ grows as $(1+z)^{-1}$, we can convert a distribution of ν to a distribution of collapse redshifts. The relatively small fractional dispersion in ν for peaks means that it is possible to speak of a reasonably well-defined epoch of galaxy formation.

CLUSTERING OF PEAKS Another important calculation that can be performed with density peaks is to estimate the clustering of cosmological objects. Peaks have some inbuilt clustering as a result of the statistics of the linear density field: they are 'born clustered'. For galaxies, this clustering amplitude is greatly altered by the subsequent dynamical evolution of the density field, but this is not true for clusters of galaxies, which are the largest nonlinear systems at the current epoch. We recognize clusters simply because they are the most spectacularly large galaxy systems to have undergone gravitational collapse; this has an important consequence, as first realized by Kaiser (1984). The requirement that these systems have become nonlinear by the present means that they must have been associated with particularly high peaks in the initial conditions. If we thus confine ourselves to peaks above some **density threshold** in ν, the statistical correlations can be very strong – especially for the richer clusters corresponding to high peaks.

The main effect is easy to work out, using the **peak–background split**. Here, one conceptually decomposes the density field into short-wavelength terms, which generate the peaks, plus terms of much longer wavelength, which modulate the peak number density. Consider the large-wavelength field as if it were some extra perturbation δ_+; if

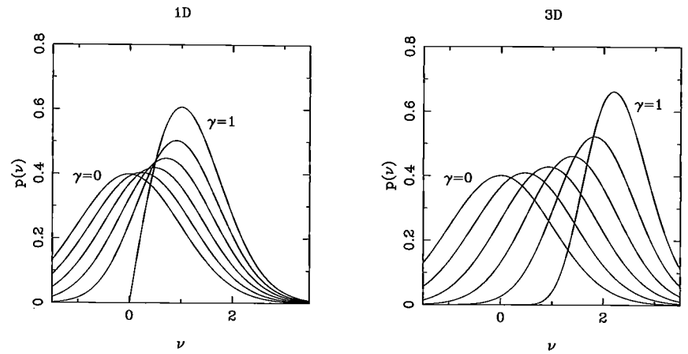

Figure 16.2. The distribution of the 'height' (relative to the rms) of maxima in Gaussian density fields, in one and three dimensions, as a function of the spectral parameter γ, for $\gamma = 0$, 0.2, 0.4, 0.6, 0.8, 1. As $\gamma \to 0$, we recover the unconstrained Gaussian case. For γ close to unity, most peaks are quite overdense ($\nu \simeq 2$) and the fractional dispersion in ν, which controls the dispersion in collapse times, is much smaller than in the unconstrained case.

we select all peaks above a threshold ν in the final field, this corresponds to taking all peaks above $\delta = \nu\sigma_0 - \delta_+$ in the initial field. This varying effective threshold will now produce more peaks in the regions of high δ_+, leading to amplification of the clustering pattern. For high peaks, $P(>\nu) \propto \nu^2 e^{-\nu^2/2}$; the exponential is the most important term, leading to a perturbation $\delta P/P \simeq \nu(\delta_+/\sigma_0)$. Hence, we obtain the high-peak amplification factor for the correlation function:

$$\xi_{\rm pk}(r) \simeq \frac{\nu^2}{\sigma_0^2} \xi_{\rm mass}(r). \tag{16.51}$$

It is important to realize that the process as described need have nothing to do with biased galaxy formation; it works perfectly well if galaxy light traces mass exactly in the universe. Clusters occur at special places in the mass distribution, so there is no reason to expect their correlations to be the same as those of the mass field.

In more detail, the exact clustering of peaks is just an extension of the calculation of the number density of peaks. We want to find the density of peaks of height ν_2 at a distance r from a peak of height ν_1. This involves a 6×6 covariance matrix for the fields and first and second derivatives even in 1D (20×20 in 3D). Moreover, most of the elements in this matrix are non-zero, so that the analytical calculation of ξ is sadly not feasible (see Lumsden, Heavens & Peacock 1989). However, a closely related calculation is easier to solve: the correlations of **thresholded regions**. Assume that objects form with unit probability in all regions whose density exceeds some threshold value, so that we need to deal with the correlation function of a modified density field that is constant