Statistics of fields for astronomers

Robert Crittenden

February 5, 2018

Abstract

Lecture notes from short course in statistics for the graduate students of the ICG. These notes build on another set of notes on purely statistical topics; see those for further background.

1 Introduction

Physicists and astronomers, even very theoretical ones, always want to know if their theories have any relation to the real universe. At some point they must compare the predictions of their theory to the observations, and this is where an understanding of statistics becomes essential.

Given the limited time, I will not be giving rigorous derivations, but will instead focus on giving an intuitive understanding of the topics. The goal is to give a general treatment that will provide working knowledge of the subject and how it is applied.

2 Statistics of fields

Thus far the discussion has been as general as possible, but it's worth spending some time discussing statistics which characterise fields. Applications of this arise all the time in astronomy: e.g., in analysing time stream data (1-d), images on the sky (2-d) or density fields in space (3-d).

For fields, the data are indexed by a position variable, $\mathbf{x}$ (or in one dimension, it is sometimes taken as a time variable, t.) These data may be well determined or they may contain noise. One usually assumes one is seeing a random sample, and trying to find characteristics of the underlying distribution.

A few key assumptions make it possible to probe the underlying distribution. The first is *homogeneity,* which requires that any joint probability distributions remain the same when the set of positions, $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3} \cdots$, is translated (but not rotated). This implies that the probabilities depend on the relative, not absolute, positions. (This is sometimes called *stationarity* for one dimensional fields, particularly in the time domain.) The second property is *isotropy*, which implies that any joint probabilities are unchanged when the set of positions are rotated. Finally, *ergodicity* implies that all the information about an underlying distribution can be learnt from a single (if infinite) realisation of the field. Ergodicity is obviously of great importance in a cosmological context, since we have only one realisation of the Universe!

Ergodicity means that we can imagine an infinite single field which we are sampling at a few finite points. From homogeneity, any statistical properties should be invariant of where we measure, so if we average over a big enough volume, we should find the underlying statistical properties of the field. Thus in effect, we can replace expectations by a volume average.

2.1 Correlation functions

For example, consider a scalar density field $\rho(\mathbf{x})$. The mean of the field can be found by

$$\bar{\rho} \equiv \langle \rho(\mathbf{x}) \rangle = \lim_{V \to \infty} \frac{1}{V} \int_V d^N \mathbf{x_0} \, \rho(\mathbf{x_0}), \tag{1}$$

which is the one point moment, and it must be independent of position by homogeneity. We could also measure the one point distribution function, which tells us what fraction of random samples the density lies in some range.

Let's now consider higher order moments of the field, but for simplicity subtract off the mean of the field; i.e., we will focus on $\delta(\mathbf{x}) = (\rho(\mathbf{x}) - \bar{\rho})/\bar{\rho}$. (In the earlier language, we focus on the central moments.)

The two point correlation is simply the covariance between the field measured at two points,

$$\xi(\mathbf{x}) = \langle \delta(\mathbf{x_0}) \delta(\mathbf{x_1}) \rangle, \tag{2}$$

where $\mathbf{x} = \mathbf{x_1} - \mathbf{x_0}$. Again, homogeneity implies it can only be a function of the relative positions of the points. If in addition the field is isotropic, then the correlation is a function of the distance between the points alone, $\xi(|\mathbf{x}|)$. For discrete fields, the correlation function describes the excess probability (compared to a random field) to see a pair of objects in two small volume elements:

$$dP(\mathbf{x_0}, \mathbf{x_1}) = \bar{\rho}^2 (1 + \xi(\mathbf{x})) dV_0 dV_1. \tag{3}$$

Finally, the variance of the field is the two point correlation function evaluated at the same point, $\xi(0)$.

The three point function is defined similarly,

$$\zeta(\mathbf{x_1}, \mathbf{x_2}) = \langle \delta(\mathbf{x_0}) \delta(\mathbf{x_0} + \mathbf{x_1}) \delta(\mathbf{x_0} + \mathbf{x_2}) \rangle. \tag{4}$$

Again, isotropy here means that the function only depends on the triangle of the relative positions of the points. There are three degrees of freedom, uniquely defined by the lengths of the edges.

We can define arbitrarily N point correlation functions in exactly the same way, and potentially they can tell us something fresh about the distribution. However, if we wish to focus on the new information, it makes sense to define connected correlation functions, completely analogously to the cumulants defined earlier. All the information of Gaussian fields is contained in the two point correlation function; any higher order correlations are simple functions of it, and can be found using Wick's theorem.

2.2 Fourier space statistics

We can also consider an alternate basis for the fields, and consider the statistics in this basis. The most useful basis to consider is Fourier space, defined as

$$\delta_{\mathbf{k}} = \frac{1}{V} \int d^N \mathbf{x} \delta(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{5}$$

This has the inverse,

$$\delta(\mathbf{x}) = \frac{V}{(2\pi)^N} \int d^N\mathbf{k}\delta_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}}. \tag{6}$$

We can now consider the two, three and higher point moments in Fourier space. The two point moments is

$$\langle \delta_{\mathbf{k}}\delta_{\mathbf{k}'} \rangle = \frac{1}{V^2} \int d^N\mathbf{x} d^N\mathbf{x}' \langle \delta(\mathbf{x})\delta(\mathbf{x}')\rangle e^{i(\mathbf{k}\cdot\mathbf{x}+\mathbf{k}'\cdot\mathbf{x}')}. \tag{7}$$

We can make the change of variable $\mathbf{x}'' = \mathbf{x}' - \mathbf{x}$ and use the fact that the two point function is only a function of the relative position $\mathbf{x}''$ to show,

$$\begin{aligned}
\langle \delta_{\mathbf{k}}\delta_{\mathbf{k}'} \rangle &= \frac{1}{V^2} \int d^N\mathbf{x} d^N\mathbf{x}'' \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{x}'')\rangle e^{i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}+i\mathbf{k}'\cdot\mathbf{x}''} \\
&= \delta_D(\mathbf{k}+\mathbf{k}') \frac{(2\pi)^N}{V^2} \int d^N\mathbf{x}'' e^{i\mathbf{k}'\cdot\mathbf{x}''} \xi(\mathbf{x}'') \\
&\equiv \frac{(2\pi)^N}{V} \delta_D(\mathbf{k}+\mathbf{k}') P(\mathbf{k}).
\end{aligned} \tag{8}$$

Here, we have integrated over $\mathbf{x}$ to get the Dirac delta function and defined the *power spectrum* $P(\mathbf{k})$ as the Fourier transform of the two point correlation function. (Be warned, there are many slightly different conventions for this definition.) The correlation function is simply the inverse transform of the power spectrum:

$$\xi(\mathbf{x}) = \frac{V}{(2\pi)^N} \int d^N\mathbf{k} P(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}}. \tag{9}$$

If we further assume isotropy, the correlation depends only on the separation distance, and so the power spectrum depends only on the magnitude, k. In these cases the angular integrals can usually be simply performed. In two dimensions, this is

$$\xi(x) = \frac{2\pi V}{(2\pi)^2} \int_0^\infty k dk P(k) J_0(kx), \tag{10}$$

where $J_0(z)$ is the ordinary Bessel function. In three dimensions, this is instead,

$$\xi(x) = \frac{4\pi V}{(2\pi)^3} \int_0^\infty k^2 dk P(k) j_0(kx), \tag{11}$$

where $j_0(z) = \sin(z)/z$ is a spherical Bessel function. The variance of the field in either case is given by

$$\sigma^2 = \xi(0) = \frac{V}{(2\pi)^N} \int d^N\mathbf{k} P(k) \propto \int k^{N-1} dk P(k). \tag{12}$$

Often we consider power spectra which follow a power law in k, $P(k) \propto k^n$, which are called *scale independent*. Two are of special significance: the *white noise* spectrum where the power is constant ($n = 0$), corresponding to a delta function for the correlation function, and the *scale invariant* spectrum ($n = -N$), corresponding to equal contributions per log interval of k. The scale invariant spectrum is both infra-red and ultra-violet log divergent. Redder spectra ($n < -N$) are infra-red divergent, while bluer spectra ($n > -N$) are ultra-violet divergent. Note that in large scale structure, we usually have near scale invariance in the potential field ($n = -3$), which corresponds to a bluer index for the density fluctuations ($n = 1$).

The analog to the three point correlation function is called the *bispectrum*, defined as

$$\langle \delta_{\mathbf{k_1}} \delta_{\mathbf{k_2}} \delta_{\mathbf{k_3}} \rangle \equiv \frac{(2\pi)^N}{V} \delta_D(\mathbf{k_1} + \mathbf{k_2} + \mathbf{k_3}) B(\mathbf{k_1}, \mathbf{k_2}), \tag{13}$$

where the Dirac delta function comes from homogeneity as above. The bispectrum is the transform of $\zeta(\mathbf{x_1}, \mathbf{x_2})$, and given isotropy is a function only of the shape of the triangle defined between $\mathbf{k_1}, \mathbf{k_2}$ and $\mathbf{k_3}$. Thus it is sometimes written as $B(k_1, k_2, k_3)$. Higher order correlations can be similarly defined.

Note that the power spectrum can be calculated from the magnitudes of the $\delta_{\mathbf{k}}$ alone, without any phase information. Since all the information of a Gaussian field is in the power spectrum, there should be none in the phases. Any phase correlations are a sign of non-Gaussianity.

2.3 Window functions

In real life its rarely possible to sample a field at an infinitesimal point; because of observational constraints like resolution issues, we actually measure a field smoothed over some scale. The kind of smoothing depends on the experiment, but in general the observed field is the underlying convolved with some window function, $W(\mathbf{y})$:

$$\delta^{obs}(\mathbf{x}) = \frac{1}{V} \int d^N \mathbf{y} W(\mathbf{y}) \delta(\mathbf{x} + \mathbf{y}) \tag{14}$$

The observed two point function can be found with a double convolution of the underlying one:

$$\xi^{obs}(\mathbf{x}) = \frac{1}{V^2} \int d^N \mathbf{y} d^N \mathbf{y'} W(\mathbf{y}) W(\mathbf{y'}) \xi(\mathbf{x} + \mathbf{y} - \mathbf{y'}). \tag{15}$$

This can be a bit hard to evaluate directly, but it simplifies significantly in Fourier space, where convolutions become products. Thus,

$$\delta^{obs}_{\mathbf{k}} = W_{\mathbf{k}} \delta_{\mathbf{k}}, \tag{16}$$

where $W_{\mathbf{k}}$ is the Fourier transform of $W(\mathbf{x})$. Assuming an isotropic window function, the observed power spectrum is $P^{obs}(k) = W^2(k)P(k)$; so, for example, the observed correlation function in three dimensions is given by,

$$\xi^{obs}(x) = \frac{4\pi V}{(2\pi)^3} \int_0^\infty k^2 dk W^2(k) P(k) j_0(kx). \tag{17}$$

Typical window functions include Gaussian functions and tophat functions, and these have the effect of damping any structure smaller than the smoothing length; that is, the high k modes are suppressed. Conventionally, the window function is usually defined such that

$$\frac{1}{V} \int d^N \mathbf{y} W(\mathbf{y}) = 1. \tag{18}$$

For example, the 3-d Gaussian window function in real space is

$$W(\mathbf{x}) = \frac{V}{(2\pi)^{3/2}R^3} e^{-r^2/2R^2} \tag{19}$$

and its Fourier transform is simply $W_{\mathbf{k}} = e^{-k^2R^2/2}$ which $\to 1$ for $k \to 0$.

2.4 Selection function

It is also impossible to sample a field at every point in space. The above approach can also be used to take into account limits to the geometry of the sample. Here though, it is a product in real space, and a convolution in Fourier space. These constraints make the measurements less sensitive to large scale (low k) power. The observations should not be sensitive to wavenumbers much smaller than the inverse of the characteristic scale (D) of the survey.

The sample geometry is characterised by the selection function, $f(\mathbf{x})$, which is normalised similarly to the window function. For galaxy surveys, this is often the product of a sky mask times the redshift selection function. The observed structure is simply the product of the true structure times the selection function, $\delta^{obs}(\mathbf{x}) \equiv \delta(\mathbf{x})f(\mathbf{x})$. In Fourier space, this becomes a convolution,

$$\delta^{obs}_{\mathbf{k}} = \frac{V}{(2\pi)^N} \int d^N\mathbf{k}' \delta_{\mathbf{k}'} f_{\mathbf{k}-\mathbf{k}'}. \tag{20}$$

Note that the selection function explicitly breaks breaks translation invariance, meaning that the two-point statistics are no longer diagonal in Fourier space. However, their expectations can be calculated similarly,

$$\langle \delta^{obs}_{\mathbf{k}} \delta^{obs*}_{\mathbf{k}'} \rangle = \frac{V}{(2\pi)^N} \int d^N\mathbf{k}'' P_{\mathbf{k}''} f_{\mathbf{k}-\mathbf{k}''} f^*_{\mathbf{k}'-\mathbf{k}''}. \tag{21}$$

Rather than trying to invert this expression directly for the true power spectrum, it is more common to use search some parameterisation of the underlying power spectrum for one which can reproduce the observed two-point statistics.

Yet one more complication exists when sampling a field over a finite patch, which is that we usually do not observe the fractional over density, but rather the total density. To construct overdensity, we would need to know the true background density of the field, whereas usually we only know the density observed locally. If our inferred density contrast is $\tilde{\delta}(\mathbf{x})$, it must be related to the true density contrast through, $\rho(\mathbf{x}) = \bar{\rho}(1 + \delta(\mathbf{x})) = \bar{\rho}_{local}(1 + \tilde{\delta}(\mathbf{x}))$, where

$$\bar{\rho}_{local} = \bar{\rho} + \frac{1}{V} \int d^N\mathbf{x} f(\mathbf{x})(\bar{\rho}\delta(\mathbf{x})). \tag{22}$$

Thus, the inferred density is given by,

$$\tilde{\delta}(\mathbf{x}) = \frac{\bar{\rho}}{\bar{\rho}_{local}} \times \left(\delta(\mathbf{x}) - \frac{1}{V} \int d^N\mathbf{x} f(\mathbf{x})\delta(\mathbf{x}) \right) \tag{23}$$

The prefactor will give a small correction in the normalisation of the inferred power spectrum, and we will drop it for simplicity.

Again, the observed field is only measured over a finite region, so the true observable is given by $\tilde{\delta}^{obs}(\mathbf{x}) = \tilde{\delta}(\mathbf{x}) f(\mathbf{x})$. In Fourier space, it can be written as

$$\begin{aligned} \tilde{\delta}^{obs}_{\mathbf{k}} &= \frac{V}{(2\pi)^N} \int d^N\mathbf{k}' \delta_{\mathbf{k}'} f_{\mathbf{k}-\mathbf{k}'} - f_{\mathbf{k}} \frac{1}{V} \int d^N\mathbf{x} f(\mathbf{x}) \delta(\mathbf{x}) \\ &= \frac{V}{(2\pi)^N} \int d^N\mathbf{k}' \delta_{\mathbf{k}'} \left(f_{\mathbf{k}-\mathbf{k}'} - f_{\mathbf{k}} f_{\mathbf{k}'} \right). \end{aligned} \tag{24}$$

By definition, we are subtracting off the local mean and so the average of this new field should be zero; in Fourier space, this corresponds to $\tilde{\delta}^{obs}_{\mathbf{0}} = 0$, which follows from above since $f_{\mathbf{0}} = 1$. This is known as the *integral constraint,* and the above equation demonstrates that the observables will not depend on modes much larger than the region which is probed.

Note that the effective power spectrum window is somewhat different from that implied by Peacock (eq. 16.127). There, the effective window for $\langle \tilde{\delta}^{obs}_{\mathbf{k}} \tilde{\delta}^{obs*}_{\mathbf{k}} \rangle$ is $|f_{\mathbf{k}-\mathbf{k}'}|^2 - |f_{\mathbf{k}}|^2 |f_{\mathbf{k}'}|^2$, whereas squaring the above expression gives $|f_{\mathbf{k}-\mathbf{k}'} - f_{\mathbf{k}} f_{\mathbf{k}'}|^2$. These do not appear to be equivalent and evaluating them for exponential selection functions they seem to differ by about a factor of two for small k and k', though they both obey the integral constraint. (Perhaps I have erred, or perhaps the Peacock analysis ignores the fact that the mean is not any number, but is itself a function of the density field and so is correlated with it.)

3 Projection from 3-D to 2-D

In astronomy, three dimensional information is often only observed in projection, with great uncertainty in the radial direction (parallel to the line of sight.) Therefore, it is useful to understand how the true three-dimensional statistics are translated into projected quantities.

3.1 Planar projection

Let's assume that the three-dimensional density field, $\rho_{3D}(\mathbf{x})$, is projected by a radial function $F(x_\parallel)$ that describes the probability that an object at a given distance is included into the projected two-dimensional map. Alternatively, for continuous fields this function describes the radial weighting of the three dimensional scalar field. Thus,

$$\rho_{2D}(\mathbf{x}_\perp) \equiv \int dx_\parallel F(x_\parallel) \rho_{3D}(x_\parallel, \mathbf{x}_\perp). \tag{25}$$

The radial selection function has an effective thickness, $L_\parallel \equiv \int dx_\parallel F(x_\parallel)$, that relates the mean densities; that is, $\bar{\rho}_{2D} = L_\parallel \bar{\rho}_{3D}$. The effective weighting of the density contrast is normalised, so that

$$\delta_{2D}(\mathbf{x}_\perp) \equiv \frac{\rho_{2D}(\mathbf{x}_\perp) - \bar{\rho}_{2D}}{\bar{\rho}_{2D}} = \int dx_\parallel \frac{F(x_\parallel)}{L_\parallel} \delta_{3D}(x_\parallel, \mathbf{x}_\perp). \tag{26}$$

In real space, the two and three dimensional correlation functions are simply related:

$$\xi_{2D}(x_\perp) = \int dx_\parallel \frac{F(x_\parallel)}{L_\parallel} \int dx'_\parallel \frac{F(x'_\parallel)}{L_\parallel} \xi_{3D}\left(\sqrt{(x_\parallel - x'_\parallel)^2 + x_\perp^2} \right). \tag{27}$$

We can relate the power spectra similarly. The two-dimensional Fourier amplitudes are given by

$$\begin{aligned}
\delta_{2D}(\mathbf{q}_\perp) &= \frac{1}{V_{2D}} \int d^2\mathbf{x}_\perp \delta_{2D}(\mathbf{x}_\perp) e^{i\mathbf{q}_\perp\cdot\mathbf{x}_\perp} \qquad (28)\\
&= \frac{1}{V_{2D}} \int d^2\mathbf{x}_\perp \int dx_\parallel \frac{F(x_\parallel)}{L_\parallel} \delta_{3D}(x_\parallel, \mathbf{x}_\perp) e^{i\mathbf{q}_\perp\cdot\mathbf{x}_\perp}\\
&= \frac{V_{3D}}{V_{2D}(2\pi)^3} \int d^2\mathbf{x}_\perp \int dx_\parallel \frac{F(x_\parallel)}{L_\parallel} \int d^3\mathbf{k}\,\delta(\mathbf{k}) e^{i(\mathbf{q}_\perp-\mathbf{k}_\perp)\cdot\mathbf{x}_\perp - ik_\parallel x_\parallel}\\
&= \frac{V_{3D}}{V_{2D}(2\pi)} \int dx_\parallel \frac{F(x_\parallel)}{L_\parallel} \int dk_\parallel\,\delta(k_\parallel, \mathbf{q}_\perp) e^{-ik_\parallel x_\parallel}.
\end{aligned}$$

The two point correlation is,

$$\begin{aligned}
\langle \delta_{2D}(\mathbf{q}_\perp)\delta_{2D}(\mathbf{q}'_\perp)\rangle &\equiv \frac{(2\pi)^2}{V_{2D}} \delta^{(2)}(\mathbf{q}_\perp + \mathbf{q}'_\perp) P_{2D}(\mathbf{q}_\perp) \qquad (29)\\
&= \frac{V_{3D}(2\pi)}{V_{2D}^2} \int dx_\parallel \frac{F(x_\parallel)}{L_\parallel} \int dx'_\parallel \frac{F(x'_\parallel)}{L_\parallel}\\
&\quad \times \int dk_\parallel\, e^{-ik_\parallel(x_\parallel + x'_\parallel)} P(k_\parallel, \mathbf{q}_\perp) \delta^{(2)}(\mathbf{q}_\perp + \mathbf{q}'_\perp).
\end{aligned}$$

Thus, the power spectra are related by

$$\begin{aligned}
P_{2D}(\mathbf{q}_\perp) &= \frac{V_{3D}}{V_{2D}(2\pi)} \int dx_\parallel \frac{F(x_\parallel)}{L_\parallel} \int dx'_\parallel \frac{F(x'_\parallel)}{L_\parallel}\\
&\quad \times \int dk_\parallel\, e^{-ik_\parallel(x_\parallel + x'_\parallel)} P(k_\parallel, \mathbf{q}_\perp). \qquad (30)
\end{aligned}$$

3.2 The Limber approximation

Above we have shown the exact expressions for the two-dimensional correlation and power spectra, but they can be considerably simplified using Limber's approximation, which assumes that the weight functions are relatively thick in comparison to the coherence length of the correlation function. We can then assume that the weight functions depend only on the average, while the correlation function is integrated over,

$$\xi_{2D}(x_\perp) = \int d\bar{x}_\parallel \frac{F^2(\bar{x}_\parallel)}{L_\parallel^2} \int d\Delta x_\parallel \xi_{3D}\left(\sqrt{(\Delta x_\parallel)^2 + x_\perp^2}\right). \qquad (31)$$

If the correlation function evolves in time, it is evaluated at the mean distance, $\bar{x}_\parallel$. This approximation is only valid in the thick shell limit, and implicitly we have assumed the flat sky limit, though the latter is not required.

In Fourier space, the Limber approximation effectively sets $k_\parallel \sim 0$. (That is, assuming the second line is effectively zero for large $\Delta x_\parallel$.) This leads to the standard Limber approximation for power spectra,

$$P_{2D}(q_\perp) \simeq \frac{V_{3D}}{V_{2D}} \int dx_\parallel \frac{F^2(x_\parallel)}{L_\parallel^2} P(0, q_\perp). \qquad (32)$$

3.3 Full sky statistics and the small angle approximation

Above we have assumed we observe a three-dimensional volume projected down to a flat plane, which should be a reasonable approximation for surveys covering a small part of the sky. For larger surveys, one needs to use a basis defined for the celestial sphere, and it is more appropriate to calculate the multipole moments as is typically done for the CMB.

We define the projected map on the sky by

$$s(\hat{\mathbf{n}}) = \int d\chi \frac{F(\chi)}{L_\chi} \delta_{3D}(\chi\hat{\mathbf{n}}) \tag{33}$$

where χ is the co-moving distance and $\int d\chi F(\chi) = L_\chi$. In analogue to the Fourier expansion, we can expand the projected map into the orthogonal spherical harmonic functions, $Y_{\ell m}(\hat{\mathbf{n}})$, so that

$$s(\hat{\mathbf{n}}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{\mathbf{n}}) \tag{34}$$

where

$$a_{\ell m} = \int d\Omega_{\mathbf{n}} s(\hat{\mathbf{n}}) Y^*_{\ell m}(\hat{\mathbf{n}}). \tag{35}$$

Here, the ℓ value corresponds roughly with the angular scale of the fluctuations ($\theta \sim \pi/\ell$); for each ℓ, there are $2\ell+1$ values of $m = -\ell, -\ell+1, \cdots, \ell-1, \ell$ that correspond to different orientations.

Rotational invariance implies that the two point expectations of these amplitudes take the form $\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \delta_{\ell\ell'} \delta_{mm'}$. The two-point correlation function is simply related to these,

$$C(\theta) = \langle s(\hat{\mathbf{n}}) s(\hat{\mathbf{n}}') \rangle = \sum_\ell C_{\ell m} Y_{\ell m}(\hat{\mathbf{n}}) Y^*_{\ell m}(\hat{\mathbf{n}}') = \frac{1}{4\pi} \sum_\ell (2\ell+1) C_\ell P_\ell(\cos\theta) \tag{36}$$

where $\cos\theta = \hat{\mathbf{n}} \cdot \hat{\mathbf{n}}'$ and $P_\ell(x)$ are the Legendre polynomials.

We can relate the amplitudes to the Fourier amplitudes of the three dimensional field:

$$\begin{aligned} a_{\ell m} &= \int d\Omega_{\mathbf{n}} \int d\chi \frac{F(\chi)}{L_\chi} \frac{V_{3D}}{(2\pi)^3} \int d^3\mathbf{k}\, \delta(\mathbf{k})\, e^{i\mathbf{k}\cdot\chi\hat{\mathbf{n}}}\, Y^*_{\ell m}(\hat{\mathbf{n}}) \\ &= \frac{4\pi V_{3D}}{(2\pi)^3} \int d\chi \frac{F(\chi)}{L_\chi} \int d^3\mathbf{k}\, \delta(\mathbf{k}) i^\ell j_\ell(k\chi) Y^*_{\ell m}(\hat{\mathbf{k}}) \end{aligned} \tag{37}$$

where the second line follows from the plane wave expansion and $j_\ell(x)$ are spherical Bessel functions.

The two-point expectation of these is thus,

$$\begin{aligned} C_\ell &= \langle a_{\ell m} a^*_{\ell m} \rangle \\ &= \frac{2V_{3D}}{\pi} \int d\chi \frac{F(\chi)}{L_\chi} \int d\chi' \frac{F(\chi')}{L_\chi} \int k^2 dk\, P_{3D}(k) j_\ell(k\chi) j_\ell(k\chi'). \end{aligned} \tag{38}$$

The Limber approximation in this case effectively assumes that the power spectrum is slowly varying compared to the spherical Bessel functions, and so can be pulled out of the k integral (evaluating it where the argument of the integral peaks, at $k = (\ell + \frac{1}{2})/\chi$.) By the orthogonality of the spherical Bessel's functions, the k becomes a Dirac delta function, and the expression simplifies considerably,

$$C_\ell = V_{3D} \int d\chi \frac{F^2(\chi)}{L_\chi^2 \chi^2} P_{3D}(k = (\ell + \frac{1}{2})/\chi). \tag{39}$$

We can treat small patches of the sky as if they were effectively flat, so for small angles there should be a direct correspondence between the multipole spectrum and the two-dimensional spectrum defined above. (For example, see J. R. Bond Les Houches review (1993), p. 515.) One can choose flat coordinates that correspond the angular coordinates around a given pole, $(x, y) = (r\cos\phi, r\sin\phi)$ where $r = 2\sin(\theta/2)$. In this way the solid angle around a point corresponds to the flat area, i.e. $d\Omega = \sin\theta d\theta d\phi \sim r dr d\phi = dx dy$. Equating expressions for the correlation function,

$$C(\theta) = \frac{1}{4\pi} \sum_\ell (2\ell + 1) C_\ell P_\ell(\cos\theta) \rightarrow \frac{1}{2\pi} \int d\ell (\ell + \frac{1}{2}) C_\ell P_\ell(\cos\theta) \tag{40}$$

$$\simeq \frac{V_{2D}}{(2\pi)} \int_0^\infty q dq P_{2D}(q) J_0(qr) \tag{41}$$

In the limit of large ℓ and small r, $P_\ell(\cos\theta) \simeq J_0((\ell + \frac{1}{2})r)$; this implies a relationship between the spectra of the form

$$\ell(\ell + 1) C_\ell \simeq V_{2D} q^2 P_{2D}(q) \tag{42}$$

where $(\ell + \frac{1}{2}) \simeq qr$.

3.4 Convergence power spectra

An important application of this is in the predicted convergence power spectrum, where we follow the derivation of Bartelmann and Scheider, 1996. The convergence felt by a population at some coming distance is given by,

$$\kappa_{\rm eff}(\theta, w) = \frac{3H_0^2\Omega_0}{2c^2} \int dw' \frac{f(w')f(w - w')}{f(w)a} \delta(f(w')\theta, w') \tag{43}$$

Integrating over a population of galaxies in some redshift bin, we have

$$\tilde{\kappa}_{\rm eff}(\theta, i) = \frac{3H_0^2\Omega_0}{2c^2} \int dw \bar{W}_i(w) f(w) \frac{\delta(f(w')\theta, w')}{a(w)} \tag{44}$$

where

$$\bar{W}_i(w) = \int dw' G(w') \frac{f(w' - w)}{f(w')} \tag{45}$$

and $G(w)dw = p_i(z)dz$ describes the probability distribution in a given bin.

4 Discrete fields

One is often dealing with discrete fields in astronomy, the most obvious example being the observed density of galaxies. Thus, it is useful to consider some issues that arise uniquely in the discrete context.

We can divide the

An example is the interpretation of correlation functions. Correlation functions can be interpreted

5 Wavelets and other transformations

$\delta(\mathbf{x})$ is localised in real space, and $\delta_{\mathbf{k}}$ is localised in Fourier space. We can also consider transformations of the data which are localised in neither, and these are generically called wavelets. One can then perform the same statistical measures on the amplitudes in the wavelet basis.

For example, one can smooth the field with a many Gaussian window functions, centred at different positions and using a range of smoothing lengths, and use these amplitudes to represent the data instead. Ideally the wavelet basis should contain the same information as field, with the same number of amplitudes so there is no redundant information.

In practise wavelets are often chosen on the basis of how fast they are to compute. There are a number of fast wavelet transforms for discrete data which are typically used. While one can easily measure the probability distributions in a wavelet basis, its generally not particularly a good basis for finding analytic predictions. Generally they are used for comparing to simulated data sets.

Wavelets are often a good way of compressing data; one simply keeps those which have an amplitude greater than a particular threshold. This is necessarily a *lossy* compression (some information is lost); but with the right wavelet choice you can ensure that the important information is stored more compactly.

6 Measures of non-Gaussianity

Often one wants to know if a particular map is Gaussian. To do this, one chooses some (sometimes arbitrary) statistics of the map and compares them to a collection of Gaussian maps with the same power spectrum. Its best to find a statistic which will be sensitive to the form of non-Gaussianity you might expect to see. However, beware of falling into the trap of defining the statistic based on the map you wish to test; such posteriorly defined statistics are nearly impossible to interpret.

It can be useful to look at topological measures of the field. For example, consider all regions where the field is above (or below) a given threshold. What is the distribution of volumes for the disconnected regions? How are their total surface areas distributed? How many handles (holes) does a typical region have? These measures are known as *Minkowski functionals*. While useful measures, they tend to emphasises small scale features where noise is often dominant.

Other interesting statistics focus on the peaks or minima of the fields. For example, one can make a discrete map of the maxima above a given threshold, and look at their density and higher order correlation functions.

7 Advanced topics

7.1 Future advanced topics

- Resampling techniques
- Tests of isotropy

References

[1] P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences,* McGraw-Hill (1969).

[2] G. Casella, R. L. Berger, *Statistical Inference,* Duxbury (1990). An introduction to orthodox statistics.

[3] S. Dodelson, *Modern Cosmology* Academic Press (2003). See particularly chapter 11 on data analysis.

[4] E. T. Jaynes, *Probability Theory– the Logic of Science,* Cambridge 2003. An excellent discussion of the foundations of statistics.

[5] M. G. Kendall, A. Stuart, (A. O'Hagan, J. K. Ord), *Advanced Theory of Statistics,* Arnold Publishers (1998). Classic text, now includes a volume on Bayesian approach.

[6] R. Lupton, *Statistics in Theory and Practice* Princeton University Press (1993). General book written by a barefoot practising astronomer.

[7] D. J. C. MacKay, *Information theory, inference, and learning algorithms* Cambridge (2003). A good how-to book for MCMC, neural networks, data compression.

[8] J. A. Peacock, *Cosmological Physics,* Cambridge (1999).

[9] P. J. E. Peebles, *The Large Scale Structure of the Universe,* Princeton (1980). Defined the subject, but can be a hard read.

[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes – The Art of Scientific Computing* Cambridge (1992). A good, concise introduction to a number of numerical techniques written by astronomers; includes Fourier and wavelet transforms, Wiener filtering, K-S tests, etc.

[11] D. S. Sivia, *Data Analysis: A Bayesian Tutorial*, Oxford University Press (1996). Short, very clear introduction to the subject.

[12] E. Vanmarke, *Random Fields: Analysis and Synthesis*, MIT Press (1983).

Appendix A: Glossary of some statistical terms

boot-strapping - a resampling technique where a set of simulated data sets are created from an observed data set by use of resampling with replacement. This is done either by random sampling or by exhausting all ways the resampling could be done.

Gibbs' sampling - a Monte Carlo method where the random steps are taken in a single dimension, alternating in turn through every dimension.

independent component analysis - a method of blind source separation; it typically assumes the observations are an unknown linear transformation of a number of independent non-Gaussian sources. The method attempts to find the inverse transformation which maximises the inferred non-Gaussianity of the inferred sources.

jack-knife - a specific resampling technique where properties such as the variance are derived from the ensemble of smaller data sets where one or more points (or regions) of the data are omitted.

Kahunen-Loeve transform or decomposition - effectively another name for principal component analysis, it describes the transformation of data into uncorrelated components.

Kolmogorov-Smirnov test - a method of comparing whether two distributions are consistent by finding the largest value of the difference between their cumulative distribution functions. See Numerical Recipes for a nice discussion of it and extensions.

Martingale series - a random process where the expectation value of the next observation is simply given by the present observation, though possibly in a way that involves past observations.

Markov series - a random process where the next observation depends only on the present observation and not on the past.

non-parametric models - while any statistical description requires some choice of parameters to be quantified, in astronomy this generally refers to a choice which does not assume some particular model; for example, inferring bins of a power spectrum.

robust statistics - a set of statistics which may not be optimal, but is more robust against a small contimation by outliers which might be non-Gaussian or have a large variance. For example, the median may be a more robust estimator of the centre of a distribution than the mean, which could be thrown off by a single large outlier.

Appendix B: Useful Relations of Special Functions

Spherical harmonic relations:

- Definition

$$Y_{lm}(\theta,\phi) = \left(\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}\right)^{1/2} P_l^m(\cos\theta)e^{im\phi}$$

- Orthogonality -

$$\int Y_{\ell m}(\mathbf{n})Y^*_{\ell' m'}(\mathbf{n})d\Omega_{\mathbf{n}} = \delta_{\ell\ell'}\delta_{mm'}$$

- Completeness -

$$\sum_{\ell m} Y_{\ell m}(\mathbf{n})Y^*_{\ell m}(\mathbf{n}') = \delta(\phi-\phi')\delta(\cos\theta-\cos\theta')$$

- Plane wave expansion -

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi\sum_{\ell m} i^\ell j_\ell(kr)Y_{\ell m}(\widehat{\mathbf{k}})Y^*_{\ell m}(\widehat{\mathbf{r}}) = \sum_\ell (2\ell+1)i^\ell j_\ell(kr)P_\ell(\cos\theta),$$

  where $j_\ell(x)$ are the spherical Bessel functions.

Legendre relations (note $P_l(x) = P_l^0(x)$):

- Legendre equation

$$(1-x^2)\frac{d^2}{dx^2}P_l^m(x) - 2x\frac{d}{dx}P_l^m(x) + \left[l(l+1) - \frac{m^2}{1-x^2}\right]P_l^m(x) = 0$$

- Orthogonality

$$\int_{-1}^{1} P_l^m(x)P_{l'}^m(x)dx = \frac{2}{2l+1}\frac{(l+m)!}{(l-m)!}\delta_{ll'}$$

- Rodrigues' formula

$$P_l^m(x) = \frac{(-1)^m}{2^l l!}(x^2-1)^{m/2}\frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l$$

- Derivative relation

$$(1-x^2)\frac{d}{dx}P_l^m(x) = (l-m+1)P_l^m(x) - (l+1)xP_l^m(x)$$

- Recurrence relation

$$(2l+1)xP_l^m(x) = (l-m+1)P_{l+1}^m(x) + (l+m)P_{l-1}^m(x)$$

- Relation to spherical harmonics

$$(2l+1)P_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}) = 4\pi\sum_m Y_{lm}(\hat{\mathbf{k}})Y^*_{lm}(\hat{\mathbf{r}})$$

Bessel relations (useful for 2-d relations):

- Bessel equation

$$\frac{d^2}{dx^2}J_m(x) + \frac{1}{x}\frac{d}{dx}J_m(x) + \left(1 - \frac{m^2}{x^2}\right)J_m(x) = 0$$

- Orthogonality

$$\int kdkJ_m(k\rho)J_m(k\rho') = \frac{1}{\rho}\delta(\rho-\rho')$$

- Plane wave expansion

$$e^{ik\rho\cos\phi} = \sum_m i^m e^{im\phi}J_m(k\rho)$$

- Integral representation

$$J_m(x) = \frac{1}{2\pi i^m}\int_0^{2\pi} d\phi e^{ix\cos\phi - im\phi}$$

- Derivative relation

$$\frac{d}{dx}J_m(x) = \frac{1}{2}\left[J_{m-1}(x) - J_{m+1}(x)\right]$$

- Recurrence relation

$$J_{m+1}(x) = \frac{2m}{x}J_m(x) - J_{m-1}(x)$$

- Spherical Bessel functions

$$j_m(x) \equiv \left(\frac{\pi}{2x}\right)^{1/2} J_{m+\frac{1}{2}}(x)$$