Introduction to Power Spectra

Adrian Liu

January 21, 2019

1 Introduction

Power spectra are designed to quantify correlations in cosmological fields. An example of a cosmological field would be the density field, $\rho(\mathbf{r})$, which is simply the density of matter as a function of three-dimensional position $\mathbf{r}$. Another example would be the brightness temperature of the 21 cm line, $T(\mathbf{r})$.

Cosmological fields are typically correlated. By this we mean that if we imagine splitting up a three-dimensional space into a set of discrete voxels (a voxel being the 3D equivalent of a pixel), then we would expect the value of the cosmological field in one voxel to be related to its value in nearby voxels. There are a number of physical reasons for why this might be the case. One of these is gravity. Gravity draws things together, which is why galaxies are not uniformly distributed in our Universe, but instead, tend to be found near each other. Phrased in terms of correlations, the value of the density field at a voxel centred around some galaxy is likely to be positively correlated with its neighbouring voxels.

In fact, studying correlations is the key to understanding the physics that drives our Universe. There do not exist theoretical models that make predictions for the exact field that one expects to measure. This is because the fields are driven by random processes. Thus, our models cannot predict precisely what the value of, say, the density field ρ is at a particular location any more than we can say for sure that the 7th coin flip in a series of coin flips will yield tails. However, theoretical models *do* yield concrete predictions for the correlation information inherent in the fields.

2 The correlation function

One way to quantify the correlations in a field is to compute what is known as the *correlation function*. Suppose we had a field with zero mean. Using the temperature field as an example for the rest of this document, we would express this mathematically as

$$\langle T(\mathbf{r}) \rangle = 0, \tag{1}$$

where the pointy brackets $\langle \cdots \rangle$ denote an *ensemble average*. An ensemble average is a hypothetical averaging process where we imagine averaging the results from an infinite collection of universes (!), each generated from a different set of initial conditions but the same physical laws. This is of course not an average that we can actually perform in practice with real data, but it's a very useful way to think about things from a formal mathematical standpoint.

At this point you might be wondering how the average of the temperature could be zero. Wouldn't this require the field to be (statistically speaking) negative half the time? How can temperatures be negative? There are several answers to this. First, note that when we use terms like "brightness temperature", we are not referring to a physical temperature. Instead, temperature is used as a proxy for the strength of spectral emission or absorption, so a negative temperature simply signifies absorption rather than emission. Another answer is that when we quantify the statistics of a field, we usually assume that somebody has subtracted the mean of the field for us ahead of time. This is because the mean is often less interesting than the *fluctuations* about the mean. Going back to the density field for a moment, knowing that the average density of our Universe is about $10^{-26}\,\mathrm{kg/m^3}$ isn't as useful as knowing that a certain part of our Universe is denser than average, and therefore is likely to be where galaxies form. Thus, when I write $T(\mathbf{r})$ I should really be clear that I am referring to the fluctuations about the mean and write $\delta T(\mathbf{r})$, but the latter gets cumbersome so I go with the former for notational simplicity.

The correlation function between two points is defined as

$$\xi(\mathbf{r}_1, \mathbf{r}_2) \equiv \langle T(\mathbf{r}_1) T(\mathbf{r}_2) \rangle. \tag{2}$$

To make sense of this, suppose that the field is uncorrelated between our two chosen points. Then $\langle T(\mathbf{r}_1) T(\mathbf{r}_2) \rangle = \langle T(\mathbf{r}_1) \rangle \langle T(\mathbf{r}_2) \rangle = 0$. The correlation function is then zero, as one would expect. If the field is positively correlated between the two points, then if $T(\mathbf{r}_1)$ is positive (i.e., higher than average) then $T(\mathbf{r}_2)$ will also likely be positive. Similarly, if $T(\mathbf{r}_1)$ is negative, then $T(\mathbf{r}_2)$ probably will be too. Multiplying them together will then give a positive correlation function, as expected.

There are two additional simplifications that can be made. First, cosmological fields tend to have *translation invariant* statistics. This does not mean that our Universe is the same everywhere. All it means is that every point is drawn from the same probability distribution. Going back to the analogy with coin flipping, every flip follows the same probability distribution (50% heads and 50% tails) but that doesn't mean every flip gives the same result. Another way to say this is that since the statistical properties at every point are the same, there is no such thing as a special location in our Universe. It follows that the correlation function must only be a function of the *difference* between $\mathbf{r}_1$ and $\mathbf{r}_2$, and not of $\mathbf{r}_1$ and $\mathbf{r}_2$ individually. In other words, instead of our previous expression we can write

$$\xi(\mathbf{r}) \equiv \xi(\mathbf{r}_1 - \mathbf{r}_2) \equiv \langle T(\mathbf{r}_1) T(\mathbf{r}_2) \rangle, \tag{3}$$

where ξ is now a function of a single (vector) variable, the separation vector between the two points.

The second assumption that we can make is that our Universe is *statistically isotropic*, so that there is no preferred direction. A given point should be just as correlated to a point 100, Mpc away in the x direction as it is to a point 100 Mpc away in the y direction. This means that the correlation function should depend only on the magnitude r of $\mathbf{r}$, and not on its direction. This gives

$$\xi(r) \equiv \xi(|\mathbf{r}_1 - \mathbf{r}_2|) \equiv \langle T(\mathbf{r}_1) T(\mathbf{r}_2) \rangle. \tag{4}$$

While the assumption of isotropy is a good one for the actual fields of our Universe, we often discard it when dealing with real data. Real-world effects can cause the correlation function to be anisotropic. Such effects can be due to observational realities or astrophysical effects (such as "redshift space distortions". Look this up if you're curious!)

Shown in Figure 1 is a correlation function of galaxies [1]. The correlation is high at small separations between galaxies. This is expected, because galaxies like to cluster around each other under gravity. The correlation function drops off as the separation increases, although there is a curious bump at $\sim 100h^{-1}$Mpc. This length scale is known as the *baryon acoustic oscillation* scale, and is one of the most useful tools of observational cosmology.

3 The power spectrum

Whereas the correlation function lives in position space (also known as *configuration* space in cosmology), the power spectrum is a function that is defined in Fourier space. There are different ways to define it. One way to define it is to say that it is the Fourier transform of the correlation function:

$$P(\mathbf{k}) \equiv \int_{-\infty}^{\infty} d^3 r \, e^{-i\mathbf{k}\cdot\mathbf{r}} \xi(\mathbf{r}). \tag{5}$$

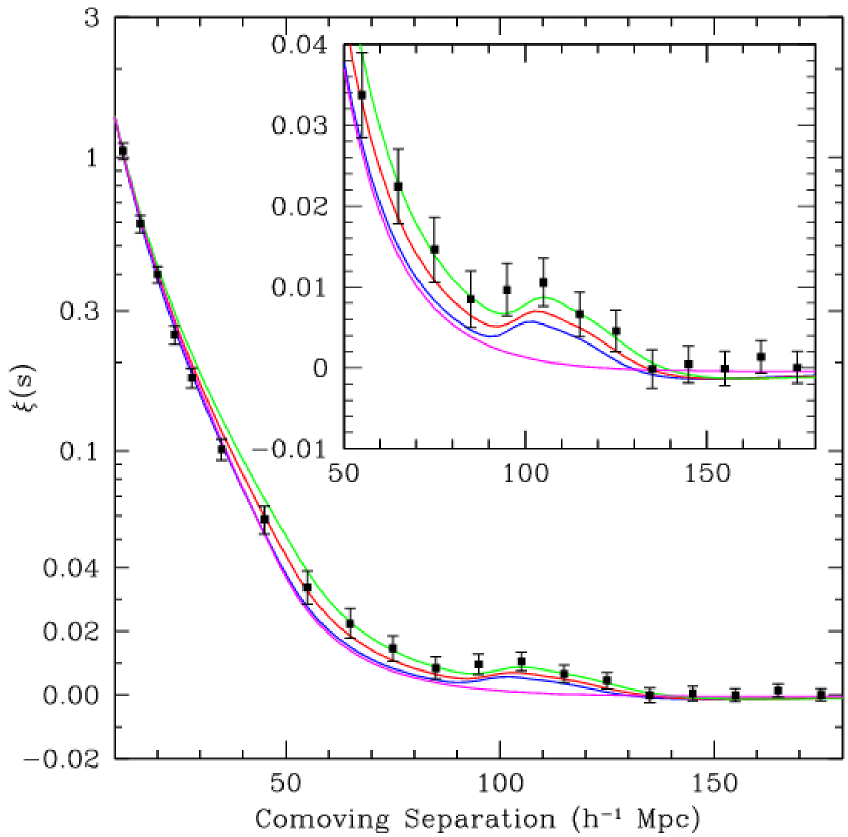

Figure 1: Correlation function of galaxies, measured using data from the Sloan Digital Sky Survey in Ref. [1].

Note that we are going back to the slightly more general form of the correlation function $\xi(\mathbf{r})$ that does not assume isotropy, rather than the one that does, $\xi(r)$. Using the inverse Fourier transform, we can also say[1]

$$\xi(\mathbf{r}) = \int_{-\infty}^{\infty} \frac{d^3k}{(2\pi)^3}\, e^{i\mathbf{k}\cdot\mathbf{r}} P(\mathbf{k}). \tag{6}$$

To push the idea of the power spectrum being the Fourier space version of the correlation function, let us actually calculate a Fourier space version of the

[1]Our Fourier convention is the one used most frequently in cosmology: going from $\mathbf{r}$ to $\mathbf{k}$ has no factors of 2π and a minus sign in the complex exponent; the reverse process gets a $(2\pi)^3$ in the denominator and a plus sign in the complex exponent.

correlation function, i.e., $\langle \widetilde{T}(\mathbf{k}_1)\widetilde{T}(\mathbf{k}_2)^*\rangle$, where[2]

$$\widetilde{T}(\mathbf{k}) = \int d^3r \, e^{-i\mathbf{k}\cdot\mathbf{r}} T(\mathbf{r}). \tag{7}$$

By transforming to Fourier space, we are decomposing our signal into different length scales. Modes with high $k \equiv |\mathbf{k}|$ are made of complex sinusoids that have very short spatial wavelength, so those modes correspond to small scales. The low k modes have long spatial wavelengths, so those modes correspond to large scales.

Just as $\langle T(\mathbf{r}_1)T(\mathbf{r}_2)\rangle$ measures the correlation between two different positions, $\langle \widetilde{T}(\mathbf{k}_1)\widetilde{T}(\mathbf{k}_2)^*\rangle$ measures the correlation between two different Fourier modes. Evaluating this explicitly gives

$$\begin{aligned}
\langle \widetilde{T}(\mathbf{k}_1)\widetilde{T}(\mathbf{k}_2)^*\rangle &= \int d^3r_1 d^3r_2 \, e^{-i\mathbf{k}_1\cdot\mathbf{r}_1} e^{i\mathbf{k}_2\cdot\mathbf{r}_2} \langle T(\mathbf{r}_1)T(\mathbf{r}_2)^*\rangle \\
&= \int d^3r_1 d^3r_2 \, e^{-i\mathbf{k}_1\cdot\mathbf{r}_1} e^{i\mathbf{k}_2\cdot\mathbf{r}_2} \langle T(\mathbf{r}_1)T(\mathbf{r}_2)\rangle \\
&= \int d^3r_1 d^3r_2 \, e^{-i\mathbf{k}_1\cdot\mathbf{r}_1} e^{i\mathbf{k}_2\cdot\mathbf{r}_2} \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} P(\mathbf{k}) \\
&= \int \frac{d^3k}{(2\pi)^3} P(\mathbf{k}) \left(\int d^3r_1 e^{-i(\mathbf{k}_1-\mathbf{k})\cdot\mathbf{r}_1} \right) \left(\int d^3r_2 e^{i(\mathbf{k}_2-\mathbf{k})\cdot\mathbf{r}_2} \right) \\
&= (2\pi)^3 \int d^3k P(\mathbf{k}) \delta^D(\mathbf{k}_1-\mathbf{k}) \delta^D(\mathbf{k}_2-\mathbf{k}),
\end{aligned} \tag{8}$$

where we used the fact that (unlike in Fourier space) the temperature field is real in configuration space to freely take its complex conjugate. We also inserted the relation between the correlation function and the power spectrum. The Dirac delta functions came about because

$$\delta^D(\mathbf{q}) = \frac{1}{(2\pi)^3} \int d^3r \, e^{i\mathbf{q}\cdot\mathbf{r}}. \tag{9}$$

They make the final integral easy to evaluate, and writing out the result nicely, we have

$$\langle \widetilde{T}(\mathbf{k})\widetilde{T}(\mathbf{k}')^*\rangle = (2\pi)^3 \delta^D(\mathbf{k}-\mathbf{k}') P(\mathbf{k}). \tag{10}$$

This is another way to define the power spectrum, and it tells us several things about the behaviour of the field in Fourier space:

- The Dirac delta function on the right hand side of the equation tells us that if $\mathbf{k} \neq \mathbf{k}'$, then $\langle \widetilde{T}(\mathbf{k})\widetilde{T}(\mathbf{k}')^*\rangle = 0$. Since $\langle \widetilde{T}(\mathbf{k})\widetilde{T}(\mathbf{k}')^*\rangle$ quantifies the correlation between $\widetilde{T}(\mathbf{k})$ and $\widetilde{T}(\mathbf{k}')$, our equation implies that different Fourier modes are uncorrelated with one another.

[2] Unless otherwise stated, all integrals range from $-\infty$ to $+\infty$.

- When $\mathbf{k} = \mathbf{k}'$, we have $\langle |\widetilde{T}(\mathbf{k})|^2 \rangle \propto P(\mathbf{k})$ (don't worry about the infinity that comes from the Dirac delta function for now—we'll deal with that later). The power spectrum therefore tells us the amplitude squared (or "power") of the field at various $\mathbf{k}$. Since different k correspond to different length scales, the functional form of the power spectrum tells us whether our field is dominated by large scale features or small features. In Figure 2, we show two simulated universes that were generated from two different power spectra. The universe on the left was generated from a power spectrum that peaks at lower k. It has more power on large scales, and one sees that all the features are coarser than the ones seen on the right.

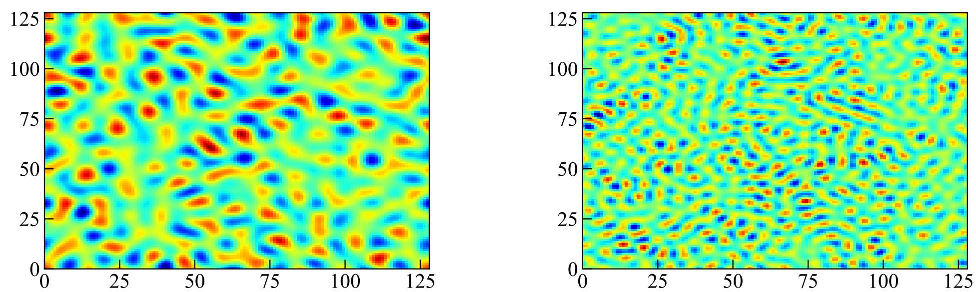

Figure 2: Two example universes generated from different power spectra. The image on the left has a power spectrum that peaks at lower k than the one on the right.

Question: Is the correlation function always positive?

Question: What are the units of $P(\mathbf{k})$?

4 The correlation function and the power spectrum from the perspective of statistics/linear algebra

As we mentioned before, the fields that we measure are random. If one has a random vector $\mathbf{z}$ (i.e., a collection of random numbers grouped into a vector), it is often helpful to describe its statistical properties by computing its *covariance matrix*, defined as

$$\mathbf{C} \equiv \langle \mathbf{z}\mathbf{z}^\dagger \rangle - \langle \mathbf{z} \rangle \langle \mathbf{z} \rangle^\dagger, \tag{11}$$

where the dagger (†) signifies the *adjoint*, i.e., complex conjugate and transpose. Another way to write this is to say that the element found at the ith row and jth column of the matrix is given by

$$C_{ij} = \langle z_i z_j^* \rangle - \langle z_i \rangle \langle z_j^* \rangle. \tag{12}$$

The covariance is the multivariable generalization of the variance. Recall that the variance of a single random variable tells us about the typical spread that one should expect as the random variable fluctuates about its mean. The covariance matrix contains variance information along its diagonal (with each diagonal element storing the variance for each one of the multiple variables), but also additional information about how the fluctuations in one variable are correlated with those in a different variable. For example, if C_{12} is positive, then we expect that if a particular draw of our first random variable z_1 is higher than the mean, then z_2 should also be high. Similarly, if C_{12} were negative, we would expect z_1 and z_2 to fluctuate in opposite directions. Of course, since we are dealing with random processes, exceptions can (and do) occur. But statistically, we expect these patterns to hold.

With this, we can see that we have just been dealing with covariance matrices this whole time. Thinking of our continuous field $T(\mathbf{r})$ as being an infinite dimensional vector, a comparison of Equations (3) and (11) reveals that the correlation function[3] is simply the continuous version of the covariance matrix, because we are assuming that the field has zero mean [and thus the second term of Equation (11) is automatically zero].

Neglecting factors of 2π, the power spectrum simply the covariance matrix written in a Fourier basis. The delta function in Equation (10) is then equivalent to the statement that in the Fourier basis, the covariance matrix is diagonal. In other words, the Fourier basis is the basis that diagonalizes the covariance matrix. That this is the case is the result of the translation invariance of our field. This can be seen by carefully examining the derivations in the previous section. But intuitively, we can understand this simply by thinking about how we would create a field that was *not* translation invariant. Suppose we wanted to create a field where a particular point were special. Perhaps the field has a very large (statistically anomalous) value at a certain location. To create a large value at a particular location, a large number of Fourier modes must conspire to all constructively interfere at that location. In other words, this requires the different modes to be correlated. If modes are uncorrelated (i.e., there are no off-

[3]In what is horribly confusing language, there is also something called a *correlation matrix* $\mathbf{R}$, which is defined as $R_{ij} \equiv C_{ij}/\sqrt{C_{ii}C_{jj}}$. Annoyingly, the discrete version of the correlation *function* is the covariance matrix, not the correlation matrix.

diagonal terms to the Fourier space covariance matrix), then the field's statistics are translation invariant.

The power spectrum and the correlation function in principle contain the same information, since one is just the other written in a different basis. The correlation function doesn't depend on the value of the position coordinates (only position *differences*), but different positions are non-independent (after all, capturing this non-independence is the whole point!). The power spectrum, on the other hand, does depend on the value of the wavenumber "coordinates" k. But the different Fourier modes are independent. Which one we use is often just a matter of convenience:

- In a practical survey of our Universe, the shape of one's survey region is usually complicated. This is often more straightforward to deal with the correlation function than it is with the power spectrum. This is because a single Fourier mode is a complex sinusoid that extends over all space. Thus, complicated boundary conditions affect every Fourier mode, whereas for the correlation function only voxels near the edge of a survey are affected.

- When fluctuations in the field (recall that when we say $T(\mathbf{r})$ we are implicitly referring to *fluctuations* in temperature) are small, the differential equations that govern their evolution can be linearized. In the linear regime, each Fourier mode evolves independently. Theoretically, then, the evolution of the power spectrum is much simpler than the evolution of the correlation function.

5 Estimating a power spectrum

In practice, Equation (10) is a slightly annoying equation for the power spectrum. There are two practical difficulties with it. First, on the left hand side, we have an ensemble average. With real data we do not have the luxury of generating an infinite number of universes in order to take this average. The other difficulty is that we would much rather have an equation where we have $P(k)$ equal to something, rather than having a Dirac delta function that's "stuck" to $P(k)$. We can't divide out the Dirac delta, since it's either 0 of ∞, neither of which are nice to manipulate.

We'll deal with the first problem later. For the second problem, the important thing is to be honest about what we're measuring. In reality, we do not measure an infinite volume. We measure

$$T^{\text{meas}}(\mathbf{r}) = \phi(\mathbf{r})T(\mathbf{r}), \tag{13}$$

where

$$\phi(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \text{ is within the survey volume} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

The power spectrum is a statistical quantity that is defined in Fourier space. Ideally, we would like to compute $\widetilde{T}(\mathbf{k})$ to plug into Equation (10). But to do that we would need to know $T(\mathbf{r})$, and we don't have that. We therefore do the next best thing and compute $\widetilde{T}^{\text{meas}}(\mathbf{k})$ instead, and hope for the best. By the convolution theorem[4] we have

$$\widetilde{T}^{\text{meas}}(\mathbf{k}) = \int \frac{d^3q}{(2\pi)^3} \widetilde{\phi}(\mathbf{k}-\mathbf{q}) \widetilde{T}(\mathbf{q}). \tag{15}$$

Now, examining Equation (10), we see that thanks to the Dirac delta, everything is zero if $\mathbf{k} \neq \mathbf{k}'$. Only when $\mathbf{k}$ and $\mathbf{k}'$ are equal to each other do we get something non-zero to measure, and in that case the left hand side reduces to $\langle |\widetilde{T}(\mathbf{k})|^2 \rangle$. Again, this is not something that we can measure directly. Instead, let's form $\langle |\widetilde{T}^{\text{meas}}(\mathbf{k})|^2 \rangle$. This gives

$$\begin{aligned} \langle |\widetilde{T}^{\text{meas}}(\mathbf{k})|^2 \rangle &= \left\langle \left(\int \frac{d^3q_1}{(2\pi)^3} \widetilde{\phi}(\mathbf{k}-\mathbf{q}_1) \widetilde{T}(\mathbf{q}_1) \right) \left(\int \frac{d^3q_2}{(2\pi)^3} \widetilde{\phi}(\mathbf{k}-\mathbf{q}_2) \widetilde{T}(\mathbf{q}_2) \right)^* \right\rangle \\ &= \int \frac{d^3q_1}{(2\pi)^3} \frac{d^3q_2}{(2\pi)^3} \widetilde{\phi}(\mathbf{k}-\mathbf{q}_1) \widetilde{\phi}^*(\mathbf{k}-\mathbf{q}_2) \langle \widetilde{T}(\mathbf{q}_1) \widetilde{T}^*(\mathbf{q}_2) \rangle \\ &= \int \frac{d^3q_1}{(2\pi)^3} \frac{d^3q_2}{(2\pi)^3} \widetilde{\phi}(\mathbf{k}-\mathbf{q}_1) \widetilde{\phi}^*(\mathbf{k}-\mathbf{q}_2) (2\pi)^3 \delta^D(\mathbf{q}_1 - \mathbf{q}_2) P(\mathbf{q}_1) \\ &= \int \frac{d^3q}{(2\pi)^3} |\widetilde{\phi}(\mathbf{k}-\mathbf{q})|^2 P(\mathbf{q}). \end{aligned} \tag{16}$$

In going from the first line to the second line, we used the fact that there is nothing random about ϕ (and therefore nothing random about $\widetilde{\phi}$). Those quantities are therefore unaffected by the ensemble averaging, and we can bring them inside the integral to act on the quantities that *are* random, i.e., $\widetilde{T}$. Now if our survey covers a fairly large volume, then $\phi(\mathbf{r})$ is a broad function. This means that its Fourier transform $\widetilde{\phi}$ is a very compact function. We should therefore visualize $|\widetilde{\phi}(\mathbf{k}-\mathbf{q})|^2$ as a narrow spike centred on $\mathbf{k} = \mathbf{q}$. Away from $\mathbf{k} = \mathbf{q}$, the function that we are integrating in Equation (16) is negligible. Over the range where $\mathbf{k} \approx \mathbf{q}$, the power spectrum does not change very much. We may therefore approximate it as a constant and factor it out of the integral. This gives

$$\langle |\widetilde{T}^{\text{meas}}(\mathbf{k})|^2 \rangle \approx P(\mathbf{k}) \int \frac{d^3q}{(2\pi)^3} |\widetilde{\phi}(\mathbf{k}-\mathbf{q})|^2 = P(\mathbf{k}) \int d^3r |\phi(\mathbf{r})|^2 = P(\mathbf{k}) V, \tag{17}$$

[4]Please look this up if you're not familiar with it!

where V is the total volume of our survey, and in the second step we used Parseval's Theorem[5]. We can therefore say

$$P(\mathbf{k}) \approx \frac{\langle |\widetilde{T}^{\text{meas}}(\mathbf{k})|^2 \rangle}{V}, \tag{18}$$

which is a much more user-friendly expression for the power spectrum. The power spectrum can therefore be thought of as the, ensemble-averaged, squared magnitude of the Fourier modes, divided by a volume normalization factor. We ended up with this simple expression because $\phi(\mathbf{r})$ took on the form given by Equation (14), but even for a more general case (e.g., perhaps $\phi(\mathbf{r})$ might take on intermediate values between 0 and 1 to signify low—but non-zero—sensitivity to a certain part of the sky) one can simply evaluate the integral $\int d^3 r |\phi(\mathbf{r})|^2$ to get a single number to divide by. Recalling that $T(\mathbf{r})$ is a random field (and therefore so is $\widetilde{T}^{\text{meas}}(\mathbf{k})$), the numerator of our new expression for $P(\mathbf{k})$ is the variance on a particular Fourier mode, bolstering our interpretation of the power spectrum as the variance ("power") of the field on various length scales.

In deriving the above expression, we needed to take an ensemble average. This is not something we can do with real data! What we often do instead is to invoke statistical isotropy. We say that statistical quantities (like the power spectrum) are the same in all directions. This means that $P(\mathbf{k})$ does not depend on the direction of $\mathbf{k}$, but instead, only on its magnitude k. Different Fourier modes that have the same wavenumber k but not the same wavevector $\mathbf{k}$ are therefore different measurements of the same underlying quantity, and we can average their statistics—such as their power spectrum—together. This averaging in direction can act as a proxy for the ensemble average, giving

$$\widehat{P}(k) \approx \frac{\langle |\widetilde{T}^{\text{meas}}(\mathbf{k})|^2 \rangle_{\mathbf{k} \in k}}{V}. \tag{19}$$

Note that the left hand side has now acquired a hat ($\widehat{\ }$), to signify that it is just an *estimator* of the power spectrum, not the true power spectrum. This estimate will have error bars associated with it, in a way that the true $P(k)$ does not. The extent to which $\widehat{P}(k)$ is a good approximation for $P(k)$ is determined by how well the spatial average approximates the ensemble average. Again invoking the analogy of coin flips, $P(k)$ is analogous to the true variance of the outcomes, obtained from an infinite number of coin flips. It is not a random quantity, even though the outcome of each individual coin flip is random (just as $\widetilde{T}(\mathbf{k})$ is random). The estimator $\widehat{P}(k)$ is analogous to an estimated variance from a finite number of flips.

[5]Another thing to look up if you haven't encountered this before!

It has an error associated with it, but converges to the true $P(k)$ as the number of samples approaches infinity.

Question: Is the error on $\widehat{P}(k)$ likely to be smaller for low k or high k? Thinking about this question will provide intuition for a phenomenon known as *cosmic variance.*

Exercise: Equation (19) provides a practical recipe for computing the power spectrum given a box of data:

1. Fourier transform the box (using, say, an off-the-shelf routine like `np.fft`).
2. Compute the squared magnitude of each mode.
3. Average the results over shells of constant k.
4. Divide by the volume V.

Code this up! To make things easier to visualize, it might be easier to start with a 2D universe rather than a 3D one. This will also take some careful thinking regarding the discretization of the sky. If you want to tackle this head on in a rigorous manner, do the next exercise first. But if might be worthwhile to pursue the two simultaneously.

Exercise: In practice $T^{\rm meas}$ isn't really the measured temperature field, because measurements are necessarily discrete—we measure T at a series of discrete voxel locations. We can capture this by saying we measure

$$T_i = \int d^3 r \phi_i(\mathbf{r}) T(\mathbf{r}), \tag{20}$$

where

$$\phi_i(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \text{ is within the } i^{\rm th} \text{ voxel} \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

With a discrete set of voxels, we will need to perform a discrete Fourier transform rather than a continuous one. For example, we might define $\widetilde{T}_j$ to be the value of the $j^{\rm th}$ (Fourier space) voxel after having put $\{T_i\}$ through `np.fft`. Write down the equivalent of Equation (19) for an estimator of the power spectrum in terms of $\{T_i\}$.

Exercise: So far we have focused on measuring the power spectrum given some data. Also useful is the reverse process: the simulation of some mock data given some theoretical power spectrum. Since our Universe is random, this mock data is not unique. What we are generating is a *realization* that is consistent with the specified statistical properties. Roughly speaking, one generates a realization from $P(k)$ by doing the following:

1. Initialize a grid in Fourier space.
2. For each grid point in Fourier space, generate a random complex number drawn from a Gaussian distribution with mean zero.
3. *Mystery step*
4. Scale each random number appropriately according to $P(k)$.
5. Fourier transform the box to position space.

This is how Figure 2 was generated. Code up your own version. One thing you'll have to figure out is how to make sure that the final mock data is *real*. In general, if I put some random numbers on a grid and I Fourier transform them, I will end up with complex numbers. To make sure my answer is real, I need to place some restrictions on the random numbers that I generate. This is what the "Mystery Step" is for, and you'll need to figure out what that is.

References

[1] Daniel J. Eisenstein, Idit Zehavi, David W. Hogg, Roman Scoccimarro, Michael R. Blanton, Robert C. Nichol, Ryan Scranton, Hee-Jong Seo, Max Tegmark, Zheng Zheng, Scott F. Anderson, Jim Annis, Neta Bahcall, Jon Brinkmann, Scott Burles, Francisco J. Castander, Andrew Connolly, Istvan Csabai, Mamoru Doi, Masataka Fukugita, Joshua A. Frieman, Karl Glazebrook, James E. Gunn, John S. Hendry, Gregory Hennessy, Zeljko Ivezić, Stephen Kent, Gillian R. Knapp, Huan Lin, Yeong-Shang Loh, Robert H. Lupton, Bruce Margon, Timothy A. McKay, Avery Meiksin, Jeffery A. Munn, Adrian Pope, Michael W. Richmond, David Schlegel, Donald P. Schneider, Kazuhiro Shimasaku, Christopher Stoughton, Michael A. Strauss, Mark SubbaRao, Alexander S. Szalay, István Szapudi, Douglas L. Tucker, Brian Yanny, and Donald G. York. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *Astrophysical Journal*, 633:560–574, November 2005.