Let us look at the Fourier-space correlation of a *stationary* function $f(x)$. By stationary, we mean that the properties do not depend on time. In particular:

$$\langle f(x)f(x+\delta)\rangle = g(\delta)$$

for some function g. With just this assumption, we can say quite a lot about the properties of the Fourier transform of f. We also note that the quantity g is known as the *correlation function* of our original function f.

Of course, one can work with either the continuous or discrete forms of the Fourier transform. This note will use the discrete form (which is what one uses in practice in data analysis) to enhance your familiarity with it. Recall the fundamental definition of the discrete Fourier transform (DFT):

$$F(k) = \sum_{x=0}^{N-1} f(x)\exp(-2\pi ixk/N)$$

For simplicity, we will assume that f is strictly real, as most functions that arise in data analysis are. With this form, we can now write down the covariance of two different modes in the DFT:

$$\left\langle F(k)F^\dagger(k')\right\rangle = \left\langle \sum \exp(-2\pi ikx/N)f(x)\sum \exp(2\pi ik'x'/N)f(x')\right\rangle$$

$$= \left\langle \sum\sum \exp(-2\pi ikx/N)\exp(2\pi ik'x'/N)f(x)f(x')\right\rangle$$

We can set $x' \equiv x + \delta_x$, and similarly $k' \equiv k + \delta_k$ and re-write a bit more neatly. Since the sums go over all data points, summing over x' and summing over δ_x are equivalent. This leaves us with

$$\left\langle \sum\sum \exp(2\pi i(kx + k\delta_x + \delta_k x + \delta_k\delta_x - kx)/N)f(x)f(x+\delta_x)\right\rangle$$

Now, regroups and make the interior sum over x and the exterior sum over δ_x. This gives:

$$\left\langle \sum_{\delta_x}\exp(2\pi i(k\delta_x + \delta_k\delta_x)/N)\sum_x \exp(2\pi i\delta_k x/N)f(x)f(x+\delta_x)\right\rangle$$

Now, by our assumption of stationarity, we know that the expectation of $f(x)f(x+\delta_x) = g(\delta_x)$, so the interior sum becomes

$$\sum_x \exp(2\pi i\delta_k x/N)g(\delta_x)$$

Since g is constant, the sum of it against any sinusoid is zero, unless the period of the sinusoid is zero. So, we know that δ_k must be equal to zero for the expectation to not vanish. In other words, even if our function has a quite complicated correlation structure, as long as that structure is time invariant, the individual Fourier modes that make up a realization of that function are independent. So, it becomes much easier to work with stationary functions by taking their Fourier transforms.

We can also take this one step further to derive the Wiener-Khinchin theorem. Since we know that $\delta_k = 0$, then the interior sum just becomes

$$\sum_x g(\delta_x) = Ng(\delta_x)$$

Inserting that into the full sum, we have

$$\langle |F(k)|^2 \rangle = \left\langle N \sum_{\delta_x} \exp(2\pi i k \delta_x / N) g(\delta_x) \right\rangle$$

The quantity on the left is just the *power spectrum* of our function (this is pretty much the definition of the power spectrum). The quantity on the right is now just the DFT of $g(\delta_x)$, the correlation function. In words, the Wiener-Khinchin theorem states that *the Fourier transform of the correlation function is the power spectrum.* This result is actually quite a bit more general than we have shown here. For instance, on the sphere, the power spectrum C_l is the Legendre transform of the angular correlation function $C(\theta)$.

A few further points are in order. Note that we only required that f be stationary - in particular, it did not have to be Gaussian. If f *is* Gaussian, then either the power spectrum or correlation function is a complete description of the properties of f. In the power spectrum description, this means that the amplitudes of $F(k)$ are Gaussian distributed, with random phases (technically, the real and imaginary amplitudes are independent and Gaussian distributed).

The Wiener-Khinchin theorem also provides a useful way of estimating the power spectrum in the presence of cuts/missing data. This is an extremely common situation in practice. One example is measuring the power spectrum of the CMB - there are foreground signals (galactic dust, extragalactic point sources, etc.) that contaminate the CMB signal in localized patches of the sky that we wish to exclude from the analysis. The publicly available (Pol)Spice package estimates $C(\theta)$ using only the uncut regions of the sky and transforms it to estimate $C(l)$. Of course, the missing regions introduce correlations between the different l modes (why?) but that is unavoidable on the a cut sky.