Notes on Fourier Analysis

ASTP 611-01: Statistical Methods for Astrophysics*

Spring Semester 2014

Contents

Wednesday, January 29, 2014

0 Preliminaries

0.1 Outline

- **Part Zero:** Fourier Analysis (Gregory Appendix B; Numerical Recipes, Chapters 12-13, Arfken, Weber & Harris, Chapter 20)
  1. Continuous and discrete Fourier transforms
  2. Spectral analysis of random data
- **Part One:** Probability Theory (Gregory Chapters 1, 2, 5)
  1. Fundamentals of probability theory (Bayesian and Frequentist interpretations) (Gregory Ch 1 & 2)
  2. Definitions and applications of probability distributions. (Gregory Ch 5)
  3. Specific probability distributions (Gregory Ch 5)

4. Gaussian approximation to probability distributions (including multivariate Gaussian and the Fisher matrix) (Gregory Ch 5; Sivia Ch)
5. Sums of random variables and the central limit theorem (Gregory Ch 5)

- **Part Two:** Statistical Inference (Gregory Chapters 3, 6, 7)
   1. Model selection, hypothesis testing, and parameter estimation: overview. Philosophy in the frequentist and Bayesian interpretations (Gregory Ch 6 & 7)
   2. Frequentist inference (chi-squared, confidence intervals, upper limits, composite hypotheses) (Gregory Ch 6 & 7)
   3. Bayesian inference (Bayes factor, plausible intervals, marginalization; connection between Bayesian and frequentist results) (Gregory Ch 3; Sivia Ch 2, 3 & 4)

1 Fourier Analysis

See Gregory, Appendix B, Numerical Recipes, Chapters 12-13, or Arfken, Weber & Harris, Chapter 20

1.1 Continuous Fourier Transform

You should be familiar[1] with the Fourier series for a function $h(t)$ defined on an interval

$$-\frac{T}{2} \leq t \leq \frac{T}{2} \tag{1.1}$$

namely[2]

$$h(t) = \sum_{n=-\infty}^{\infty} c_n \exp\left(\frac{i2\pi nt}{T}\right) \tag{1.2}$$

where the Fourier coëfficients are given by

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} h(t) \exp\left(-\frac{i2\pi nt}{T}\right) dt \tag{1.3}$$

Note that if we write

$$f_n = \frac{n}{T} = n\,\delta f \tag{1.4}$$

so that

$$h(t) = \sum_{n=-\infty}^{\infty} c_n\, e^{i2\pi f_n t} \tag{1.5}$$

the frequencies are spaced closer together for larger T. If we write

$$h(t) = \sum_{n=-\infty}^{\infty} \underbrace{T c_n}_{\widetilde{h}(f_n)} e^{i2\pi f_n t}\,\delta f \tag{1.6}$$

and take the limit as $T \to \infty$ so that the sum becomes an integral, we get

$$h(t) = \int_{-\infty}^{\infty} \widetilde{h}(f)\, e^{i2\pi f t}\, df \tag{1.7}$$

and the inverse

$$\widetilde{h}(f_n) = T c_n = \int_{-T/2}^{T/2} h(t)\, e^{-i2\pi f_n t}\, dt \tag{1.8}$$

[1] If you are unfamiliar, or a little rusty, with this, you should work through the exercises on Fourier series

[2] We'll be working almost exclusively with Fourier analysis based on complex exponentials; to relate this to trigonometric Fourier analysis, you can use the Euler relation $e^{i\theta} = \cos\theta + i\sin\theta$ and the related equations $e^{-i\theta} = \cos\theta - i\sin\theta$, $\cos\theta = (e^{i\theta} + e^{-i\theta})/2$, and $\sin\theta = (e^{i\theta} - e^{-i\theta})/(2i)$.

becomes

$$\widetilde{h}(f) = \int_{-\infty}^{\infty} h(t)\, e^{-i2\pi f t}\, dt \tag{1.9}$$

Note that the orthogonality relation

$$\int_{-T/2}^{T/2} e^{i2\pi(n_2-n_1)t/T}\, dt = T\,\delta_{n_1 n_2} = \begin{cases} T & \text{if } n_1 = n_2 \\ 0 & \text{if } n_1 \neq n_2 \end{cases} , \tag{1.10}$$

in terms of the Kronecker delta $\delta_{n_1 n_2}$, becomes, in the limit of infinite T,

$$\int_{-\infty}^{\infty} e^{i2\pi(f_2-f_1)t}\, dt = \delta(f_2 - f_1) \tag{1.11}$$

where $\delta(f_2 - f_1)$ is the Dirac delta function defined by

$$\int_{-\infty}^{\infty} \delta(f_2 - f_1)\, H(f_1)\, df_1 = H(f_2) \ ; \tag{1.12}$$

we can check that we got the normalization right by noting

$$\sum_{n_1=-\infty}^{\infty} T\,\delta_{n_1 n_2}\,\delta f = \sum_{n_1=-\infty}^{\infty} \delta_{n_1 n_2} = 1 \tag{1.13}$$

1.1.1 Convolution

A common physical situation is for one quantity, as a function of time, to be linearly related to another quantity, which we can write as:

$$g(t) = \int_{-\infty}^{\infty} A(t,t')\, h(t')\, dt' \ . \tag{1.14}$$

If the mapping of $h(t)$ onto $g(t)$ is *stationary*, e.g., doesn't depend on any time-dependent external factors, it can be written even more simply:

$$g(t) = \int_{-\infty}^{\infty} A(t-t')\, h(t')\, dt' \ . \tag{1.15}$$

This relationship is known as a *convolution* and is sometimes written $g = A * h$. It can be written even more simply if we substitute in the form of $A(t-t')$ and $h(t')$ in terms of their Fourier transforms:

$$h(t') = \int_{-\infty}^{\infty} \widetilde{h}(f)\, e^{i2\pi f t'}\, df \tag{1.16}$$

$$A(t-t') = \int_{-\infty}^{\infty} \widetilde{A}(f')\, e^{i2\pi f'(t-t')}\, df' \tag{1.17}$$

(where we have used different names for the two frequency integration variables so we don't mix them up) to get

$$\begin{aligned} g(t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widetilde{A}(f')\,\widetilde{h}(f)\, e^{i2\pi[f'(t-t')+ft']}\, df\, df'\, dt' \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widetilde{A}(f')\,\widetilde{h}(f)\, e^{i2\pi f' t} \int_{-\infty}^{\infty} e^{i2\pi(f-f')t'}\, dt'\, df'\, df \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widetilde{A}(f')\,\widetilde{h}(f)\, e^{i2\pi f' t}\, \delta(f-f')\, df'\, df \\ &= \int_{-\infty}^{\infty} \widetilde{A}(f)\,\widetilde{h}(f)\, e^{i2\pi f t}\, df \end{aligned} \tag{1.18}$$

which means that

$$\widetilde{g}(f) = \widetilde{A}(f)\,\widetilde{h}(f) \ , \tag{1.19}$$

i.e., convolution in the time domain is equivalent to multiplication in the frequency domain.

1.1.2 Properties of the Fourier Transform

Here are a number of important and useful properties obeyed by Fourier transforms, and handy Fourier transforms of specific functions:

- If $h(t)$ is real, then $\widetilde{h}(-f) = \widetilde{h}^*(f)$.

- If $h(t) = h_0$, a constant, then
$$\widetilde{h}(f) = \int_{-\infty}^{\infty} h_0 e^{-i2\pi f t} dt = h_0 \delta(f) \tag{1.20}$$
- If $h(t) = h_0\delta(t - t_0)$, then $\widetilde{h}(f) = h_0 e^{-i2\pi f t_0}$
- If $h(t)$ is a square wave
$$h(t) = \begin{cases} h_0 & \frac{-\tau}{2} < t < \frac{\tau}{2} \\ 0 & |t| > \frac{T}{2} \end{cases} \tag{1.21}$$
then
$$\widetilde{h}(f) = h_0 \frac{2\pi f \tau}{\pi f} = 2h_0 \tau \operatorname{sinc} 2f\tau \tag{1.22}$$
where
$$\operatorname{sinc} x = \frac{\sin \pi x}{\pi x} \tag{1.23}$$
is the normalized sinc function.
- If $h(t)$ is a Gaussian
$$h(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2\sigma^2} \tag{1.24}$$
then its Fourier transform is also a Gaussian:
$$\widetilde{h}(f) = e^{-(2\pi f)^2/2\sigma^{-2}} \tag{1.25}$$
Note that the narrower the Gaussian is in the time domain, the wider the corresponding Gaussian is in the frequency domain. This is related to the Heisenberg uncertainty principle.
- Dimensionally, the units of $\widetilde{h}(f)$ are the units of $h(t)$ times time (or divided by frequency). We'll usually say, e.g., if $h(t)$ has units of "gertrudes", $\widetilde{h}(f)$ has units of "gertrudes per hertz".
-
$$\int_{-\infty}^{\infty} h^*(t)\, g(t)\, dt = \int_{-\infty}^{\infty} \widetilde{h}^*(f)\, \widetilde{g}(f)\, df \tag{1.26}$$

It's a useful exercise, and not too hard, to demonstrate each of these.

Thursday, January 30, 2014

1.2 Discrete Fourier Transform

Recall that an ordinary Fourier series could be written in the form (1.6) relating a finite-duration $h(t)$ to its Fourier components $\widetilde{h}(f_n)$, with inverse relationship (1.8). The time variable t is continuously-defined with finite duration, while the frequency f_n takes on only a discrete set of values, but ranges from $-\infty$ to ∞. This situation is summarized as follows:

	Resolution	Extent
t	continuous	duration T
f	discrete, $\delta f = \frac{1}{T}$	infinite

When we took the limit as $T \to \infty$ to define the inverse Fourier transform (1.7) and the Fourier transform (1.9) we ended up with both frequency and time being continuously defined from $-\infty$ to ∞:

	Resolution	Extent
t	continuous	infinite
f	continuous	infinite

In an experimental situation, on the other hand, not only is the duration finite, but the time is also discretely sampled. Consider the simplest case of N samples separated by a fixed sampling time of δt so that the total duration is $T = N\,\delta t$:
$$h_j = h(t_j) = h(t_0 + j\,\delta t) \qquad j = 0, 1, \ldots, N-1 \tag{1.27}$$

we'd like to define the Fourier transform[3]

$$\widetilde{h}(f_k) = \int_{t_0}^{t_o+T} h(t)\, e^{-i2\pi f_k (t-t_0)}\, dt \tag{1.28}$$

but we don't have access to the full function $h(t)$, only the discrete samples $\{h_j\}$. The best we can do, then, is approximate the integral by a sum and see what we get:

$$\sum_{j=0}^{N-1} h_j\, e^{-i2\pi f_k (t_j - t_0)}\, \delta t = \sum_{j=0}^{N-1} h_j\, e^{-i2\pi (k\,\delta f)(j\,\delta t)}\, \delta t = \sum_{j=0}^{N-1} h_j\, e^{-i2\pi jk/N}\, \delta t \tag{1.29}$$

where in the last step we've used the fact that

$$\delta f\, \delta t = \frac{\delta t}{T} = \frac{1}{N} \tag{1.30}$$

Now, (1.29) is the discrete approximation to the Fourier transform, so we could call it something like $\widetilde{h}_k$. But if you're a computer manipulating a set of numbers $\{h_j\}$, you don't really need to know the physical sampling rate, except for the factor of δt in (1.29). So the standard definition of the discrete Fourier transform leaves this out:

$$\widehat{h}_k = \sum_{j=0}^{N-1} h_j\, e^{-i2\pi jk/N} \tag{1.31}$$

[3] Note that for most real-world time-series data, all times (for example GPS time–written as the number of seconds since 00:00:00 UTC on January 6, 1980–or Modified Julian Date) are measured relative to some reference point, so $t = 0$ doesn't have any absolute meaning, and different ways of writing the time have different zeros, just like Fahrenheit and Celsius temperatures do. So really expressions like (1.7) and (1.9) should have $t - t_0$ rather than t in the exponential, where t_0 is some reference time to be specified. This is a little subtle when it comes to the issue of convolution, though, since the $A(t - t')$ appearing in (1.15) is a time difference rather than a time, and thus $A(0)$ is meaningful, and the inverse Fourier transform (1.17) would not need to refer to a t_0, even if (1.16) did.

In principle (1.31) can be used to define the discrete Fourier transform for any integer k. However, we can see that not all of the $\widehat{h}_k$ are independent; in particular,

$$\widehat{h}_{k+N} = \sum_{j=0}^{N-1} h_j\, e^{-i2\pi jk/N}\, e^{-i2\pi j} = \sum_{j=0}^{N-1} h_j\, e^{-i2\pi jk/N} = \widehat{h}_k \tag{1.32}$$

where we have used the fact that

$$e^{-i2\pi j} = \cos 2\pi j - i \sin 2\pi j = 1 \tag{1.33}$$

since j is an integer. This means there are only N independent $\widehat{h}_k$ values, which is not surprising, since we started with N samples $\{h_j\}$. One choice is to let k go from 0 to $N-1$, and we can use that to calculate the inverse transform by starting with

$$\sum_{k=0}^{N-1} \widehat{h}_k\, e^{i2\pi jk/N} = \sum_{k=0}^{N-1} \sum_{\ell=0}^{N-1} h_\ell\, e^{i2\pi (j-\ell)k/N} \tag{1.34}$$

If we recall that

$$1 - a^N = (1-a)\,(1 + a + a^2 + \ldots + a^{N-1}) \tag{1.35}$$

we can see that

$$\begin{aligned} \sum_{k=0}^{N-1} e^{i2\pi (j-\ell)k/N} &= \sum_{k=0}^{N-1} \left(e^{i2\pi (j-\ell)/N}\right)^k \\ &= \begin{cases} N & \text{if } j = \ell \bmod N \\ \frac{1-e^{i2\pi(j-\ell)}}{1-e^{i2\pi(j-\ell)/N}} = 0 & \text{if } j \neq \ell \bmod N \end{cases} \end{aligned} \tag{1.36}$$

i.e.,

$$\sum_{k=0}^{N-1} e^{i2\pi (j-\ell)k/N} = N\, \delta_{j,\ell \bmod N} \tag{1.37}$$

so

$$\sum_{k=0}^{N-1} \widehat{h}_k\, e^{i2\pi jk/N} = \sum_{\ell=0}^{N-1} h_\ell\, N\, \delta_{j,\ell \bmod N} = N h_j \tag{1.38}$$

and the inverse transform is

$$h_j = \frac{1}{N} \sum_{k=0}^{N-1} \widehat{h}_k\, e^{i2\pi jk/N} \tag{1.39}$$

Note that the asymmetry between the forward and reverse transform arose because we left out the factor of δt from (1.31); if we write

$$(\widehat{h}_k\, \delta t) = \sum_{j=0}^{N-1} h_j\, e^{-i2\pi jk/N}\, \delta t \tag{1.40}$$

then the inverse transform is

$$h_j = \sum_{k=0}^{N-1} (\widehat{h}_k\, \delta t)\, e^{i2\pi jk/N}\, \frac{1}{N\, \delta t} = \sum_{k=0}^{N-1} (\widehat{h}_k\, \delta t)\, e^{i2\pi jk/N}\, \delta f \tag{1.41}$$

which restores the notational symmetry of the continuous Fourier transform.

1.2.1 Nyquist Frequency and Aliasing

In this discussion we'll assume the number of samples N is even; the generalization to odd N is straightforward.

We saw above that if you take the discrete Fourier transform of N data points $\{h_j\}$, the periodicity $\widehat{h}_{k+N} = \widehat{h}_k$ means that only N of the Fourier components are independent. We implicitly considered those to be $\{\widehat{h}_k | k = 0, 1, \ldots, N-1\}$, which is certainly convenient if you're a computer, but it doesn't really make the most physical sense.

For example, consider the behavior of the discrete Fourier transform if the original time series is real, so that $h_j^* = h_j$.

$$\text{if } h_j^* = h_j, \qquad \widehat{h}_k^* = \left(\sum_{j=0}^{N-1} h_j\, e^{-i2\pi jk/N} \right)^* = \sum_{j=0}^{N-1} h_j\, e^{i2\pi jk/N} = \widehat{h}_{-k} = \widehat{h}_{N-k} \tag{1.42}$$

If we confine our attention to $0 \le k \le N-1$, the appropriate symmetry relation is $\widehat{h}_{N-k} = \widehat{h}_k^*$, which means the second half of the list of Fourier components is determined by the first. But this seems a little bit removed from the corresponding symmetry property $\widetilde{h}(-f) = \widetilde{h}^*(f)$ from the continuous Fourier transform.

To better keep positive and negative frequencies together, we'd like to consider the physically interesting set of N Fourier components to be

$$\left\{ \widehat{h}_k \,\middle|\, k = -\frac{N}{2}, \ldots, \frac{N}{2} - 1 \right\} . \tag{1.43}$$

It's a matter of convention that we include $-N/2$ rather than $N/2$ in the list. It makes things more convenient for `fftshift()` functions in SciPy, matlab, etc., which move the Fourier components $\{\widehat{h}_{N/2}, \ldots, \widehat{h}_{N-1}\}$ to the front of a vector so they can represent $\{\widehat{h}_{-N/2}, \ldots, \widehat{h}_{-1}\}$.

Note now that the reality condition becomes

$$\text{if } h_j^* = h_j, \qquad \widehat{h}_k^* = \widehat{h}_{-k} \tag{1.44}$$

which means that all of the negative components $\{\widehat{h}_{-N/2+1}, \ldots, \widehat{h}_{-1}\}$ of the DFT of a real series are just the complex conjugates of the corresponding positive components. The reality condition also enforces

$$\widehat{h}_0 = \widehat{h}_0^* \in \mathbb{R} \tag{1.45a}$$

$$\widehat{h}_{-N/2} = \widehat{h}_{N/2}^* = \widehat{h}_{-N/2}^* \in \mathbb{R} \tag{1.45b}$$

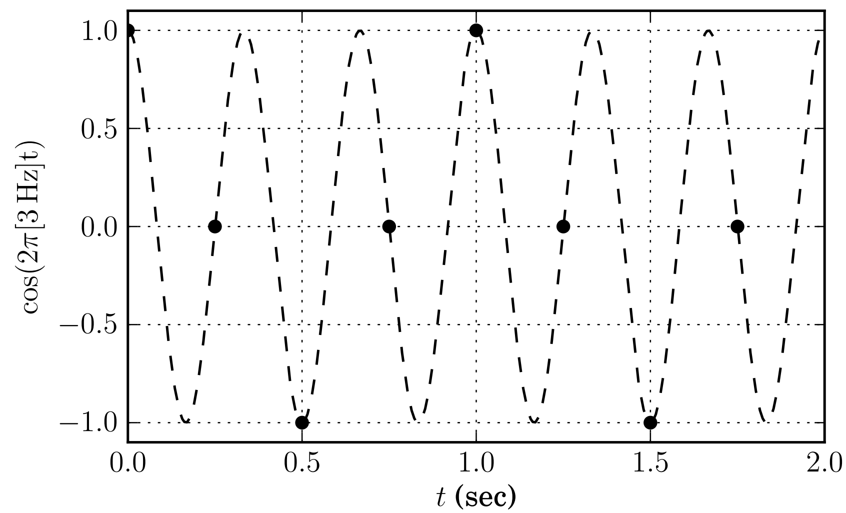
So from N real samples $\{h_j | j = 0, \ldots, N-1\}$ we get a discrete Fourier transform completely described by 2 real components $\widehat{h}_0$ and $\widehat{h}_{-N/2} = \widehat{h}_{N/2}$ and $\frac{N}{2} - 1$ complex components $\{\widehat{h}_k | k = 1, \ldots, \frac{N}{2} - 1\}$.

The frequency corresponding to the last Fourier component,

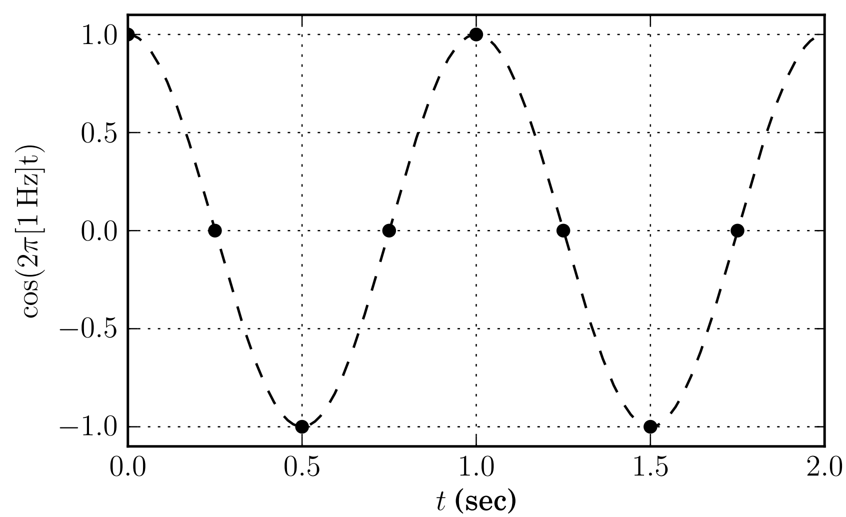
$$\left|f_{-N/2}\right| = \left|f_{N/2}\right| = \frac{N}{2}\delta f = \frac{1}{2\,\delta t} \tag{1.46}$$

is half of the sampling frequency $1/\delta t$, and is known as the *Nyquist frequency*.[4] It is the highest frequency which can be resolved in the discrete Fourier transform of a series sampled with a sampling time δt. Of course, as we've seen, frequencies above the Nyquist frequency, which correspond to Fourier components with $k > N/2$, aren't invisible, they just show up in the same place as lower frequencies. For example, consider a cosine wave with a frequency of 3 Hz, sampled with a time step of $\delta t = 0.25$ sec:

[4]There's an unfortunate bit of linguistic confusion. If you know the sampling frequency $1/\delta t$, then the Nyquist frequency $1/(2\delta t)$ is the highest independent frequency in your Fourier transform. On the other hand, if you're trying to discretely sample a continuous time series which is *band-limited* with some highest frequency so that $\widetilde{h}(f)$ vanishes for $|f| > f_c$, then the lowest sample rate $1/\delta t$ that you can use to avoid aliasing is $2f_c$; if $1\delta t > 2f_c$ then $f_c < 1/(2\delta t)$. Somewhat confusingly, this is called the *Nyquist rate*.



If we just look at the dots, they don't look like a 3 Hz cosine wave, but rather like one with a frequency of 1 Hz. And indeed, we'd get the exact same samples if we sampled a 1 Hz at the same rate:



This is because $f = 3\,\mathrm{Hz}$ is above the Nyquist frequency at this sampling rate, which is $f_{\mathrm{Ny}} = 2\,\mathrm{Hz}$. The higher-frequency cosine wave has been *aliased* down to a frequency of $f_{\mathrm{Ny}} - f = -1\,\mathrm{Hz}$.

Note, in closing, that the range of independent frequencies, from $-f_{\mathrm{Ny}}$ to $+f_{\mathrm{Ny}}$, is $2\,|f_{\mathrm{Ny}}| = \frac{1}{\delta t}$ so we can fill in the table for time-frequency resolution and extent:

	Resolution	Extent
t	discrete, δt	duration T
f	discrete, $\delta f = \frac{1}{T}$	finite, $2\,|f_{\mathrm{Ny}}| = \frac{1}{\delta t}$

Tuesday, February 4, 2014

2 Spectral Analysis of Random Data

2.1 Amplitude Spectrum

Given a real time series $h(t)$ we know how to construct its Fourier transform

$$\widetilde{h}(f) = \int_{-\infty}^{\infty} h(t)\, e^{-i2\pi ft}\, dt \tag{2.1}$$

or the equivalent discrete Fourier transform from a set of samples $\{h_j | j = 0, \ldots, N-1\}$:

$$\widehat{h}_k = \sum_{j=0}^{N-1} h_j\, e^{-i2\pi jk/N} \tag{2.2}$$

where $\widehat{h}_k\, \delta t \sim \widetilde{h}(f_k)$.

Think about the physical meaning of $\widetilde{h}(f)$, by breaking this complex number up into an amplitude and phase

$$\widetilde{h}(f) = A(f)\, e^{i\phi(f)}\ . \tag{2.3}$$

If $h(t)$ is real, the condition $\widetilde{h}(-f) = \widetilde{h}^*(f) = A(f)e^{-i\phi(f)}$ means $A(-f) = A(f)$ and $\phi(-f) = -\phi(f)$. Thus we can write the inverse Fourier transform as

$$\begin{aligned} h(t) &= \int_{-\infty}^{0} A(f)\, e^{i[2\pi ft+\phi(f)]}\, df + \int_{0}^{\infty} A(f)\, e^{i[2\pi ft+\phi(f)]}\, df \\ &= \int_{0}^{\infty} A(f)\, \left(e^{i[2\pi ft+\phi(f)]} + e^{-i[2\pi ft+\phi(f)]}\right)\, df \\ &= \int_{0}^{\infty} 2\,A(f) \cos\left(2\pi ft + \phi(f)\right)\, df \end{aligned} \tag{2.4}$$

So $A(f)$ is a measure of the amplitude at frequency f and $\phi(f)$ is the phase.

Note also that

$$\int_{-\infty}^{\infty} \big(h(t)\big)^2\, dt = \int_{-\infty}^{\infty} |\widetilde{h}(f)|^2\, df \tag{2.5}$$

This works pretty well if the properties of $h(t)$ are deterministic. But suppose $h(t)$ is modelled as random, i.e., it depends on a lot of factors we don't know about, and all we can really do is make statistical statements. What is a sensible description of the spectral content of h?

2.2 Random Variables

Consider a *random variable* x. Its value is not known, but we can talk about statistical expectations as to its value. We will have a lot more to say about this soon, but for now imagine we have a lot of chances to measure x in an ensemble of identically-prepared systems. The hypothetical average of all of those imagined measurements is called the *expectation value* and we write it as $\langle x\rangle$. (Another notation is $E[x]$.) We can also take some known function f and talk about the expectation value $\langle f(x)\rangle$

corresponding to a large number of hypothetical measurements of $f(x)$.

The expectation value of x itself is the *mean*, sometimes abbreviated as μ. (This name is taken by analogy to the mean of an actual finite ensemble.) Since x is random, though, x won't always have the value $\langle x\rangle$. We can talk about how far off x typically is from its average value $\langle x\rangle$. Note however that

$$\langle x-\mu\rangle = \langle x\rangle - \langle\mu\rangle = \langle x\rangle - \mu = 0 \tag{2.6}$$

since the expectation value is a linear operation (being the analogue of an average), and the expectation value of a non-random quantity is just the quantity itself. So instead of the mean deviation from the mean, we need to consider the mean square deviation from the mean

$$\left\langle (x-\mu)^2\right\rangle \tag{2.7}$$

This is called the variance, and is sometimes written σ^2. Its square root, the root mean square (RMS) deviation from the mean, is the standard deviation.

2.2.1 Random Sequences

Now imagine we have a bunch of random variables $\{x_j\}$; in principle each can have its own mean $\mu_j = \langle x_j\rangle$ and variance $\sigma_j^2 = \langle (x_j-\mu_j)^2\rangle$. But we can also think about possible correlations ${\sigma^2}_{j\ell} = \langle (x_j-\mu_j)(x_\ell-\mu_\ell)\rangle$; if the variables are uncorrelated ${\sigma^2}_{j\ell} = \delta_{j\ell}\,\sigma_j^2$, but that need not be the case.

Think specifically about a series of N samples which are all uncorrelated random variables with zero mean and the same variance:

$$\langle x_j\rangle = 0; \qquad \langle x_j x_\ell\rangle = \delta_{j\ell}\,\sigma^2 \; ; \tag{2.8}$$

what are the characteristics of the discrete Fourier transform of this sequence?

$$\widehat{x}_k = \sum_{j=0}^{N-1} x_j\, e^{-i2\pi jk/N} \tag{2.9}$$

Well,

$$\langle \widehat{x}_k\rangle = \sum_{j=0}^{N-1} \overbrace{\langle x_j\rangle}^{0} e^{-i2\pi jk/N} = 0 \tag{2.10}$$

and

$$\begin{aligned}\langle \widehat{x}_k \widehat{x}_\ell^*\rangle &= \sum_{j=0}^{N-1}\sum_{n=0}^{N-1} \langle x_j x_n\rangle\, e^{-i2\pi(jk-n\ell)/N} = \sum_{j=0}^{N-1} \sigma^2 e^{-i2\pi j(k-\ell)/N} \\ &= N\sigma^2 \delta_{k,\ell \bmod N}\end{aligned} \tag{2.11}$$

so the Fourier components are also uncorrelated random variables with variance

$$\left\langle |\widehat{x}_k|^2\right\rangle = N\sigma^2 \tag{2.12}$$

Note this is independent of k, which is maybe a bit surprising. After all, it's natural to think all times are alike, but all frequencies need not be. Random data like this, which is the same at all frequencies, is called "white noise". To gain some more insight into white and colored noise, it helps to think about the same thing in the idealization of the continuous Fourier transform.

2.3 Continuous Random Data

Think about the continuous-time analog to the random sequence considered in section 2.2.1. This is a time series $x(t)$ which is characterized by statistical expectations. In particular we could talk about its expectation value

$$\langle x(t)\rangle = \mu(t) \tag{2.13}$$

and the expectation value of the product of samples taken at possibly different times

$$\langle x(t)x(t')\rangle = K_x(t,t') \tag{2.14}$$

(it's conventional not to subtract out $\mu(t)$ here). Note that $\langle x(t)\rangle$ is *not* the time-average of a particular instantiation of $x(t)$, although the latter may sometimes be used to estimate the former.

2.3.1 White Noise

Now, for white noise we need the continuous-time equivalent of (2.8), in which the data is uncorrelated with itself except at the very same time. In the case of continuous time, the sensible thing is the Dirac delta function, so white noise is characterized by

$$\langle x(t)x(t')\rangle = K_0\,\delta(t-t') \tag{2.15}$$

where K_0 is a measure of how "loud" the noise is. In the frequency domain, this means

$$\begin{aligned}\langle \widetilde{x}(f)\widetilde{x}^*(f')\rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\langle x(t)x(t')\rangle\; e^{-i2\pi(ft-f't')}\,dt\,dt' \\ &= K_0\int_{-\infty}^{\infty} e^{-i2\pi(f-f')t}\,dt = K_0\,\delta(f-f')\ .\end{aligned} \tag{2.16}$$

2.3.2 Colored Noise

A lot of times, the quantity measured in an experiment is related to some starting quantity by a linear convolution, so even if we start out with white noise, we could end up dealing with something that has different properties. If we consider some random variable $h(t)$ which is produced by convolving white noise with a deterministic response function $R(t-t')$, so that

$$h(t) = \int_{-\infty}^{\infty} R(t-t')\,x(t')\,dt' \tag{2.17}$$

and in the frequency domain

$$\widetilde{h}(f) = \widetilde{R}(f)\widetilde{x}(f) \tag{2.18}$$

we have

$$\langle \widetilde{h}(f)\widetilde{h}^*(f')\rangle = K_0\,|\widetilde{R}(f)|^2\,\delta(f-f')\ . \tag{2.19}$$

Note that even in this case, $\langle|\widetilde{h}(f)|^2\rangle$ blows up, so simply looking at the magnitudes of Fourier components is not the most useful thing to do. However, the quantity $K_0\,|\widetilde{R}(f)|^2$ which multiplies the delta function can be well-behaved, and gives a useful spectrum. We'll see that this is the power spectral density $S_h(f)$, which we'll define more carefully in a bit.

We can also go back into the time domain in this example, and calculate the autocorrelation

$$\begin{aligned}K_h(t,t') &= \langle h(t)h(t')\rangle \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(t-t_1)\,R(t'-t_1')\;\langle x(t_1)x(t_1')\rangle\;dt_1\,dt_1' \\ &= \int_{-\infty}^{\infty} K_0 R(t-t_1)\,R(t'-t_1)\,dt_1\end{aligned} \tag{2.20}$$

which is basically the convolution of the response function with itself, time reversed; unlike in the case of white noise, where $K_x(t,t')$ was a delta function, this will in general be finite.

Note that in this example, the autocorrelation $K_h(t,t')$ is unchanged by shifting both of its arguments:

$$\begin{aligned}K_h(t+\tau,t'+\tau) &= K_0\int_{-\infty}^{\infty} R(t-[t_1-\tau])\,R(t'-[t_1-\tau])\,dt_1 \\ &= K_0\int_{-\infty}^{\infty} R(t-t_2)\,R(t'-t_2)\,dt_2 = K_h(t,t')\end{aligned} \tag{2.21}$$

where we make the change of integration variables from t_1 to $t_2 = t_1 - \tau$. This means that in this colored noise case the autocorrelation is a function only of $t-t'$.

2.4 Wide-Sense Stationary Data

We now turn to a general formalism which incorporates our observations about colored noise. A random time series $h(t)$ is called *wide-sense stationary* if it obeys

$$\langle h(t) \rangle = \mu = \text{constant} \tag{2.22}$$

and

$$\langle h(t)h(t') \rangle = K_h(t-t') \ . \tag{2.23}$$

Clearly, both our white noise and colored noise examples were wide-sense stationary. The appearance of a convolution in the time-domain (2.20) a product in the frequency domain (2.19) suggests to us that the Fourier transform of the auto-correlation function $K_h(t-t')$ is a useful quantity. We thus define the *power spectral density* as

$$S_h(f) = \int_{-\infty}^{\infty} K_h(\tau)\, e^{-i2\pi f\tau}\, d\tau \ . \tag{2.24}$$

Note that by the construction (2.23) the autocorrelation is even in its argument $[K_h(\tau) = K_h(-\tau)]$ so the PSD of real data will be real and even in f.

We can then show that for a general wide-sense stationary process,

$$\begin{aligned}\langle \widetilde{h}(f)\widetilde{h}^*(f') \rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \langle h(t)h(t') \rangle \ e^{-i2\pi(ft-f't')}\, dt\, dt' \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_h(\tau)\, e^{-i2\pi f\tau} e^{-i2\pi(f-f')t'}\, d\tau\, dt' \\ &= \delta(f-f')\, S_h(f) \ ,\end{aligned} \tag{2.25}$$

where we make the change of variables from t to $\tau = t - t'$.

2.4.1 Symmetry Properties of the Auto-Correlation and the PSD

We can see from the definition (2.23) and the result (2.25) that, for real data, both $K_h(\tau)$ and $S_h(f)$ are real and even, and in particular that $S_h(f) = S_h(-f)$. (Of course the fact that $K_h(\tau)$ and $S_h(f)$ are Fourier transforms of each other means that once we know the symmetry properties of one, we can deduce the symmetry properties of the other.) Because it's defined at both positive and negative frequencies, the power spectral density $S_h(f)$ that we've been using is called the *two-sided PSD*. Since the distinction between positive and negative frequencies depends on things like the sign convention for the Fourier transform, it is sometimes considered more natural to define a *one-sided PSD* which is defined only at non-negative frequencies, and contains all of the power at the corresponding positive and negative frequencies:

$$S_h^{\text{1-sided}}(f) = \begin{cases} S_h(0) & f = 0 \\ S_h(-f) + S_h(f) & f > 0 \end{cases} \tag{2.26}$$

Apparently, for real data, $S_h^{\text{1-sided}}(f) = 2S_h(f)$ when $f > 0$.

If the original time series is not actually real, there is a straightforward generalization of the definition of the autocorrelation function:

$$\langle h(t)h^*(t') \rangle = K_h(t-t') \tag{2.27}$$

The PSD is then defined as the Fourier transform of this, and (2.25) holds as before. Examination of (2.27) and (2.25) shows that, for complex data, the symmetry properties which remain are

- $K_h(-\tau) = K_h^*(\tau)$
- $S_h(f)$ is real.

Thursday, February 6, 2014

2.5 Power Spectrum Estimation

Suppose we have a stretch of data, duration T, sampled at intervals of δt, from a wide-sense stationary data stream,

$$h_j = h(t_0 + j\,\delta t) \tag{2.28}$$

where the autocorrelation

$$K_h(t-t') = \langle h(t)h(t')\rangle \tag{2.29}$$

and its Fourier transform (the power spectral density or PSD)

$$S_h(f) = \int_{-\infty}^{\infty} K_h(\tau)\, e^{-i2\pi f\tau}\, d\tau \tag{2.30}$$

are unknown. How do we estimate $S_h(f)$? One idea, keeping in mind that

$$\langle \widetilde{h}(f)\widetilde{h}^*(f')\rangle = \delta(f-f')\, S_h(f)\ , \tag{2.31}$$

is to use the discrete Fourier components to construct

$$|\widehat{h}_k|^2 \tag{2.32}$$

this is, up to normalization, the *periodogram*. Now, we're going to have to be a little careful about just using the identification

$$\widehat{h}_k\,\delta t \sim \widetilde{h}(f_k) \tag{2.33}$$

where

$$f_k = k\,\delta f = \frac{k}{T}\ ; \tag{2.34}$$

after all, taking that literally would mean setting f and f' both to f_k and evaluating the delta function at zero argument. So we need to be a little more careful about the approximation that relates the discrete and continuous Fourier transforms.

2.5.1 The Discretization Kernel $\Delta_N(x)$

If we substitute the continuous inverse Fourier transform into the discrete forward Fourier transform, we find

$$\begin{aligned}\widehat{h}_k &= \sum_{j=0}^{N-1} h(t_0 + j\,\delta t)\, e^{-i2\pi jk/N} = \sum_{j=0}^{N-1}\int_{-\infty}^{\infty} \widetilde{h}(f)\, e^{-i2\pi jk/N}\, e^{i2\pi f\,j\,\delta t}\, df \\ &= \int_{-\infty}^{\infty}\left(\sum_{j=0}^{N-1} e^{-i2\pi j(f_k-f)\delta t}\right)\widetilde{h}(f)\, df = \int_{-\infty}^{\infty}\Delta_N([f_k-f]\delta t)\,\widetilde{h}(f)\, df\end{aligned} \tag{2.35}$$

where we have defined[5]

$$\Delta_N(x) = \sum_{j=0}^{N-1} e^{-i2\pi jx}\ . \tag{2.36}$$

Let's look at some of the properties of this $\Delta_N(x)$. First, if x is an integer,

$$\Delta_N(x) = \sum_{j=0}^{N-1} e^{-i2\pi jx} = \sum_{j=0}^{N-1} 1 = N \qquad (x\in\mathbb{Z}) \tag{2.37}$$

Next, note that $\Delta_N(x)$ is periodic in x with period 1, since

$$\begin{aligned}\Delta_N(x+1) &= \sum_{j=0}^{N-1} e^{-i2\pi j(x+1)} = \sum_{j=0}^{N-1} e^{-i2\pi jx}e^{-i2\pi j} \\ &= \sum_{j=0}^{N-1} e^{-i2\pi jx} = \Delta_N(x)\ .\end{aligned} \tag{2.38}$$

[5]This is closely related to the *Dirichlet kernel*

$$\sum_{k=-n}^{n} e^{-ikx} = \frac{\sin([2n+1]x/2)}{\sin(x/2)}$$

.

Note that this is not surprising for something that relates a discrete to a continuous Fourier transform; incrementing $[f_k - f]\delta t$ by 1 is the same as decrementing f by $1/\delta t$, which is twice the Nyquist frequency. This is just the usual phenomenon of aliasing, where many continuous frequency components, separated at intervals of $1/\delta t$, are mapped onto the same discrete component.

Note also that

$$\Delta_N(\ell/N) = \sum_{j=0}^{N-1} e^{-i2\pi j\ell/N} = 0 \qquad \ell \in \mathbb{Z} \text{ and } \ell \neq 0 \bmod N \tag{2.39}$$

of course, it's sort of cheating to quote that result from before, since we got it by actually evaluating the sum, so let's do that again. Since

$$(1 - a^N) = (1 - a) \sum_{j=0}^{N-1} a^j \ , \tag{2.40}$$

if we set $a = e^{-i2\pi x}$, we get, for $x \notin \mathbb{Z}$ (which means $a \neq 1$),

$$\Delta_N(x) = \frac{1 - e^{-i2\pi N x}}{1 - e^{-i2\pi x}} = \frac{e^{-i\pi N x}}{e^{-i\pi x}} \frac{e^{i\pi N x} - e^{-i\pi N x}}{e^{i2\pi x} - e^{-i2\pi x}} = e^{-i\pi(N-1)x} \frac{\sin \pi N x}{\sin \pi x} \tag{2.41}$$

so, to summarize,

$$\Delta_N(x) = \begin{cases} N \ , & x \in \mathbb{Z} \\ e^{-i\pi(N-1)x} \left(\frac{\sin \pi N x}{\sin \pi x}\right) \ , & x \notin \mathbb{Z} \end{cases} \tag{2.42}$$

2.5.2 Derivation of the Periodogram

Now that we have a more precise relationship between $\widehat{h}_k$ and $\widetilde{h}(f)$, we can think about $|\widehat{h}_k|^2$, and in particular its expectation value:

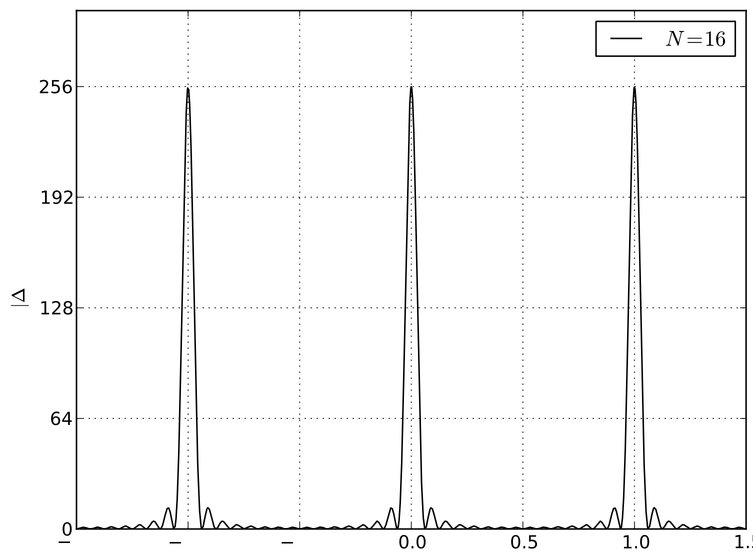
$$\begin{aligned} \langle |\widehat{h}_k|^2 \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Delta_N([f_k - f]\delta t) \Delta_N^*([f_k - f']\delta t) \langle \widetilde{h}(f) \widetilde{h}^*(f') \rangle \, df \, df' \\ &= \int_{-\infty}^{\infty} |\Delta_N([f_k - f]\delta t)|^2 \ S_h(f) \, df \end{aligned} \tag{2.43}$$

We can get a handle on

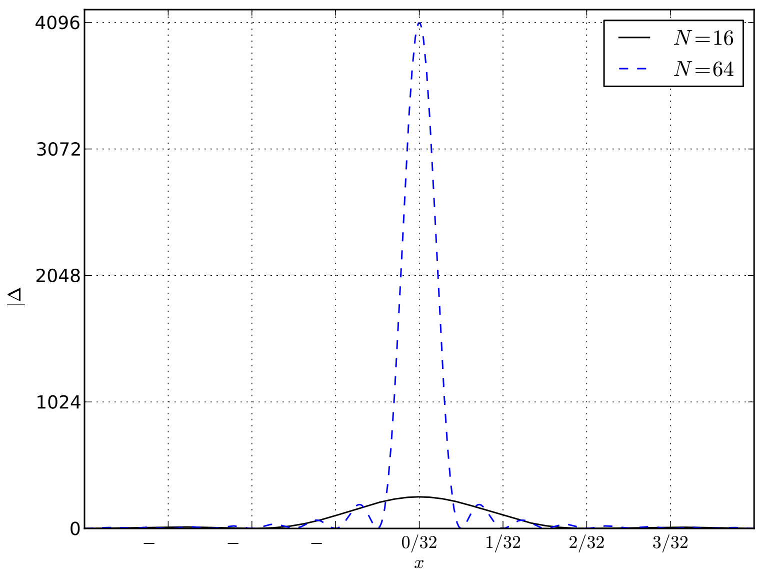
$$|\Delta_N(x)|^2 = \left(\frac{\sin \pi N x}{\sin \pi x}\right)^2 \tag{2.44}$$

by looking at some plots of it in NumPy/matplotlib; see the ipython notebook `http://ccrg.rit.edu/~whelan/courses/2014_1sp_ASTP_611/data/notes_fourier_kernel.ipynb`

The upshot is that $|\Delta_N(x)|^2$ is an approximation to a sum of Dirac delta functions (one peaked at each integer value of x):



and this approximation is better for higher N:



We can write this situation as

$$|\Delta_N(x)|^2 \approx \mathcal{N}_N \sum_{s=-\infty}^{\infty} \delta(x+s) \qquad (2.45)$$

To get the overall normalization constant $\mathcal{N}_N$, we have to integrate both sides of (2.45), using the fact that

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1 \qquad (2.46)$$

Of course, we don't actually want to integrate $|\Delta_N(x)|^2$ from $-\infty$ to ∞, because it's periodic, and we're bound to get something infinite when we add up the contributions from the infinite number of peaks. And likewise, the right-hand side contains an infinite number of delta functions. So we should just integrate over one cycle, using

$$\int_{-1/2}^{1/2} \delta(x)\, dx = 1 \qquad (2.47)$$

and choose $\mathcal{N}_N$ so that

$$\int_{-1/2}^{1/2} |\Delta_N(x)|^2\, dx = \mathcal{N}_N \sum_{s=-\infty}^{\infty} \int_{-1/2}^{1/2} \delta(x+s)\, dx = \mathcal{N}_N \qquad (2.48)$$

We could explicitly evaluate (or look up) the integral of $|\Delta_N(x)|^2 = \left(\frac{\sin \pi N x}{\sin \pi x}\right)^2$, but it turns out to be easier to evaluate it as

$$\mathcal{N}_N = \int_{-1/2}^{1/2} \Delta_N(x)[\Delta_N(x)]^*\, dx = \int_{-1/2}^{1/2} \sum_{j=0}^{N-1} e^{-i2\pi j x} \sum_{\ell=0}^{N-1} e^{i2\pi \ell x}\, dx$$
$$= \sum_{j=0}^{N-1} \sum_{\ell=0}^{N-1} \int_{-1/2}^{1/2} e^{-i2\pi(j-\ell)x}\, dx\ . \qquad (2.49)$$

But the integral is straighforward:

$$\int_{-1/2}^{1/2} e^{-i2\pi(j-\ell)x}\, dx = \begin{cases} 1 & j=\ell \\ \frac{\sin(\pi[j-\ell])}{\pi(j-\ell)} & j \neq \ell \end{cases} = \delta_{j\ell} \qquad (2.50)$$

so

$$\int_{-1/2}^{1/2} e^{-i2\pi(j-\ell)x}\, dx = \delta_{j\ell} \qquad j,\ell \in \mathbb{Z} \qquad (2.51)$$

and the normalization constant is

$$\mathcal{N}_N = \sum_{j=0}^{N-1} \sum_{\ell=0}^{N-1} \delta_{j\ell} = N \qquad (2.52)$$

and

$$|\Delta_N(x)|^2 \approx N \sum_{s=-\infty}^{\infty} \delta(x-s) \tag{2.53}$$

We can now substitute this back into (2.43) and find

$$\begin{aligned}
\left\langle \left|\widehat{h}_k\right|^2 \right\rangle &= \int_{-\infty}^{\infty} |\Delta_N([f_k-f]\delta t)|^2 \; S_h(f)\, df \\
&\approx N \sum_{s=-\infty}^{\infty} \int_{-\infty}^{\infty} \delta([f_k-f]\delta t - s)\, S_h(f)\, df \\
&= \frac{N}{\delta t} \sum_{s=-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(f-[f_k-s/\delta t])\, S_h(f)\, df \\
&= \frac{N}{\delta t} \sum_{s=-\infty}^{\infty} S_h(f_k - s/\delta t)\ .
\end{aligned} \tag{2.54}$$

That means that the correct definition of the *periodogram* is

$$P_k := \frac{\delta t}{N} \left|\widehat{h}_k\right|^2 \ . \tag{2.55}$$

Its expectation value is

$$\langle P_k \rangle = \int_{-\infty}^{\infty} \frac{\delta t}{N} |\Delta_N([f_k-f]\delta t)|^2 \; S_h(f)\, df \approx \sum_{s=-\infty}^{\infty} S_h(f_k - s/\delta t) \tag{2.56}$$

2.5.3 Shortcomings of the Periodogram

There are a few ways in which the periodogram is not quite an ideal estimate of the underlying PSD $S_h(f)$:

1. As noted above, it's not actually an approximation to the PSD at f_k, but to that, plus the PSD at $f_k + 1/\delta t$ plus the PSD at $f_k + 2/\delta t$ etc. This is the usual aliasing phenomenon; since $1/\delta t$ is twice the Nyquist frequency, we can avoid it by doing some sort of analog processing of our original time series so that $S_h(f) = 0$ if $|f|$ is above the Nyquist frequency, and then confining attention to k between $-N/2$ and $N/2-1$ so that f_k is between minus Nyquist and Nyquist. We'll assume we've done that.

2. The function $|\Delta_N([f_k-f]\delta t)|^2$ is not actually the greatest approximation to the Dirac delta function, because of the "ringing" in the side lobes outside of the main peak. This means the periodogram estimate at a given frequency will be "contaminated" with data from nearby frequencies to a greater degree than necessary. This phenomenon is called *spectral leakage*. The source of this problem is that by sampling $h(t)$ only from t_0 to t_0+T, we've effectively multiplied it by a rectangular window in the time domain

$$W_r(t) = \begin{cases} 0 & t < t_0 \\ 1 & t_0 \le t < T \\ 0 & t \ge T \end{cases} \tag{2.57}$$

which means the Fourier transform is the convolution of $\widetilde{W}(f)$ with $\widetilde{h}(f)$. We know that the Fourier transform of a rectangle is not the nicest thing in the world, so we're better off multiplying the data by a window which more smoothly rises from 0 to 1 and then goes back down again, since its Fourier transform will stretch out less in frequencies. We won't elaborate on that further right now, but see, e.g., Section 13.4 of *Numerical Recipes* for more.

3. While P_k has the right expected mean (2.56), its expected variance

$$\left\langle (P_k - \langle P_k \rangle)^2 \right\rangle = \left\langle P_k^2 \right\rangle - \langle P_k \rangle^2 \ , \tag{2.58}$$

which is a measure of the square of the typical error associated with the estimate, is larger than we'd like. We can look at

$$\left\langle P_k^2 \right\rangle = \left(\frac{\delta t}{N}\right)^2 \left\langle \widehat{h}_k \widehat{h}_k^* \widehat{h}_k \widehat{h}_k^* \right\rangle \; ; \tag{2.59}$$

now, we can't actually evaluate this without saying more about the properties of $h(t)$ than we've specified. We've talked about the expectation value and the autocorrelation, but not the full distribution of probabilities of possible values. We'll soon develop the machinery to consider such things, but for now, we'll just state that for some choices of that underlying distribution

$$\left\langle \widehat{h}_k \widehat{h}_k^* \widehat{h}_k \widehat{h}_k^* \right\rangle \sim 2 \left\langle \widehat{h}_k \widehat{h}_k^* \right\rangle \left\langle \widehat{h}_k \widehat{h}_k^* \right\rangle \tag{2.60}$$

and if that's the case

$$\left\langle P_k^2 \right\rangle \sim 2 \left(\frac{\delta t}{N} \left|\widehat{h}_k\right|\right)^2 = 2 \left\langle P_k \right\rangle^2 \tag{2.61}$$

but that means the expected mean square error is the square of the expectation value itself:

$$\left\langle (P_k - \langle P_k \rangle)^2 \right\rangle \sim \langle P_k \rangle^2 \approx [S_h(f_k)]^2 \tag{2.62}$$

so we have an estimate of the power whose typical error is the same size as the estimate itself!

Note that this is independent of T, which means you don't get any better of an estimate of the PSD of a wide-sense stationary process by including more data in the periodogram. This is perhaps not so surprising if we recall that the discrete Fourier transform was developed in the context of deterministic data. Having a longer stretch of data produces higher resolution in the frequency domain, because $\delta f = \frac{1}{T}$. So that means if you construct a periodogram from 50 seconds of data, you get not-very-accurate PSD estimates at frequencies separated by 0.02 Hz. If you use 200 seconds, you get PSD estimates at more frequencies, just 0.005 Hz apart, but they're not any more accurate; they still have RMS expected errors equal to their expectation values. However, if the underlying PSD $S_h(f)$ doesn't vary much as a function of frequency, then $S_h(43.000\,\text{Hz})$, $S_h(43.005\,\text{Hz})$, $S_h(43.010\,\text{Hz})$ etc may be more or less the same, and so the corresponding periodograms will be estimates of roughly the same quantity. So you'd want to average those together to get a more accurate estimate. I.e., you want to lower the frequency resolution.

You can get a lower frequency resolution by doing your Fourier transforms over a shorter time, i.e., by breaking up the time T into $\mathcal{N}_c$ chunks, each of duration $T/\mathcal{N}_c$, then doing discrete Fourier transforms over the $N = \frac{T}{\mathcal{N}_c \delta t}$ samples in each chunk. The periodogram from the αth chunk is then

$$P_{\alpha k} = \frac{\delta t}{N} \left|\widehat{h}_k\right|^2 \; ; \tag{2.63}$$

its frequency resolution is

$$\delta f = \frac{1}{N\,\delta t} = \frac{\mathcal{N}_c}{T} \; . \tag{2.64}$$

Each periodogram has expected mean

$$\langle P_{\alpha k} \rangle \approx P(f_k) = P(k \delta f) \tag{2.65}$$

and variance

$$\left\langle P_{\alpha k}^2 \right\rangle - \langle P_{\alpha k} \rangle \approx [P(f_k)]^2 \tag{2.66}$$

as before. But we can take the average of all $\mathcal{N}_c$ periodograms

$$\overline{P}_k = \frac{1}{\mathcal{N}_c} \sum_{\alpha=0}^{\mathcal{N}_c - 1} P_{\alpha k} \tag{2.67}$$

and since averaging independent random variables lowers the resulting variance,

$$\left\langle \overline{P}_k \right\rangle = P(f_n) \tag{2.68}$$

and variance

$$\left\langle \overline{P}_k^2 \right\rangle - \left\langle \overline{P}_k \right\rangle \approx \frac{[P(f_k)]^2}{\mathcal{N}_c} \; . \tag{2.69}$$

So the error in the PSD estimation shrinks like $1/\sqrt{\mathcal{N}_c}$. We can get a more accurate estimate of the PSD by breaking the time up into chunks and averaging, but at the expense of a coarser frequency resolution. The appropriate tradeoff depends on the sharpness of the features that we want to resolve in the spectrum.