Discrete Fourier Transform

Overview

1. Review of spectral representation
2. Harmonic regression
3. Discrete Fourier transform
4. Examples of the DFT
5. Periodogram
6. Hibert space perspective
7. Examples

Review of Spectral Representation

Spectral representation of a stationary process $\{X_t\}$ is

$$X_t = \int_{-1/2}^{1/2} e^{2\pi\imath\, t\lambda} Z(d\lambda)\ , \tag{1}$$

where Z represents a right-continuous, complex-valued random process with orthogonal increments for which Var $Z(d\lambda) = dF(\lambda)$. The covariances are

$$\gamma(h) = E\ X_{t+h}\ \overline{X_t} = \int_{-1/2}^{1/2} e^{2\pi\imath\, h\lambda} dF(\lambda)\ . \tag{2}$$

If F is absolutely continuous with derivative $dF(\lambda)/d\lambda = f(\lambda)$, then

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi\imath\, h\lambda} f(\lambda) d\lambda \quad \text{and} \quad f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi\imath\, \lambda h} \tag{3}$$

For real-valued processes, $f(\lambda)$ is symmetric about $\lambda = 0$.

Anova interpretation Since the variance of $\{X_t\}$ has the representation

$$\mathrm{Var}(X_t) = \gamma(0) = \int_{-1/2}^{1/2} f(\lambda) d\lambda \ ,$$

the spectral density represents a decomposition of variance into frequency intervals. Just as a pdf measures probability in intervals, the "power spectrum" $f(\lambda)$ shows the distribution of variance over frequency.

Review: Harmonic regression

Regression on sinusoids Regression models with sines and cosines are the underlying statistical models used in frequency domain analysis of time series. Consider a regression model that mimics the random phase model,

$$\begin{aligned} X_t &= \mu + R\cos(2\pi\lambda t + \varphi) + w_t \\ &= \mu + R\left(\cos(2\pi\lambda t)\cos\varphi) - \sin(2\pi\lambda t)\sin\varphi\right) + w_t \\ &= \mu + A\cos 2\pi\lambda t + B\sin 2\pi\lambda t + w_t \end{aligned}$$

where $A = R\cos\varphi$ and $B = -R\sin\varphi$; $R^2 = A^2 + B^2$ is the squared amplitude of the sinusoid at frequency λ. The period associated with the frequency λ is $1/\lambda$. (Alternatively, one can use *angular* frequencies, being 2π times the usual frequency. I intend to reserve the symbol ω for angular frequencies.)

Aliasing The frequency is restricted to the range $-1/2 < \lambda \le 1/2$. For discrete data, frequencies outside of this range are *aliased* into this range. For example, suppose that $\frac{1}{2} < (\lambda = 1 - \delta) < 1$, then

$$\begin{aligned} \cos(2\pi\lambda t) &= \cos(2\pi(1-\delta)t) \\ &= \cos(2\pi t)\cos(2\pi\delta t) + \sin(2\pi t)\sin(2\pi\delta t) \\ &= \cos 2\pi\delta t \ . \end{aligned}$$

A sampled sinusoid with frequency higher than $\frac{1}{2}$ appears as a sinusoid with frequency in the interval $[0, 1/2]$. $\frac{1}{2}$ is known as the *folding frequency*; we have to see two samples to estimate the energy associated with a sinusoid.

Fourier frequencies and orthogonality The frequency $-\frac{1}{2} < \lambda \le \frac{1}{2}$ is known as a Fourier frequency if the associated sinusoid completes an integer number of cycles in the observed length of data. Since the period is $1/\lambda$, Fourier frequencies have the form (assuming n is even)

$$\lambda_j = \frac{j}{n}\,, \quad j = 0, 1, 2, \ldots, n/2, \tag{4}$$

Because of aliasing, the set of j's stop at $n/2$. The advantage of considering the Fourier frequencies is that they generate an orthogonal set of regressors. For sines/cosines, we have

$$\begin{aligned}
\sum_{t=1}^{n} \cos^2(2\pi\lambda_j t) &= \begin{cases} n & j = 0, n/2 \\ n/2 & j = 1, \ldots, n/2-1 \end{cases} \\
\sum_{t=1}^{n} \sin^2(2\pi\lambda_j t) &= \begin{cases} 0 & j = 0, n/2 \\ n/2 & j = 1, \ldots, n/2-1 \end{cases} \\
\sum_{t=1}^{n} \cos(2\pi\lambda_k t)\sin(2\pi\lambda_j t) &= 0
\end{aligned}$$

Harmonic regression If we use Fourier frequencies in our harmonic regression, the regression coefficients are easily found since "$X'X$" is diagonal. Consider the coefficients in the harmonic regression (n even)

$$X_t = A_0 + \sum_{j=0}^{n/2-1} A_j \cos(2\pi\lambda_j t) + B_j \sin(2\pi\lambda_j t) + A_{n/2} \tag{5}$$

where we define the coefficients (which are also the least squares estimates)

$$\begin{aligned}
A_0 &= \sum_t X_t/n & A_{n/2} &= \sum_t X_t(-1)^t/n \\
A_j &= \frac{2}{n}\sum_t X_t \cos 2\pi\lambda_j t & B_j &= \frac{2}{n}\sum_t X_t \sin 2\pi\lambda_j t\ .
\end{aligned}$$

Note that $B_0 = B_{n/2} = 0$; there is no imaginary/sine component for these terms. The sum of squares captured by a specific sine/cosine pair at frequency λ_j ($j \ne 0, n/2$) is (recall in OLS regression that the regression SS is $\hat\beta' X' X \hat\beta$)

$$\text{Regr SS}_j = \frac{n}{2}(A_j^2 + B_j^2) = \frac{n}{2}\,R_j^2\ . \tag{6}$$

The amplitude of the fitted sinusoid R_j determines the variance explained by this term in a regression model.

Orthogonal transformation Since the harmonic regression 5 includes all $1 + \frac{n}{2}$ Fourier frequencies from zero to $\frac{1}{2}$, this regression fits n parameters to n observations $X_1, \ldots, X_n$. This is not estimation; it's a transformation. The model fits perfectly. Thus the variation in the fitted values is exactly that of the original data, and we obtain the following decomposition of the variance by adding up the regression sum-of-squares (6) attributed to each frequency:

$$\sum_t X_t^2 = n(R_0^2 + R_{n/2}^2) + \frac{n}{2} \sum_{j=1}^{n/2-1} R_j^2, \tag{7}$$

The weights on R_0 and $R_{n/2}$ differ since there is no sine term at these frequencies.

Hilbert space The data $\boldsymbol{X} = (X_1, \ldots, X_n)'$ form a vector in n-dimensional space. The usual basis for this space is the set of vectors

$$\mathbf{1}_j = (0 \ \ldots \ 0\ 0\ 1_j\ 0\ 0 \ \ldots \ 0).$$

Thus we can write $\boldsymbol{X} = \sum_t X_t \mathbf{1}_t$. The harmonic model uses a different orthogonal basis, namely the sines and cosines associated with the Fourier frequencies. The "saturated" harmonic regression (5) represents $\boldsymbol{X}$ in this new basis. The coordinates of $\boldsymbol{X}$ in this basis are the coefficients A_j and B_j. Since we are writing the same vector $\boldsymbol{X}$ in two different coordinate systems (that are both orthogonal), the length of the vector does not change. Thus we must have the equivalence of lengths evident in (7), which is Parseval's equality in the context of harmonic regression.

Changing to complex variables leads to the discrete Fourier transform.

Discrete Fourier transform

Definition The discrete Fourier transform (DFT) of the real-valued n-term sequence $X_0, \ldots, X_{n-1}$ is defined as (zero-based indexing on the data from 0 to $n-1$ is more convenient with the DFT)

$$J_{n,j} = \frac{1}{n}\sum_{t=0}^{n-1} X_t e^{-2\pi\imath\,\lambda_j t}, \quad j = 0, 1, 2, \ldots, n-1. \tag{8}$$

The DFT is the set of harmonic regression coefficients, written using complex variables. For $j = 0, 1, \ldots, \frac{n}{2}$,

$$\begin{aligned} J_{n,j} &= \frac{1}{n}\sum_{t=0}^{n-1} X_t e^{-2\pi\imath\,\lambda_j t} \\ &= \frac{1}{2}\left(\frac{2}{n}\sum_{t=0}^{n-1} X_t \cos(2\pi\lambda_j t) - iX_t \sin(2\pi\lambda_j t)\right) \\ &= \frac{1}{2}(A_j - iB_j) \end{aligned}$$

Caution: There are many conventions for the leading divisor. **S&S** define the DFT with leading divisor $1/\sqrt{n}$ and **R** omits this factor altogether. Always test your software (*e.g.*, take the transform of the sequence 1,1 and see if the leading term is 1, $\sqrt{2}$, or 2).

Matrix form As with harmonic regression, the DFT amounts to a change of basis transformation. Define the $n \times n$ matrix $F_{n,jk} = e^{2\pi\imath\,jk/n}$ and note $F_n^* F_n = n\, I_n$ (* denotes the conjugate of the transpose). We can then express the transform as

$$\boldsymbol{J}_n = \frac{1}{n} F_n^* \boldsymbol{Y} \ . \tag{9}$$

Fast Fourier transform (FFT) is an algorithm for evaluating the matrix multiplication (9) (which appears to be of order n^2) in order $n \log n$ operations by a clever recursion (which is basically Horner's rule for evaluating a polynomial). Here's the idea.

The DFT of a sequence $\{x_0, x_1\}$ of length $n = 2$ is easy: $J_2, 0 = (x_0 + x_1)/2$ and $J_2, 1 = (x_1 - x_2)/2$. Now consider the DFT of the sequence $\{x_0, x_1, x_2, x_3$:

$$\begin{aligned} \sum_{t=0}^{3} x_t e^{-2\pi\imath\,jt/4} &= (x_0 + x_2 e^{-2\pi\imath\,2j/4}) + (x_1 e^{-2\pi\imath\,j/4} + x_3 e^{-2\pi\imath\,3j/4} \\ &= (x_0 + x_2 e^{-2\pi\imath\,j/2}) + e^{-2\pi\imath\,j/4}(x_1 + x_3 e^{-2\pi\imath\,j/2}) \end{aligned}$$

You can see that the DFT of a sequence of 4 numbers can be written in terms of two DFTs of length 2, applied to the even-indexed and odd-indexed elements. This recursion works in general: the DFT of a sequence of n can be written (for even n) as a sum of the DFT of the even and odd-indexed elements. So what? Count the number of operations: if $n = 2^N$ is a power of two, you require $O(n \log_2 n)$ operations rather than $O(n^2)$. Wavelets take this one step farther, requiring order $O(n)$ operations.

Properties of the DFT

Linearity Since it's a linear transformation (matrix multiplication, a change of basis), the DFT is a linear operator. *e.g.*, the DFT of a sum is the sum of the DFT's:

$$J_{n,j}^{x+y} = \frac{1}{n} \sum_t (x_t + y_t) e^{-2\pi\imath\, \lambda_j t} = J_{n,j}^x + J_{n,j}^y .$$

Thus, once we understand how the DFT behaves for some simple series, we can understand it for any others that are sums of these simple cases.

Real-valued data Since we begin with n real-valued observations X_t, but obtain n complex values $J_{n,j}$, the DFT has a redundancy (symmetry):

$$\begin{aligned} \overline{J}_{n,n-j} &= \frac{1}{n} \sum_{t=0}^{n-1} X_t e^{2\pi\imath\, \lambda_{n-j} t} \\ &= \frac{1}{n} \sum_{t=0}^{n-1} X_t e^{-2\pi\imath\, \lambda_j t} \\ &= J_{n,j} \; . \end{aligned}$$

You can see this result in the harmonic regression as well. Whereas frequencies in the harmonic regression (5) goes from 0 to 1/2, frequencies in the DFT span 0 to $(n-1)/n$. For frequencies above $\lambda_{n/2} = \frac{1}{2}$, $J_{n,j} = A_j + \imath\, B_j$ $(j > n/2)$. One can exploit this symmetry to obtain the transform of two real-valued series at once from one application of the FFT.

Inversion We can recover the data from the DFT by inverting the transform,

$$\begin{aligned}\sum_j J_{n,j} e^{2\pi\imath\,\lambda_j t} &= \frac{1}{n}\sum_{j,s} X_s e^{2\pi\imath\,(\lambda_j t - \lambda_j s)} \\ &= \frac{1}{n}\sum_s X_s \sum_j e^{2\pi\imath\,\lambda_j (t-s)} \\ &= X_t \qquad (10)\end{aligned}$$

where the last step follows from the orthogonality at the Fourier frequences, $\sum_j e^{2\pi\imath\,\lambda_j(t-s)} = 0$ for $s \neq t$, and otherwise is n. The relationship (10) is the DFT version (or discrete-time version) of the spectral representation (1). Using the matrix form, multiplying both sides of (9) by F_n gives $\boldsymbol{Y} = F_n \boldsymbol{J}_n$ immediately.

Variance decomposition As in harmonic regression, we can associate a variance with $J_{n,j}$. In particular,

$$\begin{aligned}\sum_t X_t^2 &= \sum_t |X_t|^2 = \sum_t |\sum_j J_{n,j} e^{2\pi\imath\,\lambda_j t}|^2 \\ &= \sum_{j,k} J_{n,j}\overline{J}_{n,k} \sum_t e^{2\pi\imath\,(\lambda_j - \lambda_k)t} \\ &= n\sum_j J_{n,j}\overline{J}_{n,j} = n\sum_{j=0}^{n-1} |J_{n,j}|^2,\end{aligned}$$

which is a much "neater" formula than that offered in the real-valued harmonic regression model in (6). In matrix form, this is easier still:

$$\sum_t X_t^2 = \boldsymbol{Y}^*\boldsymbol{Y} = (F_n\boldsymbol{J}_n)^*(F_n\boldsymbol{J}_n) = \boldsymbol{J}_n^*(F_n^*F_n)\boldsymbol{J}_n = n\sum_j |J_{n,j}|^2$$

Convolutions If the input data are a product, $x_t = y_t z_t$, the DFT has again a very special form. Using the inverse transform we find that the transform of the product is the *convolution* of the transforms,

$$\begin{aligned}J_{n,j}^x &= \frac{1}{n}\sum_t y_t z_t e^{-2\pi\imath\,\lambda_j t} \\ &= \frac{1}{n}\sum_t y_t \left(\sum_k J_{z,k} e^{2\pi\imath\,\lambda_k t}\right) e^{-2\pi\imath\,\lambda_j t}\end{aligned}$$

$$
\begin{aligned}
&= \sum_k J_{z,k} \left(\frac{1}{n} \sum_t y_t e^{-2\pi\imath\,\lambda_{j-k} t} \right) \\
&= \sum_{k=0}^{n-1} J^z_{n,k}\, J^y_{n,j-k}
\end{aligned}
$$

Recall the comparable property of r.v.'s: the MGF of a sum of two ind. r.v.'s is the product of the MGF's and the distribution of the sum is the convolution.

Special Cases of the DFT

Constant. If the series $X_t = k$ for all t, then

$$J_{n,j} = \frac{1}{n} \sum_t X_t e^{-2\pi\imath\,\lambda_j t} = \frac{k}{n} \sum_t e^{-2\pi\imath\,\lambda_j t}$$

which is zero unless $j = 0$, in which case $J_0 = k$. Hence a constant input generates a single "spike" in the output at frequency zero.

Spike. If the input is zero except for a single non-zero value k at index s, then $J_{n,j} = \frac{k}{n} e^{-2\pi\imath\,\lambda_j s}$. The amplitude of the DFT is constant, with the phase a linear function of the location of the single spike.

Sinusoid. If $X_t = k e^{2\pi\imath\,\lambda t}$, then we obtain a multiple of the *Dirichlet kernel*,

$$J_{n,j} = \frac{1}{n} \sum_t e^{2\pi\imath\,(\lambda-\lambda_j) t} = e^{2\pi\imath\,(\lambda-\lambda_j)\frac{(n-1)}{2}} D_n(\lambda - \lambda_j)\ ,$$

If $\lambda = \lambda_k$ is a Fourier frequency, only $J_{n,k}$ is non-zero. The version of the Dirichlet kernel used here is (set up for frequencies rather than angular frequencies)

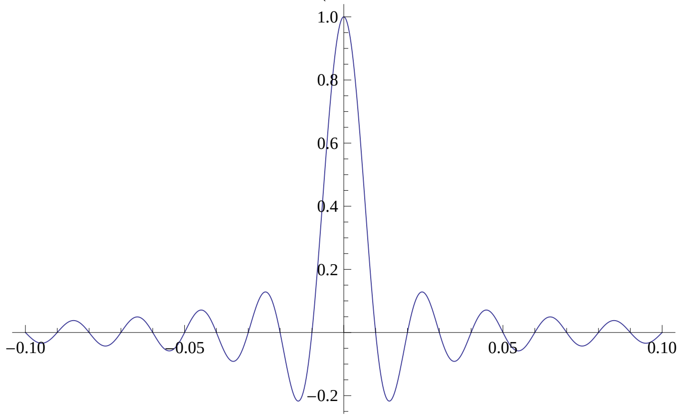
$$D_n(\lambda) = \frac{\sin(n\pi\lambda)}{n\ \sin(\pi\lambda)} \approx \frac{\sin(n\pi\lambda)}{n\ \pi\lambda} \text{ if } \lambda \approx 0. \tag{11}$$

The Dirichlet kernel arises as a sum of complex exponentials. In particular

$$D_n(\lambda) = \tfrac{1}{n} \sum_{j=-(n-1)/2}^{(n-1)/2} e^{2\pi\imath\,\lambda j} = \frac{e^{-2\pi\imath\,\lambda(n-1)/2}}{n} \sum_{j=0}^{n-1} e^{2\pi\imath\,\lambda j} \tag{12}$$

Some definitions of this kernel omit the leading factor $1/n$ so that $D_n(0) = n$ and $\int D_n(\lambda) d\lambda = 1$.

Here's a plot of the Dirichlet kernel. Notice that $D_n(\lambda_j) = 0$; it's zeros are at the Fourier frequencies ($n = 100$, so these are multiples of 0.01).



Note that $D_n(\lambda)$ does not have the "delta function" property.

Boxcar. If the input is the step function (or "boxcar"),

$$X_t = 1, t = 0, 1, ..., m-1, \quad X_t = 0,\ t = m, m+1, ..., n-1,$$

then $|J_{n,j}| = \frac{m}{n} D_m(\lambda_j)$.

Periodic function. Suppose that the input data X_t is composed of K repetitions of the sequence of H points x_t ($n = KH$). Then the DFT of X_t is (write $t = h + kH$)

$$\begin{aligned}
J_{n,j}^X &= \frac{1}{n} \sum_t X_t e^{-2\pi\imath\lambda_j t} \\
&= \frac{1}{KH} \sum_{k=0}^{K-1} \sum_{h=0}^{H-1} x_h e^{-2\pi\imath j(h+kH)/(KH)} \\
&= \left(\frac{1}{K} \sum_k e^{-2\pi\imath\,\frac{jk}{K}} \right) \frac{1}{H} \sum_h x_h e^{-2\pi\imath\,\frac{jh}{KH}} \\
&= D_K(j/K) \frac{1}{H} \sum_h x_h e^{-2\pi\imath\,\frac{jh}{KH}} \\
&= \begin{cases} 0 & \text{for} \quad j \neq 0, K, 2K, \ldots, (H-1)K. \\ J_\ell^x & \text{for} \quad j = \ell K. \end{cases}
\end{aligned}$$

The transform of the y's is zero except at multiples of K/n, which is known as the *fundamental frequency.*

Periodogram

Definition The periodogram I_n is the decomposition of variation associated with the harmonic regression and DFT,

$$\begin{aligned} I_n(\lambda_j) &= nJ_j\overline{J}_j \\ &= n|J_j|^2 \\ &= \frac{1}{n}|\sum_t x_t e^{-2\pi\imath\lambda_j t}|^2 \end{aligned} \tag{13}$$

The DFT sum-of-squares at the Fourier frequency λ_j is (see 6):

$$nJ_j\overline{J}_j = \begin{cases} \frac{n}{4}(A_j^2 + B_j^2) & j \neq 0, n/2 \\ nA_j^2 & j = 0, n/2 \end{cases}$$

At frequencies $\lambda_j \neq 0, \frac{1}{2}$, the DFT splits the variation assigned by the harmonic regression in keeping with the symmetry of the DFT around $\lambda = \frac{1}{2}$.

Statistical properties The relationship $I_n(\lambda_j) = n|J_{n,j}|^2 = \frac{n}{4}(A_j^2 + B_j^2)$ suggests that the asymptotic distribution of $I_n(\lambda_j)$ is a multiple of a χ^2 random variable with two degrees of freedom when the data are white noise,

$$I_n(\lambda_j) \propto \chi_2^2\ .$$

At $j = 0$ or $n/2$, the r.v. is χ_1^2. It is easy to see that the $I_n(\lambda_j)$ are uncorrelated when X_t is white noise; the transform is an orthogonal rotation of the data. The key property of the transformation is that the ordinates are uncorrelated even if the input data are *not* uncorrelated. We'll see why in the next class.

Hilbert space perspective

Motivation Remove most of the superficial complexity associated with the DFT by a change of notation and point of view.

Define The relevant Hilbert space is $\mathbb{C}^n$ with the usual inner product

$$\langle x, y\rangle = \sum_i x_i \overline{y}_i$$

Define the orthonormal vectors (you might add an n to remind you that these vectors lie in $\mathbb{C}^n$)

$$e_k = \left(1, e^{2\pi\imath\,\lambda_k}, e^{2\pi\imath\,2\lambda_k}, e^{2\pi\imath\,3\lambda_k}, \ldots, e^{2\pi\imath\,(n-1)\lambda_k}\right)/\sqrt{n} \tag{14}$$

The vectors $e_k,\ k = 0, 1, \ldots, n-1$ form a basis for $\mathbb{C}^n$.

DFT To obtain the DFT of a vector $x \in \mathbb{C}^n$, observe that

$$\sqrt{n}\; J_{n,j} = \frac{1}{\sqrt{n}} \sum_t e^{-2\pi\imath\,\lambda_j t} x_t = \langle x, e_j\rangle$$

(With the $\{e_k\}$ defined to be orthonormal, we obtain the normalization of the DFT as defined in the **S&S** textbook.) Hence, you can see that the DFT of x is the collection of inner products of x with this basis. Since $\{e_k\}$ are orthonormal, we can write

$$x = \sum_j \langle x, e_j\rangle e_j \; .$$

Basis matrix It also follows that $\|x\|^2 = \sum_j \langle x, e_j\rangle^2$. Other inner-product operations also follow, such as moving between points in the time domain and those in the frequency domain. Define the linear operator $T(x) = (\langle x, e_0\rangle, \langle x, e_1\rangle, \ldots, \langle x, e_{n-1}\rangle)$ (the matrix with rows $e_0, e_1, \ldots, e_{n-1}$). This operator is symmetric with $T^*T = I$. Hence, Tx is the DFT and inner products are perserved,

$$\langle x, y\rangle = \langle x, T^*Ty\rangle = \langle Tx, Ty\rangle$$

Convolutions? Convolution is *not* a "natural" property of a Hilbert space because convolution requires the notion of a product. Hilbert spaces don't. For products, we need to move from Hilbert spaces to objects known as algebras.

Define the product $x \cdot y$ in $\mathbb{C}^n$ element-wise. Under this definition, we start to see some special properties of the basis $\{e_k\}$. (Up to now, you can do everything with another orthonomal basis.) Notice that $e_k \cdot e_j = e_{k+j \bmod n}$: the collection $\{e_k\}$ form a group. The neat properties of the DFT when applied to stationary processes come from (a) the algebraic properties of this group and (b) the fact that $\{e_k\}$ are very nearly eigenvectors of *all* Toeplitz matrices.

Examples in R

Variable star data. This integer time series is reported to be the magnitude of a variable star observed on 600 successive nights (Whittaker and Robinson, 1924). Bloomfield (1976) shows that this data is essentially the sum of two sinusoids plus round-off error! The variable star data is in the file `varstar.dat`.

Raw periodogram suggests a much richer structure with power at many frequencies. The problem is leakage from the peaks. Unless the frequencies in the data occur *exactly* at Fourier frequencies, there will be leakage of power to nearby frequencies.