Sequential Data - 1D

Ronald N Bracewell (1921 – 2007)

- PhD in ionospheric research, Cavendish Lab (Ratcliffe).
- 1949 -1954 Radiophysics Laboratory of CSIRO; a founding father of radio astronomy, wrote the first textbook with Pawsey.
- aperture synthesis, radio astronomy - many fundamental papers on restoration, reconstruction, interferometry (1954-1974).
- co-discoverer of strong polarization in Cen A (NGC5128),1962

- the 'Bracewell probe' - autonomous interstellar vehicles for communication with alien civilizations.

- solar physics, especially the sunspot cycle and the solar interior, stimulated by discovery in South Australia of laminated sediments with rich record of astronomical Precambrian cyclicity.

- 1983 a new factorization of the discrete Fourier transform matrix :***The Hartley Transform (1986);*** also ***chirplets***

- **Bracewell 1965, 1980, 1999 – *The Fourier Transform and its applications***

Last time ....

We looked at three aspects of surveys/detections:

(1) How to use **maximum likelihood** (and Bayes if we wish) to map out space distribution more accurately than good ol' 1/Vmax.

(2) **Censored data / survival analysis** – how to use data from re-surveys when it's in the form of upper or lower limits:
- the censoring must be random;
- two algorithms are available to work out the luminosity function for censored variables;
- comparison of normalized luminosity distributions can be done for two censored variables.

(3) **The confusion limit** – the result of finite resolution 'adding up' the faint sources into a continuum:
- crucial for surveys – cf the 2C source counts, Big Bang vs. Steady State;
- confusion limit can be used via P(D) analysis to obtain population information below the level at which individual sources can be seen.

1D (Sequential Data) Statistics

Many observations consist of sequential data :

- **intensity vs position as a single-beam/pixel is scanned across the sky,**
- **signal variation along a row/column on a 2D (e.g. CCD) detector,**
- **single-slit spectra,**
- **time-measurements of intensity (or any other property like the stock market).**

1D (Sequential Data) Statistics

What do we want to do? (and this is just the start….)

- establish a baseline, so that signal on this baseline can be analyzed

- detect signal, identification for example of a spectral line or source in the data for which the noise may be comparable in magnitude to the signal

- filter, to improve signal-to-noise ratio

- quantify the noise

- period-search; find periodicities in the data

- trend-finding; can we predict the future behaviour of subsequent data?

- correlation of time series to find correlated signal between antenna pairs, or to find spectral lines

- modelling; many astronomical systems give us our data convolved with some instrumental function, and we want to get back to the true data.

1D Statistics - Data Transformations

Distinctive aspect of analyses: the feature of interest only emerges after a transformation, e.g.

(a) **filtering to find the feature, or**
(b) **transformation may be an integral part of the data, as in periodicity search, or spectral-line correlator.**

Expansions will be in **orthogonal functions**, e.g. Fourier series.
(Close affinity with PCA - the main features can be extracted from a jumble of data. What is extracted depends entirely on the **basis set** used. It's art and craft.)

- A scan ***f(t)*** ; ***t*** is a sequential or ordering index, e.g. time, space, wavelength.
- ***f*** is sampled at discrete intervals, thus $f(t_1), f(t_2), \ldots.$
- The set will be described by some sort of multivariate distribution function
- If Gaussian, **covariance matrix** of the ***f*** 's will be a sufficient description.

Data Transformations continued

Long scans in ***f(t)*** may be represented by

$$f(t) = \int_{-\infty}^{\infty} F(\omega)\mathcal{B}(t,\omega)\, d\omega \quad \text{or} \quad f(t) = \sum_i F_{\omega_i}\mathcal{B}(t,\omega_i)$$

in which the basis functions are **B** and the expansion coefficients are ***F***, the variable ***ω*** changing from continuous to discrete.

To be useful, we need transformations which can be reversed.
We get equations like

$$F_{\omega_j} = \sum_i f_{t_i}\mathcal{B}'(t_i,\omega)$$

with sampling at discrete values of ***t***, and with some simple relationship between **B** and **B'**. If **B** is the exponential function, we have the **Fourier transforms and series**.

Data Transformations continued continued

If our scan ***f*** is a **random variable**, then the coefficients ***F*** are random, and will have different values for each of the (discrete) values of ***ω: ω₁, ω₂ …***

The covariance matrix ***C*** of the coefficients describes ***F***, **if the stats are Gaussian**. The components of ***F*** are then described by a **multivariate Gaussian**.

A **basis set** giving a **diagonal *C*** is very **efficient at capturing the variance** in the data => data variation is compressed into the smallest number of coefficients F_ω.
=> use in data compression, noise isolation.

Requiring that C_F be diagonal leads to the **Karhunen-Loeve equation** which, for our discrete case, is an eigenvalue problem: $R\vec{\mathcal{B}} = \lambda\vec{\mathcal{B}}.$

The matrix ***R*** is $R_{ij} = E[\,f_n f_{n+(i-j)}]$, closely related to the autocorrelation function, and ***R*** is just the covariance matrix of the original data components $f(t_i)$.

Data Transformations concluded

With a model for the statistics of our data, we can construct ***R*** and solve the ***Karhunen-Loeve equations***. The eigenvectors **B** will be **discretized basis functions**, and **they may be the familiar sines and cosines of Fourier analysis**.

Or not: for e.g. **optimum data compression**, we may want tailor-made functions.

=> Chebyshev polynomials, wavelets…a modelling problem,

We might be able to '**Bayesiate'** from start to finish, finding basis functions and F, optimized and hands-off.

Fourier Analysis

The Fourier transform is king. Why?

(1) Most physical processes at both macro and micro levels involve **oscillation and frequency**: orbits of galaxies, stars or planets, atomic transitions at particular frequencies, spatial frequencies on the sky as measured by correlated output from pairs of telescopes.

We want **the frequencies composing data streams**; just the **amplitudes** of these frequency components may be the answer (as in the case of detection of a spectral line).

(2) In many physical sciences there is frequent need to measure **a single signal from a data series**. In measuring a specific attribute of this signal such as redshift, the power of Fourier analysis has long been recognized

(3) But it really comes down to one simple fact –

The existence of the Fast Fourier Transform (FFT).

(I'll come back to it.)

Fourier Analysis

Solutions to many questions posed of the data lie in taking the one-dimensional scan to pieces in a Fourier analysis:

Any continuous function may be represented as the sum of sines and cosines:

$$f(t) = \int_{-\infty}^{+\infty} F(\omega) \exp^{-i\omega t} \, dt$$

where ***F***, representing the phased amplitudes of the sinusoidal components of ***f***, is known as the **Fourier Transform (FT).**

Fourier Analysis - a bunch of well-known things

- ♠ The **FT** of a sine is a **delta function in the frequency domain** – c.f. period search.

- ♠ The **FT** of ***f* ¤*g***, the **cross-correlation or convolution** of functions ***f*** and ***g***, is ***F x G*** - c.f. stable instrumental profile convolved with line width.

- ♠ The **FT** of $f(t+\tau)$ is just the transform of ***f*** times a simple exponential $e^{-i\omega\tau}$. Use of this **shift theorem** has measured many redshifts, maybe millions.

- ♠ The **Wiener-Khinchine theorem** states that the **power spectrum** $|F(\omega)|^2$ and the **autocorrelation function** $\int f(\tau)\, f(t+\tau)\, d\tau$ are **Fourier pairs**. The autocorrelation function is very closely related to the covariance matrix and hence is a fundamental statistical quantity. Its relationship to the power spectrum is the basis of every digital spectrometer.

- ♠ Closely related is **Parseval's theorem**; this relates the variance of ***f*** and the variance in the mean of ***f***, to the power spectrum – cf cases where we have **correlated noise**, especially the prevalent and pernicious **"1/*f*" noise**.

- ♠ The **FT of a Gaussian is another Gaussian**. Given the prevalence of Gaussians everywhere, this is a very convenient result.

Fourier - Uniform Sampling

The **Discrete Fourier Transform (DFT)** has special features:

If the function sampled ***N*** times at uniform intervals ***Δt*** in the spatial (observed) frame, the total length in the ***t-direction*** is ***L = Δt x (N-1).***

Result is the **continuous function** multiplied by the **'comb'** function, producing a ***f'(t)*** which (with the interval in spatial frequency as ***Δv = 2 π Δt***) may be represented either as a sum of sines and cosines

$$f'(t) = A_n \Sigma \sin(n\Delta\nu) + B_n \Sigma \cos(n\Delta\nu)$$

or as a cosine series

$$f'(t) = A'_n \Sigma \cos(n\Delta\nu + \Phi'_n)$$

with amplitudes A'_n and phases Φ'_n given by $A'_n = \sqrt{A^2{}_n + B^2{}_n}\,,\ \phi'_n = \arctan(\frac{A_n}{B_n})\,.$

In the latter formulation, obtaining the DFT produces - by virtue of the 2π cyclic nature of sine and cosine - a **'FT plane'** for ***f'(t)*** which shows the amplitudes **mirror-imaged about zero frequency**, with a sampling in spatial frequency at intervals of ***2π / [Δt (N - 1)]*** and a repetition of the pattern at intervals of ***2π / Δt***.

Fourier - Uniform Sampling continued

There are **five criteria for successful discrete-sampling**:

1. The **Nyquist criterion** or **Nyquist limit** guarantees **no** information at spatial frequencies above $\pi/\Delta t$. The sampling interval Δt sets the highest spatial frequency $2\pi/\Delta t$ retained; higher frequencies present in the data are lost.

2. The **Sampling theorem: any** bandwidth-limited function can be specified **exactly** by regularly-sampled values provided that the sample interval does not exceed **a critical length** (approximately half the FWHM resolution), i.e. for an instrumental half-width B, $f'(t) \rightarrow f(t)$ if $\Delta t < B/2$. Any physical system is band-pass limited**, preventing** full recovery of the signal.

3. To avoid any ambiguity - **aliasing** - in the reconstruction of the scan from its DFT, the **sampling interval must be small enough** for the amplitude coefficients of components at frequencies as high as $\pi/\Delta t$ to be effectively zero. Otherwise there's a tangle with the negative tail of the repeating function → **ambiguity.**

4. The lowest frequencies are $2\pi/(N\Delta t)$. Such low-frequency components may be real or instrumental; but to find signal the scan length must exceed the width of single resolved features by **> 10**.

5. The integration time per sample must be **long enough for decent s/n**.

Fourier Transforms - Statistical Properties

For **data assessment or model-fitting in the Fourier domain**, we need to know **the probability distribution of the Fourier components** and their derived properties.

For the comparatively simple case where the “data" $\boldsymbol{f}$ are pure Gaussian noise, of known covariance $\boldsymbol{C_f}$, there are **analytical results** for the **Fourier components**, for the **power spectrum** and the **auto-** and **cross-correlation functions.**

There is a discussion of this case in W&J pp237-241. No systematic signal was present in these model data. And note that **in real life the input distribution functions are unlikely to be Gaussian**.

Fourier Transforms - Statistical Properties, more

Thus for reliable **error estimation – detailed Monte Carlo simulation**, building in the mess of real observation, is **essential**.

The analytic results of the W&J pages provide some guidance:

- **power spectra** will have problems of **consistency and bias**
- **correlation functions** will contain **highly correlated errors**
- **detail will have to be sacrificed** in estimating **response functions**.

The take-home point - we need a **reasonable idea of basic statistical properties** – power spectrum or correlation function – to make progress in understanding our data when it is in the form of scans.

The Fast Fourier Transform - the FFT

FFT – Cooley and Tukey 1965

Does the transform of *N* points in a time proportional to *N log N*, rather than the N^2 timing of a brute-force implementation. This is a monumental cpu saver.

Quirky (see Bracewell, or Numerical Recipes)
- **typical (?) arrangement of input / output data**
- **normalization**

Critical to most image processing; certainly to the design of radio telescopes

Algorithm was apparently known to Gauss – even before Fourier had discovered his series. It may be the most used algorithm on the planet. (Think about every .jpg image for a start.)

Example - Redshifts from Cross-Correlation

Tonry and Davis 1979

Galaxy spectrum $g(n)$ **with** $n = A \ln\lambda + B$, n **is bin number.**

Template spectrum $t(n)$**, zero redshift, instrumentally-broadened.**

Set up DFTs $G(k) = \Sigma_n\, g(n) \exp(-2\pi i n k/N)$ **, and** equiv for $T(k)$

Then FT for cross-correlation $c(n) = g ¤ t(n)$ **is** $C(k) = (1/N\sigma_g\sigma_t)\, G(k)\, T^*(k)$

Now set $g(n) = \alpha\, t(n) ¤ b\,(n - \delta)$

i.e. the galaxy spectrum is a multiple of the template spectrum convolved with a broadening function shifted by δ**. This function accounts for the** velocity dispersion and the redshift**, and we urgently seek its parameters.**

Assume $b(n)$ **Gaussian, and likewise for** $c(n)$ **, centered at** δ

Minimizing $X^2(\alpha, \delta; b) = \Sigma_n\, [\alpha\, t ¤ b\,(n - \delta) - g(n)\,]^2$

is equivalent to **maximizing** $(1/\sigma_{t \times b})\, c ¤ b(\delta)$

Example - Redshifts from Cross-Correlation

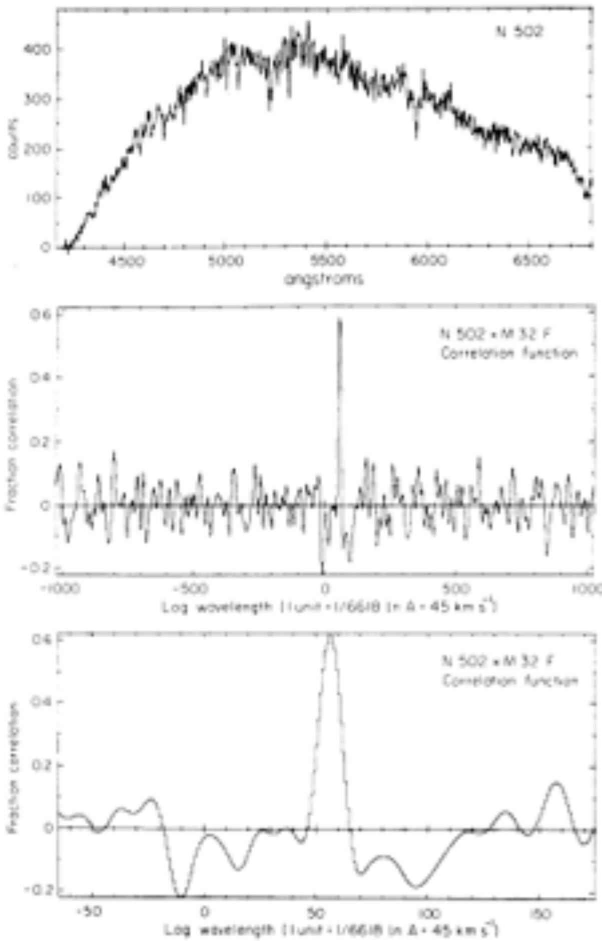

FIG. 10. Same as Fig. 8 for NGC 502.

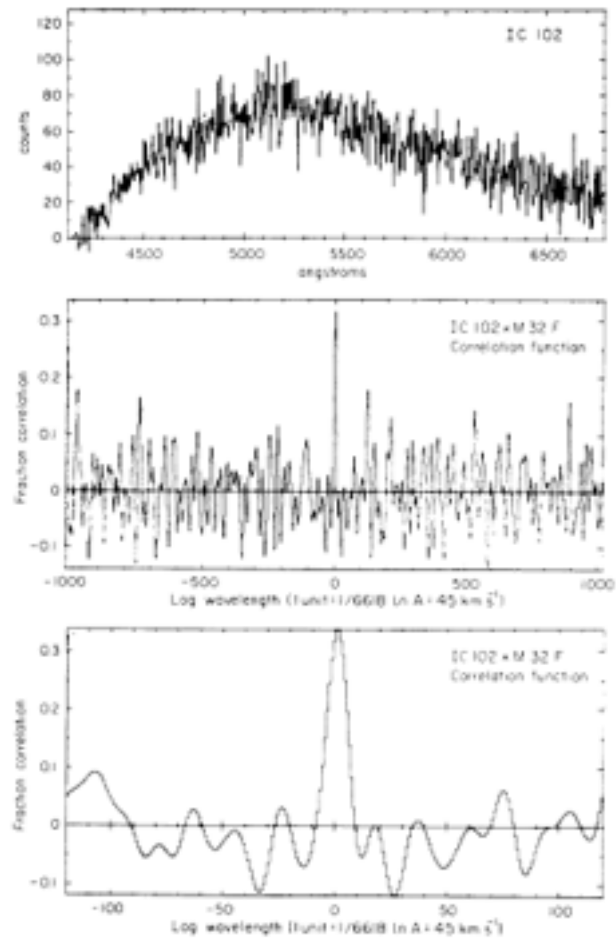

FIG. 11. Same as Fig. 8 for IC102.