Non-Parametric Analysis of NMR data

Daniel O’Donovan
CCPN - Dept. of Biochemistry
Introduction

NMR Data

Fourier Transform

Non-Parametric Analysis

maximise \( S = - \sum p_i \log p_i \)
Intro: NMR

- Method for obtaining physical, chemical, electronic and structural information about a molecule.
Small molecules - often easily determined
Larger, more complex molecules require increased resolution and dimensionality.
Limit where time restrains the amount of data that it is reasonable to collect.

Before
Limit where time restrains the amount of data that it is reasonable to collect.

New methods allow for increased dimensionality and resolution.
Intro: NMR

- Non-Uniform Sampling (NUS) method for increasing resolution without increasing time involves recording non-continuously.

- NUS data cannot be processed using regular Fourier Transforms.

- Require new methods to process this data.
Intro: Non-Parametric Methods

- Non-parametric; make no (few) assumptions about model.
- Favourite non-parametric method:
Non-parametric; make no (few) assumptions about model.

Favourite non-parametric method:

Maximum Entropy
Non-parametric; make no (few) assumptions about model.

Favourite non-parametric method:

Maximum Entropy

Entropy equivalent to (negative) information
Non-parametric; make no (few) assumptions about model.

Favourite non-parametric method:

**Maximum Entropy**

*Entropy equivalent to (negative) information*

*Maximising Entropy minimises information*
Intro: Non-Parametric Methods

- Non-parametric; make no (few) assumptions about model.
- Favourite non-parametric method:
  
  **Maximum Entropy**
  
  - *Entropy* equivalent to *(negative) information*
  
  - Maximising Entropy *minimises* information
  
  - Analogy... *(courtesy of Ray Freeman)*
Terrible crime committed in Cambridge at night

King's College Cambridge at night
Photo by "jgraham"
Non-Parametric Methods

Description from witness 1:

Hat
Non-Parametric Methods

Description from witness 1: Hat

Description from witness 2: Glasses

Thursday, 24 June 2010
Non-Parametric Methods

Description from witness 1:
Hat

Description from witness 2:
Glasses

Description from witness 3:
Book

Thursday, 24 June 2010
Non-Parametric Methods

Description from witness 1:
Hat

Description from witness 2:
Glasses

Description from witness 3:
Book

Description from witness 4:
Beard

Thursday, 24 June 2010
Non-Parametric Methods

Description from witness 1:
Hat

Description from witness 2:
Glasses

Description from witness 3:
Book

Least Committal Solution:

Description from witness 4:
Beard

Maximum Entropy:
Simplest

Thursday, 24 June 2010
Non-Parametric Methods

- In NMR, equivalent to having data set with parts missing.
- Task for Maximum Entropy is to reconstruct these missing parts.
Non-Parametric Methods

◆ In NMR, equivalent to having data set with parts missing.

◆ Task for Maximum Entropy is to reconstruct these missing parts.

Maximal Entropy spectrum
Non-Parametric Methods

- In NMR, equivalent to having data set with parts missing.
- Task for Maximum Entropy is to reconstruct these missing parts.

Maximal Entropy spectrum

Inverse Fourier Transform
Non-Parametric Methods

- In NMR, equivalent to having data set with parts missing.
- Task for Maximum Entropy is to reconstruct these missing parts.

Maximal Entropy spectrum → Inverse Fourier Transform → Calculate fit (Chi^2)
Non-Parametric Methods

- In NMR, equivalent to having data set with parts missing.
- Task for Maximum Entropy is to reconstruct these missing parts.

Maximal Entropy spectrum → Inverse Fourier Transform → Calculate fit (Chi^2) → Repeat until convergence
MaxEnt

Typically have sharper peaks and less noise than regular FFT spectra.
MaxEnt

- Maximum Entropy Advantages
  - Current state-of-the-art Maximum Entropy uses Bayesian methods to explore spectra.
MaxEnt

- Maximum Entropy Advantages
  - Current state-of-the-art Maximum Entropy uses Bayesian methods to explore spectra.

- Maximum Entropy Disadvantages
  - Fundamentally flawed: quality of result depends on resolution - should be resolution independent
MaxEnt

♦ Maximum Entropy Advantages

♦ Current state-of-the-art Maximum Entropy uses Bayesian methods to explore spectra.

♦ Maximum Entropy Disadvantages

♦ Fundamentally flawed: quality of result depends on resolution - should be resolution independent

♦ Alternatives

♦ Considering Independent Component Analysis (ICA)

♦ Sophisticated Markov Chain Monte Carlo methods (MCMC)
Code

- **MemSys** C Library (*MaxEnt Data Consultants, Drs. S. Gull and John Skilling (Dept. Astronomy))

- Powerful exploration algorithms

- Largely linear algebra operations

- Vector processing optimised using **OpenMP**, (also MPI and CUDA)

- Optimised Fourier Transform Library **FFTW3**

- Python (**ctypes**) wrapping code
Code: OpenMP

- Available in Fortran and C, through GCC 4.2+ and ICC
- In C, implemented with `#pragma` statements
- Trivial example:
Code: OpenMP

- Available in Fortran and C, through GCC 4.2+ and ICC
- In C, implemented with `#pragma` statements
- Trivial example:

```c
for ( i = 0; i < HUGE; i++ )
{
    x[i] = a * y[i];
}
```
Available in Fortran and C, through GCC 4.2+ and ICC

In C, implemented with `#pragma` statements

Trivial example:

```c
for (i = 0; i < HUGE; i++)
{
    x[i] = a * y[i];
}
```

```c
#pragma omp parallel for \
    shared(HUGE, x, a, y) \
    private(i)
for (i = 0; i < HUGE; i++)
{
    x[i] = a * y[i];
}
```
Code: OpenMP

- Available in Fortran and C, through GCC 4.2+ and ICC

- In C, implemented with `#pragma` statements

- Trivial example:

```
for ( i = 0; i < HUGE; i++ )
{
    x[i] = a * y[i];
}
```

```
#pragma omp parallel for \
    shared( HUGE, x, a, y) \
    private( i )
for ( i = 0; i < HUGE; i++ )
{
    x[i] = a * y[i];
}
```

- Invoked (GCC):

```
$ gcc -fopenmp file.c -o file.o -lgomp
```
Code: FFTW3

- FFTW3 the *Fastest Fourier Transform in the West*
- Highly optimised C library for calculating discrete Fourier Transforms
- Very easy to use and very fast

Thursday, 24 June 2010
Python: general purpose and high level

- Great for prototyping VERY quickly
- Has plotting and math libraries
- Great for calling optimised code libraries
- Several ways to interface with foreign libraries
- CCPN API and software written in Python
- Available from http://www.ccpn.ac.uk/
Example of Python with `ctypes`

# Load Python module ctypes
```python
>>> import ctypes
```

# Load regular C shared library ‘MaxEnt.so’
```python
>>> cLib = ctypes.cdll.LoadLibrary( ‘MaxEnt.so’ )
```

# Run C function ‘runUnitTests’ with no arguments
```python
>>> cLib.runUnitTests( None )
Tests Completed Successfully!
```

Also possible to pass and return ints, floats, arrays, pointers and structures
With Thanks

CCPN - Collaborative Computing Project for NMR

Prof. Ernest Laue, Tim Stevens, Wayne Boucher, Rasmus Fogh, John Ionides and Alan da Silva