Statistics and EEG Analysis

John Lind

EEG Workshop
Department of Psychology
May 19, 2017

EEG Analysis:

Topics:

1. The Model
2. The Spectrum and Cross-Spectrum
3. Power, Coherency and Phase
4. Regression
5. Analysis of Power
6. Discriminant Analysis
7. Canonical Correlation Analysis
8. Artifacts, Outliers and Cluster Analysis
9. Principal Components and Factor Analysis

Complex Variables

Complex variables:

$$z = a + ib, \qquad i = \sqrt{-1}$$

Conjugate:

$$Conj(z) = a - ib$$

Modulus:

$$\begin{aligned} Mod(z) &= |z|^{\frac{1}{2}} = \sqrt{a^2 + b^2} \\ &= [(Re\, z)^2 + (Im\, z)^2]^{\frac{1}{2}} \end{aligned}$$

Hermitian Matrix: (Example: Complex-valued covariance matrix)

$$\boldsymbol{S} = \begin{bmatrix} 2 & 3 - 3i \\ 3 + 3i & 5 \end{bmatrix} = \boldsymbol{S}^*$$

'*' denotes conjugate transpose

Euler's formula:

$$e^{ix} = cos(x) + i\, sin(x)$$

The Data

April 17, 2017 4:40 PM

The Model

Data: An r-vector-valued time series sampled from $t = 0, \ldots, T-1$. (Usually mean-deviated and linear trends removed).

$$\boldsymbol{X(t)} = \begin{bmatrix} x_{10} & x_{11} & \cdots & \cdots & x_{1,T-1} \\ x_{20} & x_{21} & \cdots & \cdots & x_{2,T-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{r0} & x_{r1} & \cdots & \cdots & x_{r,T-1} \end{bmatrix}$$

The model:

$$x(t) = \sum_{i=1}^{k} A_i cos(\omega_i t + \phi_i) + \epsilon_t$$

where $E[\epsilon_t] = 0$ and $E[\epsilon_t^2] = \sigma_\epsilon^2$

The Model

Can rewrite the model as:

$$x(t) = \sum_{i=1}^{k} (b_{i,1} cos\omega_i t + b_{i,2} sin\omega_i t) + \epsilon_t$$

where

$b_{i,1} = A_i cos\phi_i$ and $b_{i,2} = -A_i sin\phi_i$.

Solve for $b_{i,1}$ and $b_{i,2}$, $\quad i = 1, ...k$?

$$min \sum_{t=0}^{T-1} \epsilon_t^2 = min \sum_{t=0}^{T-1} (x(t) - \sum_{i=1}^{k} (b_{i,1} cos\omega_i t + b_{i,2} sin\omega_i t))^2$$

The Spectral Estimates

The O.L.S. solution:

$$\hat{b}_{i,1} = \frac{2}{T}\sum_{t=0}^{T-1} x(t) cos\omega_i t$$

and

$$\hat{b}_{i,2} = \frac{2}{T}\sum_{t=0}^{T-1} x(t) sin\omega_i t.$$

$\hat{b}_{i,1}$ and $\hat{b}_{i,2}$ are the cosine and sine transforms of $x(t)$.

The Discrete Fourier Transform:

$$\begin{aligned} d_x(\frac{2\pi_j}{T}) &= \sum_{t=0}^{T-1}(x(t)cos(\frac{2\pi jt}{T}) - i\ x(t)sin(\frac{2\pi jt}{T})) \\ &= \sum_{t=0}^{T-1} x(t)exp(\frac{-i2\pi jt}{T}), \quad j = 0, ..., T-1 \end{aligned}$$

Discrete Fourier Transform

Consider a $T \times T$ matrix with columns $exp(\frac{-i2\pi jt}{T}) \quad j = 0, \dots T-1$

$$\Psi = \begin{array}{c c} & \begin{array}{ccccccc} 0 & 1 & 2 & \dots & j & \dots & T-1 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ t \\ \vdots \\ T-1 \end{array} & \begin{bmatrix} 1 & 1 & 1 & 1 & exp(\frac{-i2\pi j0}{T}) = 1 & 1 & 1 \\ 1 & \vdots & \vdots & \vdots & exp(\frac{-i2\pi j1}{T}) & \vdots & \vdots \\ 1 & \vdots & \vdots & \vdots & exp(\frac{-i2\pi j2}{T}) & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \vdots & \vdots & \vdots & exp(\frac{-i2\pi jt}{T}) & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \vdots & \vdots & \vdots & exp(\frac{-i2\pi j(T-1)}{T}) & \vdots & \vdots \end{bmatrix} \end{array}$$

The first row and column contain 1's because either j or $t = 0$ in these locations and therefore the corresponding $exp(0) = 1$.

Discrete Fourier Transform

The Fourier coefficients, $\boldsymbol{F_x}$ can now be obtained by calculating $r \times T$ dot products $\boldsymbol{F_x} = \boldsymbol{X(t)}\boldsymbol{\Psi}$

$$\boldsymbol{F_x} = \qquad \boldsymbol{X(t)} \qquad\qquad \boldsymbol{\Psi}$$

$$= \begin{bmatrix} x_{10} & x_{11} & \cdots & x_{1,T-1} \\ x_{20} & x_{21} & \cdots & x_{2,T-1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{r0} & x_{r1} & \cdots & x_{r,T-1} \end{bmatrix} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & \dots & j & \dots & T-1 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & e^{\left(\frac{-i2\pi j0}{T}\right)} & 1 & 1 \\ 1 & \vdots & \vdots & \vdots & e^{\left(\frac{-i2\pi j1}{T}\right)} & \vdots & \vdots \\ 1 & \vdots & \vdots & \vdots & e^{\left(\frac{-i2\pi j2}{T}\right)} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \vdots & \vdots & \vdots & e^{\left(\frac{-i2\pi j(T-1)}{T}\right)} & \vdots & \vdots \end{bmatrix} \end{array}$$

The Spectrum Coefficients

The complex-valued spectrum coefficients $\boldsymbol{\omega} = \omega_0, \ldots, \omega_{T-1}$

$$
\boldsymbol{F_x} = \begin{array}{c c}
 & \begin{array}{ccccccc} \omega_0 & \omega_1 & \omega_2 & \ldots & \omega_j & \ldots & \omega_{T-1} \end{array} \\
\begin{array}{c} ch_1 \\ ch_2 \\ \vdots \\ ch_r \end{array} & \begin{bmatrix} f_{11} & f_{12} & f_{13} & \vdots & \boldsymbol{f_{1j}} & \cdots & f_{1,T-1} \\ f_{21} & f_{22} & f_{23} & \vdots & \boldsymbol{f_{2j}} & \cdots & f_{2,T-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{r1} & f_{r2} & f_{r3} & \vdots & \boldsymbol{f_{rj}} & \cdots & f_{r,T-1} \end{bmatrix}
\end{array}
$$

where $\boldsymbol{f_{ij}} = a + ib$.

In practice the rows of $\boldsymbol{F_x}$ are calculated using the fft.

Spectrum Coefficients

$\hat{\boldsymbol{F}}_{\boldsymbol{x}} =$

```
       w0         w1         w2          w3         w9        w10        w11 ->
ch1 0+0i  6.1+ 1.8i -0.9+2.7i  -4.8-7.4i  .  1.0+5.7i (34.0+17.3i) -1.0-0.7i ->
ch2 0+0i  2.8+ 2.8i  1.7-4.3i   0.9-9.4i  . -5.8+0.0i (31.0+ 9.1i) -9.7-6.4i ->
ch3 0+0i  0.7+10.2i -1.6+0.7i  -1.2+0.9i  . -5.9+3.5i (18.8-20.8i) -1.3+2.2i ->
ch4 0+0i  2.6+ 8.6i 10.0+5.3i -10.5+5.2i  . 10.1+4.5i (18.7+ 6.3i) -0.2+1.2i ->
ch5 0+0i -1.4+14.6i  4.0+6.8i  -4.3+5.5i  .  7.8+9.7i (28.0+10.0i) -1.7-6.8i ->
ch6 0+0i -5.2- 2.7i  0.5-8.9i   2.6-4.5i  . 11.3+2.4i (33.0+17.4i) -1.3-5.6i ->
```

$\uparrow\uparrow$

Usually scale $\hat{\boldsymbol{F}}_{\boldsymbol{x}}$ as $\frac{1}{\sqrt{T}}\hat{\boldsymbol{F}}_{\boldsymbol{x}}$

Now can select $\boldsymbol{x}$, a $r \times 1$ complex-valued vector at frequency ω. For example ω_{10} is $a + ib$ for each channel and 10 Hz.

Can calculate $\boldsymbol{x}'s$ for k segments and place in the rows of a $k \times r$ matrix, $\boldsymbol{X}$.

The Cross-Spectrum matrix

For k vectors arranged in a $k \times r$ matrix $\boldsymbol{X}$

the cross-spectrum matrix at frequency ω is obtained as

$$\boldsymbol{S} = \frac{1}{k}\boldsymbol{X}^*\boldsymbol{X}$$

$$= \frac{1}{k} \qquad \boldsymbol{X}^* \qquad \boldsymbol{X}$$

$$\boldsymbol{S}_{r\times r} = \frac{1}{k}\begin{bmatrix} x_{11}^* & x_{12}^* & x_{13}^* & \dots & x_{1k}^* \\ x_{21}^* & x_{22}^* & x_{23}^* & \dots & x_{2k}^* \\ x_{31}^* & x_{32}^* & x_{33}^* & \dots & x_{3k}^* \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{r1}^* & x_{r2}^* & x_{r3}^* & \dots & x_{rk}^* \end{bmatrix}\begin{bmatrix} x_{11} & x_{12} & x_{13} & \vdots & x_{1r} \\ x_{21} & x_{22} & x_{23} & \vdots & x_{2r} \\ x_{31} & x_{32} & x_{33} & \vdots & x_{3r} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{k1} & x_{k2} & x_{k3} & \dots & x_{kr} \end{bmatrix}$$

where $x_{ij} = a + ib$ and $x_{ji}^* = Conj(x_{ij})$.
$\boldsymbol{S}$ is of rank k.

Example

Cross Spectrum Matrix $\boldsymbol{S}(\omega)$:

	ch1	ch2	ch3	ch4	ch5	c
ch1	22.58+ 0.00i	15.31- 1.78i	1.64-17.92i	-0.34-14.78i	13.46- 1.73i	17.69- 1.8
ch2	15.31+ 1.78i	17.66+ 0.00i	4.73-16.41i	0.85-13.34i	11.71+ 1.60i	15.86+ 1.8
ch3	1.64+17.92i	4.73+16.41i	21.28+ 0.00i	15.52- 4.12i	3.15+13.26i	4.09+17.2
ch4	-0.34+14.78i	0.85+13.34i	15.52+ 4.12i	15.41+ 0.00i	-0.12+11.85i	0.28+15.3
ch5	13.46+ 1.73i	11.71- 1.60i	3.15-13.26i	-0.12-11.85i	14.55+ 0.00i	14.66- 0.9
ch6	17.69+ 1.89i	15.86- 1.89i	4.09-17.24i	0.28-15.30i	14.66+ 0.95i	21.29+ 0.0

For frequency ω, $\boldsymbol{S}$ is complex-valued and Hermitian.

- Diagonal contains the variances (power)
- Off-diagonal contains the covariances (cross-spectrum)
- All information contained in $\boldsymbol{S}$

Data Tapers and Windows

Data Taper:
In practice, before applying the fft, a data taper is usually applied to each channel to reduce 'spectral leakage' into neighbouring frequencies.

Example: Cosine bell (the Hamming or Hanning taper):

$$h_t = .5\left[1 + cos\left(\frac{2 * \pi(t - \bar{t})}{T}\right)\right]$$

where $\bar{t} = (T + 1)/2$ and T is the number of points in the data segment.

Data Tapers and Windows (cont'd)

Frequency Windows:
After calculating the spectrum or cross-spectrum matrix for each frequency, it is common practice to 'smooth' the estimates by averaging over neighbouring frequencies.

Example:
If the EEG alpha band is defined as 8-12 Hz, a smoothed estimate of the cross-spectrum matrix can be obtained by averaging over the alpha frequencies such as

$$\boldsymbol{S}_{alpha} = \frac{1}{5}(\boldsymbol{S}_8 + \boldsymbol{S}_9 + \boldsymbol{S}_{10} + \boldsymbol{S}_{11} + \boldsymbol{S}_{12}).$$

As defined earlier, each of the cross-spectrum matrices $\boldsymbol{S}_8, ..., \boldsymbol{S}_{12}$ above are calculated from k data segments.

Coherence

For a specified frequency, ω calculate coherency as

$$|r_{ij}| = \frac{|s_{ij}|}{\sqrt{s_{ii}^2 s_{jj}^2}} = \left[\frac{c_{ij}^2 + q_{ij}^2}{s_{ii}^2 s_{jj}^2}\right]^{\frac{1}{2}}$$

where $c_{ij} = Re(s_{ij})$ and $q_{ij} = Im(s_{ij})$.

Example:

$$r_{12} = \frac{15.31 - 1.78i}{\sqrt{22.58 + 0i}\sqrt{17.66 + 0i}}$$

$$|r_{12}| = \sqrt{\frac{15.31^2 + 1.78^2}{(2.58)(17.66)}} = .772$$

The Coherence Matrix

The $r \times r$ complex-valued coherence matrix, $\boldsymbol{R}$ can be obtained as

$$r_{ij} = s_{ij} / \sqrt{s_{ii}^2} \sqrt{s_{jj}^2}, \qquad i, j = 1, \ldots, r$$

or,

$$\boldsymbol{R} = \boldsymbol{D}^{-1/2} \boldsymbol{S} \boldsymbol{D}^{-1/2}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{s_{11}^2}} & & 0 \\ & \frac{1}{\sqrt{s_{22}^2}} & \\ 0 & & \frac{1}{\sqrt{s_{33}^2}} \end{bmatrix} \begin{bmatrix} s_{11}^2 & s_{12} & s_{13} \\ s_{21} & s_{22}^2 & s_{23} \\ s_{31} & s_{32} & s_{33}^2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{s_{11}^2}} & & 0 \\ & \frac{1}{\sqrt{s_{22}^2}} & \\ 0 & & \frac{1}{\sqrt{s_{33}^2}} \end{bmatrix}$$

Examples

Complex Coherence Matrix ($\boldsymbol{R}$)

	ch1	ch2	ch3	ch4	ch5	ch6
ch1	1.00+0.00i	0.77-0.09i	0.07-0.82i	-0.02-0.79i	0.74-0.10i	0.81-0.09i
ch2	0.77+0.09i	1.00+0.00i	0.24-0.85i	0.05-0.81i	0.73+0.10i	0.82+0.10i
ch3	0.07+0.82i	0.24+0.85i	1.00+0.00i	0.86-0.23i	0.18+0.75i	0.19+0.81i
ch4	-0.02+0.79i	0.05+0.81i	0.86+0.23i	1.00+0.00i	-0.01+0.79i	0.02+0.84i
ch5	0.74+0.10i	0.73-0.10i	0.18-0.75i	-0.01-0.79i	1.00+0.00i	0.83-0.05i
ch6	0.81+0.09i	0.82-0.10i	0.19-0.81i	0.02-0.84i	0.83+0.05i	1.00+0.00i

Coherency Matrix: Modulus($\boldsymbol{R}$)

	ch1	ch2	ch3	ch4	ch5	ch6
ch1	1.0000000	0.7719745	0.8211096	0.7924772	0.7490382	0.8115489
ch2	0.7719745	1.0000000	0.8808886	0.8107135	0.7370743	0.8235586
ch3	0.8211096	0.8808886	1.0000000	0.8868031	0.7744842	0.8325135
ch4	0.7924772	0.8107135	0.8868031	1.0000000	0.7918235	0.8449067
ch5	0.7490382	0.7370743	0.7744842	0.7918235	1.0000000	0.8346244
ch6	0.8115489	0.8235586	0.8325135	0.8449067	0.8346244	1.0000000

Coherency: Hypothesis tests

To test $H_o : \rho_{ij}^2 = 0$ at frequency ω:

$$\frac{|r_{ij}|^2}{1 - |r_{ij}|^2}(k-1) \sim F_{\alpha,2,2k-2}$$

where k is the number of segments or number of averaged frequencies.

Can show

$$\frac{|r_{ij}|^2}{1 - |r_{ij}|^2} = \frac{SSq's\ Regression}{SSq's\ Residual}$$

For C.I.'s apply Fisher's r to z transformation to $|r_{ij}|$

Phase

For frequency ω

$$\hat{\phi}(\omega) = tan^{-1}\left(\frac{-\hat{q}_{ij}(\omega)}{\hat{c}_{ij}(\omega)}\right)$$

Variance:

$$var(\hat{\phi}(\omega)) = \frac{C}{2T}\left\{\frac{1}{|r_{ij}(\omega)|^2} - 1\right\}$$

When $|r_{ij}(\omega)|^2$ near 0 ?

Phase (cont'd):

When $\rho_{ij}(\omega) = 0$, $\hat{\phi}(\omega)$ has a uniform distribution in $[-\pi, \pi]$, i.e. can be anywhere in the interval and is meaningless.

When $\rho_{ij}(\omega) \neq 0$, can approximate with a normal distribution.

Partial Coherence

Wish to remove the correlation of one set of channels from a second set. Ex: Remove correlations of channels 5-6 (possibly EOG, EMG) from 1-4.

$$
\boldsymbol{S} = \begin{array}{c|c}
 & \begin{array}{cccc|cc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\
\begin{array}{c} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array} &
\left[\begin{array}{cccc|cc}
s_{1,1}^2 & s_{1,2} & s_{1,3} & s_{1,4} & s_{1,5} & s_{1,6} \\
s_{2,1} & s_{2,2}^2 & s_{2,3} & s_{2,4} & s_{2,5} & s_{2,6} \\
s_{3,1} & s_{3,2} & s_{3,3}^2 & s_{3,4} & s_{3,5} & s_{3,6} \\
s_{4,1} & s_{4,2} & s_{4,3} & s_{4,4}^2 & s_{4,5} & s_{4,6} \\
\hline
s_{5,1} & s_{5,2} & s_{5,3} & s_{5,4} & s_{5,5} & s_{5,6} \\
s_{6,1} & s_{6,2} & s_{6,3} & s_{6,4} & s_{6,5} & s_{6,6}^2
\end{array}\right]
\end{array}
$$

Partial Coherency (cont'd):

Partition of $\boldsymbol{S}$:

$$\boldsymbol{S} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array} \begin{array}{c} \begin{array}{cccc|cc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \left[\begin{array}{c|c} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \hline \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{array}\right] \end{array}$$

Solution:

$$\boldsymbol{S}_{11.2} = \boldsymbol{S}_{11} - \boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}$$

Regression

At frequency ω, wish to predict activity at one lead location from several other locations. For example ch6 from ch2,...,ch5.

The regression equation:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{b_o}$$

O.L.S. solution:

$$\hat{\boldsymbol{b}} = (\boldsymbol{X}^*\boldsymbol{X})^{-1}\boldsymbol{X}^*\boldsymbol{y}$$

$$\boldsymbol{S} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array} \begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \left[\begin{array}{ccccc|c} s_{1,1}^2 & s_{1,2} & s_{1,3} & s_{1,4} & s_{1,5} & s_{1,6} \\ s_{2,1} & s_{2,2}^2 & s_{2,3} & s_{2,4} & s_{2,5} & s_{2,6} \\ s_{3,1} & s_{3,2} & s_{3,3}^2 & s_{3,4} & s_{3,5} & s_{3,6} \\ s_{4,1} & s_{4,2} & s_{4,3} & s_{4,4}^2 & s_{4,5} & s_{4,6} \\ s_{5,1} & s_{5,2} & s_{5,3} & s_{5,4} & s_{5,5} & s_{5,6} \\ \hline s_{6,1} & s_{6,2} & s_{6,3} & s_{6,4} & s_{6,5} & s_{6,6}^2 \end{array}\right] \end{array}$$

Regression (cont'd)

O.L.S. solution:

Regression coefficients (complex-valued):

$$\hat{\boldsymbol{b}} = (\boldsymbol{X}^* \boldsymbol{X})^{-1} \boldsymbol{X}^* \boldsymbol{y}$$

$$\boldsymbol{S} = \begin{array}{cc} & \begin{array}{cccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \begin{array}{c} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array} & \left[\begin{array}{c|c} \boldsymbol{X}^* \boldsymbol{X} & \boldsymbol{X}^* \boldsymbol{y} \\ \hline \boldsymbol{y}^* \boldsymbol{X} & \boldsymbol{y}^* \boldsymbol{y} \end{array} \right] \end{array}$$

$$\hat{R}^2 = \frac{1}{\boldsymbol{y}^* \boldsymbol{y}} \boldsymbol{y}^* \boldsymbol{X} (\boldsymbol{X}^* \boldsymbol{X})^{-1} \boldsymbol{X}^* \boldsymbol{y}$$

Regression: R example

Wish to predict ch_6 from channels 1-5:

The regression equation:

$ch_6 = b_1 ch_1 + b_2 ch_2 + ... + b_5 ch_5$

R output: (From Coherence matrix gives complex valued beta weights)

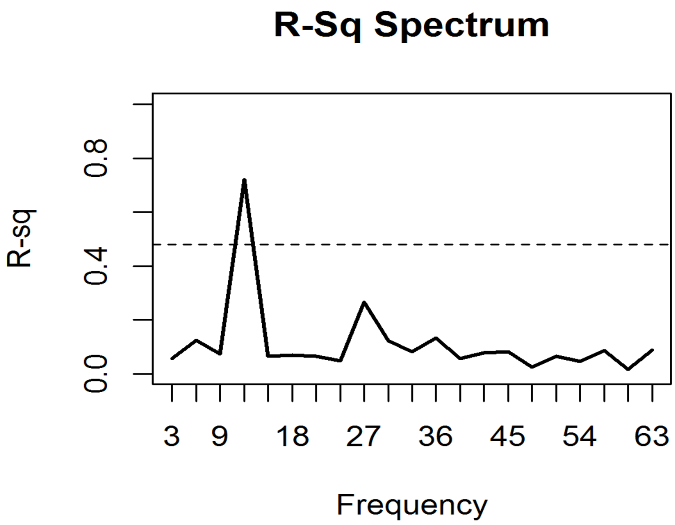
```
> round(R,2)
              ch1          ch2          ch3          ch4          ch5          ch6
ch1  1.00+0.00i -0.24+0.22i  0.14-0.09i  0.16+0.61i  0.30-0.14i  0.14+0.19i
ch2 -0.24-0.22i  1.00+0.00i  0.59-0.06i  0.19-0.41i -0.46-0.37i -0.25-0.26i
ch3  0.14+0.09i  0.59+0.06i  1.00+0.00i -0.10-0.36i  0.04-0.39i -0.35+0.14i
ch4  0.16-0.61i  0.19+0.41i -0.10+0.36i  1.00+0.00i -0.26-0.30i -0.09-0.41i
ch5  0.30+0.14i -0.46+0.37i  0.04+0.39i -0.26+0.30i  1.00+0.00i  0.12-0.01i
ch6  0.14-0.19i -0.25+0.26i -0.35-0.14i -0.09+0.41i  0.12+0.01i  1.00+0.00i
```

Regression: R example (cont'd)

The Regression Coefficients:

```
> b <- solve(R[1:5,1:5]) %*% R[1:5,6]
> round(b,2)
           [,1]
ch1 -0.49+1.01i
ch2 -0.31-0.64i
ch3  0.64-0.21i
ch4 -0.80-1.27i
ch5 -0.67-0.76i
> round(Mod(b),2)
    [,1]
ch1 1.12
ch2 0.71
ch3 0.68
ch4 1.50
ch5 1.01
```

Regression (cont'd)



One-way Manova layout:

When have a non-zero mean signal, define for frequency ω

- $y_{ijk}(\omega)$ DFT of *ith* individual in *jth* group and *kth* channel
- $\bar{\boldsymbol{y}}_{.jk}(\omega)$ DFT of mean signal for *jth* group and *kth* channel
- $\bar{\boldsymbol{y}}_{..p}(\omega)$ DFT of grand mean signal (p channels)

A_1	A_2	A_3
$y_{111}, \dots, y_{11p}$	$y_{121}, \dots, y_{12p}$	$y_{131}, \dots, y_{13p}$
$y_{211}, \dots, y_{21p}$	$y_{221}, \dots, y_{22p}$	$y_{231}, \dots, y_{23p}$
$\vdots$	$\vdots$	$\vdots$
$y_{n11}, \dots, y_{n1p}$	$y_{n21}, \dots, y_{n2p}$	$y_{n31}, \dots, y_{n3p}$
$\bar{y}_{.11}, \dots, \bar{y}_{.1p}$	$\bar{y}_{.21}, \dots, \bar{y}_{.2p}$	$\bar{y}_{.31}, \dots, \bar{y}_{.3p}$

Grand mean: $\bar{y}_{..1}, \dots, \bar{y}_{..p}$

Analysis of Power: (cont'd).

The between ($\boldsymbol{H}$) and and within ($\boldsymbol{E}$) SSCP matrices are:

$$\boldsymbol{H} = \sum_{j}^{J} n_j \begin{bmatrix} \vdots \\ \bar{\boldsymbol{y}}_{.j} - \bar{\boldsymbol{y}}_{...} \\ \vdots \end{bmatrix} \begin{bmatrix} \cdots & \bar{\boldsymbol{y}}^*_{.j} - \bar{\boldsymbol{y}}^*_{...} & \cdots \end{bmatrix}$$

$$\boldsymbol{E} = \sum_{j}^{J} \sum_{i}^{n_j} \begin{bmatrix} \vdots \\ \bar{\boldsymbol{y}}_{ij} - \bar{\boldsymbol{y}}_{.j} \\ \vdots \end{bmatrix} \begin{bmatrix} \cdots & \bar{\boldsymbol{y}}^*_{ij} - \bar{\boldsymbol{y}}^*_{.j} & \cdots \end{bmatrix}$$

Hypothesis Tests:

Roy's greatest characteristic root of $\boldsymbol{HE}^{-1}$

Wilk's Lambda:

$$\Lambda = \frac{|\boldsymbol{E}|}{|\boldsymbol{H} + \boldsymbol{E}|}$$

where

$$\chi^2_{2(J-1)p} = -2(\sum n_j - J - p - 1) log \Lambda$$

with $2(J-1)p$ d.f.

Analysis of Power: (cont'd.)

Hypothesis Tests: 2 group case (Hotelling's T^2)

$$\hat{T}^2 = \frac{n_1 n_2}{n_1 + n_2} [\bar{\boldsymbol{y}}_{.1} - \bar{\boldsymbol{y}}_{.2}]^H S_{pool.}^{-1} [\bar{\boldsymbol{y}}_{.1} - \bar{\boldsymbol{y}}_{.2}]$$

where $S_{pool.} = (E_1 + E_2)/(n_1 + n_2 - 2)$.

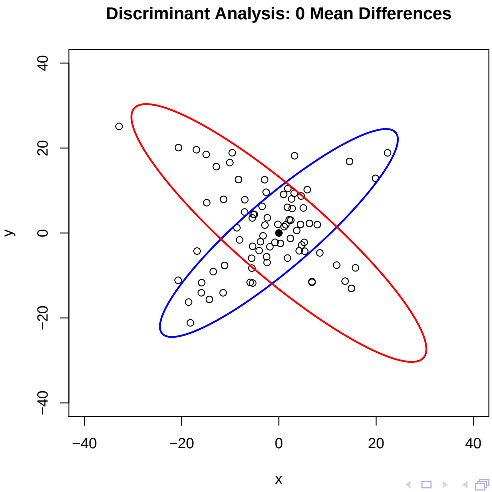
Test with

$$F_{2p,2(n_1+n_2-p-1)} = \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} \hat{T}^2$$

Can generate T^2 and F spectral plots.

Extend to mixed model and factorial designs.

Objective: Classify individuals with unknown group membership into one of several populations.



Classifying Vectors:

Calculate Mahalanobis distances from each group centroid and assign to the group with the shortest distance.

$$d_q(\boldsymbol{x}_i) = \boldsymbol{x}_i^* \boldsymbol{\Sigma}_1^{-1} \boldsymbol{x}_i - \boldsymbol{x}_i^* \boldsymbol{\Sigma}_2^{-1} \boldsymbol{x}_i$$

Assign to population 1 if $d_q(\boldsymbol{x}_1) < 0$ and population 2 if $d_q(\boldsymbol{x}_1) \geq 0$. In practice replace $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ with the sample average group spectral matrices $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$.

Classifying Cross Spectrum Matrices:

For a $r \times 1$ complex vector $\boldsymbol{x}_i$ and $r \times r$ cross-spectrum matrix $\boldsymbol{S}$, can write

$$\boldsymbol{x}_i^* \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_i = tr\left\{\boldsymbol{\Sigma}^{-\mathbf{1}} \boldsymbol{x}_i \boldsymbol{x}_i^*\right\}$$

Therefore can write $d_q(\boldsymbol{x}_i)$ as

$$\begin{aligned} d_q(\boldsymbol{x}_i) &= \boldsymbol{x}_i^* \boldsymbol{\Sigma}_1^{-1} \boldsymbol{x}_i - \boldsymbol{x}_i^* \boldsymbol{\Sigma}_2^{-1} \boldsymbol{x}_i \\ &= tr\left\{(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1}) \boldsymbol{x}_{\boldsymbol{i}} \boldsymbol{x}_{\boldsymbol{i}}^*\right\} \end{aligned}$$

Replace $\boldsymbol{x}_{\boldsymbol{i}} \boldsymbol{x}_{\boldsymbol{i}}^*$ with $\frac{1}{k} \sum_i^k \boldsymbol{x}_{\boldsymbol{i}} \boldsymbol{x}_{\boldsymbol{i}}^* = \boldsymbol{S}_i$

$$d_q(\boldsymbol{S}_i) = tr\left\{(\boldsymbol{\Sigma}_{\mathbf{1}}^{-1} - \boldsymbol{\Sigma}_2^{-1}) \boldsymbol{S}_i\right\}, \quad i = 1, ..., n_i$$

Canonical Correlations

Wish to evaluate the correlation between 2 sets of recording channels. Ex. max correlation of ch 1-4 and 5-6.

For p channels in first set and q in second set, maximize correlation between u and v.

$$\boldsymbol{a}_i'\boldsymbol{X_1} = u \quad v = \boldsymbol{b}_i'\boldsymbol{X_2}$$

Will obtain $s = min(p, q)$ orthogonal combinations:

$$\begin{aligned} \boldsymbol{a}_i'\boldsymbol{X_1} = u_1 &\quad v_1 = \boldsymbol{b}_i'\boldsymbol{X_2} \\ \vdots \quad &\quad \quad \vdots \\ \boldsymbol{a}_s'\boldsymbol{X_1} = u_s &\quad v_s = \boldsymbol{b}_s'\boldsymbol{X_2} \end{aligned}$$

Canonical Correlation Analysis (cont'd)

Solution: $max\ r^2_{u_j,v_j}$ are eigenvalues of $\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}$.
The eigenvectors of $\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}$ and $\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}$ are the coefficients $\boldsymbol{a}_j$ and $\boldsymbol{b}_j$ for the corresponding u_j and v_j.

Example:

Partition of $\boldsymbol{R}$:

$$\boldsymbol{R} = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array} \begin{array}{c} \begin{array}{cccc|cc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array} \\ \left[\begin{array}{c|c} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \hline \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{array}\right] \end{array}$$

Canonical Correlation Analysis (cont'd)

R Output for $R_{11}^{-1} R_{12} R_{22}^{-1} R_{21}$:

```
> round(R,2)
            ch1         ch2         ch3         ch4         ch5         ch6
ch1  1.00+0.00i -0.04+0.11i -0.03+0.14i -0.13+0.19i -0.01+0.48i -0.01-0.07i
ch2 -0.04-0.11i  1.00+0.00i -0.23-0.06i  0.66+0.23i  0.39+0.38i  0.45+0.18i
ch3 -0.03-0.14i -0.23+0.06i  1.00+0.00i -0.16+0.14i  0.01+0.13i -0.03+0.08i
ch4 -0.13-0.19i  0.66-0.23i -0.16-0.14i  1.00+0.00i  0.50+0.31i  0.40-0.32i
ch5 -0.01-0.48i  0.39-0.38i  0.01-0.13i  0.50-0.31i  1.00+0.00i -0.10+0.03i
ch6 -0.01+0.07i  0.45-0.18i -0.03-0.08i  0.40+0.32i -0.10-0.03i  1.00+0.00i
> e.out <- eigen( solve(R[1:4,1:4]) %*% R[1:4,5:6] %*% solve(R[5:6,5:6])
                       %*% R[5:6,1:4] )
> lambda1 <- round(e.out[[1]],2)
> V1 <- round(e.out[[2]], 2)
> lambda1
[1] 0.89+0i 0.31+0i 0.00+0i 0.00+0i
> V1
            [,1]        [,2]        [,3]        [,4]
[1,] 0.33+0.26i -0.20+0.32i -0.54+0.04i -0.23+0.26i
[2,] 0.24+0.53i  0.65+0.00i -0.31-0.45i -0.48+0.17i
[3,] 0.15+0.22i  0.07+0.16i  0.12-0.21i  0.63+0.00i
[4,] 0.64+0.00i -0.62+0.18i  0.59+0.00i  0.23-0.41i
```

Output for $R_{22}^{-1} R_{21} R_{11}^{-1} R_{12}$:

```
> e.out <- eigen( solve(R[5:6,5:6]) %*% R[5:6,1:4] %*% solve(R[1:4,1:4])
                    %*% R[1:4,5:6] )
> lambda2 <- round(e.out[[1]],2)
> V2 <- round(e.out[[2]], 2)
> lambda2
[1] 0.89+0i 0.31+0i
> V2
           [,1]       [,2]
[1,] 0.79+0.00i -0.3+0.52i
[2,] 0.37+0.49i  0.8+0.00i
>
> round(sqrt(Re(lambda1)),2)
[1] 0.94 0.56 0.00 0.00
```

Linkage for largest canonical correlation (.94):

```
> round(Mod(V1[,1]),2)
[1] 0.42 0.58 0.27 0.64
> round(Mod(V2[,1]),2)
[1] 0.79 0.61
```

Artifacts, Outliers and Cluster Analysis

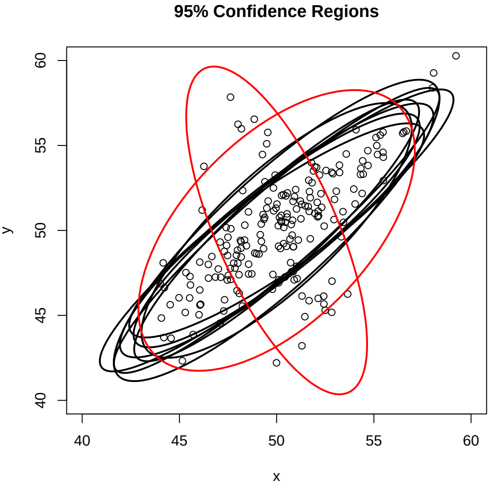
Artifacts/Outliers:

1. Spikes
2. Change in frequency
3. Misclassified individuals

Result in biased spectra

How to remove effects?

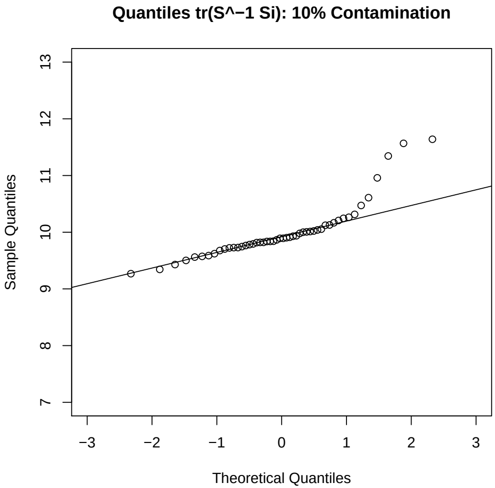
Matrix Outliers:



Resampling methods: Minimum Covariance Determinant (MCD)

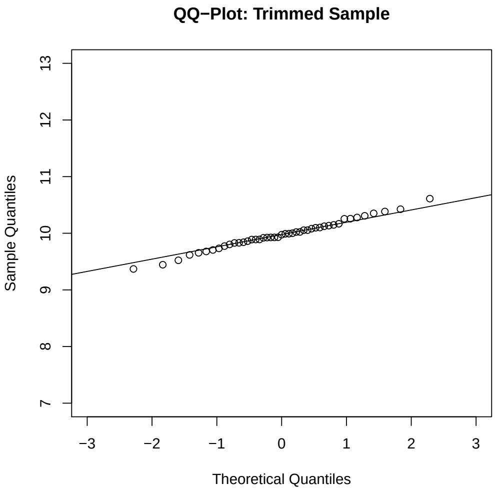
1. Draw a subsample of matrices and compute the mean spectral matrix of subsample, $\boldsymbol{S_{AVG}}$
2. Calculate $d_i = tr\{\boldsymbol{S}_{\boldsymbol{AVG}}^{-1}\boldsymbol{S}_i\}, i = 1, ..., n$
3. Drop largest $d_i's$
4. Calculate new $\boldsymbol{S_{AVG}}$ with largest $d_i's$ removed
5. Compare determinant of new $\boldsymbol{S_{AVG}}$ with previous $\boldsymbol{S_{AVG}}$ and if smaller replace with new $\boldsymbol{S_{AVG}}$
6. repeat

N=50 10×10 **matrices, 10% contamination**



Artifacts, Outliers and Cluster Analysis (cont'd):

10% Trimmed Sample with MCD



Principal Components

Purpose:

1. Data Reduction
2. Identify groups of correlated leads

Complex-valued coherency matrix, $\boldsymbol{R}$ decompose as

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^*$$

$\boldsymbol{\Lambda}$: $r \times r$ diagonal matrix of eigenvalues (real valued)
$\boldsymbol{V}$: $r \times r$ matrix of eigenvectors (complex-valued)
Eigenvectors are unique up to a rotation.

Use modulus of the $\boldsymbol{v}_j$ for interpretation.

Common Factor Model - Time Domain:

$$\boldsymbol{y}(t) = \boldsymbol{F}\boldsymbol{x}(t) + \boldsymbol{U}\boldsymbol{v}(t),\ \ t = 0, ..., T$$

$\boldsymbol{y}(t)$: p-vector valued observed series

$\boldsymbol{x}(t)$: $3r$-vector valued latent variables (factors)

$\boldsymbol{v}(t)$: p-vector unique latent variables

$\boldsymbol{F}$: $p \times 3r$ lead-field matrix (known constants)

$\boldsymbol{U}$: $p \times p$ diagonal matrix unique latent variable coefficients

Factor Analysis and EEG Source Location (cont'd)

Dipole Source Model (2 sources):

$$\boldsymbol{x}(t) = \begin{bmatrix} \gamma_1(t) sin(\theta_1) cos(\theta_2) \\ \gamma_1(t) sin(\theta_1) sin(\theta_2) \\ \gamma_1(t) cos(\theta_1) \\ \gamma_2(t) sin(\theta_3) cos(\theta_4) \\ \gamma_2(t) sin(\theta_3) sin(\theta_4) \\ \gamma_2(t) cos(\theta_3) \end{bmatrix}$$

Common Factor Model - Frequency Domain:

$$\boldsymbol{Z}_f = \boldsymbol{F}\boldsymbol{w}_f + \boldsymbol{U}\boldsymbol{v}_f, \ f = 0, ..., T/2 - 1$$

$$\boldsymbol{w}_f = \begin{bmatrix} w_{1,f} sin(\theta_1) cos(\theta_2) \\ w_{1,f} sin(\theta_1) sin(\theta_2) \\ w_{1,f} cos(\theta_1) \\ w_{2,f} sin(\theta_3) cos(\theta_4) \\ w_{2,f} sin(\theta_3) sin(\theta_4) \\ w_{2,f} cos(\theta_3) \end{bmatrix}$$

Factor Analysis and EEG Source Location (cont'd)

Cross-Spectrum Matrix, Σ_f:

$$
\begin{aligned}
\Sigma_f &= \boldsymbol{ZZ}' \\
&= (\boldsymbol{Fw}_f + \boldsymbol{Uv}_f)(\boldsymbol{Fw}_f + \boldsymbol{Uv}_f)' \\
&= \boldsymbol{Fw}_f \boldsymbol{w}_f' \boldsymbol{F}' + \mathbf{U_f^2}
\end{aligned}
$$

$\boldsymbol{S}$ has a Wishart distribution with log-likelihood

$$f(\boldsymbol{S}, \Sigma) = ln|\Sigma| + tr(\Sigma^{-1}\boldsymbol{S}) - ln|\boldsymbol{S}| - p$$

To estimate $w_{1,f}, w_{2,f}, \theta_{1,f}, \theta_{2,f}, \theta_{3,f}, \theta_{4,f}$ substitute $\boldsymbol{Fww'F}' + \boldsymbol{U}^2$ for Σ and minimize the log-likelihood

$$
\begin{aligned}
f(\boldsymbol{S}, \boldsymbol{Fw}_f \boldsymbol{w}_f' \boldsymbol{F}' + \mathbf{U_f^2}) = {} & ln|\boldsymbol{Fw}_f \boldsymbol{w}_f' \boldsymbol{F}' + \mathbf{U_f^2}| \\
& + tr((\boldsymbol{Fw}_f \boldsymbol{w}_f' \boldsymbol{F}' + \mathbf{U_f^2})^{-1}\mathbf{S}) - \mathbf{ln}|\mathbf{S}| - \mathbf{p}
\end{aligned}
$$

Factor Analysis and EEG Source Location (cont'd)

Source Locations:

Search over sets of dipole locations for 'best fit', i.e., where log-likelihood is a minimum.

References

Shumway, R.H. and Stoffer, D.S. (2011). *Time series analysis and its applications* (3^{rd} *ed.*). New York:Springer.

http://www.stat.pitt.edu/stoffer/tsa4/tsaEZ.pdf