General linear model: theory of linear model & advanced applications in statistics

Dr C. Pernet
Uni. of Edinburgh
Overview

• Linear algebra 2: Projection matrices
• ANOVAs using projections
• Multivariate Regressions
• Linear time invariant model (fMRI)
• A word on generalized linear model
Linear Algebra again!
Lecture 1: we write equations as matrices – using matrices we can find all of the unknown (the Bs, or regression coefficients) in 1 single operation $B = \text{pinv}(X)Y$

What we do with matrices, is to find how many of each vectors in $X$ we need to be as close as possible to $Y$ ($Y = XB + e$)
Linear Algebra and Statistics

- $Y = 3$ observations $X = 2$ regressors
- $Y = XB + E \rightarrow Y^\delta = XB$

SS total = variance in $Y$
SS effect = variance in $XB$
SS error = variance in $E$
$R^2 = \frac{SS \text{ effect}}{SS \text{ total}}$
$F = \frac{SS \text{ effect}/df}{SS \text{ error}/dfe}$

We can find $Y^\delta$ by computing $B$
Can we think of another way?
Linear Algebras: Projections

Why project? XB = Y may have no solution, the closest solution is a vector located in X space that is the closest to Y. With a bit of math we can find P = Xinv(X’X)X’
Projection and Least square

\[ y = \beta x + c \]

- \( P \) projects the points on the line
- Minimizing the distance \((^2)\)
- is projecting at perpendicular angles

\[ Y = Y^ + e \]
\[ Y^ = PY \]
\[ e = (I-P)Y \]
1 way ANOVA

- \( u_1 = \text{rand}(10,1) + 11.5; u_2 = \text{rand}(10,1) + 7.2; u_3 = \text{rand}(10,1) + 5; Y = [u_1; u_2; u_3]; \)
- \( x_1 = [\text{ones}(10,1); \text{zeros}(20,1)]; x_2 = [\text{zeros}(10,1); \text{ones}(10,1); \text{zeros}(10,1)]; x_3 = [\text{zeros}(20,1); \text{ones}(10,1)]; X = [x_1 \ x_2 \ x_3 \ \text{ones}(30,1)]; \)
- Lecture 1 solution:
  - \( B = \text{pinv}(X) \ast Y \) and \( \text{Yhat} = X \ast B \)
- Now the solution:
  - \( P = X \ast \text{pinv}(X) \) and \( \text{Yhat2} = P \ast Y \)
1 way ANOVA

- What to use? Both!
- Projections are great because we can now constrain $Y^\wedge$ to move along any combinations of the columns of $X$
- Say you now want to contrast gp1 vs gp2? $C = [1 \ -1 \ 0 \ 0]$
- Compute $B$ so we have $XB$ based on the full model $X$ then using $P(C(X))$ we project $Y^\wedge$ onto the constrained model
1 way ANOVA

- \( R = \text{eye}(Y) - P; \) % projection on error space
- \( C = \text{diag}([1 \ -1 \ 0 \ 0]); \) % our contrast
- \( C_0 = \text{eye}(\text{size}(X,2)) - C*\text{pinv}(C); \) % the opposite of \( C \)
- \( X_0 = X*C_0; \) % the opposite of \( C \) into \( X \)
- \( R_0 = \text{eye}(\text{size}(Y,1)) - (X_0*\text{pinv}(X_0)); \) % projection on \( E \)
- \( M = R_0 - R; \) % finally our projection matrix
- \( \text{SSeffect} = (\text{Betas}'*X'*M*X*\text{Betas}); \) % ~ 93.24
- \( F = (\text{SSeffect} / \text{rank}(X)-1) / (\text{SSerror} / \text{size}(Y,1)-\text{rank}(X))); \)
1 way ANOVA

- $SS_{total} = \text{norm}(Y - \text{mean}(Y)).^2$;
- $SS_{error} = \text{norm}(\text{Res} - \text{mean}(\text{Res})).^2$;
- $F = \frac{SS_{effect}}{\text{df}(C)} / \frac{SS_{error}}{\text{dfe}(X)}$
- $\text{df} = \text{rank}(C) - 1$;
- $\text{dfe} = \text{length}(Y) - \text{rank}(X)$;
Code Summary

- \( Y = XB + e \% \text{ master equation} \)
- \( B = \text{pinv}(X)Y \% \text{ betas} \)
- \( P = X\text{pinv}(X) \% \text{ projection onto } X \)
- \( R = I - P; \% \text{ projection on null}(X) \)
- \( \text{SSerror} = Y'RY; \% \text{ makes sense} \)
- \( C = \text{diag}(	ext{Contrast}) \rightarrow C0 \rightarrow R0 \)
- \( M = R - R0 \% \text{ projection on } C(X) \)
- \( \text{SSeffect} = B'X'MXB \% \text{ our effect for } C \)
- \( F = (\text{SSeffect}/\text{rank}(C)-1) / (\text{SSerror}/\text{rank}(X)-1); \)
Any ANOVAs
1 way

- $F \rightarrow C = \text{diag}([1 \ 1 \ 1 \ 0])$

How much the factor explains of the data (thus test against the mean)
How much the factor $i$ explains of the data (thus test against the other factors and the mean).
N way

- Interaction: multiply columns
- $C = \text{diag}([0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1 \ 0])$

How much the interaction explains of the data (thus test against the main effects and the mean)
Repeated Measure

- \( S \rightarrow C_s = \text{diag}([\text{ones}(10,1) \ 0 \ 0 \ 0 \ 0]) \)
- \( F \rightarrow C = \text{diag}([\text{zeros}(10,1) \ 1 \ 1 \ 1 \ 0]) \)

The specificity of repeated measures is the subject effect

Note is this model SS error is the SS subject – there is no more grand mean, but a mean per subject.
Multivariate Stats

Is this really more difficult?
Mutivariate stats

- Before we had one series of Y and a model X --> find B and various effects using C
- Now we have a set of Ys and a model X --> still want to find B and look at various effects
- IMPORTANT: the same model applies to all Ys
Multivariate regression

- We have one experimental conditions and plenty of measures – e.g. show emotional pictures and the subjects have to rate 1,2,3,4 – beside this subjective measurement the researcher measures RT, heart rate, pupil size
- Pblm: RT, heart rate, pupil size are likely to be correlated so doing 3 independent tests is not very informative.
Multivariate regression

- This can be solved easily since we apply the same model $X$ to all the data $Y$
- $Y = XB$ and $B = \text{pinv}(X)Y$
- $B$ is now a matrix and the coef. for each $Y$ are located on the diagonal of $B$
- $SS_{\text{total}}, SS_{\text{effect}}$ and $SS_{\text{error}}$ are also matrices called sum square (on the diagonal) and cross products matrices
Multivariate regression

- Load carbig
- \( Y = [\text{Acceleration} \ \text{Displacement}] \)
- \( X = [\text{Cylinders} \ \text{Weights} \ \text{ones(length(Weigth),1)}] \)
- Then apply the same technique as before
- Note the difference in results
- Multivariate test depends on eigen values ~ PCA
- Take a matrix and find a set orthogonal vectors (eigen vectors) and weights (eigen values) such as \( Ax = \lambda x \)

Multivariate regression

- 4 tests in multivariate analyses, all relying on the eigenvalues $\lambda$ of $\text{inv}(E)^*H$
  - Roy $\theta = \lambda_1 / 1 + \lambda_1$
  - Wilk $\Lambda = \prod(1/1 + \lambda_i)$
  - Lawley-Hotelling $U = \Sigma \lambda_i$
  - Pillai $V = \Sigma (\lambda_i / 1 + \lambda_i)$
Convolution model

Application to fMRI
General linear convolution model

- \[ y(t) = X(t)\beta + e(t) \]
- \[ X(y) = u(t) \otimes h(\tau) = \int u(t-\tau) h(\tau) \, d\tau \]

- The data \( y \) are expressed as a function of \( X \) which varies in time (\( X(t) \)) but \( \beta \) are time-invariant parameters (= linear time invariant model)

- \( X \) the design matrix describes the occurrence of neural events or stimuli \( u(t) \) convolved by a function \( h(\tau) \) with \( \tau \) the peristimulus time
General linear convolution model

- $X(y) = u(t) \otimes h(\tau)$

- Say you have a stimulus ($u$) occurring every 8/12 sec and you measure the brain activity in between ($y$)

- If you have an a priori idea of how the brain response is (i.e. you have a function $h$ which describes this response) then you can use this function instead of 1s and 0s
General linear convolution model

- $X(y) = u(t) \otimes h(\tau)$

$u = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$;
General linear convolution model

- \( X(y) = u(t) \otimes h(\tau) \)

\[
\begin{align*}
u &= [0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \; ; \\
h &= \text{spm}_\text{hrf}(1)
\end{align*}
\]
General linear convolution model

- $X(y) = u(t) \otimes h(\tau)$

$$u = [0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 ];$$

$h = \text{spm_hrf}(1)$

$X = \text{conv}(u,h);$
General linear convolution model

- $y(t) = X(t)\beta + e(t)$
- $X$ has now two conditions $u1$ and $u2$ ...
- And we search the beta parameters to fit $Y$
General linear convolution model

\[ \text{Data} = \text{cond. 1} \times \beta_1 + \text{cond. 2} \times \beta_2 + \text{cst} + e \]
General linear convolution model

\[
\begin{align*}
\text{fMRI (one voxel)} &= \text{Design matrix} \times \text{Betas} + \text{error} \\
\beta_1 + \beta_2 + e_u
\end{align*}
\]
Generalized linear model

Move between distributions
Generalized linear models

- You still have your responses $Y$ but they follow a distribution that may be binomial, poisson, gamma, etc.

- You still make a model (design matrix) $X$ and you search for a coefficient vector $\beta$

- Here in addition, there is a link function $f(.)$ such as $f(Y) = X\beta$
Generalized linear models

- Usually, $Y$ is a normally distributed response variable with a constant variance and for a simple case can be represented as a straight line ($y = c + ax$) with Gaussian distributions about each point.
In a generalized linear model, the mean of the response is modelled as a monotonic nonlinear transformation of a linear function of the predictors, $g(b_0 + b_1 x_1 + ...)$.

The inverse of the transformation $g$ is known as the "link" function.

**Gamma distributed data**

**Link function**

See glmfit
References

- **Linear algebra**
- MIT open course by prof. Strang (best course ever!)

- Elementary Linear Algebra
  By Howard Anton (best book ever! 10th edition)
References

- **Stats with matrices**
- Plane Answers to complex questions by Ronald Christensen

Methods of Multivariate Analysis by Alvin Rencher