**Matrix solution for Generalised least squares**

Firstly, let me just point out that, to the best of my knowledge, GLS is appropriate for normal data only. It is a generalisation of ordinary least squares (OLS) that ‘fixes’ non-independent residuals. From scant reading I believe there may be methods available for other response distributions (see the attached paper by Carl et al. 2007, Ecography), but I am unfamiliar with them. It’s likely that one could use the method for poisson data that has been log-transformed, since the residuals of log-transformed poisson data often behave normally. I doubt it would ever be appropriate for binomial data (but further reading may prove me wrong).

Typically in OLS regression, we have an equation for a response modelled as

y = **X**β **+** ε

We try to solve this problem

Min (y - **X**β) = Min **(**ε).

This will yield:

y = **X** β **+** ε, where ε ~ N(0, **Ω**)

ε ~ N(0, **Ω**) has mean expected value E[ε| **X** = 0], and covariance structure Cov[ε| **X**] = **Ω.**

If data points are fully independent of one another, then **Ω** = **I** σi

**I** is the indentity matrix with diagonals all = 1 and off-diagonals all = 0.

For a sample size n=3, this means:

This is what we assume when we use ordinary linear models…!!! This is because the models require the assumption that the residuals are IID: identically independently distributed

In reality the off-diagonals are hardly ever zeros, so a symmetric covariance matrix looks more like this.

In some cases the problem is severe and systematic and needs to be corrected.

Common problems where we encounter residual dependence include:

Clustered data/ hierarchical data (addressed using mixed models)

Spatial data

Temporal data

Phylogenetic data

When we have this problem, the model looks like this after regression:

Y = **X** β **+** ε, where E[ε| X = 0], Var[ε| X] = **Ω**, **Ω** ≠ **I**

To fix this, we need to apply matrix transformations to restore independence.

IF the covariance matrix is non-singular then it can be inverted (i.e. **Ω-1** exists) and, because it is symmetric, it can be decomposed into non-singular matrices:

**Ω** = **C C­T** , where **C** and **C**T can also be inverted.

Importantly: **C**-1 **Ω** (**C­**-1)T = (**C**-1**C) (C­T**(**C­**-1)T) = **I I** = **I**

This means that if we transform the entire linear relationship by **C**­-1.

**C**­-1y = **C**­-1**X** β **+ C**­-1ε, Var[ε| **C**­-1X] = **C**-1 **Ω** (**C­**-1)T = I

If we work this calculation through with a lot of beautiful matrix multiplication… then:

(**C**­-1)T **C**­-1y = (**C**­-1)T **C**­-1**X** β **+** (**C**­-1)T **C**­-1ε

(**C**­-1)T **C**­-1y = (**C**­-1)T **C**­-1**X** β **+** (**C**­-1)T **C**­-1ε where (**C**­-1)T**C**­-1 = **Ω-1**

**Ω-1**y = **Ω-1X** β **+ Ω-1**ε

**XTΩ-1**y = **XT Ω-1X** β **+ XT Ω-1**ε

Now to get β alone, multiply by the inverse of **XT Ω-1X** :

**(XT Ω-1X)-1 XTΩ-1**y = (**XT Ω-1X**)-1 (**XT Ω-1X)** β **+** (**XT Ω-1X**)-1 **XT Ω-1**ε

Now since, E(ε) = 0, (**XT Ω-1X**)-1 **XT Ω-1**ε = 0

So:

**(XT Ω-1X)-1 XTΩ-1**y = β … QED ;-)

It turns out that when this problem is solved analytically, the solution for the betas is a straightforward extension of the OLS solution:

The OLS solution:

The GLS solution:

It has further been shown that the estimated β are BLUE (best linear unbiased estimators).

Please note: the shape of **Ω** can be imposed based on your understanding of the likely autocorrelative relationship. e.g.

1. using a distance matrix derived from a phylogeny
2. using Euclidean distance matrix for spatial distances between points.
3. Using correlations between values separated in time.

NOTE: if you think back to lecture 2, this is the same solution structure as for the maximum likelihood method with an assumed covariance structure.