GLS

Srikar babu Pilli

October 2022

We have seen that when we are using the covariance structure $\Sigma(\overrightarrow{\theta})$,all we have to care about is how to specify $\Sigma(\overrightarrow{\theta})$ and we can leave the job of estimating the $\overrightarrow{\theta}$ by using the maximum likelihood to the computer.Though this is true ingeneral,there are two situations where we can use some simpler techniques than just using the brute force MLE technique and one of them is the Generalised Least Squares.

1 Generalised Least Squares(GLS)

We have the setup: $\Sigma(\overrightarrow{\theta}) = \sigma^2 P$,where σ^2 is unknown and P is a PD matrix.If $\Sigma(\overrightarrow{\theta})$ is of this form,which means $\overrightarrow{\theta}$ is just a single number σ^2,then we can fit the model much more easily in the following way:
Our model is $\overrightarrow{y} = X\overrightarrow{\beta} + \overrightarrow{\epsilon}$,where $\overrightarrow{\epsilon} \sim (\overrightarrow{0}, \sigma^2 P)$,We shall now take a square root of P,since it is a PD,we can write it as a product of some non-singular matrix(say,$P^{\frac{1}{2}}$) and it's transpose(say,$\mathbf{P^{\frac{1}{2}}}^T$) that is $P = P^{\frac{1}{2}}\mathbf{P^{\frac{1}{2}}}^T$.There are various ways by which we can get this and one way is by Cholesky decomposition,and this can be obtained by R using the function called cholesky(cholesky(P) will give us this $P^{\frac{1}{2}}$).After getting such $P^{\frac{1}{2}}$,we pre-multiply $\overrightarrow{y} = X\overrightarrow{\beta} + \overrightarrow{\epsilon}$ by it's inverse(let it be $P^{\frac{-1}{2}}$).So we get: $P^{\frac{-1}{2}}\overrightarrow{y} = P^{\frac{-1}{2}}X\overrightarrow{\beta} + P^{\frac{-1}{2}}\overrightarrow{\epsilon}$.Now,clearly $P^{\frac{-1}{2}}X$ is known and so we can consider it as our new design matrix.Similarly,we can consider $P^{\frac{-1}{2}}\overrightarrow{y}$ as our new data and $P^{\frac{-1}{2}}\overrightarrow{\epsilon}$ as our new error(let it be $\overrightarrow{\eta}$).Clearly,$\overrightarrow{\eta} \sim (\overrightarrow{0}, \sigma^2 P^{\frac{-1}{2}} P P^{\frac{-1}{2}})$ that is $\overrightarrow{\eta} \sim (\overrightarrow{0}, \sigma^2 I)$.So we have reduced it to our familiar Gauss-Markov setup and we can just proceed.So in this case,we do not really have to write down the likelihood function and try to maximize.

One special case of this is when P is a diagonal matrix with strictly positive known diagonal entries,that is if $P = diag(w_1, ..., w_n)$,then $P^{\frac{1}{2}} = diag(\sqrt{w_1}, ..., \sqrt{w_n})$ and when P is of this special form,it is called Weighted Least Squares(WLS).

This setup is a contrived one because,why should we not know σ^2 and yet we know what is the multiple of σ^2?

So,we may be using two different instruments to make the measurement,so we

believe that the two instruments are possibly of two different levels of precision.So it is okay to say one set is σ_1^2 and the other set is σ_2^2 but how are we ever going to know the ratio between them?
So we can never express one of them as σ^2 and the other as exactly $2\sigma^2$,we do not know the level of precision but somehow we know one is exactly double the level of precision as the other one,which is a contrite situation.So,this GLS is more like a theoretical interest.However it occurs as an intermediate step of a more general thing,which we shall see now.