Recap: Test of Hypothesis for a Subset of Liner Models

Dr.Arnab Chakraborty [*] Mrinmoy Banik [†]

4th October 2022

Abstract

We had previously discussed this topic of test of hypothesis regarding linear models so this initial section is basically a recap to that mathematical setup and also a intuitive geometrical understanding of how this test is generally carried out using ratios of excess residual to our acceptable residuals which comes under a subclass of other well known techniques like Likelihood Ratio Tests discussed in the upcoming sections.

1 Recap

Lets first quickly recall the setup of our linear models. The setup involved a general linear of the following form:-

$$\vec{y} = \mathbb{X}\vec{\beta} + \vec{\epsilon}$$
$$\vec{\epsilon} \sim \mathbb{N}(\vec{0}, \sigma^2\mathbb{I})$$
$$\vec{\beta} \in \mathbb{R}^p \quad , \quad \sigma^2 > 0$$

The usual assumptions like the normality assumption (i.e the errors are normally distributed.) apply here whenever we perform test of statistics in linear models.

[*]Department of Applied Statistical Division, Indian Statistical Institute,Kolkata Email:http://www.isical.ac.in/ arnabc/

[†]This portion of the book is latex-ed and edited by him,Student of B.Stat 3rd yr at ISI,BS2012

2 The Null and Alternate Hypothesis

We give a postulate that for some design matrix, $\mathbb{X}_0$ such it's column space lies in the column space of $\mathbb{X}$, i.e $\mathbf{col}(\mathbb{X}_0) \subseteq \mathbf{col}(\mathbb{X})$, here $\mathbb{X}_0$ for the restricted model is completely specified in the null hypothesis. Hence the alternative hypothesis is the space spanned by $\mathbf{col}(\mathbb{X}) - \mathbf{col}(\mathbb{X}_0)$.They are given below as two linear models, the left one being our known universal model and the right one is a smaller subset of this larger class of linear models. The alternate hypothesis here is implicit in the sense that it consist of the set of linear models which are subset of the larger model with design matrix $\mathbb{X}$ but not within the class of models represented by our null hypothesis, i.e $\mathbb{X}_0$.

$$\begin{aligned} \vec{y} &= \mathbb{X}\vec{\beta} + \vec{\epsilon} \qquad & \vec{y} &= \mathbb{X}\vec{\beta} + \vec{\epsilon} \\ \vec{\epsilon} &\sim \mathbb{N}(\vec{0}, \sigma^2\mathbb{I}) \qquad & \vec{\epsilon} &\sim \mathbb{N}(\vec{0}, \sigma^2\mathbb{I}) \\ \vec{\beta} &\in \mathbb{R}^p \ , \ \sigma^2 > 0 \qquad & \vec{\beta} &\in \mathbb{R}^p \ , \ \sigma^2 > 0 \end{aligned}$$

3 Geometrical Intuition with Orthogonal Projection

Look at FIGURE 1. Our data vector is represented by the Green arrow. The pink plane on the other hand represents the general linear model with design matrix $\mathbb{X}$ while the blue line along that plane represents the null hypothesis with design matrix $\mathbb{X}_0$.

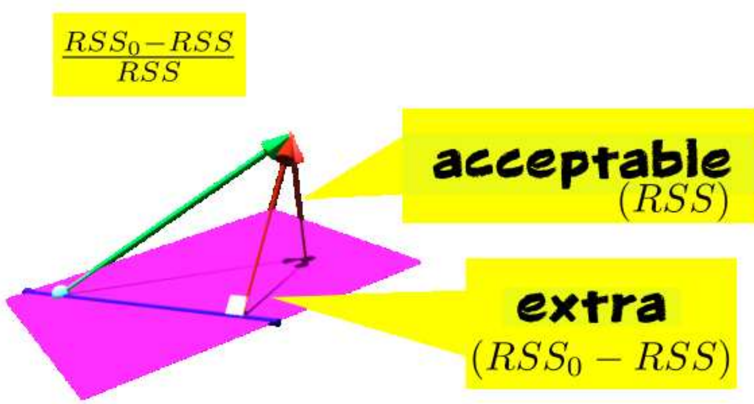

FIGURE 1: Green vector is the data vector, the pink plane is the linear model and blue line is the null hypothesis.

Now we already know that the general linear model is a good fit, i.e the data vector always lies close to the pink plane but the null hypothesis claims it indeed lies close enough to the blue line on that pink plane, that is what we are testing.

The way we carryout this test is as before by comparing the lengths of the residual sum of squares(the red arrows).The smaller one here represents the total sum of squares or **RSS** while the slanted larger red arrow represents the residual sum of squares under the null hypothesis.($\mathbf{RSS_0}$). Then we take a ratio of the extra sum of squares by the our original yardstick i.e **RSS** and reject the null hypothesis when this ration exceeds a certain value which depends on the level of significance of our test.

4 Conclusion

Here in this initial section we saw how we can test for sub-spaces of linear models by using simple geometrical intuitions. They actually are part of a wider class of well know statistical tests known as **L**ikelihood **R**atio **T**est or **LRT** wich we will study in the next section.

[——————]