Main results

Souvik Roy

October 6, 2022

1 Introduction

We now are prepared to discuss the distribution of the Residual Sum of Squares given the normality assumption. We shall use the following results discussed previously to obtain our distribution.

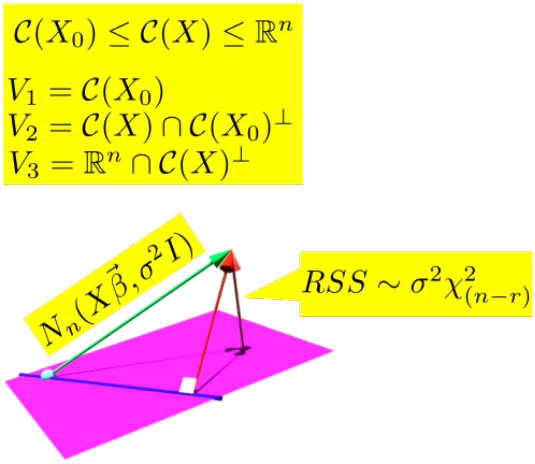
1. If $\vec{X} \sim N_n(\vec{\mu}, \sigma^2 I)$, and R is an orthogonal matrix, then, $R\vec{X} \sim N_n(R\vec{\mu}, \sigma^2 I)$.

2. If $\mathbb{R}^n = V_1 \oplus \cdots \oplus V_k$, with all the V_i's pairwise orthogonal, and $\vec{Y} = P_{V_1}\vec{Y} + P_{V_2}\vec{Y} + \cdots + P_{V_k}\vec{Y}$, then, $||P_{V_i}\vec{Y}||^2 \sim \sigma^2\chi^2_{(r)}(||P_{V_i}\vec{\mu}||^2)$ if r is the rank of the projector matrix P_{V_i}. To see this, we observe that we can express $\vec{Y}$ as a linear combination of the canonical basis vectors, and since P_{V_i} is a projector matrix, its rows form an orthonormal basis, and thus, precisely $n - r$ of the canonical basis vectors are orthogonal to the rows of P_{V_i} and thus, upon premultiplying this matrix to $\vec{Y}$, only r non zero components remain, all of which, using result 1, are independent and follow normal with the same variance σ^2. Hence, their norm squared follows $\sigma^2\chi^2_{(r)}(||P_{V_i}\vec{\mu}||^2)$.
   We shall use these results to obtain the distribution for the residual sum of squares.

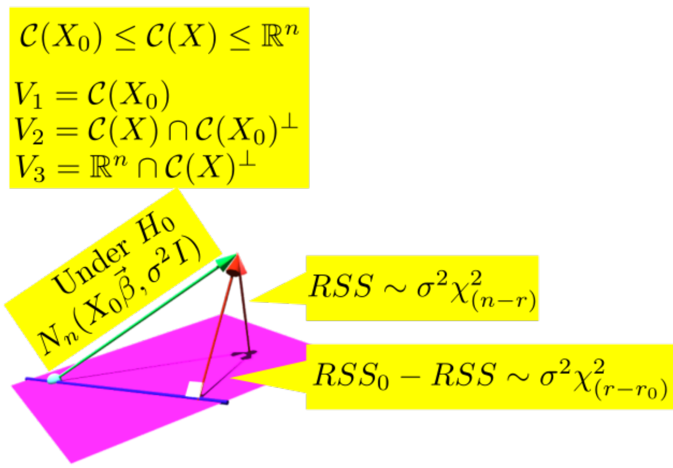
2 Main results

We observe that to apply the second result stated above, we shall write $\mathbb{R}^n$ as the direct sum of orthogonal subspaces. We already have our original model $\vec{Y} = X\vec{\beta} + \vec{\varepsilon}$, $\vec{Y} \sim N_n(X\vec{\beta}, \sigma^2 I)$ and the model under the null hypothesis, $\vec{Y} = X_0\vec{\beta} + \vec{\varepsilon}$, $\vec{Y} \sim N_n(X_0\vec{\beta}, \sigma^2 I)$, with $\mathcal{C}(X_0) \subseteq \mathcal{C}(X) \subseteq \mathbb{R}^n$. Now, we consider the three pairwise orthogonal subspaces, $V_1 = \mathcal{C}(X_0)$, $V_2 = \mathcal{C}(X) \cup \mathcal{C}(X_0)^\perp$ and $\mathcal{C}(X)^\perp$. Clearly, the direct sum of these subspaces forms $\mathbb{R}^n$, and they are pairwise orthogonal. Now, we observe that if the rank of X is r, then the dimension of $\mathcal{C}(X)^\perp$ is $n - r$. We observe that since $\hat{\vec{Y}}$ lies in $\mathcal{C}(X)$, and the residual vector is orthogonal to $\hat{\vec{Y}}$, it follows that the residual vector lies in $\mathcal{C}(X)^\perp$ which has dimension $n - r$, and hence, using result 1, we obtain that the norm of the residual vector, which indeed is the Residual Sum of Squares(RSS) since the residuals have mean 0, follows $\sigma^2\chi^2_{(n-r)}(||P_{V_3}X\vec{\beta}||^2)$. But since V_3 is orthogonal to $\mathcal{C}(X)$, it follows that, $P_{V_3}X\vec{\beta} = \vec{0}$, and thus, the chi-squared distribution above becomes a central chi-squared distribution $\sigma^2\chi^2_{(n-r)}$. We observe that this is true irrespective of the value of X_0.

The following diagram shows the case for $n = 3$.



Now, suppose the null hypothesis is true. In this case, we shall look at the residual sum of squares under the null model, and call it RSS_0. We shall use the subscript 0 to denote the corresponding quantities under the null model. Now, observe that the least squares estimate under the null model is the projection of $\vec{Y}$ onto the blue line which denotes the column space of X_0, and hence, $\hat{\vec{\varepsilon_0}}$ lies outside V_1. Now, we also observe that $\hat{\vec{Y_0}} - \hat{\vec{Y}}$ lies in $\mathcal{C}(X)$ and outside $\mathcal{C}(X_0)$, and thus, it must lie in V_2. Moreover, the norm squared of $\hat{\vec{Y_0}} - \hat{\vec{Y}}$ is nothing but $RSS_0 - RSS$, and hence, we find that the distribution under the null, of $RSS_0 - RSS$ is $\sigma^2\chi^2_{(r-r_0)}(||P_{V_2}X_0\vec{\beta}||^2)$, where r_0 is the rank of X_0. We observe that since $V_2 = \mathcal{C}(X) \cup \mathcal{C}(X_0)^\perp$, $P_{V_2}X_0\vec{\beta} = \vec{0}$ and thus, the above distribution also becomes a central chi-squared distribution $\sigma^2\chi^2_{(r-r_0)}$. The following diagram again demonstrates the case for $n = 3$.



It is important to observe that if one does not assume the null hypothesis to be true, the latter distribution is **not** a central chi-squared, but the former is still a central chi-squared under the normality

assumptions.
Now, since, the quantities $\hat{\vec{Y}}_0 - \hat{\vec{Y}}$ and $\vec{Y} - \hat{\vec{Y}}$ are orthogonal, it follows that they are stochastically independent (that is independent in probability), and thus, their ratio after a scaling suitablly, shall follow a central F distribution, with certain degrees of freedom. This is the null distribution that we shall perform our test with. If the the null is not true however, we shall have a non central F distribution, since one of the distributions will not be a central chi-squared.