Lemma: Let $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ be an orthonormal basis for $\mathbb{R}^n$, where $\boldsymbol{B} := (\boldsymbol{b}_1, \dots, \boldsymbol{b}_{n-p})$ contains a basis for $\boldsymbol{X}$'s null space (kernel). If $\boldsymbol{c}$ is random vector that satisfies:

$$\boldsymbol{B}\boldsymbol{c} \sim \mathcal{N}_n\big(\boldsymbol{0}, \sigma^2(\boldsymbol{I}_n - \boldsymbol{H})\big) \tag{1}$$

then

$$\boldsymbol{c} \sim \mathcal{N}_{n-p}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{n-p}) \tag{2}$$

where $\boldsymbol{H}$ denotes the hat matrix.

Proof: Let $\boldsymbol{c}$ be a random vector of $(n-p)$ elements. Properties of the multivariate Gaussian distribution imply that, if $\boldsymbol{B}\boldsymbol{c} \sim \mathcal{N}_n\big(\boldsymbol{0}, \sigma^2(\boldsymbol{I}_n - \boldsymbol{H})\big)$, then:

$$\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{c} \sim \mathcal{N}_{n-p}\big(\boldsymbol{0}, \sigma^2\boldsymbol{B}^T(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{B}\big) \tag{3}$$

The definition of $\boldsymbol{B}$ implies that $\boldsymbol{B}^T\boldsymbol{B} = \boldsymbol{I}_{n-p}$. Consequently,

$$\boldsymbol{c} \sim \mathcal{N}_{n-p}\big(\boldsymbol{0}, \sigma^2\boldsymbol{B}^T(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{B}\big) \tag{4}$$

By definition, $\boldsymbol{H} := \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$. $\boldsymbol{B}$'s columns are orthogonal to $\boldsymbol{X}$'s columns, so $\boldsymbol{B}^T\boldsymbol{X} = \boldsymbol{B}^T\boldsymbol{H} = \boldsymbol{0}$, meaning that the above expression can be simplified to:

$$\boldsymbol{c} \sim \mathcal{N}_{n-p}\big(\boldsymbol{0}, \sigma^2\boldsymbol{B}^T\boldsymbol{B}\big) \tag{5}$$

$$\implies \boldsymbol{c} \sim \mathcal{N}_{n-p}\big(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{n-p}\big) \tag{6}$$

Relevance to least-squares linear regression

Assume that

$$\boldsymbol{y} = \sum_{j=1}^{n} d_j \boldsymbol{b}_j \tag{7}$$

where the d_j are random variables. Since $\boldsymbol{b}_j$ lies in $\boldsymbol{X}$'s columnspace for $j = (n-p+1), \dots, n$ and is orthogonal to it otherwise, and $\boldsymbol{H}$ is a projection matrix for $\boldsymbol{X}$'s columnspace, we have:

$$\boldsymbol{H}\boldsymbol{b}_j = \begin{cases} \boldsymbol{0} & \text{for } j = 1, \dots, (n-p) \\ \boldsymbol{b}_j & \text{for } j = (n-p+1), \dots, n \end{cases} \tag{8}$$

It follows that:

$$(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y} = \sum_{j=1}^{n-p} d_j \boldsymbol{b}_j = \boldsymbol{B}\boldsymbol{d}_{1:(n-p)} \tag{9}$$

Lemma 1 then implies that if $\hat{\boldsymbol{\epsilon}} = \boldsymbol{B}\boldsymbol{d}_{1:(n-p)} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2(\boldsymbol{I}_n - \boldsymbol{H}))$, then $\boldsymbol{d}_{1:(n-p)} \sim \mathcal{N}_{n-p}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{n-p})$.

This result means that if the OLS residuals are consistent with $\mathcal{N}_n(\boldsymbol{0}, \sigma^2(\boldsymbol{I}_n - \boldsymbol{H}))$, then $\boldsymbol{y}$ can be decomposed into a centred spherical Gaussian component restricted to null($\boldsymbol{X}$) and another component, restricted to col($\boldsymbol{X}$), that can be of any distribution:

$$\boldsymbol{y} = \underbrace{\sum_{j=1}^{n-p} d_j \boldsymbol{b}_j}_{\text{Gaussian variation on null}(X)} + \underbrace{\sum_{j=n-p+1}^{n} d_j \boldsymbol{b}_j}_{\text{Unrestricted variation on col}(X)} \tag{10}$$

The assumption underpinning the normal linear model is that $\boldsymbol{y} \sim \mathcal{N}_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I}_n)$. Notice that the mean of this distribution lies on col($\boldsymbol{X}$). The residuals being consistent with this distribution verifies that the mean of $\boldsymbol{y}$ is in col($\boldsymbol{X}$) (i.e $\exists\boldsymbol{\beta} : \mathbb{E}[y] = \boldsymbol{X\beta}$), but it does not test that variation of the response in col($\boldsymbol{X}$) follows a spherical Gaussian distribution. In other words, the reverse implication very nearly holds.