ESTIMATING THE VARIANCE

DAVAR KHOSHNEVISAN

Recall the linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{1}$$

The most standard assumption on the noises is that ε_i's are i.i.d. $N(0, \sigma^2)$ for a fixed unknown parameter $\sigma > 0$. The MLE for σ^2 is

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 = \frac{1}{n}\|\boldsymbol{\varepsilon}\|^2 = \frac{1}{n}\left\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|^2. \tag{2}$$

Write $\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \mathbf{P}_{\mathscr{C}(\boldsymbol{X})}\boldsymbol{Y}$ to obtain

$$\widehat{\sigma}^2 = \frac{1}{n}\left\|\boldsymbol{Y} - \mathbf{P}_{\mathscr{C}(\boldsymbol{X})}\boldsymbol{Y}\right\|^2 = \frac{1}{n}\left\|\left(\mathbf{I}_n - \mathbf{P}_{\mathscr{C}(\boldsymbol{X})}\right)\boldsymbol{Y}\right\|^2. \tag{3}$$

Lemma 0.1. *If S denotes a subspace of $\boldsymbol{R}^n$, then $\mathbf{I}_n - \mathbf{P}_S = \mathbf{P}_{S^\perp}$, where $S^\perp$ denotes the orthogonal complement to S; i.e.,*

$$S^\perp = \left\{\boldsymbol{x} \in \boldsymbol{R}^n : \ \boldsymbol{x} \perp S\right\}. \tag{4}$$

Proof. First, let us check that if $\boldsymbol{x} \in \boldsymbol{R}^n$ then $(\mathbf{I}_n - \mathbf{P}_S)\boldsymbol{x}$ is orthogonal to $\mathbf{P}_S\boldsymbol{x}$. Indeed,

$$\begin{aligned} \left[\left(\mathbf{I}_n - \mathbf{P}_S\right)\boldsymbol{x}\right]'\mathbf{P}_S\boldsymbol{x} &= \left[\boldsymbol{x}' - \boldsymbol{x}'\mathbf{P}_S'\right]\mathbf{P}_S\boldsymbol{x} \\ &= \boldsymbol{x}'\mathbf{P}_S\boldsymbol{x} - \boldsymbol{x}'\mathbf{P}_S^2\boldsymbol{x}, \end{aligned} \tag{5}$$

because $\mathbf{P}_S' = \mathbf{P}_S$. Since $\mathbf{P}_S^2 = \mathbf{P}_S$, it follows that $(\mathbf{I}_n - \mathbf{P}_S)\boldsymbol{x}$ is orthogonal to $\mathbf{P}_S\boldsymbol{x}$, as promised.

Next, let us prove that $\mathbf{I}_n - \mathbf{P}_S$ is idempotent; i.e., a projection matrix. This too is a routine check, viz.,

$$\left(\mathbf{I}_n - \mathbf{P}_S\right)^2 = \mathbf{I}_n - 2\mathbf{P}_S + \mathbf{P}_S^2 = \mathbf{I}_n - \mathbf{P}_S, \tag{6}$$

as claimed.

We have shown, thus far, that $\mathbf{I}_n - \mathbf{P}_S$ is a projection matrix, and it projects $\boldsymbol{x} \in \boldsymbol{R}^n$ to some point in $S^\perp$. Thus, there exists a subspace T of $\boldsymbol{R}^n$ such that $\mathbf{I}_n - \mathbf{P}_S = \mathbf{P}_T$. It remains to verify that $T = S^\perp$; this follows from the fact that any $\boldsymbol{x} \in \boldsymbol{R}^n$ can be written as $\boldsymbol{x} = \mathbf{P}_S\boldsymbol{x} + (\mathbf{I}_n - \mathbf{P}_S)\boldsymbol{x}$. □

In summary, we have shown that

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \mathbf{P}_{\mathscr{C}(\boldsymbol{X})}\boldsymbol{Y}, \\ \widehat{\sigma^2} &= \frac{1}{n}\left\|\mathbf{P}_{\mathscr{C}(\boldsymbol{X})^\perp}\boldsymbol{Y}\right\|^2. \end{aligned} \tag{7}$$

Date: August 30, 2004.

It will turn out that if $S \perp T$—and under the assumption that the ε_i's are i.i.d. normals—then $\mathbf{P}_S\boldsymbol{Y}$ is *statistically independent* of $\mathbf{P}_T\boldsymbol{Y}$. Therefore, in particular, we will see soon that, in the normal-errors model,

$$\widehat{\boldsymbol{\beta}} \text{ and } \widehat{\sigma^2} \text{ are independent.} \tag{8}$$