A basis for V=span(X)

stats545.2 545.2 The multivariate normal distribution and projections in the linear model

Gunnar Stefansson

September 5, 2022

Subspaces

In the model $y = X\boldsymbol{\beta} + e$, assume $rank(X) = r$ where X is $n \times p$ and $r \leq p$

Recall $X\hat{\boldsymbol{\beta}}$ is a project so $y - X\hat{\boldsymbol{\beta}} \perp X\hat{\boldsymbol{\beta}}$ so that $y - X\hat{\boldsymbol{\beta}} \in V^{\perp} = \{v : v \perp sp(X)\}$ and $dim(V^{\perp}) = n - r$.
If $r = p$, then:

$$\underbrace{\hat{e}}_{n\times 1} = y - X\hat{\boldsymbol{\beta}} = y - X(X'X)^{-1}X'y$$

$$= (I - X(X'X)^{-1}X')y = (I - H)y$$

and $rank(I - H) = dim(V^{\perp}) = n - p$
Assume $rank(X) = r$ (X is $n \times p$ and usually $p = r$ but $r < p$ is common also).
The model $y = X\boldsymbol{\beta} + e$ is estimated using the projection $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y$ onto the subspace $sp(X)$.
We have $y - X\hat{\boldsymbol{\beta}} \perp X\hat{\boldsymbol{\beta}}$ svo $y - X\hat{\boldsymbol{\beta}} \in V^{\perp} = \{v : v \perp sp(X)\}$ and $dim(V^{\perp}) = n - r$.

$$\hat{e} = y - X\hat{\boldsymbol{\beta}} = y - X(X'X)^{-1}X'y$$

A basis for the span of X

Orthonormal basis, $\{u_1, \ldots, u_n\}$ for R^n:

Using Gram-Schmidt, first generate $u_1, \ldots, u_r$ which span $sp\{X\}$, with $rank\{X\} = r$ and the rest, $u_{r+1}, \ldots, u_n$ are chosen so that the entire set, $u_1, \ldots, u_n$ spans R^n.

$$
\begin{aligned}
X\hat{\beta} &= \hat{\zeta}_1 u_1 + \ldots \hat{\zeta}_r u_r \\
y &= \hat{\zeta}_1 u_1 + \ldots \hat{\zeta}_r u_r + \hat{\zeta}_{r+1} u_{r+1} + \ldots + \hat{\zeta}_n u_n
\end{aligned}
$$

Q-R decomposition

$$Q := \left[u_1 \vdots u_2 \vdots \ldots \vdots u_n \right]$$

is the Q in the Q-R decomposition of $X = QR$.

If

$$z = \left(\hat{\zeta}_1, \hat{\zeta}_2, \ldots, \hat{\zeta}_n \right)$$

then

$$z = Q'y$$

and hence

$$E\left[z\right] = Q'X\beta$$

$$V\left[z\right] = Q'\sigma^2 IQ = \sigma^2 I$$

Variances of coefficients

For each i we obtain

$$V\left[\hat{\zeta}_i\right] = \sigma^2$$

Expected values of coefficients

For $i = r + 1, \dots, n$ we obtain

$$E\left[\hat{\zeta}_i\right] = 0$$

Sums of squares and norms

$$SSE(F) = ||\mathbf{y} - \mathbf{X}\hat{\beta}||^2 \quad = \sum_{i=p+1}^{n} \hat{\zeta}_i^2$$

Each $\hat{\zeta}_i$ is a coordinate in an orthonormal basis, $\hat{\zeta}_i = \mathbf{y} \cdot \mathbf{u}_i$. When y_i are independent Gaussian random variables, $\hat{\zeta}_i$ also become independent. As a result, the sum of squares is related to a $\sigma^2 \cdot \chi^2$-distribution in a natural way.

Normality and independence of coeffients

Note that $\hat{\zeta}_i$ are linear combinations of the various y_j since $\hat{\zeta}_i = \mathrm{u}_i \cdot \mathrm{y}$.

When the y_i are independent Gaussian random variables, $\hat{\zeta}_i$ have zero covariance and are thus also independent.

Each $\hat{\zeta}_i$ is a coordinate in an orthonormal basis, $\hat{\zeta}_i = \mathrm{y} \cdot \mathrm{u}_i$. When Y_i are independent Gaussian random variables, $\hat{\zeta}_i$ also become independent. As a result, the sums of squares are related to $\sigma^2 \cdot \chi^2$-distributions in a natural way.

Degrees of freedom

$SSE(F)$has $n - r$ degrees of freedom.