

Orthogonal projections in multiple regression

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

Gunnar Stefansson

August 28, 2020

Subspaces

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

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

If $r = p$, then:

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

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

and $\text{rank}(\mathbf{I} - \mathbf{H}) = \dim(\mathbf{V}^\perp) = n - p$

Assume $\text{rank}(\mathbf{X}) = r$ (\mathbf{X} is $n \times p$ and usually $p = r$ but $r < p$ is common also).

The model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ is estimated using the projection $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ onto the subspace $\text{sp}(\mathbf{X})$.

We have $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \perp \mathbf{X}\hat{\boldsymbol{\beta}}$ svo $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \in \mathbf{V}^\perp = \{v : v \perp \text{sp}(\mathbf{X})\}$ and $\dim(\mathbf{V}^\perp) = n - r$.

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

Projections and bases

If \mathbf{a} is a vector then we can write a general vector \mathbf{w} in the form $\mathbf{w} = \mathbf{x} + \mathbf{y}$ where $\mathbf{x} = k\mathbf{a}$ and $\mathbf{a}'\mathbf{y} = \mathbf{a} \cdot \mathbf{y} = 0$.

In the general case,

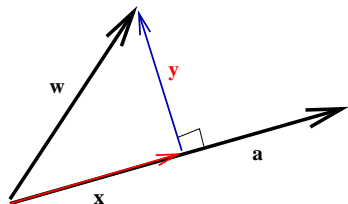
$$k = \frac{\mathbf{w} \cdot \mathbf{a}}{\|\mathbf{a}\|^2},$$

and for unit vectors \mathbf{a} we obtain

$$k = \mathbf{w} \cdot \mathbf{a}.$$

The Gram-Schmidt technique uses these projections to iteratively build an orthonormal basis, $\mathbf{u}_1, \dots, \mathbf{u}_r$ which spans the same space as a sequence of arbitrary starting vectors, $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p$.

In linear regression, the starting vectors are



A basis for the span of \mathbf{X}

Orthonormal basis, $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for \mathbf{R}^n :

Using Gram-Schmidt, first generate $\mathbf{u}_1, \dots, \mathbf{u}_r$ which span $sp\{\mathbf{X}\}$, with $rank\{\mathbf{X}\} = r$ and the rest, $\mathbf{u}_{r+1}, \dots, \mathbf{u}_n$ are chosen so that the entire set, $\mathbf{u}_1, \dots, \mathbf{u}_n$ spans \mathbf{R}^n .

$$\begin{aligned}\mathbf{X}\hat{\boldsymbol{\beta}} &= \hat{\zeta}_1\mathbf{u}_1 + \dots + \hat{\zeta}_r\mathbf{u}_r \\ \mathbf{y} &= \hat{\zeta}_1\mathbf{u}_1 + \dots + \hat{\zeta}_r\mathbf{u}_r + \hat{\zeta}_{r+1}\mathbf{u}_{r+1} + \dots + \hat{\zeta}_n\mathbf{u}_n\end{aligned}$$

Q-R decomposition

$$\mathbf{Q} := \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n \end{bmatrix}$$

is the \mathbf{Q} in the Q-R decomposition of $\mathbf{X} = \mathbf{QR}$.

If

$$\mathbf{z} = (\hat{\zeta}_1, \hat{\zeta}_2, \dots, \hat{\zeta}_n)$$

then

$$\mathbf{z} = \mathbf{Q}'\mathbf{y}$$

and hence

$$E[\mathbf{z}] = \mathbf{Q}'\mathbf{X}\boldsymbol{\beta}$$

$$V[\mathbf{z}] = \mathbf{Q}'\sigma^2\mathbf{I}\mathbf{Q} = \sigma^2\mathbf{I}$$

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

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

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 coefficients

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

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, $\hat{\zeta}_i = \mathbf{y} \cdot \mathbf{u}_i$. 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.

References Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. 1996. Applied linear statistical models. McGraw-Hill, Boston. 1408pp.

Copyright 2020, Gunnar Stefansson

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit

<http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.