02403 Introduction to Mathematical Statistics

Lecture 5: Cochrans theorem and the general linear model

DTU Compute
Technical University of Denmark
2800 Kgs. Lyngby

Agenda

1. Matrix formulation of statistical models
2. The χ^2-distribution and the multivariate normal
3. Projections and Cochrans theorem
4. The general linear model

The Aim

- Formulate simple statistical model in matrix notation
- Establish the connection between the multivariate normal and the χ^2-distribution
- Establish and (partially) prove Cochrans theorem
- Use Cochrans theorem to design test strategies

Overview

1. Matrix formulation of statistical models
2. The χ^2-distribution and the multivariate normal
3. Projections and Cochrans theorem
4. The general linear model

Example: Model

Assume that you plan take (small) sample with two observations Y_1 and Y_2 with Y_i iid. and $N(\mu, \sigma^2)$-distributed. The estimator for the mean is $\hat{\mu} = \frac{1}{2}(Y_1 + Y_2)$. Show that the model can be written as

$$\boldsymbol{Y} = \mathbf{1}\mu + \boldsymbol{\epsilon}; \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\sigma}^2 \boldsymbol{I}).$$

Further show that inserting the estimators we can write the model as

$$\boldsymbol{Y} = \mathbf{1}\hat{\mu} + \boldsymbol{r} = \boldsymbol{A}\boldsymbol{Y} + (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{Y}$$

with $\boldsymbol{A}\boldsymbol{Y} = \mathbf{1}\hat{\mu}$.

- We will be interested in the distribution related to $\boldsymbol{A}\boldsymbol{Y}$ and $(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{Y}$.

Overview

The χ^2-distribution and the multivariate normal

Recall that if $Z_i \sim N(0,1)$, $i \in \{1,...,n\}$ and iid. then

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$$

it follow directly that if $\boldsymbol{Z} \sim N_n(\mathbf{0}, \boldsymbol{I})$ then

$$\boldsymbol{Z}^T \boldsymbol{Z} \sim \chi^2(n)$$

and further that (Corollary 9.18), if $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then

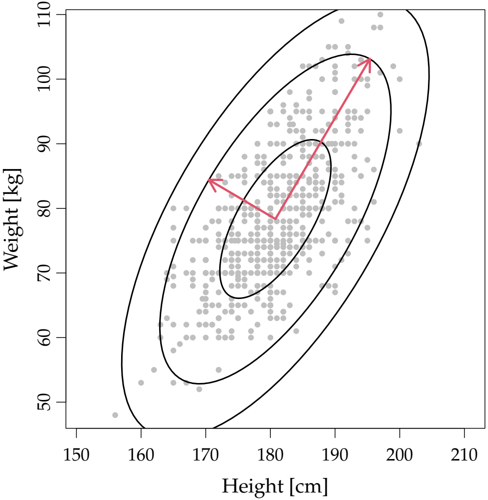
$$(\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y} - \boldsymbol{\mu}) \sim \chi^2(n).$$

The result can be used to draw probability regions.

Example

Let $[Y_1, Y_2]^T \sim N(\mathbf{0}, \boldsymbol{I})$, define $Z = Y_1 + Y_2$. For the random vector $[Y_1, Z]$ write down $\boldsymbol{\Sigma}$ and draw a 95% probability region.

Probability regions Normal distribution



A test statistics

Assume that you plan take (small) sample with two observations Y_1 and Y_2 with Y_i iid. and $N(\mu, \sigma^2)$-distributed. The estimator for the mean is $\hat{\mu} = \frac{1}{2}(Y_1 + Y_2)$. The model can be written as

$$\boldsymbol{Y} = \mathbf{1}\mu + \boldsymbol{\epsilon}; \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\sigma}^2 \boldsymbol{I}).$$

Assuming that $\mu = 0$ what is the distribution of

$$F = \frac{2\bar{Y}^2}{\boldsymbol{r}^T \boldsymbol{r}}.$$

Overview

Orthogoinal Projections

Definition (Orthogonal projections)

A matrix $\boldsymbol{P}$ is an orthogonal projection matrix if and only if

- $\boldsymbol{P}$ is symmetric, i.e. $\boldsymbol{P} = \boldsymbol{P}^T$
- $\boldsymbol{P}$ is idempotent, i.e. $\boldsymbol{P}^2 = \boldsymbol{P}$.

Orthogonal Projections: Properties

Lemma (Properties of orthogonal projection matrices)

If $\boldsymbol{P}$ is an orthogonal projection matrix, then

1. *The eigenvalues λ_i of $\boldsymbol{P}$ are either 0 or 1, and $Rank(\boldsymbol{P}) = \sum_i \lambda_i$.*
2. $Rank(\boldsymbol{P}) = Trace(\boldsymbol{P})$.

If $\boldsymbol{P}$ is a projection matrix then $\boldsymbol{I} - \boldsymbol{P}$ is also a projection matrix.

Example: Model

Assume that you plan take (small) sample with two observations Y_1 and Y_2 with Y_i iid. and $N(\mu, \sigma^2)$-distributed. The estimator for the mean is $\hat{\mu} = \frac{1}{2}(Y_1 + Y_2)$. Show that the model can be written as

$$\boldsymbol{Y} = \mathbf{1}\mu + \boldsymbol{\epsilon}; \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}).$$

Further show that inserting the estimators we can write the model as

$$\boldsymbol{Y} = \mathbf{1}\hat{\mu} + \boldsymbol{r} = \boldsymbol{A}\boldsymbol{Y} + (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{Y}$$

with $\boldsymbol{A}\boldsymbol{Y} = \mathbf{1}\hat{\mu}$.

- Refering to the example above show that $\boldsymbol{A}$ and $\boldsymbol{I} - \boldsymbol{A}$ are both orthogonal projection matrices.
- In the example above find the Rank of $\boldsymbol{A}$ and $\boldsymbol{I} - \boldsymbol{A}$.

Cochrans theorem

Theorem (Cochran's theorem)

Let $\boldsymbol{Y} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$*, and let* $\boldsymbol{H}_i$ *be orthogonal projection matrices such that*

$$\frac{1}{\sigma^2}\boldsymbol{Y}^T\boldsymbol{Y} = \frac{1}{\sigma^2}\sum_{i=1}^{K}\boldsymbol{Y}^T\boldsymbol{H}_i\boldsymbol{Y}$$

i.e. $\sum_{i=1}^{K}\boldsymbol{H}_i = \boldsymbol{I}_n$*, with* $Rank(\boldsymbol{H}_i) = p_i$*, and* $\sum_i p_i = n$ *then*

1. $\frac{1}{\sigma^2}\boldsymbol{Y}^T\boldsymbol{H}_i\boldsymbol{Y} \sim \chi^2(p_i)$
2. $\boldsymbol{Y}^T\boldsymbol{H}_i\boldsymbol{Y}$ *and* $\boldsymbol{Y}^T\boldsymbol{H}_j\boldsymbol{Y}$ *are independent for* $i \neq j$*.*

Overview

The general linear model

The general linear model is a statistical model that can be written in the form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}; \quad \boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \boldsymbol{\sigma}^2\boldsymbol{I})$$

or $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\sigma}^2\boldsymbol{I})$.

- $\boldsymbol{Y}$ are the observations
- $\boldsymbol{X}$ is the design matrix
- $\boldsymbol{\beta}$ is a vector of parameters
- $\boldsymbol{\epsilon}$ are the residuals

The general linear model: Some questions

- How do we construct the design matrix $\boldsymbol{X}$, and is it unique?
- How do we estimate $\boldsymbol{\beta}$?
- What is the best estimate of the residual variance σ^2?

Example: The design matrix

Two items A and B are weighted on a balance, first separately then together, giving the observations y_1, y_2, y_3. Assume measurement erros are iid. normal.
Write the design matrix when the parameter interpretation is

- β_1 is the weight of item 1 and β_2 is the weight of item 2.
- β_1 is the weight of item 1 and β_2 is the difference in weight between item 1 and 2.
- β_1 is average weight of item 1 and 2, and β_2 deviation between the average and the individual weights.

Least square estimator

For the general linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}; \quad \boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I})$$

we find the least square estimator by minimizing the residual sum of squares, i.e.

$$\hat{\boldsymbol{\beta}} = \text{argmin}_{\beta} RSS(\boldsymbol{\beta}),$$

with

$$RSS(\boldsymbol{\beta}) = \boldsymbol{r}^T\boldsymbol{r} = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}).$$

The least square estimator is given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y},$$

and

$$V[\hat{\boldsymbol{\beta}}] = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}.$$

Example

In the items on a scale example find the optimal parameters for each choice of the design matrix.

Orthogonal parameters

Definition (Orthogonal parametrization)

A parametrization is called orthogonal if $(\boldsymbol{X}^T\boldsymbol{X})_{ij} = 0$ for $i \neq j$.

An orthogonal parametrization imply the the covariance between parameters is zero. Strong correlation between parameters is refered to as multicollinarity.

- Which of the parametrizations in the items on a scale example are orthogonal?

The general linear model as a projection

The fitted values in a general linear model can be ritten as

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{Y},$$

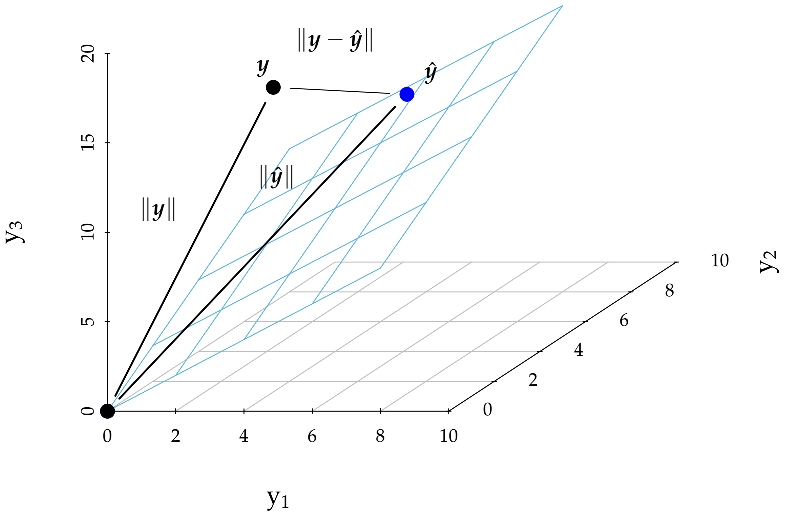
and the observed residuals can be written as

$$\boldsymbol{r} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y},$$

where

- $\boldsymbol{H}$ is an orthogonal projection matrix
- $\boldsymbol{r}$ and $\hat{\boldsymbol{Y}}$ are independent.
- The dimension of the model is $Trace(\boldsymbol{H}) = Rank(\boldsymbol{X}) = p$
- If two design matrices have the same projection matrix then the models are equivalent.

Example: Items on a scale, projections



Example: Items on a scale, geometric interpretation

The example highlight the geometric interpretation of the projections, in the example we have

- Norm of the observations

$$||\boldsymbol{y}|| = \sqrt{\sum_{i=1}^{n} y_i^2} = \sqrt{\boldsymbol{y}^T\boldsymbol{y}}$$

- Norm of fitted values

$$||\hat{\boldsymbol{y}}|| = \sqrt{\sum_{i=1}^{n} \hat{y}_i^2} = \sqrt{\boldsymbol{y}^T\boldsymbol{H}\boldsymbol{y}}$$

- Norm of residuals

$$||\boldsymbol{y}-\hat{\boldsymbol{y}}|| = \sqrt{\sum_{i=1}^{n} (y_i-\hat{y}_i)^2} = \sqrt{\boldsymbol{y}^T(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}}$$

and further as $\hat{\boldsymbol{y}}$ and $\boldsymbol{r} = \boldsymbol{y}-\hat{\boldsymbol{y}}$ are orthogonal it follows (Pythagoras) that

$$||\boldsymbol{y}||^2 = ||\hat{\boldsymbol{y}}||^2 + ||\boldsymbol{y}-\hat{\boldsymbol{y}}||^2$$

Example: Items on a scale, test strategy

Using the items on a scale example, formulate a partioning of variation of the form

$$\boldsymbol{Y}^T\boldsymbol{Y} = \boldsymbol{Y}^T\boldsymbol{H}_0\boldsymbol{Y} + \boldsymbol{Y}^T(\boldsymbol{H}_1 - \boldsymbol{H}_0)\boldsymbol{Y} + \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{Y}$$

where $\boldsymbol{H}_0$ correspond to the assumption, that the two items have the same weight. Under the assumption, what is the distribution of

$$\frac{\boldsymbol{Y}^T(\boldsymbol{H}_1 - \boldsymbol{H}_0)\boldsymbol{Y}}{\boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H}_1)\boldsymbol{Y}}.$$

Assume you have observed $\boldsymbol{y} = [10, 20, 40]$, is that unusual under the assumption?

Agenda

1. Matrix formulation of statistical models
2. The χ^2-distribution and the multivariate normal
3. Projections and Cochrans theorem
4. The general linear model