Analysis of Variance and Design of Experiments

Results from Matrix Theory and Random Variables

:::

Lecture 2

Random Vectors and Linear Estimation

Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Slides can be downloaded from http://home.iitk.ac.in/~shalab/sp

Random vectors:

Let $Y_1, Y_2, \ldots, Y_n$ **be *n* random variables then** $Y = (Y_1, Y_2, \ldots, Y_n)'$ **is called a random vector.**

- **The mean vector *Y* is**

$$E(Y) = ((E(Y_1), E(Y_2), \ldots, E(Y_n))'.$$

- **The covariance matrix or dispersion matrix of *Y* is**

$$Var(Y) = \begin{pmatrix} Var(Y_1) & Cov(Y_1, Y_2) & \ldots & Cov(Y_1, Y_n) \\ Cov(Y_2, Y_1) & Var(Y_2) & \ldots & Cov(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_n, Y_1) & Cov(Y_n, Y_2) & \ldots & Var(Y_n) \end{pmatrix}$$

which is a symmetric matrix

Random vectors:

- **If $Y_1, Y_2, \ldots, Y_n$ are independently distributed, then the covariance matrix is a diagonal matrix.**
- **If $Var(Y_i) = \sigma^2$ for all *i* = 1, 2,…,*n* then** $Var(Y) = \sigma^2 I_n$.

Linear function of random variable:

If $Y_1, Y_2, \ldots, Y_n$ **are *n* random variables, and** $k_1, k_2, .., k_n$ **are scalars,**

then $\sum_{i=1}^{n} k_i Y_i$ **is called a linear function of random variable** $Y_1, Y_2, \ldots, Y_n$**.**

If $Y = (Y_1, Y_2, \ldots, Y_n)', \; K = (k_1, k_2, \ldots, k_n)'$ **then** $K'Y = \sum_{i=1}^{n} k_i Y_i$ **,**

- **the mean** $K'Y$ **is** $E(K'Y) = K'E(Y) = \sum_{i=1}^{n} k_i E(Y_i)$ **and**
- **the variance of** $K'Y$ **is** $Var\left(K'Y\right) = K'Var\left(Y\right)K.$

Multivariate normal distribution:

A random vector $Y = (Y_1, Y_2, \ldots, Y_n)'$ **has a multivariate normal distribution with mean vector** $\mu = (\mu_1, \mu_2, \ldots, \mu_n)'$ **and dispersion matrix** $\sum$ **if its probability density function is**

$$f(Y \mid \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} (Y - \mu)' \Sigma^{-1} (Y - \mu) \right]$$

assuming $\sum$ **is a nonsingular matrix.**

Chi-square distribution:

- **If $Y_1, Y_2, \ldots, Y_k$ are independently distributed random variables following the normal distribution with common mean 0 and common variance 1, then the distribution of $\sum_{i=1}^{k} Y_i^2$ is called the χ^2- distribution with *k* degrees of freedom.**

- **The probability density function of χ^2-distribution with *k* degrees of freedom is given as**

$$f_{\chi^2}(x) = \frac{1}{\Gamma(k/2)2^{k/2}} x^{\frac{k}{2}-1} \exp\left(-\frac{x}{2}\right); \quad 0 < x < \infty.$$

Chi-square distribution:

- **If $Y_1, Y_2, \ldots, Y_k$ are independently distributed following the normal distribution with common mean 0 and common variance σ^2, then $\frac{1}{\sigma^2}\sum_{i=1}^{k} Y_i^2$ has χ^2 distribution with *k* degrees of freedom.**

- **If the random variables $Y_1, Y_2, \ldots, Y_k$ are normally distributed with non-null means $\mu_1, \mu_2, \ldots, \mu_k$ but common variance 1, then the distribution of $\sum_{i=1}^{k} Y_i^2$ has noncentral χ^2 distribution with *k* degrees of freedom and noncentrality parameter $\lambda = \sum_{i=1}^{k} \mu_i^2$.**

Chi-square distribution:

- **If** $Y_1, Y_2, \ldots, Y_k$ **are independently and normally distributed following the normal distribution with means** $\mu_1, \mu_2, \ldots, \mu_k$ **but common variance** σ^2 **then** $\frac{1}{\sigma^2}\sum_{i=1}^{k} Y_i^2$ **has non-central** χ^2 **distribution with *k* degrees of freedom and noncentrality parameter** $\lambda = \frac{1}{\sigma^2}\sum_{i=1}^{k}\mu_i^2.$

- **If *U* has a Chi-square distribution with *k* degrees of freedom then** $E(U) = k$ **and** $Var(U) = 2k.$

Chi-square distribution:

- **If *U* has a noncentral Chi-square distribution with *k* degrees of freedom and noncentrality parameter λ then**

  $E\left(U\right)=k+\lambda$ **and** $Var(U)=2k+4\lambda.$

- **If** $U_1,U_2,\ldots,U_k$ **are independently distributed random variables with each** U_i **having a noncentral Chi-square distribution with** n_i **degrees of freedom and non-centrality parameter** $\lambda_i,\ i=1,2,\ldots,k$ **then** $\sum_{i=1}^{k}U_i$ **has noncentral Chi-square distribution with** $\sum_{i=1}^{k}n_i$ **degrees of freedom and noncentrality parameter** $\sum_{i=1}^{k}\lambda_i.$

Chi-square distribution:

- **Let** $X = (X_1, X_2, ..., X_n)'$ **has a multivariate normal distribution with mean vector** μ **and positive definite covariance matrix** Σ**. Then** $X'AX$ **is distributed as noncentral** χ^2 **with *k* degrees of freedom if and only if** ΣA **is an idempotent matrix of rank *k*.**

Chi-square distribution:

Let the two quadratic forms-

- $X'A_1X$ **is distributed as** χ^2 **with** n_1 **degrees of freedom and noncentrality parameter** $\mu' A_1 \mu$ **and**
- $X'A_2X$ **is distributed as** χ^2 **with** n_2 **degrees of freedom and noncentrality parameter** $\mu' A_2 \mu.$ **.**

Then $X'A_1X$ and $X'A_2X$ **are independently distributed if**

$A_1 \Sigma A_2 = 0.$

t- distribution:

- **If**

  - ❖ ***X* has a normal distribution with mean 0 and variance 1,**
  - ❖ ***Y* has a χ^2 distribution with *n* degrees of freedom, and**
  - ❖ ***X* and *Y* are independent random variables,**

then the distribution of the statistic $T = \dfrac{X}{\sqrt{Y/n}}$ is called the *t*-distribution with *n* degrees of freedom.

The probability density function of *t* is

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1+\frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} ; \quad -\infty < t < \infty.$$

t- distribution:

- **If the mean of *X* is nonzero then the distribution of** $\dfrac{X}{\sqrt{Y/n}}$ **is called the noncentral *t* - distribution with *n* degrees of freedom and noncentrality parameter** μ **.**

F - distribution:

- **If *X* and *Y* are independent random variables with χ^2-distribution with *m* and *n* degrees of freedom respectively, then the distribution of the statistic $F = \dfrac{X/m}{Y/n}$ is called the *F*-distribution with *m* and *n* degrees of freedom. The probability density function of *F* is**

$$f_F(f) = \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{m/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} f^{\left(\frac{m-2}{2}\right)}\left(1+\left(\frac{m}{n}\right)f\right)^{-\left(\frac{m+n}{2}\right)}; \quad 0 < f < \infty.$$

F **- distribution:**

- **If *X* has a noncentral Chi-square distribution with *m* degrees of freedom and noncentrality parameter** λ; Y **has a** χ^2 **distribution with *n* degrees of freedom, and *X* and *Y* are independent random variables, then the distribution of** $F = \frac{X/m}{Y/n}$ **is the noncentral *F* distribution with *m* and *n* degrees of freedom and noncentrality parameter** λ.

Linear model:

Suppose there are *n* observations. In the linear model, we assume that these observations are the values taken by *n* random variables $Y_1, Y_2, \ldots, Y_n$ **satisfying the following conditions:**

1. $E(Y_i)$ **is a linear combination of *p* unknown parameters** $\beta_1, \beta_2, \ldots, \beta_p$

$$E(Y_i) = x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ip}\beta_p, \;\; i = 1, 2, \ldots, n$$

where $x_{ij}\,'s$ **are known constants.**

2. $Y_1, Y_2, \ldots, Y_n$ **are uncorrelated and normality distributed with variance** $Var(Y_i) = \sigma^2$.

Linear model:

The linear model can be rewritten by introducing independent normal random variables following $N(0,\sigma^2)$, as

$$Y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + ... + x_{ip}\beta_p + \varepsilon_i,\; i = 1, 2, ..., n.$$

These equations can be written using the matrix notations as

$$Y = X\beta + \varepsilon$$

- ***Y* is a *n* x 1 vector of observation,**
- ***X* is a *n* x *p* matrix of *n* observations on each of $X_1, X_2, ..., X_p$ variables,**
- **β is a $p \times 1$ vector of parameters and**
- **ε is $n \times 1$ a vector of random error components with** $\varepsilon \sim N(0, \sigma^2 I_n).$

Linear model:

Here

- ***Y* is called study or dependent variable,**
- $X_1, X_2, ..., X_p$ **are called explanatory or independent variables and**
- $\beta_1, \beta_2, ..., \beta_p$ **are called as regression coefficients.**

Alternatively since $Y \sim N(X\beta, \sigma^2 I)$ **so the linear model can also be expressed in the expectation form as a normal random variable *Y* with**

$$E(Y) = X\beta$$

$$Var(Y) = \sigma^2 I.$$

Note that β **and** σ^2 **are unknown but *X* is known.**

Estimable functions:

A linear parametric function $\lambda'\beta$ **of the parameter is said to be an estimable parametric function or estimable if there exists a linear function of random variables** $\ell' Y$ **of *Y* where** $Y = (Y_1, Y_2, ..., Y_n)'$ **such that**

$$E(\ell' Y) = \lambda' \beta$$

with $\ell = (\ell_1, \ell_2, ..., \ell_n)'$ **and** $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)'$ **being vectors of known scalars.**

Best linear unbiased estimates (BLUE):

The unbiased minimum variance linear estimate $\ell'Y$ of an estimable function $\lambda'\beta$ is called the best linear unbiased estimate (BLUE) of $\lambda'\beta$.

- **Suppose $\ell_1'Y$ and $\ell_2'Y$ are the BLUE of $\lambda_1'\beta$ and $\lambda_2'\beta$ respectively.**

  Then $(a_1\ell_1 + a_2\ell_2)'Y$ is the BLUE of $(a_1\lambda_1 + a_2\lambda_2)'\beta.$

- **If $\lambda'\beta$ is estimable, its best estimate is $\lambda'\hat{\beta}$ where $\hat{\beta}$ is any solution of the equations $X'X\beta = X'Y.$**

Least squares estimation:

The least-squares estimate of β is $Y = X\beta + \varepsilon$ is the value of β which minimizes the error sum of squares $\varepsilon'\varepsilon$.

Let $S = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta)$

$$= Y'Y - 2\beta' X'Y + \beta' X' X\beta.$$

Minimizing S with respect to β involves

$$\frac{\partial S}{\partial \beta} = 0$$

$$\Rightarrow X' X\beta = X'Y$$

which is termed as normal equation.

This normal equation has a unique solution given by,

$$\hat{\beta} = (X'X)^{-1} X'Y$$

assuming *rank*(*X*) = *p*.

Least squares estimation:

Note that $\dfrac{\partial^2 S}{\partial\beta\partial\beta'} = X'X$ **is a positive definite matrix.**

So $\hat{\beta} = (X'X)^{-1}X'Y$ **is the value of** β **which minimizes** $\varepsilon'\varepsilon$ **and is termed as ordinary least squares estimator of** β **.**

- **In this case,** $\beta_1, \beta_2, \ldots, \beta_p$ **are estimable and consequently, all the linear parametric function are estimable.**

- $$E(\hat{\beta}) = (X'X)^{-1}X'E(Y) = (X'X)^{-1}X'X\beta = \beta$$

- $$Var(\hat{\beta}) = (X'X)^{-1}X'Var(Y)X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

Least squares estimation:

- **If** $\lambda'\hat{\beta}$ **and** $\mu'\hat{\beta}$ **are the estimates of** $\lambda'\beta$ **and** $\mu'\beta$ **respectively, then**
  - $Var(\lambda'\hat{\beta}) = \lambda' Var(\hat{\beta})\lambda = \sigma^2[\lambda'(X'X)^{-1}\lambda]$
  - $Cov(\lambda'\hat{\beta}, \mu'\hat{\beta}) = \sigma^2[\mu'(X'X)^{-1}\lambda].$
- $Y - X\hat{\beta}$ **is called the residual vector**
- $E(Y - X\hat{\beta}) = 0.$

Linear model with correlated observations:

In the linear model $Y = X\beta + \varepsilon$ **with** $E(\varepsilon) = 0, Var(\varepsilon) = \Sigma$ **and** ε **is normally distributed, we find**

$$E(Y) = X\beta, Var(Y) = \Sigma.$$

Assuming Σ **to be positive definite, so we can write**

$$\Sigma = P'P$$

where *P* is a nonsingular matrix. Premultiplying $Y = X\beta + \varepsilon$ **by *P*, we get** $PY = PX\beta + P\varepsilon$

or $Y^* = X^*\beta + \varepsilon^*$

where $Y^* = PY$, $X^* = PX$ **and** $\varepsilon^* = P\varepsilon.$

Note that in this model $E(\varepsilon^*) = 0$ **and** $Var(\varepsilon^*) = \sigma^2 I.$

Distribution of $l'Y$:

In the linear model $Y = X\beta + \varepsilon,\ \varepsilon \sim N(0, \sigma^2 I)$ **consider a linear function** $\ell' Y$ **which is normally distributed with**

$$E(\ell' Y) = \ell' X\beta,$$

$$Var(\ell' Y) = \sigma^2(\ell'\ell).$$

Then

$$\frac{\ell' Y}{\sigma\sqrt{\ell'\ell}} \sim N\left(\frac{\ell' X\beta}{\sigma\sqrt{\ell'\ell}}, 1\right).$$

Further, $\dfrac{(\ell' y)^2}{\sigma^2 \ell'\ell}$ **has a noncentral Chi-square distribution with one degree of freedom and noncentrality parameter** $\dfrac{(\ell' X\beta)^2}{\sigma^2 \ell'\ell}$.

Degrees of freedom:

A linear function $l'Y$ of the observations $(\ell \neq 0)$ is said to carry one degrees of freedom.

A set of linear functions $L'Y$ where *L* is *r* x *n* matrix, is said to have *M* degrees of freedom if there exist *M* linearly independent functions in the set and no more.

Alternatively, the degrees of freedom carried by the set $L'Y$ equals rank (*L*).

When the set $L'Y$ are the estimates of $\Lambda'\beta$ the degrees of freedom of the set Λ will also be called the degrees of freedom for the estimates of $\Lambda'\beta$.

Sum of squares:

If $\ell' Y$ is a linear function of observations, then the projection of *Y* on 1 is the vector $\frac{Y'\ell}{\ell'\ell}.\ell$.

The square of this projection is called the sum of squares (SS) due to $\ell' y$ is given by $\frac{(\ell' Y)^2}{\ell'\ell}$.

Since $\ell' y$ has one degree of freedom, so the SS due $\ell' Y$ to has one degree of freedom.

Sum of squares:

The sum of squares and the degrees of freedom arising out of the mutually orthogonal sets of functions can be added together to give the sum of squares and degrees of freedom for the set of all the function together and vice versa.

Sum of squares:

Let $X = (X_1, X_2, ..., X_n)$ has a multivariate normal distribution with mean vector μ and positive definite covariance matrix Σ.

Let the two quadratic forms.

- **$X'A_1X$ is distribution χ^2 with n_1 degrees of freedom and noncentrality parameter $\mu'A_1\mu$ and**
- **$X'A_2X$ is distributed as χ^2 with n_2 degrees of freedom and noncentrality parameter $\mu'A_2\mu$.**

Then $X'A_1X$ and $X'A_2X$ are independently distributed if $A_1\Sigma A_2 = 0$.

Fisher-Cochran theorem:

If $X = (X_1, X_2, ..., X_n)'$ **has a multivariate normal distribution with mean vector** μ **and positive definite covariance matrix** Σ **and let** $X'\Sigma^{-1}X = Q_1 + Q_2 + ... + Q_k$

where $Q_i = X'A_iX$ **with** $rank(A_i) = N_i,\ i = 1, 2, ..., k.$ **Then** $Q_i's$ **are independently distributed noncentral Chi-square distribution with** N_i **degrees of freedom and noncentrality parameter** $\mu' A_i \mu$ **if and only if** $\sum_{i=1}^{k} N_i = N,$ **is which case**

$$\mu'\Sigma^{-1}\mu = \sum_{i=1}^{k} \mu' A_i \mu.$$

Derivatives of quadratic and linear forms:

Let $X=(x_1,x_2,\ldots,x_n)'$ **and** ***f*(*X*) be any function of *n* independent variables** $x_1,x_2,\ldots,x_n$ **, then**

$$\frac{\partial f(X)}{\partial X}=\begin{pmatrix}\dfrac{\partial f(X)}{\partial x_1}\\ \dfrac{\partial f(X)}{\partial x_2}\\ \vdots\\ \dfrac{\partial f(X)}{\partial x_n}\end{pmatrix}.$$

If $K=(k_1,k_2,\ldots,k_n)'$ **is a vector of constants, then** $\dfrac{\partial K'X}{\partial X}=K.$

If *A* is a $m\times n$ **matrix, then** $\dfrac{\partial X'AX}{\partial X}=2(A+A')X.$

Independence of linear and quadratic forms:

- **Let *Y* be an $n \times 1$ vector having multivariate normal distribution $N(\mu, I)$ and *B* be a $m \times n$ matrix.**

  Then the $m \times 1$ vector linear form *BY* is independent of the quadratic form $Y'AY$ if *BA* = 0 where *A* is a symmetric matrix of known elements.

- **Let *Y* be a $n \times 1$ vector having multivariate normal distribution $N(\mu, \Sigma)$ with $rank(\Sigma) = n$.**

  If $B\Sigma A = 0$, then the quadratic form $Y'AY$ is independent of linear form *BY* where *B* is a $m \times n$ matrix.