The multivariate Gauss distribution and related topics

Sandra Fortini *Bocconi University, Milano*

MSc in Finance

Outline

- Review: continuous distributions
- The normal distribution and related distributions: chi-squared, student's t, F
- The multivariate normal distribution
- Linear transformations of a standard normal vector
- Quadratic forms in a standard normal vector
- The standardized normal distribution

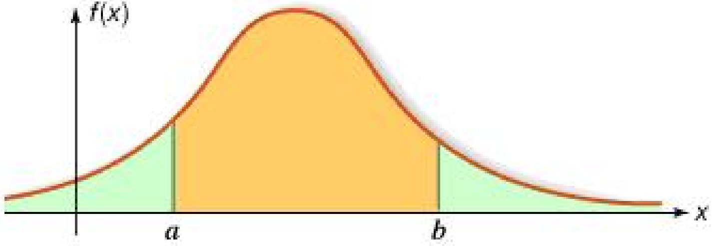
Reference: Greene W. H. *Econometric Analysis* Appendix B

Continuous distributions

A random variable X is continuous if

$$P(a < X \leq b) = \int_a^b f(x)dx$$

f is called the *density function of* X.



Prices, returns, volumes, indices etc. are continuous random variables.

Numerical characteristic of continuous distributions

- *expectation*: $\mu = E(X) = \int x f(x) dx$:
  is the mean value of X;

- *standard deviation* $\sigma = \sqrt{E(X-\mu)^2}$:=mean deviation of X from μ;

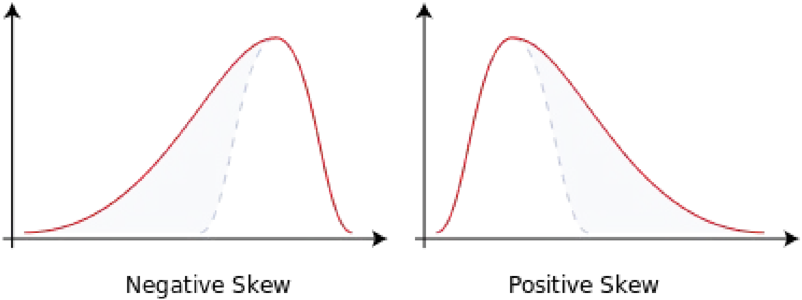
- $\sigma^2 = E(X-\mu)^2 = V(X)$: *variance* of X.

- *skewness coefficient* $\varsigma = \frac{E(X-\mu)^3}{\sigma^3}$: measures asymmetry
  symmetric distribution $\varsigma = 0$; long right tail $\varsigma > 0$; long left tail $\varsigma < 0$

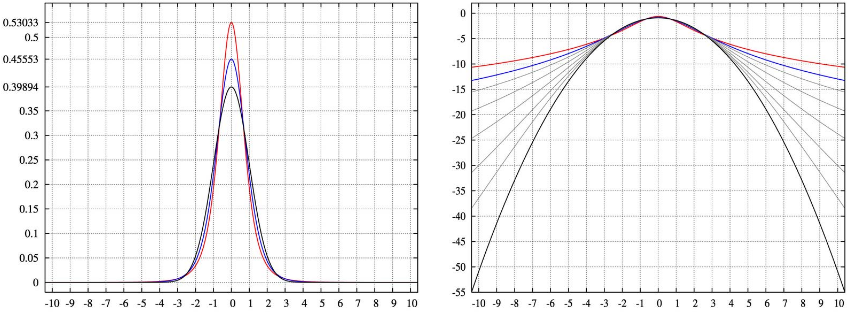
- *excess of kurtosis* $\kappa = \frac{E(X-\mu)^4}{\sigma^4} - 3$: measures tails thickness
  the larger κ, the thicker the tails.

Remark: the above parameters can be not defined or ∞.

Skewness



Different degrees of Kurtosis



On the left densities; on the right their logarithms

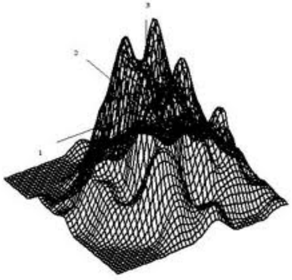
The red curve is the one with the largest Kurtosis

Multivariate distributions

The random variables $X_1, \ldots, X_n$ have continuous distribution if

$$P(a_1 < X_1 \leq b_1, \ldots, a_n < X_n \leq b_n) = \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

f is called the *joint density* of $X_1, \ldots, X_n$.



Marginal distributions

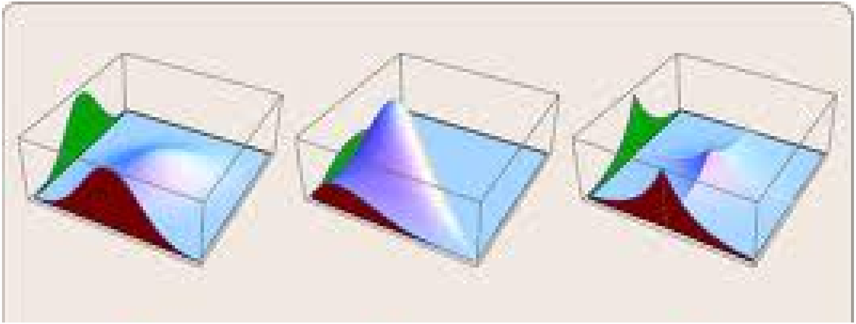
Suppose now you are interested only in the first k variables $X_1, \ldots, X_k$ $(k < n)$

The density function of $X_1, \ldots, X_k$ can be found integrating out the remaining variables:

$$f(x_1, \ldots, x_k) = \int \int \ldots \int f(\underbrace{x_1, \ldots, x_k}_{keep}, \underbrace{x_{k+1}, \ldots, x_n}_{integrate}) dx_{k+1} \ldots dx_n$$

$f(x_1, \ldots, x_k)$ is called the *marginal density* of $X_1, \ldots, X_k$.

Remark: with an abuse of notation we write f for all densities



Conditional distributions

Consider the random variables $X_1, \ldots, X_n$ with joint density f.

Split the variables into two groups: $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_n$.

Suppose you are interested in the distribution of $X_{k+1} \ldots, X_n$ in the hypothesis that $X_1 = x_1, \ldots, X_k = x_k$

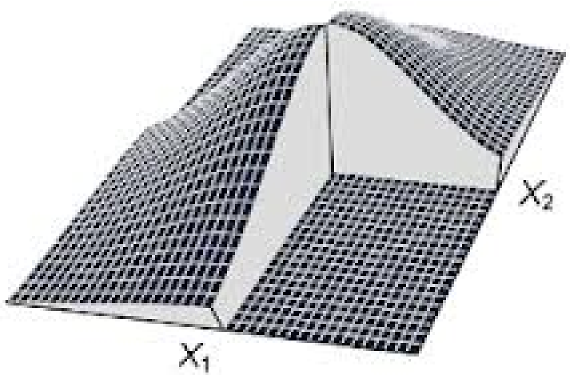
called *conditional distribution* of $X_{k+1}, \ldots, X_n$ *given that* $X_1 = x_1, \ldots, X_k = x_k$.

The *conditional density* is

$$f(x_{k+1}, \ldots, x_n | x_1, \ldots, x_k) = \frac{f(\overbrace{x_1, \ldots, x_k}^{given}, x_{k+1}, \ldots, x_n)}{f(\underbrace{x_1, \ldots, x_k}_{given})}$$

Notation: $X_{k+1} \ldots, X_n | X_1, \ldots, X_k \sim f$

Conditional density



$f(x_{k+1}, \dots, x_n | x_1, \dots, x_k)$ is obtained by looking at $f(x_1, \dots, x_n)$ as a function of $x_{k+1}, \dots, x_n$, keeping $x_1, \dots, x_k$ fixed, and then re-scaling.

Independent random variables

The random variables $X_1, \ldots, X_k$ are independent of $X_{k+1}, \ldots, X_n$ if

$$f(x_{k+1}, \ldots, x_n | x_1, \ldots, x_k) = f(x_{k+1}, \ldots, x_n)$$

This holds iff

$$f(x_1, \ldots, x_n) = f(x_1, \ldots, x_k) f(x_{k+1}, \ldots, x_n)$$

The above definition can be extended to more than two groups.

The random variables $X_1, \ldots, X_n$ are *independent* if

$$f(x_1, \ldots, x_n) = f(x_1) f(x_2) \ldots f(x_n)$$

Matrix notation

A random vector is a vector $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$ of random variables

The expected value of a random vector is defined as the vector of the expected values:

$$\mu = E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

Covariance

The covariance matrix of $\mathbf{X}$ is defined as

$$\boldsymbol{\Sigma} = V(\mathbf{X}) = E((\mathbf{X}-\mu)(\mathbf{X}-\mu)^T) = E(\mathbf{X}\mathbf{X}^T) - \mu\mu^T$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix}$$

where

$$\sigma_{ii} = V(X_i)$$

$$\sigma_{ij} = Cov(X_i, X_j) = E((X_i - \mu_i)(X_j - \mu_j))$$

Hence $\boldsymbol{\Sigma}$ contains the *variances in the diagonal* and the *covariances off the diagonal*.

Properties of expectation and variance

In matrix notation, the properties of expectation and variance can be written in a very compact way

- $E(\mathbf{AX}) = \mathbf{A}E(\mathbf{X})$
- $V(\mathbf{AX}) = \mathbf{A}V(\mathbf{X})\mathbf{A}^T$
- If X_i and X_j are independent, then $Cov(X_i, X_j) = 0$
- If $X_1, \ldots, X_n$ are independent, then $\mathbf{\Sigma}$ is diagonal.

Partitioning

Partition: $$\mathbf{X} = \left[\begin{array}{c} X_1 \\ X_2 \\ \vdots \\ X_k \\ \hline \\ X_{k+1} \\ X_{k+2} \\ \vdots \\ X_n \end{array} \right]$$

Write $\mathbf{X_1} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix}$ $\mathbf{X_2} = \begin{bmatrix} X_{k+1} \\ X_{k+2} \\ \vdots \\ X_n \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix}$

Partitioning μ and Σ

μ and $\mathbf{\Sigma}$ can be partitioned likewise:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}; \quad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}$$

Problem:
you know the density of $\mathbf{X}$*:* $f_{\mathbf{X}}(\mathbf{x})$
you know that $\mathbf{Y} = \mathbf{g}(\mathbf{X})$

want to find the density of $\mathbf{Y}$*:* $f_{\mathbf{Y}}(\mathbf{y})$.

Theorem
Let $\mathbf{g}$ *be one-to-one and continuously differentiable,* $\mathbf{J}$ = *Jacobian matrix of* $\mathbf{g}^{-1}$.
If $\det(\mathbf{J}(\mathbf{y})) \neq \mathbf{0}$ $\forall \mathbf{y}$, *then*

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y}))|\det \mathbf{J}(\mathbf{y})|$$

(Jacobian= matrix of all first-order partial derivatives)

Exercise 1

Let $X = (X_1, X_2)$with X_1, X_2 i.i.d. random variables with density

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}$$

Furthermore, let $Y = (Y_1, Y_2)$ with $Y_1 = X_1 + X_2$, $Y_2 = X_1 + 2X_2$.

a) Find the expectation and variance of X

b) Find skewness and excess kurtosis coefficients of X_1 and of X_2

c) Find the density function of $Z = X_1^2$

d) Find $E(Y)$ and $V(Y)$.

e) Write the joint density of X_1 and X_2

f) Find the joint density of Y_1, Y_2. Are Y_1 and Y_2 stochastically independent?

g) Find the marginal density of Y_1

h) Find the conditional density of Y_2, given Y_1

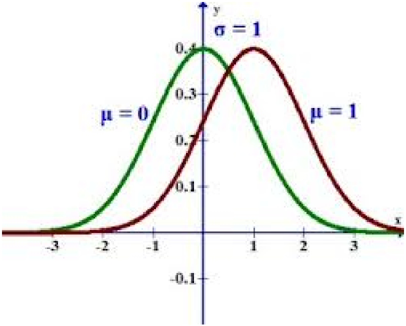
(Hint: $\int_{-\infty}^{+\infty} \exp\{-x^2\}dx = \sqrt{\pi}$)

The normal (or Gauss) distribution

Let X be a continuous random variable with $\mu = E(X)$, $\sigma^2 = V(X)$.

X has normal distribution if

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



$X \sim N[\mu, \sigma^2]$

Law of errors

The Gauss distribution is the *normal* law of errors.

Why?

It can be proved that if X is the sum of a large number of *independent* and *uniformly small* random variables, then the distribution of X is well approximated by the normal law.

This is the *Central Limit Theorem* (CLT)

The CLT explains why errors have Gaussian law:

an error has *many different sources*, each of which gives a *small contribution*. If the sources are *independent*, the error has normal distribution.

Properties of the normal distribution

The normal law is stable with respect to linear transformations:

- If X has normal distribution with mean μ and variance σ^2, then

$$aX + b \sim N[a\mu + b, a^2\sigma^2]$$

- If X has normal distribution with mean μ and variance σ^2, then

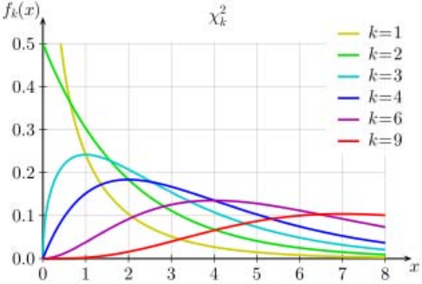
$$\frac{X - \mu}{\sigma} \sim N[0, 1]$$

The chi-squared distribution

The chi-squared distribution with n degrees of freedom has density

$$f(x) = \begin{cases} cx^{n/2-1}e^{-x/2} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where c depends on n and is such that the density integrates to 1.



$E(X) = n$ and $V(X) = 2n$

Properties of the chi-square distribution

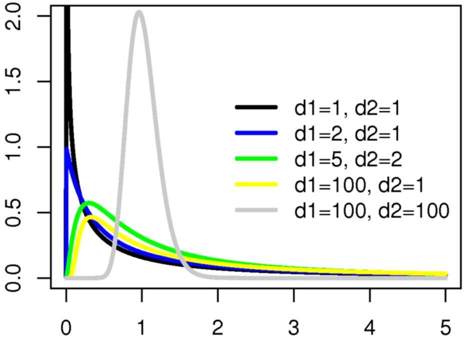
- If X_1 and X_2 are independent chi-squared variables with n_1 and n_2 degrees of freedom, respectively, then
$$X_1 + X_2 \sim \chi^2[n_1 + n_2]$$
- If $X_1, \ldots, X_n$ are independent *chi-squared*$[1]$ variables, then
$$X_1 + \cdots + X_n \sim \chi^2[n]$$
- If $Z \sim N[0, 1]$, then
$$X = Z^2 \sim \chi^2[1]$$
- If $Z_1, \ldots, Z_n$ are independent $N[0, 1]$ variables, then
$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2[n]$$
- If $Z_1, \ldots, Z_n$ are independent $N[\mu, \sigma^2]$ variables, then
$$\sum_{i=1}^{n} \frac{(Z_i - \mu)^2}{\sigma^2} \sim \chi^2[n]$$

The F distribution

The F distribution with n_1 and n_2 degrees of freedom has density

$$f(x) = \begin{cases} cx^{n_1/2-1}(1+\frac{n_1}{n_2}x)^{-\frac{n_1+n_2}{2}} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where c depends on n and is such that the density integrates to 1.



Properties of the F distribution

- If X_1 and X_2 are independent chi-squared variables with degrees of freedom parameters n_1 and n_2, then the ratio

$$F[n_1, n_2] = \frac{X_1/n_1}{X_2/n_2}$$

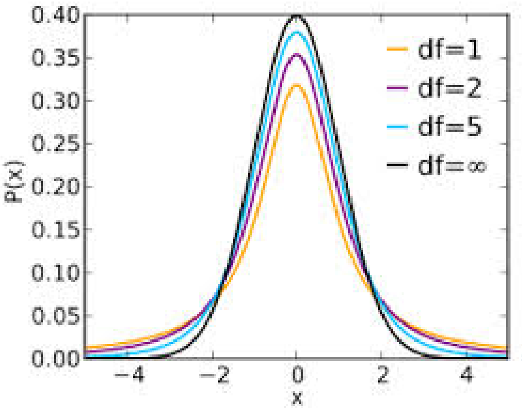
has the F-distribution with n_1 and n_2 degrees of freedom.

The t-distribution

The Student's t distribution with n degrees of freedom has density

$$f(x) = c\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

where c depends on n and is such that the density integrate to 1.



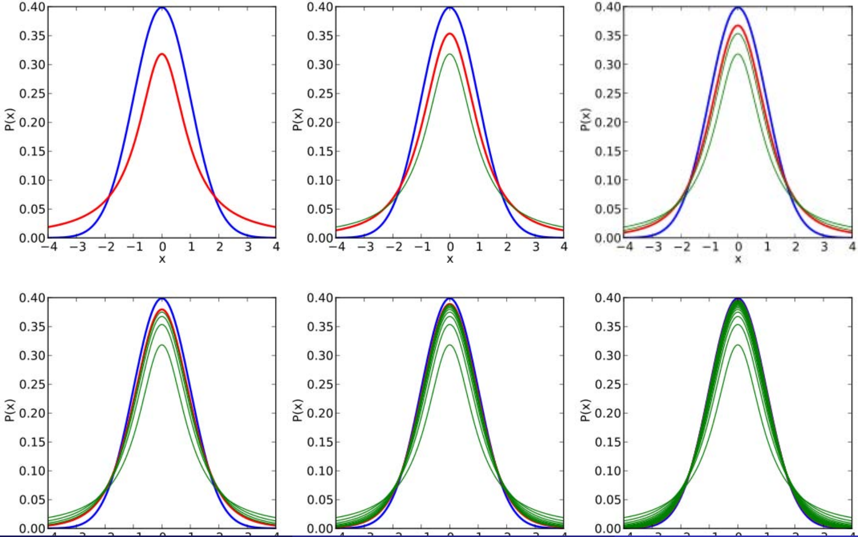
Properties of the t-distribution

- If Z is an $N[0,1]$ variable and X is $\chi^2[n]$ and is independent of Z, then the ratio
$$t = \frac{Z}{\sqrt{X/n}}$$
has t distribution with n degrees of freedom, denoted $t \sim t[n]$

- If $t \sim t[n]$, then $t^2 \sim F[1,n]$

Asymptotic distributions

- $t \sim t[n] \rightarrow N[0,1]$ as $n \rightarrow \infty$



- If $F \sim F[n_1, n_2]$, then $n_1 F \to \chi^2[n_1]$ as $n_2 \to \infty$
- If $X \sim \chi^2[n]$, then $Z = \sqrt{2X} - \sqrt{2n-1} \to N[0,1]$ as $n \to \infty$

The multivariate normal distribution

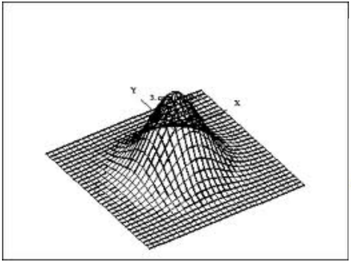
A continuous random vector $\mathbf{X}$ has multivariate normal (or Gauss) distribution if it has density

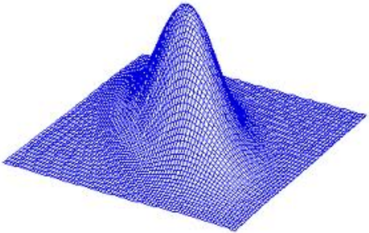
$$f(\mathbf{x}) = (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2}e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}$$

It holds $\mu = E(\mathbf{X})$, $\mathbf{\Sigma} = V(\mathbf{X})$.

The notation is $\mathbf{X} \sim N[\mu, \mathbf{\Sigma}]$.

Remark: $\mathbf{\Sigma}$ must be non-singular.





The multivariate standard normal distribution

The $N[\mathbf{0}, \mathbf{I}]$ is called *multivariate standard normal* or *spherical normal distribution*.

The density function of a multivariate standard normal vector is

$$f(\mathbf{x}) = (2\pi)^{-n/2}|\mathbf{I}|^{-1/2}e^{-\frac{1}{2}\mathbf{x}^T\mathbf{I}^{-1}\mathbf{x}}$$

$$= (2\pi)^{-n/2}e^{-\frac{1}{2}\sum_{i=1}^n x_i^2}$$

$$= \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x_i^2}{2}}\right)$$

Hence $\mathbf{X} \sim N[\mathbf{0}, \mathbf{I}]$ iff $X_1, \ldots, X_n$ are *i.i.d. variables with standard normal* distribution.

Partitioning a normal vector

Theorem
If

$$\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} \right)$$

then

- $\mathbf{X_1} \sim N[\mu_1, \mathbf{\Sigma}_{11}]$
- $\mathbf{X_2} \sim N[\mu_2, \mathbf{\Sigma}_{22}]$
- $\mathbf{X_2} | \mathbf{X_1} \sim N[\mu_{2.1}, \mathbf{\Sigma}_{22.1}]$ with

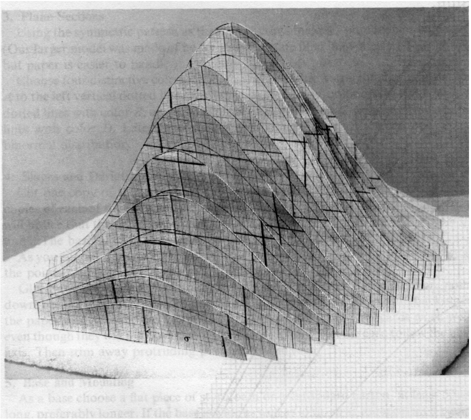
$$\mu_{2.1} = \mu_2 + \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} (\mathbf{X}_1 - \mu_1)$$

$$\mathbf{\Sigma}_{22.1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{11}^{-1} \mathbf{\Sigma}_{12}$$

Remark: $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{21} = \mathbf{0}$

Partitioning a normal vector

All marginal and conditional distributions from a multivariate normal are normal



Linear functions of a normal vector

Any linear function of jointly normally distributed random variables is normally distributed:

If $\mathbf{X} \sim N[\mu, \mathbf{\Sigma}]$, then $\mathbf{AX} + \mathbf{b} \sim N[\mathbf{A}\mu + \mathbf{b}, \mathbf{A\Sigma A}^T]$.

Quadratic forms

A *quadratic form* in the variables $x_1, \ldots, x_n$ is homogeneous polynomial of second degree in $x_1, \ldots, x_n$. A quadratic form can be written as

$$a_{11}x_1^2 + a_{12}x_1x_2 + \ldots a_{1n}x_1x_n + a_{21}x_2x_1 + \cdots + a_{nn}x_n^2 = \mathbf{x}^T\mathbf{A}\mathbf{x}$$

with $\mathbf{A}$ *symmetric* matrix.

A matrix $\mathbf{A}$ is said to be *idempotent* if $\mathbf{A}^2 = \mathbf{A}$

A quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is called *idempotent* if $\mathbf{A}$ is idempotent.

Idempotent Quadratic forms in a standard normal vector

Theorem
If $\mathbf{X} \sim N[\mathbf{0}, \mathbf{I}]$ and $\mathbf{A}$ is symmetric and idempotent, then $\mathbf{X}^T \mathbf{A} \mathbf{X}$ has a chi-squared distribution with degrees of freedom equal to the rank of $\mathbf{A}$.

Remark: for a symmetric, idempotent matrix, the rank is equal to its trace (sum of the elements in the diagonal).

Application: Distribution of S^2 in a Standard Normal Vector

Let $X_1, \ldots, X_n$ i.i.d. $\sim N[0,1]$.

Let $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$ be the sample mean
and $S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n}$ be the sample variance.

Then

$$nS^2 \sim \chi^2[n-1]$$

Proof

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{bmatrix} \sim N[\mathbf{0}, \mathbf{I}]$$

$nS^2 = \mathbf{X}^T \mathbf{A} \mathbf{X}$

where $\mathbf{A} = \begin{bmatrix} 1-1/n & -1/n & \dots & -1/n \\ -1/n & 1-1/n & \dots & -1/n \\ \vdots & \vdots & \dots & \vdots \\ -1/n & -1/n & \dots & 1-1/n \end{bmatrix}$ is symmetric and idempotent.

Hence nS^2 has a chi-squared distribution.

To find the number of degrees of freedom we need to compute rank of $\mathbf{A}$, which is equal to the trace of $\mathbf{A}$.

trace$(\mathbf{A}) = \sum_{i=1}^{n} (1 - 1/n) = n - 1.$

Independence of linear forms in a standard normal vector

Theorem
If $\mathbf{X} \sim N[\mathbf{0}, \mathbf{I}]$ and $\mathbf{AX}$ and $\mathbf{BX}$ are two linear forms, then $\mathbf{AX}$ and $\mathbf{BX}$ are independent if and only if $\mathbf{AB}^T = \mathbf{0}$.

Proof

Let $\mathbf{Z} = \mathbf{AX}$, $\mathbf{W} = \mathbf{BX}$. Then

$$\left[\begin{array}{c} \mathbf{Z} \\ \mathbf{W} \end{array}\right] = \left[\begin{array}{c} \mathbf{A} \\ \mathbf{B} \end{array}\right] \mathbf{X} \sim N\left(\left[\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array}\right], \left[\begin{array}{cc} \mathbf{AA^T} & \mathbf{AB^T} \\ \mathbf{BA^T} & \mathbf{BB^T} \end{array}\right]\right)$$

It follows that $\mathbf{Z}, \mathbf{W}$ are independent if and only if $\mathbf{AB^T} = \mathbf{0}$.

Independence of Idempotent Quadratic forms in a standard normal vector

Theorem
If $\mathbf{X} \sim N[\mathbf{0}, \mathbf{I}]$ and $\mathbf{X}^T\mathbf{A}\mathbf{X}$ and $\mathbf{X}^T\mathbf{B}\mathbf{X}$ are two idempotent quadratic forms, then $\mathbf{X}^T\mathbf{A}\mathbf{X}$ and $\mathbf{X}^T\mathbf{B}\mathbf{X}$ are independent if $\mathbf{A}\mathbf{B} = \mathbf{0}$.

Application: Ratio of Independent Idempotent Quadratic forms in a standard normal vector

Let $\mathbf{X} \sim N[\mathbf{0}, \mathbf{I}]$ and let $\mathbf{A}$ and $\mathbf{B}$ be symmetric idempotent matrices such that $\mathbf{AB} = \mathbf{0}$.

Let r_A and r_b be the ranks of $\mathbf{A}$ and $\mathbf{B}$, respectively.

According to the above theorems $\mathbf{X}^T\mathbf{AX}$ and $\mathbf{X}^T\mathbf{BX}$ are independent chi-square random variables.

The ratio of independent, chi-square random variables, each divided by the number of degrees of freedom, has F distribution.

Hence

$$\frac{\mathbf{X}^T\mathbf{AX}/r_A}{\mathbf{X}^T\mathbf{BX}/r_B} \sim F[r_A, r_B]$$

Independence of a linear and quadratic form in a standard normal vector

Theorem

Let $\mathbf{X} \sim N[\mathbf{0}, \mathbf{I}]$. A linear function $\mathbf{LX}$ and an idempotent quadratic form $\mathbf{X}^T\mathbf{AX}$ are independent if $\mathbf{LA} = \mathbf{0}$.

Application: independence of the sample mean and the sample variance in a standard normal vector

Let $X_1, \ldots, X_n$ be i.i.d $\sim N[0,1]$.

The sample mean $\overline{X}$ and the sample variance S^2 are independent random variables.

Proof

$n\overline{X} = \sum_{i=1}^{n} X_i = \mathbf{L}\mathbf{X}$ with $\mathbf{L} = \left[\begin{array}{cccc} 1 & 1 & \dots & 1 \end{array} \right]$

$nS^2 = \mathbf{X}^T \mathbf{A}\mathbf{X}$ with A as before

$$\mathbf{L}\mathbf{A} = \left[\begin{array}{cccc} 1 & 1 & \dots & 1 \end{array} \right] \left[\begin{array}{cccc} 1-1/n & -1/n & \dots & -1/n \\ -1/n & 1-1/n & \dots & -1/n \\ \vdots & \vdots & \dots & \vdots \\ -1/n & -1/n & \dots & 1-1/n \end{array} \right]$$

$$= \left[\begin{array}{cccc} 0 & 0 & \dots & 0 \end{array} \right]$$

Hence $n\overline{X}$ and nS^2 are independent and, therefore $\overline{X}$ and S^2 are independent.

The standardized normal distribution

Let $\mathbf{X} \sim N[\mu, \mathbf{\Sigma}]$.

There exists a symmetric matrix $\mathbf{A}$ such that $\mathbf{A}^2 = \mathbf{\Sigma}$.

$\mathbf{A}$ is called the *square root* of $\mathbf{\Sigma}$ and denoted by $\mathbf{\Sigma}^{1/2}$

The inverse of $\mathbf{\Sigma}^{1/2}$ is denoted by $\mathbf{\Sigma}^{-1/2}$ and satisfies

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2} = \mathbf{\Sigma}^{-1}$$

$$\mathbf{Z} = \mathbf{\Sigma}^{-1/2}(\mathbf{X} - \mu)$$

is called the standardized normal.

$$\mathbf{Z} \sim N[\mathbf{0}, \mathbf{I}]$$

$$\mathbf{Z}^T\mathbf{Z} \sim \chi^2[n]$$

Proof

$$\mathbf{X} - \mu \sim N[\mathbf{0}, \mathbf{\Sigma}]$$

hence

$$\mathbf{\Sigma}^{-1/2}(\mathbf{X} - \mu) \sim N[\mathbf{\Sigma}^{-1/2}\mathbf{0}, \mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}\mathbf{\Sigma}^{-1/2}] = N[\mathbf{0}, \mathbf{I}]$$

For the last assertion

$$(\mathbf{X} - \mu)^T\mathbf{\Sigma}^{-1}(\mathbf{X} - \mu) = (\mathbf{X} - \mu)^T\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2}(\mathbf{X} - \mu) = \mathbf{Z}^T\mathbf{Z}$$

Extended definition of multivariate normal

If $\mathbf{X} \sim N[\mu, \mathbf{\Sigma}]$, then

$$\mathbf{X} = \mu + \mathbf{\Sigma}^{1/2}\mathbf{Z}$$

with

$$\mathbf{Z} \sim N[\mathbf{0}, \mathbf{I}]$$

Definition

A random vector $\mathbf{X}$ has multivariate normal distribution if

$$\mathbf{X} = \mu + \mathbf{A}\mathbf{Z}$$

for some

$$\mathbf{Z} \sim N[\mathbf{0}, \mathbf{I}]$$

some vector μ and some matrix $\mathbf{A}$.

With the new definition, *all properties still hold.*

Two cases

Let

$$\mathbf{X} = \mu + \mathbf{AZ}$$

with

$$\mathbf{Z} \sim N[\mathbf{0}, \mathbf{I}]$$

Let $\mathbf{\Sigma} = \mathbf{V}(\mathbf{X}) = \mathbf{A}\mathbf{A}^{\mathbf{T}}$.

- If $\mathbf{\Sigma}$ is non-singular, $\mathbf{X}$ is continuous and the definition coincides with the old one.
- If $\mathbf{\Sigma}$ is singular, there exists $\mathbf{a}$ such that

$$\mathbf{a}^T \mathbf{\Sigma} \mathbf{a} = \mathbf{V}(\mathbf{a}^T \mathbf{X}) = 0$$

  Hence $\mathbf{a}^T \mathbf{X} = a_1 X_1 + \cdots + a_n X_n = c$ with c constant. In this case $\mathbf{X}$ has no density. If rank$(\mathbf{\Sigma}) = k$ it is possible to write $n - k$ of the random variables as linear functions of a constant and of the remaining k random variables, that have continuous multivariate normal distribution.

Exercise 2

Let $Y = X\beta + \epsilon$ with $\epsilon \sim N(0, \sigma^2 I)$ and X deterministic of dimension $n \times k$ and rank k. Let $M = X(X^T X)^{-1} X^T$

a) Write the joint density of $Y_1, \ldots, Y_n$.

b) Show that the MLE of (β, σ^2) is $\hat{\beta} = (X^T X)^{-1} X^T Y = \beta + (X^T X)^{-1} X^T \epsilon$, $\hat{\sigma}^2 = \hat{\epsilon}^T \hat{\epsilon}/n$ where $\hat{\epsilon} = Y - \hat{Y}$ with $\hat{Y} = X\hat{\beta}$.

c) Show that $\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$.

d) Show that $\hat{Y} = MY = X\beta + M\epsilon$ and that $\hat{Y} \sim N(X\beta, \sigma^2 M)$.

e) Show that $\hat{\epsilon} = (I - M)Y = (I - M)\epsilon$ and that $\hat{\epsilon} \sim N(0, \sigma^2 (I - M))$.

f) Show that $\hat{\epsilon}$ is independent from $\hat{\beta}$ and from $\hat{Y}$.

g) Show that $\hat{\epsilon}^T \hat{\epsilon}/\sigma^2 \sim \chi^2(n - k)$.

h) Assuming $\beta = 0$, show that $\hat{Y}^T \hat{Y}/\sigma^2 \sim \chi^2(k)$.

i) Assuming $\beta_i = 0$,show that $T_i = \hat{\beta}_i / \sqrt{[(X^T X)^{-1}]_{ii} \hat{\epsilon}^T \hat{\epsilon}/(n - k)} \sim t(n - k)$.

j) Assuming $\beta = 0$, show that $F = (\hat{Y}^T \hat{Y}/k)/(\hat{\epsilon}^T \hat{\epsilon}/(n - k)) \sim F(k, n - k)$.