The normal family

Probability calculus / Adv Stat I

Prof. Dr. Matei Demetrescu

Today's outline

The normal family

Outline

One of Gauss' many ideas

The **normal (Gaussian) family** of distributions is the most extensively used distribution in statistics and econometrics. There are three main reasons for this.

1) The normal distribution is very tractable analytically.
2) The normal density has a bell shape, whose symmetry makes it an appealing candidate to model the probability space of many experiments.
3) There is the Central Limit Theorem (which we will discuss in Chapter 5), which indicates that under mild conditions, the normal distribution can be used to approximate a large variety of distributions in large samples.

The normal distribution

Family Name: Univariate Normal

Parameterization	$(\mu, \sigma) \in \Omega = \{(\mu, \sigma) : \mu \in (-\infty, \infty), \sigma > 0\}$
Density Definition	$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$
Moments	$\mathrm{E}(X) = \mu, \quad \mathrm{Var}(X) = \sigma^2, \quad \mu_3 = 0$
MGF	$M_X(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}$

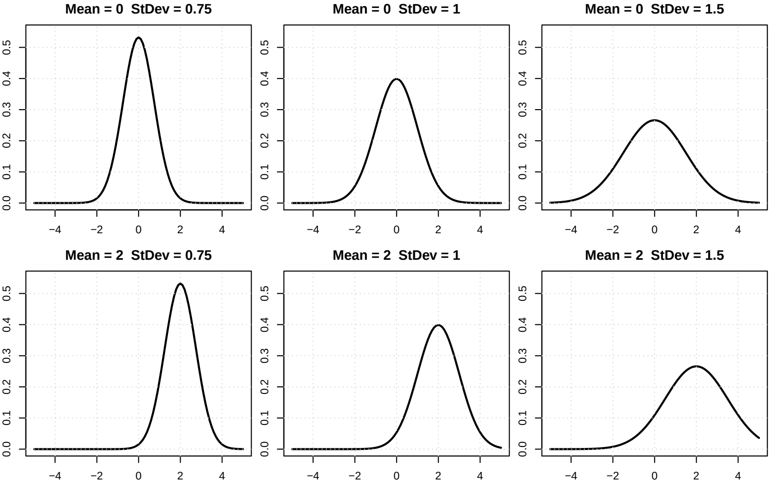
The normal family of densities is indexed by the two parameters μ and σ which correspond to the mean and the standard deviation, respectively.

In order to denote a normally distributed random variable with mean μ and variance σ^2, we will use the usual notation $X \sim \mathcal{N}(\mu, \sigma^2)$.

A normal distribution with $\mu = 0$ and $\sigma^2 = 1$ is called **standard normal distribution**, and is abbreviated by $\mathcal{N}(0, 1)$. It has density φ and cdf Φ.

The bell shape

The normal density is symmetric about its mean μ, has its maximum at $x = \mu$ and inflection points at $x = \mu \pm \sigma$:



Some properties

Theorem (4.5)

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$.

Hence, the standard normal distribution is sufficient to assign probabilities to **all** events involving Gaussian random variables.

Let $X \sim \mathcal{N}(17, 1/4)$. The probability of the event $X \in [16, 18]$ can be computed as

$$\begin{aligned} \mathrm{P}(16 \leq X \leq 18) &= \mathrm{P}\left(\frac{16-17}{(1/2)} \leq \frac{X-17}{(1/2)} \leq \frac{18-17}{(1/2)}\right) \\ &= \mathrm{P}(-2 \leq Z \leq 2) = \Phi(2) - \Phi(-2) = 0.9544, \end{aligned}$$

where $\Phi(\cdot)$ denotes the cdf of a standard normal distribution.

Relation to gamma

Normal and chi-square distribution: There is relationship between standard normal random variables and the χ^2 distribution which is subject of the following two theorems:

Theorem (4.6)

If $X \sim \mathcal{N}(0,1)$, *then* $Y = X^2 \sim \chi^2_1$.

Theorem (4.7)

Let $(X_1, \ldots, X_n)$ *independent* $\mathcal{N}(0,1)$*-distributed random variables. Then* $Y = \sum_{i=1}^n X_i^2 \sim \chi^2_n$.

(The MGF works miracles...)

Outline

The generalized normal distribution

Family Name: Generalized normal

Parameterization	$\mu \in \mathbb{R},\ \alpha, \beta \in (0, \infty)$
Density Definition	$f(x;\mu,\alpha,\beta) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)}\, e^{-\left\lvert\frac{x-\mu}{\alpha}\right\rvert^{\beta}}$
CDF	$F(x) = \frac{1}{2} + \operatorname{sgn}(x-\mu)\, \frac{\gamma\left(\frac{1}{\beta}, \left\lvert\frac{x-\mu}{\alpha}\right\rvert^{\beta}\right)}{2\Gamma\left(\frac{1}{\beta}\right)}$ where $\gamma(s,x) = \int_0^x t^{s-1} e^{-t} \mathrm{d}s$
Moments	$\mathrm{E}(X) = \mu,\ \mathrm{Var}(X) = \frac{\alpha^2\Gamma\left(\frac{3}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)}$

Other names: generalized error distribution, exponential power distribution, generalized Gaussian distribution

- Setting $\beta = 1$ leads to the Laplace (double exponential) distribution
- Setting $\beta = 2$ leads to the normal (note the missing $1/2$ in the exp.)

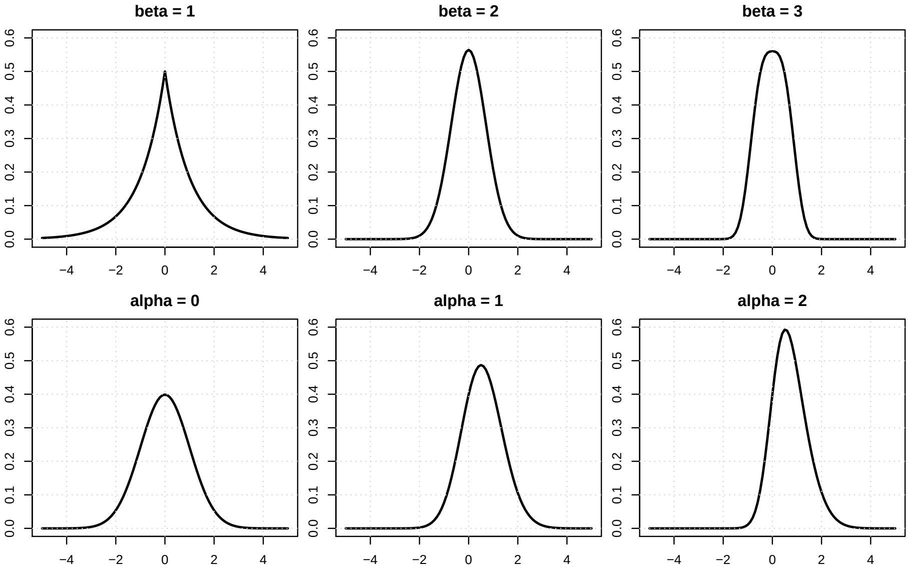
Skewed distributions

The already discussed skewed distributions are sometimes not flexible enough...

Family Name: Skew normal distribution	
Parameterization	$\alpha \in \mathbb{R}$,
Density Definition	$f\left(x\right) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\frac{x-\xi}{\omega}\right)$

- For $\alpha = 0$, symmetry is recovered.
- For $\alpha \to \pm\infty$, $f(x)$ converges to the positive (negative) half-normal distribution given by $f(x) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\mathbb{I}_{(\xi,\infty)}(x)$
- This can be generalized to $f(x) = \frac{2}{\omega}h\left(\frac{x-\xi}{\omega}\right)G\left(\alpha\frac{x-\xi}{\omega}\right)$ with h, g continuous densities, symmetric about 0 (and G the associated cdf).

Generalized (top) and skew normal (bottom) pdfs



Location-scale families

What if the shape is of secondary interest?

Family Name: Location-scale (univariate)

Parameterization	$\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$, g a pdf
Density Definition	$f(x; \mu, \sigma) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right)$
CDF	$F(x) = G\left(\frac{x-\mu}{\sigma}\right)$, G the corresponding cdf
Moments	μ, σ^2 (if g is standardized with finite variance)
MGF	$M_X(t) = e^{\mu t} M_Z(\sigma t)$

Note that the family can actually be defined for base densities that do not have finite variance (or even expectation).

If X has a location-scale distribution (with a given base g), then so does $Y = a + bX$ for any $a, b \neq 0$.

And finally: Gaussian mixtures

Another approach considers building densities from adding simpler basic elements:

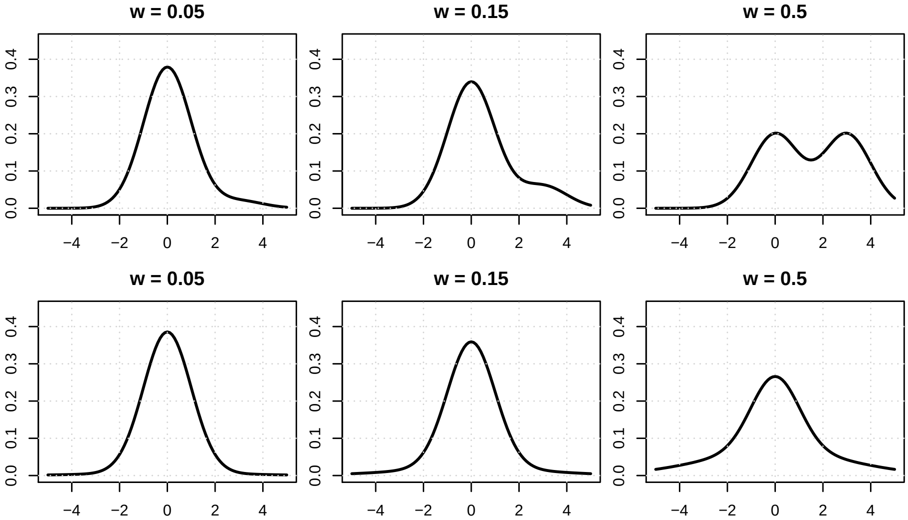
Family Name: Gaussian mixture distributions (countable)

Parameterization	$w_i \geq 0,\ \sum_{i\geq 1} w_i = 1, \quad \mu_i, \sigma_i^2$
Density definition	$f(x) = \sum_i w_i \frac{1}{\sigma_i} \phi\left(\frac{x-\mu_i}{\sigma_i}\right)$,
Moments	$\mu = \sum_i w_i \mu_i$,
	$\sigma^2 = \sum_i w_i \left((\mu_i - \bar{\mu})^2 + \sigma_i^2\right)$

Clearly, this can be extended to other base distributions (which may also be multivariate).

One may consider uncountable versions thereof with $f(x) = \int_\theta f(x;\theta)\, w(\theta)\, \mathrm{d}\theta$, w some pdf.

Some mixtures



Top: $w \cdot \mathcal{N}(0,1) + (1-w)\mathcal{N}(3,1)$; Bottom: $w \cdot \mathcal{N}(0,1) + (1-w)\mathcal{N}(0,3)$

Outline

Adding some variates

The univariate normal distribution discussed so far has a straightforward multivariate generalization.

Family Name: Multivariate Normal

Parameterization	$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_n^2 \end{pmatrix}$ $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Omega = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in \mathbf{R}^n,$ $\boldsymbol{\Sigma}$ is a $(n \times n)$ p.d. symmetric matrix$\}$
Density Definition	$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}\lvert\boldsymbol{\Sigma}\rvert^{1/2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$
Moments	$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}$, $\quad \mathrm{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$, $\quad \underset{(n\times 1)}{\boldsymbol{\mu}_3} = [\mathbf{0}]$
MGF	$M_{\boldsymbol{X}}(\boldsymbol{t}) = \exp\{\boldsymbol{\mu}'\boldsymbol{t} + (1/2)\,\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}\}$, where $\boldsymbol{t} = (t_1, \ldots, t_n)'$.

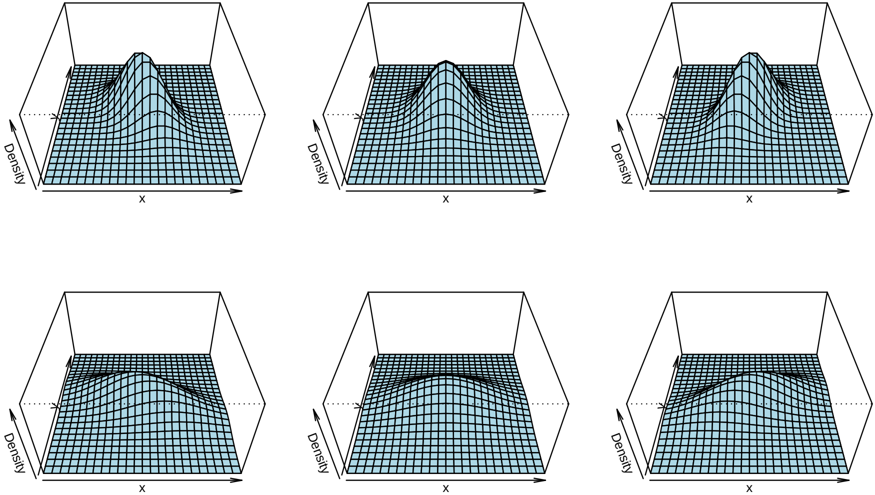
More details

The n-variate normal family of distribution is indexed by $n + n(n+1)/2$ parameters: In the mean vector ($\boldsymbol{\mu}$) n parameters and in the covariance matrix ($\boldsymbol{\Sigma}$) $n + (n^2 - n)/2$ parameters.

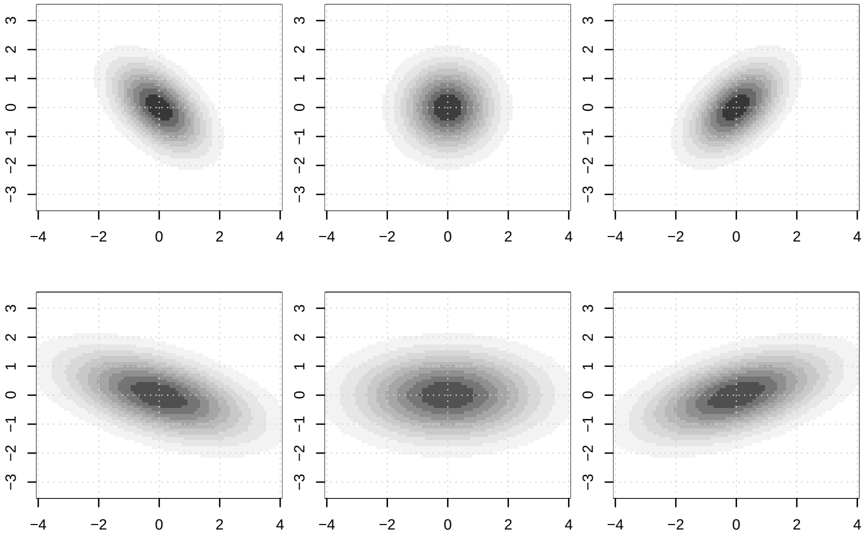
In order to illustrate graphically some of the characteristics of a multivariate Gaussian density, we consider the bivariate case with $n = 2$.

- The multivariate Gaussian density is bell-shaped and has its maximum at $\boldsymbol{x} = (x_1, x_2) = \boldsymbol{\mu} = (\mu_1, \mu_2)$.
- The iso-density contours, given by the set of points $(x_1, x_2) \in \{(x_1, x_2) : f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = c\}$, have the form of an ellipse. Its origin is given by $\boldsymbol{\mu}$ and its direction depends on $\boldsymbol{\Sigma}$.

Various bivariate normal pdfs



... and the contour plots



Properties of Multivariate Normal Distributions

A useful property is that linear combinations of a vector of multivariate normally distributed random variables are also normally distributed as stated in the following theorem.

Theorem (4.8)

Let $\boldsymbol{X}$ be an n-dimensional $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distributed random variable. Let $\mathbf{A}$ be any $(k \times n)$ matrix of constants with $\operatorname{rank}(\mathbf{A}) = k$, and let $\boldsymbol{b}$ be any $(k \times 1)$ vector of constants. Then the $(k \times 1)$ random vector $\boldsymbol{Y} = \mathbf{A}\boldsymbol{X} + \boldsymbol{b}$ is $\mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \boldsymbol{b}\ ,\ \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ distributed.

This theorem can be used to standardize a normally distributed random vector.

Standardizing in the multivariate case

- Let $\boldsymbol{Z}$ be a $\mathcal{N}(\mathbf{0}, \mathbf{I})$ distributed $(n \times 1)$ random vector, that is a vector of n uncorrelated $\mathcal{N}(0,1)$ distributed random variables.
- Then the $(n \times 1)$ random vector $\boldsymbol{Y}$ with a $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution can be represented in terms of $\boldsymbol{Z}$ as

$$\boldsymbol{Y} = \boldsymbol{\mu} + \mathbf{A}\boldsymbol{Z}, \qquad \text{where } \mathbf{A} \text{ is selected such that}^1 \qquad \mathbf{A}\mathbf{A}' = \boldsymbol{\Sigma}.$$

  This is because $\boldsymbol{Y} \sim \mathcal{N}(\mathbf{A}\mathbf{0} + \boldsymbol{\mu}\ ,\ \mathbf{A}\mathbf{I}\mathbf{A}') = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- Furthermore, the inversion of the function $\boldsymbol{Y} = \boldsymbol{\mu} + \mathbf{A}\boldsymbol{Z}$ *standardizes* the normally distributed vector $\boldsymbol{Y}$

$$\mathbf{A}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}) = \boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

[1] If $\mathbf{A}$ is a lower triangular matrix, we call it Cholesky factor, and $\mathbf{A}\mathbf{A}' = \boldsymbol{\Sigma}$ denotes the so-called Cholesky decomposition.

Margins are also normal

Theorem (4.9)

Let $\boldsymbol{Z}$ be an n-dimensional $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distributed random variable, where

$$\boldsymbol{Z} = \left[\begin{array}{c} \underset{(m \times 1)}{\boldsymbol{Z}_{(1)}} \\ \hline \underset{(n-m) \times 1}{\boldsymbol{Z}_{(2)}} \end{array}\right], \; \boldsymbol{\mu} = \left[\begin{array}{c} \underset{(m \times 1)}{\boldsymbol{\mu}_{(1)}} \\ \hline \underset{(n-m) \times 1}{\boldsymbol{\mu}_{(2)}} \end{array}\right], \; \boldsymbol{\Sigma} = \left[\begin{array}{c|c} \underset{(m \times m)}{\boldsymbol{\Sigma}_{11}} & \underset{m \times (n-m)}{\boldsymbol{\Sigma}_{12}} \\ \hline \underset{(n-m) \times m}{\boldsymbol{\Sigma}_{21}} & \underset{(n-m) \times (n-m)}{\boldsymbol{\Sigma}_{22}} \end{array}\right].$$

Then the marginal pdf of $\boldsymbol{Z}_{(1)}$ is $\mathcal{N}\left(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}\right)$, and the marginal PDF of $\boldsymbol{Z}_{(2)}$ is $N\left(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}\right)$.

Note that Theorem 4.9 can be applied to obtain the marginal pdf of *any* subset of the normal random variable $(Z_1, ..., Z_n)$ by simply ordering them appropriately in the definition of $\boldsymbol{Z}$ in the theorem.

... as are the conditional distributions

Theorem (4.10)

Let $\boldsymbol{Z}$ be an n-dimensional $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$-distributed random variable, where

$$\boldsymbol{Z} = \left[\begin{array}{c} \underset{(m\times 1)}{\boldsymbol{Z}_{(1)}} \\ \hline \underset{(n-m)\times 1}{\boldsymbol{Z}_{(2)}} \end{array}\right],\ \boldsymbol{\mu} = \left[\begin{array}{c} \underset{(m\times 1)}{\boldsymbol{\mu}_{(1)}} \\ \hline \underset{(n-m)\times 1}{\boldsymbol{\mu}_{(2)}} \end{array}\right],\ \boldsymbol{\Sigma} = \left[\begin{array}{c|c} \underset{(m\times m)}{\boldsymbol{\Sigma}_{11}} & \underset{m\times(n-m)}{\boldsymbol{\Sigma}_{12}} \\ \hline \underset{(n-m)\times m}{\boldsymbol{\Sigma}_{21}} & \underset{(n-m)\times(n-m)}{\boldsymbol{\Sigma}_{22}} \end{array}\right];$$

and let $\boldsymbol{z}^0$ be an n-dimensional vector of constants partitioned conformably with the partition $\boldsymbol{Z}$ into $\boldsymbol{z}^0_{(1)}$ and $\boldsymbol{z}^0_{(2)}$. Then,

$$\boldsymbol{Z}_{(1)}|(\boldsymbol{Z}_{(2)} = \boldsymbol{z}^0_{(2)}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left[\boldsymbol{z}^0_{(2)} - \boldsymbol{\mu}_{(2)}\right],\ \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)$$

$$\boldsymbol{Z}_{(2)}|(\boldsymbol{Z}_{(1)} = \boldsymbol{z}^0_{(1)}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{(2)} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\left[\boldsymbol{z}^0_{(1)} - \boldsymbol{\mu}_{(1)}\right],\ \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right).$$

The conditional expectation

Note that the mean of the conditional distribution given by

$$\mathrm{E}(\boldsymbol{Z}_{(1)}|\boldsymbol{Z}_{(2)} = \boldsymbol{z}_{(2)}) = \boldsymbol{\mu}_{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{z}_{(2)} - \boldsymbol{\mu}_{(2)}\right)$$

is a linear function in the 'conditioning variable' $\boldsymbol{z}_{(2)}$.

This linearity of the conditional mean is a specific feature of the multivariate normal distribution as a member of the family of *elliptically contoured distributions*.

Consider the special case where $Z_{(1)}$ is a scalar and $\boldsymbol{Z}_{(2)}$ is a $(k \times 1)$ vector. Then the conditional mean of $Z_{(1)}$ given $\boldsymbol{z}_{(2)}$, that is, the regression function of $Z_{(1)}$ on $\boldsymbol{Z}_{(2)}$ has the form

$$\mathrm{E}(Z_{(1)}|\boldsymbol{z}_{(2)}) = \underset{(1\times 1)}{a} + \underset{(1\times k)}{\boldsymbol{b}}\boldsymbol{z}_{(2)},$$

where $\quad a = \mu_{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_{(2)}, \qquad \boldsymbol{b} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}.$

Correlation and independence

The following theorem states that in the case of a normal distribution, zero covariance implies independence of the random variables, which in general is not true for other distributions.

Theorem (4.11)

Let $\boldsymbol{X} = (X_1, ..., X_n)'$ *be a* $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*-distributed random variable. Then* $(X_1, ..., X_n)$ *are independent iff* $\boldsymbol{\Sigma}$ *is a diagonal matrix with all covariances being zero.*

Quadratic forms

Sometimes, one may be interested in the behavior of so-called quadratic forms in $\boldsymbol{X}$,

$$Q = \boldsymbol{X}'\mathbf{A}\boldsymbol{X}$$

with $\mathbf{A}$ some conformable matrix.

- Means and variances of Q may be derived for Gaussian $\boldsymbol{X}$
- Quite useful: if $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} \sim \chi^2\left(\dim(\boldsymbol{X}), \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right)$$

  with $\chi^2(r, \lambda)$ a so-called non-central chi-squared distribution with r degrees of freedom and non-centrality parameter λ
- If $\lambda = 0$, the usual χ^2 with r degrees of freedom is recovered.

Outline

Coming up

More on modelling joint distributions