

$$e = \mathbf{Y} - \begin{pmatrix} \bar{Y} \\ \vdots \\ \bar{Y} \end{pmatrix}. \text{ Hence, } e^T e = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$\frac{\text{RSS}}{n-r} = \frac{\sum(Y_i - \bar{Y})^2}{\underbrace{n-1}_{=s^2 = \text{sample variance}}} = s^2$$

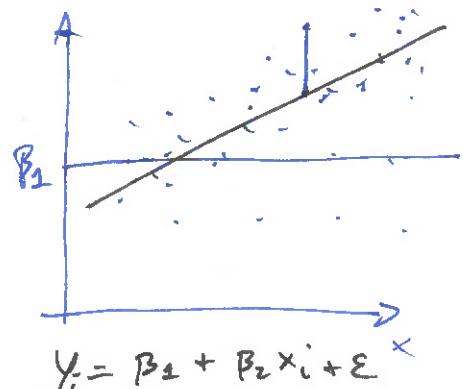
which we already know is unbiased for σ^2 .

$$E[s^2] = \sigma^2$$

Coefficient of Determination (R^2)

In the simplest model with only an intercept term, i.e. in

$$\mathbf{Y} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \beta_1 + \epsilon, \quad E\epsilon = \mathbf{0}$$



we have $\text{RSS} = \sum_{i=1}^n (Y_i - \bar{Y})^2$. Larger models, i.e. models with more columns in X will only lead to smaller RSS.

For models containing an *intercept term*, (i.e. X contains a column consisting of 1s (or any other constant)), a popular measure of the quality of a model is

$$R^2 = 1 - \frac{\text{RSS} \rightarrow \text{RSS OF YOUR MODEL}}{\sum_{i=1}^n (Y_i - \bar{Y})^2, \rightarrow \text{RSS OF THE SIMPLE MODEL}}$$

called the *coefficient of determination* or simply R^2 . A smaller RSS is "better", thus we want a large R^2 . Note: $0 \leq R^2 \leq 1$ and $R^2 = 1$ for a "perfect" model.

Remark (Intuitive interpretation) RSS/n is an estimator of σ^2 . $\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$ is an estimator of σ^2 in the model with only the intercept term (let us call this the "total variance").

Thus $\frac{\text{RSS}/n}{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} \approx \frac{\text{Variance in the model}}{\text{total variance}}$ and hence

$$R^2 \approx \frac{\text{total variance} - \text{variance in model}}{\text{total variance}}$$

Hence, $R^2 \approx$ fraction of the total variance of the data that "is explained" by the model.

$R^2 \geq 0.8$

Example 56**Boiling Point:** $R^2 = 0.995$ **Mammals:** Model $\log(\text{brain}_i) = \beta_1 + \log(\text{body}_i)\beta_2 + \epsilon_i$: $R^2 = 0.92$ Model: $\text{brain}_i = \beta_1 + \text{body}_i\beta_2 + \epsilon_i$: $R^2 = 0.87$

Note: These are unusually high values! Often, R^2 can be much smaller.

Remark Adding columns to \mathbf{X} will never decrease R^2 . Thus one should not use R^2 directly for model comparisons; one should penalise models with a larger number of parameters. More about this in Chapter 10.

10 Linear Models with Normal Theory Assumptions

In this Chapter we will again consider a linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\mathbf{E}\boldsymbol{\epsilon} = \mathbf{0}$. In order to construct confidence intervals or test hypotheses we need assumptions about the distribution of \mathbf{Y} (or equivalently about the distribution of $\boldsymbol{\epsilon}$).

We will work with the (NTA), which are $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$.

10.1 Distributional Results

We first define the multivariate normal distribution and some distributions constructed from it. After that some useful properties are shown.

10.1.1 The Multivariate Normal Distribution

The multivariate normal distribution, denoted by $N(\boldsymbol{\mu}, \Sigma)$, is a distribution of a random vector. It has two parameters: one vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and one positive semidefinite matrix $\Sigma \in \mathbb{R}^{n \times n}$. It will turn out that $\boldsymbol{\mu}$ is its expectation and Σ is its covariance.

It can be defined in several ways. In your previous courses you have defined it via the joint pdf as follows (this definition only works if Σ is positive definite):
 $\mathbf{Z} \sim N(\boldsymbol{\mu}, \Sigma)$ if \mathbf{Z} has a pdf of the form

$$f(\mathbf{z}) = \frac{1}{(\sqrt{2\pi})^n |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{z} - \boldsymbol{\mu})\right),$$

where $|\Sigma|$ denotes the determinant of Σ .

Example 57

$\mathbf{Z} \sim N(\boldsymbol{\mu}, \sigma^2 I)$ for some $\sigma^2 > 0$. Then

$$\begin{aligned} f(\mathbf{z}) &= \frac{1}{\sqrt{2\pi}^n |\sigma^2 I|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^\top (\sigma^{-2} I)(\mathbf{z} - \boldsymbol{\mu})\right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z_i - \mu_i)^2}{2\sigma^2}\right) \end{aligned}$$

Thus: Z_1, \dots, Z_n are independent, with $Z_i \sim N(\mu_i, \sigma^2)$, $i = 1, \dots, n$.

The three definitions mentioned below (which are all equivalent) will also work if Σ is only positive semidefinite.

Definition 23

- An n -variate random vector \mathbf{Z} follows a multivariate normal distribution if for all $\mathbf{a} \in \mathbb{R}^n$ the random variable $\mathbf{a}^T \mathbf{Z}$ follows a univariate normal distribution (the degenerate case $N(\cdot, 0)$ is allowed).
- Let $X_1, \dots, X_r \sim N(0, 1)$ be iid, let $\mu \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times r}$. Then $\mathbf{Z} = A\mathbf{X} + \mu \sim N(\mu, AA^T)$.
- $\mathbf{Z} \sim N(\mu, \Sigma)$ if its characteristic function $\phi : \mathbb{R}^n \rightarrow \mathbb{C}$, $\phi(\mathbf{t}) = E(\exp(i\mathbf{Z}^T \mathbf{t}))$ satisfies

$$\phi(\mathbf{t}) = \exp\left(i\mu^T \mathbf{t} - \frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t}\right) \quad \forall t \in \mathbb{R}^n.$$

where $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ is a positive semidefinite matrix.

$$E[e^{itx}]$$

CHARACTERISTIC FUNCTIONS ARE NOT EXAMINABLE

Remark (Concerning the definition via the characteristic function) You have previously met the moment generating function. One of the key results about moment generating functions is that the moment generating function uniquely identifies the distribution of a random variable.

Similarly, there is a moment generating function defined for n -variate random vectors \mathbf{X} , namely $M : \mathbb{R}^n \rightarrow \mathbb{R}, \mathbf{t} \mapsto E(\exp(\mathbf{t}^T \mathbf{X}))$. Again, M identifies the distribution of a random vector.

The characteristic function (which was used in the above definition), is often used instead of the moment generating function. It has similar properties (in particular it uniquely defines a distribution). The i in it is the complex number i . Furthermore, if $Z = X + iY$ is a complex-valued random variable then $E(Z) := E(X) + iE(Y)$.

Remark (Useful properties) Let $\mathbf{Z} \sim N(\mu, \Sigma)$. Then

- $E\mathbf{Z} = \mu$,
- $\text{cov } \mathbf{Z} = \Sigma$,
- if A is a deterministic matrix and \mathbf{b} is a deterministic vector of appropriate dimension then

$$A\mathbf{Z} + \mathbf{b} \sim N(A\mu + \mathbf{b}, A\Sigma A^T).$$

$$\text{ax} + b + N(a\mu + b, a^2\sigma^2)$$

In general: if X and Y are random variables then $\text{cov}(X, Y) = 0$ does not imply that X and Y are independent. To put it briefly: uncorrelated does not imply independence. The following lemma shows (in a general form) that *uncorrelated and jointly normal* does imply independence.

Lemma 14

For $i = 1, \dots, k$, let $A_i \in \mathbb{R}^{n_i \times n_i}$ be pos. semidef. and symmetric and let \mathbf{Z}_i be an n_i -

variate random vector with $\text{Cov}(\mathbf{Z}_i) = A_i$ for $i = 1, \dots, k$. If $\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_k \end{pmatrix} \sim N(\mu, \Sigma)$,
 for some $\mu \in \mathbb{R}^{\sum_{i=1}^k n_i}$ and $\Sigma = \text{diag}(A_1, \dots, A_k) = \begin{pmatrix} A_1 & & & 0 \\ & \ddots & & \\ 0 & & A_k & \end{pmatrix}$ then $\mathbf{Z}_1, \dots, \mathbf{Z}_k$
are independent.

Proof In the special case that all A_i are positive definite, this can be shown by using $\Sigma^{-1} = \text{diag}(A_1^{-1}, \dots, A_k^{-1})$ and $|\Sigma| = \prod_{i=1}^k |A_i|$ to factor the pdf.

The full proof works via the characteristic (or via the moment generating) functions; to show independence one needs to show that the characteristic functions can be written as product of the characteristic functions of the components, i.e. one needs to show $E \exp(it^T \mathbf{Z}) = \prod_{i=1}^k E \exp(it_i^T \mathbf{Z}_i)$ for all $t = (t_1^T, \dots, t_k^T)^T \in \mathbb{R}^n$.

Example 58

Let $k = 3$, $A_1 = 2 = (2)$, $A_2 = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$, $A_3 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Let

$$\Sigma = \begin{pmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & -0.5 & 0 & 0 \\ 0 & -0.5 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix}.$$

If $\mathbf{Z} \sim N(\mu, \Sigma)$ for some $\mu \in \mathbb{R}^5$ then $Z_1, \frac{(Z_2)}{(Z_3)}, \frac{(Z_4)}{(Z_5)}$ are independent.

$\hookrightarrow \mathbf{Z} = \begin{pmatrix} z_1 \\ \vdots \\ z_5 \end{pmatrix}, z_i \text{ is ONE DIMENSIONAL}$