Linear algebra

Sofia Olhede

EPFL

November 10, 2020

Testing for the Linear Effect

- Just as a remainder from yesterday let us look at the hypotheses:

$$H_0 : \mathbb{E}\, Y_i = \underbrace{\beta_1}_{\text{const.}} \quad \text{versus} \quad H_1 : \mathbb{E}\, Y_i = \underbrace{\beta_1 + \beta_2 x_i}_{\text{const + linear}}.$$

- In this example we have

$$\mathrm{X}_0 = \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix} = 1, \quad \mathrm{X} = \begin{pmatrix} 1 & x_1 \\ \dots & \dots \\ 1 & x_n \end{pmatrix}.$$

- Furthermore the matrix A is given by

$$A = \begin{pmatrix} 0 & 1 \end{pmatrix}.$$

- We then arrive at

$$P_0 = X_0 (X_0^T X_0)^{-1} X_0^T = \frac{1}{n} 1 1^T.$$

Testing for the Linear Effect II

- We also find

$$\mathrm{RSS} = \sum_{i=1}^{n} \left\{ Y_i - \bar{Y} + \frac{s_{xy}}{s_{xx}} \bar{x} - \frac{s_{xy}}{s_{xx}} x_i \right\}^2 \tag{1}$$

$$\mathrm{RSS}_0 = \sum_{i=1}^{n} \left\{ Y_i - \bar{Y} \right\}^2 \tag{2}$$

- We then compute

F-statistic: $$F = \frac{n-2}{n-1} \frac{\mathrm{RSS}_0 - \mathrm{RSS}}{\mathrm{RSS}},$$

and compare it to the quantiles of the F–distribution on 1 and $n-2$ degrees of freedom. This can be summarized in a table.

Some more linear algebra

If $\boldsymbol{Q}$ is an $n \times p$ real matrix, we define the column space (or *range*) of $\boldsymbol{Q}$ to be the set spanned by its columns:

$$\mathcal{M}(\boldsymbol{Q}) = \{\boldsymbol{y} \in \mathbb{R}^n : \exists \boldsymbol{\beta} \in \mathbb{R}^p, \ \boldsymbol{y} = \boldsymbol{Q}\boldsymbol{\beta}\}.$$

- Recall that $\mathcal{M}(\boldsymbol{Q})$ is a subspace of $\mathbb{R}^n$.
- The columns of $\boldsymbol{Q}$ provide a coordinate system for the subspace $\mathcal{M}(\boldsymbol{Q})$
- If $\boldsymbol{Q}$ is of full column rank (p), then the coordinates $\boldsymbol{\beta}$ corresponding to a $\boldsymbol{y} \in \mathcal{M}(\boldsymbol{Q})$ are unique.
- Allows interpretation of system of linear equations

$$\boldsymbol{Q}\boldsymbol{\beta} = \boldsymbol{y}.$$

[existence of solution $\leftrightarrow$ is $\boldsymbol{y}$ an element of $\mathcal{M}(\boldsymbol{Q})$?]
[uniqueness of solution $\leftrightarrow$ is there a unique coordinate vector $\boldsymbol{\beta}$?]

Some more linear algebra

Two further important subspaces associated with a real $n \times p$ matrix $\boldsymbol{Q}$:

- the null space (or *kernel*), $\ker(\boldsymbol{Q})$, of $\boldsymbol{Q}$ is the subspace defined as

$$\ker(\boldsymbol{Q}) = \{x \in \mathbb{R}^p : \boldsymbol{Q}\boldsymbol{x} = 0\};$$

- the orthogonal complement of $\mathcal{M}(\boldsymbol{Q})$, $\mathcal{M}^\perp(\boldsymbol{Q})$, is the subspace defined as

$$\begin{aligned} \mathcal{M}^\perp(\boldsymbol{Q}) &= \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{x} = 0, \ \forall \boldsymbol{x} \in \mathbb{R}^p\} \\ &= \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{y}^\top \boldsymbol{v} = 0, \ \forall \boldsymbol{v} \in \mathcal{M}(\boldsymbol{Q})\}. \end{aligned}$$

The orthogonal complement may be defined for arbitrary subspaces by using the second equality.

Some more linear algebra

Theorem (Spectral Theorem)

A $p \times p$ matrix $\boldsymbol{Q}$ is symmetric if and only if there exists a $p \times p$ orthogonal matrix $\boldsymbol{U}$ and a diagonal matrix $\boldsymbol{\Lambda}$ such that

$$\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top.$$

In particular:

1. *the columns of $\boldsymbol{U} = (\boldsymbol{u}_1 \ \cdots \ \boldsymbol{u}_p)$ are eigenvectors of $\boldsymbol{Q}$, i.e. there exist λ_j such that*
$$\boldsymbol{Q}\boldsymbol{u}_j = \lambda_j \boldsymbol{u}_j, \qquad j = 1, \ldots, p;$$
2. *the entries of $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ are the corresponding eigenvalues of $\boldsymbol{Q}$, which are real; and*
3. *the rank of $\boldsymbol{Q}$ is the number of non-zero eigenvalues.*

Note: if the eigenvalues are distinct, the eigenvectors are unique (up to changes in signs).

Some more linear algebra

Theorem (Singular Value Decomposition)

Any $n \times p$ real matrix can be factorised as

$$\underset{n\times p}{\boldsymbol{Q}} = \underset{n\times n}{\boldsymbol{U}}\ \underset{n\times p}{\boldsymbol{\Sigma}}\ \underset{p\times p}{\boldsymbol{V}}^{\top},$$

where $\boldsymbol{U}$ and $\boldsymbol{V}^{\top}$ are orthogonal with columns called left singular vectors and right singular vectors, respectively, and Σ is diagonal with real entries called singular values.

1. The left singular vectors are eigenvectors of $\boldsymbol{Q}\boldsymbol{Q}^{\top}$.
2. The right singular vectors are eigenvectors of $\boldsymbol{Q}^{\top}\boldsymbol{Q}$.
3. The squares of the singular values are eigenvalues of both $\boldsymbol{Q}\boldsymbol{Q}^{\top}$ and $\boldsymbol{Q}^{\top}\boldsymbol{Q}$.
4. The left singular vectors corresponding to non-zero singular values form an orthonormal basis for $\mathcal{M}(\boldsymbol{Q})$.
5. The left singular vectors corresponding to zero singular values form an orthonormal basis for $\mathcal{M}^{\perp}(\boldsymbol{Q})$.

Some more linear algebra

A matrix $\boldsymbol{Q}$ is called idempotent if $\boldsymbol{Q}^2 = \boldsymbol{Q}$.

An orthogonal projection (henceforth projection) onto a subspace $\mathcal{V}$ is a symmetric idempotent matrix $\boldsymbol{H}$ such that $\mathcal{M}(\boldsymbol{H}) = \mathcal{V}$.

Proposition

The only possible eigenvalues of a projection matrix are 0 and 1.

Proposition

Let $\mathcal{V}$ be a subspace and $\boldsymbol{H}$ be a projection onto $\mathcal{V}$. Then $\boldsymbol{I} - \boldsymbol{H}$ is the projection matrix onto $\mathcal{V}^{\perp}$.

Proof (∗).

$(\boldsymbol{I} - \boldsymbol{H})^\top = \boldsymbol{I} - \boldsymbol{H}^\top = \boldsymbol{I} - \boldsymbol{H}$ since $\boldsymbol{H}$ is symmetric and,
$(\boldsymbol{I} - \boldsymbol{H})^2 = \boldsymbol{I}^2 - 2\boldsymbol{H} + \boldsymbol{H}^2 = \boldsymbol{I} - \boldsymbol{H}$. Thus $\boldsymbol{I} - \boldsymbol{H}$ is a projection matrix.

It remains to identify the column space of $\boldsymbol{I} - \boldsymbol{H}$. Let $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ be the spectral decomposition of $\boldsymbol{H}$. Then $\boldsymbol{I} - \boldsymbol{H} = \boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top = \boldsymbol{U}(\boldsymbol{I} - \boldsymbol{\Lambda})\boldsymbol{U}^\top$. Hence the column space of $\boldsymbol{I} - \boldsymbol{H}$ is spanned by the eigenvectors of $\boldsymbol{H}$ corresponding to zero eigenvalues of $\boldsymbol{H}$, which coincides with $\mathcal{M}^{\perp}(\boldsymbol{H}) = \mathcal{V}^{\perp}$. $\square$

Some more linear algebra

Proposition

Let $\mathcal{V}$ be a subspace and $\boldsymbol{H}$ be a projection onto $\mathcal{V}$. Then $\boldsymbol{H}\boldsymbol{y} = \boldsymbol{y}$ for all $\boldsymbol{y} \in \mathcal{V}$.

Proposition

If $\boldsymbol{P}$ and $\boldsymbol{Q}$ are projection matrices onto a subspace $\mathcal{V}$, then $\boldsymbol{P} = \boldsymbol{Q}$.

Proposition

If $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ are linearly independent and are such that span$(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p) = \mathcal{V}$, then the projection onto $\mathcal{V}$ can be represented as

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$$

where $\boldsymbol{X}$ is a matrix with columns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$.

Some more linear algebra

Proposition

Let $\mathcal{V}$ be a subspace of $\mathbb{R}^n$ and $\boldsymbol{H}$ be a projection onto $\mathcal{V}$. Then

proof next slide

$$||\boldsymbol{x} - \boldsymbol{H}\boldsymbol{x}|| \leq ||\boldsymbol{x} - \boldsymbol{v}||, \qquad \forall \boldsymbol{v} \in \mathcal{V}.$$

Proof (∗).

Let $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ be the spectral decomposition of $\boldsymbol{H}$, $\boldsymbol{U} = (\boldsymbol{u}_1 \ \cdots \ \boldsymbol{u}_n)$ and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Letting $p = \dim(\mathcal{V})$,

1. $\lambda_1 = \cdots = \lambda_p = 1$ and $\lambda_{p+1} = \cdots = \lambda_n = 0$,
2. $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ is an orthonormal basis of $\mathbb{R}^n$,
3. $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$ is an an orthonormal basis of $\mathcal{V}$.

Some more linear algebra

$$\begin{aligned}
\|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{x}\|^2 &= \sum_{i=1}^{n} (\boldsymbol{x}^\top \boldsymbol{u}_i - (\boldsymbol{H}\boldsymbol{x})^\top \boldsymbol{u}_i)^2 && \text{[orthonormal basis]} \\
&= \sum_{i=1}^{n} (\boldsymbol{x}^\top \boldsymbol{u}_i - \boldsymbol{x}^\top \boldsymbol{H}\boldsymbol{u}_i)^2 && [H \text{ is symmetric}] \\
&= \sum_{i=1}^{n} (\boldsymbol{x}^\top \boldsymbol{u}_i - \lambda_i \boldsymbol{x}^\top \boldsymbol{u}_i)^2 && [u\text{'s are eigenvectors of } H] \\
&= 0 + \sum_{i=p+1}^{n} (\boldsymbol{x}^\top \boldsymbol{u}_i)^2 && \text{[eigenvalues 0 or 1]} \\
&\leq \sum_{i=1}^{p} (\boldsymbol{x}^\top \boldsymbol{u}_i - \boldsymbol{v}^\top \boldsymbol{u}_i)^2 + \sum_{i=p+1}^{n} (\boldsymbol{x}^\top \boldsymbol{u}_i)^2 && \forall \boldsymbol{v} \in \mathcal{V} \\
&= \|\boldsymbol{x} - \boldsymbol{v}\|^2.
\end{aligned}$$

$\square$

Some more linear algebra

Proposition

Let $\mathcal{V}_1 \subseteq \mathcal{V} \subseteq \mathbb{R}^n$ *be two nested linear subspaces. If* $\boldsymbol{H}_1$ *is the projection onto* $\mathcal{V}_1$ *and* $\boldsymbol{H}$ *is the projection onto* $\mathcal{V}$, *then*

$$\boldsymbol{H}\boldsymbol{H}_1 = \boldsymbol{H}_1 = \boldsymbol{H}_1\boldsymbol{H}.$$

Proof $(*)$.

First we show that $\boldsymbol{H}\boldsymbol{H}_1 = \boldsymbol{H}_1$, and then that $\boldsymbol{H}_1\boldsymbol{H} = \boldsymbol{H}\boldsymbol{H}_1$. For all $\boldsymbol{y} \in \mathbb{R}^n$ we have $\boldsymbol{H}_1\boldsymbol{y} \in \mathcal{V}_1$. But then $\boldsymbol{H}_1\boldsymbol{y} \in \mathcal{V}$, since $\mathcal{V}_1 \subseteq \mathcal{V}$.
Therefore $\boldsymbol{H}\boldsymbol{H}_1\boldsymbol{y} = \boldsymbol{H}_1\boldsymbol{y}$. We have shown that $(\boldsymbol{H}\boldsymbol{H}_1 - \boldsymbol{H}_1)\boldsymbol{y} = 0$ for all $\boldsymbol{y} \in \mathbb{R}^n$, so that $\boldsymbol{H}\boldsymbol{H}_1 - \boldsymbol{H}_1 = 0$, as its kernel is all $\mathbb{R}^n$. Hence $\boldsymbol{H}\boldsymbol{H}_1 = \boldsymbol{H}_1$.

To prove that $\boldsymbol{H}_1\boldsymbol{H} = \boldsymbol{H}\boldsymbol{H}_1$, note that symmetry of projection matrices and the first part of the proof give

$$\boldsymbol{H}_1\boldsymbol{H} = \boldsymbol{H}_1^\top\boldsymbol{H}^\top = (\boldsymbol{H}\boldsymbol{H}_1)^\top = (\boldsymbol{H}_1)^\top = \boldsymbol{H}_1 = \boldsymbol{H}\boldsymbol{H}_1.$$

□

Some more linear algebra

Definition (Non-Negative Matrix – Quadratic Form Definition)

A $p \times p$ real symmetric matrix $\boldsymbol{\Omega}$ is called non-negative definite (written $\boldsymbol{\Omega} \succeq 0$) if and only if $\boldsymbol{x}^\top \boldsymbol{\Omega} \boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^p$. If $\boldsymbol{x}^\top \boldsymbol{\Omega} \boldsymbol{x} > 0$ for all $\boldsymbol{x} \in \mathbb{R}^p \setminus \{0\}$, then we call $\boldsymbol{\Omega}$ positive definite (written $\boldsymbol{\Omega} \succ 0$).

Definition (Non-Negative Matrix – Spectral Definition)

A $p \times p$ real symmetric matrix $\boldsymbol{\Omega}$ is called non-negative definite (written $\boldsymbol{\Omega} \succeq 0$) if and only the eigenvalues of $\boldsymbol{\Omega}$ are non-negative. If the eigenvalues of $\boldsymbol{\Omega}$ are strictly positive, then $\boldsymbol{\Omega}$ is called positive definite (written $\boldsymbol{\Omega} \succ 0$).

Lemma (Little exercise)

The two definitions are equivalent.

Proposition (Non-Negative and Covariance Matrices)

Let $\boldsymbol{\Omega}$ be a real symmetric matrix. Then $\boldsymbol{\Omega}$ is non-negative definite if and only if $\boldsymbol{\Omega}$ is the covariance matrix of some random vector $\boldsymbol{Y}$.

Some more linear algebra

- Let $\boldsymbol{Y}$ be a random vector in $\mathbb{R}^d$ with covariance matrix $\boldsymbol{\Omega}$.
- Find direction $\boldsymbol{v}_1 \in \mathbb{S}^{d-1}$ such that the projection of $\boldsymbol{Y}$ onto $\boldsymbol{v}_1$ has maximal variance.
- For $j = 2, 3, \ldots, d$, find direction $\boldsymbol{v}_j \perp \{\boldsymbol{v}_1, ..., \boldsymbol{v}_{j-1}\}$ such that projection of $\boldsymbol{Y}$ onto $\boldsymbol{v}_j$ has maximal variance.

Solution: maximise $\underline{\mathrm{var}(\boldsymbol{v}_1^\top \boldsymbol{Y}) = \boldsymbol{v}_1^\top \boldsymbol{\Omega} v_1}$ over $\|\boldsymbol{v}_1\| = 1$

spectral decomposition

$$\boldsymbol{v}_1^\top \boldsymbol{\Omega} \boldsymbol{v}_1 = \boldsymbol{v}_1^\top \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^\top \boldsymbol{v}_1 = \|\boldsymbol{\Lambda}^{1/2} \boldsymbol{U}^\top \boldsymbol{v}_1\|^2 = \sum_{i=1}^{d} \lambda_i (\boldsymbol{u}_i^\top \boldsymbol{v}_1)^2 \qquad \text{[change of basis]}$$

Now $\sum_{i=1}^{d} (\boldsymbol{u}_i^\top \boldsymbol{v}_1)^2 = \|\boldsymbol{v}_1\|^2 = 1$ so we have a convex combination of $\{\lambda_j\}_{j=1}^{d}$,

$$\sum_{i=1}^{d} p_i \lambda_i, \qquad \sum_{i} p_i = 1, \quad p_i \geq 0, \quad i = 1, \ldots, d.$$

But $\lambda_1 \geq \lambda_i \geq 0$ so clearly this sum is maximised when $p_1 = 1$ and $p_j = 0$ $\forall j \neq 1$, i.e. $\boldsymbol{v}_1 = \pm \boldsymbol{u}_1$.

Iteratively, $\boldsymbol{v}_j = \pm \boldsymbol{u}_j$, i.e. principal components are eigenvectors of $\boldsymbol{\Omega}$.

Some more linear algebra

Theorem (Optimal (Linear) Dimension Reduction Theorem)

Let $\boldsymbol{Y}$ be a mean-zero random variable in $\mathbb{R}^d$ with $d \times d$ covariance $\boldsymbol{\Omega}$. Let $\boldsymbol{H}$ be the projection matrix onto the span of the first k eigenvectors of $\boldsymbol{\Omega}$. Then

$$\mathbb{E}\|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{Y}\|^2 \leq \mathbb{E}\|\boldsymbol{Y} - \boldsymbol{Q}\boldsymbol{Y}\|^2$$

for any $d \times d$ projection matrix $\boldsymbol{Q}$ or rank at most k.

Intuitively: if you want to approximate a mean-zero random variable taking values $\mathbb{R}^d$ by a random variable that ranges over a subspace of dimension at most $k \leq d$, the optimal choice is the projection of the random variable onto the space spanned by its first k principal components (eigenvectors of the covariance). "Optimal" is with respect to the mean squared error.

For the proof, use lemma below (follows immediately from spectral decomposition)

Lemma

$\boldsymbol{Q}$ is a rank k projection matrix if and only if there exist orthonormal vectors $\{\boldsymbol{v}_j\}_{j=1}^k$ such that $\boldsymbol{Q} = \sum_{j=1}^k \boldsymbol{v}_i \boldsymbol{v}_i^\top$.

Some more linear algebra

Proof of Optimal Linear Dimension Reduction (∗).

?

Write $\boldsymbol{Q} = \sum_{j=1}^{k} \boldsymbol{v}_i \boldsymbol{v}_i^\top$ for some orthonormal $\{\boldsymbol{v}_j\}_{j=1}^{k}$. Then
$\mathbb{E}\|\boldsymbol{Y} - \boldsymbol{Q}\boldsymbol{Y}\|^2 =$

$$
\begin{aligned}
&= \mathbb{E}\left[\boldsymbol{Y}^\top(\boldsymbol{I}-\boldsymbol{Q})^\top(\boldsymbol{I}-\boldsymbol{Q})\boldsymbol{Y}\right] = \mathbb{E}\left[\mathrm{tr}\{(\boldsymbol{I}-\boldsymbol{Q})\boldsymbol{Y}\boldsymbol{Y}^\top(\boldsymbol{I}-\boldsymbol{Q})^\top\}\right] \\
&= \mathrm{tr}\{(\boldsymbol{I}-\boldsymbol{Q})\mathbb{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right](\boldsymbol{I}-\boldsymbol{Q})^\top\} = \mathrm{tr}\{(\boldsymbol{I}-\boldsymbol{Q})^\top(\boldsymbol{I}-\boldsymbol{Q})\Omega\} \\
&= \mathrm{tr}\{(\boldsymbol{I}-\boldsymbol{Q})\Omega\} = \mathrm{tr}\{\boldsymbol{\Omega}\} - \mathrm{tr}\{\boldsymbol{Q}\boldsymbol{\Omega}\} = \sum_{i=1}^{d}\lambda_i - \mathrm{tr}\left\{\sum_{j=1}^{k}\boldsymbol{v}_i\boldsymbol{v}_i^\top\boldsymbol{\Omega}\right\} \\
&= \sum_{i=1}^{d}\lambda_i - \sum_{j=1}^{k}\mathrm{tr}\left\{\boldsymbol{v}_i\boldsymbol{v}_i^\top\boldsymbol{\Omega}\right\} = \sum_{i=1}^{d}\lambda_i - \sum_{j=1}^{k}\boldsymbol{v}_i\boldsymbol{\Omega}\boldsymbol{v}_i^\top \\
&= \sum_{i=1}^{d}\lambda_i - \sum_{j=1}^{k}\mathrm{var}[\boldsymbol{v}_i^\top\boldsymbol{Y}]
\end{aligned}
$$

If we can minimise this expression over all $\{\boldsymbol{v}_j\}_{j=1}^{k}$ with $\boldsymbol{v}_i^\top\boldsymbol{v}_j = \mathbf{1}\{i=j\}$, then we're done. By PCA, this is done by choosing the top k eigenvectors of $\boldsymbol{\Omega}$. □

Some more linear algebra

Corollary (Deterministic Version)

Let $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_p\} \subset \mathbb{R}^d$ be such that $\boldsymbol{x}_1 + \ldots + \boldsymbol{x}_p = 0$, and let $\boldsymbol{X}$ be the matrix with columns $\{\boldsymbol{x}_i\}_{i=1}^p$. The best approximating k-hyperplane to the points $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_p\}$ is given by the span of the first k eigenvectors of the matrix $\boldsymbol{X}\boldsymbol{X}^\top$, i.e. if $\boldsymbol{H}$ is the projection onto this span, it holds that

$$\sum_{i=1}^{p} ||\boldsymbol{x}_i - \boldsymbol{H}\boldsymbol{x}_i||^2 \leq \sum_{i=1}^{p} ||\boldsymbol{x}_i - \boldsymbol{Q}\boldsymbol{x}_i||^2$$

for any $d \times d$ projection operator $\boldsymbol{Q}$ or rank at most k.

Proof.

Define the discrete random vector $\boldsymbol{Y}$ by $\mathbb{P}[\boldsymbol{Y} = \boldsymbol{x}_i] = 1/p$, and use optimal linear dimension reduction as stated earlier. $\square$

Some more linear algebra

Definition (Multivariate Gaussian Distribution)

A random vector $\boldsymbol{Y}$ in $\mathbb{R}^d$ has the multivariate normal distribution if and only if $\boldsymbol{\beta}^\top \boldsymbol{Y}$ has the univariate normal distribution, $\forall \boldsymbol{\beta} \in \mathbb{R}^d$.

How can we use this definition to determine basic properties?

Recall that the *moment generating function* (MGF) of a random vector $\boldsymbol{W}$ in $\mathbb{R}^d$ is defined as

$$M_{\boldsymbol{W}}(\boldsymbol{\theta}) = \mathbb{E}[e^{\boldsymbol{\theta}^\top \boldsymbol{W}}], \qquad \boldsymbol{\theta} \in \mathbb{R}^d,$$

provided the expectation exists. When the MGF exists *it characterises the distribution of the random vector. Furthermore, two random vectors are independent if and only if their joint MGF is the product of their marginal MGF's.*

Some more linear algebra

Most important facts about Gaussian vectors:

1. Moment generating function of $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega})$:

$$\boxed{M_{\boldsymbol{Y}}(\boldsymbol{u}) = \exp\left(\boldsymbol{u}^\top \boldsymbol{\mu} + \frac{1}{2}\boldsymbol{u}^\top \boldsymbol{\Omega} \boldsymbol{u}\right)}.$$

2. $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{p\times 1}, \boldsymbol{\Omega}_{p\times p})$ and given $\boldsymbol{B}_{n\times p}$ and $\boldsymbol{\theta}_{n\times 1}$, then
$$\boldsymbol{\theta} + \boldsymbol{B}\,\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\theta} + \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Omega}\boldsymbol{B}^\top).$$

3. $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega})$ density, assuming $\boldsymbol{\Omega}$ nonsingular:

$$\boxed{f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{(2\pi)^{p/2}\,|\boldsymbol{\Omega}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right\}}.$$

4. Constant density isosurfaces are ellipsoidal
5. Marginals of Gaussian are Gaussian (converse NOT true).
6. $\boldsymbol{\Omega}$ diagonal $\Leftrightarrow$ independent coordinates Y_j.
7. If $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{p\times 1}, \boldsymbol{\Omega}_{p\times p})$,
$$\boldsymbol{A}\boldsymbol{Y} \text{ independent of } \boldsymbol{B}\boldsymbol{Y} \iff \boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{B}^\top = 0$$

Some more linear algebra

Proposition (Property 1: Moment Generating Function)

The moment generating function of $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega})$ *is*

$$M_{\boldsymbol{Y}}(\boldsymbol{u}) = \exp\left(\boldsymbol{u}^\top \boldsymbol{\mu} + \tfrac{1}{2}\boldsymbol{u}^\top \boldsymbol{\Omega} \boldsymbol{u}\right)$$

Proof (∗).

Let $\boldsymbol{u} \in \mathbb{R}^d$ be arbitrary. Then $\boldsymbol{u}^\top \boldsymbol{Y}$ is Gaussian with mean $\boldsymbol{u}^\top \boldsymbol{\mu}$ and variance $\boldsymbol{u}^\top \boldsymbol{\Omega} \boldsymbol{u}$. Hence it has moment generating function:

$$M_{\boldsymbol{u}^\top \boldsymbol{Y}}(t) = \mathbb{E}\left(e^{t\boldsymbol{u}^\top \boldsymbol{Y}}\right) = \exp\left\{t(\boldsymbol{u}^\top \boldsymbol{\mu}) + \frac{t^2}{2}(\boldsymbol{u}^\top \boldsymbol{\Omega} \boldsymbol{u})\right\}.$$

Now take $t = 1$ and observe that

$$M_{\boldsymbol{u}^\top \boldsymbol{Y}}(1) = \mathbb{E}\left(e^{\boldsymbol{u}^\top \boldsymbol{Y}}\right) = M_{\boldsymbol{Y}}(\boldsymbol{u}).$$

Combining the two, we conclude that

$$M_{\boldsymbol{Y}}(\boldsymbol{u}) = \exp\left(\boldsymbol{u}^\top \boldsymbol{\mu} + \frac{1}{2}\boldsymbol{u}^\top \boldsymbol{\Omega} \boldsymbol{u}\right), \quad \boldsymbol{u} \in \mathbb{R}^d.$$

□

Some more linear algebra

Proposition (Property 2: Affine Transformation)

For $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{p\times 1}, \boldsymbol{\Omega}_{p\times p})$ *and given* $\boldsymbol{B}_{n\times p}$ *and* $\boldsymbol{\theta}_{n\times 1}$, *we have*

$$\boldsymbol{\theta} + \boldsymbol{B}\,\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\theta} + \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Omega}\boldsymbol{B}^\top)$$

Proof (∗).

$$\begin{aligned}
M_{\boldsymbol{\theta}+\boldsymbol{B}\,\boldsymbol{Y}}(u) &= \mathbb{E}\left[\exp\{\boldsymbol{u}^\top(\boldsymbol{\theta}+\boldsymbol{B}\,\boldsymbol{Y})\}\right] = \exp\left\{\boldsymbol{u}^\top\boldsymbol{\theta}\right\}\mathbb{E}\left[\exp\{(\boldsymbol{B}^\top\boldsymbol{u})^\top\boldsymbol{Y}\}\right] \\
&= \exp\left\{\boldsymbol{u}^\top\boldsymbol{\theta}\right\} M_{\boldsymbol{Y}}(\boldsymbol{B}^\top\boldsymbol{u}) \\
&= \exp\left\{\boldsymbol{u}^\top\boldsymbol{\theta}\right\}\exp\left\{(\boldsymbol{B}^\top\boldsymbol{u})^\top\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{u}^\top\boldsymbol{B}\boldsymbol{\Omega}\boldsymbol{B}^\top\boldsymbol{u}\right\} \\
&= \exp\left\{\boldsymbol{u}^\top\boldsymbol{\theta} + \boldsymbol{u}^\top(\boldsymbol{B}\boldsymbol{\mu}) + \frac{1}{2}\boldsymbol{u}^\top\boldsymbol{B}\boldsymbol{\Omega}\boldsymbol{B}^\top\boldsymbol{u}\right\} \\
&= \exp\left\{\boldsymbol{u}^\top(\boldsymbol{\theta}+\boldsymbol{B}\boldsymbol{\mu}) + \frac{1}{2}\boldsymbol{u}^\top\boldsymbol{B}\boldsymbol{\Omega}\boldsymbol{B}^\top\boldsymbol{u}\right\}
\end{aligned}$$

And this last expression is the MGF of a $\mathcal{N}(\boldsymbol{\theta} + \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Omega}\boldsymbol{B}^\top)$ distribution. □

Some more linear algebra

Proposition (Property 3: Density Function)

Let $\boldsymbol{\Omega}_{p\times p}$ be nonsingular. The density of $\mathcal{N}(\boldsymbol{\mu}_{p\times 1}, \boldsymbol{\Omega}_{p\times p})$ is

$$f_Y(\boldsymbol{y}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Omega}|^{1/2}} \exp\left\{-\tfrac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right\}$$

Proof (∗).

Let $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^\top$ be a vector of iid $\mathcal{N}(0,1)$ random variables. Then, because of independence,

(a) the density of $\boldsymbol{Z}$ is

$$f_Z(\boldsymbol{z}) = \prod_{i=1}^p f_{Z_i}(z_i) = \prod_{i=1}^p \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z_i^2\right) = \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}\boldsymbol{z}^\top \boldsymbol{z}\right).$$

(b) The MGF of $\boldsymbol{Z}$ is

$$M_Z(\boldsymbol{u}) = \mathbb{E}\left\{\exp\left(\sum_{i=1}^p u_i Z_i\right)\right\} = \prod_{i=1}^p \mathbb{E}\{\exp(u_i Z_i)\} = \exp(\boldsymbol{u}^\top \boldsymbol{u}/2),$$

which is the MGF of a p-variate $\mathcal{N}(0, \boldsymbol{I})$ distribution.

Some more linear algebra

$\overset{(a)+(b)}{\Longrightarrow}$ the $\mathcal{N}(0, \boldsymbol{I})$ density is $f_Z(\boldsymbol{z}) = \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}\boldsymbol{z}^\top \boldsymbol{z}\right)$.

By the spectral theorem, $\boldsymbol{\Omega}$ admits a square root, $\boldsymbol{\Omega}^{1/2}$. Furthermore, since $\boldsymbol{\Omega}$ is non-singular, so is $\boldsymbol{\Omega}^{1/2}$.

Now observe that from our Property 2, we have $\boldsymbol{Y} \overset{d}{=} \boldsymbol{\Omega}^{1/2}\boldsymbol{Z} + \boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega})$.

By the change of variables formula,

$$
\begin{aligned}
f_Y(\boldsymbol{y}) &= f_{\boldsymbol{\Omega}^{1/2}Z+\boldsymbol{\mu}}(\boldsymbol{y}) \\
&= |\boldsymbol{\Omega}^{-1/2}| f_Z\{\Omega^{-1/2}(\boldsymbol{y}-\boldsymbol{\mu})\} \\
&= \frac{1}{(2\pi)^{p/2}\,|\boldsymbol{\Omega}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right\}.
\end{aligned}
$$

[Recall that to obtain the density of $\boldsymbol{W} = g(\boldsymbol{X})$ at $\boldsymbol{w}$, we need to evaluate f_X at $g^{-1}(\boldsymbol{w})$ but also multiply by the Jacobian determinant of g^{-1} at $\boldsymbol{w}$.]

$\square$

Some more linear algebra

Proposition (Property 4: Isosurfaces)

The isosurfaces of a $\mathcal{N}(\boldsymbol{\mu}_{p\times 1}, \boldsymbol{\Omega}_{p\times p})$ are $(p-1)$-dimensional ellipsoids centred at $\boldsymbol{\mu}$, with principal axes given by the eigenvectors of $\boldsymbol{\Omega}$ and with anisotropies given by the ratios of the square roots of the corresponding eigenvalues of $\boldsymbol{\Omega}$.

Proof (∗).

Exercise: Use Property 3, and the spectral theorem. □

Proposition (Property 5: Coordinate Distributions)

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^\top \sim \mathcal{N}(\boldsymbol{\mu}_{p\times 1}, \boldsymbol{\Omega}_{p\times p})$. Then $Y_j \sim \mathcal{N}(\mu_j, \Omega_{jj})$.

Proof (∗).

Observe that $Y_j = (0\ ,\ 0\ , \ldots ,\ \underbrace{1}_{\text{j}th \text{ position}}\ , \ldots ,\ 0\ ,\ 0)\,\boldsymbol{Y}$ and use Property 2. □

Some more linear algebra

Proposition (Property 6: Diagonal $\boldsymbol{\Omega} \iff$ Independence)

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^\top \sim \mathcal{N}(\boldsymbol{\mu}_{p\times 1}, \boldsymbol{\Omega}_{p\times p})$. Then the Y_i are mutually independent if and only if $\boldsymbol{\Omega}$ is diagonal.

Proof (∗).

Suppose that the Y_j are independent. Property 5 yields $Y_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ for some $\sigma_j > 0$. Thus the density of $\boldsymbol{Y}$ is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \prod_{j=1}^{p} f_{Y_j}(y_j) = \prod_{i=1}^{p} \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(y_j-\mu_j)^2}{\sigma_j^2}\right\}$$

$$= \frac{1}{(2\pi)^{p/2}\,|\mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \mathrm{diag}(\sigma_1^{-2}, \ldots, \sigma_p^{-2})(\boldsymbol{y}-\boldsymbol{\mu})\right\}.$$

Hence $\boldsymbol{Y} \sim \mathcal{N}\{\boldsymbol{\mu}, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)\}$, i.e. the covariance $\boldsymbol{\Omega}$ is diagonal.

Conversely, assume $\boldsymbol{\Omega}$ is diagonal, say $\boldsymbol{\Omega} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$. Then we can reverse the steps of the first part to see that the joint density $f_{\boldsymbol{Y}}(\boldsymbol{y})$ can be written as a product of the marginal densities $f_{Y_j}(y_j)$, thus proving independence. □

Some more linear algebra

Proposition (Property 7: AY, BY indep $\iff$ $A\Omega B^\top = 0$)

If $Y \sim \mathcal{N}(\mu_{p\times 1}, \Omega_{p\times p})$, and $A_{m\times p}$, $B_{d\times p}$ be real matrices. Then,

$$AY \text{ independent of } BY \iff A\Omega B^\top = 0.$$

Proof (∗). [wlog assuming $\mu = 0$ (simplifies the algebra)]

First assume $A\Omega B^\top = 0$. Let $W_{(m+d)\times 1} = \binom{AY}{BY}$ and $\theta_{(m+d)\times 1} = \binom{u_{m\times 1}}{v_{d\times 1}}$.

$$\begin{aligned}
M_W(\theta) &= \mathbb{E}[\exp\{W^\top\theta\}] = \mathbb{E}\left[\exp\left\{Y^\top A^\top u + Y^\top B^\top v\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{Y^\top(A^\top u + B^\top v)\right\}\right] = M_Y(A^\top u + B^\top v) \\
&= \exp\left\{\frac{1}{2}(A^\top u + B^\top v)^\top \Omega (A^\top u + B^\top v)\right\}
\end{aligned}$$

$$= \exp\left\{\frac{1}{2}\left(u^\top A\Omega A^\top u + v^\top B\Omega B^\top v + u^\top \underbrace{A\Omega B^\top}_{=0} v + v^\top \underbrace{B\Omega A^\top}_{=0} u\right)\right\}$$

Some more linear algebra

For the converse, assume that $\boldsymbol{A}\boldsymbol{Y}$ and $\boldsymbol{B}\boldsymbol{Y}$ are independent. Then, $\forall \boldsymbol{u}, \boldsymbol{v}$,

$$M_{\boldsymbol{W}}(\boldsymbol{\theta}) = M_{\boldsymbol{A}\boldsymbol{Y}}(\boldsymbol{u}) M_{\boldsymbol{B}\boldsymbol{Y}}(\boldsymbol{v}), \quad \forall \boldsymbol{u}, \boldsymbol{v},$$

$$\implies \exp\left\{\frac{1}{2}\left(\boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}^\top \boldsymbol{u} + \boldsymbol{v}^\top \boldsymbol{B}\boldsymbol{\Omega}\boldsymbol{B}^\top \boldsymbol{v} + \boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{B}^\top \boldsymbol{v} + \boldsymbol{v}^\top \boldsymbol{B}\boldsymbol{\Omega}\boldsymbol{A}^\top \boldsymbol{u}\right)\right\}$$

$$= \exp\left\{\frac{1}{2}\boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{\Omega}\boldsymbol{A}^\top \boldsymbol{u}\right\} \exp\left\{\frac{1}{2}\boldsymbol{v}^\top \boldsymbol{B}\boldsymbol{\Omega}\boldsymbol{B}^\top \boldsymbol{v}\right\}$$

$$\implies \exp\left\{\frac{1}{2} \times 2\boldsymbol{v}^\top \boldsymbol{A}\Omega\boldsymbol{B}^\top \boldsymbol{u}\right\} = 1$$

$$\implies \boldsymbol{v}^\top \boldsymbol{A}\Omega\boldsymbol{B}^\top \boldsymbol{u} = 0, \qquad \forall\ \boldsymbol{u}, \boldsymbol{v},$$

$$\implies \boldsymbol{A}\Omega\boldsymbol{B}^\top = 0.$$

□