Matrix Approach to SLR

Zhenisbek Assylbekov

Department of Mathematics

Regression Analysis

Random Vectors

Let $\mathbf{Y}$ be a random vector in $\mathbb{R}^n$, i.e.

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

Random Vectors

Let $\mathbf{Y}$ be a random vector in $\mathbb{R}^n$, i.e.

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

The **pdf** of $\mathbf{Y}$ is the joint pdf of $Y_1, \ldots, Y_n$:

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)$$

Random Vectors

Let $\mathbf{Y}$ be a random vector in $\mathbb{R}^n$, i.e.

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

The **pdf** of $\mathbf{Y}$ is the joint pdf of $Y_1, \ldots, Y_n$:

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)$$

It is a function s.t. for any (measurable) $B \subset \mathbb{R}^n$:

$$\Pr(\mathbf{Y} \in B) = \int_{\mathbf{y} \in B} f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) dy_1 \cdots dy_n$$

Random Vectors

Let $\mathbf{Y}$ be a random vector in $\mathbb{R}^n$, i.e.

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

The **pdf** of $\mathbf{Y}$ is the joint pdf of $Y_1, \ldots, Y_n$:

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)$$

It is a function s.t. for any (measurable) $B \subset \mathbb{R}^n$:

$$\begin{aligned} \Pr(\mathbf{Y} \in B) &= \int_{\mathbf{y} \in B} f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) dy_1 \cdots dy_n \\ &= \int_{\mathbf{y} \in B} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} \end{aligned}$$

Expectation of a Random Vector

The **expectation** of a random vector $\mathbf{Y}$ is

$$\boldsymbol{\mu} = \mathrm{E}[\mathbf{Y}] = \begin{bmatrix} \mathrm{E}[Y_1] \\ \vdots \\ \mathrm{E}[Y_n] \end{bmatrix}$$

Random Vectors

The **covariance matrix** of a random vector $\mathbf{Y}$ is an $n \times n$ matrix defined as

$$\mathrm{Var}[\mathbf{Y}] = \mathrm{E}[(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^\top]$$

Random Vectors

The **covariance matrix** of a random vector $\mathbf{Y}$ is an $n \times n$ matrix defined as

$$
\begin{aligned}
\operatorname{Var}[\mathbf{Y}] &= \operatorname{E}[(\mathbf{Y} - \operatorname{E}[\mathbf{Y}])(\mathbf{Y} - \operatorname{E}[\mathbf{Y}])^\top] \\
&= \begin{bmatrix}
\operatorname{Var}[Y_1] & \operatorname{Cov}[Y_1, Y_2] & \cdots & \operatorname{Cov}[Y_1, Y_n] \\
\operatorname{Cov}[Y_2, Y_1] & \operatorname{Var}[Y_2] & \cdots & \operatorname{Cov}[Y_2, Y_n] \\
\vdots & \vdots & \ddots & \vdots \\
\operatorname{Cov}[Y_n, Y_1] & \operatorname{Cov}[Y_n, Y_2] & \cdots & \operatorname{Var}[Y_n]
\end{bmatrix}
\end{aligned}
$$

Random Vectors

The **covariance matrix** of a random vector $\mathbf{Y}$ is an $n \times n$ matrix defined as

$$
\begin{aligned}
\mathrm{Var}[\mathbf{Y}] &= \mathrm{E}[(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^\top] \\
&= \begin{bmatrix} \mathrm{Var}[Y_1] & \mathrm{Cov}[Y_1, Y_2] & \cdots & \mathrm{Cov}[Y_1, Y_n] \\ \mathrm{Cov}[Y_2, Y_1] & \mathrm{Var}[Y_2] & \cdots & \mathrm{Cov}[Y_2, Y_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[Y_n, Y_1] & \mathrm{Cov}[Y_n, Y_2] & \cdots & \mathrm{Var}[Y_n] \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_2^2 & \cdots & \sigma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n,1} & \sigma_{n,2} & \cdots & \sigma_n^2 \end{bmatrix}
\end{aligned}
$$

Random Vectors

The **covariance matrix** of a random vector $\mathbf{Y}$ is an $n \times n$ matrix defined as

$$
\begin{aligned}
\mathrm{Var}[\mathbf{Y}] &= \mathrm{E}[(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^\top] \\
&= \begin{bmatrix}
\mathrm{Var}[Y_1] & \mathrm{Cov}[Y_1, Y_2] & \cdots & \mathrm{Cov}[Y_1, Y_n] \\
\mathrm{Cov}[Y_2, Y_1] & \mathrm{Var}[Y_2] & \cdots & \mathrm{Cov}[Y_2, Y_n] \\
\vdots & \vdots & \ddots & \vdots \\
\mathrm{Cov}[Y_n, Y_1] & \mathrm{Cov}[Y_n, Y_2] & \cdots & \mathrm{Var}[Y_n]
\end{bmatrix} \\
&= \begin{bmatrix}
\sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,n} \\
\sigma_{2,1} & \sigma_2^2 & \cdots & \sigma_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{n,1} & \sigma_{n,2} & \cdots & \sigma_n^2
\end{bmatrix} \\
&= \Sigma \qquad \text{(symmetric)}
\end{aligned}
$$

SLR

In the regression model, the only random term is $\boldsymbol{\epsilon}$ with

$$\mathrm{E}[\boldsymbol{\epsilon}] =$$

SLR

In the regression model, the only random term is $\boldsymbol{\epsilon}$ with

$$\mathrm{E}[\boldsymbol{\epsilon}] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad \text{and} \quad \mathrm{Var}[\boldsymbol{\epsilon}]$$

SLR

In the regression model, the only random term is $\boldsymbol{\epsilon}$ with

$$\mathrm{E}[\boldsymbol{\epsilon}] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad \text{and} \quad \mathrm{Var}[\boldsymbol{\epsilon}] = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n$$

SLR

In the regression model, the only random term is $\boldsymbol{\epsilon}$ with

$$\mathrm{E}[\boldsymbol{\epsilon}] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad \text{and} \quad \mathrm{Var}[\boldsymbol{\epsilon}] = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n$$

Hence for the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

SLR

In the regression model, the only random term is $\boldsymbol{\epsilon}$ with

$$\mathrm{E}[\boldsymbol{\epsilon}] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad \text{and} \quad \mathrm{Var}[\boldsymbol{\epsilon}] = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n$$

Hence for the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- $\mathrm{E}[\mathbf{Y}]$

SLR

In the regression model, the only random term is $\boldsymbol{\epsilon}$ with

$$\mathrm{E}[\boldsymbol{\epsilon}] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad \text{and} \quad \mathrm{Var}[\boldsymbol{\epsilon}] = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n$$

Hence for the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- $\mathrm{E}[\mathbf{Y}] = \mathrm{E}[\mathbf{X}\boldsymbol{\beta}] + \mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{X}\boldsymbol{\beta}$

SLR

In the regression model, the only random term is $\boldsymbol{\epsilon}$ with

$$\mathrm{E}[\boldsymbol{\epsilon}] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad \text{and} \quad \mathrm{Var}[\boldsymbol{\epsilon}] = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n$$

Hence for the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- $\mathrm{E}[\mathbf{Y}] = \mathrm{E}[\mathbf{X}\boldsymbol{\beta}] + \mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{X}\boldsymbol{\beta}$
- $\mathrm{Var}[\mathbf{Y}]$

SLR

In the regression model, the only random term is $\boldsymbol{\epsilon}$ with

$$\mathrm{E}[\boldsymbol{\epsilon}] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad \text{and} \quad \mathrm{Var}[\boldsymbol{\epsilon}] = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n$$

Hence for the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- $\mathrm{E}[\mathbf{Y}] = \mathrm{E}[\mathbf{X}\boldsymbol{\beta}] + \mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{X}\boldsymbol{\beta}$
- $\mathrm{Var}[\mathbf{Y}] = \mathrm{Var}[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}] = \mathrm{Var}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}_n$

Linear Transform of a Random Vector

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{Y}$ be a random n-vector, then

$$\mathbf{AY} =$$

Linear Transform of a Random Vector

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{Y}$ be a random n-vector, then

$$\mathbf{AY} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

Linear Transform of a Random Vector

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{Y}$ be a random n-vector, then

$$\mathbf{AY} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} Y_1 + a_{12} Y_2 + \ldots + a_{1n} Y_n \\ a_{21} Y_1 + a_{22} Y_2 + \ldots + a_{2n} Y_n \\ \vdots \\ a_{m1} Y_1 + a_{m2} Y_2 + \ldots + a_{mn} Y_n \end{bmatrix}$$

Mean of a Linear Transform

Theorem. Let $\mathbf{Y}$ be a random n-vector, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\mathrm{E}[\mathbf{AY}] = \mathbf{A}\mathrm{E}[\mathbf{Y}]$.

Proof.

$$\mathrm{E}[\mathbf{AY}] =$$

Mean of a Linear Transform

Theorem. Let $\mathbf{Y}$ be a random n-vector, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\mathrm{E}[\mathbf{AY}] = \mathbf{A}\mathrm{E}[\mathbf{Y}]$.

Proof.

$$\mathrm{E}[\mathbf{AY}] = \begin{bmatrix} \mathrm{E}[a_{11}Y_1 + a_{12}Y_2 + \ldots + a_{1n}Y_n] \\ \vdots \\ \mathrm{E}[a_{m1}Y_1 + a_{m2}Y_2 + \ldots + a_{mn}Y_n] \end{bmatrix}$$

Mean of a Linear Transform

Theorem. Let $\mathbf{Y}$ be a random n-vector, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\mathrm{E}[\mathbf{AY}] = \mathbf{A}\mathrm{E}[\mathbf{Y}]$.

Proof.

$$
\begin{aligned}
\mathrm{E}[\mathbf{AY}] &= \begin{bmatrix} \mathrm{E}[a_{11}Y_1 + a_{12}Y_2 + \ldots + a_{1n}Y_n] \\ \vdots \\ \mathrm{E}[a_{m1}Y_1 + a_{m2}Y_2 + \ldots + a_{mn}Y_n] \end{bmatrix} \\
&= \begin{bmatrix} a_{11}\mathrm{E}[Y_1] + a_{12}\mathrm{E}[Y_2] + \ldots + a_{1n}\mathrm{E}[Y_n] \\ \vdots \\ a_{m1}\mathrm{E}[Y_1] + a_{m2}\mathrm{E}[Y_2] + \ldots + a_{mn}\mathrm{E}[Y_n] \end{bmatrix}
\end{aligned}
$$

Mean of a Linear Transform

Theorem. Let $\mathbf{Y}$ be a random n-vector, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\mathrm{E}[\mathbf{AY}] = \mathbf{A}\mathrm{E}[\mathbf{Y}]$.

Proof.

$$
\begin{aligned}
\mathrm{E}[\mathbf{AY}] &= \begin{bmatrix} \mathrm{E}[a_{11}Y_1 + a_{12}Y_2 + \ldots + a_{1n}Y_n] \\ \vdots \\ \mathrm{E}[a_{m1}Y_1 + a_{m2}Y_2 + \ldots + a_{mn}Y_n] \end{bmatrix} \\
&= \begin{bmatrix} a_{11}\mathrm{E}[Y_1] + a_{12}\mathrm{E}[Y_2] + \ldots + a_{1n}\mathrm{E}[Y_n] \\ \vdots \\ a_{m1}\mathrm{E}[Y_1] + a_{m2}\mathrm{E}[Y_2] + \ldots + a_{mn}\mathrm{E}[Y_n] \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} \mathrm{E}[Y_1] \\ \mathrm{E}[Y_2] \\ \vdots \\ \mathrm{E}[Y_n] \end{bmatrix}
\end{aligned}
$$

Mean of a Linear Transform

Theorem. Let $\mathbf{Y}$ be a random n-vector, and $\mathbf{A} \in \mathbb{R}^{m\times n}$. Then $\mathrm{E}[\mathbf{AY}] = \mathbf{A}\mathrm{E}[\mathbf{Y}]$.

Proof.

$$
\begin{aligned}
\mathrm{E}[\mathbf{AY}] &= \begin{bmatrix} \mathrm{E}[a_{11}Y_1 + a_{12}Y_2 + \ldots + a_{1n}Y_n] \\ \vdots \\ \mathrm{E}[a_{m1}Y_1 + a_{m2}Y_2 + \ldots + a_{mn}Y_n] \end{bmatrix} \\
&= \begin{bmatrix} a_{11}\mathrm{E}[Y_1] + a_{12}\mathrm{E}[Y_2] + \ldots + a_{1n}\mathrm{E}[Y_n] \\ \vdots \\ a_{m1}\mathrm{E}[Y_1] + a_{m2}\mathrm{E}[Y_2] + \ldots + a_{mn}\mathrm{E}[Y_n] \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} \mathrm{E}[Y_1] \\ \mathrm{E}[Y_2] \\ \vdots \\ \mathrm{E}[Y_n] \end{bmatrix} \\
&= \mathbf{A}\mathrm{E}[\mathbf{Y}]
\end{aligned}
$$

Covariance Matrix of a Linear Transform

Theorem. Let $\mathbf{Y}$ be a random n-vector, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\mathrm{Var}[\mathbf{AY}] = \mathbf{A}\mathrm{Var}[\mathbf{Y}]\mathbf{A}^\top$.

Proof.

$$\mathrm{Var}[\mathbf{AY}] =$$

$\square$

Covariance Matrix of a Linear Transform

Theorem. Let $\mathbf{Y}$ be a random n-vector, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\mathrm{Var}[\mathbf{AY}] = \mathbf{A}\mathrm{Var}[\mathbf{Y}]\mathbf{A}^\top$.

Proof.

$$\mathrm{Var}[\mathbf{AY}] = \mathrm{E}[(\mathbf{AY} - \mathrm{E}[\mathbf{AY}])(\mathbf{AY} - \mathrm{E}[\mathbf{AY}])^\top]$$

$\square$

Covariance Matrix of a Linear Transform

Theorem. Let $\mathbf{Y}$ be a random n-vector, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\mathrm{Var}[\mathbf{AY}] = \mathbf{A}\mathrm{Var}[\mathbf{Y}]\mathbf{A}^\top$.

Proof.

$$\begin{aligned}\mathrm{Var}[\mathbf{AY}] &= \mathrm{E}[(\mathbf{AY} - \mathrm{E}[\mathbf{AY}])(\mathbf{AY} - \mathrm{E}[\mathbf{AY}])^\top] \\ &= \mathrm{E}[(\mathbf{AY} - \mathbf{A}\mathrm{E}[\mathbf{Y}])(\mathbf{AY} - \mathbf{A}\mathrm{E}[\mathbf{Y}])^\top]\end{aligned}$$

$\square$

Covariance Matrix of a Linear Transform

Theorem. Let $\mathbf{Y}$ be a random n-vector, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\mathrm{Var}[\mathbf{AY}] = \mathbf{A}\mathrm{Var}[\mathbf{Y}]\mathbf{A}^\top$.

Proof.

$$\begin{aligned}
\mathrm{Var}[\mathbf{AY}] &= \mathrm{E}[(\mathbf{AY} - \mathrm{E}[\mathbf{AY}])(\mathbf{AY} - \mathrm{E}[\mathbf{AY}])^\top] \\
&= \mathrm{E}[(\mathbf{AY} - \mathbf{A}\mathrm{E}[\mathbf{Y}])(\mathbf{AY} - \mathbf{A}\mathrm{E}[\mathbf{Y}])^\top] \\
&= \mathrm{E}[\mathbf{A}(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^\top\mathbf{A}^\top]
\end{aligned}$$

$\square$

Covariance Matrix of a Linear Transform

Theorem. Let $\mathbf{Y}$ be a random n-vector, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\mathrm{Var}[\mathbf{AY}] = \mathbf{A}\mathrm{Var}[\mathbf{Y}]\mathbf{A}^\top$.

Proof.

$$\begin{aligned}
\mathrm{Var}[\mathbf{AY}] &= \mathrm{E}[(\mathbf{AY} - \mathrm{E}[\mathbf{AY}])(\mathbf{AY} - \mathrm{E}[\mathbf{AY}])^\top] \\
&= \mathrm{E}[(\mathbf{AY} - \mathbf{A}\mathrm{E}[\mathbf{Y}])(\mathbf{AY} - \mathbf{A}\mathrm{E}[\mathbf{Y}])^\top] \\
&= \mathrm{E}[\mathbf{A}(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^\top \mathbf{A}^\top] \\
&= \mathbf{A}\mathrm{E}[(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^\top]\mathbf{A}^\top
\end{aligned}$$

$\square$

Covariance Matrix of a Linear Transform

Theorem. Let $\mathbf{Y}$ be a random n-vector, and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\mathrm{Var}[\mathbf{AY}] = \mathbf{A}\mathrm{Var}[\mathbf{Y}]\mathbf{A}^\top$.

Proof.

$$\begin{aligned}
\mathrm{Var}[\mathbf{AY}] &= \mathrm{E}[(\mathbf{AY} - \mathrm{E}[\mathbf{AY}])(\mathbf{AY} - \mathrm{E}[\mathbf{AY}])^\top] \\
&= \mathrm{E}[(\mathbf{AY} - \mathbf{A}\mathrm{E}[\mathbf{Y}])(\mathbf{AY} - \mathbf{A}\mathrm{E}[\mathbf{Y}])^\top] \\
&= \mathrm{E}[\mathbf{A}(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^\top \mathbf{A}^\top] \\
&= \mathbf{A}\mathrm{E}[(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])(\mathbf{Y} - \mathrm{E}[\mathbf{Y}])^\top]\mathbf{A}^\top \\
&= \mathbf{A}\mathrm{Var}[\mathbf{Y}]\mathbf{A}^\top
\end{aligned}$$

$\square$

Covariance Matrix is Positive Semi-definite

Theorem. Let $\mathbf{Y}$ be a random vector in $\mathbb{R}^n$. Then $\mathrm{Var}[\mathbf{Y}] \succeq \mathbf{0}$.

Covariance Matrix is Positive Semi-definite

Theorem. Let $\mathbf{Y}$ be a random vector in $\mathbb{R}^n$. Then $\mathrm{Var}[\mathbf{Y}] \succeq \mathbf{0}$.

Proof.
$\forall \mathbf{a} \in \mathbb{R}^n$:

$$\mathbf{a}^\top \mathrm{Var}[\mathbf{Y}]\mathbf{a}$$

Covariance Matrix is Positive Semi-definite

Theorem. Let $\mathbf{Y}$ be a random vector in $\mathbb{R}^n$. Then $\mathrm{Var}[\mathbf{Y}] \succeq \mathbf{0}$.

Proof.
$\forall \mathbf{a} \in \mathbb{R}^n$:

$$\mathbf{a}^\top \mathrm{Var}[\mathbf{Y}]\mathbf{a} = \mathrm{Var}[\mathbf{a}^\top \mathbf{Y}]$$

Covariance Matrix is Positive Semi-definite

Theorem. Let $\mathbf{Y}$ be a random vector in $\mathbb{R}^n$. Then $\mathrm{Var}[\mathbf{Y}] \succeq \mathbf{0}$.

Proof.
$\forall \mathbf{a} \in \mathbb{R}^n$:

$$\mathbf{a}^\top \mathrm{Var}[\mathbf{Y}]\mathbf{a} = \mathrm{Var}[\mathbf{a}^\top \mathbf{Y}] \geq 0,$$

since $\mathbf{a}^\top \mathbf{Y} = \sum_i a_i Y_i$ is a random variable with values in $\mathbb{R}^1$. $\square$

Pdf of a Linear Transform, $n = 1$

Theorem. Let X be a continuous r.v. with the pdf $f_X(x)$, and $a \neq 0$. Then the pdf of $Y = aX$ is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y}{a}\right)$$

Proof.
For $a > 0$:

$$f_Y(y)$$

$\square$

Pdf of a Linear Transform, $n = 1$

Theorem. Let X be a continuous r.v. with the pdf $f_X(x)$, and $a \neq 0$. Then the pdf of $Y = aX$ is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y}{a}\right)$$

Proof.
For $a > 0$:

$$f_Y(y) = \frac{d}{dy}[F_Y(y)]$$

□

Pdf of a Linear Transform, $n = 1$

Theorem. Let X be a continuous r.v. with the pdf $f_X(x)$, and $a \neq 0$. Then the pdf of $Y = aX$ is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y}{a}\right)$$

Proof.
For $a > 0$:

$$f_Y(y) = \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}\left[\Pr(Y \leq y)\right]$$

$\square$

Pdf of a Linear Transform, $n = 1$

Theorem. Let X be a continuous r.v. with the pdf $f_X(x)$, and $a \neq 0$. Then the pdf of $Y = aX$ is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y}{a}\right)$$

Proof.
For $a > 0$:

$$f_Y(y) = \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}\left[\Pr(Y \leq y)\right] = \frac{d}{dy}\left[\Pr(aX \leq y\right]$$

$\square$

Pdf of a Linear Transform, $n = 1$

Theorem. Let X be a continuous r.v. with the pdf $f_X(x)$, and $a \neq 0$. Then the pdf of $Y = aX$ is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y}{a}\right)$$

Proof.
For $a > 0$:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}\left[\Pr(Y \leq y)\right] = \frac{d}{dy}\left[\Pr(aX \leq y\right] \\ &= \frac{d}{dy}\left[\Pr\left(X \leq \frac{y}{a}\right)\right] \end{aligned}$$

$\square$

Pdf of a Linear Transform, $n = 1$

Theorem. Let X be a continuous r.v. with the pdf $f_X(x)$, and $a \neq 0$. Then the pdf of $Y = aX$ is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y}{a}\right)$$

Proof.
For $a > 0$:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}\left[\Pr(Y \leq y)\right] = \frac{d}{dy}\left[\Pr(aX \leq y\right] \\ &= \frac{d}{dy}\left[\Pr\left(X \leq \frac{y}{a}\right)\right] = \frac{d}{dy}\left[F_X\left(\frac{y}{a}\right)\right] \end{aligned}$$

□

Pdf of a Linear Transform, $n = 1$

Theorem. Let X be a continuous r.v. with the pdf $f_X(x)$, and $a \neq 0$. Then the pdf of $Y = aX$ is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y}{a}\right)$$

Proof.
For $a > 0$:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}\left[\Pr(Y \leq y)\right] = \frac{d}{dy}\left[\Pr(aX \leq y\right] \\ &= \frac{d}{dy}\left[\Pr\left(X \leq \frac{y}{a}\right)\right] = \frac{d}{dy}\left[F_X\left(\frac{y}{a}\right)\right] = \frac{1}{a} f_X\left(\frac{y}{a}\right) \end{aligned}$$

$\square$

Pdf of a Linear Transform, $n = 1$

Theorem. Let X be a continuous r.v. with the pdf $f_X(x)$, and $a \neq 0$. Then the pdf of $Y = aX$ is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y}{a}\right)$$

Proof.
For $a > 0$:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}\left[\Pr(Y \leq y)\right] = \frac{d}{dy}\left[\Pr(aX \leq y\right] \\ &= \frac{d}{dy}\left[\Pr\left(X \leq \frac{y}{a}\right)\right] = \frac{d}{dy}\left[F_X\left(\frac{y}{a}\right)\right] = \frac{1}{a} f_X\left(\frac{y}{a}\right) \end{aligned}$$

For $a < 0$ the proof is analogous and is left as *exercise*. $\square$

Pdf of a Linear Transform, General Case

Theorem. Let $\mathbf{X}$ be a random vector with the pdf $f_{\mathbf{X}}(\mathbf{x})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\det \mathbf{A} \neq 0$. Then the pdf of $\mathbf{Y} = \mathbf{AX}$ is given by

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

Pdf of a Linear Transform, General Case

Theorem. Let $\mathbf{X}$ be a random vector with the pdf $f_{\mathbf{X}}(\mathbf{x})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\det \mathbf{A} \neq 0$. Then the pdf of $\mathbf{Y} = \mathbf{AX}$ is given by

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

Proof.
For any $B \subset \mathbb{R}^n$,

$\square$

Pdf of a Linear Transform, General Case

Theorem. Let $\mathbf{X}$ be a random vector with the pdf $f_{\mathbf{X}}(\mathbf{x})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\det \mathbf{A} \neq 0$. Then the pdf of $\mathbf{Y} = \mathbf{A}\mathbf{X}$ is given by

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

Proof.
For any $B \subset \mathbb{R}^n$,

$$\int_{\mathbf{y} \in B} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} = \Pr(\mathbf{Y} \in B)$$

$\square$

Pdf of a Linear Transform, General Case

Theorem. Let $\mathbf{X}$ be a random vector with the pdf $f_{\mathbf{X}}(\mathbf{x})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\det \mathbf{A} \neq 0$. Then the pdf of $\mathbf{Y} = \mathbf{AX}$ is given by

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

Proof.
For any $B \subset \mathbb{R}^n$,

$$\int_{\mathbf{y} \in B} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} = \Pr(\mathbf{Y} \in B) = \Pr(\mathbf{AX} \in B)$$

$\square$

Pdf of a Linear Transform, General Case

Theorem. Let $\mathbf{X}$ be a random vector with the pdf $f_{\mathbf{X}}(\mathbf{x})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\det \mathbf{A} \neq 0$. Then the pdf of $\mathbf{Y} = \mathbf{AX}$ is given by

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

Proof.
For any $B \subset \mathbb{R}^n$,

$$\int_{\mathbf{y} \in B} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} = \Pr(\mathbf{Y} \in B) = \Pr(\mathbf{AX} \in B) = \int_{\mathbf{Ax} \in B} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

□

Pdf of a Linear Transform, General Case

Theorem. Let $\mathbf{X}$ be a random vector with the pdf $f_{\mathbf{X}}(\mathbf{x})$, and $\mathbf{A} \in \mathbb{R}^{n\times n}$ with $\det \mathbf{A} \neq 0$. Then the pdf of $\mathbf{Y} = \mathbf{AX}$ is given by

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

Proof.
For any $B \subset \mathbb{R}^n$,

$$\begin{aligned}\int_{\mathbf{y}\in B} f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y} = \Pr(\mathbf{Y} \in B) = \Pr(\mathbf{AX} \in B) &= \int_{\mathbf{Ax}\in B} f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \\ &= \int_{\mathbf{y}\in B} f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{y})|\det \mathbf{A}^{-1}|d\mathbf{y}\end{aligned}$$

$\square$

Pdf of a Linear Transform, General Case

Theorem. Let $\mathbf{X}$ be a random vector with the pdf $f_{\mathbf{X}}(\mathbf{x})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\det \mathbf{A} \neq 0$. Then the pdf of $\mathbf{Y} = \mathbf{AX}$ is given by

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

Proof.
For any $B \subset \mathbb{R}^n$,

$$\int_{\mathbf{y} \in B} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} = \Pr(\mathbf{Y} \in B) = \Pr(\mathbf{AX} \in B) = \int_{\mathbf{Ax} \in B} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$
$$= \int_{\mathbf{y} \in B} f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{y}) |\det \mathbf{A}^{-1}| d\mathbf{y}$$

Noting that $\det \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}$ concludes the proof. $\square$

Pdf of an Affine Transform

Theorem. Let X be a continuous r.v. with the pdf $f_X(x)$, and $a \neq 0$. Then the pdf of $Y = aX + b$ is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Pdf of an Affine Transform

Theorem. Let X be a continuous r.v. with the pdf $f_X(x)$, and $a \neq 0$. Then the pdf of $Y = aX + b$ is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Theorem. Let $\mathbf{X}$ be a random vector with the pdf $f_{\mathbf{X}}(\mathbf{x})$, $\mathbf{A} \in \mathbb{R}^{n\times n}$ with $\det \mathbf{A} \neq 0$, and $\mathbf{b} \in \mathbb{R}^b$. Then the pdf of $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det \mathbf{A}|} f_{\mathbf{X}}(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}))$$

Pdf of an Affine Transform

Theorem. Let X be a continuous r.v. with the pdf $f_X(x)$, and $a \neq 0$. Then the pdf of $Y = aX + b$ is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Theorem. Let $\mathbf{X}$ be a random vector with the pdf $f_{\mathbf{X}}(\mathbf{x})$, $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\det \mathbf{A} \neq 0$, and $\mathbf{b} \in \mathbb{R}^b$. Then the pdf of $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det \mathbf{A}|} f_{\mathbf{X}}(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}))$$

Proofs are left as exercises.

Multivariate Normal Distribution

Definition. A random vector $\mathbf{Y} = \begin{bmatrix} Y_1 & \dots & Y_n \end{bmatrix}^\top$ is **multivariate normal (Gaussian)** if it can be represented as

$$\mathbf{Y} = \mathbf{AZ} + \boldsymbol{\mu},$$

where

- $\mathbf{Z} = \begin{bmatrix} Z_1 & \dots & Z_k \end{bmatrix}^\top$ with $Z_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$,
- $\mathbf{A} \in \mathbb{R}^{n \times k}$ s.t. $\Sigma = \mathbf{AA}^\top$

Multivariate Normal Distribution

Definition. A random vector $\mathbf{Y} = \begin{bmatrix} Y_1 & \dots & Y_n \end{bmatrix}^\top$ is **multivariate normal (Gaussian)** if it can be represented as

$$\mathbf{Y} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu},$$

where

- $\mathbf{Z} = \begin{bmatrix} Z_1 & \dots & Z_k \end{bmatrix}^\top$ with $Z_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$,
- $\mathbf{A} \in \mathbb{R}^{n \times k}$ s.t. $\Sigma = \mathbf{A}\mathbf{A}^\top$

Notation. $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$

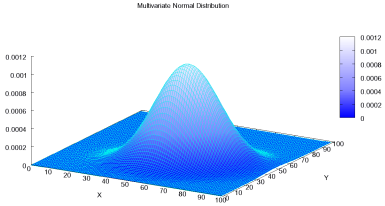
Multivariate Normal Distribution

Definition. A random vector $\mathbf{Y} = \begin{bmatrix} Y_1 & \dots & Y_n \end{bmatrix}^\top$ is **multivariate normal (Gaussian)** if it can be represented as

$$\mathbf{Y} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu},$$

where

- $\mathbf{Z} = \begin{bmatrix} Z_1 & \dots & Z_k \end{bmatrix}^\top$ with $Z_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$,
- $\mathbf{A} \in \mathbb{R}^{n \times k}$ s.t. $\Sigma = \mathbf{A}\mathbf{A}^\top$



Notation. $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$

Multivariate Normal Distribution

Theorem. $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\mathrm{rank}\boldsymbol{\Sigma} = n \quad \Rightarrow \quad$ The pdf of $\mathbf{Y}$ has the form

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\sqrt{\det \Sigma}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}.$$

Proof.
$\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$

Multivariate Normal Distribution

Theorem. $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\mathrm{rank}\boldsymbol{\Sigma} = n \quad \Rightarrow \quad$ The pdf of $\mathbf{Y}$ has the form

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\sqrt{\det \Sigma}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}.$$

Proof.
$\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma) \quad \Rightarrow \quad \exists \mathbf{A} \in \mathbb{R}^{n \times k}$, $\mathbf{Z} \in \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$:

Multivariate Normal Distribution

Theorem. $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\mathrm{rank}\boldsymbol{\Sigma} = n \quad \Rightarrow \quad$ The pdf of $\mathbf{Y}$ has the form

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\sqrt{\det \Sigma}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}.$$

Proof.

$\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma) \quad \Rightarrow \quad \exists \mathbf{A} \in \mathbb{R}^{n \times k}$, $\mathbf{Z} \in \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$:

$$\mathbf{Y} = \mathbf{AZ} + \boldsymbol{\mu} \quad \text{with } \mathbf{AA}^\top = \Sigma$$

Since $n = \mathrm{rank}\Sigma$

Multivariate Normal Distribution

Theorem. $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\mathrm{rank}\boldsymbol{\Sigma} = n \quad \Rightarrow \quad$ The pdf of $\mathbf{Y}$ has the form

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\sqrt{\det \Sigma}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})}.$$

Proof.

$\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma) \quad \Rightarrow \quad \exists \mathbf{A} \in \mathbb{R}^{n \times k}$, $\mathbf{Z} \in \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$:

$$\mathbf{Y} = \mathbf{AZ} + \boldsymbol{\mu} \quad \text{with } \mathbf{AA}^\top = \Sigma$$

Since $n = \mathrm{rank}\Sigma = \mathrm{rank}\mathbf{AA}^\top$

Multivariate Normal Distribution

Theorem. $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\mathrm{rank}\Sigma = n \quad \Rightarrow \quad$ The pdf of $\mathbf{Y}$ has the form

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\sqrt{\det \Sigma}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})}.$$

Proof.

$\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma) \quad \Rightarrow \quad \exists \mathbf{A} \in \mathbb{R}^{n\times k}$, $\mathbf{Z} \in \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$:

$$\mathbf{Y} = \mathbf{AZ} + \boldsymbol{\mu} \quad \text{with } \mathbf{AA}^\top = \Sigma$$

Since $n = \mathrm{rank}\Sigma = \mathrm{rank}\mathbf{AA}^\top = k$, we have $k = n$.

Also, $\det \Sigma = \det \mathbf{AA}^\top = \det \mathbf{A} \cdot \det \mathbf{A}^\top = (\det \mathbf{A})^2$. Thus,

$$\det \mathbf{A} = \sqrt{\det \Sigma}.$$

Multivariate Normal Distribution

Poof (cont'd).

For the $\mathbf{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}$ with $Z_i \overset{\text{iid}}{\sim} \in \mathcal{N}(0,1)$,

$$f_{\mathbf{Z}}(\mathbf{z})$$

Multivariate Normal Distribution

Poof (cont'd).

For the $\mathbf{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}$ with $Z_i \overset{\text{iid}}{\sim} \in \mathcal{N}(0, 1)$,

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n)$$

Multivariate Normal Distribution

Poof (cont'd).

For the $\mathbf{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}$ with $Z_i \overset{\text{iid}}{\sim} \in \mathcal{N}(0, 1)$,

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n) = \prod_{i=1}^{n} f_{Z_i}(z_i)$$

Multivariate Normal Distribution

Poof (cont'd).

For the $\mathbf{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}$ with $Z_i \overset{\text{iid}}{\sim} \in \mathcal{N}(0, 1)$,

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n) = \prod_{i=1}^{n} f_{Z_i}(z_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}}$$

Multivariate Normal Distribution

Poof (cont'd).

For the $\mathbf{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}$ with $Z_i \overset{\text{iid}}{\sim} \in \mathcal{N}(0,1)$,

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) = \prod_{i=1}^{n} f_{Z_i}(z_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}}$$

$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(z_1^2+\ldots+z_n^2)}$$

Multivariate Normal Distribution

Poof (cont'd).

For the $\mathbf{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}$ with $Z_i \overset{\text{iid}}{\sim} \in \mathcal{N}(0,1)$,

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) = \prod_{i=1}^{n} f_{Z_i}(z_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}}$$

$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(z_1^2+\ldots+z_n^2)} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{z}^\top \mathbf{z}}$$

Multivariate Normal Distribution

Poof (cont'd).

For the $\mathbf{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}$ with $Z_i \overset{\text{iid}}{\sim} \in \mathcal{N}(0,1)$,

$$
\begin{aligned}
f_{\mathbf{Z}}(\mathbf{z}) = f_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) &= \prod_{i=1}^{n} f_{Z_i}(z_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} \\
&= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(z_1^2+\ldots+z_n^2)} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{z}^\top \mathbf{z}}
\end{aligned}
$$

Since $\mathbf{Y} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu}$, and $\Sigma = \mathbf{A}\mathbf{A}^\top$, we have

Multivariate Normal Distribution

Poof (cont'd).

For the $\mathbf{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}$ with $Z_i \overset{\text{iid}}{\sim} \in \mathcal{N}(0, 1)$,

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n) = \prod_{i=1}^{n} f_{Z_i}(z_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}}$$

$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(z_1^2+\ldots+z_n^2)} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{z}^\top \mathbf{z}}$$

Since $\mathbf{Y} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu}$, and $\Sigma = \mathbf{A}\mathbf{A}^\top$, we have

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\det \mathbf{A}} f_{\mathbf{Z}}(\mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu}))$$

Multivariate Normal Distribution

Poof (cont'd).

For the $\mathbf{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}$ with $Z_i \overset{\text{iid}}{\sim} \in \mathcal{N}(0,1)$,

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) = \prod_{i=1}^{n} f_{Z_i}(z_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}}$$
$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(z_1^2+\ldots+z_n^2)} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{z}^\top \mathbf{z}}$$

Since $\mathbf{Y} = \mathbf{AZ} + \boldsymbol{\mu}$, and $\Sigma = \mathbf{AA}^\top$, we have

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\det \mathbf{A}} f_{\mathbf{Z}}(\mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})) = \frac{1}{\sqrt{\det \Sigma}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top (\mathbf{A}^{-1})^\top \mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})}$$

Multivariate Normal Distribution

Poof (cont'd).

For the $\mathbf{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}$ with $Z_i \overset{\text{iid}}{\sim} \in \mathcal{N}(0,1)$,

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) = \prod_{i=1}^{n} f_{Z_i}(z_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}}$$

$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(z_1^2+\ldots+z_n^2)} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{z}^\top \mathbf{z}}$$

Since $\mathbf{Y} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu}$, and $\Sigma = \mathbf{A}\mathbf{A}^\top$, we have

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\det \mathbf{A}} f_{\mathbf{Z}}(\mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})) = \frac{1}{\sqrt{\det \Sigma}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top (\mathbf{A}^{-1})^\top \mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})}$$

$$= \frac{1}{\sqrt{\det \Sigma}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top (\mathbf{A}^\top)^{-1} \mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})}$$

Multivariate Normal Distribution

Poof (cont'd).

For the $\mathbf{Z} = \begin{bmatrix} Z_1 & \cdots & Z_n \end{bmatrix}$ with $Z_i \overset{\text{iid}}{\sim} \in \mathcal{N}(0,1)$,

$$
\begin{aligned}
f_{\mathbf{Z}}(\mathbf{z}) = f_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) &= \prod_{i=1}^{n} f_{Z_i}(z_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} \\
&= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(z_1^2+\ldots+z_n^2)} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{z}^\top \mathbf{z}}
\end{aligned}
$$

Since $\mathbf{Y} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu}$, and $\Sigma = \mathbf{A}\mathbf{A}^\top$, we have

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\det \mathbf{A}} f_{\mathbf{Z}}(\mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})) &= \frac{1}{\sqrt{\det \Sigma}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top (\mathbf{A}^{-1})^\top \mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})} \\
&= \frac{1}{\sqrt{\det \Sigma}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top (\mathbf{A}^\top)^{-1} \mathbf{A}^{-1}(\mathbf{y}-\boldsymbol{\mu})} \\
&= \frac{1}{\sqrt{\det \Sigma|}(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})}
\end{aligned}
$$

Orthogonal Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{Q}$ be an orthogonal $n \times n$ matrix ($\mathbf{Q} \in \mathcal{O}_n$). Then $\mathbf{QZ} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$.

Proof.

$\square$

Orthogonal Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{Q}$ be an orthogonal $n \times n$ matrix ($\mathbf{Q} \in \mathcal{O}_n$). Then $\mathbf{QZ} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$.

Proof.
The pdf of $\mathbf{QZ}$ is

$\square$

Orthogonal Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{Q}$ be an orthogonal $n \times n$ matrix ($\mathbf{Q} \in \mathcal{O}_n$). Then $\mathbf{QZ} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$.

Proof.
The pdf of $\mathbf{QZ}$ is

$$f_{\mathbf{QZ}}(\mathbf{w})$$

$\square$

Orthogonal Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{Q}$ be an orthogonal $n \times n$ matrix ($\mathbf{Q} \in \mathcal{O}_n$). Then $\mathbf{QZ} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$.

Proof.
The pdf of $\mathbf{QZ}$ is

$$f_{\mathbf{QZ}}(\mathbf{w}) = \frac{1}{|\det \mathbf{Q}|} f_{\mathbf{Z}}(\mathbf{Q}^{-1}\mathbf{w})$$

$\square$

Orthogonal Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{Q}$ be an orthogonal $n \times n$ matrix ($\mathbf{Q} \in \mathcal{O}_n$). Then $\mathbf{QZ} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$.

Proof.
The pdf of $\mathbf{QZ}$ is

$$f_{\mathbf{QZ}}(\mathbf{w}) = \frac{1}{|\det \mathbf{Q}|} f_{\mathbf{Z}}(\mathbf{Q}^{-1}\mathbf{w}) = \frac{1}{1} \cdot \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{Q}^{-1}\mathbf{w})^\top(\mathbf{Q}^{-1}\mathbf{w})}$$

$\square$

Orthogonal Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{Q}$ be an orthogonal $n \times n$ matrix ($\mathbf{Q} \in \mathcal{O}_n$). Then $\mathbf{QZ} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$.

Proof.
The pdf of $\mathbf{QZ}$ is

$$
\begin{aligned}
f_{\mathbf{QZ}}(\mathbf{w}) &= \frac{1}{|\det \mathbf{Q}|} f_{\mathbf{Z}}(\mathbf{Q}^{-1}\mathbf{w}) = \frac{1}{1} \cdot \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{Q}^{-1}\mathbf{w})^\top(\mathbf{Q}^{-1}\mathbf{w})} \\
&= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{Q}^\top\mathbf{w})^\top(\mathbf{Q}^\top\mathbf{w})}
\end{aligned}
$$

$\square$

Orthogonal Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{Q}$ be an orthogonal $n \times n$ matrix ($\mathbf{Q} \in \mathcal{O}_n$). Then $\mathbf{QZ} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$.

Proof.
The pdf of $\mathbf{QZ}$ is

$$f_{\mathbf{QZ}}(\mathbf{w}) = \frac{1}{|\det \mathbf{Q}|} f_{\mathbf{Z}}(\mathbf{Q}^{-1}\mathbf{w}) = \frac{1}{1} \cdot \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{Q}^{-1}\mathbf{w})^\top(\mathbf{Q}^{-1}\mathbf{w})}$$
$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{Q}^\top\mathbf{w})^\top(\mathbf{Q}^\top\mathbf{w})} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{w}^\top\mathbf{Q})(\mathbf{Q}^\top\mathbf{w})}$$

$\square$

Orthogonal Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{Q}$ be an orthogonal $n \times n$ matrix ($\mathbf{Q} \in \mathcal{O}_n$). Then $\mathbf{QZ} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$.

Proof.
The pdf of $\mathbf{QZ}$ is

$$\begin{aligned} f_{\mathbf{QZ}}(\mathbf{w}) &= \frac{1}{|\det \mathbf{Q}|} f_{\mathbf{Z}}(\mathbf{Q}^{-1}\mathbf{w}) = \frac{1}{1} \cdot \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{Q}^{-1}\mathbf{w})^\top(\mathbf{Q}^{-1}\mathbf{w})} \\ &= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{Q}^\top\mathbf{w})^\top(\mathbf{Q}^\top\mathbf{w})} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{w}^\top\mathbf{Q})(\mathbf{Q}^\top\mathbf{w})} \\ &= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{w}^\top(\mathbf{Q}\mathbf{Q}^\top)\mathbf{w}} \end{aligned}$$

□

Orthogonal Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{Q}$ be an orthogonal $n \times n$ matrix ($\mathbf{Q} \in \mathcal{O}_n$). Then $\mathbf{QZ} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$.

Proof.
The pdf of $\mathbf{QZ}$ is

$$
\begin{aligned}
f_{\mathbf{QZ}}(\mathbf{w}) &= \frac{1}{|\det \mathbf{Q}|} f_{\mathbf{Z}}(\mathbf{Q}^{-1}\mathbf{w}) = \frac{1}{1} \cdot \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{Q}^{-1}\mathbf{w})^\top (\mathbf{Q}^{-1}\mathbf{w})} \\
&= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{Q}^\top\mathbf{w})^\top (\mathbf{Q}^\top\mathbf{w})} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{w}^\top\mathbf{Q})(\mathbf{Q}^\top\mathbf{w})} \\
&= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{w}^\top(\mathbf{Q}\mathbf{Q}^\top)\mathbf{w}} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{w}^\top\mathbf{w}}
\end{aligned}
$$

$\square$

Orthogonal Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{Q}$ be an orthogonal $n \times n$ matrix ($\mathbf{Q} \in \mathcal{O}_n$). Then $\mathbf{QZ} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$.

Proof.
The pdf of $\mathbf{QZ}$ is

$$
\begin{aligned}
f_{\mathbf{QZ}}(\mathbf{w}) &= \frac{1}{|\det \mathbf{Q}|} f_{\mathbf{Z}}(\mathbf{Q}^{-1}\mathbf{w}) = \frac{1}{1} \cdot \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{Q}^{-1}\mathbf{w})^\top(\mathbf{Q}^{-1}\mathbf{w})} \\
&= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{Q}^\top\mathbf{w})^\top(\mathbf{Q}^\top\mathbf{w})} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\mathbf{w}^\top\mathbf{Q})(\mathbf{Q}^\top\mathbf{w})} \\
&= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{w}^\top(\mathbf{Q}\mathbf{Q}^\top)\mathbf{w}} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{w}^\top\mathbf{w}}
\end{aligned}
$$

Hence, $\mathbf{QZ} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$. $\square$

Linear Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$. If $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\mathrm{rank}\mathbf{A} = \min(m, n)$, then

$$\mathbf{AY} \sim \mathcal{N}_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}^\top)$$

Proof.
Since $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, we can write

$$\mathbf{Y} = \mathbf{BZ} + \boldsymbol{\mu},$$

Linear Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$. If $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\mathrm{rank}\mathbf{A} = \min(m, n)$, then

$$\mathbf{AY} \sim \mathcal{N}_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}^\top)$$

Proof.
Since $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, we can write

$$\mathbf{Y} = \mathbf{BZ} + \boldsymbol{\mu},$$

where $\mathbf{B} \in \mathbb{R}^{n \times k}$, $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I})$, and $\mathbf{BB}^\top = \Sigma$.

Linear Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$. If $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\mathrm{rank}\mathbf{A} = \min(m, n)$, then

$$\mathbf{AY} \sim \mathcal{N}_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}^\top)$$

Proof.
Since $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, we can write

$$\mathbf{Y} = \mathbf{BZ} + \boldsymbol{\mu},$$

where $\mathbf{B} \in \mathbb{R}^{n \times k}$, $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I})$, and $\mathbf{BB}^\top = \Sigma$. Thus,

$$\mathbf{AY} = (\mathbf{AB})\mathbf{Z} + \mathbf{A}\boldsymbol{\mu},$$

Linear Transform of a Multivariate Gaussian

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$. If $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\mathrm{rank}\mathbf{A} = \min(m, n)$, then

$$\mathbf{AY} \sim \mathcal{N}_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}^\top)$$

Proof.
Since $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, we can write

$$\mathbf{Y} = \mathbf{BZ} + \boldsymbol{\mu},$$

where $\mathbf{B} \in \mathbb{R}^{n \times k}$, $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I})$, and $\mathbf{BB}^\top = \Sigma$. Thus,

$$\mathbf{AY} = (\mathbf{AB})\mathbf{Z} + \mathbf{A}\boldsymbol{\mu},$$

where $\mathbf{AB} \in \mathbb{R}^{m \times k}$ s.t. $(\mathbf{AB})(\mathbf{AB})^\top = \mathbf{ABB}^\top\mathbf{A} = \mathbf{A}\Sigma\mathbf{A}^\top$. $\square$

Non-central χ^2 distribution

Definition. Let $Y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, 1)$,

Non-central χ^2 distribution

Definition. Let $Y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, 1)$, then the random variable

$$\sum_{i=1}^{n} Y_i^2$$

is said to have **noncentral chi-square** distribution with parameters n and $\lambda = \sum_{i=1}^{n} \mu_i^2$.

Non-central χ^2 distribution

Definition. Let $Y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, 1)$, then the random variable

$$\sum_{i=1}^{n} Y_i^2$$

is said to have **noncentral chi-square** distribution with parameters n and $\lambda = \sum_{i=1}^{n} \mu_i^2$.

Notation. $\chi_n^2(\lambda)$

Non-central χ^2 distribution

Definition. Let $Y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, 1)$, then the random variable

$$\sum_{i=1}^{n} Y_i^2$$

is said to have **noncentral chi-square** distribution with parameters n and $\lambda = \sum_{i=1}^{n} \mu_i^2$.

Notation. $\chi_n^2(\lambda)$

- n is called *degrees of freedom*

Non-central χ^2 distribution

Definition. Let $Y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, 1)$, then the random variable

$$\sum_{i=1}^{n} Y_i^2$$

is said to have **noncentral chi-square** distribution with parameters n and $\lambda = \sum_{i=1}^{n} \mu_i^2$.

Notation. $\chi_n^2(\lambda)$

- n is called *degrees of freedom*
- λ is called *noncentrality parameter*

Non-central χ^2 distribution

Definition. Let $Y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, 1)$, then the random variable

$$\sum_{i=1}^{n} Y_i^2$$

is said to have **noncentral chi-square** distribution with parameters n and $\lambda = \sum_{i=1}^{n} \mu_i^2$.

Notation. $\chi_n^2(\lambda)$

- n is called *degrees of freedom*
- λ is called *noncentrality parameter*

Remark. Using matrix notation, if $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \mathbf{I})$,

Non-central χ^2 distribution

Definition. Let $Y_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, 1)$, then the random variable

$$\sum_{i=1}^{n} Y_i^2$$

is said to have **noncentral chi-square** distribution with parameters n and $\lambda = \sum_{i=1}^{n} \mu_i^2$.

Notation. $\chi_n^2(\lambda)$

- n is called *degrees of freedom*
- λ is called *noncentrality parameter*

Remark. Using matrix notation, if $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \mathbf{I})$,then $\mathbf{Y}^\top \mathbf{Y} \sim \chi_n^2(\lambda)$, with $\lambda = \boldsymbol{\mu}^\top \boldsymbol{\mu}$.

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} \sim \chi^2_r$.

Proof.

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$.

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n\times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A}\mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x}$$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A}\mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x}$$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A}\mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x}$$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A}\mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} = \lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} = \lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

So, $(\lambda^2 - \lambda)\mathbf{x} = \mathbf{0}$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} \sim \chi_r^2$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} = \lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

So, $(\lambda^2 - \lambda)\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \lambda^2 - \lambda = 0 \quad \Rightarrow \quad \lambda \in \{0, 1\}$.

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n\times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A}\mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} = \lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

So, $(\lambda^2 - \lambda)\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \lambda^2 - \lambda = 0 \quad \Rightarrow \quad \lambda \in \{0, 1\}$.

Eigendecomposition of $\mathbf{A}$:

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} \sim \chi_r^2$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} = \lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

So, $(\lambda^2 - \lambda)\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \lambda^2 - \lambda = 0 \quad \Rightarrow \quad \lambda \in \{0, 1\}$.

Eigendecomposition of $\mathbf{A}$: $\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$,

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n\times n}$ be symmetric idempotent with $\operatorname{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A}\mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} = \lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

So, $(\lambda^2 - \lambda)\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \lambda^2 - \lambda = 0 \quad \Rightarrow \quad \lambda \in \{0, 1\}$.

Eigendecomposition of $\mathbf{A}$: $\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$, where $\Lambda = \operatorname{diag}(\underbrace{1, \ldots, 1}_{r}, \underbrace{0, \ldots, 0}_{n-r})$.

Thus, $\mathbf{Z}^\top \mathbf{A}\mathbf{Z}$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} = \lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

So, $(\lambda^2 - \lambda)\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \lambda^2 - \lambda = 0 \quad \Rightarrow \quad \lambda \in \{0, 1\}$.

Eigendecomposition of $\mathbf{A}$: $\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$, where $\Lambda = \mathrm{diag}(\underbrace{1, \ldots, 1}_{r}, \underbrace{0, \ldots, 0}_{n-r})$.

Thus, $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} = \mathbf{Z}^\top \mathbf{Q}^\top \Lambda \mathbf{Q} \mathbf{Z}$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n\times n}$ be symmetric idempotent with $\operatorname{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A}\mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} = \lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

So, $(\lambda^2 - \lambda)\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \lambda^2 - \lambda = 0 \quad \Rightarrow \quad \lambda \in \{0, 1\}$.

Eigendecomposition of $\mathbf{A}$: $\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$, where $\Lambda = \operatorname{diag}(\underbrace{1, \ldots, 1}_{r}, \underbrace{0, \ldots, 0}_{n-r})$.

Thus, $\mathbf{Z}^\top\mathbf{A}\mathbf{Z} = \mathbf{Z}^\top\mathbf{Q}^\top\Lambda\mathbf{Q}\mathbf{Z} = (\mathbf{Q}\mathbf{Z})^\top\Lambda(\mathbf{Q}\mathbf{Z})$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\operatorname{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} = \lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

So, $(\lambda^2 - \lambda)\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \lambda^2 - \lambda = 0 \quad \Rightarrow \quad \lambda \in \{0, 1\}$.

Eigendecomposition of $\mathbf{A}$: $\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$, where $\Lambda = \operatorname{diag}(\underbrace{1, \ldots, 1}_{r}, \underbrace{0, \ldots, 0}_{n-r})$.

Thus, $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} = \mathbf{Z}^\top \mathbf{Q}^\top \Lambda \mathbf{Q} \mathbf{Z} = (\mathbf{Q}\mathbf{Z})^\top \Lambda (\mathbf{Q}\mathbf{Z}) = \tilde{\mathbf{Z}}^\top \Lambda \tilde{\mathbf{Z}}$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} = \lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

So, $(\lambda^2 - \lambda)\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \lambda^2 - \lambda = 0 \quad \Rightarrow \quad \lambda \in \{0, 1\}$.

Eigendecomposition of $\mathbf{A}$: $\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$, where $\Lambda = \mathrm{diag}(\underbrace{1, \ldots, 1}_{r}, \underbrace{0, \ldots, 0}_{n-r})$.

Thus, $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} = \mathbf{Z}^\top \mathbf{Q}^\top \Lambda \mathbf{Q} \mathbf{Z} = (\mathbf{Q}\mathbf{Z})^\top \Lambda (\mathbf{Q}\mathbf{Z}) = \tilde{\mathbf{Z}}^\top \Lambda \tilde{\mathbf{Z}} = \sum_{i=1}^r \tilde{Z}_i^2 \sim \chi^2_r$,

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Z}^\top \mathbf{A} \mathbf{Z} \sim \chi^2_r$.

Proof.
$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$. Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ and λ is the corresponding eigenvalue. Then

$$\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{A}\lambda\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x} = \lambda\mathbf{x} = \mathbf{A}\mathbf{x}$$

So, $(\lambda^2 - \lambda)\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \lambda^2 - \lambda = 0 \quad \Rightarrow \quad \lambda \in \{0, 1\}$.

Eigendecomposition of $\mathbf{A}$: $\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$, where $\Lambda = \mathrm{diag}(\underbrace{1, \ldots, 1}_{r}, \underbrace{0, \ldots, 0}_{n-r})$.

Thus, $\mathbf{Z}^\top \mathbf{A}\mathbf{Z} = \mathbf{Z}^\top \mathbf{Q}^\top \Lambda \mathbf{Q}\mathbf{Z} = (\mathbf{Q}\mathbf{Z})^\top \Lambda (\mathbf{Q}\mathbf{Z}) = \tilde{\mathbf{Z}}^\top \Lambda \tilde{\mathbf{Z}} = \sum_{i=1}^r \tilde{Z}_i^2 \sim \chi^2_r$, where $\tilde{\mathbf{Z}} = (\tilde{Z}_1, \ldots, \tilde{Z}_n) \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I})$. $\square$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Y}^\top \mathbf{A} \mathbf{Y} \sim \chi_r^2(\lambda)$, with $\lambda = \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}$

Proof.

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Y}^\top \mathbf{A} \mathbf{Y} \sim \chi^2_r(\lambda)$, with $\lambda = \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}$

Proof.
$\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$, where $\Lambda = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$.

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Y}^\top \mathbf{A} \mathbf{Y} \sim \chi^2_r(\lambda)$, with $\lambda = \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}$

Proof.
$\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$, where $\Lambda = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$. Thus,

$$\mathbf{Y}^\top \mathbf{A} \mathbf{Y} = (\mathbf{Q}\mathbf{Y})^\top \Lambda (\mathbf{Q}\mathbf{Y})$$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\operatorname{rank} \mathbf{A} = r$. Then $\mathbf{Y}^\top \mathbf{A} \mathbf{Y} \sim \chi_r^2(\lambda)$, with $\lambda = \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}$

Proof.
$\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$, where $\Lambda = \operatorname{diag}(1, \ldots, 1, 0, \ldots, 0)$. Thus,

$$\mathbf{Y}^\top \mathbf{A} \mathbf{Y} = (\mathbf{Q}\mathbf{Y})^\top \Lambda (\mathbf{Q}\mathbf{Y})$$

Notice that $\tilde{\mathbf{Y}} := \mathbf{Q}\mathbf{Y} \sim \mathcal{N}_n(\mathbf{Q}\boldsymbol{\mu}, \mathbf{I})$. (Why?)

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric idempotent with $\mathrm{rank}\mathbf{A} = r$. Then $\mathbf{Y}^\top \mathbf{A} \mathbf{Y} \sim \chi_r^2(\lambda)$, with $\lambda = \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}$

Proof.
$\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$, where $\Lambda = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$. Thus,

$$\mathbf{Y}^\top \mathbf{A} \mathbf{Y} = (\mathbf{Q}\mathbf{Y})^\top \Lambda (\mathbf{Q}\mathbf{Y})$$

Notice that $\tilde{\mathbf{Y}} := \mathbf{Q}\mathbf{Y} \sim \mathcal{N}_n(\mathbf{Q}\boldsymbol{\mu}, \mathbf{I})$. (Why?) Therefore,

$$\mathbf{Y}^\top \mathbf{A} \mathbf{Y} = \tilde{\mathbf{Y}}^\top \Lambda \tilde{\mathbf{Y}} = \sum_{i=1}^{r} \tilde{Y}_i^2 \sim \chi_r^2(\lambda),$$

Quadratic Form of a Multivariate Gaussian

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \mathbf{I})$, and $\mathbf{A} \in \mathbb{R}^{n\times n}$ be symmetric idempotent with $\operatorname{rank}\mathbf{A} = r$. Then $\mathbf{Y}^\top \mathbf{A}\mathbf{Y} \sim \chi_r^2(\lambda)$, with $\lambda = \boldsymbol{\mu}^\top \mathbf{A}\boldsymbol{\mu}$

Proof.
$\mathbf{A} = \mathbf{Q}^\top \Lambda \mathbf{Q}$, where $\Lambda = \operatorname{diag}(1, \ldots, 1, 0, \ldots, 0)$. Thus,

$$\mathbf{Y}^\top \mathbf{A}\mathbf{Y} = (\mathbf{Q}\mathbf{Y})^\top \Lambda (\mathbf{Q}\mathbf{Y})$$

Notice that $\tilde{\mathbf{Y}} := \mathbf{Q}\mathbf{Y} \sim \mathcal{N}_n(\mathbf{Q}\boldsymbol{\mu}, \mathbf{I})$. (Why?) Therefore,

$$\mathbf{Y}^\top \mathbf{A}\mathbf{Y} = \tilde{\mathbf{Y}}^\top \Lambda \tilde{\mathbf{Y}} = \sum_{i=1}^{r} \tilde{Y}_i^2 \sim \chi_r^2(\lambda),$$

where $\lambda = (\Lambda \mathbf{Q}\boldsymbol{\mu})^\top (\Lambda \mathbf{Q}\boldsymbol{\mu}) = \boldsymbol{\mu}^\top \mathbf{Q}^\top \Lambda^\top \Lambda \mathbf{Q}\boldsymbol{\mu} = \boldsymbol{\mu}^\top \mathbf{Q}^\top \Lambda \mathbf{Q}\boldsymbol{\mu}$
$= \boldsymbol{\mu}^\top \mathbf{A}\boldsymbol{\mu}$. $\square$

Matrix Calculus

It is useful to have differentiation rules for matrix and vector expressions.

Matrix Calculus

It is useful to have differentiation rules for matrix and vector expressions. The most important for our purposes are

$$\nabla_{\mathbf{x}}(\mathbf{a}^\top \mathbf{x}) = \mathbf{a}$$
$$\nabla_{\mathbf{x}}(\mathbf{x}^\top \mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}^\top)\mathbf{x}$$

Matrix Calculus

It is useful to have differentiation rules for matrix and vector expressions. The most important for our purposes are

$$\nabla_{\mathbf{x}}(\mathbf{a}^\top \mathbf{x}) = \mathbf{a}$$
$$\nabla_{\mathbf{x}}(\mathbf{x}^\top \mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}^\top)\mathbf{x}$$

Note that the second rule is defined only if $\mathbf{A}$ is square. Furthermore, if $\mathbf{A}$ is symmetric, we can simplify the result to $2\mathbf{A}\mathbf{x}$.

Matrix Calculus

It is useful to have differentiation rules for matrix and vector expressions. The most important for our purposes are

$$\nabla_{\mathbf{x}}(\mathbf{a}^\top \mathbf{x}) = \mathbf{a}$$
$$\nabla_{\mathbf{x}}(\mathbf{x}^\top \mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}^\top)\mathbf{x}$$

Note that the second rule is defined only if **A** is square. Furthermore, if **A** is symmetric, we can simplify the result to $2\mathbf{A}\mathbf{x}$.

Exercise. Prove the rules above.

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

Least squares objective is

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n}[Y_i - (\beta_0 + \beta_1 x_i)]^2$$

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

Least squares objective is

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n}[Y_i - (\beta_0 + \beta_1 x_i)]^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

Least squares objective is

$$\begin{aligned} Q(\beta_0, \beta_1) &= \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 x_i)]^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}^\top \mathbf{Y} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \end{aligned}$$

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

Least squares objective is

$$\begin{aligned}
Q(\beta_0, \beta_1) &= \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 x_i)]^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{Y}^\top \mathbf{Y} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\
\nabla_{\boldsymbol{\beta}} Q &= \mathbf{X}^\top \mathbf{Y} - \mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}
\end{aligned}$$

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

Least squares objective is

$$\begin{aligned} Q(\beta_0, \beta_1) &= \sum_{i=1}^{n}[Y_i - (\beta_0 + \beta_1 x_i)]^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}^\top\mathbf{Y} - \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} \\ \nabla_{\boldsymbol{\beta}} Q &= \mathbf{X}^\top\mathbf{Y} - \mathbf{X}^\top\mathbf{Y} + 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}^\top\mathbf{Y} \end{aligned}$$

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

Least squares objective is

$$\begin{aligned}
Q(\beta_0, \beta_1) &= \sum_{i=1}^{n}[Y_i - (\beta_0 + \beta_1 x_i)]^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{Y}^\top\mathbf{Y} - \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} \\
\nabla_{\boldsymbol{\beta}} Q &= \mathbf{X}^\top\mathbf{Y} - \mathbf{X}^\top\mathbf{Y} + 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}^\top\mathbf{Y} = \mathbf{0}
\end{aligned}$$

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

Least squares objective is

$$\begin{aligned}
Q(\beta_0, \beta_1) &= \sum_{i=1}^{n}[Y_i - (\beta_0 + \beta_1 x_i)]^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{Y}^\top\mathbf{Y} - \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} \\
\nabla_{\boldsymbol{\beta}} Q &= \mathbf{X}^\top\mathbf{Y} - \mathbf{X}^\top\mathbf{Y} + 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}^\top\mathbf{Y} = \mathbf{0} \\
\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^\top\mathbf{Y} \quad \Rightarrow
\end{aligned}$$

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

Least squares objective is

$$\begin{aligned}
Q(\beta_0, \beta_1) &= \sum_{i=1}^{n}[Y_i - (\beta_0 + \beta_1 x_i)]^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{Y}^\top\mathbf{Y} - \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} \\
\nabla_{\boldsymbol{\beta}} Q &= \mathbf{X}^\top\mathbf{Y} - \mathbf{X}^\top\mathbf{Y} + 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}^\top\mathbf{Y} = \mathbf{0} \\
\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^\top\mathbf{Y} \quad \Rightarrow \quad \boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}}
\end{aligned}$$

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

Least squares objective is

$$\begin{aligned}
Q(\beta_0, \beta_1) &= \sum_{i=1}^{n}[Y_i - (\beta_0 + \beta_1 x_i)]^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{Y}^\top\mathbf{Y} - \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} \\
\nabla_{\boldsymbol{\beta}} Q &= \mathbf{X}^\top\mathbf{Y} - \mathbf{X}^\top\mathbf{Y} + 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}^\top\mathbf{Y} = \mathbf{0} \\
\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^\top\mathbf{Y} \quad \Rightarrow \quad \boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}}
\end{aligned}$$

where we used $\mathbf{X}^\top\mathbf{X} = \begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}$ is non-singular (check).

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

Least squares objective is

$$\begin{aligned}
Q(\beta_0, \beta_1) &= \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 x_i)]^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{Y}^\top \mathbf{Y} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\
\nabla_{\boldsymbol{\beta}} Q &= \mathbf{X}^\top \mathbf{Y} - \mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}^\top \mathbf{Y} = \mathbf{0} \\
\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^\top \mathbf{Y} \quad \Rightarrow \quad \boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}}
\end{aligned}$$

where we used $\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}$ is non-singular (check).

The predictions are $(\hat{Y}_1, \ldots, \hat{Y}_n) = \hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$

Back to SLR

Recall the SLR in matrix form:

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

Least squares objective is

$$\begin{aligned}
Q(\beta_0, \beta_1) &= \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 x_i)]^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{Y}^\top \mathbf{Y} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} \\
\nabla_{\boldsymbol{\beta}} Q &= \mathbf{X}^\top \mathbf{Y} - \mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}^\top \mathbf{Y} = \mathbf{0} \\
\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^\top \mathbf{Y} \quad \Rightarrow \quad \boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}}
\end{aligned}$$

where we used $\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}$ is non-singular (check).

The predictions are $(\hat{Y}_1, \ldots, \hat{Y}_n) = \hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$

Random Vectors

Multivariate Gaussian Distribution

Quadratic Forms of Multivariate Gaussians

Distributions of SSE, SSR; Independence

Distribution of SSE

Theorem. $\frac{\mathrm{SSE}}{\sigma^2} \sim \chi^2_{n-2}$

Proof.

Distribution of SSE

Theorem. $\frac{\text{SSE}}{\sigma^2} \sim \chi^2_{n-2}$

Proof.

$$\text{SSE} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Distribution of SSE

Theorem. $\frac{\mathrm{SSE}}{\sigma^2} \sim \chi^2_{n-2}$

Proof.

$$\mathrm{SSE} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^\top (\mathbf{Y} - \hat{\mathbf{Y}})$$

Distribution of SSE

Theorem. $\frac{\text{SSE}}{\sigma^2} \sim \chi^2_{n-2}$

Proof.

$$\begin{aligned}\text{SSE} &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^\top(\mathbf{Y} - \hat{\mathbf{Y}}) \\ &= [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}]^\top[\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}]\end{aligned}$$

Distribution of SSE

Theorem. $\frac{\text{SSE}}{\sigma^2} \sim \chi^2_{n-2}$

Proof.

$$\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^\top (\mathbf{Y} - \hat{\mathbf{Y}}) \\
&= [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}]^\top [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}] \\
&= \mathbf{Y}^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top]^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{Y}
\end{aligned}$$

Distribution of SSE

Theorem. $\frac{\text{SSE}}{\sigma^2} \sim \chi^2_{n-2}$

Proof.

$$\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^\top (\mathbf{Y} - \hat{\mathbf{Y}}) \\
&= [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y}]^\top [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y}] \\
&= \mathbf{Y}^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top]^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top] \mathbf{Y}
\end{aligned}$$

The matrix $\mathbf{A} := \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$ is symmetric (check) and idempotent:

Distribution of SSE

Theorem. $\frac{\text{SSE}}{\sigma^2} \sim \chi^2_{n-2}$

Proof.

$$\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^\top (\mathbf{Y} - \hat{\mathbf{Y}}) \\
&= [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}]^\top [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}] \\
&= \mathbf{Y}^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top]^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{Y}
\end{aligned}$$

The matrix $\mathbf{A} := \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ is symmetric (check) and idempotent:

$$\mathbf{A}^2 = [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top][\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top]$$

Distribution of SSE

Theorem. $\frac{\text{SSE}}{\sigma^2} \sim \chi^2_{n-2}$

Proof.

$$\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^\top (\mathbf{Y} - \hat{\mathbf{Y}}) \\
&= [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}]^\top [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}] \\
&= \mathbf{Y}^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]\mathbf{Y}
\end{aligned}$$

The matrix $\mathbf{A} := \mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ is symmetric (check) and idempotent:

$$\begin{aligned}
\mathbf{A}^2 &= [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top][\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= \mathbf{I} - 2\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top + \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top
\end{aligned}$$

Distribution of SSE

Theorem. $\frac{\text{SSE}}{\sigma^2} \sim \chi^2_{n-2}$

Proof.

$$\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})^\top (\mathbf{Y} - \hat{\mathbf{Y}}) \\
&= [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}]^\top [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}] \\
&= \mathbf{Y}^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top]^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \mathbf{Y}
\end{aligned}$$

The matrix $\mathbf{A} := \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ is symmetric (check) and idempotent:

$$\begin{aligned}
\mathbf{A}^2 &= [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top][\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \\
&= \mathbf{I} - 2\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top + \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \\
&= \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top = \mathbf{A}
\end{aligned}$$

Distribution of SSE

Theorem. $\frac{\text{SSE}}{\sigma^2} \sim \chi^2_{n-2}$

Proof.

$$
\begin{aligned}
\text{SSE} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 &= (\mathbf{Y} - \hat{\mathbf{Y}})^\top (\mathbf{Y} - \hat{\mathbf{Y}}) \\
&= [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}]^\top [\mathbf{Y} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}] \\
&= \mathbf{Y}^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]^\top [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]\mathbf{Y}
\end{aligned}
$$

The matrix $\mathbf{A} := \mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ is symmetric (check) and idempotent:

$$
\begin{aligned}
\mathbf{A}^2 &= [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top][\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= \mathbf{I} - 2\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top + \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \\
&= \mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \mathbf{A}
\end{aligned}
$$

Hence, $\text{SSE} = \mathbf{Y}^\top\mathbf{A}^\top\mathbf{A}\mathbf{Y} = \mathbf{Y}^\top\mathbf{A}\mathbf{Y}$.

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$.

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\text{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A}(\sigma^{-1}\mathbf{Y})$$

$\square$

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\mathsf{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{A}}(\lambda),$$

$\square$

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\mathsf{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{A}}(\lambda),$$

where $\quad \mathrm{rank}\mathbf{A} = \mathrm{trace}\mathbf{A}$

$\square$

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\text{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A} (\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{A}}(\lambda),$$

where $\quad \mathrm{rank}\mathbf{A} = \mathrm{trace}\mathbf{A} = \mathrm{trace}[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]$

$\square$

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\mathsf{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{A}}(\lambda),$$

$$\text{where} \quad \mathrm{rank}\mathbf{A} = \mathrm{trace}\mathbf{A} = \mathrm{trace}[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]$$
$$= \mathrm{trace}\mathbf{I} - \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]$$

$\square$

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\text{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{A}}(\lambda),$$

$$\begin{aligned} \text{where} \quad \mathrm{rank}\mathbf{A} &= \mathrm{trace}\mathbf{A} = \mathrm{trace}[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\ &= \mathrm{trace}\mathbf{I} - \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\ &= n - \mathrm{trace}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}] \end{aligned}$$

$\square$

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\mathrm{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{A}}(\lambda),$$

$$\begin{aligned} \text{where} \quad \mathrm{rank}\mathbf{A} &= \mathrm{trace}\mathbf{A} = \mathrm{trace}[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\ &= \mathrm{trace}\mathbf{I} - \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\ &= n - \mathrm{trace}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}] = n - 2, \end{aligned}$$

$\square$

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\text{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{A}}(\lambda),$$

$$\begin{aligned}
\text{where} \quad \mathrm{rank}\mathbf{A} &= \mathrm{trace}\mathbf{A} = \mathrm{trace}[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= \mathrm{trace}\mathbf{I} - \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= n - \mathrm{trace}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}] = n - 2, \\
\text{and} \quad \lambda &= (\sigma^{-1}\mathbf{X}\boldsymbol{\beta})^\top \mathbf{A}(\sigma^{-1}\mathbf{X}\boldsymbol{\beta})
\end{aligned}$$

□

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\text{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A} (\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{A}}(\lambda),$$

$$\begin{aligned}
\text{where} \quad \mathrm{rank}\mathbf{A} &= \mathrm{trace}\mathbf{A} = \mathrm{trace}[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= \mathrm{trace}\mathbf{I} - \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= n - \mathrm{trace}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}] = n - 2, \\
\text{and} \quad \lambda &= (\sigma^{-1}\mathbf{X}\boldsymbol{\beta})^\top \mathbf{A} (\sigma^{-1}\mathbf{X}\boldsymbol{\beta}) \\
&= \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\mathbf{X}^\top[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]\mathbf{X}\boldsymbol{\beta}
\end{aligned}$$

$\square$

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\text{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{A}}(\lambda),$$

$$\begin{aligned}
\text{where} \quad \mathrm{rank}\mathbf{A} &= \mathrm{trace}\mathbf{A} = \mathrm{trace}[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= \mathrm{trace}\mathbf{I} - \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= n - \mathrm{trace}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}] = n - 2, \\
\text{and} \quad \lambda &= (\sigma^{-1}\mathbf{X}\boldsymbol{\beta})^\top \mathbf{A}(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}) \\
&= \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\mathbf{X}^\top[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]\mathbf{X}\boldsymbol{\beta} \\
&= \frac{1}{\sigma^2}\boldsymbol{\beta}^\top[\mathbf{X}^\top\mathbf{X} - \mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}]\boldsymbol{\beta}
\end{aligned}$$

$\square$

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\text{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{A}}(\lambda),$$

$$\begin{aligned}
\text{where} \quad \mathrm{rank}\mathbf{A} &= \mathrm{trace}\mathbf{A} = \mathrm{trace}[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= \mathrm{trace}\mathbf{I} - \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= n - \mathrm{trace}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}] = n - 2, \\
\text{and} \quad \lambda &= (\sigma^{-1}\mathbf{X}\boldsymbol{\beta})^\top \mathbf{A}(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}) \\
&= \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\mathbf{X}^\top[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]\mathbf{X}\boldsymbol{\beta} \\
&= \frac{1}{\sigma^2}\boldsymbol{\beta}^\top[\mathbf{X}^\top\mathbf{X} - \mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}]\boldsymbol{\beta} = 0
\end{aligned}$$

$\square$

Distribution of SSE

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\text{SSE}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{A}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{A}}(\lambda),$$

$$\begin{aligned}
\text{where} \quad \mathrm{rank}\mathbf{A} &= \mathrm{trace}\mathbf{A} = \mathrm{trace}[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= \mathrm{trace}\mathbf{I} - \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= n - \mathrm{trace}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}] = n-2, \\
\text{and} \quad \lambda &= (\sigma^{-1}\mathbf{X}\boldsymbol{\beta})^\top \mathbf{A}(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}) \\
&= \frac{1}{\sigma^2}\boldsymbol{\beta}^\top\mathbf{X}^\top[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top]\mathbf{X}\boldsymbol{\beta} \\
&= \frac{1}{\sigma^2}\boldsymbol{\beta}^\top[\mathbf{X}^\top\mathbf{X} - \mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}]\boldsymbol{\beta} = 0
\end{aligned}$$

Finally, $\frac{\text{SSE}}{\sigma^2} \sim \chi^2_{n-2}$. $\square$

Distribution of SSR

Theorem. $\frac{\text{SSR}}{\sigma^2} \sim \chi_1^2$ (non-central)

Proof.

Distribution of SSR

Theorem. $\frac{\text{SSR}}{\sigma^2} \sim \chi_1^2$ (non-central)

Proof.
Let $\mathbf{1} := (1, \ldots, 1)$, then $\mathbf{1}^\top \mathbf{1} = n$

Distribution of SSR

Theorem. $\frac{\text{SSR}}{\sigma^2} \sim \chi_1^2$ (non-central)

Proof.
Let $\mathbf{1} := (1, \ldots, 1)$, then $\mathbf{1}^\top \mathbf{1} = n$, and $\bar{Y} = \frac{\mathbf{1}^\top \mathbf{Y}}{\mathbf{1}^\top \mathbf{1}}$.

Distribution of SSR

Theorem. $\frac{\text{SSR}}{\sigma^2} \sim \chi_1^2$ (non-central)

Proof.
Let $\mathbf{1} := (1, \ldots, 1)$, then $\mathbf{1}^\top \mathbf{1} = n$, and $\bar{Y} = \frac{\mathbf{1}^\top \mathbf{Y}}{\mathbf{1}^\top \mathbf{1}}$.

Let $\bar{\mathbf{Y}} := (\bar{Y}, \ldots, \bar{Y})$. Then $\bar{\mathbf{Y}} = \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top \mathbf{Y}$

Distribution of SSR

Theorem. $\frac{\text{SSR}}{\sigma^2} \sim \chi_1^2$ (non-central)

Proof.
Let $\mathbf{1} := (1, \ldots, 1)$, then $\mathbf{1}^\top \mathbf{1} = n$, and $\bar{Y} = \frac{\mathbf{1}^\top \mathbf{Y}}{\mathbf{1}^\top \mathbf{1}}$.

Let $\bar{\mathbf{Y}} := (\bar{Y}, \ldots, \bar{Y})$. Then $\bar{\mathbf{Y}} = \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top \mathbf{Y}$

$$\text{SSR} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

Distribution of SSR

Theorem. $\frac{\text{SSR}}{\sigma^2} \sim \chi_1^2$ (non-central)

Proof.
Let $\mathbf{1} := (1, \ldots, 1)$, then $\mathbf{1}^\top \mathbf{1} = n$, and $\bar{Y} = \frac{\mathbf{1}^\top \mathbf{Y}}{\mathbf{1}^\top \mathbf{1}}$.

Let $\bar{\mathbf{Y}} := (\bar{Y}, \ldots, \bar{Y})$. Then $\bar{\mathbf{Y}} = \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top \mathbf{Y}$

$$\text{SSR} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = (\hat{\mathbf{Y}} - \bar{\mathbf{Y}})^\top (\hat{\mathbf{Y}} - \bar{\mathbf{Y}})$$

Distribution of SSR

Theorem. $\frac{\text{SSR}}{\sigma^2} \sim \chi_1^2$ (non-central)

Proof.
Let $\mathbf{1} := (1, \ldots, 1)$, then $\mathbf{1}^\top \mathbf{1} = n$, and $\bar{Y} = \frac{\mathbf{1}^\top \mathbf{Y}}{\mathbf{1}^\top \mathbf{1}}$.

Let $\bar{\mathbf{Y}} := (\bar{Y}, \ldots, \bar{Y})$. Then $\bar{\mathbf{Y}} = \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top \mathbf{Y}$

$$
\begin{aligned}
\text{SSR} &= \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = (\hat{\mathbf{Y}} - \bar{\mathbf{Y}})^\top (\hat{\mathbf{Y}} - \bar{\mathbf{Y}}) \\
&= [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top \mathbf{Y}]^\top [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top \mathbf{Y}]
\end{aligned}
$$

Distribution of SSR

Theorem. $\frac{\text{SSR}}{\sigma^2} \sim \chi_1^2$ (non-central)

Proof.
Let $\mathbf{1} := (1, \ldots, 1)$, then $\mathbf{1}^\top \mathbf{1} = n$, and $\bar{Y} = \frac{\mathbf{1}^\top \mathbf{Y}}{\mathbf{1}^\top \mathbf{1}}$.

Let $\bar{\mathbf{Y}} := (\bar{Y}, \ldots, \bar{Y})$. Then $\bar{\mathbf{Y}} = \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top \mathbf{Y}$

$$
\begin{aligned}
\text{SSR} &= \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = (\hat{\mathbf{Y}} - \bar{\mathbf{Y}})^\top (\hat{\mathbf{Y}} - \bar{\mathbf{Y}}) \\
&= [\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top\mathbf{Y}]^\top [\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top\mathbf{Y}] \\
&= \mathbf{Y}^\top [\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top]^\top [\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top]\mathbf{Y}
\end{aligned}
$$

Distribution of SSR

Theorem. $\frac{\text{SSR}}{\sigma^2} \sim \chi_1^2$ (non-central)

Proof.
Let $\mathbf{1} := (1, \ldots, 1)$, then $\mathbf{1}^\top \mathbf{1} = n$, and $\bar{Y} = \frac{\mathbf{1}^\top \mathbf{Y}}{\mathbf{1}^\top \mathbf{1}}$.

Let $\bar{\mathbf{Y}} := (\bar{Y}, \ldots, \bar{Y})$. Then $\bar{\mathbf{Y}} = \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top \mathbf{Y}$

$$\begin{aligned}
\text{SSR} &= \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = (\hat{\mathbf{Y}} - \bar{\mathbf{Y}})^\top (\hat{\mathbf{Y}} - \bar{\mathbf{Y}}) \\
&= [\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top\mathbf{Y}]^\top[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top\mathbf{Y}] \\
&= \mathbf{Y}^\top[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top]^\top[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top]\mathbf{Y}
\end{aligned}$$

The matrix $\mathbf{B} := \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top$ is symmetric and idempotent (check).

Distribution of SSR

Theorem. $\frac{\text{SSR}}{\sigma^2} \sim \chi_1^2$ (non-central)

Proof.
Let $\mathbf{1} := (1, \ldots, 1)$, then $\mathbf{1}^\top \mathbf{1} = n$, and $\bar{Y} = \frac{\mathbf{1}^\top \mathbf{Y}}{\mathbf{1}^\top \mathbf{1}}$.

Let $\bar{\mathbf{Y}} := (\bar{Y}, \ldots, \bar{Y})$. Then $\bar{\mathbf{Y}} = \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top \mathbf{Y}$

$$\begin{aligned}
\text{SSR} &= \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = (\hat{\mathbf{Y}} - \bar{\mathbf{Y}})^\top (\hat{\mathbf{Y}} - \bar{\mathbf{Y}}) \\
&= [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top \mathbf{Y}]^\top [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top \mathbf{Y}] \\
&= \mathbf{Y}^\top [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top]^\top [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top]\mathbf{Y}
\end{aligned}$$

The matrix $\mathbf{B} := \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top$ is symmetric and idempotent (check). Hence, $\text{SSR} = \mathbf{Y}^\top \mathbf{B}^\top \mathbf{B}\mathbf{Y} = \mathbf{Y}^\top \mathbf{B}\mathbf{Y}$.

Distribution of SSR

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$

Distribution of SSR

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$.

Distribution of SSR

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\mathrm{SSR}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{B}(\sigma^{-1}\mathbf{Y})$$

$\square$

Distribution of SSR

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\mathrm{SSR}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{B}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{B}}(\lambda),$$

$\square$

Distribution of SSR

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\mathsf{SSR}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{B}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{B}}(\lambda),$$

where $\quad \mathrm{rank}\mathbf{B} = \mathrm{trace}\mathbf{B}$

$\square$

Distribution of SSR

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\mathrm{SSR}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{B}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{B}}(\lambda),$$

where $\quad \mathrm{rank}\mathbf{B} = \mathrm{trace}\mathbf{B} = \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top]$

$\square$

Distribution of SSR

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\mathrm{SSR}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{B} (\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{B}}(\lambda),$$

where

$$\begin{aligned} \mathrm{rank}\mathbf{B} &= \mathrm{trace}\mathbf{B} = \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top] \\ &= \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] - \mathrm{trace}[\mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top] \end{aligned}$$

$\square$

Distribution of SSR

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\text{SSR}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{B}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{B}}(\lambda),$$

where

$$\begin{aligned}
\mathrm{rank}\mathbf{B} &= \mathrm{trace}\mathbf{B} = \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top] \\
&= \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] - \mathrm{trace}[\mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top] \\
&= \mathrm{trace}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}] - \mathrm{trace}[(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top\mathbf{1}]
\end{aligned}$$

$\square$

Distribution of SSR

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\mathrm{SSR}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{B}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{B}}(\lambda),$$

where

$$\begin{aligned}
\mathrm{rank}\mathbf{B} &= \mathrm{trace}\mathbf{B} = \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top] \\
&= \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] - \mathrm{trace}[\mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top] \\
&= \mathrm{trace}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}] - \mathrm{trace}[(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top\mathbf{1}] \\
&= \mathrm{trace}\mathbf{I}_2 - \mathrm{trace}1 = 2 - 1 = 1
\end{aligned}$$

$\square$

Distribution of SSR

Proof (cont'd).

Recall, $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad \Rightarrow \quad \sigma^{-1}\mathbf{Y} \sim \mathcal{N}_n(\sigma^{-1}\mathbf{X}\boldsymbol{\beta}, \mathbf{I})$. Hence,

$$\frac{\mathrm{SSR}}{\sigma^2} = (\sigma^{-1}\mathbf{Y})^\top \mathbf{B}(\sigma^{-1}\mathbf{Y}) \sim \chi^2_{\mathrm{rank}\mathbf{B}}(\lambda),$$

$$\begin{aligned}
\text{where} \quad \mathrm{rank}\mathbf{B} &= \mathrm{trace}\mathbf{B} = \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top] \\
&= \mathrm{trace}[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] - \mathrm{trace}[\mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top] \\
&= \mathrm{trace}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}] - \mathrm{trace}[(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top\mathbf{1}] \\
&= \mathrm{trace}\mathbf{I}_2 - \mathrm{trace}1 = 2 - 1 = 1
\end{aligned}$$

Finally, $\frac{\mathrm{SSR}}{\sigma^2} \sim \chi^2_1$ (non-central). $\square$

Independence between SSR and SSE

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\Sigma \succ \mathbf{0}$. Then two quadratic forms $\mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ and $\mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ are independent $\quad \Leftrightarrow \quad \mathbf{A} \Sigma \mathbf{B} = \mathbf{0}$.

Independence between SSR and SSE

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\Sigma \succ \mathbf{0}$. Then two quadratic forms $\mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ and $\mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ are independent $\quad \Leftrightarrow \quad \mathbf{A} \Sigma \mathbf{B} = \mathbf{0}$.

Theorem. SSR and SSE are independent.

Proof.

Independence between SSR and SSE

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\Sigma \succ \mathbf{0}$. Then two quadratic forms $\mathbf{Y}^\top \mathbf{A}\mathbf{Y}$ and $\mathbf{Y}^\top \mathbf{B}\mathbf{Y}$ are independent $\quad \Leftrightarrow \quad \mathbf{A}\Sigma\mathbf{B} = \mathbf{0}$.

Theorem. SSR and SSE are independent.

Proof.
$\text{SSE} = \mathbf{Y}^\top \mathbf{A}\mathbf{Y}$ with $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$

Independence between SSR and SSE

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\Sigma \succ \mathbf{0}$. Then two quadratic forms $\mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ and $\mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ are independent $\quad \Leftrightarrow \quad \mathbf{A} \Sigma \mathbf{B} = \mathbf{0}$.

Theorem. SSR and SSE are independent.

Proof.
$\mathrm{SSE} = \mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ with $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$
$\mathrm{SSR} = \mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ with $\mathbf{B} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top$

Independence between SSR and SSE

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\Sigma \succ \mathbf{0}$. Then two quadratic forms $\mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ and $\mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ are independent $\quad \Leftrightarrow \quad \mathbf{A}\Sigma\mathbf{B} = \mathbf{0}$.

Theorem. SSR and SSE are independent.

Proof.
$\text{SSE} = \mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ with $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$
$\text{SSR} = \mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ with $\mathbf{B} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top$
$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$.

Independence between SSR and SSE

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\Sigma \succ \mathbf{0}$. Then two quadratic forms $\mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ and $\mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ are independent $\quad\Leftrightarrow\quad \mathbf{A}\Sigma\mathbf{B} = \mathbf{0}$.

Theorem. SSR and SSE are independent.

Proof.

$\text{SSE} = \mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ with $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$

$\text{SSR} = \mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ with $\mathbf{B} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top$

$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$. And now let's check

$\mathbf{A}\sigma^2\mathbf{I}\mathbf{B} = \sigma^2 \mathbf{A}\mathbf{B}$

Independence between SSR and SSE

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\Sigma \succ \mathbf{0}$. Then two quadratic forms $\mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ and $\mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ are independent $\quad \Leftrightarrow \quad \mathbf{A} \Sigma \mathbf{B} = \mathbf{0}$.

Theorem. SSR and SSE are independent.

Proof.

SSE $= \mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ with $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$

SSR $= \mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ with $\mathbf{B} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top$

$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$. And now let's check

$$\mathbf{A} \sigma^2 \mathbf{I} \mathbf{B} = \sigma^2 \mathbf{A} \mathbf{B}$$

$$= \sigma^2 [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top][\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top]$$

Independence between SSR and SSE

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\Sigma \succ \mathbf{0}$. Then two quadratic forms $\mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ and $\mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ are independent $\Leftrightarrow$ $\mathbf{A} \Sigma \mathbf{B} = \mathbf{0}$.

Theorem. SSR and SSE are independent.

Proof.
SSE $= \mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ with $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$
SSR $= \mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ with $\mathbf{B} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top$
$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$. And now let's check

$$\mathbf{A}\sigma^2 \mathbf{I}\mathbf{B} = \sigma^2 \mathbf{A}\mathbf{B}$$
$$= \sigma^2 [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top][\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top]$$
$$= \sigma^2 [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top]$$

Independence between SSR and SSE

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\Sigma \succ \mathbf{0}$. Then two quadratic forms $\mathbf{Y}^\top \mathbf{A}\mathbf{Y}$ and $\mathbf{Y}^\top \mathbf{B}\mathbf{Y}$ are independent $\quad\Leftrightarrow\quad \mathbf{A}\Sigma\mathbf{B} = \mathbf{0}$.

Theorem. SSR and SSE are independent.

Proof.
$\text{SSE} = \mathbf{Y}^\top \mathbf{A}\mathbf{Y}$ with $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$
$\text{SSR} = \mathbf{Y}^\top \mathbf{B}\mathbf{Y}$ with $\mathbf{B} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top$
$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. And now let's check

$$
\begin{aligned}
&\mathbf{A}\sigma^2\mathbf{I}\mathbf{B} = \sigma^2\mathbf{A}\mathbf{B} \\
&= \sigma^2[\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top][\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top] \\
&= \sigma^2[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top\mathbf{1})^{-1}\mathbf{1}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] \\
&= \mathbf{0}
\end{aligned}
$$

Independence between SSR and SSE

Theorem. Let $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, and $\Sigma \succ \mathbf{0}$. Then two quadratic forms $\mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ and $\mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ are independent $\quad \Leftrightarrow \quad \mathbf{A} \Sigma \mathbf{B} = \mathbf{0}$.

Theorem. SSR and SSE are independent.

Proof.
$\text{SSE} = \mathbf{Y}^\top \mathbf{A} \mathbf{Y}$ with $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$
$\text{SSR} = \mathbf{Y}^\top \mathbf{B} \mathbf{Y}$ with $\mathbf{B} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top$
$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$. And now let's check

$$
\begin{aligned}
&\mathbf{A} \sigma^2 \mathbf{I} \mathbf{B} = \sigma^2 \mathbf{A} \mathbf{B} \\
&= \sigma^2 [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top][\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top] \\
&= \sigma^2 [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \\
&= \mathbf{0}
\end{aligned}
$$

Hence, SSE and SSR are independent. $\square$