1. Matrix Algebra

1.1 Partitioned matrix and its inverse

Addition and multiplication
\[
A + B = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} + \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{pmatrix}
\]
\[
AB = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}
\]

Determinant of a partitioned matrix
\[
jA_j = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = jA_{22} \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}
\]
\[
jA_{11} \begin{vmatrix} A_{12} & A_{21} \\ A_{12} & A_{21} \end{vmatrix}
\]
promvided that both \( A_{11} \) and \( A_{22} \) are non-singular.

Inverse of a partitioned matrix
\[
A^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{vmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{vmatrix}
\]
\[
i A_{11}^{-1} A_{22}^{-1} = \begin{vmatrix} A_{11}^{-1} & A_{12}^{-1} \\ A_{21}^{-1} & A_{22}^{-1} \end{vmatrix}
\]
\[
F_{11}^{-1} A_{11}^{-1} A_{12}^{-1} \begin{vmatrix} A_{11}^{-1} & A_{12}^{-1} \\ A_{21}^{-1} & A_{22}^{-1} \end{vmatrix}
\]
\[
F_{22}^{-1} A_{22}^{-1} A_{21}^{-1} \begin{vmatrix} A_{11}^{-1} & A_{12}^{-1} \\ A_{21}^{-1} & A_{22}^{-1} \end{vmatrix}
\]
\[
F_{11}^{-1} A_{11}^{-1} A_{12}^{-1} \begin{vmatrix} A_{11}^{-1} & A_{12}^{-1} \\ A_{21}^{-1} & A_{22}^{-1} \end{vmatrix}
\]
\[
F_{22}^{-1} A_{22}^{-1} A_{21}^{-1} \begin{vmatrix} A_{11}^{-1} & A_{12}^{-1} \\ A_{21}^{-1} & A_{22}^{-1} \end{vmatrix}
\]
provided that both \( A_{11} \) and \( A_{22} \) are non-singular.

An important application of inverse of partitioned matrix; Suppose we partition a matrix
\[
X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}
\]
\[
X^9X = \begin{pmatrix} X_1^9X_1 & X_1^9X_2 \\ X_2^9X_1 & X_2^9X_2 \end{pmatrix}
\]
The upper left block \((I \leq I)\) matrix can be expressed as

\[
\begin{align*}
(X^0X)^i1 &= X^0X_1i X^0X_2(X^0X_2)^i1 X^0X_1 \quad \# \\
&= X^0X_1i X^0X_2(X^0X_2)^i1 X^0X_1 \quad \# \\
&= [X^0M_2X_1]^i1
\end{align*}
\]

where \(M_2 = I \leq X_2 (X^0X_2)^i1 X^0_2 \): The matrix \(M_2\) plays quite an important and unique role in multiple regression model.

\(^2\) \(M_2\) is symmetric and idempotent.

\[
\begin{align*}
M_2M^0_2 &= I \leq X_2 (X^0X_2)^i1 X^0_2 \quad I \leq X_2 (X^0X_2)^i1 X^0_2 \\
&= I \leq X_2 (X^0X_2)^i1 X^0_2 \quad I \leq X_2 (X^0X_2)^i1 X^0_2 + X_2 (X^0X_2)^i1 X^0_2 (X^0X_2)^i1 X^0_2 \\
&= I \leq X_2 (X^0X_2)^i1 X^0_2 = M_2
\end{align*}
\]

\(^2\) \(M_2X_2 = 0\):

### 1.2 Eigenvectors and eigenvalues of real matrix

Suppose the solution of the following system of equations;

\[
A\times = \lambda \times
\]

where \(A\) is an \((n \leq n)\) square matrix, \(\times\) is a non null \((n \leq 1)\) vector and \(\lambda\) is a scalar. The \(\lambda\)'s satisfying the system of equations are called eigenvalues(characteristic values, latent values) and \(\times\)'s eigenvectors(characteristic vectors, latent vectors). We can rewrite the system as

\[
(A - \lambda I)\times = 0
\]

If the matrix \((A - \lambda I)\) is non-singular - i.e. its inverse exists, the only solution is the trivial solution, \(\times = 0\): In order for a non-trivial solution to exist, we should have a singular matrix \((A - \lambda I)\); which implies that

\[
jA - \lambda I = 0
\]

We can find the eigenvalues of a matrix \(A\) by expanding above determinant and solving the \(n\)th order polynomial equation.

**Example:**

\[
A = \begin{pmatrix}
4 & 2 \\
2 & 1
\end{pmatrix}
\]

Then,

\[
A - \lambda I = \begin{pmatrix}
4 - \lambda & 2 \\
2 & 1 - \lambda
\end{pmatrix}
\]

\(2\)
Therefore, 
\[ jA_j \cdot 1_j = (4_j \cdot ) (1_1 \cdot )_j \cdot 4 = 0 \]

The solutions for the equation is given by; 

\[ \lambda_1 = 5 \text{ and } \lambda_2 = 0 \]

For \( \lambda_1 = 5; \)

\[
\begin{pmatrix}
1 \\
2
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix} +
\begin{pmatrix}
2 \\
4
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix} = 0
\]

\[ x_1 = 2x_2 \]

One element in the eigenvector is always arbitrary. The usual practice is to normalize the vector so that we have unit length for the eigenvectors, i.e., \( x_1^2 + x_2^2 = 1 \):

Then, the eigenvector corresponding to \( \lambda_1 = 5 \) is

\[
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}
\]

For \( \lambda_2 = 0; \) it is given by

\[
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}
\]

Note that the matrix \( X \) whose columns consist of eigenvectors of a matrix is an orthogonal matrix - columns of the matrix are orthogonal;

\[
X = \begin{pmatrix}
\frac{2}{5} \\
\frac{1}{5}
\end{pmatrix}
\]

Moreover,

\[ X^0X = XX^0 = I \]

and

\[ X \cdot 1 = X^0 \]

Suppose \( X \) is \((n \times n)\) real matrix.

\[ ^2 \text{ The sum of the eigenvalues of } X \text{ is equal to the sum of its diagonal elements (trace).} \]

\[ ^2 \text{ The product of the eigenvalues of } X \text{ is equal to its determinants.} \]

\[ ^2 \text{ The rank of } X \text{ is equal to the number of its non-zero eigenvalues.} \]

\[ ^2 \text{ The eigenvalues of } X^1 \text{ are the reciprocals of those of } X; \text{ but eigenvectors are the same.} \]

\[ ^2 \text{ Each eigenvalue of an idempotent matrix is either } 0 \text{ or } 1; \]

\[ ^2 \text{ The rank of an idempotent matrix is equal to its trace.} \]

Now, we assume that \( X \) is \((n \times n)\) symmetric matrix as well as real.

\[ ^2 \text{ The eigenvalues of } X \text{ are real.} \]

\[ ^2 \text{ Eigenvectors corresponding to distinct eigenvalues are pairwise orthogonal.} \]

\[ ^2 \text{ The orthogonal matrix of eigenvectors diagonalizes } X, \text{ i.e., } X = I^0 \text{ where } I \]

is \((n \times n)\) orthogonal matrix whose columns consist of eigenvectors of \( X \) and \( I \)

is \((n \times n)\) diagonal matrix whose main diagonals consist of eigenvalues of \( X; \)

\[ ^2 \text{ Any symmetric positive definite matrix } X \text{ can be factored into } LL^0 \text{ where } L \text{ is a lower} \]
triangular matrix. It is called the Cholesky decomposition.

### 1.3 Rank of a matrix

The maximum number of linearly independent rows is equal to the maximum number of linearly independent columns. This number is the rank of the matrix, denoted by \( \text{rank}(X) \):

\[
\frac{1}{\text{rank}(X)} \cdot \min(m, n)
\]

where \( m \) and \( n \) are row and column dimensions of a matrix \( X \):

\[
\frac{1}{\text{rank}(X)} = \frac{1}{\text{rank}(X)^0}
\]

If \( \text{rank}(X) = m = n \); \( X \) is non-singular and a unique inverse \( X^{-1} \) exists.

\[
\frac{1}{\text{rank}(X)} = \frac{1}{\text{rank}(X) X^0} = \frac{1}{\text{rank}(X)}
\]

If \( P \) and \( Q \) are non-singular matrices of orders \( m \) and \( n \); then \( \frac{1}{\text{rank}(P X)} = \frac{1}{\text{rank}(Q X)} = \frac{1}{\text{rank}(X)} \):

\[
\frac{1}{\text{rank}(X)} \cdot \min(\frac{1}{\text{rank}(X)} ; \frac{1}{\text{rank}(Y)})
\]

where \( Y \) is \((n \leq l)\) matrix.

### 1.4 Kronecker products

The Kronecker product is defined as

\[
A \odot B = \\
\begin{bmatrix}
    a_{11}B & a_{12}B & \cdots & a_{1k}B \\
    a_{21}B & a_{22}B & \cdots & a_{2k}B \\
    \cdots & \cdots & \cdots & \cdots \\
    a_{n1}B & a_{n2}B & \cdots & a_{nk}B
\end{bmatrix}
\]

\[
(A \odot B)^{-1} = A^{-1} \odot B^{-1}
\]

\[
(A \odot B)^0 = A^0 \odot B^0
\]

If \( A \) is \((l \leq l)\) and \( B \) is \((n \leq n)\); then \( jA \odot B = jA^j \odot B^j \):

\[
\text{tr}(A \odot B) = \text{tr}(A) \text{tr}(B)
\]

\[
(A \odot B) (C \odot D) = AC \odot BD
\]

\[
A \odot (B + C) = A \odot B + A \odot C
\]

\[
A \odot (B - C) = (A \odot B) - (A \odot C)
\]

### 1.5 vec and vech operators

Suppose that \( X \) is \((m \leq n)\) matrix expressed as 

\[
X = \begin{bmatrix}
    X_1 & X_2 & \cdots & X_n
\end{bmatrix}
\]

where \( X_i \) is the \( i \)th column of the matrix \( X \): Then

\[
\text{vec}(X) = \\
\begin{bmatrix}
    X_1 \\
    X_2 \\
    \cdots \\
    X_n
\end{bmatrix}
\]

\[
\text{vec}(AB) = (C^0 - A) \text{vec}(B)
\]

Suppose that \( X \) is \((n \leq n)\) matrix. \text{vec} operator transforms an \((n \leq n)\) matrix into
\[
\begin{array}{cccc}
2 & a_{11} & a_{12} & a_{13} \\
2 & a_{21} & a_{22} & a_{23} \\
2 & a_{31} & a_{32} & a_{33} \\
\end{array}
\]

\[\text{v} = \text{vec} \begin{pmatrix} a_{11} \\
a_{21} \\
a_{31} \\
a_{12} \\
a_{22} \\
a_{32} \\
a_{13} \\
a_{23} \\
a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} \\
a_{21} \\
a_{31} \\
a_{12} \\
a_{22} \\
a_{32} \\
a_{13} \\
a_{23} \\
a_{33} \end{pmatrix}
\]

1.6 Matrix Differentiation

Let \( a \) and \( x \) be \((k \times 1)\) vectors and \( A \) is an \((k \times k)\) matrix.

\[
\begin{align*}
\frac{\partial (a^0x)}{\partial a} &= a \\
\frac{\partial (a^0x)}{\partial x^T} &= a^0 \\
\frac{\partial (x^0Ax)}{\partial a} &= (A + A^0)x \\
\frac{\partial (x^0Ax)}{\partial x} &= (A + A^0) \\
\frac{\partial (x^0Ax)}{\partial A} &= xx^0
\end{align*}
\]

We don’t want to prove the claim rigorously. But

\[
a^0x = \sum_{i=1}^{k} a_i x_i
\]

If you want to differentiate the function with respect to \( x \), you have to differentiate the function with respect to each element of vector \( x \) and form a vector -called gradient- with the result.

\[
\begin{align*}
\frac{\partial (a^0x)}{\partial x} &= \begin{pmatrix} \frac{\partial (a^0x)}{\partial x_1} \\
\frac{\partial (a^0x)}{\partial x_2} \\
\frac{\partial (a^0x)}{\partial x_3} \\
\end{pmatrix} \\
&= \begin{pmatrix} a_1 \\
a_2 \\
a_3 \\
\end{pmatrix}
\end{align*}
\]

You can understand \( \frac{\partial (a^0x)}{\partial x} \) simply as the transpose of \( \frac{\partial (a^0x)}{\partial a} \). For the differentiation of the quadratic form, consider the summation expression:

\[
x^0Ax = \sum_{i=1}^{k} \sum_{j=1}^{k} x_i a_{ij} x_j
\]

\[
x_1^0a_{11}x_1 + x_1 a_{12} x_2 + x_1 a_{13} x_3 + \cdots + x_1 a_{1k} x_k \\
+ x_2 a_{21} x_1 + x_2 a_{22} x_2 + x_2 a_{23} x_3 + \cdots + x_2 a_{2k} x_k \\
+ x_3 a_{31} x_1 + x_3 a_{32} x_2 + x_3 a_{33} x_3 + \cdots + x_3 a_{3k} x_k \\
+ \cdots
\]

\[
+ x_k a_{k1} x_1 + x_k a_{k2} x_2 + x_k a_{k3} x_3 + \cdots + x_k a_{kk} x_k
\]

5
Now, we have
\[
\frac{\partial (x^T A x)}{\partial x_i} = 2a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \cdots + a_{ik}x_k
\]
+ x_2a_{i1} + x_3a_{i3} + \cdots + x_ka_{ik}
\]
= \frac{\partial}{\partial x_i} \left( x_1a_{i1} + x_2a_{i2} + x_3a_{i3} + \cdots + x_ka_{ik} \right)
\]
+ a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \cdots + a_{ik}x_k
\]
= A_1x + A_1^T x
\]
where \(A_1\) is the first row of the matrix \(A\) and \(A^1\) is the first column of the matrix \(A\).

Similarly,
\[
\frac{\partial (x^T A x)}{\partial x_2} = a_{21}x_1 + 2a_{22}x_2 + a_{23}x_3 + \cdots + a_{2k}x_k
\]
+ x_1a_{21} + x_3a_{23} + \cdots + x_ka_{2k}
\]
= \frac{\partial}{\partial x_2} \left( x_1a_{21} + x_2a_{22} + x_3a_{23} + \cdots + x_ka_{2k} \right)
\]
+ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2k}x_k
\]
= A_2x + A_2^T x
\]
You see the pattern emerging from the calculation. In general,
\[
\frac{\partial (x^T A x)}{\partial x_i} = A_i + A_i^T x \quad i = 1; 2; \ldots; k
\]
We stack the vectors to get:
\[
\frac{\partial (x^T A x)}{\partial x} = \begin{bmatrix}
\frac{\partial (x^T A x)}{\partial x_1} \\
\frac{\partial (x^T A x)}{\partial x_2} \\
\vdots \\
\frac{\partial (x^T A x)}{\partial x_k}
\end{bmatrix} = A + A^T x
\]
You can verify the result for \(\frac{\partial (x^T A x)}{\partial x} = xx^T\) with a similar argument.

Consider the following minimization problem:
\[
\min_b S(b) = \min_b \left( y_1^T x b + y_2^T x b \right)
\]
\[
= \min_b y_1^T y_1 b x b + y_2^T y_2 b x b
\]
where \(y\) is \((n \times 1)\) vector, \(X\) is \((n \times K)\) matrix whose rank is \(K\) and \(b\) is \((K \times 1)\) vector. Note that \(y_1^T X\) is a \(0\) vector, \(b\) is \(x\) vector and \(X^T x\) is \(A\) matrix in the formula above. Hence,
\[
S(b) = y_1^T y_1 b x b + y_2^T y_2 b x b
\]
\[
= \min_b y_1^T y_1 b x b + y_2^T y_2 b x b
\]
\[
= \min_b y_1^T y_1 b x b + y_2^T y_2 b x b
\]
\[
= \min_b y_1^T y_1 b x b + y_2^T y_2 b x b
\]
\[
= \min_b y_1^T y_1 b x b + y_2^T y_2 b x b
\]
\[
= \min_b y_1^T y_1 b x b + y_2^T y_2 b x b
\]
\[
= \min_b y_1^T y_1 b x b + y_2^T y_2 b x b
\]
\[
= \min_b y_1^T y_1 b x b + y_2^T y_2 b x b
\]
Therefore, we can find the solution to the minimization problem as:
\[ b^* = (X^0X)^{1/2}X^0y \]
You can check the second order condition for the minimum.

2. Some distribution theory

2.1 Multivariate normal distribution

A random vector \( X \) with values in \( \mathbb{R}^p \) is multivariate normal if every linear combination of its components \( \sum_{i=1}^{p} \alpha_i X_i \) follows a normal distribution on \( \mathbb{R} \):

Every multivariate normal random vector has a finite mean \( \mathbb{E}(X) = \mu \) \( \in \mathbb{R}^p \) and a finite covariance matrix \( \text{Var}(X) = \Sigma \) : We denote the random vector as \( X \sim N(\mu, \Sigma) \).

If \( X \sim N(\mu, \Sigma) \), then \( (\alpha X + \beta) \sim N(\alpha \mu + \beta \Sigma \alpha, \alpha \Sigma \alpha) \) where \( \alpha \) is an \( (1 \times p) \) matrix and \( \beta \) is an \( (1 \times 1) \) vector. Both of them are non-stochastic.

If \( \Sigma \) is non-singular, then the density of \( X \) whose distribution is a multivariate normal is
\[
f(x) = \frac{1}{(2\pi)^{p/2} \det \Sigma} \exp \left( -\frac{1}{2} \sum_{i=1}^{p} (x_i - \mu_i)^2 \right)
\]

Suppose we can partition \( X \) into \( X_1 \) and \( X_2 \) so that
\[
X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)
\]

Then, the marginal distributions of \( X_1 \) and \( X_2 \) are
\[
X_1 \sim N(\mu_1, \Sigma_{11}) \quad \text{and} \quad X_2 \sim N(\mu_2, \Sigma_{22})
\]

The conditional distribution are given by
\[
X_1 \mid X_2 \sim N \left( \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (X_2 - \mu_2), \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right)
\]
\[
X_2 \mid X_1 \sim N \left( \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (X_1 - \mu_1), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right)
\]

2.2 Censored and truncated normal distribution

A censored normal random variable \( X^\gamma \) is defined as \( \gamma \)
\[
X = \begin{cases} X^\gamma & \text{if } X^\gamma > 0 \\ 0 & \text{otherwise} \end{cases}
\]

where \( X^\gamma \sim N(\mu, \Sigma) \): The censoring point is 0 here. But it can be any point on \( \mathbb{R} \):

The density for \( X \) is given by
\[
f(x) = \frac{\exp \left( -\frac{1}{2} \left( x - \mu \right)^T \Sigma^{-1} \left( x - \mu \right) \right)}{\int_{-\infty}^{\gamma} \exp \left( -\frac{1}{2} \left( x - \mu \right)^T \Sigma^{-1} \left( x - \mu \right) \right) dx} \quad \text{and} \quad \frac{\exp \left( -\frac{1}{2} \left( x - \mu \right)^T \Sigma^{-1} \left( x - \mu \right) \right)}{\int_{0}^{\infty} \exp \left( -\frac{1}{2} \left( x - \mu \right)^T \Sigma^{-1} \left( x - \mu \right) \right) dx}
\]
2.3 Distribution derived from the normal distribution

2.3.1 $\bar A^2$ distribution

$\bar A^2$ distribution with $n$ degrees of freedom is the probability distribution of a random variable $Y = X_1 + X_2 + \cdots + X_n$; where the random variables $X_i$ are independently distributed with respective distributions $N(1;1)$: When $X_i = 0$ for all $i$, the chi-square distribution is said to be central, which usually called chi-square distribution without the adjective central and denoted as $\bar A^2(n)$. It is non-central otherwise, which is called non-central chi-square distribution with non-centrality parameter $\lambda = \sum_{i=1}^{n} \lambda_i^2$ and denoted as $\bar A^2(n;\lambda)$. 

\[
\frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A} = \frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A} = \frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A} = \frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A}
\]

where $d = 1$ when $X > 0$ and $d = 0$ when $X \leq 0$: $\bar A$ is the cumulative distribution function of standard normal and $\bar A$ is the density function of standard normal.

Suppose that $X$ is a censored normal random variable with censoring point of 0 and the distribution of the latent variable is given by $X_n \sim \text{N}(1; \frac{1}{2})$:

(i) $E(X) = \frac{3\bar A_i}{\bar A + \frac{1}{2}}$

(ii) $\text{Var}(X) = \frac{3\bar A_i}{\bar A + \frac{1}{2}}$

where $W(z) = \int_{0}^{\frac{z}{\sqrt{2}}} \frac{d(t)}{\sqrt{2}} = z \phi(z) + \Phi(z)$

A truncated normal random variable $X$ is defined as $X_n$, if $X_n > 0$

$X = \frac{X_n}{\sqrt{2}}$, not observed otherwise.

where $X_n \sim \text{N}(1; \frac{1}{2})$: The truncation point is 0 here. But it can be any point on $\mathbb{R}$:

\[
\frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A} = \frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A} = \frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A} = \frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A}
\]

if $X > 0 = 9$

\[
\frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A} = \frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A} = \frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A} = \frac{\left(\begin{array}{c}
\beta
\end{array}\right) d x}{\bar A}
\]

where $\phi(x) = \frac{A(x)}{\sqrt{2\pi}}$, called the Mill's ratio.
The density for a random variable $X$ whose distribution is $\hat{A}^2(n)$ is
\[ f(x) = \frac{2^{\frac{1}{2}}}{\pi^{\frac{1}{2}} n^{\frac{3}{2}}} \exp \left(-\frac{1}{2} \langle x \rangle^2 \right) I_{\{x > 0\}} \]

note that $\hat{A}^2(n)$ is a Gamma distribution with parameter $\frac{n}{2}$.

When $X \sim \hat{A}^2(n_1)$; $E(X) = n + \frac{1}{2}$, and $\text{Var}(X) = 2(n + 2)$:

Let $Y$ be distributed as multivariate normal, $Y \sim N(1; \Sigma)$ where $\Sigma$ is non-singular, and $\Sigma$ be a symmetric matrix. Then, $Y \sim \hat{A}^2(n)$ if and only if $\Sigma = \hat{A}(1)$; in which case the degrees of freedom is $\text{rank}(\Sigma)$ and the non-centrality parameter is $\frac{n}{2}$:

For example, if $X$ is an $n$-variate normal random variable, $X \sim N(1; \Sigma)$, then $X \hat{A}^2(n_1)$ and $\hat{A}^2(n)$:

We have chosen $\hat{A}^2(n)$:

### 2.3.2 (Student) $t$ distribution

The $t$ distribution with $n$ degrees of freedom is the probability distribution of the random variable $Z = \frac{X}{\sqrt{\hat{A}(n)}}$ where $X \sim N(1; 1)$ and $Y \sim \hat{A}^2(n)$. $X$ and $Y$ are independent each other. The parameter $1$ is called non-centrality parameter and the (central) $t$ distribution corresponds to $\Sigma = 0$ and denoted as $t(n)$:

The density of a random variable $Z$ whose distribution is $t(n)$ is given by

\[ f(z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n} \Gamma(\frac{n}{2})} \left(1 + \frac{z^2}{n}\right)^{-\frac{n+1}{2}} \]

If $Z \sim t(n)$; $E(Z^p) = 0$ when $p$ is odd and $p < n$; $E(Z^p) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\left(\frac{n}{2}\right)}$ when $p$ is even and $p < n$; When $p = n$; $E(Z^p)$ does not exist.

Why do we need $\hat{A}^2$ and $t$? Here is a good example. Suppose $X_{ij}$s are i.i.d. with the distribution $X_i \sim N(1; 1, \Sigma)$. It is well-known that $\frac{n}{\Sigma} \sum_{i=1}^n X_i = X \sim N(1; \Sigma)$:

Moreover, $\frac{(n+1)s^2}{\Sigma} \sim \hat{A}^2(n_1)$ where $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - X_{..})^2$:

In addition, we can show that $X$ and $s^2$ are independent. Hence,

\[ \frac{n}{\Sigma} \sum_{i=1}^n X_{ij} \sim \hat{A}^2(n_1) \]

### 2.4 $F$ distribution

The $F$ distribution with $n_1$ and $n_2$ degrees of freedom is the probability distribution of the random variable $W = \frac{Y_1}{Y_2}$ where $Y_1 \sim \hat{A}^2(n_1)$ and $Y_2 \sim \hat{A}^2(n_2)$:

And, $Y_1$ and $Y_2$ are independent. The distribution is denoted as $F(n_1; n_2)$:
The density for a random variable $W$ whose distribution is $F(n_1; n_2)$ can be written as

$$f(w) = (n_1)^{n_2} (n_2)^{n_1} \frac{w^{n_1} i^{n_1} (n_1 + n_2)^{n_1 + i} i^{n_1 + n_2}}{n_1! n_2!}$$

If $W \sim F(n_1; n_2)$, $E(W^p) = \frac{n_2}{n_1} \frac{p \left( \frac{1}{1+p} \right) \left( \frac{1}{1+p} \right)}{i(1+p) i(1+p)}$.

We can define a non-central $F$ distribution with two independent non-central chi-square distributions.

If $Z \sim t(n); Z^2 = \frac{\chi^2_m}{\chi^2_{n}} = \frac{X^2 = 1}{Y}; Y \sim \chi^2(n)$ since $X^2 \sim \chi^2(1); Y \sim \chi^2(n)$ and $X$ and $Y$ are independent by definition.