6.1 Multivariable Normal Distribution

**Definition 6.1.** Let \( X \sim N(\mu, \Sigma) \) be a \( p \times 1 \) random vector, where \( \mu \) is a \( p \times 1 \) vector and \( \Sigma \) is a positive definite matrix, with probability density function:

\[
p(X) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (X - \mu)^T \Sigma^{-1} (X - \mu)\right)
\]

where \( \mathbb{E}(X) = \mu \), \( \text{Cov}(X) = \Sigma \).

Take \( X \) into two parts, i.e. \( X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}, \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \)

where \( X^{(1)} \) is \( q \times 1 \) and \( X^{(2)} \) is \( (p-q) \times 1 \), so as \( \mu^{(1)}, \mu^{(2)} \)

Let’s define \( X_{2.1} = X^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} X^{(1)} \)

**Theorem 6.1.** If \( X \sim N_p(\mu, \Sigma) \) then

1. \( X^{(1)} \sim N_q(\mu^{(1)}, \Sigma_{11}), \quad X^{(2)} \sim N_{p-q}(\mu^{(2)}, \Sigma_{22}) \)
2. \( X^{(1)} \) and \( X_{2.1} \) are independent
3. \( X^{(2)}|X^{(1)} \sim N_{p-q}(\mu^{(2)} + \Sigma_{21} \Sigma_{11}^{-1} (x^{(1)} - \mu^{(1)}), \Sigma_{22.1}) \)

where

\[
\mu_{2.1} = \mu^{(2)} - \Sigma_{21} \Sigma_{11}^{-1} \mu^{(1)}
\]

and

\[
\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}
\]

is the Schur Complement of \( \Sigma_{11} \).

**Remarks:** The Jacobian of transform \((X^{(1)}, X^{(2)}) \rightarrow (X^{(1)}, X_{2.1})\) is 1.

**Proof.**

\[
Z = \begin{pmatrix} X^{(1)} \\ X_{2.1} \end{pmatrix} = \begin{pmatrix} I & 0 \\ -\Sigma_{21} \Sigma_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}
\]

That makes \( dZ = \det \begin{pmatrix} I & 0 \\ -\Sigma_{21} \Sigma_{11}^{-1} & I \end{pmatrix} dX \)

obviously, \( \det \begin{pmatrix} I & 0 \\ -\Sigma_{21} \Sigma_{11}^{-1} & I \end{pmatrix} = 1. \)
Let $B = \begin{pmatrix} I & 0 \\ -\Sigma_{21}^{-1} & I \end{pmatrix}$, then $X = B^{-1}Z$.

The above derivations are also established if replacing $X$ with $X - \mu$. Hence, we have $X - \mu = B^{-1}Z$. Since the Jacobian from $X$ to $Z$ is 1, we can derive the p.d.f of $Z$ easily (just ignoring the constants):

$$(X - \mu)^T \Sigma^{-1} (X - \mu) = Z^T (B^{-1})^T \Sigma^{-1} B^{-1} Z$$

$$= Z^T \left( -\Sigma_{21}^{-1} + \frac{1}{2} \Sigma_{11} \Sigma_{22}^{-1} \right) Z$$

$$= Z^T \left( \Sigma^{-1}_{11} \Sigma_{22}^{-1} \right) Z$$

So $Z$ forms a Gaussian Distribution with variance matrix $\begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix}$.

Since the covariance is 0, so $X^{(1)}$ and $X_{2.1}$ are independent. Now we have proved the proposition 1 and 2 in theorem 1.

Let consider the constant part $|\Sigma|^{\frac{1}{2}}$ to confirm our conclusion.

$$|B \Sigma B^T| = \left| \begin{array}{cc} \Sigma_{11} & 0 \\ 0 & \Sigma_{22}^{-1} \end{array} \right| = |\Sigma_{11}| |\Sigma_{22}^{-1}| = |B|^2 |\Sigma|$$

$$\Rightarrow |\Sigma| = |\Sigma_{11}| |\Sigma_{22}^{-1}|$$

$$\Rightarrow |\Sigma|^{\frac{1}{2}} = |\Sigma_{11}|^{\frac{1}{2}} |\Sigma_{22}^{-1}|^{\frac{1}{2}}$$

So, the p.d.f of $Z$ is

$$p(Z) = \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma_{11}|^{\frac{1}{2}} |\Sigma_{22}^{-1}|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} Z^{(1)^T} \Sigma_{11}^{-1} Z^{(1)} \right) \exp \left( -\frac{1}{2} Z^{(2)^T} \Sigma_{22}^{-1} Z^{(2)} \right)$$

**Corollary 6.1.** $\Sigma$ is positive definite $\iff$ $\Sigma_{11}, \Sigma_{22}^{-1}$ is positive definite.

Now let’s prove the proposition 3 in theorem 1. Since $X^{(1)}$ is a constant in conditional probability, we have

$$X^{(2)} = X_{2.1} + \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}$$

$$\Rightarrow \mathbb{E}(X^{(2)} | X^{(1)}) = \mu_{2.1} + \Sigma_{21} \Sigma_{11}^{-1} X^{(1)}$$

$$\Rightarrow \text{Cov}(X^{(2)} | X^{(1)}) = \Sigma_{22.1}$$

That’s all of the proving of theorem 6.1.

**Theorem 6.2.** If $C = \Sigma^{-1}$, i.e. $C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1}$, then

1. $C_{22}^{-1} = \Sigma_{22.1}$
2. $C_{11}^{-1}C_{12} = -\Sigma_{12}\Sigma_{22}^{-1}$

Proof.

\[
\begin{pmatrix}
I & 0 \\
-\Sigma_{21}\Sigma_{11}^{-1} & I
\end{pmatrix}
\begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix}
\begin{pmatrix}
I & -\Sigma_{11}^{-1}\Sigma_{12} \\
0 & I
\end{pmatrix}
= 
\begin{pmatrix}
\Sigma_{11} & 0 \\
0 & \Sigma_{22,1}
\end{pmatrix}
\]

\[
\begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix} = 
\begin{pmatrix}
I & 0 \\
-\Sigma_{11}^{-1}\Sigma_{12} & I
\end{pmatrix}
\begin{pmatrix}
\Sigma_{11} & 0 \\
0 & \Sigma_{22,1}
\end{pmatrix}
\begin{pmatrix}
I & 0 \\
0 & I
\end{pmatrix}
\]

\[
\begin{pmatrix}
C_{11} & C_{12} \\
C_{21} & C_{22}
\end{pmatrix} = 
\begin{pmatrix}
I & -\Sigma_{11}^{-1}\Sigma_{12} \\
0 & I
\end{pmatrix}
\begin{pmatrix}
\Sigma_{11}^{-1} & 0 \\
0 & \Sigma_{22,1}^{-1}
\end{pmatrix}
\begin{pmatrix}
I & 0 \\
-\Sigma_{21}\Sigma_{11}^{-1} & I
\end{pmatrix}
\]

\[
\begin{pmatrix}
C_{11} & C_{12} \\
C_{21} & C_{22}
\end{pmatrix} = 
\begin{pmatrix}
\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1} \\
-\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & \Sigma_{22,1}^{-1}
\end{pmatrix}
\]

So, $C_{22}^{-1} = \Sigma_{22,1}$. And

\[
C_{11}^{-1}C_{12} = (\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1})(-\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1})
\]

\[
= -\Sigma_{11}\Sigma_{22}^{-1} - \Sigma_{12}\Sigma_{22,1}^{-1}
\]

\[
= -\Sigma_{11}\Sigma_{22}^{-1} - \Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}
\]

\[
= -\Sigma_{12}\Sigma_{22}^{-1}
\]

\[
\square
\]

6.2 Matrix Variate Distribution

Let $X = (X_1, X_2, \ldots, X_n)^T$, $X_i \in \mathbb{R}^p$ and $X_i \sim N(\mu_i, \Sigma)$. If

\[
p(X) = \prod_{i=1}^{n} p(X_i)
\]

\[
= \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{n}{2}}} \exp\left(\frac{1}{2} (X_i - \mu_i)^T \Sigma^{-1} (X_i - \mu_i)\right)
\]

\[
= \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1} \sum_{i=1}^{n} (X_i - \mu_i)(X_i - \mu_i)^T)\right)
\]

Suppose $\mu = (\mu_1, \mu_2, \ldots, \mu_n)^T$, then

\[
p(X) = \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1}(X - \mu)^T I(X - \mu))\right)
\]

\[
= \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \text{etr}(-\frac{1}{2} \Sigma^{-1}(X - \mu)^T I(X - \mu))\right)
\]

We call $X$ is Matrix-variate normal distributed.

Homework 1 If $\text{vec}(X^T) \sim N_{np}(\text{vec}(\mu^T), BA \otimes A)$, show the p.d.f of $X$ is

\[
\frac{1}{(2\pi)^{\frac{np}{2}}|A|^{\frac{n}{2}}|B|^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \text{tr}(A^{-1}(X - \mu)^T B^{-1}(X - \mu))\right)
\]
Definition 6.2 (Wishart Distribution). If \( S = X^T X \), where the \( n \times p \) matrix \( X \) is \( N(0, I_n \otimes \Sigma) \), then \( S \) is positive definite and is said to have the Wishart distribution with \( n \) degrees of freedom and covariance matrix \( \Sigma \). We will write that \( S \) is \( W_p(\Sigma, n) \), the subscript on \( W \) denoting the size of the matrix \( S \).

Theorem 6.3. If \( S \) is \( W_p(\Sigma, r) \) with \( r \geq p \) then the density function of \( S \) is

\[
p(S) = \frac{|S|^{\frac{r-p-1}{2}} \exp(-\frac{1}{2} \text{tr}(\Sigma^{-1} S))}{2^{\frac{p(p+1)}{4}} \pi^{\frac{p(p-1)}{4}} |\Sigma|^{\frac{r}{2}} \prod_{i=1}^{p} \Gamma(r+i-1)}, r \geq p
\]

In Bayesian statistics, in the context of the multivariate normal distribution, the Wishart distribution is the conjugate prior to the precision matrix \( \Omega = \Sigma^{-1} \), where \( \Sigma \) is the covariance matrix.

Splitting \( S \) into parts of \( q \) and \( p-q \), i.e. \( S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \)

where \( S_{11} \) is \( q \times q \) and \( S_{22} \) is \( (p-q) \times (p-q) \). So as \( \Sigma \).

Theorem 6.4. Let \( S \sim W_p(\Sigma, r) \), \( S_{11.2} = S_{11} - S_{12} S_{22}^{-1} S_{21}, \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \) then

1. \( S_{11} \sim W_q(\Sigma_{11}, r), S_{22} \sim W_{p-q}(\Sigma_{22}, r) \)
2. \( S_{11.2} \sim W_q(\Sigma_{11.2}, r-(p-q)) \)
3. \( S_{11.2} \) and \( (S_{12}, S_{22}) \) are independent.
4. \( S_{12}|S_{22} \sim N_{q,p-q}(\Sigma_{12} \Sigma_{22}^{-1} S_{22}, \Sigma_{11.2} \otimes S_{22}) \)

Proof. Making the transformation

\[
\begin{cases}
S_{11.2} = S_{11} - S_{12} S_{22}^{-1} S_{21} \\
B_{12} = S_{12} \\
B_{22} = S_{22}
\end{cases}
\]

i.e. \((S_{11}, S_{12}, S_{22}) \rightarrow (S_{11.2}, B_{12}, B_{22})\). Since

\[
(d(S_{11.2}, B_{12}, B_{22})) = (d(S_{11.2}, S_{12}, S_{22}))
= (d(S_{11} - S_{12} S_{22}^{-1} S_{21}, B_{12}, B_{22}))
= (d(S_{11}, B_{12}, B_{22}))
= (d(S_{11}, S_{12}, S_{22}))
\]

So the Jacobian is 1. Hence, we can replace \( S \) with \( S_{11.2}, B_{12}, B_{22} \).
First, we have $|S| = |S_{11,2}| |S_{22}| = |S_{11,2}| |B_{22}|$. Second, in the trace part, we have

$$tr(\Sigma^{-1} S) = tr \left( \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \right)$$

$$= tr(C_{11}S_{11}) + 2tr(C_{21}S_{12}) + tr(C_{22}S_{22})$$

$$= tr(C_{11}S_{11}) + 2tr(C_{12}B_{21}) + tr(C_{22}B_{22})$$

(since $S_{11} = S_{11,2} + S_{12} S_{22}^{-1} S_{21}$)

$$= tr(C_{11}S_{11,2}) + tr(C_{11}B_{12}B_{22}^{-1}B_{21}) + 2tr(C_{12}B_{21}) + tr(C_{22}B_{22})$$

(using theorem 6.2)

$$= tr(\Sigma_{11,2}^{-1} S_{11,2}) + tr(C_{11}B_{12}B_{22}^{-1}B_{21}) + 2tr(C_{12}B_{21}) + tr(\Sigma_{22}^{-1} B_{22})$$

$$+ tr((C_{21} C_{11}^{-1} C_{12}) B_{22})$$

We can see that $tr(\Sigma_{11,2}^{-1} S_{11,2})$ is corresponding to $p(S_{11,2})$, $tr(\Sigma_{22}^{-1} B_{22})$ is corresponding to $p(B_{22})$. And to prove $S_{11,2}$ and $(B_{12}, B_{22})$ are independent, we should have

$$p(S_{11,2}, B_{12}, B_{22}) = p(S_{11,2})p(B_{12}, B_{22}) = p(S_{11,2})p(B_{12}|B_{22})p(B_{22})$$

So, the residue terms should be corresponding to $p(B_{12}, B_{22})$, which is the 4th proposition in theorem 6.4. Now we rewritten them to show that they are corresponding to $N_{q,p-q}(\Sigma_{12} \Sigma_{22}^{-1} S_{22}, \Sigma_{11,2} \otimes S_{22})$.

$$tr(C_{21} C_{11}^{-1} C_{12} B_{22}) + tr(C_{11} B_{12} B_{22}^{-1} B_{21}) + 2tr(C_{12} B_{21})$$

$$= tr(C_{11} (B_{12} + C_{11}^{-1} C_{12} B_{22}) B_{22}^{-1} (B_{12} + C_{11}^{-1} C_{12} B_{22})^T)$$

$$= tr(\Sigma_{11,2} (B_{12} - \Sigma_{12} \Sigma_{22}^{-1} B_{22}) B_{22}^{-1} (B_{12} - \Sigma_{12} \Sigma_{22}^{-1} B_{22})^T)$$

Finally, we have $|\Sigma| = |\Sigma_{11,2}||\Sigma_{22}|$.

Now we have proved that the p.d.f of $S$ can be decomposed into terms $S_{22}$, $S_{11,2}$ and $S_{12}|S_{22}$.

The following theorem is used to solve the problem: how to sample from a Wishart distribution.

**Theorem 6.5.** Let $S \sim W_p(I_p, r)$ and $S = T^T T$ where $T = (t_{i,j})$ is an upper triangle matrix, $t_{i,i} > 0$ then

1. $t_{i,j}$ $1 \leq j \leq i \leq p$ are independently distributed.
2. $t_{i,i}^2 \sim \chi^2_{r-i+1}$
3. $t_{i,j} \sim N(0, 1)$ $1 \leq j < i \leq p$

**Proof.** First, we have

$$|S|^\frac{1}{2}(r-p-1)etr(-\frac{1}{2}S) = \left( \prod_{i=1}^{p} t_{ii}^{2(r-p-1)} \right)etr(-\frac{1}{2} \sum_{1 \leq j < i \leq p} t_{i,j}^2)$$

6 - 5
According to Theorem 5.6, we have $J(S \rightarrow T) = 2^p \prod_{i=1}^{p} \pi^{p-i+1}_{ii}$. Also, we have $\text{tr}(S) = \text{tr}(T^T T)$. Thus,

$$p(T) \propto \prod_{1 \leq j < i \leq p} \exp(-\frac{1}{2}t_{ij}^2) \prod_{i=1}^{p} (t_{ii})^{\frac{r-1}{2}} |J(S \rightarrow T)|$$

$$\propto \prod_{1 \leq j < i \leq p} \exp(-\frac{1}{2}t_{ij}^2) \prod_{i=1}^{p} (t_{ii})^{\frac{r-1}{2}} \exp(-\frac{1}{2}t_{ii}^2)$$

$\prod_{1 \leq j < i \leq p} \exp(-\frac{1}{2}t_{ij}^2)$ denote the independent standard normal distributions of $t_{i,j}$. $\prod_{i=1}^{p} (t_{ii})^{\frac{r-1}{2}} \exp(-\frac{1}{2}t_{ii}^2)$ denote the independent distributions $\chi^2_{r-1}$. \hfill \Box

Wishart distribution is a generalization to multiple dimensions of the chi-squared distribution. If $p = 1$ and $\Sigma = 1$ then this distribution is a chi-squared distribution with $r$ degrees of freedom.

**Definition 6.3.** $S^{-1}$ is said to have an inverse Wishart Distribution $W_{p}^{-1}(\Sigma, r)$ if its p.d.f. $(M = S^{-1})$

$$f(M) = \frac{|M|^{-\frac{r+p+2}{2}} \text{etr}(-\frac{1}{2}\Sigma^{-1}M^{-1})}{2^{\frac{p(p-1)}{4}} \pi^{\frac{p(p-1)}{4}} |\Sigma|^{\frac{r}{2}} \prod_{i=1}^{p} \Gamma(\frac{r+1-i}{2})} \tag{1}$$

**Theorem 6.6.** Let $A$ and $B$ be independent where $A \sim W_p(\Sigma, r_1)$, $B \sim W_p(\Sigma, r_2)$, with $r_1 \geq p$, $r_2 \geq p$. Put $A + B = T^T T$. $T$ is upper triangular. And $A = T^T U T$. Let $U$ be an $m \times m$ symmetric matrix. Then $0 < U < I$, and

1. $A + B$ and $U$ are independent
2. $A + B \sim W_p(\Sigma, r_1 + r_2)$
3. $p(u) \propto |U|^{r_1 - p - 1 + \frac{r_2 - p - 1}{2}} |I - U|^\frac{r_2 - p - 1}{2}$

$p(U)$ is called matrix-variate Beta Distribution.

**Homework 2** Prove theorem 6.6.