Lecture Notes 6: Wishart Distribution

Professor: Zhihua Zhang

6.1 Multivariable Normal Distribution

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

$$p(X) = \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{\Sigma}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(X-\mu)^T \mathbf{\Sigma}^{-1}(X-\mu))$$

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

Take X into two parts. i.e. $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$, $\quad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}$, $\quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$
where $X^{(1)}$ is $q \times 1$ and $X^{(2)}$ is $(p-q) \times 1$, so as $\mu^{(1)}, \mu^{(2)}$

Let's define $X_{2.1} = X^{(2)} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}X^{(1)}$

Theorem 6.1. *If $X \sim N_p(\mu, \mathbf{\Sigma})$ then*

1. $X^{(1)} \sim N_q(\mu^{(1)}, \mathbf{\Sigma}_{11})$, $\quad X_{(2)} \sim N_{p-q}(\mu^{(2)}, \mathbf{\Sigma}_{22})$
   $X_{2,1} \sim N_{p-q}(\mu_{2.1}, \mathbf{\Sigma}_{22.1})$
2. $X^{(1)}$ *and* $X_{2.1}$ *are indenpdent*
3. $X^{(2)}|X(1) \sim N_{p-q}(\mu^{(2)} + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(x^{(1)} - \mu^{(1)}), \mathbf{\Sigma}_{22.1})$

where

$$\mu_{2.1} = \mu^{(2)} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}\mu^{(1)}$$

and

$$\mathbf{\Sigma}_{22.1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$$

is the Schur Complement of $\mathbf{\Sigma}_{11}$.

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

Proof.

$$Z = \begin{pmatrix} X^{(1)} \\ X_{2.1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & 0 \\ -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$$

That makes $dZ = \det \begin{pmatrix} \boldsymbol{I} & 0 \\ -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1} & \boldsymbol{I} \end{pmatrix} dX$

obviously, $\det \begin{pmatrix} \boldsymbol{I} & 0 \\ -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1} & \boldsymbol{I} \end{pmatrix} = 1$. □

Let $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{I} & 0 \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{I} \end{pmatrix}$, then $X = \boldsymbol{B}^{-1}Z$.

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

$$\begin{aligned}
& (X-\mu)^T\boldsymbol{\Sigma}^{-1}(X-\mu) \\
= \;& Z^T(\boldsymbol{B}^{-1})^T\boldsymbol{\Sigma}^{-1}\boldsymbol{B}^{-1}Z \\
= \;& Z^T(\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T)^{-1}Z \\
= \;& Z^T\left[\begin{pmatrix} \boldsymbol{I} & 0 \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{I} \end{pmatrix}\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{I} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ 0 & \boldsymbol{I} \end{pmatrix}\right]^{-1} Z \\
= \;& Z^T\begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & 0 \\ 0 & \boldsymbol{\Sigma}_{22.1}^{-1} \end{pmatrix} Z
\end{aligned}$$

So Z forms a Gaussian Distribution with variance matrix $\begin{pmatrix} \boldsymbol{\Sigma}_{11} & 0 \\ 0 & \boldsymbol{\Sigma}_{22.1}^{-1} \end{pmatrix}$. 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 $|\boldsymbol{\Sigma}|^{\frac{1}{2}}$ to confirm our conclusion.

$$\begin{aligned}
& |\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T| = \begin{vmatrix} \boldsymbol{\Sigma}_{11} & 0 \\ 0 & \boldsymbol{\Sigma}_{22.1} \end{vmatrix} = |\boldsymbol{\Sigma}_{11}||\boldsymbol{\Sigma}_{22.1}| = |\boldsymbol{B}|^2|\boldsymbol{\Sigma}| \\
\Rightarrow \;& |\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{11}||\boldsymbol{\Sigma}_{22.1}| \\
\Rightarrow \;& |\boldsymbol{\Sigma}|^{\frac{1}{2}} = |\boldsymbol{\Sigma}_{11}|^{\frac{1}{2}}|\boldsymbol{\Sigma}_{22.1}|^{\frac{1}{2}}
\end{aligned}$$

So, the p.d.f of Z is

$$p(Z) = \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}_{11}|^{\frac{1}{2}}|\boldsymbol{\Sigma}_{22.1}|^{\frac{1}{2}}}\exp(-\frac{1}{2}Z^{(1)}\boldsymbol{\Sigma}_{11}^{-1}Z^{(1)})\exp(-\frac{1}{2}Z^{(2)}\boldsymbol{\Sigma}_{22.1}^{-1}Z^{(2)})$$

Corollary 6.1. $\boldsymbol{\Sigma}$ *is positive definite* $\iff$ $\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22.1}$ *is positive definite.*

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

$$\begin{aligned}
& X^{(2)} = X_{2.1} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}X^{(1)} \\
\Rightarrow \;& \mathbb{E}(X^{(2)}|X^{(1)}) = \mu_{2.1} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}X^{(1)} \\
\Rightarrow \;& Cov(X^{(2)}|X^{(1)}) = \boldsymbol{\Sigma}_{22.1}
\end{aligned}$$

That's all of the proving of theorem 6.1.

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

1. $\boldsymbol{C}_{22}^{-1} = \boldsymbol{\Sigma}_{22.1}$

2. $\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12} = -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$

Proof.

$$\begin{aligned}
&\begin{pmatrix} \boldsymbol{I} & 0 \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ 0 & \boldsymbol{I} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & 0 \\ 0 & \boldsymbol{\Sigma}_{22.1} \end{pmatrix} \\
\Longrightarrow\ & \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & 0 \\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & 0 \\ 0 & \boldsymbol{\Sigma}_{22.1} \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ 0 & \boldsymbol{I} \end{pmatrix} \\
\Longrightarrow\ & \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ 0 & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & 0 \\ 0 & \boldsymbol{\Sigma}_{22.1}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & 0 \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{I} \end{pmatrix} \\
\Longrightarrow\ & \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & -\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1} \\ -\boldsymbol{\Sigma}_{22.1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{\Sigma}_{22.1}^{-1} \end{pmatrix}
\end{aligned}$$

So , $\boldsymbol{C}_{22}^{-1} = \boldsymbol{\Sigma}_{22.1}$. And

$$\begin{aligned}
\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12} &= (\boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1})(-\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1}) \\
&= -\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{21}^{-1} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1} \\
&= -\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{21}^{-1} - \boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1} \\
&= -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}
\end{aligned}$$

□

6.2 Matrix Variate Distribution

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

$$\begin{aligned}
p(\boldsymbol{X}) &= \prod_{i=1}^{n} p(X_i) \\
&= \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(X_i - \mu_i)^T\boldsymbol{\Sigma}^{-1}(X_i - \mu_i)) \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(X_i - \mu_i)(X_i - \mu_i)^T))
\end{aligned}$$

Suppose $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots \mu_n)^T$, then

$$\begin{aligned}
p(\boldsymbol{X}) &= \frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}} \exp(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})^T I(\boldsymbol{X} - \boldsymbol{\mu}))) \\
&= \frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}} \mathrm{etr}(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})^T I(\boldsymbol{X} - \boldsymbol{\mu}))
\end{aligned}$$

We call $\boldsymbol{X}$ is Matrix-variate normal distributed.
Homework 1 If $vec(\boldsymbol{X}^T) \sim N_{np}(vec(\boldsymbol{\mu}^T), \boldsymbol{B} \bigotimes \boldsymbol{A})$, show the p.d.f of $\boldsymbol{X}$ is

$$\frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{A}|^{\frac{n}{2}}|\boldsymbol{B}|^{\frac{p}{2}}} \exp(-\frac{1}{2}\mathrm{tr}(\boldsymbol{A}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})^T\boldsymbol{B}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})))$$

Definition 6.2 (Wishart Distribution)**.** *If* $\boldsymbol{S} = \boldsymbol{X}^T\boldsymbol{X}$*, where the* $n \times p$ *matrix* $\boldsymbol{X}$ *is* $N(0, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$*, then* $\boldsymbol{S}$ *is positive definite and is said to have the Wishart distribution with* n *degrees of freedom and covariance matrix* $\boldsymbol{\Sigma}$*. We will write that* $\boldsymbol{S}$ *is* $W_p(\boldsymbol{\Sigma}, n)$*, the subscript on W' denoting the size of the matrix* $\boldsymbol{S}$*.*

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

$$p(\boldsymbol{S}) = \frac{|\boldsymbol{S}|^{\frac{r-p-1}{2}} \exp(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}))}{2^{\frac{rp}{2}} \pi^{\frac{p(p-1)}{4}} |\boldsymbol{\Sigma}|^{\frac{r}{2}} \prod_{i=1}^{p} \Gamma(\frac{r+1-i}{2})}, r \geq p$$

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

Splitting $\boldsymbol{S}$ into parts of q and $p-q$, i.e.

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{pmatrix}$$

where $\boldsymbol{S}_{11}$ is $q \times q$ and $\boldsymbol{S}_{22}$ is $(p-q) \times (p-q)$. So as $\boldsymbol{\Sigma}$.

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

1. $\boldsymbol{S}_{11} \sim W_q(\boldsymbol{\Sigma}_{11}, r), \boldsymbol{S}_{22} \sim W_{p-q}(\boldsymbol{\Sigma}_{22}.r)$
2. $\boldsymbol{S}_{11.2} \sim W_q(\boldsymbol{\Sigma}_{11.2}, r-(p-q))$
3. $\boldsymbol{S}_{11.2}$ *and* $(\boldsymbol{S}_{12}, \boldsymbol{S}_{22})$ *are independent.*
4. $\boldsymbol{S}_{12}|\boldsymbol{S}_{22} \sim N_{q,p-q}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{S}_{22}, \boldsymbol{\Sigma}_{11.2} \bigotimes \boldsymbol{S}_{22})$

Proof. Making the transformation

$$\begin{cases} \boldsymbol{S}_{11.2} = & \boldsymbol{S}_{11} - \boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21} \\ \boldsymbol{B}_{12} = & \boldsymbol{S}_{12} \\ \boldsymbol{B}_{22} = & \boldsymbol{S}_{22} \end{cases}$$

i.e. $(\boldsymbol{S}_{11}, \boldsymbol{S}_{12}, \boldsymbol{S}_{22}) \to (\boldsymbol{S}_{11.2}, \boldsymbol{B}_{12}, \boldsymbol{B}_{22})$. Since

$$\begin{aligned} (d(\boldsymbol{S}_{11.2}, \boldsymbol{B}_{12}, \boldsymbol{B}_{22})) &= (d(\boldsymbol{S}_{11.2}, \boldsymbol{S}_{12}, \boldsymbol{S}_{22})) \\ &= (d(\boldsymbol{S}_{11} - \boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}, \boldsymbol{B}_{12}, \boldsymbol{B}_{22})) \\ &= (d(\boldsymbol{S}_{11}, \boldsymbol{B}_{12}, \boldsymbol{B}_{22})) \\ &= (d(\boldsymbol{S}_{11}, \boldsymbol{S}_{12}, \boldsymbol{S}_{22})) \end{aligned}$$

So the Jacobian is 1. Hence, we can replace $\boldsymbol{S}$ with $\boldsymbol{S}_{11.2}, \boldsymbol{B}_{12}, \boldsymbol{B}_{22}$.

First, we have $|\boldsymbol{S}| = |\boldsymbol{S}_{11.2}||\boldsymbol{S}_{22}| = |\boldsymbol{S}_{11.2}||\boldsymbol{B}_{22}|$. Second, in the trace part, we have

$$\begin{aligned}
\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}) &= \operatorname{tr}\left(\begin{pmatrix}\boldsymbol{C}_{11} & \boldsymbol{C}_{12}\\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22}\end{pmatrix}\begin{pmatrix}\boldsymbol{S}_{11} & \boldsymbol{S}_{12}\\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22}\end{pmatrix}\right)\\
&= \operatorname{tr}(\boldsymbol{C}_{11}\boldsymbol{S}_{11}) + 2\operatorname{tr}(\boldsymbol{C}_{21}\boldsymbol{S}_{12}) + \operatorname{tr}(\boldsymbol{C}_{22}\boldsymbol{S}_{22})\\
&= \operatorname{tr}(\boldsymbol{C}_{11}\boldsymbol{S}_{11}) + 2\operatorname{tr}(\boldsymbol{C}_{12}\boldsymbol{B}_{21}) + \operatorname{tr}(\boldsymbol{C}_{22}\boldsymbol{B}_{22})\\
&(\text{since } \boldsymbol{S}_{11} = \boldsymbol{S}_{11.2} + \boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21})\\
&= \operatorname{tr}(\boldsymbol{C}_{11}\boldsymbol{S}_{11.2}) + \operatorname{tr}(\boldsymbol{C}_{11}\boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}) + 2\operatorname{tr}(\boldsymbol{C}_{12}\boldsymbol{B}_{21}) + \operatorname{tr}(\boldsymbol{C}_{22}\boldsymbol{B}_{22})\\
&(\text{using theorem 6.2})\\
&= \operatorname{tr}(\boldsymbol{\Sigma}_{11.2}^{-1}\boldsymbol{S}_{11.2}) + \operatorname{tr}(\boldsymbol{C}_{11}\boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}) + 2\operatorname{tr}(\boldsymbol{C}_{12}\boldsymbol{B}_{21}) + \operatorname{tr}(\boldsymbol{C}_{22}\boldsymbol{B}_{22})\\
&(\text{using } \boldsymbol{\Sigma}_{22}^{-1} = \boldsymbol{C}_{22.1} = \boldsymbol{C}_{22} - \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12})\\
&= \operatorname{tr}(\boldsymbol{\Sigma}_{11.2}^{-1}\boldsymbol{S}_{11.2}) + \operatorname{tr}(\boldsymbol{C}_{11}\boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}) + 2\operatorname{tr}(\boldsymbol{C}_{12}\boldsymbol{B}_{21}) + \operatorname{tr}(\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{B}_{22})\\
&\quad + \operatorname{tr}((\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12})\boldsymbol{B}_{22})
\end{aligned}$$

We can see that $\operatorname{tr}(\boldsymbol{\Sigma}_{11.2}^{-1}\boldsymbol{S}_{11.2})$ is corresponding to $p(\boldsymbol{S}_{11.2})$, $\operatorname{tr}(\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{B}_{22})$ is corresponding to $p(\boldsymbol{B}_{22})$. And to prove $\boldsymbol{S}_{11.2}$ and $(\boldsymbol{B}_{12}, \boldsymbol{B}_{22})$ are independent, we should have

$$p(\boldsymbol{S}_{11.2}, \boldsymbol{B}_{12}, \boldsymbol{B}_{22}) = p(\boldsymbol{S}_{11.2})p(\boldsymbol{B}_{12}, \boldsymbol{B}_{22}) = p(\boldsymbol{S}_{11.2})p(\boldsymbol{B}_{12}|\boldsymbol{B}_{22})p(\boldsymbol{B}_{22})$$

So, the residue terms should be corresponding to $p(\boldsymbol{B}_{12}, \boldsymbol{B}_{22})$, which is the 4th proposition in theorem 6.4. Now we rewritten them to show that they are corresponding to $N_{q,p-q}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{S}_{22}, \boldsymbol{\Sigma}_{11.2} \bigotimes \boldsymbol{S}_{22})$.

$$\begin{aligned}
& \operatorname{tr}(\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{B}_{22}) + \operatorname{tr}(\boldsymbol{C}_{11}\boldsymbol{B}_{12}\boldsymbol{B}_{22}^{-1}\boldsymbol{B}_{21}) + 2\operatorname{tr}(\boldsymbol{C}_{12}\boldsymbol{B}_{21})\\
= \; & \operatorname{tr}(\boldsymbol{C}_{11}(\boldsymbol{B}_{12} + \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{B}_{22})\boldsymbol{B}_{22}^{-1}(\boldsymbol{B}_{12} + \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{B}_{22})^T)\\
= \; & \operatorname{tr}(\boldsymbol{\Sigma}_{11.2}^{-1}(\boldsymbol{B}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{B}_{22})\boldsymbol{B}_{22}^{-1}(\boldsymbol{B}_{12} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{B}_{22})^T)
\end{aligned}$$

Finally, we have $|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{11.2}||\boldsymbol{\Sigma}_{22}|$.

Now we have proved that the p.d.f of $\boldsymbol{S}$ can be decomposed into terms $\boldsymbol{S}_{22}$, $\boldsymbol{S}_{11.2}$ and $\boldsymbol{S}_{12}|\boldsymbol{S}_{22}$. □

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

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

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

Proof. First, we have

$$|\boldsymbol{S}|^{\frac{1}{2}(r-p-1)} etr(-\frac{1}{2}\boldsymbol{S}) = (\prod_{i=1}^{p} t_{ii}^{\frac{2(r-p-1)}{2}}) etr(-\frac{1}{2}\sum_{1\leqslant j\leqslant i\leqslant p} t_{ij}^2)$$

According to Theorem 5.6, we have $J(\boldsymbol{S} \to \boldsymbol{T}) = 2^p \prod_{i=1}^p t_{ii}^{p-i+1}$. Also, we have $\mathrm{tr}(\boldsymbol{S}) = \mathrm{tr}(\boldsymbol{T}^T\boldsymbol{T})$. Thus,

$$
\begin{aligned}
p(\boldsymbol{T}) &\propto \prod_{1 \leqslant j \leqslant i \leqslant p} \exp(-\frac{1}{2}t_{ij}^2) \prod_{i=1}^{p} (t_{ii})^{\frac{r-p-1}{2}} |J(\boldsymbol{S} \to \boldsymbol{T})| \\
&\propto \prod_{1 \leqslant j < i \leqslant p} \exp(-\frac{1}{2}t_{ij}^2) \prod_{i=1}^{p} (t_{ii}^2)^{\frac{r-i+1}{2}} \exp(-\frac{1}{2}t_{ii}^2)
\end{aligned}
$$

$\prod_{1 \leqslant j < i \leqslant p} \exp(-\frac{1}{2}t_{i,j}^2)$ denote the independent standard normal distributions of $t_{i,j}$.
$\prod_{i=1}^{p} (t_{i,i}^2)^{\frac{r-i+1}{2}} \exp(-\frac{1}{2}t_{i,i}^2)$ denote the independent distributions χ_{r-i+1}^2. □

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

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

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

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

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

$p(\boldsymbol{U})$ is called matrix-variate Beta Distribution.

Homework 2 Prove theorem 6.6.