Matrix Variate Normal

Seoul National University
Bayesian Statistics Lab

May 17, 2023

Table of contents

We know that

(1) For univariate normal r.v. x, p.d.f. is : $(2\pi)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \quad x \in \mathbb{R}$

(2) For multivariate normal r.v. $\mathbf{x} = (x_1, \ldots, x_p)^\top$, p.d.f. is :

$$(2\pi)^{-p/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1/2}(\mathbf{x}-\mu)\right\}, \quad \mathbf{x}, \mu \in \mathbb{R}^p, \Sigma > 0$$

Question

Question1 : How about matrix normal r.m. $\mathbf{X} \in \mathbb{R}^{p \times n}$?

Definition 1

The random matrix $\mathbf{X}(p \times n)$ is said to have a matrix variate normal distribution with mean matrix $\mathbf{M}(p \times n)$ and covariance matrix $\Sigma \otimes \Psi$ where $\Sigma(p \times p) > 0$ and $\Psi(n \times n) > 0$, if $\text{vec}(\mathbf{X}^\top) \sim N_{pn}(\text{vec}(\mathbf{M}^\top, \Sigma \otimes \Psi))$.

Simply write as :

$$\mathbf{X} \sim N_{p,n}(\mathbf{M}, \Sigma \otimes \Psi).$$

Question

Question2 : How do we derive p.d.f. of Matrix Variate Normal Distribution?

Density function

Theorem 1

If $\mathbf{X} \sim N_{p,n}(\mathbf{M}, \Sigma \otimes \Psi)$, *then the p.d.f. of* $\mathbf{X}$ *is given by*

$$(2\pi)^{-np/2}|\Sigma|^{-n/2}|\Psi|^{-p/2} \exp\left\{ tr\left[-\frac{1}{2}\Sigma^{-1}(\mathbf{X}-\mathbf{M})\Psi^{-1}(\mathbf{X}-\mathbf{M})^{\top} \right] \right\},$$

where $\quad \mathbf{X} \in \mathbb{R}^{p\times n}, \quad \mathbf{M} \in \mathbb{R}^{p\times n}.$

Before we start , remark the following :

Lemma 1

For $A(p \times m), B(n \times q), C(q \times m), D(q \times n), E(m \times m)$ *and* $X(m \times n)$, *we have*

(i) $\mathit{vec}(AXB) = (B^\top \otimes A)\mathit{vec}(X)$

(ii) $\mathit{tr}(CXB) = [\mathit{vec}(C^\top)]^\top [I_q \otimes X]\mathit{vec}(X)$

(iii) $\mathit{tr}(DX^\top EXB) = [\mathit{vec}(X)]^\top [D^\top B^\top \otimes E]\mathit{vec}(X) = [\mathit{vec}(X)]^\top [BD \otimes E^\top]\mathit{vec}(X)$

Lemma 2

(i) *For nonsingular matrices* A *and* B, $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

(ii) *For* $A(m \times m)$ *and* $B(n \times n)$, $|A \otimes B| = |A|^n |B|^m$.

Proof.

Let $\mathbf{x} = \text{vec}(\mathbf{X}^\top)$ and $\mathbf{m} = \text{vec}(\mathbf{M}^\top)$.
Then according to the **Definition 1**, $\mathbf{x} \sim N_{pn}(\mathbf{m}, \Sigma \otimes \Psi)$, and its p.d.f. is

$$(2\pi)^{-np/2}|\Sigma \otimes \Psi|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^\top(\Sigma \otimes \Psi)^{-1}(\mathbf{x}-\mathbf{m})\right\}.$$

Using **Lemma 1**, we can derive the following :

$$\begin{aligned}
-\frac{1}{2}(\mathbf{x}-\mathbf{m})^\top(\Sigma \otimes \Psi)^{-1}(\mathbf{x}-\mathbf{m}) &= (\mathbf{x}-\mathbf{m})^\top(\Sigma^{-1} \otimes \Psi^{-1})(\mathbf{x}-\mathbf{m}) \quad (\because \textbf{Lemma 2}(i)) \\
&= \text{tr}[(\Sigma^{-1} \otimes \Psi^{-1})(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^\top] \\
&= \text{tr}[(\Sigma^{-1} \otimes \Psi^{-1})(\text{vec}(\mathbf{X}^\top - \mathbf{M}^\top))(\text{vec}(\mathbf{X}^\top - \mathbf{M}^\top))^\top] \\
&= \text{tr}[(\text{vec}(\mathbf{X}^\top - \mathbf{M}^\top))^\top(\Sigma^{-1} \otimes \Psi^{-1})(\text{vec}(\mathbf{X}^\top - \mathbf{M}^\top))] \\
&= \text{tr}[(\text{vec}(\mathbf{X}^\top - \mathbf{M}^\top))^\top \text{vec}(\Psi^{-1}\mathbf{X}^\top - \mathbf{M}^\top\Sigma^{-1})] \quad (\because \textbf{Lemma 1}(i)) \\
&= \text{tr}[(\mathbf{X}-\mathbf{M})\Psi^{-1}\mathbf{X}^\top - \mathbf{M}^\top\Sigma^{-1}] \quad (\because \textbf{Lemma 1}(ii)) \\
&= \text{tr}\left[-\frac{1}{2}\Sigma^{-1}(\mathbf{X}-\mathbf{M})\Psi^{-1}(\mathbf{X}-\mathbf{M})^\top\right]
\end{aligned}$$

□

Density function

Proof continued.

Also using **Lemma 2**, we can derive the following :

$$|\Sigma \otimes \Psi|^{-1/2} = (|\Sigma|^n |\Psi|^p)^{-1/2} = |\Sigma|^{-n/2}|\Psi|^{-p/2} (\because \textbf{Lemma 2}(ii))$$

Combining above results, p.d.f of $\mathbf{X}$ is

$$(2\pi)^{-np/2}|\Sigma|^{-n/2}|\Psi|^{-p/2} \exp\left\{ \text{tr}\left[-\frac{1}{2}\Sigma^{-1}(\mathbf{X}-\mathbf{M})\Psi^{-1}(\mathbf{X}-\mathbf{M})^\top \right] \right\},$$

where $\quad \mathbf{X} \in \mathbb{R}^{p\times n}, \quad \mathbf{M} \in \mathbb{R}^{p\times n}.$ □

Theorem 2

If $\mathbf{X} \sim N_{p,n}(\mathbf{M}, \Sigma \otimes \Psi)$, *then* $\mathbf{X}^\top \sim N_{n,p}(\mathbf{M}^\top, \Psi \otimes \Sigma)$.

We skip the proof since the proof suffices to show that the exponents occurring in the densities of vec$(\mathbf{X}^\top)$ and vec$(\mathbf{X})$ are equal. But this can be derived not so hard from **Lemma 1**.

Characteristic function

Theorem 3

If $\mathbf{X} \sim N_{p,n}(\mathbf{M}, \Sigma \otimes \Psi)$, *then the characteristic function of* $\mathbf{X}$ *is*

$$\phi_{\mathbf{X}}(\mathbf{Z}) = \exp\left(tr \left[i\mathbf{Z}^\top \mathbf{M} - \frac{1}{2}\mathbf{Z}^\top \Sigma \mathbf{Z} \Psi \right] \right).$$

Proof.

We know that the characteristic function of a random matrix $\mathbf{X}(p \times n)$ is denoted by

$$\begin{aligned} \phi_{\mathbf{X}}(\mathbf{Z}) &= M_{\mathbf{X}}(i\mathbf{Z}) = \mathbb{E}\left[e^{\mathrm{tr}(i\mathbf{X}\mathbf{Z}^\top)} \right] \\ &= \mathbb{E}\left[e^{\mathrm{tr}(i\mathrm{vec}(\mathbf{X}^\top)^\top \mathrm{vec}(\mathbf{Z}^\top))} \right] \quad (\because \textbf{Lemma 1}(ii)) \end{aligned}$$

$\square$

Proof continued.

Now, we know that $\text{vec}(\mathbf{X}^\top) \sim N_{pn}(\text{vec}(\mathbf{M}^\top), \Sigma \otimes \Psi)$. Hence, we can use the form of characteristic function of multivariate normal as following :

$$\begin{aligned}
\phi_{\mathbf{X}}(\mathbf{Z}) &= \exp\left[i\text{vec}(\mathbf{M}^\top)^\top \text{vec}(\mathbf{Z}^\top) - \frac{1}{2}\text{vec}(\mathbf{Z}^\top)^\top (\Sigma \otimes \Psi)\text{vec}(\mathbf{Z}^\top)\right] \\
&= \exp\left[i\text{tr}(\mathbf{M}\mathbf{Z}^\top) - \frac{1}{2}\text{vec}(\mathbf{Z}^\top)^\top \text{vec}(\Psi\mathbf{Z}^\top\mathbf{Z})\right] \quad (\because \textbf{Lemma 1}(ii),(i)) \\
&= \exp\left[i\text{tr}(\mathbf{M}\mathbf{Z}^\top) - \frac{1}{2}\text{tr}(\mathbf{Z}\Psi\mathbf{Z}^\top\Sigma)\right] \quad (\because \textbf{Lemma 1}(ii)) \\
&= \exp\left[\text{tr}\left(i\mathbf{M}\mathbf{Z}^\top - \frac{1}{2}\mathbf{Z}\Psi\mathbf{Z}^\top\Sigma\right)\right] \\
&= \exp\left[\text{tr}\left(i\mathbf{Z}^\top\mathbf{M} - \frac{1}{2}\mathbf{Z}^\top\Sigma\mathbf{Z}\Psi\right)\right]
\end{aligned}$$

$\square$

Question

Question3 : Why do we need c.f. of Matrix Variate Normal Distribution?

Properties

Theorem 4

If $\mathbf{X} \sim N_{p,n}(\mathbf{M}, \Sigma \otimes \Psi)$, $D(m \times p)$ *is of rank* $m \leq p$ *and* $C(n \times t)$ *is of rank* $t \leq n$, *then*

$$D\mathbf{X}C \sim N_{m,t}(D\mathbf{M}C, (D\Sigma D^\top) \otimes (C^\top \Psi C))$$

Proof.

We know that the characteristic function of $D\mathbf{X}C$ is

$$\phi_{D\mathbf{X}C}(\mathbf{Z}) = \mathbb{E}[e^{tr[iD\mathbf{X}C\mathbf{Z}^\top]}] = \mathbb{E}[e^{tr[i\mathbf{X}\mathbf{Z}_1^\top]}]$$

where $\mathbf{Z}_1^\top = C\mathbf{Z}^\top D$. Then by **Theorem 3**,

$$\begin{aligned}
\phi_{D\mathbf{X}C}(\mathbf{Z}) &= e^{\mathrm{tr}[iZ_1^\top \mathbf{M} - \frac{1}{2}\mathbf{Z}_1^\top \Sigma \mathbf{Z}_1 \Psi]} \\
&= e^{\mathrm{tr}[iC\mathbf{Z}^\top D\mathbf{M} - \frac{1}{2}C\mathbf{Z}^\top D\Sigma(C\mathbf{Z}^\top D)^\top \Psi]} \\
&= e^{\mathrm{tr}[i\mathbf{Z}^\top (D\mathbf{M}C) - \frac{1}{2}\mathbf{Z}^\top (D\Sigma D^\top)\mathbf{Z}(C^\top \Psi C)]},
\end{aligned}$$

which is the characteristic function of a MVND with mean $D\mathbf{M}C$ and covariance matrix $(D\Sigma D^\top) \otimes (C^\top \Psi C)$. □

Question

Question4 : Can we get marginal and conditional distribution?

Marginal distribution

Theorem 5

Let $\mathbf{X} \sim N_{p,n}(\mathbf{M}, \Sigma \otimes \Psi)$, *and partition* $\mathbf{X}, \mathbf{M}, \Sigma$ *and* Ψ *as*

$$\mathbf{X} = \begin{bmatrix} \underset{(p_1 \times n_1)}{\mathbf{X}_{11}} & \underset{(p_1 \times n_2)}{\mathbf{X}_{12}} \\ \underset{(p_2 \times n_1)}{\mathbf{X}_{21}} & \underset{(p_2 \times n_2)}{\mathbf{X}_{22}} \end{bmatrix}, \ \mathbf{M} = \begin{bmatrix} \underset{(p_1 \times n_1)}{\mathbf{M}_{11}} & \underset{(p_1 \times n_2)}{\mathbf{M}_{12}} \\ \underset{(p_2 \times n_1)}{\mathbf{M}_{21}} & \underset{(p_2 \times n_2)}{\mathbf{M}_{22}} \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} \underset{(p_1 \times p_1)}{\Sigma_{11}} & \underset{(p_1 \times p_2)}{\Sigma_{12}} \\ \underset{(p_2 \times p_1)}{\Sigma_{21}} & \underset{(p_2 \times p_2)}{\Sigma_{22}} \end{bmatrix} \ and \ \ \Psi = \begin{bmatrix} \underset{(n_1 \times n_1)}{\Psi_{11}} & \underset{(n_1 \times n_2)}{\Psi_{12}} \\ \underset{(n_2 \times n_1)}{\Psi_{21}} & \underset{(n_2 \times n_2)}{\Psi_{22}} \end{bmatrix}$$

Then, $\mathbf{X}_{11} \sim N_{p_1,n_1}(\mathbf{M}_{11}, \Sigma_{11} \otimes \Psi_{11})$.

Proof.

The result follows by using

$$D = \begin{bmatrix} I_{p_1} & 0 \end{bmatrix}, \ \ and \ \ C^\top = \begin{bmatrix} I_{n_1} & 0 \end{bmatrix}$$

in **Theorem 4**. □

Theorem 6

Let $\mathbf{X} \sim N_{p,n}(\mathbf{M}, \Sigma \otimes \Psi)$, *and partition* $\mathbf{X}, \mathbf{M}, \Sigma$ *and* Ψ *as*

$$\mathbf{X} = \begin{bmatrix} \underset{(p_1 \times n)}{\mathbf{X}_{1r}} \\ \underset{(p_2 \times n))}{\mathbf{X}_{2r}} \end{bmatrix} = \begin{bmatrix} \underset{(p \times n_1)}{\mathbf{X}_{1c}} & \underset{(p \times n_2)}{\mathbf{X}_{2c}} \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} \underset{(p_1 \times n)}{\mathbf{M}_{1r}} \\ \underset{(p_2 \times n))}{\mathbf{M}_{2r}} \end{bmatrix} = \begin{bmatrix} \underset{(p \times n_1)}{\mathbf{M}_{1c}} & \underset{(p \times n_2)}{\mathbf{M}_{2c}} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \underset{(p_1 \times p_1)}{\Sigma_{11}} & \underset{(p_1 \times p_2)}{\Sigma_{12}} \\ \underset{(p_2 \times p_1)}{\Sigma_{21}} & \underset{(p_2 \times p_2)}{\Sigma_{22}} \end{bmatrix} and \Psi = \begin{bmatrix} \underset{(n_1 \times n_1)}{\Psi_{11}} & \underset{(n_1 \times n_2)}{\Psi_{12}} \\ \underset{(n_2 \times n_1)}{\Psi_{21}} & \underset{(n_2 \times n_2)}{\Psi_{22}} \end{bmatrix}$$

Then, following holds :

(i) $\mathbf{X}_{1r} \sim N_{m,n}(\mathbf{M}_{1r}, \Sigma \otimes \Psi)$, $\mathbf{X}_{1c} \sim N_{p,t}(\mathbf{M}_{1c}, \Sigma \otimes \Psi_{11})$

(ii) $X_{2r}|X_{1r} \sim N_{p_2,n}(\mathbf{M}_{2r} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_{1r} - \mathbf{M}_{1r}), \Sigma_{22\cdot1} \otimes \Psi)$

(iii) $X_{2c}|X_{1c} \sim N_{p,n_2}(\mathbf{M}_{2c} + (\mathbf{X}_{1c} - \mathbf{M}_{1c})\Psi_{11}^{-1}\Psi_{12}), \Sigma \otimes \Psi_{22\cdot1})$,

where $\Sigma_{22\cdot1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ *and* $\Psi_{22\cdot1} = \Psi_{22} - \Psi_{21}\Psi_{11}^{-1}\Psi_{12}$.

Proof.

(i)It can be proved by using same logic as in the proof of **Theorem 5**:

$$\begin{cases} for \quad X_{1r} \Rightarrow use \quad D = (I_{p_1} \quad 0), \quad C^\top = (I_n \quad 0) \\ for \quad X_{1c} \Rightarrow use \quad D = (I_p \quad 0), \quad C^\top = (I_{n_1} \quad 0) \end{cases}$$

□

Proof continued.

(ii)We will use the logic : $pdf(y_1, y_2) = pdf(y_1)pdf(y_2|y_1)$

$$\text{Let } \Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix},$$

where $\Sigma^{11} = \Sigma_{11\cdot 2}^{-1}$, $\Sigma^{22} = \Sigma_{22\cdot 1}^{-1}$, $\Sigma^{12} = -\Sigma_{11\cdot 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} = -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1} = (\Sigma^{21})^\top$.
Then we can derive the following :

$$\begin{aligned}
(\mathbf{X}-\mathbf{M})^\top\Sigma^{-1}(\mathbf{X}-\mathbf{M}) &= \begin{bmatrix}(\mathbf{X}_{1r}-\mathbf{M}_{1r})^\top & (\mathbf{X}_{2r}-\mathbf{M}_{2r})^\top\end{bmatrix} \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{1r}-\mathbf{M}_{1r} \\ \mathbf{X}_{2r}-\mathbf{M}_{2r} \end{bmatrix} \\
&= (\mathbf{X}_{1r}-\mathbf{M}_{1r})^\top\Sigma^{11}(\mathbf{X}_{1r}-\mathbf{M}_{1r}) + (\mathbf{X}_{1r}-\mathbf{M}_{1r})^\top\Sigma^{12}(\mathbf{X}_{2r}-\mathbf{M}_{2r}) \\
&\quad + (\mathbf{X}_{2r}-\mathbf{M}_{2r})^\top\Sigma^{21}(\mathbf{X}_{1r}-\mathbf{M}_{1r}) + (\mathbf{X}_{2r}-\mathbf{M}_{2r})^\top\Sigma^{22}(\mathbf{X}_{2r}-\mathbf{M}_{2r}) \\
&= (\mathbf{X}_{1r}-\mathbf{M}_{1r})^\top(\Sigma^{11}-\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21})(\mathbf{X}_{1r}-\mathbf{M}_{1r}) \\
&\quad + (\mathbf{X}_{1r}-\mathbf{M}_{1r})^\top\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}(\mathbf{X}_{1r}-\mathbf{M}_{1r}) + (\mathbf{X}_{1r}-\mathbf{M}_{1r})^\top\Sigma^{12}(\mathbf{X}_{2r}-\mathbf{M}_{2r}) \\
&\quad + (\mathbf{X}_{2r}-\mathbf{M}_{2r})^\top\Sigma^{21}(\mathbf{X}_{1r}-\mathbf{M}_{1r}) + (\mathbf{X}_{2r}-\mathbf{M}_{2r})^\top\Sigma^{22}(\mathbf{X}_{2r}-\mathbf{M}_{2r})
\end{aligned}$$

$\square$

Proof continued.

$$= (\mathbf{X}_{1r} - \mathbf{M}_{1r})^\top \Sigma_{11}^{-1} (\mathbf{X}_{1r} - \mathbf{M}_{1r}) \\ + (\mathbf{X}_{2r} - \mathbf{M}_{2r} - \Sigma_{21}\Sigma_{11}^{-1})^\top \Sigma_{22\cdot 1}^{-1} (\mathbf{X}_{2r} - \mathbf{M}_{2r} - \Sigma_{21}\Sigma_{11}^{-1})$$

This is because

(1) $\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21} = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}(\Sigma_{22\cdot 1})\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}$

(2) $\Sigma^{11} - \Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21} = \Sigma_{11}^{-1}$ ($\because$ Woodbury matrix identity)

Thus, the density of $\mathbf{X}$ can be written as

$$f(\mathbf{X}) = (2\pi)^{-np/2}|\Sigma|^{-n/2}|\Psi|^{-p/2} e^{tr[-\frac{1}{2}(\mathbf{X}-\mathbf{M})^\top \Sigma^{-1}(\mathbf{X}-\mathbf{M})\Psi^{-1}]}$$

$$= (2\pi)^{-np_1/2}|\Sigma_{11}|^{-n/2}|\Psi|^{-p_1/2} e^{tr[-\frac{1}{2}(\mathbf{X}_{1r}-\mathbf{M}_{1r})^\top \Sigma_{11}^{-1}(\mathbf{X}_{1r}-\mathbf{M}_{1r})\Psi^{-1}]}$$

$$\times (2\pi)^{-np_2/2}|\Sigma_{11}|^{-n/2}|\Psi|^{-p_2/2} e^{tr[-\frac{1}{2}(\mathbf{X}_{2r}-\mathbf{M}_{2r}-\Sigma_{21}\Sigma_{11}^{-1})^\top \Sigma_{22\cdot 1}^{-1}(\mathbf{X}_{2r}-\mathbf{M}_{2r}-\Sigma_{21}\Sigma_{11}^{-1})]}$$

Hence we can conclude the following :

$$\begin{cases} \mathbf{X}_{1r} \sim N_{p_1,n}(\mathbf{M}_{1r}, \Sigma \otimes \Psi) \\ \mathbf{X}_{2r}|\mathbf{X}_{1r} \sim N_{p_2,n}(\mathbf{M}_{2r} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_{1r} - \mathbf{M}_{1r}), \Sigma_{22\cdot 1} \otimes \Psi) \end{cases}$$

□

Proof continued.

Using **Theorem 2**, we know that

$$\mathbf{X}^\top = \begin{bmatrix} \mathbf{X}_{1c}^\top \\ \mathbf{X}_{2c}^\top \end{bmatrix} \sim N_{n,p}\left(\begin{bmatrix} \mathbf{M}_{1c}^\top \\ \mathbf{M}_{2c}^\top \end{bmatrix}, \Psi \otimes \Sigma \right)$$

Then using same logic as in (ii), we get

$$\mathbf{X}_{2c}^\top | \mathbf{X}_{1c}^\top \sim N_{n_2,p}(\mathbf{M}_{2c}^\top + \Psi_{21}\Psi_{11}^{-1}(\mathbf{X}_{1c}^\top - \mathbf{M}_{2c}^\top), \Psi_{22\cdot1} \otimes \Sigma)$$

Again using **Theorem 2**, we get

$$\mathbf{X}_{2c} | \mathbf{X}_{1c} \sim N_{p,n_2}(\mathbf{M}_{2c} + (\mathbf{X}_{1c} - \mathbf{M}_{1c})\Psi_{11}^{-1}\Psi_{12}, \Sigma \otimes \Psi_{22\cdot1})$$

$\square$

Another Proof.

(ii)We will use the logic : $X_2|X_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$
From **Definition 1**, we have

$$\text{vec}(\mathbf{X}^\top) = \begin{bmatrix} \text{vec}(\mathbf{X}_{1r}^\top) \\ \text{vec}(\mathbf{X}_{2r}^\top) \end{bmatrix} \sim N_{pn}\left(\begin{bmatrix} \text{vec}(\mathbf{M}_{1r}^\top) \\ \text{vec}(\mathbf{M}_{2r}^\top) \end{bmatrix}, \begin{bmatrix} \Sigma_{11} \otimes \Psi & \Sigma_{12} \otimes \Psi \\ \Sigma_{21} \otimes \Psi & \Sigma_{22} \otimes \Psi \end{bmatrix} \right)$$

Then we have

$$\text{vec}(\mathbf{X}_{2r}^\top)|\text{vec}(\mathbf{X}_{1r}^\top) \sim N_{p_2 n}\left(\text{vec}(\mathbf{M}_{2r}^\top) + (\Sigma_{21} \otimes \Psi)(\Sigma_{11} \otimes \Psi)^{-1}\text{vec}(\mathbf{X}_{1r}^\top - \mathbf{M}_{1r}^\top), \mathbf{V}_{22\cdot 1} \right),$$

where $\mathbf{V}_{22\cdot 1} = (\Sigma_{22} \otimes \Psi) - (\Sigma_{21} \otimes \Psi)(\Sigma_{11} \otimes \Psi)^{-1}(\Sigma_{12} \otimes \Psi)$. □

Another Proof continued.

For mean part,

(1) $(\Sigma_{21} \otimes \Psi)(\Sigma_{11} \otimes \Psi)^{-1} = (\Sigma_{21}\Sigma_{11}^{-1} \otimes I_p)$

(2)

$$\begin{aligned}(\Sigma_{21} \otimes \Psi)(\Sigma_{11} \otimes \Psi)^{-1}\text{vec}\left([\mathbf{X}_{1r} - \mathbf{M}_{1r}]^\top\right) &= (\Sigma_{21}\Sigma_{11}^{-1} \otimes I_p)\text{vec}\left([\mathbf{X}_{1r} - \mathbf{M}_{1r}]^\top\right) \\ &= \text{vec}\left([\mathbf{X}_{1r} - \mathbf{M}_{1r}]^\top \Sigma_{11}^{-1}\Sigma_{12}\right)\end{aligned}$$

(3)

$$\begin{aligned}\text{mean} &\equiv \text{vec}\left(\mathbf{M}_{2r}^\top\right) + (\Sigma_{21} \otimes \Psi)(\Sigma_{11} \otimes \Psi)^{-1}\text{vec}\left([\mathbf{X}_{1r} - \mathbf{M}_{1r}]^\top\right) \\ &= \text{vec}\left([\mathbf{M}_{2r} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_{1r} - \mathbf{M}_{1r})]^\top\right)\end{aligned}$$

□

Another Proof continued.

For covariance matrix part,

(1) $(\Sigma_{21} \otimes \Psi)(\Sigma_{11} \otimes \Psi)^{-1}(\Sigma_{12} \otimes \Psi) = (\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \otimes \Psi)$

(2)

$$\begin{aligned}(\Sigma_{22} \otimes \Psi) - (\Sigma_{21} \otimes \Psi)(\Sigma_{11} \otimes \Psi)^{-1}(\Sigma_{12} \otimes \Psi) &= ([\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}] \otimes \Psi) \\ &= (\Sigma_{22\cdot 1} \otimes \Psi)\end{aligned}$$

□

Another Proof continued.

Combining above results, we get

$$\therefore \text{vec}(\mathbf{X}_{2r}^\top)|\text{vec}(\mathbf{X}_{1r}^\top) \sim N_{p_2 n}\left(\text{vec}\left([\mathbf{M}_{2r} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_{1r} - \mathbf{M}_{1r})]^\top\right), \Sigma_{22\cdot 1} \otimes \Psi\right)$$

By **Definition 1**, we get

$$\therefore \mathbf{X}_{2r}|\mathbf{X}_{1r} \sim N_{p_2,n}\left(\mathbf{M}_{2r} + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_{1r} - \mathbf{M}_{1r}), \Sigma_{22\cdot 1} \otimes \Psi\right),$$

which is the same result we get from the previous proof. □

Discussion

(1) Why isn't matrix variate normal popular?

(2) Relationship between matrix variate normal and wishart

Reference

A.K. Gupta, D.K. Nagar (1999). Matrix Variate Distributions. CHAPMAN HALL/CRC