Note on Multivariate conditional Gaussian distribution

Hyungjin Chung

March 9, 2021

1 Inverse of matrix partition

Before diving into calculating the multivariate conditional Gaussian distribution, it is first necessary to lay out the useful theorems that will be needed to proceed with the proof. The notations that are used in Murphy [2012] will be adopted.

Theorem 1. *Define the following compartmentalized matrix*

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix},$$

where each of the components $\boldsymbol{E}$, $\boldsymbol{F}$, $\boldsymbol{G}$, and $\boldsymbol{H}$are matrices that are invertible. Then, we have

$$\boldsymbol{M}^{-1} = \begin{pmatrix} (\boldsymbol{M}/\boldsymbol{H})^{-1} & -(\boldsymbol{M}/\boldsymbol{H})^{-1}\boldsymbol{F}\boldsymbol{H}^{-1} \\ -\boldsymbol{H}^{-1}\boldsymbol{G}(\boldsymbol{M}/\boldsymbol{H})^{-1} & \boldsymbol{H}^{-1} + \boldsymbol{H}^{-1}\boldsymbol{G}(\boldsymbol{M}/\boldsymbol{H})^{-1}\boldsymbol{F}\boldsymbol{H}^{-1} \end{pmatrix}$$

Here, we used the *Schur complement* for the ease of notation, which are defined as follows:

$$\begin{aligned} (\boldsymbol{M}/\boldsymbol{H}) &:= \boldsymbol{E} - \boldsymbol{F}\boldsymbol{H}^{-1}\boldsymbol{G} \\ (\boldsymbol{M}/\boldsymbol{E}) &:= \boldsymbol{H} - \boldsymbol{G}\boldsymbol{E}^{-1}\boldsymbol{F}. \end{aligned}$$

Proof. It is convenient to first zero-out the entries outside of the diagonal.

$$\begin{pmatrix} \boldsymbol{I} & -\boldsymbol{F}\boldsymbol{H}^{-1} \\ 0 & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix} = \begin{pmatrix} \boldsymbol{E} - \boldsymbol{F}\boldsymbol{H}^{-1}\boldsymbol{G} & 0 \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix} \tag{1}$$

Similarly, we do the same for the bottom left entry:

$$\begin{pmatrix} \boldsymbol{E} - \boldsymbol{F}\boldsymbol{H}^{-1}\boldsymbol{G} & 0 \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & 0 \\ -\boldsymbol{H}^{-1}\boldsymbol{G} & \boldsymbol{I} \end{pmatrix} = \begin{pmatrix} \boldsymbol{E} - \boldsymbol{F}\boldsymbol{H}^{-1}\boldsymbol{G} & 0 \\ 0 & \boldsymbol{H} \end{pmatrix} \tag{2}$$

Then, using eq. 1, 2, we have

$$\underbrace{\begin{pmatrix} \boldsymbol{I} & -\boldsymbol{F}\boldsymbol{H}^{-1} \\ 0 & \boldsymbol{I} \end{pmatrix}}_{\boldsymbol{X}} \underbrace{\begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix}}_{\boldsymbol{M}} \underbrace{\begin{pmatrix} \boldsymbol{I} & 0 \\ -\boldsymbol{H}^{-1}\boldsymbol{G} & \boldsymbol{I} \end{pmatrix}}_{\boldsymbol{Z}} = \underbrace{\begin{pmatrix} \boldsymbol{E} - \boldsymbol{F}\boldsymbol{H}^{-1}\boldsymbol{G} & 0 \\ 0 & \boldsymbol{H} \end{pmatrix}}_{\boldsymbol{W}} \tag{3}$$

Subsequently, by taking the inverse of both sides,

$$\boldsymbol{Z}^{-1}\boldsymbol{M}^{-1}\boldsymbol{X}^{-1} = \boldsymbol{W}^{-1} \tag{4}$$

$$\boldsymbol{M}^{-1} = \boldsymbol{Z}\boldsymbol{W}^{-1}\boldsymbol{X} \tag{5}$$

By expanding eq. 5, we have

$$\begin{pmatrix} \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{I} & 0 \\ -\boldsymbol{H}^{-1}\boldsymbol{G} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} (\boldsymbol{M}/\boldsymbol{H})^{-1} & 0 \\ 0 & \boldsymbol{H}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{F}\boldsymbol{H}^{-1} \\ 0 & \boldsymbol{I} \end{pmatrix} \tag{6}$$

$$= \begin{pmatrix} (\boldsymbol{M}/\boldsymbol{H})^{-1} & 0 \\ -\boldsymbol{H}^{-1}\boldsymbol{G}(\boldsymbol{M}/\boldsymbol{H})^{-1} & \boldsymbol{H}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{F}\boldsymbol{H}^{-1} \\ 0 & \boldsymbol{I} \end{pmatrix} \tag{7}$$

$$= \begin{pmatrix} (\boldsymbol{M}/\boldsymbol{H})^{-1} & -(\boldsymbol{M}/\boldsymbol{H})^{-1}\boldsymbol{F}\boldsymbol{H}^{-1} \\ -\boldsymbol{H}^{-1}\boldsymbol{G}(\boldsymbol{M}/\boldsymbol{H})^{-1} & \boldsymbol{H}^{-1} + \boldsymbol{H}^{-1}\boldsymbol{G}(\boldsymbol{M}/\boldsymbol{H})^{-1}\boldsymbol{F}\boldsymbol{H}^{-1} \end{pmatrix} \tag{8}$$

□

2 Multivariate Conditional Gaussian Distribution

Now that we have Theorem 1, we are ready to proceed. Suppose that $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$ is jointly Gaussian with parameters

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \tag{9}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \tag{10}$$

$$\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix}. \tag{11}$$

The marginal probability distributions read

$$p(\boldsymbol{x}_1) = \mathcal{N}(\boldsymbol{x}_1|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})p(\boldsymbol{x}_2) = \mathcal{N}(\boldsymbol{x}_2|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}) \tag{12}$$

Finally, the theorem for the multivariate conditional Gaussian distribution is given.

Theorem 2. *The parameters* $\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}$ *defined in*

$$p(\boldsymbol{x}_1|\boldsymbol{x}_2) = \mathcal{N}(\boldsymbol{x}_1|\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})$$

are given as

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2),$$

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

Proof. We can get the conditional distribution by factoring the joint distribution as follows:

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2) = p(\boldsymbol{x}_1|\boldsymbol{x}_2)p(\boldsymbol{x}_2) \tag{13}$$

By neglecting the normalization constants and considering only the exponential terms, we have

$$
\begin{aligned}
E &= \exp\{-\frac{1}{2}\begin{pmatrix}\boldsymbol{x}_1-\boldsymbol{\mu}_1\\ \boldsymbol{x}_2-\boldsymbol{\mu}_2\end{pmatrix}^T \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{x}_1-\boldsymbol{\mu}_1\\ \boldsymbol{x}_2-\boldsymbol{\mu}_2\end{pmatrix}\}\\
&= \exp\{-\frac{1}{2}\begin{pmatrix}\boldsymbol{x}_1-\boldsymbol{\mu}_1\\ \boldsymbol{x}_2-\boldsymbol{\mu}_2\end{pmatrix}^T \begin{pmatrix}\boldsymbol{I} & 0\\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{I}\end{pmatrix}\begin{pmatrix}(\boldsymbol{\Sigma}/\boldsymbol{\Sigma}_{22})^{-1} & 0\\ 0 & \boldsymbol{\Sigma}_{22}^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\\ 0 & \boldsymbol{I}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_1-\boldsymbol{\mu}_1\\ \boldsymbol{x}_2-\boldsymbol{\mu}_2\end{pmatrix}\}\\
&= \exp\{-\frac{1}{2}\begin{pmatrix}\boldsymbol{x}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)\\ \boldsymbol{x}_2-\boldsymbol{\mu}_2\end{pmatrix}^T\begin{pmatrix}(\boldsymbol{\Sigma}/\boldsymbol{\Sigma}_{22})^{-1} & 0\\ 0 & \boldsymbol{\Sigma}_{22}^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_1-\boldsymbol{\mu}_1-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2)\\ \boldsymbol{x}_2-\boldsymbol{\mu}_2\end{pmatrix},
\end{aligned}
$$

where the first equality comes from eq. 6. By computing the quadratic form inside the exponential, we get

$$
\begin{aligned}
E =& \exp\{-\frac{1}{2}(\boldsymbol{x}_1-\boldsymbol{\mu}-\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}(\boldsymbol{x}_2-\boldsymbol{\mu}))(\boldsymbol{\Sigma}/\boldsymbol{\Sigma}_{22})^{-1}\times\\
&(\boldsymbol{x}_1-\boldsymbol{\mu}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}))\}\times\\
&\exp\{-\frac{1}{2}(\boldsymbol{x}_2-\boldsymbol{\mu})^T\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu})\}.
\end{aligned}
$$

From inspection, it is seen that the third line of the equation is the exponential term of $p(\boldsymbol{x}_2)$, while the first two lines consist $p(\boldsymbol{x}_1|\boldsymbol{x}_2)$. Since $\boldsymbol{\Sigma}_{22}$ and subsequently $\boldsymbol{\Sigma}_{22}^{-1}$ are positive semidefinite, and $(\boldsymbol{\Sigma}_{12})^T = \boldsymbol{\Sigma}_{21}$, we see the equality $(\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^T = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$. Hence, by inspection, we have

$$
\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2-\boldsymbol{\mu}_2), \tag{14}
$$

$$
\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{15}
$$

□

References

Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.