Conditional Gaussian and Marginal Gaussian

Xinghu Yao

September 22, 2018

1 Question

Given a joint Gaussian distribution $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Lambda} \equiv \boldsymbol{\Sigma}^{-1}$ and

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} \tag{1}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}, \quad \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{bb} \end{pmatrix} \tag{2}$$

Try yo derive the following conditional distribution and marginal distribution:

$$p(\mathbf{x}_a|\mathbf{x}_b) = \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}_{a|b}, \boldsymbol{\Lambda}_{aa}^{-1}\right) \tag{3}$$

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}\mathbf{x}_b - \boldsymbol{\mu}_b. \tag{4}$$

$$p(\mathbf{x}_n) = \mathcal{N}(\mathbf{x}_a|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa}). \tag{5}$$

2 Conditional Gaussian

According to the definition of condition distribution, we have $p(\mathbf{x}_a|\mathbf{x}_b) = \frac{p(\mathbf{x}_a, \mathbf{x}_b)}{p(\mathbf{x}_b)}$. Thus, through fixing $\mathbf{x}_b$ to the observed value and normalizing the resulting expression with $p(\mathbf{x}_b)$, we can get the conditional distribution $p(\mathbf{x}_a|\mathbf{x}_b)$.

We firstly consider the quadratic form in the exponent of the Gaussian distribution give by Eq. (1) and Eq. (2). In fact, we have

$$\begin{aligned}
&-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \\
&= -\frac{1}{2}\begin{pmatrix} \mathbf{x}_a - \boldsymbol{\mu}_a \\ \mathbf{x}_b - \boldsymbol{\mu}_b \end{pmatrix}^T \begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{pmatrix} \begin{pmatrix} \mathbf{x}_a - \boldsymbol{\mu}_a \\ \mathbf{x}_b - \boldsymbol{\mu}_b \end{pmatrix} \\
&= -\frac{1}{2}\left[(\mathbf{x}_a - \boldsymbol{\mu}_a)^T, (\mathbf{x}_b - \boldsymbol{\mu}_b)^T\right] \begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{pmatrix} \begin{pmatrix} \mathbf{x}_a - \boldsymbol{\mu}_a \\ \mathbf{x}_b - \boldsymbol{\mu}_b \end{pmatrix} \\
&= -\frac{1}{2}(\mathbf{x}_a - \boldsymbol{\mu}_a)^T \boldsymbol{\Lambda}_{aa}(\mathbf{x}_a - \boldsymbol{\mu}_a) - \frac{1}{2}(\mathbf{x}_a - \boldsymbol{\mu}_a)^T \boldsymbol{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b) \\
&-\frac{1}{2}(\mathbf{x}_b - \boldsymbol{\mu}_b)^T \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a) - \frac{1}{2}(\mathbf{x}_b - \boldsymbol{\mu}_b)^T \boldsymbol{\Lambda}_{bb}(\mathbf{x}_b - \boldsymbol{\mu}_b).
\end{aligned} \tag{6}$$

We see that as a function of $\mathbf{x}_a$, this is a quadratic form, thus the conditional distribution $p(\mathbf{x}_a|\mathbf{x}_b)$ is a Gaussian distribution. Noticing that the exponent of a general Gaussian distribution $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be written

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = -\frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} + \mathbf{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + const \tag{7}$$

where 'const' denotes terms which are independent of $\mathbf{x}$. Consider the functional dependence of Eq. (6) on $\mathbf{x}_b$ in which $\mathbf{x}_a$ is regarded as a constant. The second order of $\mathbf{x}_a$ can be written

$$-\frac{1}{2}\mathbf{x}_a^T\Lambda_{aa}\mathbf{x}_a \tag{8}$$

Comparing Eq. (7) and Eq. (8), we can immediately get the covariance matrix of $p\left(\mathbf{x}_a|\mathbf{x}_b\right)$ is given by

$$\boldsymbol{\Sigma}_{a|b} = \Lambda_{aa}^{-1}. \tag{9}$$

Now consider the linear form of $\mathbf{x}_a$ in Eq. (6)

$$\mathbf{x}_a^T \left\{\boldsymbol{\Lambda}_{aa}\boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{ab}\left(\mathbf{x}_b - \boldsymbol{\mu}_b\right)\right\} \tag{10}$$

Comparing Eq. (7) Eq. (10), we can get $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \mathbf{x}_a^T \left\{\boldsymbol{\Lambda}_{aa}\boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{ab}\left(\mathbf{x}_b - \boldsymbol{\mu}_b\right)\right\}$. Thus, we have

$$\begin{aligned}\boldsymbol{\mu}_{a|b} &= \boldsymbol{\Sigma}_{a|b}\left\{\boldsymbol{\Lambda}_{aa}\boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{ab}\left(\mathbf{x}_b - \boldsymbol{\mu}_b\right)\right\} \\ &= \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{ab}^{-1}\boldsymbol{\Lambda}_{ab}\left(\mathbf{x}_b - \boldsymbol{\mu}_b\right)\end{aligned} \tag{11}$$

Combing Eq. (9) and Eq. (11), we can get $p\left(\mathbf{x}_a|\mathbf{x}_b\right) = \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}_{a|b}, \boldsymbol{\Lambda}_{aa}^{-1}\right)$ where $\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}\mathbf{x}_b - \boldsymbol{\mu}_b$.

3 Marginal Gaussian

In fact, the Marginal Gaussian distribution can be written as

$$p\left(\mathbf{x}_a\right) = \int p\left(\mathbf{x}_a, \mathbf{x}_b\right) d\mathbf{x}_b \tag{12}$$

Picking out those terms only involve $\mathbf{x}_b$ in the joint distribution $p\left(\mathbf{x}_a, \mathbf{x}_b\right)$, we have

$$-\frac{1}{2}\mathbf{x}_b^T\boldsymbol{\Lambda}_{bb}\mathbf{x}_b + \mathbf{x}_b^T\mathbf{m} = -\frac{1}{2}\left(\mathbf{x}_b - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m}\right)^T\boldsymbol{\Lambda}_{bb}\left(\mathbf{x}_b - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m}\right) + \frac{1}{2}\mathbf{m}^T\boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m} \tag{13}$$

where $\mathbf{m} = \boldsymbol{\Lambda}_{bb}\boldsymbol{\mu}_b - \boldsymbol{\Lambda}_{ba}\left(\mathbf{x}_a - \boldsymbol{\mu}_a\right)$ Now we turn to consider the general Gaussian distribution, which is

$$\mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \frac{1}{\left(2\pi\right)^{D/2}}\frac{1}{\left|\boldsymbol{\Sigma}\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(\mathbf{x} - \boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{x} - \boldsymbol{\mu}\right)\right\} \tag{14}$$

From Eq. (14) we can see the coefficient of Gaussian distribution is independent of the mean and only governed by the determinant of the covariance matrix. Back to Eq. 13, we can see the integration over $\mathbf{x}$ is as follows

$$\int \exp\left\{-\frac{1}{2}\left(\mathbf{x}_b - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m}\right)^T\boldsymbol{\Lambda}_{bb}\left(\mathbf{x}_b - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m}\right)\right\} \mathrm{d}\mathbf{x}_b. \tag{15}$$

This integration is irrelevant with the mean so we can margin out $mathbfx_b$ after the integration.Combing the $\mathbf{x}_b^T\mathbf{m}$ in Eq. (13) with the remaining terms from Eq. (6), we have

$$\begin{aligned}&\frac{1}{2}\left[\boldsymbol{\Lambda}_{bb}\boldsymbol{\mu}_b - \boldsymbol{\Lambda}_{ba}\left(\mathbf{x}_a - \boldsymbol{\mu}_a\right)\right]^T\boldsymbol{\Lambda}_{bb}^{-1}\left[\boldsymbol{\Lambda}_{bb}\boldsymbol{\mu}_b - \boldsymbol{\Lambda}_{ba}\left(\mathbf{x}_a - \boldsymbol{\mu}_a\right)\right] - \frac{1}{2}\mathbf{x}_a^T\boldsymbol{\Lambda}_{aa}\mathbf{x}_a + \mathbf{x}_a^T\left(\boldsymbol{\Lambda}_{aa}\boldsymbol{\mu}_a + \boldsymbol{\Lambda}_{ab}\boldsymbol{\mu}_b\right) + \mathrm{const} \\ &= \frac{1}{2}\mathbf{x}_a^T\left(\boldsymbol{\Lambda}_{aa} - \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}\right)\mathbf{x}_a + \mathbf{x}_a^T\left(\boldsymbol{\Lambda}_{aa} - \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\right)^{-1}\boldsymbol{\mu}_a + \mathrm{const}\end{aligned} \tag{16}$$

where 'const' denotes quantities independent of $\mathbf{x}_a$. Comparing Eq. (16) with Eq. (7), we can get the covariance of $p\left(\mathbf{x}_a\right)$ is

$$\boldsymbol{\Sigma}_a = \left(\boldsymbol{\Lambda}_{aa} - \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}\right)^{-1} = \boldsymbol{\Sigma}_{aa} \tag{17}$$

and the mean of $p\left(\mathbf{x}_a\right)$ is given by

$$\boldsymbol{\mu}_a = \boldsymbol{\Sigma}_a\left(\boldsymbol{\Lambda}_{aa} - \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}\right)\boldsymbol{\mu}_a \tag{18}$$

We can summarized Eq. (17) and Eq. (18) as

$$p\left(\mathbf{x}_a\right) = \mathcal{N}\left(\mathbf{x}_a|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa}\right) \tag{19}$$