Marginal and Conditional Distributions of Multivariate Gaussians

Patrick Yin

May 29, 2020

1 Introduction

This note derives the marginal and conditional distributions of multivariate gaussians. This derivation requires knowledge of the Woodbury Formula and block-wise matrix inversion. The linear algebra can get rather complicated, but the essence of the derivation comes down to breaking down the joint distribution of the multivariate gaussian into the product of two multivariate normal distributions, one of which should be in terms of only one of the two subvectors that form the Jointly Gaussian Random Vector (i.e. Equation 17).

2 Lemma for Derivation

Let A be a symmetric matrix that can be expressed as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

such that A_{11} and A_{22} are invertible square matrices.

Lemma 1 (Determinant of block symmetric matrix)**.** *Following the assumption above,*

$$|A| = |A_{11}||A_{22} - A_{21}A_{11}^{-1}A_{12}| = |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}|$$

Proof. Via matrix multiplication, we can see that

$$A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & I \end{bmatrix} \begin{bmatrix} I & A_{11}^{-1}A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix} = \begin{bmatrix} I & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & 0 \\ A_{22}^{-1}A_{21} & I \end{bmatrix}$$

Since $|AB| = |A||B|$ and the determinant of a triangular matrix is the product of the determinants of its two diagonal blocks, we have proved the lemma. $\square$

3 Derivation of Marginal and Conditional Distributions of Multivariate Gaussian

Let X be an n-dimensional Jointly Gaussian (JG) Random Vector such that

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N(\mu, \Sigma) \tag{1}$$

where

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

X_1 is a n_1-dimensional JG Random Vector and X_2 is a n_2-dimensional JG Random Vector such that $n_1 + n_2 = n$.

Theorem 2 (Marginal Distribution of Multivariate Gaussian)**.**

$$X_1 \sim N(\mu_1, \Sigma_{11}) \tag{2}$$

$$X_2 \sim N(\mu_2, \Sigma_{22}) \tag{3}$$

Theorem 3 (Conditional Distribution of Multivariate Gaussian)**.**

$$X_1|X_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \tag{4}$$

$$X_2|X_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}) \tag{5}$$

Note that since Σ is symmetric, $\Sigma_{12} = \Sigma_{21}^T$.

Proof. The joint density of X is:

$$f_X(x) = f_{X_1,X_2}(x_1, x_2) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)} \tag{6}$$

We can break down the exponential term further:

$$\begin{aligned}(X-\mu)^T\Sigma^{-1}(X-\mu) &= \begin{bmatrix} X_1 - \mu_1 & X_2 - \mu_2 \end{bmatrix} \begin{bmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix} \\ &= (X_1 - \mu_1)^T\Sigma_{11}^*(X_1 - \mu_1) \\ &\quad + 2(X_1 - \mu_1)^T\Sigma_{12}^*(X_2 - \mu_2) \\ &\quad + (X_2 - \mu_2)^T\Sigma_{22}^*(X_2 - \mu_2)\end{aligned} \tag{7}$$

such that

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{bmatrix} \tag{8}$$

Applying block matrix inverse and Woodbury Formula, we find that:

$$\begin{aligned}\Sigma_{11}^* &= (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} \\ &= \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{21}\Sigma_{11}^{-1}\end{aligned} \tag{9}$$

$$
\begin{aligned}
\Sigma_{22}^* &= (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \\
&= \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}
\end{aligned} \tag{10}
$$

$$
\Sigma_{12}^* = -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} = (\Sigma_{21}^*)^T \tag{11}
$$

For the last equation, $\Sigma_{12}^* = (\Sigma_{21}^*)^T$ because applying block matrix inverse for the $(\Sigma_{21}^*)^T$ term of $(\Sigma^T)^{-1}$ will result in the same expression as the block matrix inverse for the Σ_{12}^* term of $(\Sigma)^{-1}$. Since $\Sigma^T = \Sigma$, $\Sigma_{12}^* = (\Sigma_{21}^*)^T$.

Let $A := \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Plugging in our derived values for Σ_{11}^*, Σ_{12}^*, and Σ_{22}^* into equation 7, we have:

$$
\begin{aligned}
(X-\mu)^T\Sigma^{-1}(X-\mu) &= (X_1-\mu_1)^T(\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}A^{-1}\Sigma_{21}\Sigma_{11}^{-1})(X_1-\mu_1) \\
&\quad - 2(X_1-\mu_1)^T(\Sigma_{11}^{-1}\Sigma_{12}A^{-1})(X_2-\mu_2) \\
&\quad + (X_2-\mu_2)^T(A^{-1})(X_2-\mu_2) \\
&= (X_1-\mu_1)^T\Sigma_{11}^{-1}(X_1-\mu_1) \\
&\quad + (X_1-\mu_1)^T\Sigma_{11}^{-1}\Sigma_{12}A^{-1}\Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1) \\
&\quad - 2(X_1-\mu_1)^T(\Sigma_{11}^{-1}\Sigma_{12}A^{-1})(X_2-\mu_2) \\
&\quad + (X_2-\mu_2)^T(A^{-1})(X_2-\mu_2)
\end{aligned} \tag{12}
$$

We want to simplify equation 12 further. Note that the latter 3 terms of equation 12 all have common components. Let $u^T = (X_1-\mu_1)^T\Sigma_{11}^{-1}\Sigma_{12}$, $M = A^{-1}$, and $v = X_2 - \mu_2$. Then the latter 3 terms of equation 12 break down as such:

$$
\begin{aligned}
&(X_1-\mu_1)^T\Sigma_{11}^{-1}\Sigma_{12}A^{-1}\Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1) \\
&- 2(X_1-\mu_1)^T(\Sigma_{11}^{-1}\Sigma_{12}A^{-1})(X_2-\mu_2) + (X_2-\mu_2)^T(A^{-1})(X_2-\mu_2) \\
&= u^TMu - 2u^TMv + v^TMv \\
&= u^TMu - u^TMv - u^TMv + v^TMv \\
&= u^TM(u-v) - (u-v)^TMv \\
&= u^TM(u-v) - v^TM(u-v) \\
&= (u-v)^TM(u-v) \\
&= (v-u)^TM(v-u) \\
&= ((X_2-\mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1))^TA^{-1}((X_2-\mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1))
\end{aligned} \tag{13}
$$

Plugging in what we got in equation 13 back into equation 12, we get

$$
\begin{aligned}
(X-\mu)^T\Sigma^{-1}(X-\mu) &= (X_1-\mu_1)^T\Sigma_{11}^{-1}(X_1-\mu_1) \\
&\quad + ((X_2-\mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1))^T \\
&\quad A^{-1}((X_2-\mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1))
\end{aligned} \tag{14}
$$

Let $b := \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)$, then

$$\begin{aligned}(X-\mu)^T\Sigma^{-1}(X-\mu) = (X_1-\mu_1)^T\Sigma_{11}^{-1}(X_1-\mu_1) \\ + (X_2-b)^T A^{-1}(X_2-b)\end{aligned} \tag{15}$$

For clarity, let us define

$$\begin{cases} g_1(X_1) := (X_1-\mu_1)^T\Sigma_{11}^{-1}(X_1-\mu_1) \\ g_2(X_1, X_2) := (X_2-b)^T A^{-1}(X_2-b) \end{cases} \tag{16}$$

Then, the joint density of X can be broken down as such:

$$\begin{aligned} f_X(x) &= f_{X_1,X_2}(x_1, x_2) \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)} \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(g_1(x_1)+g_2(x_1,x_2))} \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma_{11}|^{1/2}|A|^{1/2}} e^{-\frac{1}{2}(g_1(x_1)+g_2(x_1,x_2))} \\ &= \frac{1}{(2\pi)^{n_1/2}|\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}g_1(x_1)} \frac{1}{(2\pi)^{n_2/2}|A|^{1/2}} e^{-\frac{1}{2}g_2(x_1,x_2)} \\ &= \frac{1}{(2\pi)^{n_1/2}|\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}(x_1-\mu_1)^T\Sigma_{11}^{-1}(x_1-\mu_1)} \\ &\quad \frac{1}{(2\pi)^{n_2/2}|A|^{1/2}} e^{-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)} \\ X &\sim N(\mu_1, \Sigma_{11}) N(b, A) \end{aligned} \tag{17}$$

In the fourth line, $|\Sigma| = |\Sigma_{11}||A|$ is a direct consequence of lemma 1.

The marginal distribution of X_1 is

$$\begin{aligned} f_{X_1}(x_1) &= \int f_{X_1,X_2}(x_1, x_2) dx_2 \\ &= \frac{1}{(2\pi)^{n_1/2}|\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}(x_1-\mu_1)^T\Sigma_{11}^{-1}(x_1-\mu_1)} \\ &\quad \int \frac{1}{(2\pi)^{n_2/2}|A|^{1/2}} e^{-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)} dx_2 \\ &= \frac{1}{(2\pi)^{n_1/2}|\Sigma_{11}|^{1/2}} e^{-\frac{1}{2}(x_1-\mu_1)^T\Sigma_{11}^{-1}(x_1-\mu_1)} \\ X_1 &\sim N(\mu_1, \Sigma_{11}) \end{aligned} \tag{18}$$

The conditional distribution of $X_2|X_1$ is

$$
\begin{aligned}
f_{X_2|X_1}(x_2|x_1) &= \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)} \\
&= \frac{1}{(2\pi)^{n_2/2}|A|^{1/2}} e^{-\frac{1}{2}(x_2-b)^T A^{-1}(x_2-b)} \\
X_2|X_1 &\sim N(b,A) \\
&\sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1-\mu_1), \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})
\end{aligned}
\tag{19}
$$

The marginal distribution of X_2 and conditional distribution of $X_1|X_2$ can be derived by writing equation 12 in terms of $(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$ instead of $(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}$ and following the same derivation procedure. □