Chapter 1 Conditionals and Marginals

Monday, February 4, 2019 4:35 PM

Suppose that $x \sim N(\mu, \Sigma)$ has a multivariate normal distribution. Let x_1 comprise the first k elements of x, and x_2 the last $p-k$. We will assume that μ and Σ have been partitioned comfortably with x:

$$\mu = (\mu_1, \mu_2)^T \quad \text{and } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\Sigma_{21} = \Sigma_{12}^T$ as Σ is a symmetric matrix.

A) Define the marginal distribution of x_1. (Remember your result about affine transformations).

Let $\underset{\sim}{x} \sim N(\mu, \Sigma)$ where $\underset{\sim}{x} = [x_1, \dots, x_p]^T$. Let $\underset{\sim}{x}_1 = [x_1, \dots, x_k]^T$ and $\underset{\sim}{x}_2 = [x_{k+1}, \dots x_p]^T$. Since $\underset{\sim}{x}$ is MVN, it can be written as an affine transformation

$$\underset{\sim}{x} = L\underset{\sim}{z} + \mu \quad \text{where } \underset{\sim}{z} \sim MVN(0, I)$$

Define a matrix $A_{p \times p}$

$$A = \begin{bmatrix} I_{k \times k} & 0_{k \times (p-k)} \\ 0_{(p-k) \times k} & 0_{(p-k) \times (p-k)} \end{bmatrix}$$

Then $A\underset{\sim}{x} = \underset{\sim}{x}_1$. Thus $\underset{\sim}{x}_1$ is multivariate normal since it can be written as an affine transformation

$$\begin{aligned} \underset{\sim}{x}_1 &= AL\underset{\sim}{z} + A\mu \\ &= ALz + \mu_1 \end{aligned}$$

The mean and covariance of $\underset{\sim}{x}_1$ is

$$\begin{aligned} E(\underset{\sim}{x}_1) &= E(A\underset{\sim}{x}) \\ &= A\mu \\ &= \mu_1 \end{aligned} \qquad \begin{aligned} Cov(A\underset{\sim}{x}) &= A\,Cov(\underset{\sim}{x})A^T \\ &= A\Sigma A^T \\ &= \Sigma_{11} \end{aligned}$$

b) Let $\Omega = \Sigma^{-1}$ be the inverse covariance matrix, or precision matrix of x, and partition Ω just as you did Σ:

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^T & \Omega_{22} \end{pmatrix}$$

Using identities for the inverse of a partitioned matrix, express each block of Ω in terms of blocks of Σ.

We can start with the inverse partition matrix identity

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{bmatrix}$$

$$\text{Let } \Omega = \Sigma^{-1} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12}^{T} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} \overset{A}{\Sigma_{11}} & \overset{B}{\Sigma_{12}} \\ \underset{C}{\Sigma_{21}} & \underset{D}{\Sigma_{22}} \end{pmatrix}^{-1}$$

$$\text{Then } \Omega_{11} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{21}\Sigma_{11}^{-1}$$

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

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

Alternatively,

We have that $\Omega\Sigma = \Sigma\Omega = I$

$$\text{Therefore: } \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = I$$

$$\begin{pmatrix} \Omega_{11}\Sigma_{11} + \Omega_{12}\Sigma_{21} & \Omega_{11}\Sigma_{12} + \Omega_{12}\Sigma_{22} \\ \Omega_{21}\Sigma_{11} + \Omega_{22}\Sigma_{21} & \Omega_{21}\Sigma_{12} + \Omega_{22}\Sigma_{22} \end{pmatrix} = \begin{bmatrix} I_{k\times k} & O_{k\times(p-k)} \\ O_{(p-k)\times k} & I_{(p-k)\times(p-k)} \end{bmatrix}$$

$$\begin{aligned} \Omega_{11}\Sigma_{12} + \Omega_{12}\Sigma_{22} &= O_{k+(p-k)} \\ \Omega_{21}\Sigma_{11} + \Omega_{22}\Sigma_{21} &= O_{(p-k)\times k} \\ \Omega_{11}\Sigma_{11} + \Omega_{12}\Sigma_{21} &= I_{k\times k} \\ \Omega_{21}\Sigma_{12} + \Omega_{22}\Sigma_{22} &= I_{(p-k)\times(p-k)} \end{aligned}$$

Solving for Ω_{12}, we have

$$\begin{aligned} \Omega_{11}\Sigma_{12} + \Omega_{12}\Sigma_{22} &= O_{k\times(p-k)} \\ \Omega_{12}\Sigma_{22} &= -\Omega_{11}\Sigma_{12} \\ \Omega_{12} &= -\Omega_{11}\Sigma_{12}\Sigma_{22}^{-1} \end{aligned}$$

We can plug in this value of Ω_{12} into $\Omega_{11}\Sigma_{11} + \Omega_{12}\Sigma_{21} = I_{k\times k}$ to solve for Ω_{12}

$$\Omega_{11}\Sigma_{11} + (-\Omega_{11}\Sigma_{12}\Sigma_{22}^{-1})\Sigma_{21} = I_{k\times k}$$

$$\Omega_{11}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) = I$$

$$\Omega_{11}\Sigma_{11} + (-\Omega_{11}\Sigma_{12}\Sigma_{22}^{-1})\Sigma_{21} = I_{k\times k}$$
$$\Omega_{11}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) = I_{k\times k}$$
$$\Rightarrow \quad \Omega_{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} \quad \text{and} \quad \Omega_{12} = -(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}$$

Similarly, we can solve for Ω_{22},

$$\text{First,} \quad \Omega_{21} = \Omega_{12}^T = -\Sigma_{22}^{-1}\Sigma_{12}^T(\Sigma_{11}^T - \Sigma_{21}^T(\Sigma_{22}^{-1})^T\Sigma_{12}^T)^{-1} \quad \text{using } (A^{-1})^T = (A^T)^{-1}$$
$$= -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$$

$$\Omega_{21}\Sigma_{12} + \Omega_{22}\Sigma_{21} = I_{(p-k)\times(p-k)}$$
$$\Omega_{22}\Sigma_{21} = I - \Omega_{21}\Sigma_{12}$$
$$\Omega_{22} = \Sigma_{21}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{21}^{-1}$$

$$\text{Thus,} \quad \Omega = \begin{bmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & -(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & \Sigma_{21}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{21}^{-1} \end{bmatrix}$$

c) Derive the conditional distribution for x_1, given x_2, in terms of the partitioned elements of x, μ, and Σ. Hints: work with densities on a log scale, ignore constants that don't affect x_1, and remember complete the square. Explain briefly how one may interpret this conditional distribution as a linear regression on x_2, where the regression matrix can be read off the precision matrix.

The definition of conditional distribution is

$$p(\underset{\sim}{x}_1 \mid \underset{\sim}{x}_2) = \frac{p(\underset{\sim}{x}_1, \underset{\sim}{x}_2)}{p(\underset{\sim}{x}_2)}$$

The joint density of $\underset{\sim}{x}_2$ is just the density of $\underset{\sim}{x}$. We can find the marginal distribution of $\underset{\sim}{x}_2$ using the same method as in part a). This time, define the matrix A to be

$$A = \begin{bmatrix} 0_{k\times k} & 0_{k\times(p-k)} \\ 0_{(p-k)\times k} & I_{(p-k)\times(p-k)} \end{bmatrix}$$

Thus, $A\underset{\sim}{x} = \underset{\sim}{x}_2$ and $E(\underset{\sim}{x}_2) = \mu_2$ and $Cov(\underset{\sim}{x}_2) = \Sigma_{22}$. We can now solve for the conditional $p(x_1 \mid x_2)$. For the sake of computation, I will work with the log densities.

$$f(\underset{\sim}{x}) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left(-\tfrac{1}{2}(\underset{\sim}{x} - \mu)^T\Sigma^{-1}(\underset{\sim}{x} - \mu)\right)$$

$$f(x_1, x_2) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\begin{pmatrix}\underset{\sim}{x}_1 - \mu_1 \\ \underset{\sim}{x}_2 - \mu_2\end{pmatrix}^T \Sigma^{-1}\begin{pmatrix}\underset{\sim}{x}_1 - \mu_1 \\ \underset{\sim}{x}_2 - \mu_2\end{pmatrix}\right)$$

$$f(x_2) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left(-\tfrac{1}{2}(x_2 - \mu_2)^T\Sigma^{-1}(x_2 - \mu_2)\right)$$

$$f(x_1, x_2) = |2\pi\Sigma|^{1/2} \exp(-\tfrac{1}{2}(\underset{\sim}{x}_2-\mu_2) \Sigma (\underset{\sim}{x}_2-\mu_2))$$

$$f(x_2) = \frac{1}{|2\pi\Sigma_{22}|^{1/2}} \exp\left(-\frac{1}{2}(\underset{\sim}{x}_2-\mu_2)^T \Sigma_{22}^{-1} (\underset{\sim}{x}_2-\mu_2)\right)$$

$$\frac{f(x_1,x_2)}{f(x_2)} = \frac{\frac{1}{|2\pi\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\begin{pmatrix}\underset{\sim}{x}_1-\mu_1\\ \underset{\sim}{x}_2-\mu_2\end{pmatrix}^T \Sigma^{-1}\begin{pmatrix}\underset{\sim}{x}_1-\mu_1\\ \underset{\sim}{x}_2-\mu_2\end{pmatrix}\right)}{\frac{1}{|2\pi\Sigma_{22}|^{1/2}} \exp\left(-\frac{1}{2}(\underset{\sim}{x}_2-\mu_2)^T \Sigma_{22}^{-1} (\underset{\sim}{x}_2-\mu_2)\right)}$$

$$\ln\left(\frac{f(x_1,x_2)}{f(x_2)}\right) = \ln(f(x_1,x_2)) - \ln(f(x_2))$$

$$= \underbrace{-\frac{1}{2}\ln(|2\pi\Sigma|)}_{c_1} - \frac{1}{2}\begin{pmatrix}\underset{\sim}{x}_1-\mu_1\\ \underset{\sim}{x}_2-\mu_2\end{pmatrix}^T \Sigma^{-1}\begin{pmatrix}\underset{\sim}{x}_1-\mu_1\\ \underset{\sim}{x}_2-\mu_2\end{pmatrix} \underbrace{- \frac{1}{2}\ln(|2\pi\Sigma_{22}|)}_{c_2}$$
$$- \frac{1}{2}(\underset{\sim}{x}_2-\mu_2)^T \Sigma_{22}^{-1} (\underset{\sim}{x}_2-\mu_2)$$

using $\Omega = \Sigma^{-1}$
$$= c_1 + c_2 - \frac{1}{2}\left[\begin{pmatrix}\overset{a}{\overbrace{\underset{\sim}{x}_1-\mu_1}}\\ \underbrace{\underset{\sim}{x}_2-\mu_2}_{b}\end{pmatrix}^T \Omega \begin{pmatrix}\overset{a}{\overbrace{\underset{\sim}{x}_1-\mu_1}}\\ \underbrace{\underset{\sim}{x}_2-\mu_2}_{b}\end{pmatrix} - (\underbrace{\underset{\sim}{x}_2-\mu}_{b})^T \Sigma_{22}^{-1} (\underbrace{\underset{\sim}{x}_2-\mu_2}_{b})\right]$$

$$= c_1 + c_2 - \frac{1}{2}\left[a^T\Omega_{11}a + b^T\Omega_{21}a + a^T\Omega_{12}b + \underbrace{b^T\Omega_{22}b - b^T\Sigma_{22}^{-1}b}_{c_3}\right]$$

Since $\Omega_{21}^T = \Omega_{12}$
$$= c_1 + c_2 - \frac{1}{2}\left[a^T\Omega_{11}a + 2a^T\Omega_{12}b + c_3\right]$$

$$= c_1 + c_2 - \frac{1}{2}\left[(x_1-\mu_1)^T\Omega_{11}(x_1-\mu_1) + 2(x_1-\mu_1)^T\Omega_{12}(\underset{\sim}{x}_2-\mu_2) + c_3\right]$$

At this point, we are trying to construct a density for $\underset{\sim}{x}_1$ so we can drop all the constants that do not contain $\underset{\sim}{x}_1$.

$$\ln(f(\underset{\sim}{x}_1|\underset{\sim}{x}_2)) \propto -\frac{1}{2}\left[(\underset{\sim}{x}_1-\mu_1)^T\Omega_{11}(\underset{\sim}{x}_1-\mu_1) + 2(\underset{\sim}{x}_1-\mu_1)^T\Omega_{12}(\underset{\sim}{x}_2-\mu_2)\right]$$

In order for this to be the kernel of a MVN density, I need it in the form
$$-\frac{1}{2}\left[(\underset{\sim}{x}_1-m)^T A (\underset{\sim}{x}_1-m)\right]$$

To solve for A and m, I will need to complete the square.

$$\ln(f(\underset{\sim}{x}_1|\underset{\sim}{x}_2)) \propto -\frac{1}{2}\left[(\underset{\sim}{x}_1-\mu_1)^T\Omega_{11}(\underset{\sim}{x}_1-\mu_1) + 2(\underset{\sim}{x}_1-\mu_1)^T\Omega_{12}(\underset{\sim}{x}_2-\mu_2)\right]$$

$$= -\frac{1}{2}\left[(\underset{\sim}{x}_1^T\Omega_{11} - \mu_1^T\Omega_{11})(\underset{\sim}{x}_1-\mu_1) + 2(\underset{\sim}{x}_1^T\Omega_{12} - \mu_1^T\Omega_{12})(\underset{\sim}{x}_2-\mu_2)\right]$$

$$= -\frac{1}{2}\left[\underset{\sim}{x}_1^T\Omega_{11}\underset{\sim}{x}_1 - 2\underset{\sim}{x}_1^T\Omega_{11}\mu_1 + \mu_1^T\mu_1 + 2(\underset{\sim}{x}_1^T\Omega_{12}\underset{\sim}{x}_2 - \underset{\sim}{x}_1^T\Omega_{12}\mu_2 - \mu_1^T\Omega_{12}\underset{\sim}{x}_2 + \mu_1^T\Omega_{12}\mu_2)\right]$$

Again, I will drop all constants,
$$\propto -\frac{1}{2}\left[\underset{\sim}{x}_1^T\Omega_{11}\underset{\sim}{x}_1 - 2\underset{\sim}{x}_1^T\Omega_{11}\mu_1 + 2(\underset{\sim}{x}_1^T\Omega_{12}\underset{\sim}{x}_2 - \underset{\sim}{x}_1^T\Omega_{12}\mu_2)\right]$$

$$= -\frac{1}{2}\left[\underset{\sim}{x}_1^T\Omega_{11}\underset{\sim}{x}_1 - 2\underset{\sim}{x}_1^T(\underbrace{\Omega_{11}\mu_1 + \Omega_{12}\underset{\sim}{x}_2 - \Omega_{12}\mu_2}_{b})\right]$$

$$= -\frac{1}{2}\left[\underset{\sim}{x}_1^T\Omega_{11}\underset{\sim}{x}_1 - 2b^T\underset{\sim}{x}_1\right]$$

$$= -\frac{1}{2}\left[x^T\Omega\, x - 2b^Tx + b^T\Omega^{-1}b - b^T\Omega^{-1}b\right]$$

$$= -\frac{1}{2}\left[\underset{\sim}{x}_1^T \Omega_{11} \underset{\sim}{x}_1 - 2b^T \underset{\sim}{x}_1\right]$$

$$= -\frac{1}{2}\left[\underset{\sim}{x}_1^T \Omega_{11} \underset{\sim}{x}_1 - 2b^T \underset{\sim}{x}_1 + b^T \Omega_{11}^{-1} b - b^T \Omega_{11}^{-1} b\right]$$

$$\propto -\frac{1}{2}\left[\underset{\sim}{x}_1^T \Omega_{11} \underset{\sim}{x}_1 - 2b^T \underset{\sim}{x}_1 + b^T \Omega_{11}^{-1} b\right]$$

$$= -\frac{1}{2}\left[(\underset{\sim}{x}_1 - \Omega_{11}^{-1} b)^T \Omega_{11} (\underset{\sim}{x}_1 - \Omega_{11}^{-1} b)\right]$$

$$\sim \text{Normal}\left(\Omega_{11}^{-1}(\Omega_{11}\mu_1 + \Omega_{12}\underset{\sim}{x}_2 - \Omega_{12}\mu_2),\ \Omega_{11}^{-1}\right)$$

$$= \text{Normal}\left(\mu_1 + \Omega_{11}^{-1}\Omega_{12}\underset{\sim}{x}_2 - \Omega_{11}^{-1}\Omega_{12}\mu_2,\ \Omega_{11}^{-1}\right)$$