Multivariate Normal Exercise
Solutions

March 2, 2014

In this exercise, we will show, for a simple case, that the conditional and marginal distributions of a multivariate normal distribution are themselves normal. We will consider the simple case when $\mu = 0$ and when we wish to condition or marginalize solely over the final dimension.

The inverse of the covariance matrix Σ is called the precision matrix, which we here give the name D:

$$D \equiv \Sigma^{-1}$$

First, we break the variable x and the precision matrix D into their block representations, where we split off only the last dimension:

$$x = \begin{bmatrix} x_a \\ \chi_b \end{bmatrix} \text{ and } D = \begin{bmatrix} D_{aa} & d \\ d^T & \delta \end{bmatrix}$$

Note that we are using the convention that capital letters are matrices, lower case latin letters are vectors, and lower case greek letters are scalars.

Conditioning

1. Show that
$$x^T \Sigma^{-1} x = x_a^T D_{aa} x_a + 2\chi_b x_a^T d + \delta \chi_b^2.$$
2. To condition over the final dimension, we simply hold χ_b constant. (Why?) With χ_b held constant, find constants μ and K such that
$$x^T \Sigma^{-1} x = (x_a - \mu)^T D_{aa} (x_a - \mu) + K.$$
3. Conclude that a multivariate gaussian with mean zero conditioned on the last dimension is itself an unnormalized multivariate gaussian with potentially non-zero mean.

Solution: *First, note that there was a typo in the original problem statement. The $-2\chi_b x_a^T d$ term should have been positive, as shown above.*

1. For those with no exposure to blocked matrix manipulations, the beauty of blocked matrices is that if the matrices are of the correct dimensions and blocked correctly, then multiplication and addition happen exactly as if it were a normal matrix. If this is not obvious to you, work out a couple 3x3 examples of the following calculations to convince yourself it is true. We proceed:

   By the symmetry of Σ, D must be symmetric. This should be intuitive, but for the vigilant who want proof (not required for this class): For any invertible matrix A, $I = (AA^{-1})^T = (A^{-1})^T A^T$ and similarly from the right side, so by def $(A^T)^{-1} = (A^{-1})^T \equiv A^{-T}$. Hence, by the symmetry of Σ, $I = \Sigma D = \Sigma^T D \Rightarrow D = \Sigma^{-T} = D^T$.

Hence, we are justified in our labeling above with d and d^T in their respective places, and

$$\begin{aligned}
x^T Dx &= \begin{bmatrix} x_a^T & \chi_b \end{bmatrix} \begin{bmatrix} D_{aa} & d \\ d^T & \delta \end{bmatrix} \begin{bmatrix} x_a \\ \chi_b \end{bmatrix} \\
&= \begin{bmatrix} x_a^T & \chi_b \end{bmatrix} \begin{bmatrix} D_{aa} x_a + \chi_b d \\ d^T x_a + \chi_b \delta \end{bmatrix} \\
&= x_a^T (D_{aa} x_a + \chi_b d) + \chi_b (d^T x_a + \chi_b \delta) \\
&= x_a^T D_{aa} x_a + 2\chi_b x_a^T d + \delta \chi_b^2
\end{aligned}$$

as was to be shown.

2. Expanding the rhs,

$$(x_a - \mu)^T D_{aa} (x_a - \mu) + K = x_a^T D_{aa} x_a - 2\mu^T D_{aa} x_a + \mu^T D_{aa} \mu + K$$

We now compare like terms with the result from part 1, using the symmetry of D_{aa}. The quadratic term in x_a is obvious. For the linear term in x_a:

$$\begin{aligned}
-2\mu^T D_{aa} x_a &= 2\chi_b d^T x_a \\
\Rightarrow -2\mu^T D_{aa} &= 2\chi_b d^T \\
\Rightarrow \mu &= -\chi_b D_{aa}^{-1} d
\end{aligned}$$

which gives the value of μ. For the constant term in x_a:

$$\begin{aligned}
\delta \chi_b^2 &= \mu^T D_{aa} \mu + K \\
&= (-\chi_b D_{aa}^{-1} d)^T D_{aa} (-\chi_b D_{aa}^{-1} d) + K \\
&= \chi_b^2 d^T D_{aa}^{-T} d + K \\
\Rightarrow K &= \delta \chi_b^2 - \chi_b^2 d^T D_{aa}^{-T} d.
\end{aligned}$$

which is indeed constant with respect to x_a, so we are done. Notice that what we have done here is simply completing the square for the result of part 1.

3. The formula for the pdf $p(x)$ of an arbirary vector $x \in R^n$ for a multivariate normal distribution with mean zero is given by:

$$p(x) = \frac{1}{(2\pi)^d |\Sigma|^{1/2}} e^{-\frac{1}{2} x^T D x}$$

We have just shown that holding χ_b constant (conditioning on χ_b) reduces this to

$$p(x_a) = \frac{e^{-K/2}}{(2\pi)^d |\Sigma|^{1/2}} e^{-\frac{1}{2}(x_a - \mu)^T D_{aa} (x_a - \mu)}$$

which is an (unnormalized) multivariate gaussian of $x_a \in \mathbb{R}^{n-1}$ with mean μ which is only zero in special cases.

Marginalizing

1. To marginalize out the final dimension, we integrate over χ_b:

$$p(x_a) = \int_{-\infty}^{\infty} p(x)\, d\chi_b.$$

Perform this integral and conclude that the marginalization of a multivariate gaussian is itself a multivariate gaussian with the same mean.

2. Argue that the marginalized gaussian is already normalized.

Solution:

1.

$$\begin{aligned}
p(x_a) &= \int_{-\infty}^{\infty} p(x)\, d\chi_b \\
&= \int_{-\infty}^{\infty} \frac{1}{(2\pi)^d |\Sigma|^{1/2}} e^{-\frac{1}{2}x^T D x}\, d\chi_b \\
&= \frac{1}{(2\pi)^d |\Sigma|^{1/2}} \int_{-\infty}^{\infty} e^{x_a^T D_{aa} x_a + 2\chi_b x_a^T d + \delta \chi_b^2}\, d\chi_b \\
&= \frac{1}{(2\pi)^d |\Sigma|^{1/2}} e^{x_a^T D_{aa} x_a} \int_{-\infty}^{\infty} e^{2\chi_b x_a^T d + \delta \chi_b^2}\, d\chi_b
\end{aligned}$$

and, completing the square with respect to χ_b,

$$\begin{aligned}
&= \frac{1}{(2\pi)^d |\Sigma|^{1/2}} e^{x_a^T D_{aa} x_a} \int_{-\infty}^{\infty} e^{\delta(\chi_b^2 + 2\chi_b \delta^{-1} x_a^T d + (\delta^{-1} x_a^T d)^2 - (\delta^{-1} x_a^T d)^2)}\, d\chi_b \\
&= \frac{1}{(2\pi)^d |\Sigma|^{1/2}} e^{x_a^T D_{aa} x_a - \delta^{-1}(x_a^T d)^2)} \int_{-\infty}^{\infty} e^{\delta(\chi_b + \delta^{-1} x_a^T d)^2}\, d\chi_b
\end{aligned}$$

But it is a well known result that for any b,

$$\int_{-\infty}^{\infty} e^{a(\chi-b)^2}\, d\chi = \sqrt{\frac{\pi}{a}}$$

and furthermore,

$$\delta^{-1}(x_a^T d)^2 = \delta^{-1} x_a^T d d^T x_a = x_a^T \left(\frac{dd^T}{\delta}\right) x_a.$$

So finally,

$$p(x_a) = \frac{\sqrt{\pi/\delta}}{(2\pi)^d |\Sigma|^{1/2}} e^{x_a^T \left(D_{aa} - \frac{dd^T}{\delta}\right) x_a}$$

One more thing. We mentioned in class that when marginalizing over a variable, you simply remove the rows and columns of the Σ matrix and then invert. Represent Σ in its block form:

$$\Sigma = \begin{bmatrix} \Sigma_{aa} & s \\ s^T & \sigma \end{bmatrix}$$

It turns out that the $D_{aa} - \frac{dd^T}{\delta}$ found above is referred to as the Schur complement of D, and one significant reason this is interesting is that we have precisely that

$$\Sigma_{aa}^{-1} = D_{aa} - \frac{dd^T}{\delta},$$

as claimed during class.

2. This is obvious by the definition of marginalization:

$$\int_{\mathbb{R}^{n-1}} p(x_a)\, dx_a = \int_{\mathbb{R}^{n-1}} \int_{-\infty}^{\infty} p(x)\, d\chi_b\, dx_a = \int_{\mathbb{R}^n} p(x)\, dx = 1$$

since the original distribution was normalized.