Bayes' theorem for Gaussian variables

Xinghu Yao

September 22, 2018

1 Question

Given a marginal Gaussian distribution for $\mathbf{x}$ and a conditional Gaussian distribution for $\mathbf{y}$ given $\mathbf{x}$ in the form

$$p(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}\right) \tag{1}$$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}\left(\mathbf{y}|\mathbf{A}\mathbf{x}+\mathbf{b}, \mathbf{L}^{-1}\right) \tag{2}$$

the marginal distribution of $\mathbf{y}$ and the conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ are given by

$$p(\mathbf{y}) = \mathcal{N}\left(\mathbf{y}|\mathbf{A}\boldsymbol{\mu}+\mathbf{b}, \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{T}\right) \tag{3}$$

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}\left(\mathbf{x}|\boldsymbol{\Sigma}\left\{\mathbf{A}^{T}\mathbf{L}\mathbf{y}-\mathbf{b}+\boldsymbol{\Lambda}\boldsymbol{\mu}\right\}, \boldsymbol{\Sigma}\right) \tag{4}$$

where

$$\boldsymbol{\Sigma} = \left(\boldsymbol{\Lambda}+\mathbf{A}^{T}\mathbf{L}\mathbf{A}\right)^{-1} \tag{5}$$

2 Solution

First, it is easy (just follow the definition of matrix multiplication) to prove the following equation for the inverse of a partitioned matrix

$$\left(\begin{array}{cc} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{array}\right) = \left(\begin{array}{cc} \mathbf{M} & -\mathbf{M}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{M} & \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{C}\mathbf{M}\mathbf{B}\mathbf{N}^{-1} \end{array}\right) \tag{6}$$

where we have defined $\mathbf{M} = (\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}$.
We define $\mathbf{z} = (\mathbf{x}, \mathbf{y})^{T}$ as the joint distribution over $\mathbf{x}$ and $\mathbf{y}$ and consider the log of $\mathbf{z}$

$$\begin{aligned} \ln p(\mathbf{z}) &= \ln p(\mathbf{x}) + \ln p(\mathbf{y}|\mathbf{x}) \\ &= -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{T})\boldsymbol{\Lambda}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{1}{2}(\mathbf{y}-\mathbf{A}\mathbf{x}-\mathbf{b})^{-1}\mathbf{L}(\mathbf{y}-\mathbf{A}\mathbf{x}-\mathbf{b}) + \text{const} \end{aligned} \tag{7}$$

The second order terms in Eq. 7 can be written as

$$\begin{aligned} &-\frac{1}{2}\boldsymbol{x}^{\boldsymbol{T}}(\boldsymbol{\Lambda}+\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{L}\boldsymbol{A})\boldsymbol{x} - \frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{L}\mathbf{y} + \frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{L}\mathbf{A}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}\mathbf{L}\mathbf{y} \\ &= -\frac{1}{2}\left(\begin{array}{c} \mathbf{x} \\ \mathbf{y} \end{array}\right)^{T}\left(\begin{array}{cc} \boldsymbol{\Lambda}+\mathbf{A}^{\mathbf{T}}\mathbf{L}\mathbf{A} & -\mathbf{A}^{\mathbf{T}}\mathbf{L} \\ -\mathbf{L}\mathbf{A} & \mathbf{L} \end{array}\right)\left(\begin{array}{c} \mathbf{x} \\ \mathbf{y} \end{array}\right) \\ &= -\frac{1}{2}\mathbf{z}^{T}\mathbf{R}\mathbf{z} \end{aligned} \tag{8}$$

The linear terms in Eq. 7 can be written as

$$\mathbf{x}^{T}\boldsymbol{\Lambda}\boldsymbol{\mu} - \mathbf{x}^{T}\mathbf{A}^{T}\mathbf{L}\mathbf{b} + \mathbf{y}^{T}\mathbf{L}\mathbf{b} = \left(\begin{array}{c} \mathbf{x} \\ \mathbf{y} \end{array}\right)^{T}\left(\begin{array}{c} \mathbf{A}\boldsymbol{\mu}-\mathbf{A}^{T}\mathbf{L}\mathbf{b} \\ \mathbf{L}\mathbf{b} \end{array}\right). \tag{9}$$

Thus, the inverse matrix of covariance matrix can be written as

$$\mathbf{R} = \left(\begin{array}{cc} \boldsymbol{\Lambda}+\mathbf{A}^{\mathbf{T}}\mathbf{L}\mathbf{A} & -\mathbf{A}^{\mathbf{T}}\mathbf{L} \\ -\mathbf{L}\mathbf{A} & \mathbf{L} \end{array}\right) \tag{10}$$

And using the Eq. 6, we have

$$\operatorname{cov}[\mathbf{z}] = \mathbf{R}^{-1} = \begin{pmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{\Lambda}^{-1}\mathbf{A}^T \\ \mathbf{A}\boldsymbol{\Lambda}^{-1} & \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^T \end{pmatrix}. \tag{11}$$

Similar to the former report, we have

$$\mathbb{E}[\mathbf{z}] = \mathbf{R}^{-1} \begin{pmatrix} \mathbf{A}\boldsymbol{\mu} - \mathbf{A}^T\mathbf{L}\mathbf{b} \\ \mathbf{L}\mathbf{b} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{A}\boldsymbol{\mu} + \mathbf{b} \end{pmatrix}. \tag{12}$$

Making use of Eq. 11 and Eq. 12, we can get the mean and covariance of the marginal distribution $p(\mathbf{y})$, which is

$$\mathbb{E}[\mathbf{y}] = \mathbf{A}\boldsymbol{\mu} + \mathbf{b} \tag{13}$$

$$\operatorname{cov}[\mathbf{y}] = \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^T. \tag{14}$$

This means $p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^T)$. By using the last reports results, we have

$$\mathbb{E}[\mathbf{x}|\mathbf{y}] = (\boldsymbol{\Lambda} + \mathbf{A}^{\mathbf{T}}\mathbf{L}\mathbf{A}^{-1}) \left\{\mathbf{A}^T\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\right\} \tag{15}$$

$$\operatorname{cov}[\mathbf{x}|\mathbf{y}] = \boldsymbol{\Lambda} + \mathbf{A}^{\mathbf{T}}\mathbf{L}\mathbf{A}^{-1}. \tag{16}$$

This means $p(\mathbf{x}|\mathbf{y}) = \mathcal{N}\left(\mathbf{x}|\left(\boldsymbol{\Lambda} + \mathbf{A}^{\mathbf{T}}\mathbf{L}\mathbf{A}\right)^{-1}\left\{\mathbf{A}^T\mathbf{L}(\mathbf{y} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\mu}\right\}, \left(\boldsymbol{\Lambda} + \mathbf{A}^{\mathbf{T}}\mathbf{L}\mathbf{A}\right)^{-1}\right)$.