Linear Transformation of Multivariate Normal Distribution: Marginal, Joint and Posterior

Li-Ping Liu
EECS, Oregon State University
Corvallis, OR 97330
liuli@eecs.oregonstate.edu

Abstract

Suppose $\mathbf{x}$ is from normal distribution $\mathcal{N}(\mu_x, \Sigma_x)$ and $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, where $\mathbf{b}$ is from $\mathcal{N}(0, \Sigma_b)$. In this note, we show that the joint distribution of $(\mathbf{x}^T, \mathbf{y}^T)^T$, marginal distribution $\mathbf{y}$ and the posterior distribution $\mathbf{x}|\mathbf{y}$. These distributions play key roles in analysis of Kalman filter. If we let $A = I$, then the calculation in this notes also apply for the Baysian analysis of the mean of normal distribution. Instead of calculating integrals, we uses several special properties of normal distribution to make the derivation.

1 Linear transform of random variable from normal distribution

Suppose $\mathbf{x} \sim \mathcal{N}(\mu_x, \Sigma_x)$ and $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, where $\mathbf{b} \sim \mathcal{N}(0, \Sigma_b)$.

Since $\mathbf{x}$ and $\mathbf{b}$ is from normal distribution, $\mathbf{y}$ and $(\mathbf{x}^T, \mathbf{y}^T)^T$ are also from normal distribution. To find parameters of these normal distributions, we only need to find its mean and the variance.

We first calculate the marginal distribution of $\mathbf{y}$. The mean μ_y and variance Σ_y of $\mathbf{y}$ are

$$\begin{aligned} \mu_y &= E[\mathbf{y}] = E[A\mathbf{x} + \mathbf{b}] = AE[\mathbf{x}] + E[\mathbf{b}] = A\mu_x, && (1) \\ \Sigma_y &= \text{Var}(A\mathbf{x} + \mathbf{b}) = \text{Var}(A\mathbf{x}) + \text{Var}(\mathbf{b}) = A\Sigma_x A^T + \Sigma_b, && (2) \end{aligned}$$

so $\mathbf{y} \sim \mathcal{N}(A\mu_x, A\Sigma_x A^T + \Sigma_b)$.

The joint distribution of $(\mathbf{x}^T, \mathbf{y}^T)^T$ is also normal distribution. Its mean is

$$\mu_{xy} = \begin{pmatrix} \mu_x \\ A\mu_x \end{pmatrix}. \quad (3)$$

The covariance of $\mathbf{x}$ and $\mathbf{y}$ is

$$\begin{aligned} \text{Cov}(\mathbf{x}, \mathbf{y}) &= E[\mathbf{x}\mathbf{y}^T] - E[\mathbf{x}]E[\mathbf{y}^T] \\ &= E[\mathbf{x}\mathbf{x}^T A^T + \mathbf{x}\mathbf{b}^T] - \mu_x \mu_x^T A^T \\ &= E[\mathbf{x}\mathbf{x}^T]A^T - \mu_x \mu_x^T A^T \\ &= \Sigma_x A^T. && (4) \end{aligned}$$

In this derivation, we used the fact that the covariance between $\mathbf{x}$ and $\mathbf{b}$ are 0. The variance matrix of $(\mathbf{x}^T, \mathbf{y}^T)^T$ is

$$\begin{aligned} \Sigma_{xy} &= \begin{bmatrix} \text{Var}(\mathbf{x}) & \text{Cov}(x, y) \\ \text{Cov}(y, x) & \text{Var}(\mathbf{y}) \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_x & \Sigma_x A^T \\ A\Sigma_x & A\Sigma_x A^T + \Sigma_b \end{bmatrix}. && (5) \end{aligned}$$

Then we show the conditional distribution $(\mathbf{x}|y)$. Apply Thereom b in [1], we get the parameter of the conditional distribution $\mathbf{x}|\mathbf{y}$ as below.

$$\mu_{x|y} = \mu_x + \Sigma_x A^T (A \Sigma_x A^T + \Sigma_b)^{-1} (\mathbf{y} - A\mu_x) \tag{6}$$

$$\Sigma_{x|y} = \Sigma_x - \Sigma_x A^T (A \Sigma_x A^T + \Sigma_b)^{-1} A \Sigma_x \tag{7}$$

References

[1] http://fourier.eng.hmc.edu/e161/lectures/gaussianprocess/node7.html