0.1 Marginal probability and Conditional probability

0.1.1 Abstract

In this section, we study the marginal probability and conditional probability of multivariate Gaussian distribution

0.1.2 Prior Knowledge

In the previous chapter, we derived the probability density function of multivariate Gaussian distribution:

$$x \backsim N(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$$

Now let's split the random variable x into two parts:

$$x \in \mathcal{R}^p \quad x_a \in \mathcal{R}^m \quad x_b \in \mathcal{R}^n \quad m+n=p$$

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

0.1.3 Theorem

$$X \sim N(\mu, \Sigma) \quad Y = AX + B \Longrightarrow Y \sim N(A\mu + B, A\Sigma A^T)$$

0.1.4 Derive Marginal Probability

$$\begin{aligned}
x_a &= \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} x_a \\ x_b \end{pmatrix} + 0 \\
E[x_a] &= \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} = \mu_a \\
Var[x_a] &= \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix} \\
&= \Sigma_{aa} \\
&\therefore x_a \sim N(\mu_a, \Sigma_{aa})
\end{aligned}$$

0.1.5 Derive conditional probability

Let's set:

$$\begin{cases} x_{b.a} = x_b - \Sigma_{ba}\Sigma_{aa}^{-1}x_a \\ \mu_{b.a} = \mu_b - \Sigma_{b.a}\Sigma_{aa}^{-1}\mu_a \\ \Sigma_{bb.a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \end{cases}$$

$$\begin{aligned} x_{b.a} &= x_b - \Sigma_{ba}\Sigma_{bb}^{-1}x_a \\ &= \begin{pmatrix} -\Sigma_{ba}\Sigma_{bb}^{-1} & I \end{pmatrix} \begin{pmatrix} x_a \\ x_b \end{pmatrix} + 0 \end{aligned}$$

$$\begin{aligned} E[x_{b.a}] &= \begin{pmatrix} -\Sigma_{ba}\Sigma_{aa}^{-1} & I \end{pmatrix} \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \\ &= \mu_b - \Sigma_{ba}\Sigma_{aa}^{-1}\mu_a \\ &= \mu_{b.a} \end{aligned}$$

$$\begin{aligned} Var[x_{b.a}] &= \begin{pmatrix} -\Sigma_{ba}\Sigma_{aa}^{-1} & I \end{pmatrix} \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \begin{pmatrix} -\Sigma_{aa}^{-1}\Sigma_{ba}^T \\ I \end{pmatrix} \\ &= \begin{pmatrix} 0 & \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \end{pmatrix} \begin{pmatrix} -\Sigma_{aa}^{-1}\Sigma_{ba}^T \\ I \end{pmatrix} \\ &= \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \\ &= \Sigma_{bb.a} \end{aligned}$$

$$\therefore x_{b.a} \sim N(\mu_{b.a}, \Sigma_{bb.a})$$

$$\begin{cases} \mu_{b.a} = \mu_b - \Sigma_{ba}\Sigma_{aa}^{-1}\mu_a \\ \Sigma_{bb.a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \end{cases}$$

We slightly modify the formula:

$$\begin{aligned} x_{b.a} &= x_b - \Sigma_{ba}\Sigma_{aa}^{-1}x_a \\ x_b|x_a &= x_{b.a} + \Sigma_{ba}\Sigma_{aa}^{-1}x_a \\ &= Ix_{b.a} + C \end{aligned}$$

Here, all parts of the covariance matrix Σ_{ba} Σ_{aa} can be calculated, so it can be regarded as a constant.

And what we're asking for here is $x_b|x_a$, so x_a is also known, so the second term of the above formula can be regarded as a constant.

therefore:

$$E[x_b|x_a] = IE[x_{b.a}] + C = \mu_{b.a} + \Sigma_{ba}\Sigma_{aa}^{-1}x_a$$

$$Var[x_b|x_a] = IVar[x_{b.a}]I^T = \Sigma_{bb.a}$$

then we get the conditional probability of multivariate gaussian distribution:

$$x_b|x_a \sim N(\mu_{b.a} + \Sigma_{ba}\Sigma_{aa}^{-1}x_a, \Sigma_{bb.a})$$