EE 325 Probability and Random Processes 11-Nov-2019

Section 20: Conditional pdf of jointly Gaussian vectors

Animesh Kumar Email: `animesh@ee.iitb.ac.in`

As discussed, jointly Gaussian vectors are routinely used in communication theory, signal processing, control theory, and machine learning. Therefore, for analyzing electrical engineering systems, conditional distribution of jointly Gaussian random vectors should be understood.

1 The setup

Let $\vec{U} = \begin{bmatrix} \vec{X} \\ \vec{Y} \end{bmatrix}$ be an $(n+m)$-dimensional zero-mean jointly Gaussian vector with a covariance matrix $K_{\vec{U}}$. The conditional distribution properties of $\vec{X}|\vec{Y} = \vec{y}$ will be established. Similar procedure can be emulated to obtain the distribution for $\vec{Y}|\vec{X} = \vec{x}$. Finally, an linear prediction plus innovation interpretation will be provided.

2 Derivation of conditional distribution

From the definition of conditional distribution, the PDF of $\vec{X}|\vec{Y} = \vec{y}$ is given by

$$f_{\vec{X}|\vec{Y}}(\vec{x}|\vec{y}) = \frac{f_{\vec{X},\vec{Y}}(\vec{x},\vec{y})}{f_{\vec{Y}}(\vec{y})}. \tag{1}$$

The joint PDF of $\vec{X}, \vec{Y}$ is given by the distribution of $\vec{U}$ (broken into its constituent components). The distribution of $\vec{Y}$ is jointly Gaussian as well, with covariance matrix as a sub-part of $K_{\vec{U}}$. In detail, $K_{\vec{U}}$ can be written in the block matrix stucture as

$$K_{\vec{U}} = \begin{bmatrix} K_{\vec{X}} & K_{\vec{X}\vec{Y}} \\ K_{\vec{Y}\vec{X}} & K_{\vec{Y}} \end{bmatrix}, \tag{2}$$

where $K_{\vec{X}\vec{Y}} := \mathbb{E}[\vec{X}\vec{Y}^T]$ is the cross-covariance matrix for zero-mean vectors $\vec{X}$ and $\vec{Y}$. Observe that $K_{\vec{X}\vec{Y}} = K_{\vec{Y}\vec{X}}^T$. Due to the block structure of $K_{\vec{U}}$, a block structure is also obtained for $K_{\vec{U}}^{-1}$:

$$K_{\vec{U}}^{-1} = \begin{bmatrix} B & C \\ C^T & D \end{bmatrix}. \tag{3}$$

The next step involves algebraic computations to obtain the conditional PDF in (1). Note that

$$\begin{aligned}
f_{\vec{X}|\vec{Y}}(\vec{x}|\vec{y}) &= \frac{1}{(2\pi)^{(n+m)/2}} \exp\left(-\frac{\vec{u}^T K_U^{-1} \vec{u}}{2}\right) \Bigg/ \frac{1}{(2\pi)^{m/2}} \exp\left(-\frac{\vec{y}^T K_{\vec{Y}}^{-1} \vec{y}}{2}\right) && (4)\\
&= f_1(\vec{y}) \exp\left(-\frac{\vec{u}^T K_U^{-1} \vec{u}}{2}\right) && (5)\\
&= f_1(\vec{y}) \exp\left(-\frac{\vec{x}^T B\vec{x} + \vec{x}^T C\vec{y} + \vec{y}^T C^T\vec{x} + \vec{y}^T D\vec{y}}{2}\right) && (6)\\
&= f_2(\vec{y}) \exp\left(-\frac{\vec{x}^T B\vec{x} + \vec{x}^T C\vec{y} + \vec{y}^T C^T\vec{x}}{2}\right) && (7)\\
&= f_3(\vec{y}) \exp\left(-\frac{(\vec{x} + B^{-1}C\vec{y})^T B(\vec{x} + B^{-1}C\vec{y})^T}{2}\right) && (8)
\end{aligned}$$

where $f_1(\vec{y}), f_2(\vec{y}), f_3(\vec{y})$ are functions of $\vec{y}$ but do not depend on $\vec{x}$. The approach is to find the function form of the conditional distribution of $\vec{X}|\vec{Y} = \vec{y}$ first, and then find the constant by normalization of PDF.

From (8), and by comparisons with a Gaussian random vector's PDF, we get that $\vec{X}|\vec{Y} = \vec{y} \sim \mathcal{N}(-B^{-1}C\vec{y}, B^{-1})$. That automatically fixes $f_3(\vec{y})$ as $\frac{1}{(2\pi)^{n/2}\sqrt{\det(B^{-1})}}$. The only remaining part is to obtain the values of B and C in terms of $K_{\vec{U}}$. Since

$$\begin{bmatrix} K_{\vec{X}} & K_{\vec{X}\vec{Y}} \\ K_{\vec{Y}\vec{X}} & K_{\vec{Y}} \end{bmatrix} \begin{bmatrix} B & C \\ C^T & D \end{bmatrix} = I_{n+m}, \tag{9}$$

so $B^{-1} = K_{\vec{X}} - K_{\vec{X}\vec{Y}}K_{\vec{Y}}^{-1}K_{\vec{Y}\vec{X}}$ and $-B^{-1}C = K_{\vec{X}\vec{Y}}K_{\vec{Y}}^{-1}$. Finally,

$$\vec{X}|\vec{Y} = \vec{y} \sim \mathcal{N}(K_{\vec{X}\vec{Y}}K_{Y}^{-1}, K_{\vec{X}} - K_{\vec{X}\vec{Y}}K_{\vec{Y}}^{-1}K_{\vec{Y}\vec{X}}). \tag{10}$$

In the above expression, since the covariance of $\vec{X}|\vec{Y} = \vec{y}$ does not depend on the value of $\vec{y}$ (only the mean does!), it motivates an innovation formula as follows. For jointly Gaussian random vectors $\vec{X}$ and $\vec{Y}$,

$$\vec{X} = G\vec{Y} + \vec{V} \tag{11}$$

where $\vec{Y}$ and $\vec{V}$ are independent, $G = K_{\vec{X}\vec{Y}}K_{\vec{Y}}^{-1}$ and $\vec{V} \sim \mathcal{N}(\vec{0}, K_{\vec{X}} - K_{\vec{X}\vec{Y}}K_{\vec{Y}}^{-1}K_{\vec{Y}})$. Yes, you read it right: the vectors $\vec{V}$ and $\vec{Y}$ are independent. This expression explains the expression for $\mathbb{E}(\vec{X}|\vec{Y} = \vec{y})$ and its covariance matrix as well.

One method (and an intuitive one!) is to find G by the uncorrelatedness of $\vec{Y}$ and $\vec{V}$ (which follows from their independence). Since $\mathbb{E}(\vec{Y}\vec{V}^T) = 0$, so $\mathbb{E}[\vec{Y}(\vec{X} - G\vec{Y})^T] = 0$. This gives an expression for G. The covariance for $\vec{V}$ can then be found out using $K_{\vec{X}} = GK_{\vec{Y}}G^T + K_{\vec{V}}$.

An exercise will help to understand some of this. Let $Y_1 = X + W_1$ and $Y_2 = X + W_2$, where $X \sim \mathcal{N}(0, 1)$, $\vec{W} \sim \mathcal{N}(\vec{0}, \sigma^2 I_2)$. It is desired to find conditional mean of $\mathbb{E}(X|\vec{Y} = \vec{y})$ given the observation $\vec{Y}$, and the variance of $X|\vec{Y} = \vec{y}$. This is the starting point in a denoising problem. Solve it to practice this lecture. Then increase the size of $\vec{Y}$. You should observe that the variance of $X|\vec{Y} = \vec{y}$ decreases to zero as the size of $\vec{Y}$ increases.

Most of the material here and more can be found in the excellent textbook by Robert Gallager [1].

References

[1] R. G. Gallager: *Stochastic Processes*, Cambridge University Press.