EE 520: Random Processes **Fall 2021**

Lecture 12
Gaussian Random Vectors

Instructor Name: John Lipor

Recommended Reading: Pishro-Nik: 6.1.1, 6.1.5; Gubner: 9.1 - 9.5

Last week we organized finite collections of random variables into vectors, called *random vectors*. In this lecture, we focus on the specific case where the elements of the random vectors are Gaussian.

1 The Multivariate Normal Distribution

Recall the univariate Gaussian PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Further, for n independent Gaussian RVs, the joint PDF is

$$\begin{aligned} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right) \\ &= (2\pi)^{-n/2}\left(\sigma_1^2\sigma_2^2\ldots\sigma_n^2\right)^{-1/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right). \qquad (1) \end{aligned}$$

In the above, since the X_i's are independent, their covariance matrix is diagonal, i.e.,

$$C_X = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ldots & 0 \\ 0 & 0 & 0 & \ldots & \sigma_n^2 \end{bmatrix}$$

Define the vectors $X = [X_1 \; X_2 \; \ldots X_n]^T \in \mathbb{R}^n$ and $\mu = [\mu_1 \; \mu_2 \; \ldots \mu_n]^T \in \mathbb{R}^n$. Then we have

$$C_X^{-1}(x-\mu) = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & 0 & \ldots & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ldots & 0 \\ 0 & 0 & 0 & \ldots & \frac{1}{\sigma_n^2} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ \vdots \\ x_n-\mu_n \end{bmatrix} = \begin{bmatrix} \frac{x_1-\mu_1}{\sigma_1^2} \\ \frac{x_2-\mu_2}{\sigma_2^2} \\ \vdots \\ \frac{x_n-\mu_n}{\sigma_n^2} \end{bmatrix}.$$

With this in mind, we can begin writing the PDF (1) in matrix form. First note that

$$\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right) = \exp\left(-\frac{1}{2}(x-\mu)^T C_X^{-1}(x-\mu)\right).$$

To write the term $(2\pi)^{-n/2}\left(\sigma_1^2\sigma_2^2\ldots\sigma_n^2\right)^{-1/2}$ more compactly, we use the following two facts.

Fact 1. Let $A \in \mathbb{R}^{n\times n}$ with eigenvalues $\lambda_1,\ldots,\lambda_n$. Then $\det(A) = \prod_{i=1}^{n}\lambda_i$.

Fact 2. Let $A \in \mathbb{R}^{n\times n}$ be a diagonal matrix. Then the eigenvalues of A are the diagonal elements of A.

Putting these facts together, we see that

$$\prod_{i=1}^{n} \sigma_i^2 = \det(C_X).$$

We can therefore write the joint PDF of Gaussian RVs in matrix-vector notation as

$$f_X(x) = (2\pi)^{-n/2} \det(C_X)^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)^T C_X^{-1}(x-\mu)\right), \tag{2}$$

where we note that the variables X and x are vectors of length n. We have shown that the above provides a compact form of the PDF for independent Gaussian RVs. It turns out that the above is the correct PDF for dependent Gaussian RVs under a few assumptions. First, we require that C_X be invertible (since the PDF requires taking an inverse). Second, we require the RVec X to satisfy the following.

Definition 1. A random vector $X = [X_1\ X_2\ \dots\ X_n]^T$ is said to be **Gaussian** if

$$\sum_{i=1}^{n} c_i X_i, \quad c_i \in \mathbb{R}$$

is a scalar Gaussian RV. In other words, X is a Gaussian RVec if every linear combination of its elements is a Gaussian RV. If X has mean vector μ and covariance C, we write $X \sim \mathcal{N}(\mu, C)$.

Example 1. If $X_1, \dots, X_n$ are independent Gaussian RVs, then the RVec X is a Gaussian RVec.

Example 2. Every subvector of a Gaussian RVec is also a Gaussian RVec.

Summarizing the above, if $X = [X_1\ X_2\ \dots\ X_n]^T$ is Gaussian and its covariance matrix C_X is invertible, then X has PDF defined by (2).

1.1 Affine Transformations

Let $X \sim \mathcal{N}(\mu, C_X)$, $A \in \mathbb{R}^{r\times n}$, and $b \in \mathbb{R}^r$. What is the distribution of $Y = AX + b$? First consider $Z = AX$. For Z to be Gaussian, we need $c^T Z$ to be a scalar Gaussian RV for any $c \in \mathbb{R}^r$. Note that

$$c^T Z = c^T(AX) = (c^T A)X,$$

but $c^T A$ has size $1 \times n$, and hence $c^T Z$ is a linear combination of the elements of X. Therefore Z is Gaussian (since we defined X to be Gaussian). It is also easily checked that adding the constant $c^T b$ to Z gives another Gaussian, and hence $Y = AX + b$ is Gaussian. Next, we need to determine its mean and variance.

$$\mathbb{E}[Y] = \mathbb{E}[AX + b] = A\mathbb{E}[X] + b = A\mu + b.$$

$$\begin{aligned}
\mathbb{E}[YY^T] &= \mathbb{E}\left[(AX+b)(AX+b)^T\right] \\
&= \mathbb{E}\left[AXX^T A + 2AXb^T + bb^T\right] \\
&= AR_X A^T + 2A\mu b + bb^T.
\end{aligned}$$

To get C_Y, we subtract $\mathbb{E}[Y]\mathbb{E}[Y]^T$ from R_Y. Note that

$$\begin{aligned}
\mathbb{E}[Y]\mathbb{E}[Y]^T &= (A\mu + b)(A\mu + b)^T \\
&= A\mu\mu^T A^T + 2A\mu b^T + bb^T.
\end{aligned}$$

Therefore

$$\begin{aligned} C_Y &= AR_X A^T + 2A\mu b^T + bb^T - (A\mu\mu^T A^T + 2A\mu b^T + bb^T) \\ &= A\left(R_X - \mu\mu^T\right) A^T \\ &= AC_X A^T. \end{aligned}$$

Putting the above together, we see that

$$X \sim \mathcal{N}(\mu, C_X) \implies Y = AX + b \sim \mathcal{N}\left(A\mu + b, AC_X A^T\right).$$

1.2 Uncorrelated Implies Independent

Jointly Gaussian RVs have the useful property that if they are uncorrelated, they are independent. Note that in general (i.e., for other distributions) we only have the reverse implication. To show this, we use the MGF of a Gaussian RVec.

Proposition 1. Let $X \sim \mathcal{N}(\mu, C_X)$. Then the MGF of X is

$$M_X(s) = \mathbb{E}\left[e^{s^T X}\right] = \exp\left(s^T \mu + \frac{1}{2} s^T C_X s\right),$$

where the second equality follows by taking $Y = s^T X$ and applying the affine transformation formula above.

We now prove that uncorrelated Gaussian RVs are independent. Let $X \sim \mathcal{N}(\mu, C_X)$ have uncorrelated elements. Then

$$\begin{aligned} M_X(s) &= \exp\left(\sum_{i=1}^n s_i\mu_i + \frac{1}{2} s_i^2 \sigma_i^2\right) \\ &= \prod_{i=1}^n \exp\left(s_i\mu_i + \frac{1}{2} s_i^2 \sigma_i^2\right) \\ &= \prod_{i=1}^n M_{X_i}(s_i), \end{aligned}$$

which is the MGF of n independent Gaussian RVs. Since the MGF corresponds uniquely to the PDF, we conclude that the PDF of X is that of n independent Gaussian RVs.

1.3 Conditional Expectation and Probability

Recall our study of MMSE estimation, where we showed the *orthogonality principle*, i.e., that $\mathbb{E}[X|Y]$ satisfies

$$\mathbb{E}\left[h(Y)(X - E[X \mid Y])\right] = 0$$

for all functions $h(\cdot)$. For random vectors, the same property holds in the following form

$$\mathbb{E}\left[h(Y)^T (X - E[X \mid Y])\right] = 0,$$

where $X, Y \in \mathbb{R}^n$ are RVecs.

Proposition 2. Let X, Y be such that $[X\ Y]^T$ is a Gaussian RVec. Then

$$\mathbb{E}\left[X \mid Y = y\right] = A(Y - \mu_y) + \mu_x,$$

where A solves $AC_Y = C_{XY}$.

Proof. We show that the proposed solution satisfies the orthogonality principle. For simplicity, assume $\mu_X = 0$ and $\mu_Y = 0$. Now observe that the vector

$$\begin{bmatrix} X - AY \\ Y \end{bmatrix} = \begin{bmatrix} I & -A \\ 0 & I \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}$$

is Gaussian, since it is a linear transformation of the Gaussian vector $[X \ Y]^T$. Next, let's look at the correlation between the top and bottom entries.

$$\begin{aligned} \mathbb{E}\left[(X - AY)Y^T\right] &= \mathbb{E}\left[XY^T\right] - A\mathbb{E}\left[YY^T\right] \\ &= C_{XY} - AC_Y \\ &= C_{XY} - C_{XY} = 0, \end{aligned}$$

where the third line follows since A solves $AC_Y = C_{XY}$. Hence $(X - AY)$ and Y are uncorrelated and therefore independent. Therefore, for any function $h(\cdot)$

$$\begin{aligned} \mathbb{E}\left[h(Y)^T(X - AY)\right] &= \mathbb{E}\left[h(Y)\right]^T \mathbb{E}\left[X - AY\right] \\ &= \mathbb{E}\left[h(Y)\right]^T 0 = 0, \end{aligned}$$

which completes the proof. $\square$

In general, in the above case we have that

$$X \mid Y = y \sim \mathcal{N}\left(\mathbb{E}[X \mid Y = y], C_{X|Y}\right)$$

where

$$C_{X|Y} = C_X - AC_{YX}$$

and A solves $AC_Y = C_{XY}$. This fact is utilized when computing the posterior distribution on predictions when performing Gaussian process regression (a popular tool in machine learning).