CST 3526 Stochastic Process Lecture 14 - 12/18/2025

Lecture 14: Gaussian Processes I

Lecturer: Ziqiao Wang

In this lecture, we continue our study of Gaussian processes. We have already seen that any marginal of a jointly Gaussian vector is still Gaussian. We will further learn that (i) any linear transformation of a Gaussian random vector remains Gaussian, and (ii) the conditional distribution of a Gaussian is also Gaussian. Finally, we will present a necessary and sufficient condition under which a Gaussian random process is also a Markov process.

1 Properties of Multiple Gaussian Random Variables (Cont.)

Multivariate Gaussian distributions enjoy many nice properties that other distributions do not possess. Understanding these properties is very important for the study of Gaussian processes.

Linear Transformations

A random vector following a joint Gaussian distribution preserves its basic statistical structure under any linear transformation; only the parameters change accordingly. **This property is one of the most appealing aspects of multivariate Gaussian distributions**.

Theorem 1.1. *Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and let $\mathbf{C}$ be any $m \times n$ matrix. Define $\mathbf{Y} = \mathbf{C}\mathbf{X}$. Then $\mathbf{Y}$ is jointly Gaussian with mean $\mathbf{C}\boldsymbol{\mu}$ and covariance $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T$.*

Proof. Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)^T$. Then

$$\begin{aligned}
\Phi_{\mathbf{Y}}(\boldsymbol{\omega}) &= \mathbb{E}\left[e^{j\boldsymbol{\omega}^T\mathbf{Y}}\right] = \mathbb{E}\left[e^{j\boldsymbol{\omega}^T\mathbf{C}\mathbf{X}}\right] \\
&= \mathbb{E}\left[e^{j(\mathbf{C}^T\boldsymbol{\omega})^T\mathbf{X}}\right] \\
&= \exp\left(j(\mathbf{C}^T\boldsymbol{\omega})^T\boldsymbol{\mu} - \frac{1}{2}(\mathbf{C}^T\boldsymbol{\omega})^T\boldsymbol{\Sigma}(\mathbf{C}^T\boldsymbol{\omega})\right) \\
&= \exp\left(j\boldsymbol{\omega}^T(\mathbf{C}\boldsymbol{\mu}) - \frac{1}{2}\boldsymbol{\omega}^T(\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)\boldsymbol{\omega}\right).
\end{aligned}$$

Hence, $\mathbf{Y}$ is an m-dimensional jointly Gaussian random vector with mean $\mathbf{C}\boldsymbol{\mu}$ and covariance $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T$. □

This property is known as the invariance of the Gaussian distribution under linear transformations. It is also a characterization property of the multivariate Gaussian distribution: one can determine whether a random vector is jointly Gaussian by checking this invariance.

Theorem 1.2 (Characterization by Linear Projections)**.** *A random vector $\mathbf{X} = (X_1, \ldots, X_n)^T$ is jointly Gaussian if and only if for any $\mathbf{c} \in \mathbb{R}^n$, the scalar random variable*

$$Y = \mathbf{c}^T\mathbf{X}$$

is univariate Gaussian.

Proof. Let

$$Y = \mathbf{c}^T\mathbf{X}, \quad \mu_Y = \mathbf{c}^T\boldsymbol{\mu}, \quad \mathrm{Var}(Y) = \mathbf{c}^T\boldsymbol{\Sigma}\mathbf{c}.$$

Then

$$\Phi_Y(\omega) = \exp\left(j\omega \mathbf{c}^T \boldsymbol{\mu} - \frac{1}{2}\omega^2 \mathbf{c}^T \boldsymbol{\Sigma} \mathbf{c}\right). \tag{3-8}$$

Setting $\omega = 1$,

$$\mathbb{E}\left[e^{j\mathbf{c}^T \mathbf{X}}\right] = \exp\left(j\mathbf{c}^T \boldsymbol{\mu} - \frac{1}{2}\mathbf{c}^T \boldsymbol{\Sigma} \mathbf{c}\right). \tag{3-9}$$

Since $\mathbf{c}$ is arbitrary, $\mathbf{X}$ must be jointly Gaussian. □

The theorem suggests an alternative definition for jointly Gaussian random vectors:

Definition 1.1. A random vector $\mathbf{X}$ is said to be *jointly Gaussian* if and only if every linear combination

$$Z = \mathbf{a}^T \mathbf{X}$$

is a Gaussian random variable.

Example 1. Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$. The sample mean is $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and the sample variance $\bar{S} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Then, we have i) $\mathbb{E}\left[\bar{X}\right] = \mu$; ii) $\mathbb{E}\left[\bar{S}\right] = \sigma^2$; and iii) $\bar{X} \perp\!\!\!\perp \bar{S}$.

The first two results do not necessarily need the Gaussian assumption. The first property is clear, to show the second one, we can see that

$$\mathbb{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = \mathbb{E}\left[\sum_{i=1}^n X_i^2 - n(\bar{X})^2\right]. \tag{1}$$

Notice that $n\mathbb{E}\left[X_i^2\right] = n(\mu^2 + \sigma^2)$ and

$$\begin{aligned}\mathbb{E}\left[\bar{X}^2\right] = \mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n X_i X_j\right] &= \frac{1}{n^2}\left(\sum_{i=j} \mathbb{E}\left[X_i^2\right] + \sum_{i \neq j} \mathbb{E}\left[X_i X_j\right]\right) \\ &= \frac{1}{n}(\mu^2 + \sigma^2) + \frac{1}{n^2}(n^2 - n)\mu^2 = \frac{1}{n}\sigma^2 + \mu^2.\end{aligned}$$

Plugging into Eq. (1), we have $\mathbb{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = n(\mu^2 + \sigma^2) - (\sigma^2 + n\mu^2) = (n-1)\sigma^2$. Hence, $\mathbb{E}\left[\bar{S}\right] = \sigma^2$.

To verify the third result, we need the Gaussian assumption. Let $X = [X_1, X_2, \ldots, X_n]^T$, an n-dimension vector $c = [\frac{1}{n}, \ldots, \frac{1}{n}]$, and let an $n \times n$ matrix

$$A = \begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1 - \frac{1}{n} \end{pmatrix}.$$

Thus, $\bar{X} = c \cdot X$ and define $R = X - \bar{X}\mathbf{1} = [X_1 - \bar{X}, \ldots, X_n - \bar{X}]^T = AX$. Note that $(n-1)\bar{S} = R^T R$. Since $\mathbb{E}\left[\bar{X}R_i\right] = 0$ (c and A are orthogonal), we have $\bar{X} \perp\!\!\!\perp \bar{S}$, where we use the fact that uncorrelatedness implies independence for Gaussian.

Conditional Distributions

If $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ follows a jointly Gaussian distribution, then the conditional distributions $P_{X_1|X_2}$ and $P_{X_2|X_1}$ are also Gaussian. This result can be verified by direct calculation.

Assume the mean vector is $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, and the covariance matrix is $\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$.

The joint PDF is

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{(2\pi)^{n/2}(\det \Sigma_X)^{1/2}} \exp\left(-\frac{1}{2} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \Sigma_X^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right).$$

Using (LDV decomposition and) block matrix inversion,

$$\Sigma_X^{-1} = \begin{pmatrix} I & 0 \\ -\Sigma_{12}\Sigma_{22}^{-1} & I \end{pmatrix} \begin{pmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix}.$$

Define

$$\tilde{\mu}_1 = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \qquad \tilde{\Sigma}_{11} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Then

$$f_{X_1|X_2}(x_1 \mid x_2) = \frac{1}{(2\pi)^{n_1/2}(\det \tilde{\Sigma}_{11})^{1/2}} \exp\left(-\frac{1}{2}(x_1 - \tilde{\mu}_1)^T \tilde{\Sigma}_{11}^{-1}(x_1 - \tilde{\mu}_1) \right).$$

Hence,

$$\mathbb{E}\left[X_1 \mid X_2\right] = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \tag{2}$$

and

$$\Sigma_{X_1|X_2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \tag{3}$$

Remark 1.1. *To better understand the conditional mean and variance of a Gaussian distribution, let's focus on the one-dimensional case. The conditional mean is* $\mathbb{E}\left[X_1|X_2\right] = \mu_1 + \frac{\sigma_{21}}{\sigma_{22}}(X_2 - \mu)$. *The second term is precisely the linear projection of* X_1 *onto* X_2, *i.e., the best linear approximation of* X_1 *based on* X_2 *in the mean-squared error sense. Meanwhile, we already know (from our second lecture) that the conditional expectation* $\mathbb{E}[X_1 \mid X_2]$ *is the optimal estimator of* X_1 *given* X_2 *among all measurable functions of* X_2. *Therefore, in the Gaussian case, the optimal linear estimator and the optimal (unrestricted) estimator coincide.*

Moreover, the conditional variance is $\mathrm{Var}(X_1 \mid X_2) = \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21}$, *which is clearly nonnegative by the Cauchy–Schwarz inequality (equivalently, by the bound* $|\rho_{12}| \leq 1$ *on the correlation coefficient).*

Example 2 (Linear Gaussian System)**.** In practice, we often want to infer an intrinsic (latent) state or pattern X of a system—something that is not directly observable—from measurements or data Y. A common approach is the linear observation model $Y = AX + \epsilon$, where ϵ represents measurement noise. In the linear Gaussian setting, we assume

$$X \sim \mathcal{N}(\mu_X, \Sigma_X) \quad \text{(serving as a prior on } X\text{)}, \qquad \epsilon \sim \mathcal{N}(0, \Sigma_N),$$

and $X \perp\!\!\!\perp \epsilon$.

Note that X and Y are jointly Gaussian. Indeed,

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} I & 0 \\ A & I \end{pmatrix} \begin{pmatrix} X \\ \epsilon \end{pmatrix},$$

and (X, ϵ) is jointly Gaussian (due to the independence). Therefore, the conditional distribution $X \mid Y$ is also Gaussian, so it suffices to compute its conditional mean and covariance. Applying Eq.(2) and Eq.(3) obtains

$$\mathbb{E}[X|Y] = \mu_X + \Sigma_X A^T \left(A\Sigma_X A^T + \Sigma_N\right)^{-1} (Y - A\mu_X), \tag{4}$$

$$\Sigma_{X|Y} = \Sigma_X - \Sigma_X A^T \left(A\Sigma_X A^T + \Sigma_N\right)^{-1} A\Sigma_X. \tag{5}$$

Hence, the posterior distribution of X is $P_{X|Y} = \mathcal{N}\big(\mathbb{E}[X \mid Y], \Sigma_{X|Y}\big)$.

Example 3. Assume $X, Y \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Compute: (i) $\mathrm{E}[X+Y \mid X-Y]$; (ii) $\mathrm{E}[(X+Y)^2 \mid X-Y]$; (iii) $\mathrm{E}[\cos(X+Y) \mid X-Y]$.

Since $(X+Y, X-Y)$ is a linear transformation of (X, Y), it is jointly Gaussian. Indeed,

$$\begin{pmatrix} X+Y \\ X-Y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Let $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Because $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}(0, I)$, we have

$$\begin{pmatrix} X+Y \\ X-Y \end{pmatrix} \sim \mathcal{N}\big(0, AIA^\top\big) = \mathcal{N}(0, 2I).$$

In particular, $X+Y$ and $X-Y$ are independent (equivalently, they are uncorrelated: $\mathrm{E}[(X+Y)(X-Y)] = \mathrm{E}[X^2 - Y^2] = 0$). Therefore, conditioning on $X-Y$ does not change the distribution of $X+Y$, and

$$\mathrm{E}[X+Y \mid X-Y] = \mathrm{E}[X+Y] = 0, \qquad \mathrm{E}[(X+Y)^2 \mid X-Y] = \mathrm{E}[(X+Y)^2] = \mathrm{Var}(X+Y) = 2.$$

For (iii), independence again implies

$$\mathrm{E}[\cos(X+Y) \mid X-Y] = \mathrm{E}[\cos(X+Y)].$$

Using $\cos(\theta) = \frac{e^{j\theta} + e^{-j\theta}}{2}$ and the characteristic function of a Gaussian,

$$\mathbb{E}[\cos(X+Y)] = \mathbb{E}\left[\frac{e^{j(X+Y)} + e^{-j(X+Y)}}{2}\right] = \frac{\Phi_{X+Y}(1) + \Phi_{X+Y}(-1)}{2} = \frac{2e^{-1}}{2} = e^{-1}.$$

2 Gaussian Process

Definition 2.1. If every finite distribution of a random process is that of a collection of jointly Gaussian random variables, we say that the random process is *Gaussian*.

Gaussian random processes are important in applications since many "noise"-type processes are modeled by Gaussian processes. This is invariably justified on the basis of the Central Limit Theorem. Gaussian random processes have the important property that a Gaussian random process remains a Gaussian process under any linear transformation.

Theorem 2.1. *A real Gaussian process is uniquely specified by giving the mean of the process and its autocorrelation or autocovariance functions.*

Proof. To uniquely specify the process, we need only specify the finite distributions of the process. To specify the distribution of a finite collection of jointly Gaussian random variables, we only need to specify the mean vector and covariance matrix. Clearly the mean of the process gives us the mean vector of any finite collection of random variables of the process, while the mean and autocorrelation functions determine the autocovariance function (if it is not explicitly given) and hence the elements of the covariance matrix. □

Corollary 2.1. *A Gaussian random process is wide-sense stationary if and only if the process is (strictly) stationary.*

3 Gauss-Markov Property

If a random process is a real Gaussian process and simultaneously satisfies the Markov property, then the process is called a *Gauss–Markov process*. For this class of processes, the autocovariance function exhibits a specific structure. For simplicity, we assume throughout that the mean of the process is zero.

Theorem 3.1 (Gauss–Markov Characterization)**.** *A zero-mean Gaussian process $X(t)$ is a Markov process if and only if its autocovariance function $C_X(t,s)$ satisfies*

$$C_X(t_1,t_3) = \frac{C_X(t_1,t_2)\,C_X(t_2,t_3)}{C_X(t_2,t_2)}, \qquad \forall\, t_1 \le t_2 \le t_3. \tag{6}$$

Proof. We first prove the necessity.

Let
$$(X_1,X_2,X_3) = \big(X(t_1),X(t_2),X(t_3)\big).$$
By definition,
$$C_X(t_1,t_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_3 f_{X_1,X_3}(x_1,x_3)\,dx_1dx_3.$$
This can be rewritten as
$$C_X(t_1,t_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_3 f_{X_1,X_2,X_3}(x_1,x_2,x_3)\,dx_2dx_1dx_3.$$

By the Markov property,
$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = f_{X_3|X_2}(x_3|x_2)\,f_{X_2,X_1}(x_2,x_1),$$
hence
$$C_X(t_1,t_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_3 f_{X_3|X_2}(x_3|x_2)\,f_{X_2,X_1}(x_2,x_1)\,dx_3dx_2dx_1.$$

Using Eq. (2), we obtain
$$C_X(t_1,t_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{C_X(t_3,t_2)}{C_X(t_2,t_2)}\,x_2x_1 f_{X_2,X_1}(x_2,x_1)\,dx_1dx_2.$$

Therefore,
$$C_X(t_1,t_3) = \frac{C_X(t_1,t_2)\,C_X(t_2,t_3)}{C_X(t_2,t_2)}, \qquad \forall\, t_1 \le t_2 \le t_3.$$
Thus, necessity is proved.

We now prove sufficiency.

It suffices to show that for all n and for all $t_n,\dots,t_1$,
$$f_{X_n|X_{n-1},\dots,X_1}(x_n|x_{n-1},\dots,x_1) = f_{X_n|X_{n-1}}(x_n|x_{n-1}).$$

Define
$$\mathbf{Y}_n = (X_n,X_{n-1},\dots,X_1)^T.$$
Then
$$f_{X_n|X_{n-1},\dots,X_1}(x_n|x_{n-1},\dots,x_1) = \frac{f_{\mathbf{Y}_n}(\mathbf{y}_n)}{f_{\mathbf{Y}_{n-1}}(\mathbf{y}_{n-1})}.$$

The covariance matrix of $\mathbf{Y}_n$ is
$$\Sigma_{\mathbf{Y}_n} = \mathbb{E}\left[\mathbf{Y}_n\mathbf{Y}_n^T\right] = \begin{pmatrix} \mathbb{E}(X_n^2) & \mathbb{E}(X_nX_{n-1}) & \mathbb{E}(X_n\mathbf{Y}_{n-2}^T) \\ \mathbb{E}(X_nX_{n-1}) & \mathbb{E}(X_{n-1}^2) & \mathbb{E}(X_{n-1}\mathbf{Y}_{n-2}^T) \\ \mathbb{E}(\mathbf{Y}_{n-2}X_n) & \mathbb{E}(\mathbf{Y}_{n-2}X_{n-1}) & \mathbb{E}(\mathbf{Y}_{n-2}\mathbf{Y}_{n-2}^T) \end{pmatrix}.$$

From the condition in the theorem statement, we have

$$\mathbb{E}(X_n\mathbf{Y}_{n-2}) = \frac{\mathbb{E}(X_nX_{n-1})}{\mathbb{E}(X_{n-1}^2)}\mathbb{E}(X_{n-1}\mathbf{Y}_{n-2}).$$

Using the decorrelation method, we write

$$\Sigma_{\mathbf{Y}_n} = A\begin{pmatrix}\sigma^2 & 0 & 0\\ 0 & \mathbb{E}(X_{n-1}^2) & \mathbb{E}(X_{n-1}\mathbf{Y}_{n-2}^T)\\ 0 & \mathbb{E}(\mathbf{Y}_{n-2}X_{n-1}) & \mathbb{E}(\mathbf{Y}_{n-2}\mathbf{Y}_{n-2}^T)\end{pmatrix}A^T,$$

where

$$A = \begin{pmatrix}1 & -\dfrac{\mathbb{E}(X_nX_{n-1})}{\mathbb{E}(X_{n-1}^2)} & 0\\ 0 & 1 & 0\\ 0 & 0 & I\end{pmatrix}, \quad \sigma^2 = \mathbb{E}(X_n^2) - \frac{\mathbb{E}(X_nX_{n-1})^2}{\mathbb{E}(X_{n-1}^2)}.$$

Therefore,

$$\exp\big(-\tfrac{1}{2}\mathbf{Y}_n^T\Sigma_{\mathbf{Y}_n}^{-1}\mathbf{Y}_n\big) = \exp\bigg(-\frac{1}{2\sigma^2}\Big(X_n - \frac{\mathbb{E}(X_nX_{n-1})}{\mathbb{E}(X_{n-1}^2)}X_{n-1}\Big)^2\bigg)\exp\Big(-\tfrac{1}{2}\mathbf{Y}_{n-1}^T\Sigma_{\mathbf{Y}_{n-1}}^{-1}\mathbf{Y}_{n-1}\Big).$$

Hence,

$$f_{X_n|X_{n-1},\ldots,X_1}(x_n|x_{n-1},\ldots,x_1) = \frac{1}{\sqrt{2\pi}\sigma}\exp\bigg(-\frac{1}{2\sigma^2}\Big(x_n - \frac{\mathbb{E}(X_nX_{n-1})}{\mathbb{E}(X_{n-1}^2)}x_{n-1}\Big)^2\bigg) = f_{X_n|X_{n-1}}(x_n|x_{n-1}).$$

Thus, sufficiency is proved. $\square$