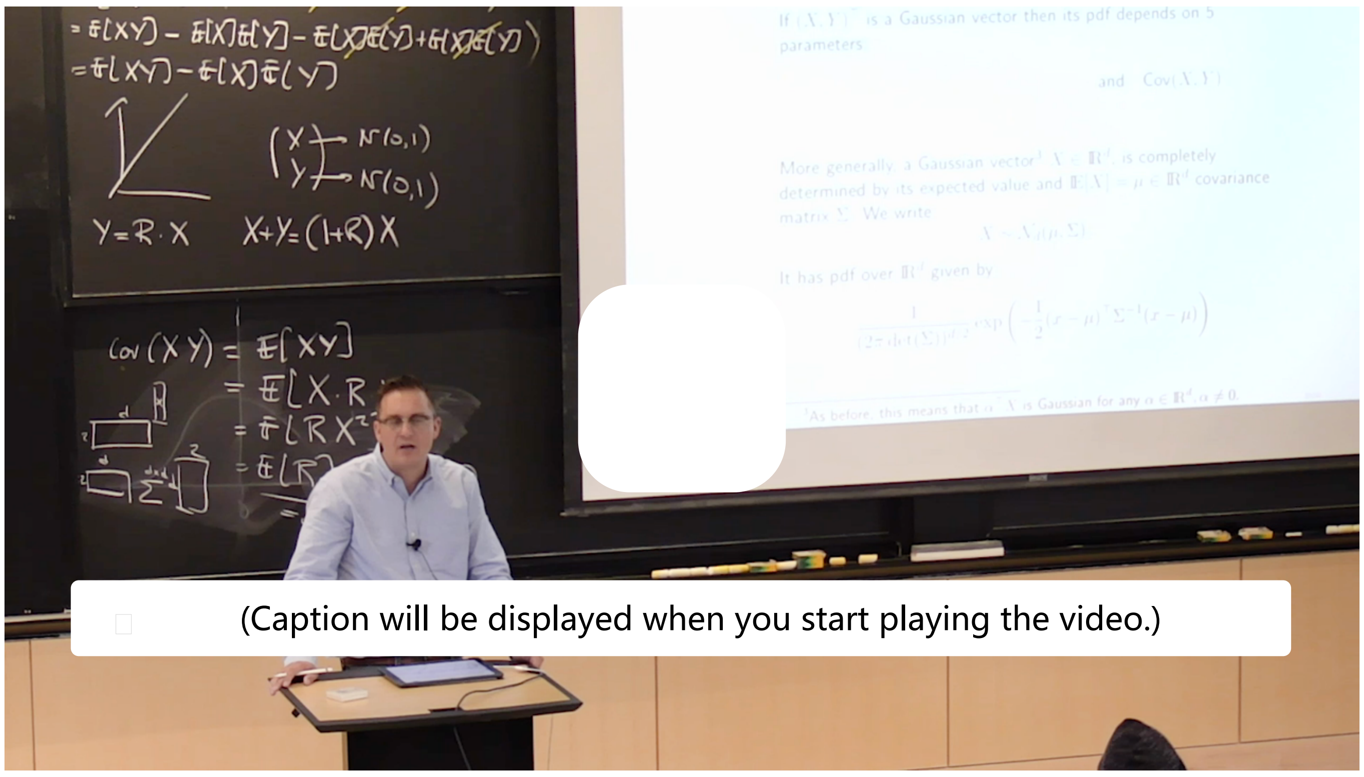
9. Multivariate Gaussian Distribution

Note: Now is a good time to review Gaussian random variables from Lecture 2.

Video Note: In the slide of the video below, there is a typo in the formula of the pdf of the multivariate Gaussian distribution: the exponent d in overall scaling factor should apply only to 2π, rather than $2\pi\det\Sigma$. The correct version is in the note below the video. (The unannotated slides in the resource section have also been corrected).

Multivariate Gaussian Distribution: Definition



Start of transcript. Skip to the end.

So now that I have a covariance matrix,

I can actually talk about a multivariate Gaussian

distribution, just like--

if I want to describe a Gaussian--

so what is nice about the Gaussian

is that it's described by only two parameters--

its mean and its variance.

视频
下载视频文件

字幕
下载 SubRip (.srt) file
下载 Text (.txt) file

Multivariate Gaussian Random Variable

A random vector $\mathbf{X} = (X^{(1)}, \ldots, X^{(d)})^T$ is a **Gaussian vector** , or **multivariate Gaussian or normal variable** , if any linear combination of its components is a (univariate) Gaussian variable or a constant (a “Gaussian" variable with zero variance), i.e., if $\alpha^T\mathbf{X}$ is (univariate) Gaussian or constant for any constant non-zero vector $\alpha \in \mathbb{R}^d$.

The distribution of $\mathbf{X}$, the d**-dimensional Gaussian or normal distribution** , is completely specified by the vector mean $\mu = \mathbb{E}[\mathbf{X}] = (\mathbb{E}[X^{(1)}], \ldots, \mathbb{E}[X^{(d)}])^T$ and the $d \times d$ covariance matrix Σ. If Σ is invertible, then the pdf of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)}, \quad \mathbf{x} \in \mathbb{R}^d$$

where $\det(\Sigma)$ is the determinant of the Σ, which is positive when Σ is invertible.

If $\mu = \mathbf{0}$ and Σ is the identity matrix, then $\mathbf{X}$ is called a **standard normal random vector** .

Note that when the covariant matrix Σ is diagonal, the pdf factors into pdfs of univariate Gaussians, and hence the components are independent.

Linear Transformation of a Multivariate Gaussian Random Vector

1/1 point (graded)

Consider the 2-dimensional Gaussian $\mathbf{X} = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$ with covariance matrix $\Sigma_X = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$ and mean $\mu_{\mathbf{X}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Consider the vector $\alpha = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, so that $Y = \alpha^T \mathbf{X}$ is a 1-dimensional Gaussian.

What is the variance $\textsf{Var}(Y)$ of Y?

$\textsf{Var}(Y) =$ 2

Answer: 2

Solution:

One way to answer this is to notice that $Y = X^{(1)} - X^{(2)}$, so

$$\textsf{Var}(Y) = \textsf{Cov}(Y, Y) = \textsf{Var}(X^{(1)}) + \textsf{Var}(X^{(2)}) - 2\textsf{Cov}(X^{(1)}, X^{(2)}) = 1 + 5 - 4 = 2.$$

Another way is to define the matrix $M \triangleq \alpha^T = \begin{pmatrix} 1 & -1 \end{pmatrix}$, and apply the formula $\Sigma_Y = M\Sigma_{\mathbf{X}}M^T = 2$.

提交 你已经尝试了1次（总共可以尝试3次）

Answers are displayed within the problem

Singular Covariance Matrices

1/1 point (graded)

Consider again a 2-dimensional Gaussian $\mathbf{X} = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$. But instead, Σ_X is $\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ and $\alpha = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$, what is the variance $\textsf{Var}(Y)$ of $Y = \alpha^T \mathbf{X}$?

$\textsf{Var}(Y) =$ 0

Answer: 0

This result tells us that the Gaussian $(X^{(1)}, X^{(2)})^T$ is actually a one-dimensional Gaussian, orthogonal to the direction of $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$.

Solution:

Define a matrix $M = \alpha^T$. We have $\Sigma_Y = M\Sigma_X M^T = 0$, since M^T is a column vector in the nullspace of Σ_X.

Such a Gaussian (with a singular covariance matrix) is sometimes referred to as a **degenerate** Gaussian.

提交 你已经尝试了2次（总共可以尝试3次）

Answers are displayed within the problem

(Optional) Diagonalization of the Covariance Matrix

Let Σ be a covariance matrix of size $d \times d$. Note that its entries are all real numbers with diagonal elements being non-negative. Σ has the following properties:

- Σ is symmetric. That is, $\Sigma = \Sigma^T$.
- Σ is diagonalizable to a diagonal matrix D via a transformation $D = U\Sigma U^T$, where U is an orthogonal matrix (recall that a square matrix A is orthogonal if $AA^T = A^T A = I$, where I is the identity matrix). This implies that $\Sigma = U^T D U$.

- Moreover, Σ is positive semidefinite. That is, the diagonal matrix D has diagonal entries that are all non-negative.

- Σ has a unique square root. That is, there exists a matrix $\Sigma^{\frac{1}{2}}$ that is unique such that $\Sigma^{\frac{1}{2}} \cdot \Sigma^{\frac{1}{2}} = \Sigma$.

- If Σ is of size $d \times d$, then it has d orthonormal eigenvectors (even if there are repeated eigenvalues). Furthermore, if U is a matrix with rows corresponding to the orthonormal eigenvectors, then the diagonal matrix $D = U\Sigma U^T$ contains the eigenvalues of Σ along its diagonal. Therefore, diagonalization of a symmetric matrix involves finding its eigenvalues and the orthonormal eigenvectors.

- If Σ is positive definite, i.e. the diagonal matrix $D = U\Sigma U^T$ has diagonal entries that are all strictly positive, then it is invertible and the inverse Σ^{-1} satisfies the following: $\Sigma^{-\frac{1}{2}} \cdot \Sigma^{-\frac{1}{2}} = \Sigma^{-1}$, where $\Sigma^{-\frac{1}{2}}$ is the inverse of the square root of Σ.

Hide

(Optional) Gaussian Random Vectors I

0 points possible (ungraded)

Recall from an earlier part of this lecture that the covariance between two random variables being 0 does not necessarily imply that the random variables are independent. However, this is true if the random variables are multivariate Gaussian.

Let $\mathbf{X}$ be a Gaussian random vector with mean μ and covariance Σ. Assume that Σ is positive definite. Determine if the following statement is true or false.

"There exists a vector B and a matrix A such that $A(\mathbf{X} + B)$ is a Gaussian random vector whose components are independent and each of mean $\mathbf{0}$".

- (•) True
- ( ) False

Hint: Refer to the note above on diagonalization of the covariance matrix.

Solution:

True. First, in order to remove the effect of μ we can set $B = -\mu$ to make the individual Gaussian random variables be of zero mean. Let $\widehat{\mathbf{X}} = \mathbf{X} - \mu$. From an earlier problem we know that the covariance matrix of $\widehat{\mathbf{X}}$ is the same as Σ.

From the above note on covariance matrices we can see that there exists an orthogonal matrix U such that $D = U\Sigma U^T$.

Consider the following transformation: $\mathbf{Y} = U\widehat{\mathbf{X}}$.

The covariance matrix of $\mathbf{Y}$ is (from an earlier problem)

$$U\Sigma U^T,$$

which is precisely equal to the diagonal matrix D. Therefore, $\mathbf{Y}$ has component Gaussian random variables that are uncorrelated and hence independent.

提交 你已经尝试了1次（总共可以尝试1次）

Answers are displayed within the problem

讨论

显示讨论

主题： Unit 3 Methods of Estimation:Lecture 10: Consistency of MLE, Covariance Matrices, and Multivariate Statistics / 9. Multivariate Gaussian Distribution