

Multivariate Normal Random Variables [Ross S7.8]

Definition of Multivariate Normal

Let Z_1, Z_2, \dots, Z_n be independent $\sim \mathcal{N}(0, 1)$.

Definition of Multivariate Normal

Let Z_1, Z_2, \dots, Z_n be independent $\sim \mathcal{N}(0, 1)$.

Then, define X_1, X_2, \dots, X_m by

$$X_1 = a_{11}Z_1 + \cdots + a_{1n}Z_n + \mu_1$$

$$X_2 = a_{21}Z_1 + \cdots + a_{2n}Z_n + \mu_2$$

$$\vdots \quad \quad \quad \vdots$$

$$X_m = a_{m1}Z_1 + \cdots + a_{mn}Z_n + \mu_m$$

Definition of Multivariate Normal

Let Z_1, Z_2, \dots, Z_n be independent $\sim \mathcal{N}(0, 1)$.

Then, define X_1, X_2, \dots, X_m by

$$\begin{aligned} X_1 &= a_{11}Z_1 + \cdots + a_{1n}Z_n + \mu_1 \\ X_2 &= a_{21}Z_1 + \cdots + a_{2n}Z_n + \mu_2 \\ &\vdots \quad \quad \quad \vdots \\ X_m &= a_{m1}Z_1 + \cdots + a_{mn}Z_n + \mu_m \end{aligned}$$

We say that X_1, \dots, X_m are **multivariate normal** (or **jointly Gaussian**).

We can write this in vector form as $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$:

We can write this in vector form as $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$:

$$\underbrace{\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix}}_{\mathbf{X}} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}}_{\mathbf{Z}} + \underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix}}_{\boldsymbol{\mu}}$$

We can write this in vector form as $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$:

$$\underbrace{\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix}}_{\mathbf{X}} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}}_{\mathbf{Z}} + \underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix}}_{\boldsymbol{\mu}}$$

Now, let \mathbf{B} be a $k \times m$ matrix, and $\boldsymbol{\nu}$ a column vector of length k . Then

$$\begin{aligned} \mathbf{Y} &= \mathbf{BX} + \boldsymbol{\nu} \\ &= (\mathbf{BA})\mathbf{Z} + (\mathbf{B}\boldsymbol{\mu} + \boldsymbol{\nu}) \end{aligned}$$

We can write this in vector form as $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$:

$$\underbrace{\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix}}_{\mathbf{X}} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}}_{\mathbf{Z}} + \underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix}}_{\boldsymbol{\mu}}$$

Now, let \mathbf{B} be a $k \times m$ matrix, and $\boldsymbol{\nu}$ a column vector of length k . Then

$$\begin{aligned} \mathbf{Y} &= \mathbf{BX} + \boldsymbol{\nu} \\ &= (\mathbf{BA})\mathbf{Z} + (\mathbf{B}\boldsymbol{\mu} + \boldsymbol{\nu}) \end{aligned}$$

So \mathbf{Y} is multivariate Gaussian too: an affine transformation of a multivariate Gaussian is again multivariate Gaussian!

Marginal Distribution of X_i

Since X_i is a sum of independent Gaussian random variables

→ X_i is Gaussian [Proposition ?? in Notes #26]

Marginal Distribution of X_i

Since X_i is a sum of independent Gaussian random variables

→ X_i is Gaussian [Proposition ?? in Notes #26]

Also:

$$\begin{aligned} E[X_i] &= E[a_{i1}Z_1 + \cdots + a_{in}Z_n + \mu_i] \\ &= a_{i1}E[Z_1] + \cdots + a_{in}E[Z_n] + \mu_i \\ &= \mu_i \end{aligned}$$

Marginal Distribution of X_i

Since X_i is a sum of independent Gaussian random variables

→ X_i is Gaussian [Proposition ?? in Notes #26]

Also:

$$\begin{aligned} E[X_i] &= E[a_{i1}Z_1 + \cdots + a_{in}Z_n + \mu_i] \\ &= a_{i1}E[Z_1] + \cdots + a_{in}E[Z_n] + \mu_i \\ &= \mu_i \end{aligned}$$

$$Var[X_i] = Var[a_{i1}Z_1 + \cdots + a_{in}Z_n + \mu_i]$$

Marginal Distribution of X_i

Since X_i is a sum of independent Gaussian random variables

→ X_i is Gaussian [Proposition ?? in Notes #26]

Also:

$$\begin{aligned} E[X_i] &= E[a_{i1}Z_1 + \cdots + a_{in}Z_n + \mu_i] \\ &= a_{i1}E[Z_1] + \cdots + a_{in}E[Z_n] + \mu_i \\ &= \mu_i \end{aligned}$$

$$\begin{aligned} Var[X_i] &= Var[a_{i1}Z_1 + \cdots + a_{in}Z_n + \mu_i] \\ &= Var[a_{i1}Z_1 + \cdots + a_{in}Z_n] \end{aligned}$$

Marginal Distribution of X_i

Since X_i is a sum of independent Gaussian random variables

→ X_i is Gaussian [Proposition ?? in Notes #26]

Also:

$$\begin{aligned} E[X_i] &= E[a_{i1}Z_1 + \cdots + a_{in}Z_n + \mu_i] \\ &= a_{i1}E[Z_1] + \cdots + a_{in}E[Z_n] + \mu_i \\ &= \mu_i \end{aligned}$$

$$\begin{aligned} Var[X_i] &= Var[a_{i1}Z_1 + \cdots + a_{in}Z_n + \mu_i] \\ &= Var[a_{i1}Z_1 + \cdots + a_{in}Z_n] \\ &= a_{i1}^2 Var[Z_1] + \cdots + a_{in}^2 Var[Z_n] \end{aligned}$$

Marginal Distribution of X_i

Since X_i is a sum of independent Gaussian random variables

→ X_i is Gaussian [Proposition ?? in Notes #26]

Also:

$$\begin{aligned} E[X_i] &= E[a_{i1}Z_1 + \cdots + a_{in}Z_n + \mu_i] \\ &= a_{i1}E[Z_1] + \cdots + a_{in}E[Z_n] + \mu_i \\ &= \mu_i \end{aligned}$$

$$\begin{aligned} Var[X_i] &= Var[a_{i1}Z_1 + \cdots + a_{in}Z_n + \mu_i] \\ &= Var[a_{i1}Z_1 + \cdots + a_{in}Z_n] \\ &= a_{i1}^2 Var[Z_1] + \cdots + a_{in}^2 Var[Z_n] \\ &= a_{i1}^2 + \cdots + a_{in}^2 \end{aligned}$$

A single Gaussian random variable U is uniquely specified by:

- its mean $E[U]$
- and its variance $Var[U]$.

Similarly:

The joint distribution of a multivariate Gaussian (normal) depends only on:

- the means $E[X_i]$ for $i = 1, \dots, m$
- and the co-variances $Cov[X_i, X_j]$ for $i = 1, \dots, m$ and $j = 1, \dots, m$

What happened to $Var[X_1]$, $Var[X_2]$, etc?

$Var[X_1] = Cov[X_1, X_1]$, so these are in the second bullet.

Common Notation

For random variables X_1, \dots, X_m , it is common to define:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} \quad \text{[random vector]}$$

$$\boldsymbol{\mu} = E[\mathbf{X}] = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_m] \end{pmatrix} \quad \text{[mean vector]}$$

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \quad \text{[covariance matrix]}$$

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \quad \text{[covariance matrix]}$$

$$= E\left[\begin{pmatrix} (X_1 - \mu_1)(X_1 - \mu_1) & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_m - \mu_m) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)(X_2 - \mu_2) & \cdots & (X_2 - \mu_2)(X_m - \mu_m) \\ \vdots & \vdots & \ddots & \vdots \\ (X_m - \mu_m)(X_1 - \mu_1) & (X_m - \mu_m)(X_2 - \mu_2) & \cdots & (X_m - \mu_m)(X_m - \mu_m) \end{pmatrix}\right]$$

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \quad \text{[covariance matrix]}$$

$$= E\left[\begin{pmatrix} (X_1 - \mu_1)(X_1 - \mu_1) & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_m - \mu_m) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)(X_2 - \mu_2) & \cdots & (X_2 - \mu_2)(X_m - \mu_m) \\ \vdots & \vdots & \ddots & \vdots \\ (X_m - \mu_m)(X_1 - \mu_1) & (X_m - \mu_m)(X_2 - \mu_2) & \cdots & (X_m - \mu_m)(X_m - \mu_m) \end{pmatrix}\right]$$

$$= \begin{pmatrix} Cov[X_1, X_1] & Cov[X_1, X_2] & \cdots & Cov[X_1, X_m] \\ Cov[X_2, X_1] & Cov[X_2, X_2] & \cdots & Cov[X_2, X_m] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_m, X_1] & Cov[X_m, X_2] & \cdots & Cov[X_m, X_m] \end{pmatrix}$$

Also, note that

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$$

Also, note that

$$\begin{aligned}\Sigma &= E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \\&= E[\mathbf{X}\mathbf{X}^T - \boldsymbol{\mu}\mathbf{X}^T - \mathbf{X}\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T] \\&= E[\mathbf{X}\mathbf{X}^T] - E[\boldsymbol{\mu}\mathbf{X}^T] - E[\mathbf{X}\boldsymbol{\mu}^T] + E[\boldsymbol{\mu}\boldsymbol{\mu}^T] \\&= E[\mathbf{X}\mathbf{X}^T] - \boldsymbol{\mu}E[\mathbf{X}^T] - E[\mathbf{X}]\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T \\&= E[\mathbf{X}\mathbf{X}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T \\&= E[\mathbf{X}\mathbf{X}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T\end{aligned}$$

If X_1, \dots, X_m are jointly Gaussian with μ and Σ , we write $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$.

If X_1, \dots, X_m are jointly Gaussian with μ and Σ , we write $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$.

It can be shown that if Σ is invertible, then

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu)}$$

Note: as expected, this depends only on μ and Σ .

Covariance Matrix

Say Z_1, \dots, Z_n are independent $\sim \mathcal{N}(0, 1)$. Then

$$\mu_Z = E[\mathbf{Z}] = \mathbf{0}$$

$$\begin{aligned} \Sigma_Z &= \begin{pmatrix} \text{Cov}[Z_1, Z_1] & \text{Cov}[Z_1, Z_2] & \cdots & \text{Cov}[Z_1, Z_n] \\ \text{Cov}[Z_2, Z_1] & \text{Cov}[Z_2, Z_2] & \cdots & \text{Cov}[Z_2, Z_n] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}[Z_n, Z_1] & \text{Cov}[Z_n, Z_2] & \cdots & \text{Cov}[Z_n, Z_n] \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = I \end{aligned}$$

Effect of Affine transformation on Covariance Matrix

Let X have mean μ_X and co-variance matrix Σ_X .

Let B be a matrix, and ν a column vector.

Let $Y = BX + \nu$. Then

Effect of Affine transformation on Covariance Matrix

Let \mathbf{X} have mean $\boldsymbol{\mu}_X$ and co-variance matrix Σ_X .

Let B be a matrix, and $\boldsymbol{\nu}$ a column vector.

Let $\mathbf{Y} = B\mathbf{X} + \boldsymbol{\nu}$. Then

$$\boldsymbol{\mu}_Y = E[\mathbf{Y}] = E[B\mathbf{X} + \boldsymbol{\nu}] = BE[\mathbf{X}] + \boldsymbol{\nu} = B\boldsymbol{\mu}_X + \boldsymbol{\nu}$$

$$\Sigma_Y = E[\mathbf{Y}\mathbf{Y}^T] - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^T$$

$$\begin{aligned}\Sigma_Y &= E[\mathbf{Y}\mathbf{Y}^T] - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^T \\&= E[(B\mathbf{X} + \boldsymbol{\nu})(B\mathbf{X} + \boldsymbol{\nu})^T] - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^T \\&= E[B\mathbf{X}\mathbf{X}^T B^T + B\mathbf{X}\boldsymbol{\nu}^T + \boldsymbol{\nu}\mathbf{X}^T B^T + \boldsymbol{\nu}\boldsymbol{\nu}^T] - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^T \\&= BE[\mathbf{X}\mathbf{X}^T]B^T + BE[\mathbf{X}]\boldsymbol{\nu}^T + \boldsymbol{\nu}E[\mathbf{X}^T]B^T + \boldsymbol{\nu}\boldsymbol{\nu}^T - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^T \\&= BE[\mathbf{X}\mathbf{X}^T]B^T + B\boldsymbol{\mu}_X\boldsymbol{\nu}^T + \boldsymbol{\nu}\boldsymbol{\mu}_X^T B^T + \boldsymbol{\nu}\boldsymbol{\nu}^T - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^T\end{aligned}$$

$$\begin{aligned}
\Sigma_Y &= E[\mathbf{Y}\mathbf{Y}^T] - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^T \\
&= E[(B\mathbf{X} + \boldsymbol{\nu})(B\mathbf{X} + \boldsymbol{\nu})^T] - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^T \\
&= E[B\mathbf{X}\mathbf{X}^T B^T + B\mathbf{X}\boldsymbol{\nu}^T + \boldsymbol{\nu}\mathbf{X}^T B^T + \boldsymbol{\nu}\boldsymbol{\nu}^T] - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^T \\
&= BE[\mathbf{X}\mathbf{X}^T]B^T + BE[\mathbf{X}]\boldsymbol{\nu}^T + \boldsymbol{\nu}E[\mathbf{X}^T]B^T + \boldsymbol{\nu}\boldsymbol{\nu}^T - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^T \\
&= BE[\mathbf{X}\mathbf{X}^T]B^T + B\boldsymbol{\mu}_X\boldsymbol{\nu}^T + \boldsymbol{\nu}\boldsymbol{\mu}_X^T B^T + \boldsymbol{\nu}\boldsymbol{\nu}^T - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y^T \\
&= BE[\mathbf{X}\mathbf{X}^T]B^T + B\boldsymbol{\mu}_X\boldsymbol{\nu}^T + \boldsymbol{\nu}\boldsymbol{\mu}_X^T B^T + \boldsymbol{\nu}\boldsymbol{\nu}^T - (B\boldsymbol{\mu}_X + \boldsymbol{\nu})(B\boldsymbol{\mu}_X + \boldsymbol{\nu})^T \\
&= BE[\mathbf{X}\mathbf{X}^T]B^T + B\boldsymbol{\mu}_X\boldsymbol{\nu}^T + \boldsymbol{\nu}\boldsymbol{\mu}_X^T B^T + \boldsymbol{\nu}\boldsymbol{\nu}^T \\
&\quad - (B\boldsymbol{\mu}_X\boldsymbol{\mu}_X^T B^T + B\boldsymbol{\mu}_X\boldsymbol{\nu}^T + \boldsymbol{\nu}\boldsymbol{\mu}_X^T B^T + \boldsymbol{\nu}\boldsymbol{\nu}^T) \\
&= BE[\mathbf{X}\mathbf{X}^T]B^T - B\boldsymbol{\mu}_X\boldsymbol{\mu}_X^T B^T \\
&= B(E[\mathbf{X}\mathbf{X}^T] - \boldsymbol{\mu}_X\boldsymbol{\mu}_X^T)B^T \\
&= B\Sigma_X B^T
\end{aligned}$$

Not all square matrices can be covariance matrices.

Below, is a general condition.

Not all square matrices can be covariance matrices.

Below, is a general condition.

Proposition

- a) A covariance matrix Σ is i) symmetric and ii) positive semi-definite.*
- b) Any matrix Σ that is symmetric and positive semi-definite is the covariance matrix of $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$ for some choice of matrix \mathbf{A} .*

Why?

$$\begin{aligned} i) \quad \Sigma^T &= E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]^T \\ &= E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]^T \\ &= E[(\mathbf{X} - \boldsymbol{\mu}^T)^T (\mathbf{X} - \boldsymbol{\mu})^T] \\ &= E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \\ &= \Sigma \end{aligned}$$

$$\begin{aligned} i) \quad \Sigma^T &= E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]^T \\ &= E[((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T)^T] \\ &= E[(\mathbf{X} - \boldsymbol{\mu}^T)^T (\mathbf{X} - \boldsymbol{\mu})^T] \\ &= E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \\ &= \Sigma \end{aligned}$$

$$\begin{aligned} ii) \quad \mathbf{v}^T \Sigma \mathbf{v} &= \mathbf{v}^T E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \mathbf{v} \\ &= E[\mathbf{v}^T (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T \mathbf{v}] \\ &= E[|(\mathbf{X} - \boldsymbol{\mu})^T \mathbf{v}|^2] \\ &\geq 0 \end{aligned}$$

b) Since Σ is symmetric, it can be diagonalized as $\Sigma = UDU^T$ where D is diagonal.

b) Since Σ is symmetric, it can be diagonalized as $\Sigma = UDU^T$ where D is diagonal.

The diagonal entries of D are ≥ 0 since Σ is positive semi-definite.

b) Since Σ is symmetric, it can be diagonalized as $\Sigma = UDU^T$ where D is diagonal.

The diagonal entries of D are ≥ 0 since Σ is positive semi-definite.

Then $\Sigma = UD^{1/2}D^{1/2}U^T$.

b) Since Σ is symmetric, it can be diagonalized as $\Sigma = UDU^T$ where D is diagonal.

The diagonal entries of D are ≥ 0 since Σ is positive semi-definite.

Then $\Sigma = UD^{1/2}D^{1/2}U^T$.

Let $A = UD^{1/2}$.

Then

$$\begin{aligned}\Sigma_X &= A\Sigma_ZA^T \\ &= AA^T \\ &= UD^{1/2}(UD^{1/2})^T \\ &= UD^{1/2}(D^{1/2})^TU^T \\ &= UD^{1/2}D^{1/2}U^T \\ &= \Sigma\end{aligned}$$