37. Properties of Expectations

Multivariate Normal Random Variables [Ross S7.8]

Definition of Multivariate Normal

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

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

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

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

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

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

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

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

So $\boldsymbol{Y}$ 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

$\rightarrow X_i$ is Gaussian [Proposition 26.1 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 $i = 1, \ldots, m$
- and the co-variances $Cov[X_i, X_j]$ for $i = 1, \ldots, m$ and $j = 1, \ldots, m$

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 $X_1, \ldots, X_m$, it is common to define:

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} \qquad \textbf{[random vector]}$$

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

$$\begin{aligned}
\Sigma &= E[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T] \qquad \textbf{[covariance matrix]} \\
&= E[\begin{pmatrix} (X_1-\mu_1)(X_1-\mu_1) & (X_1-\mu_1)(X_2-\mu_2) & \cdots & (X_1-\mu_1)(X_m-\mu_m) \\ (X_2-\mu_2)(X_1-\mu_1)] & (X_2-\mu_2)(X_2-\mu_2) & \cdots & (X_2-\mu_2)(X_m-\mu_m) \\ \vdots & \vdots & \ddots & \vdots \\ (X_m-\mu_m)(X_1-\mu_1) & (X_m-\mu_2)(X_m-\mu_2) & \cdots & (X_m-\mu_m)(X_m-\mu_m) \end{pmatrix}] \\
&= \begin{pmatrix} Cov[X_1, X_1] & Cov[X_1, X_2] & \cdots & Cov[X_1, X_m] \\ Cov[X_2, X_1] & Cov[X_2, X_2] & \cdots & Cov[X_2, X_m] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_m, X_1] & Cov[X_m, X_2] & \cdots & Cov[X_m, X_m] \end{pmatrix}
\end{aligned}$$

For special bivariate case of $\boldsymbol{X} = (X_1, X_2)^T$:

$$\begin{aligned}
\Sigma &= E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T] \\
&= E[\begin{pmatrix} (X_1 - \mu_1)(X_1 - \mu_1) & (X_1 - \mu_1)(X_2 - \mu_2) \\ (X_2 - \mu_2)(X_1 - \mu_1)] & (X_2 - \mu_2)(X_2 - \mu_2) \end{pmatrix}] \\
&= \begin{pmatrix} Cov[X_1, X_1] & Cov[X_1, X_2] \\ Cov[X_2, X_1] & Cov[X_2, X_2] \end{pmatrix} \\
&= \begin{pmatrix} Var[X_1] & Cov[X_1, X_2] \\ Cov[X_1, X_2] & Var[X_2] \end{pmatrix}
\end{aligned}$$

Back to general case. Also, note that

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

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

It can be shown that if Σ is invertible, then

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

Note: as expected, this depends only on $\boldsymbol{\mu}$ and Σ.

Covariance Matrix

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

$$\begin{aligned}
\boldsymbol{\mu}_Z &= E[\boldsymbol{Z}] = \mathbf{0} \\
\Sigma_Z &= \begin{pmatrix} Cov[Z_1, Z_1] & Cov[Z_1, Z_2] & \cdots & Cov[Z_1, Z_n] \\ Cov[Z_2, Z_1] & Cov[Z_2, Z_2] & \cdots & Cov[Z_2, Z_n] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[Z_n, Z_1] & Cov[Z_n, Z_2] & \cdots & Cov[Z_n, Z_n] \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = I
\end{aligned}$$

Effect of Affine transformation on Covariance Matrix

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

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

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

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

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

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

Not all square matrices can be covariance matrices.

Below, is a general condition.

Proposition 37.1 *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 $\boldsymbol{X} = A\boldsymbol{Z} + \boldsymbol{\mu}$ for some choice of matrix A.

Why?

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

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

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 it Σ is positive semi-definite.

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

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

Then

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