- Recall the mgf of $\mathbf{X} = (X_1, \ldots, X_k)'$,

$$M_{\mathbf{X}}(\boldsymbol{t}) = M_{\mathbf{X}}(t_1, \ldots, t_k) = \mathrm{E}e^{t_1X_1+\cdots+t_kX_k} = \mathrm{E}e^{\sum_{i=1}^k t_iX_i} = \mathrm{E}e^{\boldsymbol{t}'\mathbf{X}}$$

  for $\boldsymbol{t} = (t_1, \ldots, t_k)' \in \mathbb{R}^k$.

  (The mgf of $\boldsymbol{X}$ exists if the expected value exists for all $\boldsymbol{t}$ in some open neighborhood of $\mathbf{0} = (0, \ldots, 0)' \in \mathbb{R}^k$.)

- Recall also that if $M_{\mathbf{X}}(\boldsymbol{t}) = M_{\mathbf{Y}}(\boldsymbol{t})$ for all $\boldsymbol{t}$ in some open neighborhood of $\mathbf{0}$ then $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same distribution.

- *Result*: If $k \times 1$ random vector $\boldsymbol{X}$ has mgf $M_{\mathbf{X}}(\boldsymbol{t})$, then for a given $\ell \times k$ matrix $\boldsymbol{A}$ and given $\boldsymbol{b} \in \mathbb{R}^\ell$, the $\ell \times 1$ random vector $\mathbf{Y} = \boldsymbol{A}\mathbf{X} + \boldsymbol{b}$ has mgf

$$\boxed{M_{\mathbf{Y}}(\boldsymbol{s}) = e^{\boldsymbol{s}'\boldsymbol{b}}\mathrm{E}e^{\boldsymbol{A}'\boldsymbol{s}}} \quad \boldsymbol{s} = (s_1, \ldots, s_\ell)' \in \mathbb{R}^\ell$$

  (provided $M_{\mathbf{Y}}(\boldsymbol{s})$ exists in an open neighborhood of $\mathbf{0} \in \mathbb{R}^\ell$)

- *Recall*: If $X_1, \ldots, X_k$ are independent with mgfs $M_{X_i}(\cdot)$ then the $k \times 1$ random vector $\boldsymbol{X}$ has mgf

$$M_{\mathbf{X}}(\boldsymbol{t}) = M_{\mathbf{X}}(t_1, \ldots, t_k) = \mathrm{E}e^{\sum_{i=1}^k t_iX_i} = \prod_{i=1}^k M_{X_i}(t_i)$$

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix}_{k\times 1} \qquad (t)_{k\times 1}$$

Multivariate Normal Distribution

Introduction

Important Example: If $X_1, \ldots, X_k$ are independent with standard normal distributions, find the mgfs of

$M_{\mathbf{Y}}(s_1, s_2, \ldots, s_\ell)$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}, \qquad \mathbf{Y} = \underbrace{\begin{pmatrix} Y_1 \\ \vdots \\ Y_\ell \end{pmatrix}}_{\ell\times 1} = \underset{\ell\times k}{\boldsymbol{A}}\underset{k\times 1}{\mathbf{X}} + \underset{\ell\times 1}{\boldsymbol{b}}$$

for a given $\ell \times k$ matrix $\boldsymbol{A}$ and given $\boldsymbol{b} \in \mathbb{R}^\ell$.

$\mathbf{t} = (t_1, t_2, \ldots, t_k)'$, $M_{\mathbf{X}}(\mathbf{t}) \underset{\text{by the last result}}{=} \prod_{i=1}^{k} M_{X_i}(t_i) = \prod_{i=1}^{k} e^{t_i^2/2} = \underbrace{e^{\frac{\sum_{i=1}^{k} t_i^2}{2}}} = e^{\frac{\mathbf{t}'\mathbf{t}}{2}}$

$\mathbf{t}'\mathbf{t} = (t_1, t_2, \ldots, t_k)_{1\times k} \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_k \end{pmatrix}_{k\times 1} = \sum_{i=1}^{k} t_i^2$

$$\boxed{M_{\mathbf{X}}(\mathbf{t}) \overset{\circledast}{=} e^{\frac{\mathbf{t}'\mathbf{t}}{2}}}$$

$s = \begin{pmatrix} s_1 \\ \vdots \\ s_\ell \end{pmatrix}_{\ell\times 1}$

$\mathbf{s} = (s_1, s_2, \ldots, s_\ell)'$, $M_{\mathbf{Y}}(\mathbf{s}) \underset{\text{from the Past result}}{=} e^{s'b} \underbrace{M_{\mathbf{X}}(A's)} \overset{\circledast}{=} e^{s'b} e^{\frac{s'AA's}{2}}$

$= \underbrace{e^{s'b} e^{\frac{s'\mathrm{Var}(Y)s}{2}}}$

Note: $(AB)' = B'A'$

Notice two things from this example:

1. Because $\mathrm{E}\mathbf{X} = \mathbf{0} \in \mathbb{R}^k$ and $\mathrm{Var}(\mathbf{X}) = \boldsymbol{I}_k$ (the $k \times k$ identity matrix), then

   we had: $Y = AX + b$

   $E\mathbf{X} = \begin{pmatrix} EX_1 \\ \vdots \\ EX_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{k\times 1}$

   $$\mathrm{E}\mathbf{Y} = \underset{\ell\times k}{\boldsymbol{A}}\underset{k\times 1}{\mathrm{E}\mathbf{X}} + \boldsymbol{b} = \boldsymbol{b}, \qquad \mathrm{Var}(\mathbf{Y}) = \boldsymbol{A}\underset{k\times k}{\cancel{\mathrm{Var}(\mathbf{X})}}\boldsymbol{A}' = \underset{\ell\times\ell}{\boldsymbol{A}\boldsymbol{A}'}$$

   $E(Y) = b$

   $\mathrm{Var}(\mathbf{X}) = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}_{k\times k}$

2. Because distributions can be uniquely identified from mgfs, $\boldsymbol{Y}$ must have the same distribution as any

   $Y = AX + b$ and $W = DZ + b$ have the same dist

   $$\mathbf{W} = \begin{pmatrix} W_1 \\ \vdots \\ W_\ell \end{pmatrix} = \boldsymbol{D}\boldsymbol{Z} + \boldsymbol{b}$$

   $E(W) = D\,\cancel{E(Z)}^{\,0} + b = b$

   $\mathrm{Var}(W) = D\,\cancel{\mathrm{Var}(Z)}\,D' = DD'$

   where $\boldsymbol{D}$ is $\ell \times k$ with $\boldsymbol{D}\boldsymbol{D}' = \boldsymbol{A}\boldsymbol{A}'$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_k)'$ is a vector of k independent standard normal variables.

The last point above allows us to *define* **multivariate normal distributions**.

Multivariate Normal Distribution

The multivariate normal definition

Multivariate normal distribution is widely used in statistical theory and practice

Definition: A **multivariate normal distribution** $MVN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the distribution of a random vector $\mathbf{X} = (X_1, \ldots, X_k)'$ defined by

Recall: $Z \sim N(0,1) \Rightarrow \mu + \sigma Z \sim N(\mu, \sigma^2)$

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{P}'\boldsymbol{Z}$$

$k \times k$

where $\boldsymbol{P}$ is any $s \times k$ matrix ($s \leq k$) such that $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{\Sigma}$ and where $\boldsymbol{Z} = (Z_1, \ldots, Z_s)'$ denotes a vector with $Z_1, \ldots, Z_s$ as iid $N(0, 1)$ random variables.

$$E(X) = \mu + P'\underbrace{\cancel{E(Z)}}_{0} = \mu$$

$$Var(X) = Var(\mu + P'Z) = P'\underbrace{\cancel{Var(Z)}}_{I}(P')' = P'IP = P'P = \Sigma$$

Notes on the definition:

- $\boldsymbol{\mu}$ is a $\mathbb{R}^k$ vector denoting the mean of the distribution
- $\boldsymbol{\Sigma}$ is a symmetric, non-negative definite $k \times k$ matrix denoting the variance/covariance matrix of the distribution

  $Var(X) = P' Var(Z) P = P'P$
- For any non-negative-definite $\boldsymbol{\Sigma}$, a matrix $\boldsymbol{P}$ always exists where $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{\Sigma}$.
- For any positive-definite $k \times k$ matrix $\boldsymbol{\Sigma}$, the matrix $\boldsymbol{P}$ in the definition will necessarily be $k \times k$ and non-singular.

If Σ is non-negative + Σ^{-1} exists $\Rightarrow$ P' and P will be $k \times k$ matrices + P^{-1} exists

Multivariate Normal Distribution

Summary of multivariate normal properties to follow

1. Linear combinations of a MVN variable are again MVN

2. If any linear combination of $\boldsymbol{X}$ is always normal then $\boldsymbol{X}$ must be MVN

3. Subvectors of a MVN variable are again MVN

   Marginal distributions of a MVN are normal

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix}_{k\times 1} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \\ X_{p+1} \\ \vdots \\ X_k \end{pmatrix} = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \qquad X^{(1)} \to p\times 1, \quad X^{(2)} \to (k-p)\times 1$$

strong-Property

4. If $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{pmatrix}$ is MVN, then

   $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ are independent iff $\text{Cov}(\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}) = \boldsymbol{0}$

5. If $\boldsymbol{X}$ is MVN with non-singular $\text{Var}(\boldsymbol{X}) = \boldsymbol{\Sigma}$, then we can write out the joint pdf of $\boldsymbol{X}$

6. If $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{pmatrix}$ is MVN, the conditional distribution of $\boldsymbol{X}^{(1)} | \boldsymbol{X}^{(2)} = \boldsymbol{x}^{(2)}$ is MVN

MGF: If $\mathbf{X} = (X_1, \ldots, X_k)' \sim MVN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the moment generating function of $\boldsymbol{X}$ is given by

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \mathrm{E}e^{\boldsymbol{t}'\boldsymbol{X}} = \mathrm{E}e^{\sum_{i=1}^{k} t_i X_i} = e^{\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}}, \quad \text{for any } \boldsymbol{t} = (t_1, \ldots, t_k)' \in \mathbb{R}^k$$

Multivariate Normal Distribution

Transformation results

Result 1: Suppose $\underline{\boldsymbol{X} \sim MVN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})}$ and define $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{X}$ for a given $\boldsymbol{a} \in \mathbb{R}^m$ and $m \times k$ matrix $\boldsymbol{B}$. Then, $\boldsymbol{Y} \sim MVN_m(\boldsymbol{a} + \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}')$.

Proof: $X \sim MVN(\mu, \Sigma) \xrightarrow[\text{def}]{\text{by the}} \underset{\sim}{X} = \underbrace{\mu + P'Z}$ where $\Sigma = P'P$.

Now, $Y = a + B\underset{\sim}{X} = a + B(\mu + P'Z) = \underbrace{a + B\mu + BP'Z}_{Y \text{ must have the MVN}}$

$E(Y) = a + B\mu$

$Var(Y) = Var(BP'Z) = BP'\,Var(Z)(BP')' = BP'\underbrace{I P}_{\Sigma}B' = B\Sigma B'$

Result 2: Suppose $\boldsymbol{X} \sim MVN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and consider a partition of $\boldsymbol{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ as

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

where $\boldsymbol{X}^{(1)}$ is $p \times 1$ and $\boldsymbol{X}^{(2)}$ is $(k-p) \times 1$. Then,

$$\boldsymbol{X}^{(1)} \sim MVN_p(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11}) \quad \text{and} \quad \boldsymbol{X}^{(2)} \sim MVN_{k-p}(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22})$$

Result 3: $\boldsymbol{X}$ is MVN if and only if $\boldsymbol{a}'\boldsymbol{X} = \sum_{i=1}^{k} a_i X_i$ is normal for any $\boldsymbol{a} = (a_1, \ldots, a_k)' \in \mathbb{R}^k$