Back to Probability: Mean Vector and Covariance Matrix

For these topics (and also for bivariate normal distribution), we shall use column vectors (e.g., $\underset{\sim}{X} = (X_1, X_2, \ldots, X_m)^T$, $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_n)^T$) to denote random vectors.

Let $\underset{m\times 1}{\underset{\sim}{X}}$, $\underset{n\times 1}{\underset{\sim}{Y}}$ be two random (column) vectors defined on the same sample space Ω (i.e., each X_i and each Y_j is a map $\Omega \to \mathbb{R}$). If necessary, you may assume that $\begin{pmatrix} \underset{\sim}{X} \\ \underset{\sim}{Y} \end{pmatrix}$ is a $(m+n)$-dimensional ~~discrete~~ discrete | cont random vector.

Defn: The mean vector of $\underset{\sim}{X}$ is defined as

$$\mu_{\underset{\sim}{X}} = E(\underset{\sim}{X}) = (E(X_1), E(X_2), \ldots, E(X_m))^T \in \mathbb{R}^{m\times 1}$$

provided each X_i has finite mean.

Defn: Assume that each X_i and each Y_j have finite covariance. Then the covariance matrix between $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ is defined as $\underset{m\times n}{\mathrm{Cov}(\underset{\sim}{X}, \underset{\sim}{Y})} = \left(\left(\mathrm{Cov}(X_i, Y_j)\right)\right)_{\substack{1\le i\le m \\ 1\le j\le n}}$.

Clearly $Cov(\underset{\sim}{X}, \underset{\sim}{Y}) \in \mathbb{R}^{m \times n}$.

Defn: Assume that each X_i has finite 2^{nd} moment ($\Rightarrow$ for each pair (i,j) with $1 \leq i \leq j \leq m$, the r.v.s X_i and X_j have finite covariance).

Then the variance-covariance matrix or dispersion matrix of $\underset{\sim}{X}$ is defined as

$$Var(\underset{\sim}{X}) = Disp(\underset{\sim}{X}) := \Big(\Big(Cov(X_i, X_j) \Big)\Big)_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$$

$$\in \mathbb{R}^{m \times m}.$$

Remark: Clearly, the i^{th} diagonal element of $Var(\underset{\sim}{X})$ is simply $Var(X_i)$ (for each i) and the off-diagonal elements are cross covariances.

Exc: For each of the following, assume RHS exists. Show that LHS also exists and equals RHS.

1) $E(A\underset{\sim}{X} + \underset{\sim}{a}) = A E(\underset{\sim}{X}) + \underset{\sim}{a}$ $\quad \forall A \in \mathbb{R}^{p \times m}$ and $\forall \underset{\sim}{a} \in \mathbb{R}^{p \times 1}$.
$(p \in \mathbb{N})$

~~Clearly~~ $\cancel{\mathrm{Cov}(\underset{\sim}{X}, \underset{\sim}{Y}) \in \mathbb{R}^{m \times n}}$

- (Just a fancy restatement of linearity of expectation.)

2) $\mathrm{Cov}(A\underset{\sim}{X} + \underset{\sim}{a}, \; B\underset{\sim}{Y} + \underset{\sim}{b}) = A \, \mathrm{Cov}(\underset{\sim}{X}, \underset{\sim}{Y}) B^T$

$\forall \; A \in \mathbb{R}^{p \times m}, \; \forall \; \underset{\sim}{a} \in \mathbb{R}^{p \times 1}, \; \forall \; B \in \mathbb{R}^{q \times n},$

and $\forall \; \underset{\sim}{b} \in \mathbb{R}^{q \times 1} \; (p, q \in \mathbb{N})$.

(Just a fany linear algebraic way of writing the bilineary of covariance.)

3) $\mathrm{Var}(A\underset{\sim}{X} + \underset{\sim}{a}) = A \, \mathrm{Var}(\underset{\sim}{X}) \, A^T$

$\forall \; A \in \mathbb{R}^{p \times m}$ and $\forall \; \underset{\sim}{a} \in \mathbb{R}^{p \times 1} \; (p \in \mathbb{N})$.

4) If $m = n$, then

$$\mathrm{Var}(\underset{\sim}{X} + \underset{\sim}{Y}) = \mathrm{Var}(\underset{\sim}{X}) + \mathrm{Var}(\underset{\sim}{Y}) + \mathrm{Cov}(\underset{\sim}{X}, \underset{\sim}{Y}) + \mathrm{Cov}(\underset{\sim}{Y}, \underset{\sim}{X}).$$

Bivariate Normal Distribution

Example: Suppose $X_1, X_2 \overset{iid}{\sim} N(0,1)$, and

$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathbb{R}^{2\times 2}$ is nonsingular.

Define $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \underset{\sim}{Y} := A\underset{\sim}{X}$, where $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$.

In other words,

$$Y_1 := a_{11} X_1 + a_{12} X_2 \text{, and}$$

$$Y_2 := a_{21} X_1 + a_{22} X_2 .$$

Find the joint distn of Y_1 and Y_2.

Solution: A joint pdf of $\underset{\sim}{X}$ is

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2}, \quad x_1, x_2 \in \mathbb{R}$$

$$= \frac{1}{2\pi} e^{-\frac{1}{2}\underset{\sim}{x}^T\underset{\sim}{x}}, \quad \underset{\sim}{x} \in \mathbb{R}^{2\times 1}$$

(by ② of Pg 248) $\equiv \mathbb{R}^2$.

We shall use the bivariate change of joint density formula.

In this case,

$$I = \mathbb{R}^2 \quad (\text{open} + \text{path-connected}),$$

and $g : I \to \mathbb{R}^2$ is defined by

$$g(\underset{\sim}{x}) = A\underset{\sim}{x}, \quad \underset{\sim}{x} \in I.$$

Exc: Using the nonsingularity of A, show that g is one-to-one and $g(I) = \mathbb{R}^2 =: J$.
Also show that $g^{-1} : \underset{\mathbb{R}^2}{\overset{\|}{J}} \to \underset{\mathbb{R}^2}{\overset{\|}{I}}$ is given by

$$g^{-1}(\underset{\sim}{y}) = A^{-1}\underset{\sim}{y}, \quad \underset{\sim}{y} \in J = \mathbb{R}^2.$$

Exc: Show that $\dfrac{dg^{-1}(\underset{\sim}{y})}{d\underset{\sim}{y}} = \det\left(J_{g^{-1}}(\underset{\sim}{y})\right)$

$$= \det(A^{-1}) \quad \forall\, \underset{\sim}{y} \in \underset{\mathbb{R}^2.}{\overset{\|}{J}}$$

Therefore, $\underset{\sim}{Y} = g(\underset{\sim}{X}) = A\underset{\sim}{X}$ is a cont random vector with a joint pdf

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = f_{\underset{\sim}{X}}(g^{-1}(\underset{\sim}{y})) \left|\frac{dg^{-1}(\underset{\sim}{y})}{d\underset{\sim}{y}}\right|, \quad \underset{\sim}{y} \in \mathbb{R}^2.$$

$$= f_{\underset{\sim}{X}}(A^{-1}\underset{\sim}{y})\,|\det(A^{-1})|, \quad \underset{\sim}{y} \in \mathbb{R}^2$$

$$= \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(A^{-1}\underset{\sim}{y})^T(A^{-1}\underset{\sim}{y})\right\} |\det(A^{-1})|, \quad \underset{\sim}{y} \in \mathbb{R}^2$$

$$= \frac{1}{2\pi|\det(A)|} \exp\left\{-\frac{1}{2}\underset{\sim}{y}^T(A^T)^{-1}A^{-1}\underset{\sim}{y}\right\}, \quad \underset{\sim}{y} \in \mathbb{R}^2$$

$\left[\because \det(A^{-1}) = \frac{1}{\det(A)}\right.$ and $\left.(A^{-1})^T = (A^T)^{-1}\right]$

$$= \frac{1}{2\pi|\det(A)|} \exp\left\{-\frac{1}{2}\underset{\sim}{y}^T(AA^T)^{-1}\underset{\sim}{y}\right\}, \quad \underset{\sim}{y} \in \mathbb{R}^2$$

$[\because (AB)^{-1} = B^{-1}A^{-1}$ if B is also nonsing$]$

$$\Rightarrow f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}\underset{\sim}{y}^T\Sigma^{-1}\underset{\sim}{y}\right\}, \quad \underset{\sim}{y} \in \mathbb{R}^2,$$

where $\Sigma = AA^T \Rightarrow \begin{cases} \Sigma \text{ is pd (since } A \text{ is nonsing)} \\ |\det(A)| = +\sqrt{\det(\Sigma)} \end{cases}$

(by Remark 2 of Pg 251)

__Continuation:__ Fix a vector $\underset{\sim}{\mu} = (\mu_1, \mu_2)^T \in \mathbb{R}^{2\times 1}$.

Let $\underset{\sim}{Z} = \underset{\sim}{Y} + \underset{\sim}{\mu} = A\underset{\sim}{X} + \underset{\sim}{\mu}$, where $X_1, X_2 \overset{iid}{\sim} N(0,1)$.

__Exc:__ Show that $\underset{\sim}{Z}$ is a cont random vector with a joint pdf $f_{\underset{\sim}{Z}}(\underset{\sim}{z}) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}(\underset{\sim}{z}-\underset{\sim}{\mu})\Sigma^{-1}(\underset{\sim}{z}-\underset{\sim}{\mu})\right\}$, $\underset{\sim}{z} \in \mathbb{R}^2$.

Remark: In light of the Thm in Pg 251, it ~~foll~~ follows from Pg 255 - 257 that $\forall \underset{\sim}{\mu} \in \mathbb{R}^{2\times 1}$ and $\forall$ pd matrix $\Sigma \in \mathbb{R}^{2\times 2}$,

$$h(\underset{\sim}{z}) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}(\underset{\sim}{z}-\underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{z}-\underset{\sim}{\mu})\right\},$$

$$\underset{\sim}{z} \in \mathbb{R}^2$$

is a valid joint pdf on $\mathbb{R}^2$.

Def^n: Let $\underset{\sim}{\mu} = (\mu_1, \mu_2)^T \in \mathbb{R}^{2\times 1} \cong \mathbb{R}^2$ be a vector and $\Sigma \in \mathbb{R}^{2\times 2}$ be a pd matrix. Then a cont random vector $\underset{\sim}{X} = (X_1, X_2)^T$ is said to follow a bivariate normal distribution with parameters $\underset{\sim}{\mu}$ and Σ if $\underset{\sim}{X}$ has a joint pdf

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x}-\underset{\sim}{\mu})\right\}, \quad \underset{\sim}{x} \in \mathbb{R}^2.$$

Remark: We shall figure out soon that $E(\underset{\sim}{X}) = \underset{\sim}{\mu}$ and $Var(\underset{\sim}{X}) = \Sigma$. (See Cor 2 of Pg 259)

Notation: $\underset{\sim}{X} \sim N_2(\underset{\sim}{\mu}, \Sigma)$.

Facts on Bivariate Normal Distribution

① $\underset{\sim}{X} = (X_1, X_2)^T \sim N_2(\underset{\sim}{0}, I_2)$ iff (standard bivariate normal distribution)
$X_1, X_2 \overset{iid}{\sim} N(0,1)$.

Exc: Verify ①.

② If $\underset{\sim}{X} \sim N_2(\underset{\sim}{0}, I_2)$, then for all nonsingular $A \in \mathbb{R}^{2\times 2}$ and for all $\underset{\sim}{\mu} \in \mathbb{R}^{2\times 1}$, we have $A\underset{\sim}{X} + \underset{\sim}{\mu} \sim N_2(\underset{\sim}{\mu}, AA^T)$.

Proof: Use the computations in Pg 255 - 257 + Exc in Pg 258.

Cor 1: $\underset{\sim}{X} \sim N_2(\underset{\sim}{0}, I_2) \Rightarrow A\underset{\sim}{X} \sim N_2(\underset{\sim}{0}, I_2)$ $\forall$ $\underbrace{\text{orthogonal matrix}}_{\text{(i.e. } AA^T = A^TA = I)}$ $A \in \mathbb{R}^{2\times 2}$.
(Rotational invariance of std bivariate normal distn)

Cor 2: $\underset{\sim}{Z} \sim N_2(\underset{\sim}{\mu}, \Sigma) \Rightarrow E(\underset{\sim}{Z}) = \underset{\sim}{\mu}$ and $Var(\underset{\sim}{Z}) = \Sigma$.

Proof: Exc (Hint: Use ② + Exc in Pg 253 - 254.)

3) If $\underset{\sim}{Z} \sim N_2(\underset{\sim}{\mu}, \Sigma)$, then $\forall$ nonsingular $B \in \mathbb{R}^{2\times 2}$, we have

$$B\underset{\sim}{Z} \sim N_2(B\underset{\sim}{\mu}, B\Sigma B^T)$$

Proof: Shall use "equality of distⁿ trick".

$\Sigma \in \mathbb{R}^{2\times 2}$ is pd $\Rightarrow \exists$ nonsing $A \in \mathbb{R}^{2\times 2}$ such that $\Sigma = AA^T$.

This means that $\underset{\sim}{Z} \sim N_2(\underset{\sim}{\mu}, AA^T)$.

$$\xrightarrow[\text{Pg } 259]{\text{Fact } 2} \quad \underset{\sim}{Z} \overset{d}{=} A\underset{\sim}{X} + \underset{\sim}{\mu}, \text{ where } \underset{\sim}{X} \sim N_2(\underset{\sim}{0}, I_2).$$

$$\xrightarrow[\text{Pg } 244]{\text{Thm } 3} \quad B\underset{\sim}{Z} \overset{d}{=} B(A\underset{\sim}{X} + \underset{\sim}{\mu}) = BA\underset{\sim}{X} + B\underset{\sim}{\mu}.$$

Now $\underset{2\times 2}{A}, \underset{2\times 2}{B}$ are nonsingular

$\Rightarrow$ BA is also nonsingular

Hence by Fact 2 of Pg 259,

$$BA\underset{\sim}{X} + B\underset{\sim}{\mu} \sim N_2(B\underset{\sim}{\mu},\ BA(BA)^T).$$

Note that $BA(BA)^T = B\,AA^TB = B\Sigma B.$

Therefore $BA\underset{\sim}{X} + B\underset{\sim}{\mu} \sim N_2(B\underset{\sim}{\mu},\ B\Sigma B^T).$

On the other hand, $B\underset{\sim}{Z} \overset{d}{=} BA\underset{\sim}{X} + B\underset{\sim}{\mu}$

$$\underset{\text{Pg } 243}{\overset{\text{Thm 2}}{\Longrightarrow}} \quad B\underset{\sim}{Z} \sim N_2(B\underset{\sim}{\mu},\ B\Sigma B^T).$$

4 If $\underset{\sim}{Z} \sim N_2(\underset{\sim}{\mu}, \Sigma)$, then $\forall$ nonsingular $B \in \mathbb{R}^{2\times 2}$ and $\forall$ $\underset{\sim}{v} \in \mathbb{R}^{2\times 1}$, we have

$$B\underset{\sim}{Z} + \underset{\sim}{v} \sim N_2(B\underset{\sim}{\mu} + \underset{\sim}{v},\ B\Sigma B^T).$$

Proof: Exc.

5 Suppose $\underset{\sim}{X} = (X_1, X_2)^T \sim N_2(\underset{\sim}{\mu}, \Sigma)$. ~~Show~~ Then ~~that~~ $X_1 \perp\!\!\!\perp X_2$ if and only if $\text{Cov}(X_1, X_2) = 0$

Proof: Exc. [**Hint**: Use a joint pdf of X_1 and X_2.]

Remark: See Remark 3 of Pg 185.