$$\left.\begin{aligned} X &\sim N(\mu_x, \sigma_x^2) \\ Y &= \alpha_0 + \beta_0 X + \varepsilon \end{aligned}\right\}$$

ε is independent of X.

$\varepsilon \sim N[0, \sigma_\varepsilon^2]$

Derived the joint distribution of (X, Y) in terms of $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)^{1/2}} \exp\left[-\frac{1}{2(1-\rho^2)} E\right]$$

$$\boxed{E = \left(\frac{X-\mu_X}{\sigma_X}\right)^2 + \left(\frac{Y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right)}$$

normalized forms of (X, Y)

$BVN[\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho]$

<u>Exercise</u>: $\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} a_1 X + b_1 Y \\ a_2 X + b_2 Y \end{pmatrix}$

$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}$ is non-singular matrix,

then $\begin{pmatrix} W_1 \\ W_2 \end{pmatrix}$ has a BVN distribution.

How? Use COVT.

Parameters: $(\mu_{W_1}, \mu_{W_2}, \sigma^2_{W_1}, \sigma^2_{W_2}, \rho_{W_1 W_2})$

these can be worked out completely in terms of the parameters we start with.

Specifically, if $W_1 = \dfrac{X - \mu_X}{\sigma_X}$

$W_2 = \dfrac{Y - \mu_Y}{\sigma_Y}$

Then: $\boxed{W_1 \sim N(0,1),\ W_2 \sim N(0,1)}$

$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \sim BVN\left[(0, 0, 1, 1, \rho)\right]$

[check]

$$f_{W_1,W_2}(w_1,w_2) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left[-\frac{1}{2(1-\rho^2)}(w_1^2+w_2^2-2\rho w_1 w_2)\right]$$

We'll show that X can also be expressed as a linear model in Y.

To that end, first show that W_1 is expressible as a linear model in W_2.

$$f_{W_1,W_2}(w_1,w_2)$$

$$= \underbrace{\left[\frac{1}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2}w_2^2\right]\right]}_{\text{marginal of } W_2}$$

$$\times \underbrace{\frac{\exp\left[-\frac{1}{2(1-\rho^2)}\left(w_1^2+w_2^2-2\rho w_1 w_2 - w_2^2(1-\rho^2)\right)\right]}{(2\pi)^{1/2}(1-\rho^2)^{1/2}}}_{\text{conditional of } W_1 \mid W_2 = w_2}$$

$$f_{W_1|W_2=w_2}(w_1)$$

$$= \frac{1}{(2\pi)^{1/2}(1-\rho^2)^{1/2}} \exp\left[-\frac{1}{2}(w_1 - \rho w_2)^2\right]$$

(look at $w_1^2 + w_2^2 - 2\rho w_1 w_2$

$- w_2^2(1-\rho^2)$

$= w_1^2 + w_2^2\rho^2 - 2\rho w_1 w_2$

$= (w_1 - \rho w_2)^2$)

$W_1 | W_2 = w_2 \sim N[\rho w_2, 1-\rho^2]$

To get at LM representation,

define: $\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} W_2 \\ W_1 - \rho W_2 \end{pmatrix}$

$W_1 - \rho W_2 = \tilde{\varepsilon}$.

again use COVT to show that

W_2 and $\tilde{\varepsilon}$ are independent and so:

$w_1 = \rho w_2 + \tilde{\varepsilon}$ with $\tilde{\varepsilon}$ ind. of w_2, and $\tilde{\varepsilon} \sim N[0, 1-\rho^2]$.

This gives an LM representation.

so: $$\frac{X - \mu_X}{\sigma_X} = \rho \cdot \frac{Y - \mu_Y}{\sigma_Y} + \tilde{\varepsilon}$$

simplify to see that:

$$X = \tilde{\alpha}_0 + \tilde{\beta}_0 Y + \tilde{\tilde{\varepsilon}}$$

where $Y \sim N[\mu_Y, \sigma_Y^2]$ and

$\tilde{\tilde{\varepsilon}} \perp\!\!\!\perp Y$.

Next, we talk about the general multivariate normal distribution in p dimensions for some $p \geq 2$.

We'd like to define the notion of a random variable $X \sim N_p(\underline{\mu}, \underline{\underline{\Sigma}})$

$\mu = \mathbb{E}X$, $\underline{\underline{\Sigma}} = \text{Disp}(X)$

$$= \mathbb{E}\left[(X-\mu)(X-\mu)^T\right]$$

Σ is a positive-definite matrix.

Given any Σ p.d. there exists a $p \times p$ matrix B of full rank such ~~that~~ $\Sigma = BB^T$.

Since Σ is positive definite, we can express $\Sigma = U \Lambda U^T$

where $U_{p \times p}$ is an orthonormal matrix (i.e $UU^T = U^T U = I$) and Λ is a diagonal matrix: $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$ where $\lambda_i > 0 \; \forall i$. $\left(\underline{\Sigma U = \Lambda U} \right)$

eigen-vectors eigen-values

$$\Sigma = \underline{U \Lambda U^T} \to B.$$

$$= (U \Lambda^{1/2} U^T) U \Lambda^{1/2} U^T$$

$$\Lambda^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_p^{1/2}).$$

Then $\Sigma = B^2 \quad (B^T = B)$ B is called the symmetric squared root of Σ.

Let $Z \sim N_p(0, I_p)$: by this 9 mean that $Z = \begin{pmatrix} z_1 \\ \vdots \\ z_p \end{pmatrix}$ where

$z_1, z_2, \ldots, z_p$ are iid $N(0,1)$.

$$f_{\underset{\sim}{Z}}(z_1, z_2, \ldots, z_p) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z_i^2}$$

$$\to = \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2} \|\underset{\sim}{z}\|^2\right)$$

$$\|\underset{\sim}{z}\| = \sqrt{\sum z_i^2}$$

$$= \frac{1}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2} \underset{\sim}{z}^T \underset{\sim}{z}\right]$$

Have defined the canonical multivariate normal random vector.

Now define the class of non-singular p-variate normal distribution as the class of distributions of all non-singular affine transformations of Z, i.e

consider $X = \underset{\sim}{\mu} + BZ$ where

$\mu \in \mathbb{R}^p$ and $B_{p \times p}$ is a non-singular matrix with BB^T being denoted Σ.

Use the COVT to show:

$\mathbb{E}X = \mu$

$Disp(X) = BB^T = \Sigma$

$$f_{\underset{\sim}{X}}(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}}$$

$$\times \exp\left[-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right]$$

$x \in \mathbb{R}^p$.

density function

$$\text{Disp}(X) = E\left[(X-\mu)(X-\mu)^T\right]$$
$$= \mathbb{E}\left[BZ(BZ)^T\right]$$
$$= \mathbb{E}\left[BZZ^TB^T\right]$$
$$= B\,\underline{E\left[ZZ^T\right]}\,B^T$$
$$= BIB^T = \Sigma.$$

<u>Easy check</u>: any affine transformation of X is also MVN (multivariate normal).

$$Y = \eta + CX$$ where C is non-singular.

Then,

$$Y \sim \text{MVN}\left[\eta + C\mu,\ C\Sigma C^T\right]$$

$\boxed{\text{Prove using CDVT.}}$

Yet another application of COVT:

Suppose $X \sim N(\mu, \Sigma)$.

and $\Sigma = B^2$ (as indicated before)

and define $Y = B^{-1}(X - \mu)$

$= \Sigma^{-1/2}(X - \mu)$

(notation: $B \equiv \Sigma^{1/2}$).

$\longleftrightarrow$

show ~~that~~ $Y \sim N[0, I_p]$ i.e

$Y_1, Y_2, \ldots, Y_p$ are i.i.d $N[0,1]$.

Moment generating functions:

$\longleftrightarrow$

Suppose $Z \sim N[0, I_p]$.

so: $f_Z(z) = \dfrac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2} z^T z\right)$

Calculate: $M_Z(t) = \mathbb{E}\left[e^{t^T Z}\right] \leftarrow$

Show this is $\exp\left(\|t\|^2/2\right)$

$= \exp\left(\frac{1}{2} t^T t\right)$.

if $X \sim N_p(\mu, \Sigma)$,
use the fact that $X = \mu + BZ$
for B, the symmetric square root of Σ,
and $Z \sim N_p(0, I_p)$ to conclude

that:

$$M_X(t) = \mathbb{E}\left[e^{t^T X}\right] = e^{\mu t + \frac{1}{2} t^T \Sigma t}.$$

Critical property of Normal Distributions:

For normal random vectors (X_1, X_2)
(i.e (X_1, X_2) have a joint normal
distribution) independence is
equivalent to $Cov(X_1, X_2) = 0$.

$$\Sigma_{12} = \underline{Cov(X_1, X_2)} = \left[Cov(X_{1i}, X_{2j})\right]_{\substack{1 \le i \le p_1 \\ 1 \le j \le p_2}}$$

where $X_1 = (X_{11}, \ldots, X_{1p_1})^T$
$X_2 = (X_{21}, \ldots, X_{2p_2})^T$.

$$X_{p\times 1} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \underline{p = p_1 + p_2}.$$

$$\text{Let } x_p \sim N[\mu, \Sigma].$$

$$= N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right]$$

$$\Sigma_{11},\ p_1\times p_1 = \text{Disp}(X_1)$$

$$= \mathbb{E}\left[(X_1-\mu_1)(X_1-\mu_1)^T\right]$$

$$\Sigma_{22},\ p_2\times p_2 = \mathbb{E}\left[(X_2-\mu_2)(X_2-\mu_2)^T\right]$$

$$\Sigma_{12},\ p_1\times p_2 = \mathbb{E}\left[(X_1-\mu_1)(X_2-\mu_2)^T\right]$$

$$\Sigma_{21},\ p_2\times p_1 = \Sigma_{12}^T.$$

$$\mu_1 = \mathbb{E}X_1,\quad \mu_2 = \mathbb{E}X_2$$

(I) Suppose X_1 and X_2 are independent

Then trivially, $\Sigma_{12} = 0$.

$$\Sigma_{12} = \mathbb{E}\left[\begin{pmatrix} x_{11} - \mu_{11} \\ \vdots \\ x_{1p_1} - \mu_{1p_1} \end{pmatrix} \begin{pmatrix} x_{21} - \mu_{21}, \dots, x_{2p_2} - \mu_{2p_2} \end{pmatrix} \right]$$

column row

$$(\Sigma_{12})_{i,j}$$

$$= \mathbb{E}\left[(x_{1i} - \mu_{1i})(x_{2j} - \mu_{2j}) \right]$$

$$= 0$$

because X_{1i} is independent of X_{2j}.

Next suppose ~~that~~ $\Sigma_{12} = 0$.

$$f_{X_1, X_2}(x_1, x_2)$$

$$= \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} (\underline{x} - \underline{\mu})^T \Sigma^{-1} (x - \mu) \right]$$

$$\boxed{\underline{X} = (X_1, X_2)}$$

look at:

$$(x-\mu)^T \Sigma^{-1} (x-\mu)$$

$$|\Sigma| = |\Sigma_{11}|\,|\Sigma_{22}|$$

$$= \left[(x_1-\mu_1)^T, (x_2-\mu_2)^T\right]$$

$$\times \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}^{-1} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

$$= \left[(x_1-\mu_1)^T, (x_2-\mu_2)^T\right] \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

$$\left.\begin{aligned} = & \, (\underline{x_1}-\mu_1)^T \Sigma_{11}^{-1} (x_1-\mu_1) \\ & + (x_2-\mu_2)^T \underline{\Sigma_{22}}^{-1} (x_2-\mu_2) \end{aligned}\right\}$$

Simplify the joint density:

$$f_{X_1, X_2}(x_1, x_2)$$

$$= \frac{1}{(2\pi)^{p/2} |\Sigma_{11}|^{1/2} |\Sigma_{22}|^{1/2}}$$

$$\times \exp\left[-\frac{1}{2}(x_1 - \mu_1)^T \Sigma_{11}^{-1} (x_1 - \mu_1)\right]$$

$$\times \exp\left[-\frac{1}{2}(x_2 - \mu_2)^T \Sigma_{22}^{-1} (x_2 - \mu_2)\right]$$

Joint density splits as product of functions involving X_1 and X_2 respectively

Easily deduced that:

$$f_{x_1}(x_1) = \frac{1}{(2\pi)^{p_1/2} |\Sigma_{11}|^{1/2}} \exp(\star)$$

$$\star = -\frac{1}{2}(x_1 - \mu_1)^T \Sigma_{11}^{-1} (x_1 - \mu_1)$$

X_1, X_2 indep.
$X_1 \sim N(\mu_1, \Sigma_{11})$
$X_2 \sim N(\mu_2, \Sigma_{22})$

$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$

When $\Sigma_{12} \neq 0$, what can we say about conditional distributions?

How do we know in general ~~that~~ $X_1 \sim N(\mu_1, \Sigma_{11})$?

I could integrate out w.r.t. X_2!

or: $\mathbb{E}[\exp(t_1^T X_1)]$ $\quad t_1 : p_1 \times 1$

$$= \mathbb{E}\left[\exp\left[(t_1^T, \underset{=}{0}^T)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\right]\right]$$

where $\underset{\sim}{0}$ is a $p_2 \times 1$ vector of 0's.

$$= \mathbb{E}\left[(t_1^T, 0^T)\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \frac{1}{2}(t_1^T, 0^T)\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} t_1 \\ 0 \end{pmatrix}\right]$$

$$= \mathbb{E}\left[\exp\left(t_1^T \mu_1 + \frac{1}{2} t_1^T \Sigma_{11} t_1\right)\right]$$

precisely mgf of $N(\mu_1, \Sigma_{11})$!