Lecture Notes for Theory of Linear Models

Ch 3 and 4 in Rencher and Schaalje's book

- Random vector and matrix
- Multivariate Normal (MVN) Distribution

Longhai Li

Department of Mathematics and Statistics
University of Saskatchewan

Motivation:

$$y = X\beta + \varepsilon \quad , \quad \varepsilon_i \sim N(0, \sigma^2)$$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \sim N(X\beta, \sigma^2 I_n)$$

$$\hat{y} = Py \quad , \quad e = y - \hat{y} = (I_n - P)y$$

Random Vector: A vector whose elements are random variables. E.g.,

$$\mathbf{x}_{k\times 1} = \begin{pmatrix} x_1 & x_2 & \cdots & x_k \end{pmatrix}^T,$$

where $x_1, \ldots, x_k$ are each random variables.

Random Matrix: A matrix whose elements are random variables. E.g., $\mathbf{X}_{n\times k} = (x_{ij})$, where $x_{11}, x_{12}, \ldots, x_{nk}$ are each random variables.

Expected Value: The expected value (population mean) of a random matrix (vector) is the matrix (vector) of expected values. For $\mathbf{X}_{n\times k}$,

$$\mathrm{E}(\mathbf{X}) = \begin{pmatrix} \mathrm{E}(x_{11}) & \mathrm{E}(x_{12}) & \cdots & \mathrm{E}(x_{1k}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}(x_{n1}) & \mathrm{E}(x_{n2}) & \cdots & \mathrm{E}(x_{nk}) \end{pmatrix}.$$

$$E\left(\begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}\right) = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_k) \end{bmatrix}$$

(Population) Variance-Covariance Matrix: For a random vector $\mathbf{x}_{k\times 1} = (x_1, x_2, \ldots, x_k)^T$, the matrix

$$\sigma_{ij} = \mathrm{cov}(x_i, x_j)$$

$$\mathrm{Var}(x) = \begin{pmatrix} \mathrm{var}(x_1) & \mathrm{cov}(x_1,x_2) & \cdots & \mathrm{cov}(x_1,x_k) \\ \mathrm{cov}(x_2,x_1) & \mathrm{var}(x_2) & \cdots & \mathrm{cov}(x_2,x_k) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(x_k,x_1) & \mathrm{cov}(x_k,x_2) & \cdots & \mathrm{var}(x_k) \end{pmatrix} = \Sigma_x$$

$$\equiv \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{pmatrix}$$

is called the variance-covariance matrix of $\mathbf{x}$ and is denoted $\mathrm{var}(\mathbf{x})$ or $\Sigma_{\mathbf{x}}$ or sometimes Σ when it is clear which random vector is being referred to.

$$\sigma_{ij} = \mathrm{cov}(x_i, x_j) = \mathrm{E}[(x_i - \mu_i)(x_j - \mu_j)]$$

$$\sigma_{ii} = \mathrm{var}(x_i) = \mathrm{E}[(x_i - \mu_i)^2]$$

$$\mathrm{var}(\mathbf{x}) = \mathrm{E}[(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T].$$

Var(x) is symmetric

(Population) Correlation Matrix: For a random variable $\mathbf{x}_{k\times 1}$, the population correlation matrix is the matrix of correlations among the elements of $\mathbf{x}$:

$$\mathrm{corr}(\mathbf{x}) = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{21} & 1 & \cdots & \rho_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & 1 \end{pmatrix},$$

where $\rho_{ij} = \mathrm{corr}(x_i, x_j)$.

$$= \frac{\mathrm{cov}(x_i, x_j)}{\sqrt{V(x_i)}\sqrt{V(x_j)}}$$

$$(X-u_X)\cdot(X-u_X)' \quad , \quad u_X=(u_1,\cdots,u_n)' \quad u_i=E(x_i)$$

$$= \begin{bmatrix} x_1-u_1 \\ \vdots \\ x_n-u_n \end{bmatrix} (x_1-u_1, \cdots, x_n-u)$$

$$= \begin{bmatrix} (x_1-u_1)^2 & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \ddots & & & \vdots & \\ \vdots & & \ddots & \cdots & a_{ij} & \cdots \\ & & & & \vdots & \\ a_{n1} & \cdots & & & & (x_n-u_n)^2 \end{bmatrix} \quad a_{ij} = (x_i-u_i)(x_j-u_j)$$

(Population) Covariance Matrix: For random vectors $\mathbf{x}_{k\times 1} = (x_1, \ldots, x_k)^T$, and $\mathbf{y}_{n\times 1} = (y_1, \ldots, y_n)^T$ let $\sigma_{ij} = \text{cov}(x_i, y_j)$, $i = 1, \ldots, k$, $j = 1, \ldots, n$. The matrix

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kn} \end{pmatrix} = \begin{pmatrix} \text{cov}(x_1, y_1) & \cdots & \text{cov}(x_1, y_n) \\ \vdots & \ddots & \vdots \\ \text{cov}(x_k, y_1) & \cdots & \text{cov}(x_k, y_n) \end{pmatrix} = \Sigma_{x,y}$$

is the **covariance matrix** of $\mathbf{x}$ and $\mathbf{y}$ and is denoted $\text{cov}(\mathbf{x}, \mathbf{y})$, or sometimes $\Sigma_{\mathbf{x},\mathbf{y}}$.

- Notice that the $\text{cov}(\cdot, \cdot)$ function takes two arguments, each of which can be a scalar or a vector.
- In terms of vector/matrix algebra, $\text{cov}(\mathbf{x}, \mathbf{y})$ has formula

$$\text{cov}(\mathbf{x}, \mathbf{y}) = \text{E}[(\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})(\mathbf{y} - \boldsymbol{\mu}_\mathbf{y})^T].$$

- Note that $\text{var}(\mathbf{x}) = \text{cov}(\mathbf{x}, \mathbf{x})$.

$(x - u_x)\cdot(y - u_y) = (a_{ij})_{k\times n}$, where,

$a_{ij} = (x_i - u_i^x)\cdot(y_j - u_j^y)$

$$\text{corr}(\mathbf{x},\mathbf{y}) = \begin{pmatrix} \text{corr}(x_1,y_1) & \cdots & \text{corr}(x_1,y_n) \\ \vdots & \ddots & \vdots \\ \text{corr}(x_k,y_1) & \cdots & \text{corr}(x_k,y_n) \end{pmatrix}.$$

- Notice that $\text{corr}(\mathbf{x}) = \text{corr}(\mathbf{x},\mathbf{x})$.
- For random vectors $\mathbf{x}_{k\times 1}$ and $\mathbf{y}_{n\times 1}$, let

  $\boldsymbol{\rho}_{\mathbf{x}} = \text{corr}(\mathbf{x})$, $\Sigma_{\mathbf{x}} = \text{var}(\mathbf{x})$, $\boldsymbol{\rho}_{\mathbf{x},\mathbf{y}} = \text{corr}(\mathbf{x},\mathbf{y})$, $\Sigma_{\mathbf{x},\mathbf{y}} = \text{cov}(\mathbf{x},\mathbf{y})$,
  $\mathbf{V}_{\mathbf{x}} = \text{diag}(\text{var}(x_1),\dots,\text{var}(x_k))$, and $\mathbf{V}_{\mathbf{y}} = \text{diag}(\text{var}(y_1),\dots,\text{var}(y_n))$

  The relationship between $\boldsymbol{\rho}_{\mathbf{x}}$ and $\Sigma_{\mathbf{x}}$ is

  $$\Sigma_{\mathbf{x}} = \mathbf{V}_{\mathbf{x}}^{1/2}\boldsymbol{\rho}_{\mathbf{x}}\mathbf{V}_{\mathbf{x}}^{1/2}$$
  $$\boldsymbol{\rho}_{\mathbf{x}} = (\mathbf{V}_{\mathbf{x}}^{1/2})^{-1}\Sigma_{\mathbf{x}}(\mathbf{V}_{\mathbf{x}}^{1/2})^{-1}$$

  and the relationship between the covariance and correlation matrices of $\mathbf{x}$ and $\mathbf{y}$ is

  $$\Sigma_{\mathbf{x},\mathbf{y}} = \mathbf{V}_{\mathbf{x}}^{1/2}\boldsymbol{\rho}_{\mathbf{x},\mathbf{y}}\mathbf{V}_{\mathbf{y}}^{1/2}$$
  $$\boldsymbol{\rho}_{\mathbf{x},\mathbf{y}} = \mathbf{V}_{\mathbf{x}}^{-1/2}\Sigma_{\mathbf{x},\mathbf{y}}\mathbf{V}_{\mathbf{y}}^{-1/2}$$

$$\rho_x = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \ddots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix} =$$

$$\begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & \cdots & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\sigma_{nn}}} \end{pmatrix} \Sigma_x \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & \cdots & 0 \\ 0 & & \frac{1}{\sqrt{\sigma_{nn}}} \end{pmatrix}$$

$$\sigma_{ii} = \text{Var}(x_i)$$

Basic Properties of Mean and Variance of Random Vector

$E(X') = [E(X)]'$

$E(\mathbf{X}+\mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})$.

$E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}$.

- In particular, $E(\mathbf{AX}) = \mathbf{A}\boldsymbol{\mu}_{\mathbf{x}}$.

$\text{cov}(\mathbf{x}, \mathbf{y}) = \text{cov}(\mathbf{y}, \mathbf{x})^T$.

$E((x - u_x)\cdot(y - u_y)')$

$\text{cov}(\mathbf{x}+\mathbf{c}, \mathbf{y}+\mathbf{d}) = \text{cov}(\mathbf{x}, \mathbf{y})$.

$\text{cov}(\mathbf{Ax}, \mathbf{By}) = \mathbf{A}\text{cov}(\mathbf{x}, \mathbf{y})\mathbf{B}^T$

$cov(ax, by) = a\cdot b\, cov(x,y)$

$\text{cov}(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}_1) = \text{cov}(\mathbf{x}_1, \mathbf{y}_1) + \text{cov}(\mathbf{x}_2, \mathbf{y}_1)$

$\text{var}(\mathbf{x}_1 + \mathbf{c}) = \text{cov}(\mathbf{x}_1 + \mathbf{c}, \mathbf{x}_1 + \mathbf{c}) = \text{cov}(\mathbf{x}_1, \mathbf{x}_1) = \text{var}(\mathbf{x}_1)$.

$\text{var}(\mathbf{Ax}) = \mathbf{A}\text{var}(\mathbf{x})\mathbf{A}^T$.

$\text{var}(\mathbf{x}_1 + \mathbf{x}_2) = \text{cov}(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_1 + \mathbf{x}_2) = \text{var}(\mathbf{x}_1) + \text{cov}(\mathbf{x}_1, \mathbf{x}_2) + \text{cov}(\mathbf{x}_2, \mathbf{x}_1) + \text{var}(\mathbf{x}_2)$.

$$\text{var}(\sum_{i=1}^{n} \mathbf{x}_i) = \sum_{i=1}^{n} \text{var}(\mathbf{x}_i), \quad \text{if } \mathbf{x}_1, \ldots, \mathbf{x}_n \text{ are independent.}$$

$$cov(Ax, By) = E((Ax - Au_x)\cdot(By - Bu_y)')$$
$$= A\cdot E((x-u_x)\cdot(y-u_y)')B'$$
$$= A\cdot cov(x,y)\, B'$$

proof of $E(AXB) = A\,E(X)\,B$

$E(AXB) = A\,E(X) \cdot B$

$E(AX) = A\,E(X) \quad (*_1)$ ✓

$E(X'A') = E(X') \cdot A'$

$E(XB) = E(X)\,B$

$A = (a_{ij})_{n \times p}, \quad X = (x_{ij})_{p \times m}$

$A \cdot X = A \cdot (x_1, x_2, \cdots, x_m)$

$$E(AX) = E([Ax_1, \cdots, Ax_m])$$
$$= [E(Ax_1), \cdots, E(Ax_m)]$$
$$= [\underbrace{A\,E(x_1)}, \cdots, A\,E(x_m)] = A \cdot E(X)$$

$$A = \begin{pmatrix} a_1' \\ \vdots \\ a_n' \end{pmatrix}$$

$$A x_j = \begin{pmatrix} a_1' x_j \\ \vdots \\ a_n' x_j \end{pmatrix}$$

$$E(Ax_j) = \begin{pmatrix} E(a_1' x_j) \\ \vdots \\ E(a_n' x_j) \end{pmatrix}$$

$$= \begin{pmatrix} a_1' E(x_j) \\ \vdots \\ a_n' E(x_j) \end{pmatrix} = A \cdot E(x_j)$$

$$E(a_i' x_j) = E\left(\sum_{k=1}^{p} a_{ik} \cdot x_{kj} \right)$$

$$= \sum_{k=1}^{p} a_{ik} \cdot E(x_{kj})$$

$$= \underline{a_i' E(x_j)}$$

where $a_i = (a_{i1}, \cdots, a_{ip})'$

$$Var(x_1+x_2) = Var(x_1) + Cov(x_1, x_2) + Cov(x_2, x_1) + Var\, x_2$$

pf:

$$Var(x_1+x_2) = E[((x_1 + x_2 - u_1 - u_2)) \cdot (x_1+x_2-u_1-u_2)']$$

$$= E((x_1-u_1)\cdot(x_1-u_1)' + (x_1-u_1)\cdot(x_2-u_2)' + (x_2-u_2)\cdot(x_1-u_1)' + (x_2-u_2)(x_2-u_2)')$$

$$= Var(x_1) + \underline{Cov(x_1, x_2)} + \underline{Cov(x_2, x_1)} + Var(x_2)$$

Definition of Multivariate Normal Distribution

Independent Standard Normal $z \sim N(0, I_p)$

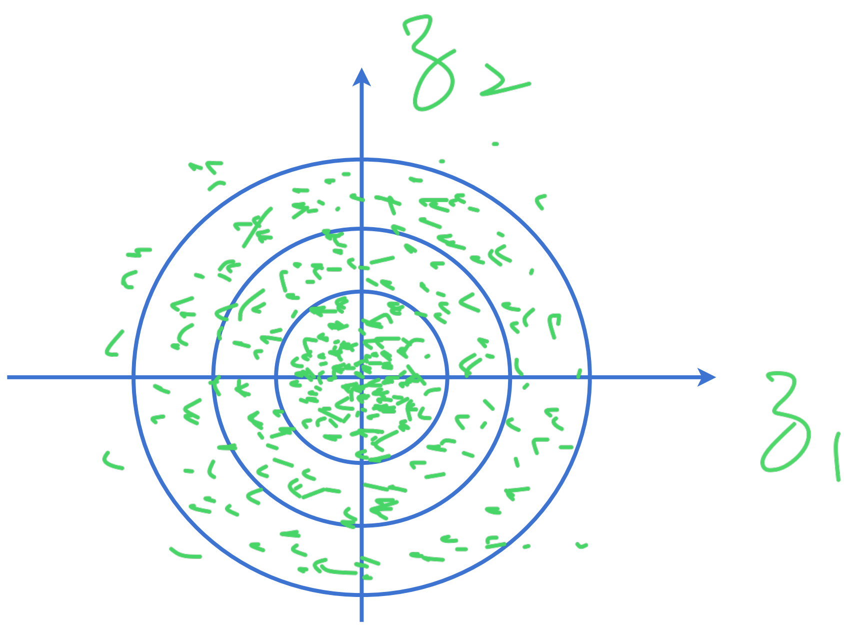
$z = (z_1, \cdots z_n)' \sim N(0, I_n)$

PDF:

$$f(z) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}}$$

$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^{n} z_i^2}{2}}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{\|z\|^2}{2}}$$



$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \qquad \begin{array}{l} Cov(z_i, z_j) = 0 \\ E(z_i) = 0 \\ V(z_i) = 1 \end{array}$$

$$\mu_z = E(z) = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\Sigma_z = Cov(z) = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}}_{I_n}$$

$$z \sim N_n(0, I_n)$$

Definition with a Linear Transformation

Multivariate Normal Distribution: A random vector $\mathbf{y}_{n\times 1}$ is said to have a multivariate normal distribution if $\mathbf{y}$ has the same distribution as

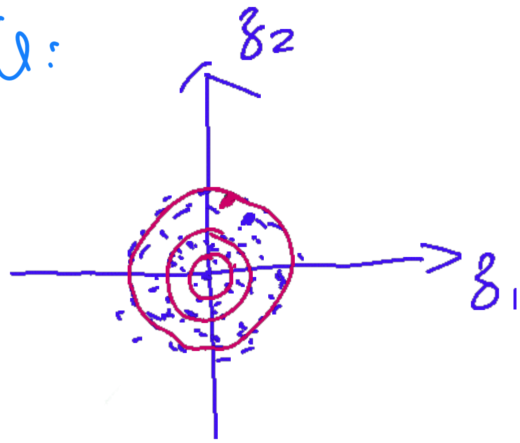
$$\mathbf{A}_{n\times p}\mathbf{z}_{p\times 1} + \boldsymbol{\mu}_{n\times 1} \equiv \mathbf{x}$$

$z \sim N_p(0, I_p)$

where, for some p, $\mathbf{z}$ is a vector of independent $N(0,1)$ random variables, $\mathbf{A}$ is a matrix of constants, and $\boldsymbol{\mu}$ is a vector of constants.

$E(x) = \mu$, $Var(x) = A \cdot A' = \Sigma$

Example:



$$f(z_1, z_2) = \frac{1}{2\pi} \cdot e^{-\frac{z_1^2+z_2^2}{2}} = \frac{1}{2\pi} e^{-\frac{\|z\|^2}{2}} \checkmark$$

$z = (z_1, z_2)'$ $\qquad$ $z_1, z_2 \overset{iid}{\sim} N(0,1)$

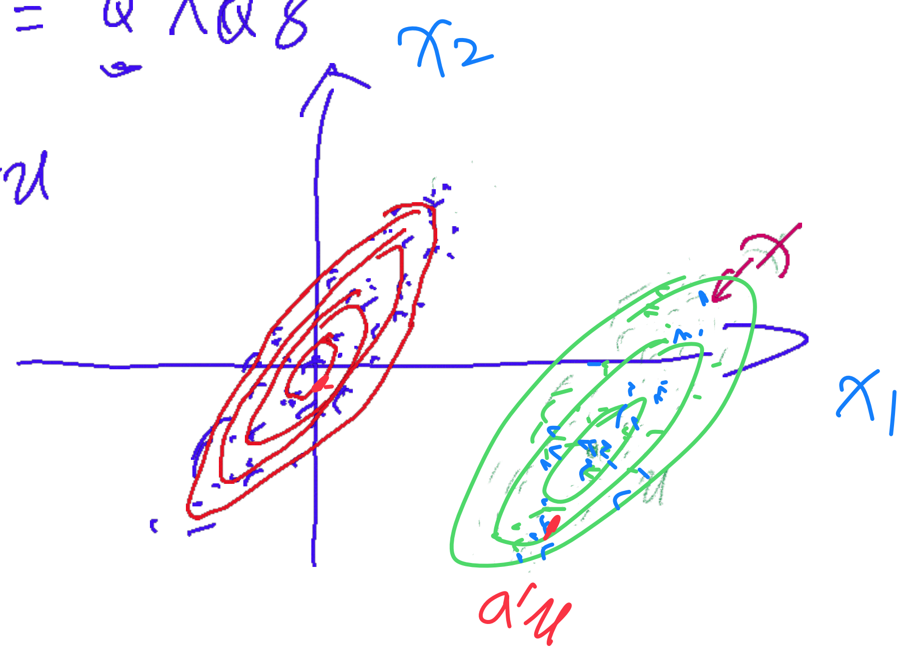
Assume $\underset{\sim}{A}$ is a symmetric matrix

$$A = (q_1, q_2) \cdot \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \cdot \begin{pmatrix} q_1' \\ q_2' \end{pmatrix}$$

$$= \underset{\sim}{Q} \cdot \Lambda \cdot \underset{\sim}{Q}'$$

$$Az = \underset{\sim}{Q} \Lambda Q' z$$

$Az + u$



$a'x$

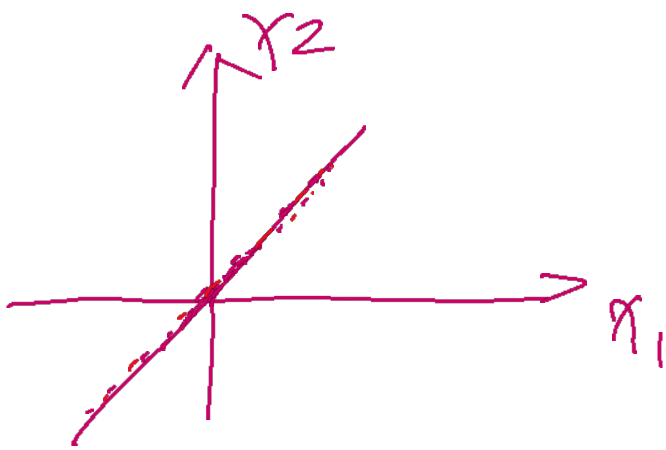
Example:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A_1 z = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

$$= \begin{pmatrix} z_1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A_2 z = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

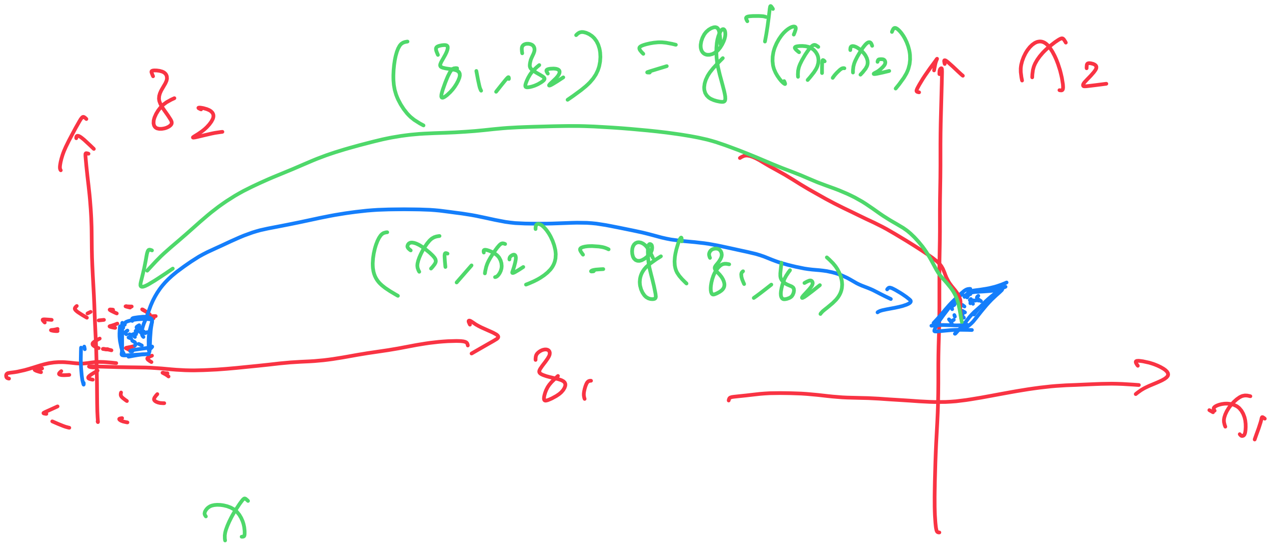
$$= \begin{pmatrix} z_1 \\ z_1 \end{pmatrix}$$



$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$$

such x don't have a <u>P.D.F.</u>

General formular for the PDF of Transformed Random Variables



$$f_X(x_1, x_2) \cdot |\Delta(x_1, x_2)| = f_Z(z_1, z_2) \cdot |\Delta(z_1, z_2)|$$

$$f_X(x_1, x_2) = f_Z(z_1, z_2) \cdot \left\| \frac{\Delta(z_1, z_2)}{\Delta(x_1, x_2)} \right\|$$

$$= f(z_1, z_2) \cdot \left\| \frac{\partial(z_1, z_2)}{\partial(x_1, x_2)} \right\|_+$$

$$\left| \begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} \end{pmatrix} \right|_+$$

Probability Density Function

$x = Az + \mu \iff z = A^{-1}(x - \mu)$

Define $g(\mathbf{z}) = \mathbf{A}\mathbf{z} + \boldsymbol{\mu}$ to be the transformation from $\mathbf{z}$ to $\mathbf{x}$. For $\mathbf{A}$ a $p \times p$ full rank matrix, $g(\mathbf{z})$ is a 1–1 function from $\mathcal{R}^p$ to $\mathcal{R}^p$ so that we can use the following change of variable formula for the density of $\mathbf{x}$:

$$f_{\mathbf{x}}(\mathbf{x}) = f_{\mathbf{z}}\{g^{-1}(\mathbf{x})\}\text{abs}\left(\left|\frac{\partial g^{-1}(\mathbf{x})}{\partial \mathbf{x}^T}\right|\right) = f_{\mathbf{z}}\{\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}\text{abs}(|\mathbf{A}^{-1}|).$$

$$f_{\mathbf{x}}(\mathbf{x}) = (2\pi)^{-p/2}\underbrace{\text{abs}(|\mathbf{A}|)}_{=|\mathbf{A}|}{}^{-1} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

Note that $\Sigma = \text{var}(\mathbf{x})$ is equal to $\text{var}(\mathbf{A}\mathbf{z}+\boldsymbol{\mu}) = \text{var}(\mathbf{A}\mathbf{z}) = \mathbf{A}\mathbf{I}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T$, so

$$|\Sigma| = |\mathbf{A}\mathbf{A}^T| = |\mathbf{A}|^2 \quad \Rightarrow \quad |\mathbf{A}| = |\Sigma|^{1/2}.$$

In addition, $\text{E}(\mathbf{x}) = \text{E}(\mathbf{A}\mathbf{z} + \boldsymbol{\mu}) = \boldsymbol{\mu}$.

So, a multivariate normal random vector of dimension p with mean $\boldsymbol{\mu}$ and p.d. var-cov matrix Σ has density

$$f_{\mathbf{x}}(\mathbf{x}) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T(\Sigma)^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}, \quad \text{for all } \mathbf{x} \in \mathcal{R}^p.$$

$$f_z(z) = \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{\sum z_i^2}{2}} = \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{\|z\|^2}{2}}$$

$$\left|\frac{\partial z}{\partial x}\right| = |A^{-1}| = |A|^{-1} = |\Sigma|^{-\frac{1}{2}}$$

where $\Sigma = A \cdot A'$

Note: P.D.F. may not exist for all multivariate normal.

Definition of M.G.F.

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

$$M_X(t) = E\left(e^{tX}\right)$$

$$M_X(t) = E\left(e^{\sum_{i=1}^{n} t_i X_i}\right)$$

$$= E\left(e^{t'X}\right)$$

Moment Generating Function

The m.g.f. of a random vector $\mathbf{x}$ is $m_{\mathbf{x}}(\mathbf{t}) = \mathrm{E}(e^{\mathbf{t}^T\mathbf{x}})$. So, for $\mathbf{x} = \mathbf{Az}+\boldsymbol{\mu} \sim N_n(\boldsymbol{\mu}, \mathbf{AA}^T = \Sigma)$, the m.g.f. of $\mathbf{x}$ is

$$m_{\mathbf{x}}(\mathbf{t}) = \mathrm{E}[\exp\{\mathbf{t}^T(\mathbf{Az}+\boldsymbol{\mu})\}] = e^{\mathbf{t}^T\boldsymbol{\mu}}\mathrm{E}(e^{\overset{u'}{\mathbf{t}^T\mathbf{A}}\mathbf{z}}) = e^{\mathbf{t}^T\boldsymbol{\mu}}m_{\mathbf{z}}(\underset{\downarrow}{\overset{u}{\mathbf{A}^T\mathbf{t}}}). \qquad (*)$$

$t = (t_1, \ldots, t_p)'$

The m.g.f. of a standard normal r.v. z_i is $m_{z_i}(u) = e^{u^2/2}$, so the m.g.f. of $\mathbf{z}$ is

$$m_{\mathbf{z}}(\mathbf{u}) = \prod_{i=1}^{p}\exp(u_i^2/2) = e^{\mathbf{u}^T\mathbf{u}/2}. = e^{\frac{\|u\|^2}{2}}$$

$$= E(e^{u'z})$$

Substituting into (*) we get

$$m_{\mathbf{x}}(\mathbf{t}) = e^{\mathbf{t}^T\boldsymbol{\mu}}\exp\{\frac{1}{2}(\mathbf{A}^T\mathbf{t})^T(\mathbf{A}^T\mathbf{t})\} = \underline{e^{\mathbf{t}^T\boldsymbol{\mu}}}\exp(\underline{\frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}}).$$

$$= \exp(t'u + \tfrac{1}{2}t^T\Sigma t)$$

We now list two important properties of moment generating functions.

1. If two random vectors have the same moment generating function, they have the same ~~density~~. distribution.
2. Two random vectors are independent if and only if their joint moment generating function factors into the product of their two separate moment generating functions; that is, if $\mathbf{y}' = (\mathbf{y}_1', \mathbf{y}_2')$ and $\mathbf{t}' = (\mathbf{t}_1', \mathbf{t}_2')$, then $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent if and only if

$$t'y = (t_1', t_2')\cdot\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$M_{\mathbf{y}}(\mathbf{t}) = M_{\mathbf{y}_1}(\mathbf{t}_1)M_{\mathbf{y}_2}(\mathbf{t}_2). \qquad (4.23)$$

$$M_y(t) = E(e^{t_1'y_1 + t_2'y_2})$$
$$= E(e^{t_1'y_1}\cdot e^{t_2'y_2}) = E(e^{t_1'y_1})E(e^{t_2'y_2})$$
$$= M_{y_1}(t_1)\cdot M_{y_2}(t_2)$$

Constructing Multivariate Normal (MVN) Random Vector

Theorem: Let $\boldsymbol{\mu}$ be an element of $\mathcal{R}^n$ and Σ an $n \times n$ symmetric p.s.d. matrix. Then there exists a multivariate normal distribution with ~~mean~~ $\boldsymbol{\mu}$ and var-cov matrix Σ.

Proof: Since Σ is symmetric and p.s.d., there exists a $\mathbf{B}$ so that $\Sigma = \mathbf{B}\mathbf{B}^T$ (e.g., the Cholesky decomposition). Let $\mathbf{z}$ be an $n \times 1$ vector of independent standard normals. Then $\mathbf{x} = \mathbf{B}\mathbf{z} + \boldsymbol{\mu} \sim N_n(\boldsymbol{\mu}, \Sigma)$. $\blacksquare$

Another approach:

$$\Sigma = Q \cdot \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} Q'$$

$$= \Sigma^{\frac{1}{2}} \cdot \Sigma^{\frac{1}{2}}$$

where $\Sigma^{\frac{1}{2}} = Q \cdot \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} Q'$

$\Sigma^{\frac{1}{2}}$ is symmetric

Let $A = \Sigma^{\frac{1}{2}}$

A isn't unique.

Linear Transformation of Multivariate Normal

Linear Transformation

$$\begin{cases} E(y) = C \cdot \mu + d \\ Var(y) = C \cdot \Sigma \cdot C^T \end{cases}$$

Theorem: Let $\mathbf{x} \sim N_n(\boldsymbol{\mu}, \Sigma)$ where Σ is ~~p.d.~~ p.s.d. Let $\mathbf{y}_{r\times 1} = \mathbf{C}_{r\times n}\mathbf{x} + \mathbf{d}$ for $\mathbf{C}$ and $\mathbf{d}$ containing constants. Then $\mathbf{y} \sim N_r(\mathbf{C}\boldsymbol{\mu} + \mathbf{d}, \mathbf{C}\Sigma\mathbf{C}^T)$.

Proof: By definition, $\mathbf{x} = \mathbf{A}\mathbf{z} + \boldsymbol{\mu}$ for some $\mathbf{A}$ such that $\mathbf{A}\mathbf{A}^T = \Sigma$, and $\mathbf{z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$. Then

$$\begin{aligned} \mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{d} = \mathbf{C}(\mathbf{A}\mathbf{z} + \boldsymbol{\mu}) + \mathbf{d} &= (\mathbf{C}\mathbf{A})\mathbf{z} + (\mathbf{C}\boldsymbol{\mu} + \mathbf{d}) \\ &= (\mathbf{C}\mathbf{A})\mathbf{z} + (\mathbf{C}\boldsymbol{\mu} + \mathbf{d}). \end{aligned}$$

So, by definition, $\mathbf{y}$ has a multivariate normal distribution with mean $\mathbf{C}\boldsymbol{\mu} + \mathbf{d}$ and var-cov matrix $(\mathbf{C}\mathbf{A})(\mathbf{C}\mathbf{A})^T = \mathbf{C}\Sigma\mathbf{C}^T$. ∎

Note that *y* may not have a PDF

Simple corollaries of this theorem are that if $\mathbf{x} \sim N_n(\boldsymbol{\mu}, \Sigma)$, then

i. any subvector of $\mathbf{x}$ is multivariate normal too, with mean and variance given by the corresponding subvector of $\boldsymbol{\mu}$ and submatrix of Σ, respectively, and (Marginal distribution of MVN)

Pf:

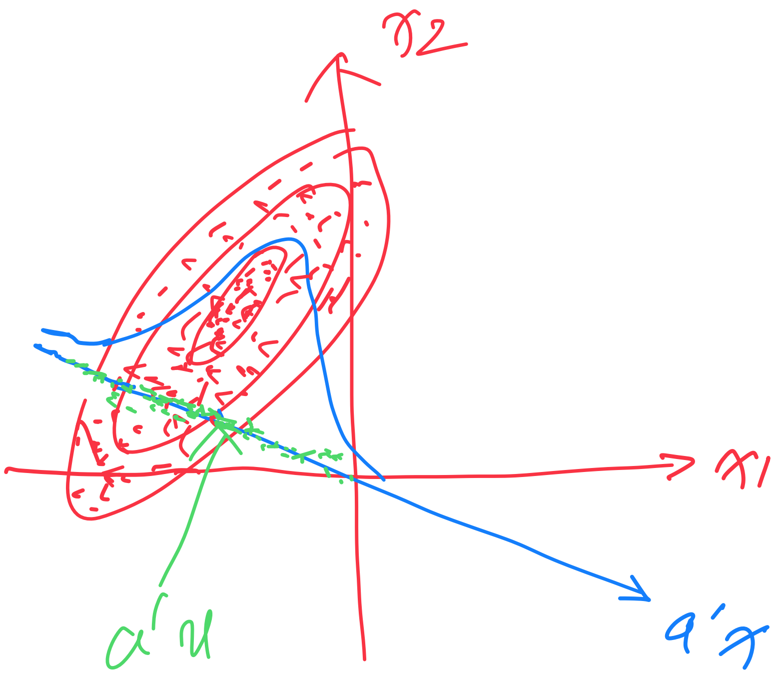
i) $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \}r$, $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

$C = (\underset{r\times r}{I_r}, \underset{r\times(n-r)}{0})$, $Cx = (I_r, 0) \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1$

$Cx = x_1$, $Cu = u_1$, $C\Sigma C' = \Sigma_{11}$

$x_1 \sim N_r(u_1, \Sigma_{11})$ $\quad (I_r, 0_1) \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I_r \\ 0_1' \end{pmatrix}$

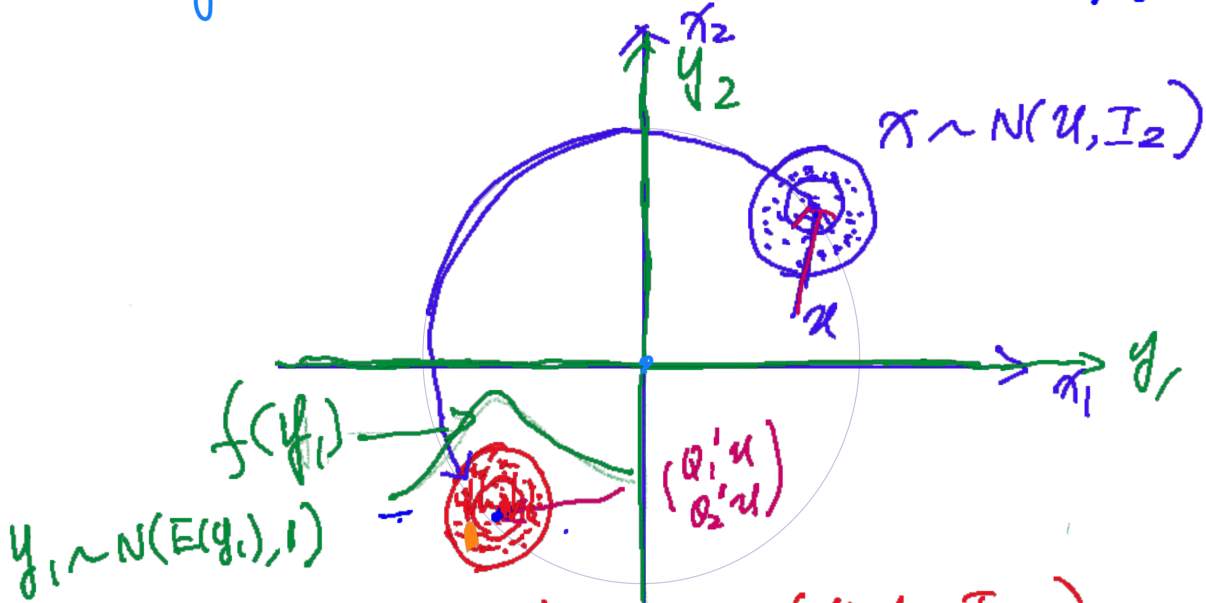
ii. any linear combination $\mathbf{a}^T\mathbf{x} \sim N(\mathbf{a}^T\boldsymbol{\mu}, \mathbf{a}^T\Sigma\mathbf{a})$ (univariate normal) for $\mathbf{a}$ a vector of constants.



$V(a'x) = a'\Sigma a$

iii) $\Sigma = I_n$, $C = Q$ with $Q'Q = QQ' = I_n$.

$y = Q'x \sim N(Q'u, I_n)$ ⟵ $Q' \cdot I_n \cdot Q = I_n$



$\underline{y} = Q'x \sim N(Q'u, I_2)$

iv) $Q_1 = (q_1, \cdots, q_r)$, $q_i \perp q_j$, $r \le n$

for $i \ne j$, $\|q_i\| = 1$, $x \sim N(u, I_n)$

$$Q_1' x \sim N(Q_1' u,\ Q_1' I_n Q_1 = I_r)$$

Pf: $Q = (Q_1, Q_2)$, $QQ' = Q'Q = I_n$

$$Q'x = \begin{pmatrix} Q_1'x \\ Q_2'x \end{pmatrix} \sim N_n\left(\begin{pmatrix} Q_1'u \\ Q_2'u \end{pmatrix}, \begin{pmatrix} \overset{Q_1'Q_1}{I_r} & 0 \\ 0 & I_{n-r} \end{pmatrix} \right)$$

$Q'Q$

v) $y \sim N_n(\mu, \Sigma)$

$$\Sigma^{-\frac{1}{2}} y \sim N_n(\Sigma^{-\frac{1}{2}}\mu, I_n)$$

$$\Sigma^{-\frac{1}{2}}(y-\mu) \sim N_n(0, I_n)$$

$\underline{Pf}$:

$$Var(\Sigma^{-\frac{1}{2}} y) = \Sigma^{-\frac{1}{2}} \cdot \Sigma \cdot \Sigma^{-\frac{1}{2}}$$

$$= I_n$$

Independence in MVN

$\Sigma_{12}=0$

Theorem: Let $\mathbf{y}_{n\times 1}$ have a multivariate normal distribution, and partition $\mathbf{y}$ as

$$\mathbf{y} = \begin{pmatrix} \underset{p\times 1}{\mathbf{y}_1} \\ \cdots \\ \underset{(n-p)\times 1}{\mathbf{y}_2} \end{pmatrix}.$$

$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$

Then $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent if and only if $\mathrm{cov}(\mathbf{y}_1, \mathbf{y}_2) = \mathbf{0}$.

Proof: 1^{st}, independence implies 0 covariance: Suppose $\mathbf{y}_1, \mathbf{y}_2$ are independent with means $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$. Then

$$\mathrm{cov}(\mathbf{y}_1, \mathbf{y}_2) = \mathrm{E}\{(\mathbf{y}_1 - \boldsymbol{\mu}_1)(\mathbf{y}_2 - \boldsymbol{\mu}_2)^T\} = \mathrm{E}\{(\mathbf{y}_1 - \boldsymbol{\mu}_1)\}\mathrm{E}\{(\mathbf{y}_2 - \boldsymbol{\mu}_2)^T\} = \mathbf{0}(\mathbf{0}^T) = \mathbf{0}.$$

2^{nd}, 0 covariance and normality imply independence: To do this we use the fact that two random vectors are independent if and only if their joint m.g.f. is the product of their marginal m.g.f.'s. Suppose $\mathrm{cov}(\mathbf{y}_1, \mathbf{y}_2) = \mathbf{0}$. Let $\mathbf{t}_{n\times 1}$ be partitioned as $\mathbf{t} = (\mathbf{t}_1^T, \mathbf{t}_2^T)^T$ where $\mathbf{t}_1$ is $p \times 1$. Then $\mathbf{y}$ has m.g.f.

$$m_{\mathbf{y}}(\mathbf{t}) = \exp(\underbrace{\mathbf{t}^T \boldsymbol{\mu}}_{=\mathbf{t}_1^T \boldsymbol{\mu}_1 + \mathbf{t}_2^T \boldsymbol{\mu}_2}) \exp(\frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t}),$$

where

$$\Sigma = \mathrm{var}(\mathbf{y}) = \begin{pmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \mathrm{var}(\mathbf{y}_1) & \mathbf{0} \\ \mathbf{0} & \mathrm{var}(\mathbf{y}_2) \end{pmatrix}$$

Because of the form of Σ, $\mathbf{t}^T \Sigma \mathbf{t} = \mathbf{t}_1^T \Sigma_{11} \mathbf{t}_1 + \mathbf{t}_2^T \Sigma_{22} \mathbf{t}_2$, so

$$\begin{aligned} m_{\mathbf{y}}(\mathbf{t}) &= \exp(\mathbf{t}_1^T \boldsymbol{\mu}_1 + \frac{1}{2}\mathbf{t}_1^T \Sigma_{11} \mathbf{t}_1 + \mathbf{t}_2^T \boldsymbol{\mu}_2 + \frac{1}{2}\mathbf{t}_2^T \Sigma_{22} \mathbf{t}_2) \\ &= \exp(\mathbf{t}_1^T \boldsymbol{\mu}_1 + \frac{1}{2}\mathbf{t}_1^T \Sigma_{11} \mathbf{t}_1) \exp(\mathbf{t}_2^T \boldsymbol{\mu}_2 + \frac{1}{2}\mathbf{t}_2^T \Sigma_{22} \mathbf{t}_2) = m_{\mathbf{y}_1}(\mathbf{t}_1) m_{\mathbf{y}_2}(\mathbf{t}_2). \end{aligned}$$

Cor: $X \sim N(\mu, \Sigma)$,
$AX \perp BX$ iff $A \Sigma B' = 0$

Pf: $\mathrm{Cov}(AX, BX) = \underline{A \Sigma B'}$

Cor: $X \sim N(\mu, I_n)$. $AX \perp BX \Leftrightarrow \underline{A \cdot B' = 0}$

$$t'\Sigma^{-1}t$$

$$= (t_1', t_2')\begin{pmatrix}\Sigma_{11} & 0 \\ 0 & \Sigma_{22}\end{pmatrix}^{-1}\begin{pmatrix}t_1 \\ t_2\end{pmatrix}$$

$$= (t_1', t_2')\begin{pmatrix}\Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1}\end{pmatrix}\begin{pmatrix}t_1 \\ t_2\end{pmatrix}$$

$$= (t_1'\Sigma_{11}^{-1}, \; t_2'\Sigma_{22}^{-1}) \cdot \begin{pmatrix}t_1 \\ t_2\end{pmatrix}$$

$$= t_1'\Sigma_{11}^{-1}t_1 + t_2'\Sigma_{22}^{-1}t_2$$

$\boxed{\text{Conditional Multivariate Normal}}$

An important Lemma: Constructing independent Random Vector

Lemma: Let $\mathbf{y} \sim N_n(\boldsymbol{\mu}, \Sigma)$ where we have the partitioning

$$\mathbf{y} = \begin{pmatrix} \underset{p\times 1}{\mathbf{y}_1} \\ \cdots \\ \underset{(n-p)\times 1}{\mathbf{y}_2} \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where $\Sigma_{21} = \Sigma_{12}^T$. Let $\mathbf{y}_{2|1} = \mathbf{y}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathbf{y}_1$. Then $\mathbf{y}_1$ and $\mathbf{y}_{2|1}$ are independent with

$$\mathbf{y}_1 \sim N_p(\boldsymbol{\mu}_1, \Sigma_{11}), \quad \mathbf{y}_{2|1} \sim N_{n-p}(\boldsymbol{\mu}_{2|1}, \Sigma_{22|1}),$$

where

$$\boldsymbol{\mu}_{2|1} = \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1, \quad \text{and} \quad \Sigma_{22|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

Proof: We can write $\mathbf{y}_1 = \mathbf{C}_1\mathbf{y}$ where $\mathbf{C}_1 = (\mathbf{I}, \mathbf{0})$ and we can write $\mathbf{y}_{2|1} = \mathbf{C}_2\mathbf{y}$ where $\mathbf{C}_2 = (-\Sigma_{21}\Sigma_{11}^{-1}, \mathbf{I})$, so by the theorem on the bottom of p. 72, both $\mathbf{y}_1$ and $\mathbf{y}_{2|1}$ are normal. Their mean and variance-covariances are $\mathbf{C}_1\boldsymbol{\mu} = \boldsymbol{\mu}_1$ and $\mathbf{C}_1\Sigma\mathbf{C}_1^T = \Sigma_{11}$ for $\mathbf{y}_1$, and $\mathbf{C}_2\boldsymbol{\mu} = \boldsymbol{\mu}_{2|1}$ and $\mathbf{C}_2\Sigma\mathbf{C}_2^T = \Sigma_{22|1}$ for $\mathbf{y}_{2|1}$. Independence follows from the fact that these two random vectors have covariance matrix $\text{cov}(\mathbf{y}_1, \mathbf{y}_{2|1}) = \text{cov}(\mathbf{C}_1\mathbf{y}, \mathbf{C}_2\mathbf{y}) = \mathbf{C}_1\Sigma\mathbf{C}_2^T = \mathbf{0}$.

$$\begin{pmatrix} y_1 \\ y_{2|1} \end{pmatrix} = \begin{pmatrix} I_p & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{n-p} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$\Sigma_{11}$$

$$\cdot \begin{pmatrix} -\Sigma_{11}^{-1}\Sigma_{12} \\ I_{n-p} \end{pmatrix}$$

$$= -\Sigma_{12} + \Sigma_{12} = 0$$

$$C = \begin{pmatrix} I_p & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{n-p} \end{pmatrix}$$

$$C \cdot \Sigma \cdot C'$$

$$= \begin{pmatrix} I_p & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{n-p} \end{pmatrix} \cdot \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \cdot C'$$

$$= \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ 0 & \Sigma_{22} - \Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix} \begin{pmatrix} I_p & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I_{n-p} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{12}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}$$

$$C \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ \underset{\overset{\uparrow}{y_{2|1}}}{y_2 - \Sigma_{21}\Sigma_{11}^{-1} y_1} \end{pmatrix}$$

Conditional MVN

Theorem: For $\mathbf{y}$ defined as in the previous theorem, the conditional distribution of $\mathbf{y}_2$ given $\mathbf{y}_1$ is

$$\mathbf{y}_2|\mathbf{y}_1 \sim N_{n-p}(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1), \Sigma_{22|1}).$$

z_2 $\quad z_1$

$$y_2 = (y_2 - \Sigma_{21}\Sigma_{11}^{-1} y_1) + \Sigma_{21}\Sigma_{11}^{-1} y_1$$

Proof: Since $\mathbf{y}_{2|1}$ is independent of $\mathbf{y}_1$, its conditional distribution for a given value of $\mathbf{y}_1$ is the same as its marginal distribution, $\mathbf{y}_{2|1} \sim N_{n-p}(\boldsymbol{\mu}_{2|1}, \Sigma_{22|1})$. Notice that $\mathbf{y}_2 = \mathbf{y}_{2|1} + \Sigma_{21}\Sigma_{11}^{-1}\mathbf{y}_1$. Conditional on the value of $\mathbf{y}_1$, $\Sigma_{21}\Sigma_{11}^{-1}\mathbf{y}_1$ is constant, so the conditional distribution of $\mathbf{y}_2$ is that of $\mathbf{y}_{2|1}$ plus a constant, or $(n-p)-$variate normal, with mean

$$\boldsymbol{\mu}_{2|1} + \Sigma_{21}\Sigma_{11}^{-1}\mathbf{y}_1 = \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1 + \Sigma_{21}\Sigma_{11}^{-1}\mathbf{y}_1 = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1),$$

and var-cov matrix $\Sigma_{22|1}$. ∎ $\quad \Sigma_{22|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$

Remarks:

1) A link to least square estimator (coming back)

$$E(y_2|y_1) = u_2 + (y_1 - u_1)'\Sigma_{11}^{-1}\Sigma_{12}$$

Let $y_2 = y$, $y_1 = x$, $u_1 = u_x$, $u_2 = u_y$, $\Sigma_{11}^{-1}\Sigma_{12}$

$$E(y|x) = u_y + (x - u_x)$$

$$y|x \sim \alpha_0 + X_c\alpha_1 + \varepsilon$$

$$\hat{\alpha}_1 = (X_c'X_c)^{-1}X_c'y = \hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}$$

$$\hat{\alpha}_0 = \bar{y} = \hat{u}_y$$

In simple linear regression (Q7 of HW1)

$$\hat{\alpha}_1 = S_{xx}^{-1}S_{xy}, \quad \hat{\alpha}_0 = \bar{y}$$

Rao-Blackwell formular

2) A link to Variance decomposition

$$V(y_2) = E(V(y_2|y_1)) + V(E(y_2|y_1))$$

$$E(y_2|y_1) = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(y_1 - \mu_1)$$

$$V(E(y_2|y_1)) = (\Sigma_{21}\Sigma_{11}^{-1})(\Sigma_{11})(\Sigma_{11}^{-1}\Sigma_{12})$$

$$= \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

$$V(y_2) = \Sigma_{22}$$

$$E(V(y_2|y_1)) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

when (y_1, y_2) follows M.N.,

$$V(y_2|y_1) = E(V(y_2|y_1))$$

Example

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \right)$$

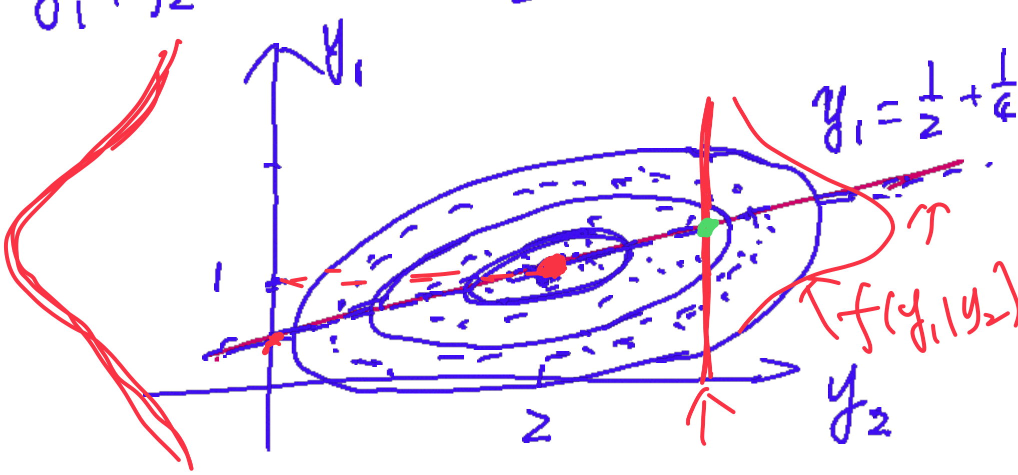
$$\Sigma_{12}\Sigma_{22}^{-1} = 1 \times \frac{1}{4} = \frac{1}{4},$$

$$E(y_1|y_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2)$$
$$= 1 + \frac{1}{4}\cdot(y_2 - 2) = \frac{1}{2} + \frac{1}{4}y_2$$

$$V(y_1|y_2) = 2 - 1 \times \frac{1}{4} \times 1 = \frac{7}{4}$$

$$\text{or } \underbrace{E(V(y_1|y_2))}_{=\,V(y_1|y_2)} = V(y_1) - V(E(y_1|y_2))$$
$$= 2 - \left(\frac{1}{4}\right)^2 \times 4 = \frac{7}{4}$$

$$y_1|y_2 \sim N\left(\frac{1}{2} + \frac{1}{4}y_2, \frac{7}{4}\right).$$



$y_1 = \frac{1}{2} + \frac{1}{4}y_2 + \varepsilon$ with $\varepsilon \sim N(0, \frac{7}{4})$

Example

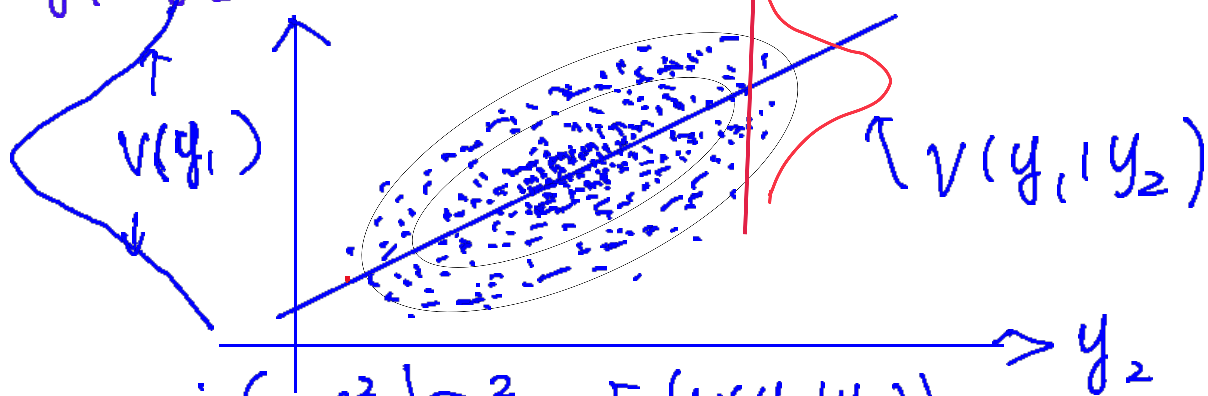
$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right)$$

$$\Sigma_{12}\Sigma_{22}^{-1} = \rho\sigma_1\sigma_2(\sigma_2^2)^{-1} = \rho \cdot \frac{\sigma_1}{\sigma_2}$$

$$E(y_1|y_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2)$$
$$= \mu_1 + \rho \cdot \frac{\sigma_1}{\sigma_2} \cdot (y_2 - \mu_2)$$

$$V(y_1|y_2) = \sigma_1^2 - \rho\sigma_1\sigma_2 \cdot (\sigma_2^2)^{-1} \cdot \rho\sigma_1\sigma_2$$
$$= \sigma_1^2(1-\rho^2).$$

$$y_1|y_2 \sim N(E(y_1|y_2), \sigma_1^2(1-\rho^2))$$



$$\sigma_1^2 < \begin{matrix} (1-\rho^2)\sigma_1^2 = E(V(y_1|y_2)) \\ \rho^2\sigma_1^2 = V(E(y_1|y_2)) \end{matrix}$$

$$\rho^2 = \frac{V(E(y_1|y_2))}{V(y_1)}$$

Example 4.4c. To illustrate Corollary 1 to Theorem 4.4d, let $\mathbf{v}$ be $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are as given in Example 4.4b. If $\mathbf{v}$ is partitioned as $\mathbf{v} = (y, x_1, x_2, x_3)'$, then $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are partitioned as follows:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix} = \left(\begin{array}{c} 2 \\ \hline 5 \\ -2 \\ 1 \end{array}\right),$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_y^2 & \boldsymbol{\sigma}'_{yx} \\ \boldsymbol{\sigma}_{yx} & \boldsymbol{\Sigma}_{xx} \end{pmatrix} = \left(\begin{array}{c|ccc} 9 & 0 & 3 & 3 \\ \hline 0 & 1 & -1 & 2 \\ 3 & -1 & 6 & -3 \\ 3 & 2 & -3 & 7 \end{array}\right).$$

By (4.33), we have

$$\begin{aligned} E(y|x_1, x_2, x_3) &= \mu_y + \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x) \\ &= 2 + (0, 3, 3)\begin{pmatrix} 1 & -1 & 2 \\ -1 & 6 & -3 \\ 2 & -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} x_1 - 5 \\ x_2 + 2 \\ x_3 + 1 \end{pmatrix} \\ &= \tfrac{95}{7} - \tfrac{12}{7}x_1 + \tfrac{6}{7}x_2 + \tfrac{9}{7}x_3. \end{aligned}$$

By (4.34), we obtain

$$\begin{aligned} \text{var}(y|x_1, x_2, x_3) &= \sigma_y^2 - \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx} \\ &= 9 - (0, 3, 3)\begin{pmatrix} 1 & -1 & 2 \\ -1 & 6 & -3 \\ 2 & -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ 3 \\ 3 \end{pmatrix} \\ &= 9 - \tfrac{45}{7} = \tfrac{18}{7}. \end{aligned}$$

Hence $y|x_1, x_2, x_3$ is $N(\frac{95}{7} - \frac{12}{7}x_1 + \frac{6}{7}x_2 + \frac{9}{7}x_3, \frac{18}{7})$. Note that $\text{var}(y|x_1, x_2, x_3) = \frac{18}{7}$ is less than $\text{var}(y) = 9$, which illustrates (4.35). □

Partial Correlation: Suppose $\mathbf{v} \sim N_{p+q}(\boldsymbol{\mu}, \Sigma)$ and let $\mathbf{v}$, $\boldsymbol{\mu}$ and Σ be partitioned as

$$\mathbf{v} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \boldsymbol{\mu}_{\mathbf{y}} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{pmatrix},$$

where $\mathbf{x} = (v_1, \ldots, v_p)^T$ is $p \times 1$ and $\mathbf{y} = (v_{p+1}, \ldots, v_{p+q})^T$ is $q \times 1$.

Recall that the conditional var-cov matrix of $\mathbf{y}$ given $\mathbf{x}$ is

$$\mathrm{var}(\mathbf{y}|\mathbf{x}) = \Sigma_{\mathbf{yy}} - \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}} \equiv \Sigma_{\mathbf{y}|\mathbf{x}}.$$

Let $\sigma_{ij|1,\ldots,p}$ denote the $(i,j)^{\text{th}}$ element of $\Sigma_{\mathbf{y}|\mathbf{x}}$.

Then the **partial correlation coefficient** of y_i and y_j given $\mathbf{x} = \mathbf{c}$ is defined by

$$\rho_{ij|1,\ldots,p} = \frac{\sigma_{ij|1,\ldots,p}}{[\sigma_{ii|1,\ldots,p}\sigma_{jj|1,\ldots,p}]^{1/2}}$$

Multiple Correlation: Suppose $\mathbf{v} \sim N_{p+1}(\boldsymbol{\mu}, \Sigma)$ and let $\mathbf{v}$, $\boldsymbol{\mu}$ and Σ be partitioned as

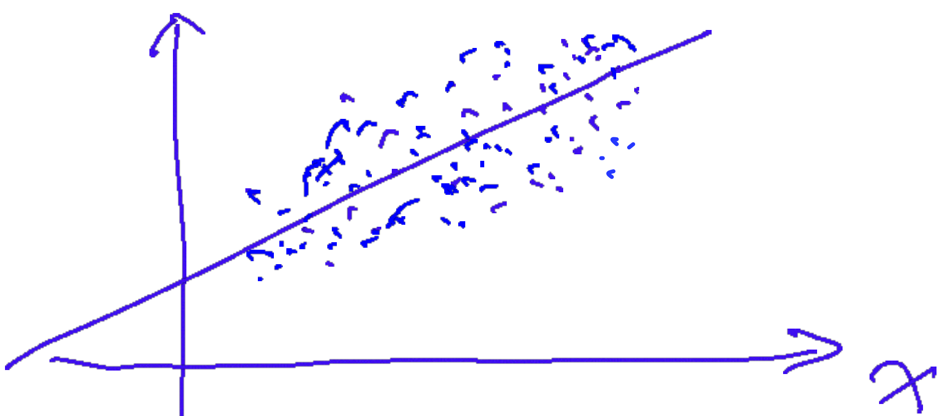
$$\mathbf{v} = \begin{pmatrix} \mathbf{x} \\ y \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_\mathbf{x} \\ \mu_y \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y} \\ \boldsymbol{\sigma}_{y\mathbf{x}} & \sigma_{yy} \end{pmatrix},$$

where $\mathbf{x} = (v_1, \ldots, v_p)^T$ is $p \times 1$, and $y = v_{p+1}$ is a scalar random variable.

Recall that the conditional mean of y given $\mathbf{x}$ is

$$E(y|x) = \mu_y + \boldsymbol{\sigma}_{y\mathbf{x}} \Sigma_{\mathbf{xx}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})$$

$$\rho_{y|x}^2 = \frac{V(E(y|x))}{Var(y)} = \frac{\sigma_{yx} \Sigma_{xx}^{-1} \sigma_{xy}}{\sigma_{yy}}$$



$$Var(y) = Var(E(y|x)) + E(Var(y|x))$$

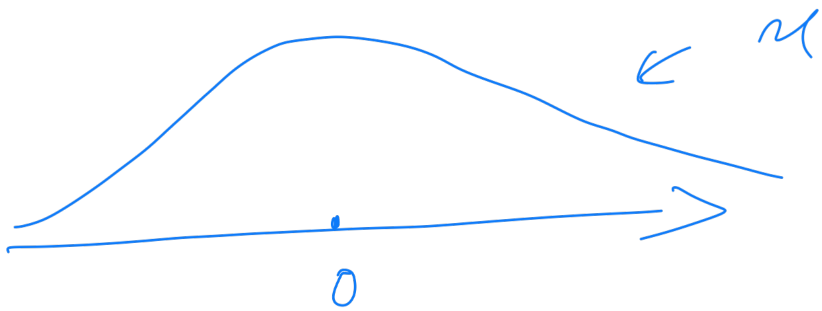
$\rho_{y|x}^2$ is the proportion of $Var(y)$ that can be explained by $E(y|x)$

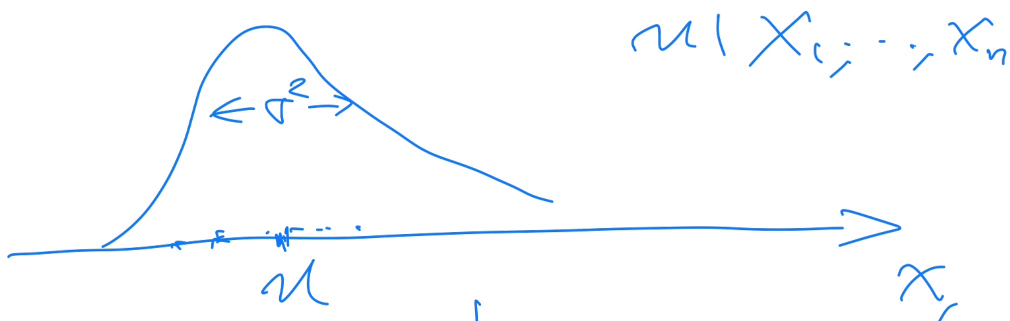
$\rho_{y|x} = \sqrt{\rho_{y|x}^2}$: multiple correlation coef.

Hints on Q11 of HW2:

$$\mu \mid X \sim N(\mu_1, \sigma_1^2)$$

$$\mu_1 = \frac{\frac{n}{\sigma^2}\bar{x} + \frac{1}{\sigma_0^2}\cdot 0}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} \quad \checkmark$$





$$\frac{1}{\sigma_1^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \quad \checkmark$$