① Finish Chapter 4 – GLS

② Distributional theory

Exam 1 : Chapters 1 – 4 (M) + Appendices
Chapter 2 (R)

+ HW 1 & 2

$S \to R \longrightarrow S'$

$\mathrm{cov}(e) = \sigma^2 I$ $\qquad$ $RS = S' \Leftrightarrow S = R^{-1}S'$

- Th

✓ $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable in model 2 if and only if $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable in model 3.

$\mathrm{Cov}(e) = \sigma^2 V$

✓ $\hat{\boldsymbol{\beta}}_{GLS}$ is a generalized least square estimate of $\boldsymbol{\beta}$ if and only if

$$\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}. = \hat{y}$$

For any estimable function there exists a unique generalized least square estimate, $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}_{GLS}$.

* Ordinary least square estimate (OLS)

$Py = X\hat{\beta} = \hat{y}$ & $P = X(X^TX)^{-1}X^T$

* GLS $\qquad$ $\phi$ $\qquad Ry = RX\beta + Re$

$= V^{-1}$ $\qquad y^* = X^*\beta + e^*$

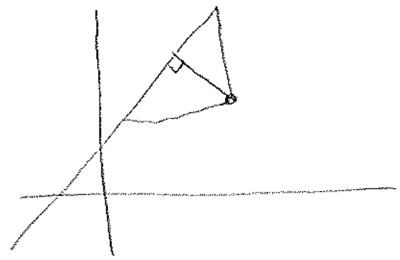
$X^T(R^TR)X$

$(RX)((RX)^T(RX))^{-}(RX)^T(Ry) = (RX)\hat{\beta}_{GLS} = (R\hat{y})$

$\Rightarrow RX(X^TV^{-1}X)^{-}X^TV^{-1}y = RX\hat{\beta}_{GLS}$ $\Big\}$ b/c $R$ is nonsingular

$A$

$\Rightarrow \underbrace{X(X^TV^{-1}X)^{-}X^TV^{-1}y}_{= \hat{\beta}_{GLS}} = X\hat{\beta}_{GLS} = \hat{Y}$



- Th (contd)

$\surd$ For an estimable function $\boldsymbol{\lambda}^T\boldsymbol{\beta}$, the generalized least squares estimate is the BLUE of $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ (Th 4.2).

$\Rightarrow$ $\underline{\underline{\lambda^T\hat{\beta}_{GLS}}}$

$P = X(X^TX)^- X^T$: the perpendicular projection matrix onto $\mathcal{C}(X)$

$\surd$ $\mathbf{A} = \mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}^T\mathbf{V}^{-1}$ is a projection matrix onto $C(\mathbf{X})$ (Not perpendicular)

- $P \neq A \Rightarrow$ A is not the perpendicular ~~ma~~ projection matrix onto $\mathcal{C}(X)$
- $Ay = X\hat{\beta}_{GLS}$

- Th (contd)

✓ $\text{Cov}(\mathbf{X}\hat{\boldsymbol{\beta}}_{GLS}) = \sigma^2\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T$. $= \text{Cov}(\hat{y})$

✓ If $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable, then the generalized least squares estimate has $\text{Var}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}_{GLS}) = \sigma^2\boldsymbol{\lambda}^T(\mathbf{X}^T\underline{\underline{\mathbf{V}}}^{-1}\mathbf{X})^{-}\boldsymbol{\lambda}$.

$\text{Var}(\lambda^T\hat{\beta}_{OLS}) = \sigma^2\lambda^T(X^TX)^{-}\lambda$

- Th (contd)

√ Estimation of $\sigma^2$

$$\hat{\sigma}^2_{GLS} = \frac{(\overbrace{\mathbf{Ry}}^{y^*} - \overbrace{\mathbf{RX}}^{X^*}\hat{\beta}_{GLS})^T(\overbrace{\mathbf{Ry}}^{y^*} - \overbrace{\mathbf{RX}}^{X^*}\hat{\beta}_{GLS})}{(n-r)},$$

where $r(\underbrace{\mathbf{RX}}_{X^*}) = r(\mathbf{X}) = r$.

$$\Downarrow$$

$$= \frac{(R(y - X\hat{\beta}_{GLS}))^T R(y - X\hat{\beta}_{GLS})}{(n-r)}$$

$$= \frac{(y - X\hat{\beta}_{GLS})^T V^{-1}(y - X\hat{\beta}_{GLS})}{(n-r)}$$

# AMS 256
# Monahan Chapter 5: Distributional Theory & Ronald Chapter 1: Introduction

Spring 2016

† The form of linear models is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

- $\mathbf{y}$: $n \times 1$ vector of observations (random)
- $\mathbf{X}$: $n \times p$ matrix of known constants (*design* matrix) with $r(\mathbf{X}) = r$
- $\boldsymbol{\beta}$: $p \times 1$ vector of unobservable parameters
- $\mathbf{e}$: $n \times 1$ vector of unobservable random errors

**So far we have assumed** the Gauss-Markov model assumption.

$E(\mathbf{e}) = \mathbf{0}$, $\text{Cov}(\mathbf{e}) = \sigma^2 I$( or $\sigma\mathbf{V}$) ($\sigma^2$ unknown, $\mathbf{V}$ known p.d matrix)

n-variate normal distribution

$\Rightarrow$ We will assume $\mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 I)$. So then let's take a look at some distributions related to the multivariate normal distribution.

♠ Warning for Sloppy Notation: We usually use $X$ to denote a random variable. Note that we use $x$, $\mathbf{x}$ and $\mathbf{X}$ to denote a random variable, a vector and a matrix, respectively.

† Def: A random variable $z$ has the standard normal distribution, denoted by $z \sim \mathrm{N}(0,1)$, whose density is

$$p(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).$$

$*$ Its moment generating function (m.g.f) is

$$m_z(t) \equiv \mathrm{E}(e^{tz}) = \int_{\mathbb{R}} e^{tz} p(z)\, dz = \exp(\frac{t^2}{2}).$$

$*$ Then we can construct a more general distribution from the standard normal distribution using the transformation, $x = \mu + \sigma z$.

$\sim N(0,1)$

† Def 5.1 A random variable $x$ has the normal distribution with mean $\mu$ and variance $\sigma^2$, denoted by $x \sim \mathrm{N}(\mu, \sigma^2)$ whose density is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

$E(X) = \mu$

$Var(X) = \sigma^2$

$**$ Its m.g.f is

$$m_x(t) = \mathrm{E}(e^{tx}) = \int_{\mathbb{R}} e^{tx} p(x)\, dx = \exp(t\mu + \frac{t^2\sigma^2}{2}).$$

- Recall!

If their densities exist, two random vectors are independent iff their joint density is equal to the product of their marginal densities.

p-dim

Let $\mathbf{z} = [z_1, \ldots, z_p]^T$ be a random vector with $z_1, \ldots, z_p$ independent identically distributed (i.i.d) $N(0, 1)$ random variables.

$\Rightarrow \quad E(\mathbf{z}) = \mathbf{0}, \quad \text{and} \quad \mathrm{Cov}(\mathbf{z}) = I.$

$\Rightarrow$ Let's find their **joint density distribution.**

† Def 5.2 Let $\mathbf{z}$ be a $p \times 1$ vector with each component $z_i$, $i = 1, \ldots, p$ independently distributed with $z_i \sim \mathrm{N}(0,1)$. Then $\mathbf{z}$ has the standard multivariate normal distribution, denoted by $\mathbf{z} \sim \mathrm{N}_p(\mathbf{0}, I_p)$, in $p$ dimensions. The joint density of the standard multivariate normal can be written then as

$$\boxed{p(\mathbf{z})} = \prod_{i=1}^{p} \overbrace{\frac{1}{(2\pi \underset{1}{\cancel{\sigma^2}})^{-1/2}} \exp\left(-\frac{z_i^2}{2}\right)}^{= p(z_i)} = \frac{1}{(2\pi \underset{1}{\cancel{\sigma^2}})^{-p/2}} \exp\left(-\sum_{i=1}^{p} \frac{z_i^2}{2}\right)$$

$$= \frac{1}{(2\pi \underset{1}{\cancel{\sigma^2}})^{-p/2}} \exp\left(-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right).$$

$**$ Its m.g.f is

$$m_{\mathbf{z}}(\mathbf{t}) = \mathrm{E}(e^{\mathbf{t}^T\mathbf{z}}) = \int_{\mathbb{R}^p} e^{\mathbf{t}^T\mathbf{z}} p(\mathbf{z})\, d\mathbf{z} = \exp(\frac{\mathbf{t}^T\mathbf{t}}{2}).$$

$$= \prod_{i=1}^{p} \exp\left(\frac{t_i^2}{2}\right)$$

$$\overset{q\times 1}{\mathbf{x}} = \overset{q\times 1}{\boldsymbol{\mu}} + \overset{q\times p}{\mathbf{A}}\overset{p\times 1}{\mathbf{z}}$$

♣ Let an $q-$dimensional vector $\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z}$, for some $p$, some $q \times p$ matrix $\mathbf{A}$, and some $q$ vector $\boldsymbol{\mu}$. What is the distribution of $\mathbf{x}$?

$Var(x) = Var(Az + \cancel{\mu}) = A\,Var(z)\,A^T = AA^T$ $\quad E(x) = E(Az + \mu) = \mu$

† Def: $\mathbf{x}$ has an $q-$dimensional multivariate normal distribution if $\mathbf{x}$ has the same distribution as $\mathbf{A}\mathbf{z} + \boldsymbol{\mu}$, i.e. $\mathbf{x} \boxed{\sim} \mathbf{A}\mathbf{z} + \boldsymbol{\mu}$, for some $p$, some $q \times p$ matrix $\mathbf{A}$, and some $q$ vector $\boldsymbol{\mu}$. We indicate the $\Sigma$ multivariate normal distribution of $\mathbf{x}$ by writing $\mathbf{x} \sim N_q(\boldsymbol{\mu}, \underbrace{\mathbf{A}\mathbf{A}^T}_{q\times q = V})$.

⇒ Let $\mathbf{V} = \mathbf{A}\mathbf{A}^T$. A multivariate normal distribution depends only on its mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$.

⇒ Is the distribution of $\mathbf{x}$ well defined?

$q \times r$ $\quad$ $\text{rank}(V) = r$ $\quad$ $A = \underline{Q_r}\,\Lambda_r\,Q_r^T$

$X = Az + \underline{\mu}$ $\quad$ $A: q \times r$ $\quad$ $q \times r$

$q \times q$ $\quad$ $q \times q$ $\quad$ $Bz + \mu$

- Let's start with $\mathbf{V}$ (nonnegative definite).

$q \times q$

- We may be able to write $\mathbf{V} = \mathbf{A}\mathbf{A}^T$ and $\mathbf{V} = \mathbf{B}\mathbf{B}^T$ with $\mathbf{A} \neq \mathbf{B}$. Then we don't know which one to take for $\mathbf{x}$.

- Further the length of $\mathbf{z}$ could change between $\mathbf{x} = \mathbf{A}\mathbf{z} + \boldsymbol{\mu}$ and $\mathbf{x} = \mathbf{B}\mathbf{z} + \boldsymbol{\mu}$ .

- Bottom line: It does *not* matter! Really??

We know that any two random vectors with the same moment generating function have the same distribution.

Let's take a look at the mgf of the multivariate normal distribution.

$$m_{\mathbf{x}}(\mathbf{t}) = \exp\left(\mathbf{t}^T\underline{\boldsymbol{\mu}} + \frac{\mathbf{t}^T\underline{\mathbf{V}}\mathbf{t}}{2}\right).$$

Aha!!! This has only $\boldsymbol{\mu}$ and $\mathbf{V}$!

$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ $V = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

$var(X_i) = \sigma_i^2$

$cov(X_1, X_2) = \rho\sigma_1\sigma_2$

$Cor(X_1, X_2) = \rho$

$f(X) = \frac{1}{(2\pi)^{-2/2} |V|^{1/2}} \exp\left(-\frac{1}{2}(X-\mu)^T V^{-1} (X-\mu)\right)$ $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbb{R}^2$

$\Rightarrow$ $\rho = \pm 1$

$\underline{\rho = 1} \Rightarrow V = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

$\sigma_2/\sigma_1$

- **Singular V**

$|V| = 0$

$\nexists f(X)$

$\surd$ Leading to the singular multivariate normal distribution.

$\surd$ The probability mass lies in a subspace and the pdf may not exit.

- **Nonsingular V** ($\Leftrightarrow$ **V** positive definite) (Cor 5.2)

$\underline{A} = Q\Lambda^{1/2}Q^T$

$\surd$ We call nonsingular **x**.

$\surd$ We can write $\mathbf{V} = \mathbf{A}\mathbf{A}^T$, with **A** nonsingular.

$\surd$ $\mathbf{x} = \mathbf{A}\mathbf{z} + \mu$ involves nonsingular transformation and has a density

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{q/n}|\mathbf{V}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right).$$

- Let's connect this to what we have

  √ $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$ and $\mathbf{e} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 I)$.

  $$\Rightarrow \quad \mathbf{y} \sim \mathrm{N}_n(\mathbf{X}\beta, \sigma^2 I).$$

  $y = \underset{A}{I} e + \underbrace{X\beta}_{\mu}$

  √ $\mathbf{y} \sim \mathrm{N}_n(\mathbf{X}\beta, \mathbf{V})$ for pd $\mathbf{V}$.

  $$\Rightarrow \quad \mathbf{V}^{-1/2}\mathbf{y} \sim \mathrm{N}_n(\underbrace{\mathbf{V}^{-1/2}}\mathbf{X}\beta, I).$$

  √ Note if $\mathbf{V} = \sigma^2 I$,

  $$\mathbf{V}^{-1/2} = \frac{1}{\sigma} I \quad \text{and} \quad \frac{1}{\sigma}\mathbf{y} \sim \mathrm{N}_n(\frac{1}{\sigma}\mathbf{X}\beta, I).$$

$\hat{Y} = \rho 4$

$y = Bx = b_1 x_1 + b_2 x_2 + \cdots + b_q x_q$

- $B$ as $\begin{bmatrix} b_1 & 0 & 0 & 0 \\ 0 & b_2 & 0 & \cdots \\ 0 & 0 & b_3 & \cdots \end{bmatrix}_{q \times p}$

$q \le p$

$\Rightarrow$ linear combination of normal random variable has a normal distribution

- Linear transformation of normal random variables has a normal distribution

- **Result 5.3** If $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$ and $\mathbf{y} = \mathbf{a} + \mathbf{Bx}$ where $\mathbf{a}$ is $q \times 1$ and $\mathbf{B}$ is $q \times p$, then $\mathbf{y} \sim N_q(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{BVB}^T)$.

$$m_y(t) = E(e^{t^T y}) = E\left(e^{t^T(a + Bx)}\right) = e^{t^T a} E\left(e^{\overbrace{t^T B x}^{= (B^T t)^T x}}\right)$$

$$= e^{t^T a} \cdot e^{(B^T t)^T \mu + \frac{(B^T t)^T V (B^T t)}{2}}$$

$$= e^{t^T(a + B\mu) + \frac{t^T B V B^T t}{2}} \Rightarrow \text{mgf for } N_q(a + B\mu, BVB^T)$$

- **Cor 5.1** If $\mathbf{x}$ is multivariate normal ($\mathbf{x} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$), then the joint distribution of any subset is multivariate normal.

Let $\underset{p\times 1}{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{matrix} p_1 \times 1 \\ p_2 \times 1 \end{matrix}$, $\quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \begin{matrix} p_1 \times 1 \\ p_2 \times 1 \end{matrix}$ $\quad V = \begin{array}{cc} p_1 \times p_1 & p_1 \times p_2 \\ \left[\begin{matrix} V_{11} \\ V_{21} \end{matrix}\right. & \left.\begin{matrix} V_{12} \\ V_{22} \end{matrix}\right] \\ p_2 \times p_1 & p_2 \times p_2 \end{array}$

$p_1 + p_2 = p$ $\qquad V_{21} = V_{12}^T$

① $X_1$ is marginally $N_{p_1}(\mu_1, V_{11})$ $\qquad a = \mathbf{0}$

② $X_2$ is marginally $N_{p_2}(\mu_2, V_{22})$ $\qquad B = [\underbrace{I_{p_1}}_{p_1 \times p_1} \ \underbrace{0}_{p_1 \times p_2}]_{p_1 \times p}$

$Y = a + BX$

$= \mathbf{0} + [I_{p_1} \ 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$= X_1$

③ Conditional distribution of $X_1$ given $X_2$ is $N_{p_1}(\mu_{1|2}, V_{1|2})$

$$\mu_{1|2} = \mu_1 + V_{12} V_{22}^{-1} (X_2 - \mu_2)$$

$$V_{1|2} = V_{11} - V_{12} V_{22}^{-1} V_{21}$$

- Result 5.4 If $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$ and $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T$, then $\text{Cov}(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{0}$ iff $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent.

$x = Az' + \mu = Bz + \mu$

$(\Leftarrow)$ $x_1$ & $x_2$ indep $\Rightarrow$ $\text{Cov}(x_1, x_2) = 0$

$(\Rightarrow)$ $\text{Cov}(x_1, x_2) = 0$ $\Rightarrow$ $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ $\quad V = A \cdot A^T = \begin{bmatrix} A_1 A_1^T & A_1 A_2^T \\ A_1^T A_2 & A_2 A_2 \end{bmatrix}$

$V = BB^T = AA^T$

$x = Bz + \mu = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} z + \mu$, $z$ iid

- Cor 5.3 Let $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$, and $\mathbf{y}_1 = \mathbf{a}_1 + \mathbf{B}_1\mathbf{x}$, $\mathbf{y}_2 = \mathbf{a}_2 + \mathbf{B}_2\mathbf{x}$. $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent iff $\mathbf{B}_1\mathbf{V}\mathbf{B}_2^T = \mathbf{0}$.

$\hat{y} = Py$

$\hat{e} = (I - P)y$

- Let's connect this to what we have $y \sim N_n(X\beta, \sigma^2 I)$

$$\mathbf{y} = \mathbf{P}\mathbf{y} + (I - \mathbf{P})\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}}$$

where $\mathbf{P}$ is the orthogonal projection operator onto $C(\mathbf{X})$.

- ✓ What is the distribution of $\hat{\mathbf{y}}$?

$$\hat{y} = Py \sim N_n(\underbrace{P\overbrace{X\beta}^{\in C(X)}}_{= X\beta}, \underbrace{\sigma^2 P I P^T}_{= \sigma^2 P})$$

$Var(Py) = P\,Var(y)\,P^T = P \cdot \sigma^2 I\, P^T = \sigma^2 P \cdot P^T = \sigma^2 P$

- ✓ What is the distribution of $\hat{\mathbf{e}}$?

$$\hat{e} = (I-P)y \sim N_n(\underbrace{(I-P)\overbrace{X\beta}^{\in C(X)}}_{=0}, \sigma^2 (I-P))$$

- ✓ What is their joint distribution?

$Cov(\hat{y}, \hat{e}) = 0$

$\hat{y} = \underset{a_1}{0} + \underset{B_1}{P}y$

$\hat{e} = \underset{a_2}{0} + \underset{B_2}{(I-P)}y$

$B_1 V B_2^T \overset{?}{=} 0$

$= P\,\sigma^2 I\,(I-P)$

$= P \cdot \sigma^2 (I-P)$

$= \sigma^2 (P - P^2)$

$= 0$

$$\begin{bmatrix} \hat{y} \\ \hat{e} \end{bmatrix} \sim N_{2n}\left( \begin{bmatrix} X\beta \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} P & 0 \\ 0 & I-P \end{bmatrix} \right)$$

- Let's connect this to what we have

$E(\hat{\beta}) = \beta$

$Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$

$$\mathbf{y} \sim \mathrm{N}_n(\underset{n \times p}{\mathbf{X}}\underset{p \times 1}{\boldsymbol{\beta}}, \sigma^2 I)$$

√ What is the distribution of $\hat{\boldsymbol{\beta}}$ for full rank $\mathbf{X}$?

$\hat{\beta} = (X^T X)^{-1} X^T y \quad \rightsquigarrow \quad \hat{\beta} \sim N_p(\beta, \sigma^2 (X^T X)^{-1})$

√ What is the distribution of $\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}}$ for $\boldsymbol{\lambda}^T \boldsymbol{\beta}$ estimable?

$\lambda^T \hat{\beta} = \lambda^T (X^T X)^- X^T y$

$E(\lambda^T \hat{\beta}) = \lambda^T \beta$

$Var(\lambda^T \hat{\beta}) = \sigma^2 \lambda^T (X^T X)^- \lambda$

$\Rightarrow \quad \lambda^T \hat{\beta} \sim N(\lambda^T \beta, \sigma^2 \lambda^T (X^T X)^- \lambda)$

- Recall!

$\surd$ Let $\mathbf{x}$ be an $p-$dimensional random vector and let $\mathbf{A}$ be an $p \times p$ symmetric matrix. A *quadratic form* is a random variable defined by $\mathbf{x}^T\mathbf{A}\mathbf{x}$ for some $\mathbf{x}$ and $\mathbf{A}$.

$\surd$ We have

$$\mathbf{y}^T\mathbf{y} = \hat{\mathbf{y}}^T\hat{\mathbf{y}} + \hat{\mathbf{e}}^T\hat{\mathbf{e}} = \mathbf{y}^T\mathbf{P}\mathbf{y} + \mathbf{y}^T(I - \mathbf{P})\mathbf{y}.$$

$**$ Observe that $\mathbf{y}^T\mathbf{y}$, $\mathbf{y}^T\mathbf{P}\mathbf{y}$ and $\mathbf{y}^T(I-\mathbf{P})\mathbf{y}$ are quadratic forms. What are their distributions?

$\surd$ Now let's take a look at the distributions of quadratic forms in multivariate normal vectors ($\mathbf{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \mathbf{V})$).

- Recall! (contd) $z_i, i=1,\ldots,p$ iid $N(0,1)$

√ Def 5.5: Let $\mathbf{z} \sim N_p(\mathbf{0}, I_p)$, then $u = \underbrace{\mathbf{z}^T\mathbf{z}}_{z^T I z} = \sum_{i=1}^{p} z_i^2$ has the chi-square distribution with $p$ degrees of freedom, denoted by $u \sim \chi^2(p)$.

† Def: Let $z_1, \ldots, z_p$ be independent with $z_i \sim \mathrm{N}(\mu_i, 1)$. Then

$$u = \sum_{i=1}^{p} z_i^2$$

has a *noncentral chi-square distribution* with $\boxed{p}$ degrees of freedom and noncentrality parameter $\phi = \sum_{i=1}^{p} \mu_i^2/2$. We write $u \sim \chi^2(p, \phi)$.

df

$E(u) = p + 2\phi$

$Var(u) = 2p + 8\phi$

√ Another definition for the noncentral chi-square distribution is in Def 5.6.

$J \sim Poi(\phi)$

$U \mid J \sim \chi^2_{\cancel{p+2}}(p + 2j)$

$\Rightarrow$ the marginal distribution of $u$ is $\chi^2(p, \phi)$

$p(u) = \sum_{j=0}^{\infty} p(u, j) = \sum_{j=0}^{\infty} p(j)\, p(u \mid j)$

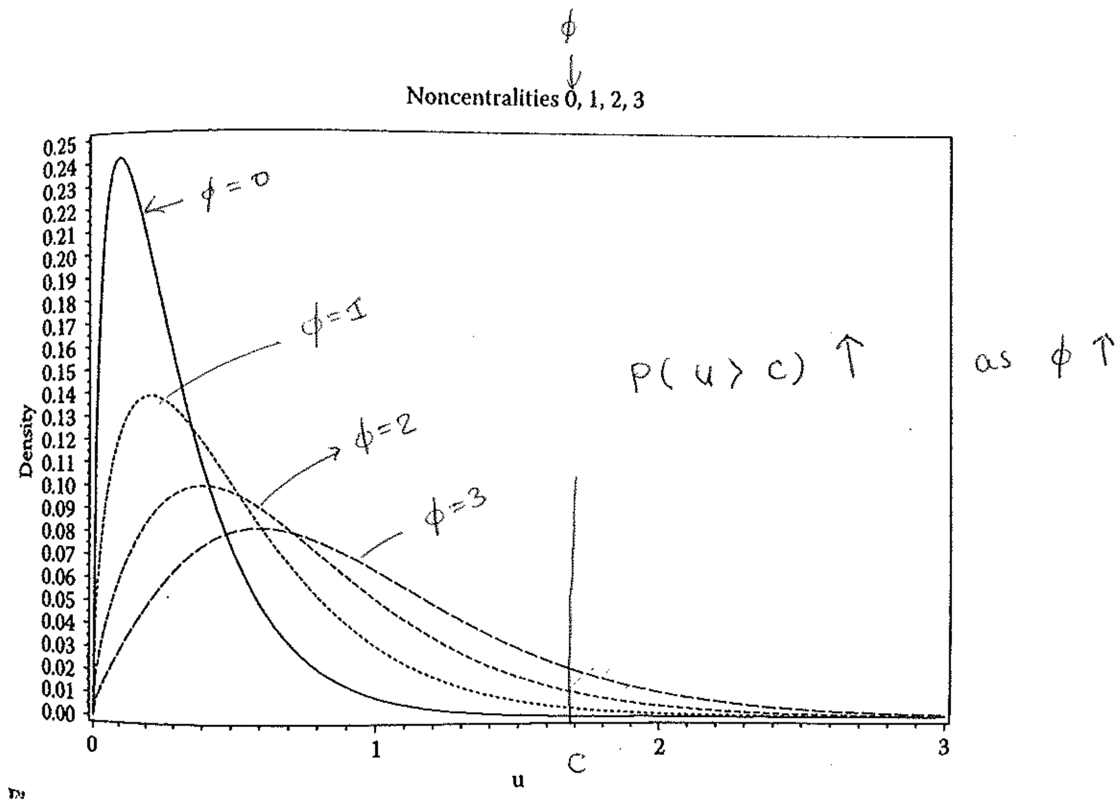
♣ Some immediate results:

$\surd$ $x \sim \chi^2(r, \phi)$ and $y \sim \chi^2(s, \delta)$ with $x$ and $y$ independent
$\Rightarrow x + y \sim \chi^2(r + s, \phi + \delta)$.

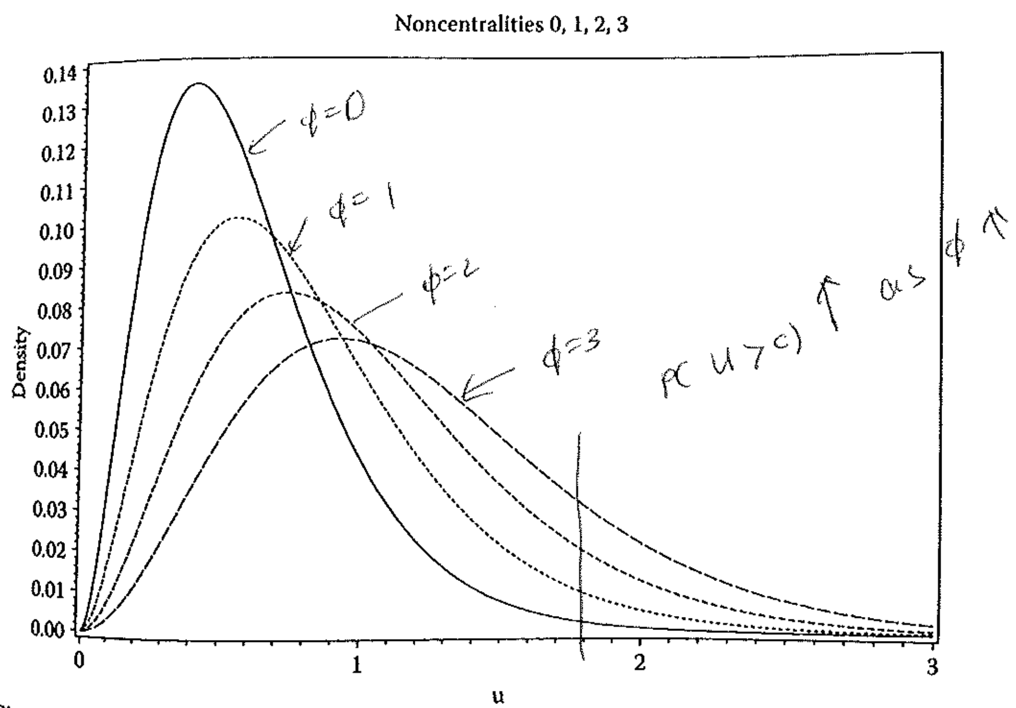
$\surd$ A central $\chi^2$ distribution is a distribution with a noncentrality parameter of zero, i.e. $\chi^2(r, 0)$.

$\surd$ $\mathbf{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, I) \quad \Rightarrow \quad \mathbf{x}^T\mathbf{x} \sim \chi^2(p, \boldsymbol{\mu}^T\boldsymbol{\mu}/2)$ (Result 5.9)

• Figure 5.1: Noncentral $\chi^2$ densities with $\boxed{p = 3 \text{ degrees of freedom}}$. Increasing $\phi$ flattens the density and shifts the distribution to the right



• Figure 5.2: Noncentral $\chi^2$ densities with $p = 6$ degrees of freedom. Increasing $\phi$ flattens the density and shifts the distribution to the right



- Recall: We say that $x$ is stochastically larger than $y$ if $F_x(t) \leq F_y(t)$ for all $t$.

- Result 5.11 Let $u \sim \chi^2(p, \phi)$, then $P(u > c)$ is strictly increasing in $\phi$ for fixed $p$ and $c > 0$.

$u_1 \sim \chi^2(p, \phi_1)$ $\quad \phi_1 < \phi_2$

$u_2 \sim \chi^2(p, \phi_2)$

$\Rightarrow$ $u_1$ is stochastically larger than $u_2$

$\Rightarrow$ related to the power of a test