ex) $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2 + \varepsilon_i$

has $Y_i = \underline{w}_i^T \beta + \varepsilon_i$ where

$\underline{w}_i = (1, x_{i1}, x_{i1}^2)^T$. so this

is a linear model.

ex) $Y_i = \beta_0 + \beta_1 e^{-\beta_2 x_i} + \varepsilon_i$ can't

be written as $Y_i = \underline{w}_i^T \beta + \varepsilon_i$

and is a nonlinear regression model.

3) p36 Let $\underset{\sim}{\theta} = X \underset{\sim}{\eta} \in C(X)$ where

treat as known

$Y_i = \underset{\sim}{x}_i^T \underset{\sim}{\eta} + e_i(\underset{\sim}{\eta})$ where $e_i(\underset{\sim}{\eta})$ depends

on $\underset{\sim}{\eta}$. Note $\varepsilon_i = e_i(\beta)$. Least Squares

minimizes $\sum_{i=1}^{n} e_i^2(\underset{\sim}{\eta})$ wrt $\underset{\sim}{\eta}$, and

the **least squares estimator** $\hat{\beta}$ is the

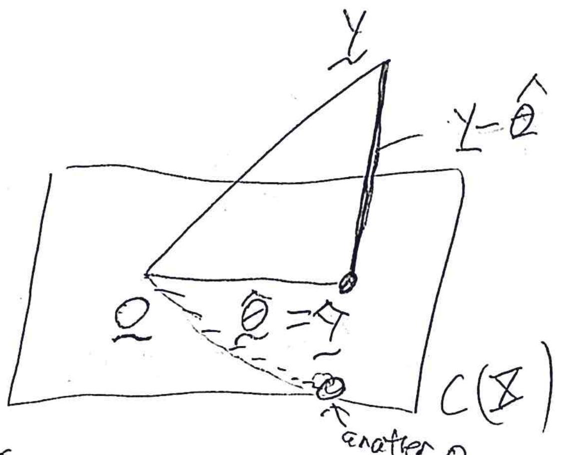
value of $\underset{\sim}{\eta} \in \mathbb{R}^p$ that achieves the min.

$$\hat{\beta} = \underset{\eta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} e_i^2(\underset{\sim}{\eta}) = \underset{\eta \in \mathbb{R}^p}{\operatorname{argmin}} \| \overbrace{\underset{\sim}{Y} - X \underset{\sim}{\eta}}^{\underset{\sim}{e}(\underset{\sim}{\eta})} \|^2$$

If $\underset{\sim}{\theta} = X\underset{\sim}{\beta} = $ then

$\hat{\underset{\sim}{\theta}} = \underset{\theta \in C(X)}{\text{argmin}} \|\underline{Y} - \underline{\theta}\|^2$. Let ith residual $e_i = e_i(\hat{\beta})$.

4) p 36 Geometrically $\hat{\underline{\theta}} = \hat{\underset{\sim}{Y}}$ is the minimum value if $\underline{Y} - \hat{\underline{\theta}} \perp C(X)$



see cover of ~~Seber for~~ better plot

5) $\hat{\underline{\theta}} = X\hat{\underline{\beta}} = P_X \underset{\sim}{Y} = \hat{\underline{Y}}$ where $H = P_X = X(X^TX)^{-1}X^T$ is the projection matrix onto $C(X)$.

6) p36-7 Algebraic proof. Let $\hat{\underline{Y}} = \hat{\underline{\theta}} = P_X \underset{\sim}{Y} \in C(X)$

$\underset{\sim}{e} = (I - P_X)\underline{Y} \in [C(X)]^\perp$ and $\underline{\theta} \in C(X)$.

(Then $\underline{Y} = \hat{\underset{\sim}{Y}} + \underset{\sim}{e} = P_X \underset{\sim}{Y} + (I - P_X)\underset{\sim}{Y}$.)

Now $\underline{Y} - \underline{\theta} = \underbrace{(\underline{Y} - \hat{\underline{\theta}}) + (\hat{\underline{\theta}} - \underline{\theta})}_{\underset{\sim}{0}}$

and $(\underline{Y}-\underline{\hat{\theta}})'(\underline{\hat{\theta}}-\underline{\theta}) = (\underline{Y}-P_x\underline{Y})'(P_x\underline{Y}-P_x\underline{\theta})$

$P_x\underline{\theta}=\underline{\theta}$ since $\underline{\theta}\in C(\boldsymbol{X})$

$$= \underline{Y}^T \underbrace{(I-P_x)P_x}_{\substack{0 \\ \text{matrix}}} (\underline{Y}-\underline{\theta}) = \underset{\text{scalar}}{0}$$

$$\text{so } \|\underset{\sim}{Y}-\underset{\sim}{\theta}\|^2 = (\underline{Y}-\underline{\hat{\theta}}+\underline{\hat{\theta}}-\underline{\theta})^T(\underline{Y}-\underline{\hat{\theta}}+\underline{\hat{\theta}}-\underline{\theta})$$

$$= \|\underline{Y}-\underline{\hat{\theta}}\|^2 + \|\underline{\hat{\theta}}-\underline{\theta}\|^2 + \underbrace{2(\underline{Y}-\underline{\hat{\theta}})^T(\underline{\hat{\theta}}-\underline{\theta})}_{0}$$

$$\geq \|\underline{Y}-\underline{\hat{\theta}}\|^2 \quad \text{with equality}$$

iff $\|\underline{\hat{\theta}}-\underline{\theta}\|^2=0$ or $\underline{\hat{\theta}}=\underline{\theta}$.

7) p37 $\underline{Y}-\underline{\hat{\theta}} \perp C(\boldsymbol{X}) \Rightarrow \underline{Y}-\underline{\hat{\theta}} \in [C(\boldsymbol{X})]^{\perp}$

$= N(\boldsymbol{X}^T)$. So $\boldsymbol{X}^T(\underline{Y}-\underline{\hat{\theta}}) = \underset{\sim}{0}$ or $\boldsymbol{X}^T\boldsymbol{X}\underline{\beta} \overset{set}{=} \boldsymbol{X}^T\underline{Y}$

So $\boldsymbol{X}^T\underset{\sim}{\hat{\theta}} = \underbrace{\boldsymbol{X}^T\boldsymbol{X}\underset{\sim}{\hat{\beta}} = \boldsymbol{X}^T\underline{Y}}_{\text{normal equations}}$.

and the <u>least squares estimator</u> of $\underset{\sim}{\beta}$

is $\hat{\beta} = (X^T X)^{-1} X^T \underset{\sim}{Y}$.

← Be able to do similar calculations on Q 3.

18) Suppose $\underline{Y} = X \underset{\sim}{\beta} + \underset{\sim}{\epsilon} = \hat{\underline{Y}} + \underline{e}$

where X is full rank, $E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$, $\text{cov}(\underset{\sim}{\epsilon}) = \sigma^2 I$

Let $P = P_X$ = projection matrix on $C(X)$,

Then $\hat{\underset{\sim}{Y}} = P\underset{\sim}{Y} = \hat{\underset{\sim}{\theta}}$, $\underline{e} = \underset{\sim}{Y} - \hat{\underset{\sim}{Y}} = (I-P)\underline{Y}$, and

$PX = X$ so $X'P = X'$,

i) $X'\underset{\sim}{e} = X'(I-P)\underset{\sim}{Y} = 0\underset{\sim}{Y} = \underset{\sim}{0}$

Predictors and residuals are orthogonal

ii) $E(\underline{Y}) = X\underset{\sim}{\beta}$

iii) $\text{cov}(\underline{Y}) = \text{cov}(\underset{\sim}{\epsilon}) = \sigma^2 I$

iv) $\text{cov}(\underset{\sim}{e}, \hat{\underline{Y}}) = 0$ since

← fitted values are residuals are uncorrelated

$\text{cov}(\underline{e}, \hat{\underline{Y}}) = E[(\underline{e} - E\underline{e})(\hat{\underline{Y}} - E\hat{\underline{Y}})']$

$= E[((I-P)\underline{Y} - (I-P)E\underline{Y})(P\underline{Y} - PEY)']$

$= E[(I-P)(\underline{Y} - E\underline{Y})(P(Y - EY))']$

$= E[(I-P)(\underline{Y} - E\underline{Y})(\underline{Y} - E\underline{Y})' P]$

$= (I-P)\sigma^2 I P = \sigma^2 (I-P)P = \underset{n \times n}{0}$

9) $\underset{\sim}{Y} = X\beta + \underset{\sim}{\varepsilon}$ X full rank, $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$

a) $E(\hat{\beta}) = E(X'X)^{-1}X'\underset{\sim}{Y} =$

$$(X'X)^{-1}X'E\underset{\sim}{Y} = (X'X)^{-1}X'X\beta = \beta,$$

(unbiased)

b) If $\operatorname{cov}(\varepsilon) = \operatorname{cov}(Y) = \sigma^2 I$,

then $\operatorname{cov}(\hat{\beta}) = \operatorname{cov}\left[\underbrace{(X'X)^{-1}X'}_{A}\underset{\sim}{Y}\right]$

$$= (X'X)^{-1}X^{T}\sigma^2 I\, X(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1}.$$

10) The linear estimator $\underset{\sim}{a}'\underset{\sim}{Y}$ of $\underset{\sim}{c}'\underset{\sim}{\theta}$ is the best linear unbiased estimator

<u>BLUE</u> of $\underset{\sim}{c}'\underset{\sim}{\theta}$ if $E\,\underline{a}'Y = \underline{c}'\underline{\theta}$

and if for any other unbiased linear estimator $\underset{\sim}{b}'\underset{\sim}{Y}$ of $\underset{\sim}{c}'\underset{\sim}{\theta}$,

$$\operatorname{var}(\underline{a}'Y) \le \operatorname{var}(\underline{b}'Y).$$

11) p42 Let $\hat{\theta} = X\hat{\beta}$ be the least squares (2a5)

Think omitted the uniqueness

estimator of $X\beta$. (X full rank)

a) $c'\hat{\theta}$ is the unique BLUE of $c'\theta$

b) $a'\hat{\beta}$ is the BLUE of $a'\beta$ for every vector a. roughly 11) is the Gauss Markov theorem

proof b) $\theta = X\beta$ so $\beta = (X'X)^{-1}X'\theta \; (= (X'X)^{-1}X'\underbrace{X\beta}_{\theta})$

and $\hat{\beta} = (X'X)^{-1}X'\underbrace{X\hat{\beta}}_{\hat{\theta}} = (X'X)^{-1}X'\hat{\theta}$

Let $c' = a'(X'X)^{-1}X'$. Then $c'\hat{\theta} = a'\hat{\beta}$

is the BLUE of $c'\theta = a'\beta$ by a).

§3.3 12) p44 Let $Y = X\beta + \varepsilon$, X full rank, ε_i iid with mean 0 and variance σ^2.

a) $s^2 = \dfrac{(Y-\hat{Y})'(Y-\hat{Y})}{n-p} = \dfrac{RSS}{n-p} = \dfrac{\sum e_i^2}{n-p}$ = MSE is

unbiased estimator of σ^2 where RSS =

residual sum of squares = $\sum_{i=1}^{n} e_i^2 = \ldots$

b) Let $h_i = H_{ii}$ where $\underset{n \times n}{H} = P_X$ be the ith leverage. If $\max h_i \to 0$ as $n \to \infty$ and $E\varepsilon_i^4 = \gamma < \infty$ then

MSE is a $\sqrt{n}$ consistent estimator of σ^2. $\sqrt{n}\,(MSE - \sigma^2) = O_p(1)$

which implies that $n^{\delta}(MSE - \sigma^2) \xrightarrow{P} 0$ if $0 < \delta \le \frac{1}{2}$.

Su and Cook (p692, 2012) Conjecture: result holds if $E\varepsilon_i^2 = \sigma$

12] Write $\hat{\underline{\theta}} \xrightarrow{P} \underline{\theta}$ if $\hat{\underline{\theta}}$ is a consistent estimator of $\underline{\theta}$.

§3.4 13] Let $Y_i = \mu + \varepsilon_i$ so $\underset{\sim}{Y} = \underset{\sim}{1}\,\mu + \underset{\sim}{\varepsilon}$. $\underset{n \times 1}{\beta}$ ← scalar

Then the least squares estimator

$$\hat{\mu} = (\underset{\sim}{1}'\underset{\sim}{1})^{-1}\underset{\sim}{1}'\underset{\sim}{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y}.$$

If the Y_i are iid $N(\mu, \sigma^2)$ so $\underset{\sim}{Y} \sim N_n(\mu\underset{\sim}{1}, \sigma^2 I)$,

then $\overline{Y} \sim N(\mu, \frac{\sigma^2}{n})$, so $\overline{Y} - \mu \sim N(0, \frac{\sigma^2}{n})$ and

$\sqrt{n}\,(\overline{Y} - \mu) \sim N(0, \sigma^2)$.

215

If the Y_i are iid with mean μ and variance σ^2 so $E(\underset{\sim}{Y}) = \mu \underset{\sim}{1}$ and $Cov(\underset{\sim}{Y}) = \sigma^2 I_n$,

then $\sqrt{n}(\bar{Y} - \mu) \xrightarrow{D} N(0, \sigma^2)$ by the central limit theorem (CLT).

use normal approx for prob and percentiles.

Will get similar results for $\hat{\underset{\sim}{\beta}}$.

14) p47 Suppose $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$, X full rank p, $\underset{\sim}{\epsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$

$\underset{\sim}{Y} \sim N_n(X\underset{\sim}{\beta}, \sigma^2 I_n)$.

Don't forget c) and d).

a) $\hat{\underset{\sim}{\beta}} \sim N_p(\underset{\sim}{\beta}, \sigma^2 (X'X)^{-1})$

b) $$\frac{(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})' X'X (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})}{\sigma^2} \sim \chi^2_p.$$

c) $\hat{\underset{\sim}{\beta}} \perp\!\!\!\perp MSE$

d) $\frac{RSS}{\sigma^2} = \frac{(n-p)MSE}{\sigma^2} \sim \chi^2_{n-p}$.

proof) a) $\hat{\underset{\sim}{\beta}} = \underbrace{(X'X)^{-1}X'}_{\text{constant}} \underset{\sim}{Y} \sim$

$N_p[(X'X)^{-1}X'X\underset{\sim}{\beta}, (X'X)^{-1}X' \sigma^2 I_n X (X'X)^{-1}]$

$\sim N_p[\underset{\sim}{\beta}, \sigma^2 (X'X)^{-1}]$.

b) $(\hat{\beta}-\beta)' \dfrac{X'X(\hat{\beta}-\beta)}{\sigma^2} = (\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})^T [Cov(\hat{\beta})]^{-1} (\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})$

= pop Mahalanobis distance of $\hat{\underset{\sim}{\beta}} \sim \chi^2_p$

by Th 2.9.

c) $Cov(\hat{\beta}, Y - X\hat{\beta}) = Cov(\hat{\beta}, e) = 0$

so $\hat{\beta} \perp\!\!\!\perp Y - X\hat{\beta}$ and $\hat{\beta} \perp\!\!\!\perp \|\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}}\|^2 = RSS$

so $\hat{\underset{\sim}{\beta}} \perp\!\!\!\perp$ MSE.

d) $RSS = \hat{e}^T \hat{e} \overset{(I-P)^2 = I-P}{=} Y^T(I-P)Y$

$= (\underset{\sim}{Y} - X\beta)^T (I-P)(\underset{\sim}{Y} - X\beta)$

since $X^T P = X^T$ and $PX = X$

$= \underset{\sim}{\varepsilon}^T (I-P)\underset{\sim}{\varepsilon}$. Now $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$

so $\frac{1}{\sigma}\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, I_n)$

and $\dfrac{\underset{\sim}{\varepsilon}^T (I-P)\underset{\sim}{\varepsilon}}{\sigma^2} \sim \chi^2_{n-p}$ by Th 2.7

since $I-P$ is idempotent of rank $n-p$.

15) Let $Y_i = \underline{x}_i^T \underline{\beta} + \epsilon_i$ where the (22.5)

ϵ_i are ind with $E\epsilon_i = 0$ and $V\epsilon_i = \sigma^2$.

Assume $\max h_i \xrightarrow{P} 0$ as $n \to \infty$. Then

a) $\hat{Y}_i = \underline{x}_i^T \hat{\underline{\beta}} \xrightarrow{P} \underline{x}_i^T \underline{\beta}$ for $i = 1, \ldots, n$ as $n \to \infty$.

b) All of the least squares estimators $\underline{a}^T \hat{\underline{\beta}}$ are asymptotically normal where $\underline{a}$ is any constant $p \times 1$ vector.

---

c) LS CLT Suppose the ϵ_i are iid, $\frac{X^T X}{n} \to W^{-1}$.

Then $\sqrt{n}(\hat{\underline{\beta}} - \underline{\beta}) \xrightarrow{D} N_p(\underline{0}, \sigma^2 W)$

Also, $(X^T X)^{\frac{1}{2}}(\hat{\underline{\beta}} - \underline{\beta}) \xrightarrow{D} N_p(\underline{0}, \sigma^2 I_p)$.

Sen and Singer p280 DasGupta p486 (Freedman 1981)

16) If $\underline{Z}_n \xrightarrow{D} N_k(\underline{\mu}, \Sigma)$ then

$\underset{m \times k}{A} \underline{Z}_n + \underline{b} \xrightarrow{D} N_m(A\underline{\mu} + \underline{b}, A \Sigma A^T)$,

provided RHS does not depend on n.

17) $\underline{Z}_n \xrightarrow{D} N_k(\underline{\mu}, \Sigma)$

$\underline{Z}_n \sim N_k(\underline{\mu}, \Sigma)$ often behave similarly. See 16). But RHS can't depend on n for $\xrightarrow{D}$.

18) * $\underline{Z}_n \xrightarrow{D} \underline{Z}$, $\underline{Z}$ is the limiting distribution

LM 25

19) Let $\hat{\Sigma}_n \xrightarrow{P} \Sigma$ where $\hat{\Sigma}_n = \hat{\Sigma}_n^T > 0$ (is a symmetric positive definite covariance matrix estimator.

If $\underset{\sim}{W}_n \xrightarrow{D} N_K(\underset{\sim}{\mu}, \Sigma)$, then

$$\underset{\sim}{W}_n - \underset{\sim}{\mu} \xrightarrow{D} N_K(\underset{\sim}{0}, \Sigma) \text{ and}$$

$$\underbrace{\hat{\Sigma}_n^{-1/2} (\underset{\sim}{W}_n - \underset{\sim}{\mu})}_{\underset{\sim}{Z}_n} \xrightarrow{D} N_K(\underset{\sim}{0}, I_K)$$

$\underset{\sim}{Z}_n$ like a Z score $\frac{Y-\mu}{\sigma}$

So $\underset{\sim}{Z}_n^T \underset{\sim}{Z}_n = (\underset{\sim}{W}_n - \underset{\sim}{\mu})^T \hat{\Sigma}_n^{-1} (\underset{\sim}{W}_n - \underset{\sim}{\mu}) \xrightarrow{D} \chi^2_K.$

$\Sigma^{-1} (W_n - \mu) \xrightarrow{D} N_K(0, \ldots)$

20) From 15 If $\frac{(X'X)}{n} \to W^{-1}$ and the ϵ_i are iid,

$\sqrt{n}(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \xrightarrow{D} N_p(\underset{\sim}{0}, \underbrace{\sigma^2 W}_{\Sigma})$. Here $X = \underset{n \times p}{X_n}$. (suppressed)

So $\frac{X'X}{n}$ is a consistent estimator of W^{-1}

So $\hat{\sigma}^2 n(X'X)^{-1} = \hat{\Sigma}_n$.

and $n(X'X)^{-1}$ is a consistent estimator of W

So $\underbrace{\frac{1}{\sigma} \frac{(X'X)^{\frac{1}{2}}}{\sqrt{n}}}_{\hat{\Sigma}_n^{-\frac{1}{2}} \text{ if } \sigma \text{ is known}} \sqrt{n}(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \xrightarrow{D} N_p(\underset{\sim}{0}, I_p)$

$(\frac{1}{\sigma} \hat{W}^{-\frac{1}{2}} =)$

or $(X'X)^{\frac{1}{2}}(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \xrightarrow{D} N_p(\underset{\sim}{0}, \sigma^2 I_p)$

Hence $(\underset{\sim}{X}_n'\underset{\sim}{X}_n)^{-\frac{1}{2}} (\underset{\sim}{X}_n'\underset{\sim}{X}_n)^{\frac{1}{2}}(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})$

$$= \hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta} \; \underset{\uparrow}{\approx} \; N_p\left(\underset{\sim}{0},\ \sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\right)$$

not $\xrightarrow{D}$ because $\underset{\sim}{X}_n'\underset{\sim}{X}_n$ depends on n

or $\hat{\underset{\sim}{\beta}} \approx N_p\left[\underset{\sim}{\beta},\ MSE\,(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\right]$

✓ Sometimes write $\hat{\underset{\sim}{\beta}} \sim \underbrace{AN}_p\left[\underset{\sim}{\beta},\ MSE\,(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\right]$

asymptotically normal or
approximately normal

where the normal approximation on the RHS now can depend on n.

$$\text{So } \underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n) \Rightarrow \hat{\underset{\sim}{\beta}} \sim N_p\left(\underset{\sim}{\beta},\ \sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\right)$$

and ε_i iid with $E(\underset{\sim}{\varepsilon})=0$ $\operatorname{cov}(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$

$\max h_i \to 0$, and $\frac{\underset{\sim}{X}^T\underset{\sim}{X}}{n} \to W^{-1}$ $\Rightarrow \hat{\underset{\sim}{\beta}} \sim AN_p\left[\underset{\sim}{\beta},\ \sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\right]$

$\overline{Y} \sim N(\mu, \frac{\sigma^2}{n})$ or $\overline{Y} \overset{CLT}{\sim} AN(\mu, \frac{\sigma^2}{n}) \sim AN(\mu, \frac{S^2}{n})$ where S^2 = samp

21} since $\underbrace{\frac{1}{\sigma}(\underline{X}'\underline{X})^{\frac{1}{2}}(\hat{\underline{\beta}}-\underline{\beta})}_{\underline{z}_n} \xrightarrow{D} N_p(\underline{0}, I_p)$.

$$\underbrace{\frac{1}{\sigma^2}(\hat{\underline{\beta}}-\underline{\beta})'(\underline{X}'\underline{X})(\hat{\underline{\beta}}-\underline{\beta})}_{\underline{z}_n'\underline{z}_n} \xrightarrow{D} \chi^2_p.$$

Note $\Sigma^{-\frac{1}{2}}(\underline{w}_n-\underline{\mu}) \xrightarrow{D} N_m(\underline{0}, I_m)$

$$LHS = (\Sigma^{-\frac{1}{2}} - \hat{\Sigma}_n^{-\frac{1}{2}} + \hat{\Sigma}_n^{-\frac{1}{2}})(\underline{w}_n-\underline{\mu})$$

$$= \underbrace{(\Sigma^{-\frac{1}{2}} - \hat{\Sigma}_n^{-\frac{1}{2}})(\underline{w}_n-\underline{\mu})}_{o_p(1)O_p(1) \to 0} + \hat{\Sigma}_n^{-\frac{1}{2}}(\underline{w}_n-\underline{\mu})$$

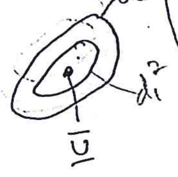
so $\hat{\Sigma}_n^{-\frac{1}{2}}(\underline{w}_n-\underline{\mu}) \xrightarrow{D} N_m(\underline{0}, I_m)$.

22} ✓ p40, 464: rank(P_x) = tr(P_x)

CEW p161-2

$MD_i^2 = d^2$

hyperellipsoids of $\underline{u}_i$



If $\underset{n\times p}{\underline{X}}$ is full rank, $\text{rank}(\underset{n\times n}{P_x}) = p$

So $P_x \neq 0$ but is singular unless $n = p$

in which case $P_x = I_n$.

23} Let $\underline{x} = (1 \;\; \underline{u}^T)^T$, $\bar{\underline{u}} = \frac{1}{n}\sum_{i=1}^{n} \underline{u}_i$, $C_u = \overbrace{\text{cov}(\underline{u})}^{\text{sample covariance matrix}} = \frac{1}{n-1}\sum(\underline{u}_i-\bar{\underline{u}})(\underline{u}_i-\bar{\underline{u}})$

$MD_i^2 = (\underline{u}_i - \bar{\underline{u}})' C_u^{-1} (\underline{u}_i - \bar{\underline{u}})$. Then $h_i = \frac{1}{n-1} MD_i^2 + \frac{1}{n}$

§3.5 節, Maximum likelihood 2) $\underset{\sim}{Y} \sim N_n(X\beta, \sigma^2 I)$

pdf

So $L(\beta, \sigma^2) = f(y \mid \beta, \sigma^2) =$

$$\frac{1}{(2\pi)^{n/2}} \frac{1}{\underbrace{\sqrt{\det(\sigma^2 I_n)}}_{\sigma^n}} \exp\left[-\frac{1}{2}(y - X\beta)^T (\sigma^2 I)^{-1} (y - X\beta)\right]$$

$\det(\mathrm{diag}(a_1, \dots, a_n)) = \prod_{i=1}^{n} a_i$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \|y - X\beta\|^2\right]$$

Fix σ^2, then maximizing $L(\beta, \sigma^2)$ is equivalent to maximizing $\exp\left(-\frac{1}{2\sigma^2}\|y - X\beta\|^2\right)$, which is equivalent to minimizing $\|y - X\beta\|^2$. But the least squares estimator minimizes $\|y - X\beta\|^2$

So $\hat{\beta}$ is the MLE of β.

Let $Q = \|y - X\hat{\beta}\|^2$. Can find $\hat{\sigma}^2$ by maximizing the profile likelihood

$$L_p(\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} Q\right);$$