WEEK 3 - ESTIMATORS

Suppose we have a r.vec $\underset{\sim}{Y}$ w/ dist'n $F(\underset{\sim}{y}; \underset{\sim}{\theta})$, where θ is the vector of parameters that describe the CDF.

Our goal is to recover θ based on Y.
Let $\hat{\theta} = g(Y)$ be an estimator of θ.

Any $g(Y)$ can be an estimator, but some are better than others.

Ex. $X \sim U(0, \theta)$. You might conclude $\hat{\theta} = 2X$ or $\hat{\theta} = \frac{3}{2}X$.
or even $\hat{\theta} = \frac{2}{3}X$

So what's a good criterion for an estimator? Several poss.:

Define a loss function $L(\hat{\theta}, \theta)$. For any given $\hat{\theta} = g(Y)$, we can calculate $E[L(\hat{\theta}, \theta)] \equiv$ risk of an estimator
(what are we integrating over?)

Might want to min risk.
Problem: often risk depends on θ so "best" estimator will depend on θ. How to deal?

"defensive estimating" $\quad \min\limits_{\hat{\theta}} \max\limits_{\theta \in \Theta} R(\hat{\theta}, \theta)$

min-max over a ~~possible~~ relevant range

Integrate over possible θ using a wt. function $w(\theta)$

$$r(\hat{\theta}) = \int R(\hat{\theta}, \theta)\, w(\theta)\, d\theta$$ wt. avg risk or Bayes risk

and then you'd $\min_{\hat{\theta}} r(\hat{\theta})$

By far the most common: quadratic loss $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$
which gives risk $R(\hat{\theta}, \theta) = E[(\hat{\theta} - \theta)^2]$ mean-squared error

$$E[(\hat{\theta} - \theta)^2] = E[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2]$$
$$= \underbrace{E[(\hat{\theta} - E(\hat{\theta}))^2]}_{Var(\hat{\theta})} + \underbrace{E[(E(\hat{\theta}) - \theta)^2]}_{Bias(\hat{\theta})^2} + 2E[(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)]$$

Let $X = (\hat{\theta} - \theta)^2$. For any r.v. X,

$$var(X) = E(X^2) - (E(X))^2$$
$$var(\hat{\theta}) = m.s.e. - [E(\hat{\theta}) - \theta]^2$$
$$var(\hat{\theta}) = mse - bias^2 \qquad bias = E(\hat{\theta}) - \theta$$
$$mse = var(\hat{\theta}) + bias^2$$

An estimator is unbiased if $E(\hat{\theta}) = \theta$

Unbiased isn't always best: Ex. Sample variance for iid normal sampl

Common estimator: $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$

can show this is unbiased ~~(should convince yourself)~~ (see Greene B11.4)

So $MSE = var(s^2) = \frac{2\sigma^4}{n-1}$

Less common: $\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$ or $\hat{\sigma}^2 = \frac{n-1}{n} s^2$

biased: $E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2$ $\qquad$ bias $= -\frac{1}{n} \sigma^2$

more precise!

but has lower $var(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)^2 var(s^2)$
('coz it's smaller)

$\left(\frac{2}{n} - \frac{1}{n^2}\right) \sigma^4$

and $MSE(\hat{\sigma}^2) = \frac{1}{n^2} \sigma^4 + \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1} = \frac{2(n-1)+1}{n^2} \sigma^4$

Finite dist'ns for $\hat{\theta}$ often hard to determine, so we work with large sample results

Def. $\hat{\theta}$ consistent if $\text{plim}\ \hat{\theta} = \theta$ $\quad (\hat{\theta} \xrightarrow{P} \theta)$

$\hat{\theta}$ strongly consistent if $\hat{\theta} \xrightarrow{a.s.} \theta$

We'll try to apply a CLT to get $a_n(\hat{\theta} - \gamma) \xrightarrow{d} N(0, b)$

for a sequence $a_n \to 0$

approx. dist'd as or asymp. dist'd as

We often write $\hat{\theta} \overset{a}{\sim} N\left(\gamma, \frac{b}{a_n^2}\right)$ Realize this is not the pfd. form by statisticians

Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators.

$\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if $var(\hat{\theta}_1) < var(\hat{\theta}_2)$

Asymptotic efficiency is often measured using $\frac{b}{a_n^2}$ above

Cramér-Rao

lower bound on any unbiased estimator

Suppose Y has density $f(y;\theta)$. Then $\int f(y,\theta)\,dy = 1$

$\frac{d}{d\theta}$ each side (assume support indep of θ) $\quad \int \frac{\partial f(y,\theta)}{\partial\theta}\,dy = 0$

$$\int \frac{\partial f(y,\theta)/\partial\theta}{f(y,\theta)} f(y,\theta)\,dy = 0$$

$$\int \underbrace{\frac{\partial \ln f(y,\theta)}{\partial\theta}} f(y,\theta)\,dy = 0$$

Define this as the score fn: $\int S(y,\theta)\, f(y,\theta)\,dy = 0$

Note that $E(S(y,\theta)) = 0$ by def'n of expectation

Differentiate again: $\int \frac{\partial S(y,\theta)}{\partial\theta} f(y,\theta)\,dy + \int S(y,\theta)^2 f(y,\theta)\,dy =$

$$E\left(\frac{\partial S(y,\theta)}{\partial\theta}\right) + E\left[[S(y,\theta)]^2\right] = 0 \qquad (*)$$

If it exists, Define the information $I(\theta) = -E\left[\frac{\partial S(y,\theta)}{\partial\theta}\right]$

$$= -E\left[\frac{\partial^2 \ln f(y,\theta)}{\partial\theta^2}\right]$$

$$= E\left[[S(y,\theta)]^2\right] \qquad \text{from } (*)$$

$$= var(S(y,\theta)) \qquad \text{since } E(S(y,\theta)) = 0$$

Now... let $\hat{\theta} = g(y)$ be unbiased for θ. Then

$$\int g(y) f(y,\theta)\,dy = \theta$$

$\frac{d}{d\theta}$ and assume support indep of θ: $\int g(y) S(y,\theta) f(y,\theta)\,dy = 1$

so $E\left[g(y) S(y,\theta)\right] = 1$

since $cov(x,y) = E(xy) - E(x)E(y)$, rewrite as

$$E\left[g(y)\right] \underbrace{E\left[S(y,\theta)\right]}_{0} + cov(\hat{\theta}, S(y,\theta)) = 1$$

$$\Rightarrow var(\hat{\theta})\, var(S(y,\theta)) \geq 1$$

($cov(x,y) = \rho_{xy}\, var(x) var(y)$ ~~correlation~~ $\rho_{xy} \leq 1$

$cov(x,y) \leq var(x) var(y)$)

$$var(\hat{\theta}) \geq \frac{1}{var(S(y,\theta))} = I(\theta)^{-1}$$

pretty cool...

also works analogously for vector θ

Now we're ready to do MLE:

Define the likelihood function $\underset{k\times 1}{\mathcal{L}(\underset{\sim}{\theta})} = f(\underset{\sim}{Y}, \underset{\sim}{\theta})$ cond'l on observing $\underset{\sim}{Y}$.

if $Y_1, \ldots, Y_n \overset{iid}{\sim} f(y, \theta)$, then

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} f(y_i, \theta)$$

and the log-likelihood

$$L_n(\theta) = \ln(\mathcal{L}_n(\theta)) = \sum_{i=1}^{n} \ln f(y_i, \theta)$$

MLE: solves $\hat{\theta}_{MLE} = \arg\max_{\theta} L_n(\theta)$

Define $s_i(\theta) = \dfrac{\partial \ln f(Y_i, \theta)}{\partial \theta}$

Then the Score $S_n(\theta) = \sum_{i=1}^{n} s_i(\theta)$

and $I(\theta) = E\left[S_n(\theta) S_n(\theta)' \right]$ <u>vector</u> <u>version</u>

Then it's possible to show, under suitable regularity conditions, that

$$\hat{\theta}_{mle} \xrightarrow{P} \theta_0 \quad (\text{true value})$$

and

$$\sqrt{n}(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$$

so ~~we~~ MLE achieves the l.b.

Do <u>some</u> examples

iid $N(\mu, \sigma^2)$

uniform$(0, \theta)$ max $Y_1, Y_2, \ldots, Y_n$

Method of Moments

no dist'n required, just $\underset{\sim}{Y}_i \overset{iid}{\sim} (\underset{\sim}{\mu}, \Sigma)$

$$\hat{\mu}_{MOM} = \frac{1}{n} \Sigma Y_i = \bar{Y}$$

CLT tells us $\hat{\mu}_{MOM} \xrightarrow{a.s.} \mu \qquad \sqrt{n}(\hat{\mu}_{MOM} - \mu) \xrightarrow{d} N(0, \Sigma)$

of course, we need an estimate of Σ to do anything here

now suppose we want to recover the underlying params θ

If: $\underset{\ell \times 1}{\underset{\sim}{\mu}} = h(\underset{k \times 1}{\underset{\sim}{\theta}})$ $\quad \underline{k \leq \ell}$ $\quad$ overfitting: ℓ obs, k unknowns

$$\hat{\theta}_{MOM} = \underset{\theta}{\text{argmin}} \; (\bar{Y} - h(\theta))'(\bar{Y} - h(\theta))$$

min. distance btw. sample means and population means

can show $\hat{\theta}_{MOM}$ is consistent and asymp. normal

we'll stop there for the moment and return when we do GMM

Confidence Intervals

iid $N(\mu, \sigma^2)$ $\qquad z = \dfrac{\sqrt{n}(\bar{x} - \mu)}{s} \sim t_{n-1}$

we want to say something about the possible μ. Define a confidence level $1 - \alpha$, say 95% ($\alpha = 5\%$). Then the $1 - \alpha$ C.I. is given by

$$\{ \mu : z \in [t_{n-1}^{.025}, t_{n-1}^{.975}] \}$$

This is a random interval

Note that there are other 95% CI's. How to choose:

Symmetry, min. C.I width

APPENDIX C
CHAP. 17 OF GREENE
DEGROOT AND SCHERVISH

Hypothesis Testing

For now: $H_0: Y \sim F_{\theta_0}(Y)$ null

$H_A: Y \sim F_{\theta_A}(Y)$ alternative

only 2 possible parm values
which is right?

Type I error: ~~falsely reject null~~ null is true, we reject it

Type II error: ~~fail to reject null~~ null is false, we fail to reject it

Consider a test based on Y.

For $Y \in W$ (the critical region) we reject H_0

$Y \notin W$ we accept H_0

Goal: Choose W. Could do this w/a loss function. Usual approach:

fix the size of the test $\equiv \Pr(\text{Type I error})$

def power $\equiv \Pr(\text{reject } H_0 \text{ correctly}) = 1 - \Pr(\text{Type II error})$

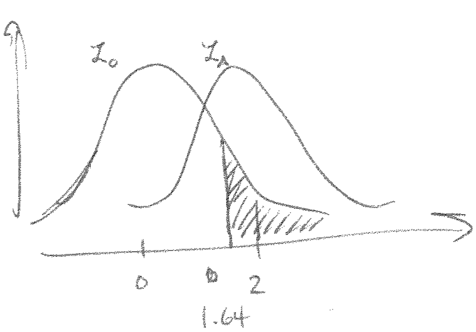
choose test to max power | size

Neyman-Pearson Lemma

M.P. test w/size α has crit. region $W_\alpha = \left\{ Y \;\middle|\; \frac{\mathcal{L}_A(Y)}{\mathcal{L}_0(Y)} > c_\alpha \right\}$

where c_α chosen s.t. $\Pr\left\{ \frac{\mathcal{L}_A(Y)}{\mathcal{L}_0(Y)} > c_\alpha \;\middle|\; Y \sim F_{\theta_0} \right\} = \alpha$

Ex. $N(0,1)$ vs $N(2,1)$



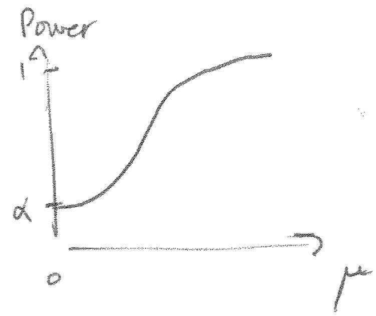
essentially
1-sided test

note: still reject at same spot if $N(3,1)$ is the A

Since the crit region is the same for all $\mu_A > 0$, the LR test is <u>uniformly</u> <u>most</u> <u>powerful</u> for $H_0: \mu = 0$ vs. $H_A: \mu > 0$.

with N-P lemma
So we basically end up w/ one-sided tests

Power function



Simple

Composite

LR for composite nulls vs. composite alt's. $LR = \dfrac{\max_{\theta \in \Theta_A} \mathcal{L}(\theta)}{\max_{\theta \in \Theta_0} \mathcal{L}(\theta)}$

Consistent if for fixed size, Power $\to 1$ as $n \to \infty$

biased test if Power < size for some $\theta \in \Theta_A$

invariant if same test can be applied to a given xformation of data

$N(\mu, \sigma^2)$ $H_0: \mu = 0$ vs $H_A: \mu > 0$

t-test ~~invariant~~ scale invariant : Same test for $X = aY$, $a > 0$

<u>LR</u> test

Recall that the NP lemma showed that the most powerful test against a simple alternative was a likelihood-ratio test.

The problem: we don't know what the crit. value should be 'coz we don't know the distn of the LR statistic in general

So we'll develop some large-sample results on the dist'n of LR

if $\hat{\theta}$ is the MLE for θ_0, then the max $LR = \mathcal{L}(\hat{\theta}) / \mathcal{L}(\theta_0)$

$$L_n(\theta) = \ln(\mathcal{L}(\theta))$$

Define: Likelihood Ratio Statistic $= 2 \ln(LR) = 2\left[L_n(\hat{\theta}) - L_n(\theta_0)\right]$
ξ_{LR}

This is a monotonic xformation of LR, so N-P tells us we can reject if $\xi_{LR} > c$ for some critical value c

Using the fact that L_n is thrice differentiable (one of the regularity conditions for MLE), the following form of the mean value thm holds:

$\exists\, \tilde{\theta}$ btw $\hat{\theta}$ and θ_0 s.t.

$$L_n(\theta_0) = L_n(\hat{\theta}) + (\theta_0 - \hat{\theta})' \cancel{\underbrace{\frac{\partial L_n(\hat{\theta})}{\partial \theta}}_{0}} + \frac{1}{2}(\theta_0 - \hat{\theta})' \frac{\partial^2 L_n(\tilde{\theta})}{\partial \theta^2} (\theta_0 - \hat{\theta})$$

So

$$\xi_{LR} = -(\hat{\theta} - \theta_0)' \frac{\partial^2 L_n(\tilde{\theta})}{\partial \theta \partial \theta'} (\hat{\theta} - \theta_0)$$

$$= -\sqrt{n}(\hat{\theta} - \theta_0)' \left[\frac{1}{n} \frac{\partial^2 L_n(\tilde{\theta})}{\partial \theta \partial \theta'}\right] \sqrt{n}(\hat{\theta} - \theta_0)$$

but we know that MLE converges to an asymp. normal:

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \bar{I}(\theta_0)^{-1})$$

where $\bar{I}$ is the information matrix for a single obs.

Also $\quad -\frac{1}{n} \sum \frac{\partial^2 \ln(\tilde{\theta})}{\partial \theta \partial \theta'} = -\frac{1}{n} \sum \frac{\partial^2 \ln f(Y_i, \tilde{\theta})}{\partial \theta \partial \theta'} \xrightarrow{d} \frac{1}{n} I(\theta_0)$

$= \bar{I}(\theta_0)$

So

$\xi_{LR} \xrightarrow{d} \chi^2_K \quad$ because it's just a quadratic form $X'\Sigma^{-1}X \quad \underset{k\times 1}{X} \sim N(0, \Sigma)$

is just the sum of k squared std. normals what to remember about LR:

need to calculate L under restricted and unrestricted θ

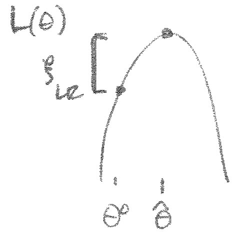
not really useful in N-P test of simple null vs. simple alt.

not $\theta = 1$ vs. $\theta = 2$

but $\theta = 1$ vs. $\theta =$ unrestricted

the 2 models have to be nested

LR graphically:



Wald test

main advantage: only requires $L_n(\hat{\theta})$ (unrestricted), not $L_n(\theta_0)$

The Wald stat is

$$\xi_W = (\hat{\theta} - \theta_0)' \left[\cancel{Var(\hat{\theta} - \theta_0)}\right]^{-1} (\hat{\theta} - \theta_0) \sim \chi^2_k$$

$$\left[- \frac{\partial^2 L_n(\hat{\theta})}{\partial\theta\partial\theta'} \right]$$

only difference (from LR): evaluated at $\hat{\theta}$, not $\tilde{\theta}$

but $\hat{\theta}, \tilde{\theta} \xrightarrow{P} \theta_0$ under H_0, so $\xi_W \xrightarrow{P} \xi_{LR} \sim \chi^2_k$

what is this middle term? Recall that

$$-E\left[\frac{\partial^2 L_n(\theta_0)}{\partial\theta\partial\theta'}\right] = I(\theta_0)$$

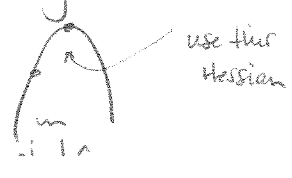
and $Var(\hat{\theta}) = I(\theta_0)^{-1}$ under H_0

asymptotically we ~~can drop~~ converge to the expectation and we can evaluate the derivative at $\hat{\theta}$, so

$$\xi_W = (\hat{\theta} - \theta_0)' \left[Var(\hat{\theta})\right]^{-1} (\hat{\theta} - \theta_0)$$

which is just the usual quadratic form

Graph



We can also use a Wald test w/ nonlinear restrictions.

Suppose $H_0: \underset{k\times 1}{c(\theta)} = \underset{k\times 1}{q}$

Then

$$\xi_W = (c(\hat{\theta}) - q)' \left[\text{aVar}(c(\hat{\theta}) - q)\right]^{-1} (c(\hat{\theta}) - q)$$

All we need to do is linearize w/ the delta method:

$$\text{Var}(c(\hat{\theta}) - q) = \frac{\partial c(\hat{\theta})}{\partial \theta'} \text{var}(\hat{\theta}) \left[\frac{\partial c(\hat{\theta})}{\partial \theta'}\right]'$$

If the restrictions are linear, the derivatives are constant:

$$H_0: \underset{k\times p}{R} \underset{p\times 1}{\theta} = \underset{k\times 1}{q}$$

Then

$$\xi_W = (\underset{k\times p}{R} \underset{p\times 1}{\hat{\theta}} - \underset{k\times 1}{q})' \left[R \, \text{avar}(\hat{\theta}) R'\right]^{-1} (R\hat{\theta} - q) \sim \chi^2_k$$

<u>Ex.</u> m-variate normal

$$\underset{p\times 1}{X_i} \overset{iid}{\sim} N(\mu, \Omega)$$

$$\log \mathcal{L} = \sum_i -\frac{1}{2} p \log(2\pi) - \frac{1}{2} \log|\Omega| - \frac{1}{2}(X_i - \mu)' \Omega^{-1} (X_i - \mu)$$

$$= -\frac{1}{2} pn \log(2\pi) - \frac{1}{2} n \log|\Omega| - \frac{1}{2} \sum_i (X_i - \mu)' \Omega^{-1} (X_i - \mu)$$

MLE:

$$\frac{\partial \log \mathcal{L}}{\partial \mu} = \sum_i \Omega^{-1}(X_i - \mu) = \underset{\sim}{0} \Rightarrow \hat{\mu} = \bar{X}$$

$$\frac{\partial \log \mathcal{L}}{\partial \Omega} = \text{eek} \Rightarrow \hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})'$$

(easy to show in univariate case)

and $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N(0, \hat{\Omega})$

Ex. univariate normal

$$X_i \overset{iid}{\sim} N(\mu, \sigma^2) \qquad \mathcal{L} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\log \mathcal{L} = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

MLE

$$\frac{\partial \log \mathcal{L}}{\partial \mu} = \sum_{i=1}^{n}\frac{(x_i-\mu)}{\sigma^2} = 0 \quad \Rightarrow \quad \hat{\mu} = \bar{X}$$

$$\frac{\partial \log \mathcal{L}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = \sum \frac{(X_i-\hat{\mu})^2}{n}$$

$$\frac{\partial^2 \log L}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 \log L}{\partial \mu \partial \sigma^2} = -\frac{\sum(x_i-\mu)}{\sigma^4}$$

$$\frac{\partial^2 \log L}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum(x_i-\mu)^2$$

$$I(\theta) = -E\left[\frac{\partial^2 \log \mathcal{L}}{\partial\theta\partial\theta'}\right] = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

$$I^{-1}(\theta) = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix} = Var\begin{pmatrix} \hat{\mu}, \\ \hat{\sigma}^2 \end{pmatrix}$$

and we can run with that for the Wald stat eval'd at $\hat{\theta}$

LM or Score test

based on just the restricted model

~~Define~~

Idea: max $\log \mathcal{L}$ s.t. constraints $c(\theta) - q = 0$

write down the Lagrangean:

$$\max \ \log \mathcal{L}(\theta) + \lambda'(c(\theta) - q)$$

F.O.C.

$$\frac{\partial}{\partial \theta}: \quad \frac{\partial \log \mathcal{L}(\theta)}{\partial \theta} + C'\lambda = 0 \qquad C = \frac{\partial c(\theta)}{\partial \theta}$$

$$\frac{\partial}{\partial \lambda} \quad c(\theta) - q = 0$$

If H_0 valid, then $\dfrac{\partial \log \mathcal{L}(\theta)}{\partial \theta} = 0$, and $\underbrace{\dfrac{\partial \log \mathcal{L}(\hat\theta_0)}{\partial \theta_0}}_{\text{score eval'd at restricted MLE}} = 0$

Def.

$$\xi_{LM} = \left(\frac{\partial \log \mathcal{L}(\hat\theta_0)}{\partial \hat\theta_0}\right)' \left[\mathrm{I}(\hat\theta)\right]^{-1} \left(\frac{\partial \log \mathcal{L}(\hat\theta_0)}{\partial \hat\theta}\right)$$

$$= S_n(\theta_0)' \left[-\frac{\partial^2 L_n(\theta_0)}{\partial \theta_0 \partial \theta_0'}\right]^{-1} S_n(\theta_0)$$

skip (

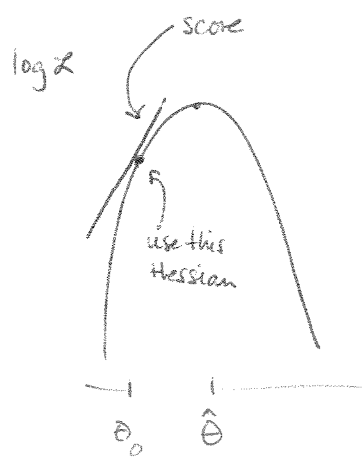
$$\cancel{= \frac{1}{\sqrt{n}} S_n(\theta_0)' \left[-\frac{1}{n} \frac{\partial^2 L_n(\theta_0)}{\partial \theta_0 \partial \theta_0}\right]^{-1} \frac{1}{\sqrt{n}} S_n(\theta_0)}$$

Since under H_0

$$S_n(\theta_0) \xrightarrow{d} \mathrm{I}(\theta_0)$$

$$\frac{\partial^2 L_n(\theta_0)}{\partial \theta_0 \partial \theta_0'} \xrightarrow{d} \mathrm{I}(\theta_0)^{-1}$$

then $\xi_{LM} \xrightarrow{d} \chi^2_K$



what if we don't want to restrict all params?

Partition $\underset{p\times 1}{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}\begin{matrix} k\times 1 \\ (p-k)\times 1 \end{matrix}$ $\quad H_0: \theta_1 = \theta_{10}$ vs $\theta_1 \neq \theta_{10}$

let $\hat{\theta}$ be the MLE (unrestricted) $\quad \hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}$

$\tilde{\theta} \equiv \underset{\theta}{\arg\max}\ L(\theta)$ (restricted) $\quad \tilde{\theta} = \begin{bmatrix} \theta_{10} \\ \tilde{\theta}_2 \end{bmatrix}$

s.t. $\theta_1 = \theta_{10}$

Let's guess that we can ignore θ_2 completely to get the Wald.

$$\xi_W = (\hat{\theta}_1 - \theta_{10})' \, avar(\hat{\theta}_1)^{-1} (\hat{\theta}_1 - \theta_{10}) \sim \chi^2_k$$

what are these?

Recall $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1})$ under H_0

Partition $\mathcal{I}^{-1} = \begin{bmatrix} \mathcal{I}^{11} & \mathcal{I}^{12} \\ \mathcal{I}^{21} & \mathcal{I}^{22} \end{bmatrix}$

Then $\sqrt{n}(\hat{\theta}_1 - \theta_{10}) \xrightarrow{d} N(0, \mathcal{I}^{11})$

so

$$\xi_W = \sqrt{n}(\hat{\theta}_1 - \theta_{10})' (n\mathcal{I}_{11})^{-1} \sqrt{n}(\hat{\theta}_1 - \theta_{10}) \sim \chi^2_k$$

and we're golden (just ~~...~~ have to invert whole information matrix which can be hard)

Similar results for LM and LR