Hypothesis Testing in the Maximum Likelihood Context

MLE gives us an "asymptotically" optimal estimation principle. Might anticipate that constructing tests based on MLE's will lead to tests which perform well, at least asymptotically.

Let's review some basic aspects of hypothesis testing, & see how MLE can help us.

Problem : $H_0 : \underset{\sim}{\theta} \in \Omega_0$

$H_1 : \underset{\sim}{\theta} \in \Omega_1$

(If $\Omega_0 \subset \Omega_1$, hypotheses are "nested".)

e.g.: $y = \beta_1 + \beta_2 x + \varepsilon$

$H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$.

Definition: A "Test" is a decision rule that leads us to reject / not reject a stated hypothesis.

Specifically : $T_n = T_n(y_1, \ldots, y_n)$ is a "test statistic" – a fctn. of the random $\{y_i\}$ data. Our test (rule) is of the form:

Reject H_0 if $T_n \in C$

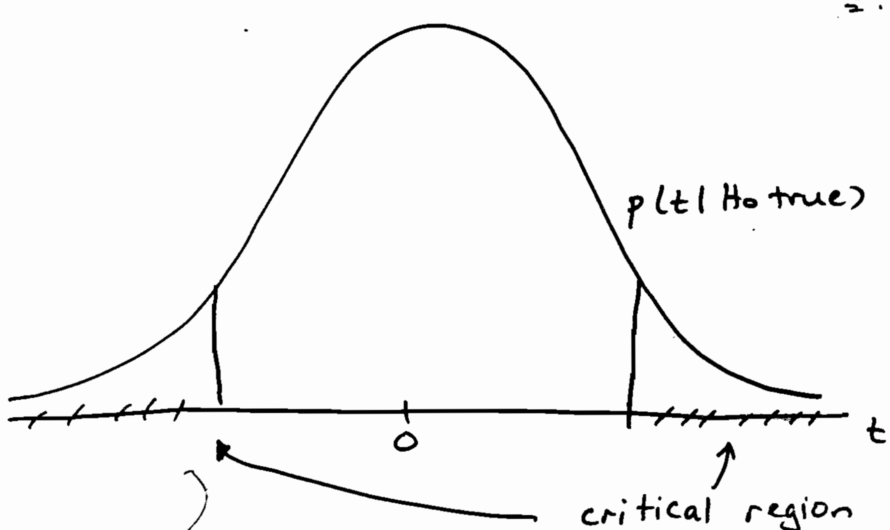
where C is the "critical region" (or 'rejection region') of the sample space.

Usually T_n is an estimator of θ, or a fctn. of such an estimator.

(This where MLE may enter the story.)

e.g. $H_0 : \beta_2 = 0$; $H_1 : \beta_2 \neq 0$

$$T_n = t_{n-k} = \frac{\tilde{\beta}_2 - 0}{s.e.(\tilde{\beta}_2)}$$



Recall, 2 types of errors we may incur:

Decision	True State of Nature: H_0 True	H_0 False
"Accept" H_0	✓	Type II
Reject H_0	Type I	✓

We call Pr. (Type I error) the "size" of the test ("significance level")

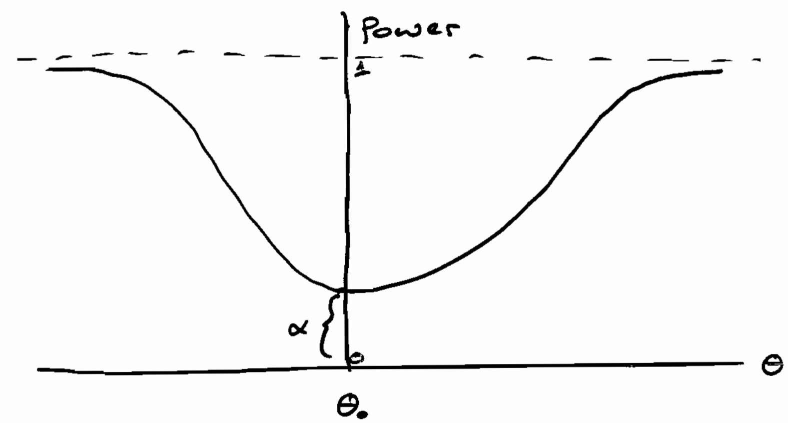
$\alpha = Pr.[T_n \in C \mid \theta \in \Omega_0]$

In classical hypothesis testing we set (upper limit for) value of α. Then try to construct a test which has small Pr. (Type II error). That is, we try to construct test so that:

$Pr.[T_n \in \bar{C} \mid \underline{\theta} \in \Omega_1]$ is small; or:

Power $= Pr.[T_n \in C \mid \underline{\theta} \in \Omega_1]$ is large.

[Power $= 1 - Pr.$ (Type II error)].



$H_0: \theta = \theta_0$; H_1 $\theta \neq \theta_0$

Recall the following test properties:

* Unbiased
* Consistent
* Most Powerful
* UMP
* Locally most Powerful.

The Big Question is:

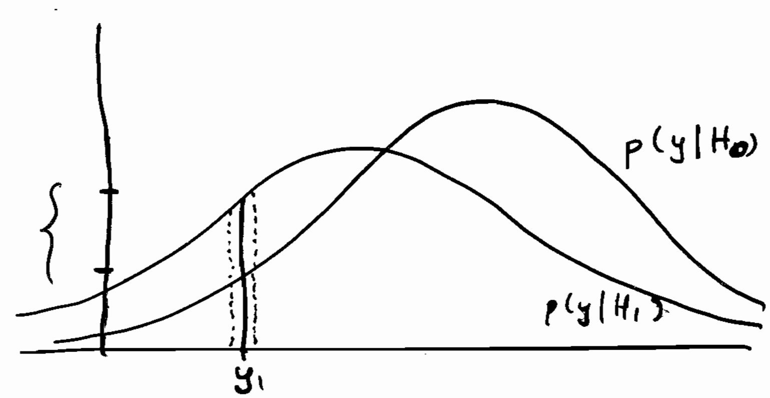
For a given choice of α, how can we construct a (uniformly most) powerful test for a particular testing problem?

Perhaps not surprisingly, this is where the Likelihood fctn. and MLE can help.

Recall – Likelihood fctn. provides a full description of random sample data.

First, some simple intuition, then the answer to the B.Q.

Suppose we have a scalar θ, & $n = 1$:



$$\Pr[\, y_1 - \varepsilon < y < y_1 + \varepsilon \mid H_1 \,]$$
$$> \Pr[\, y_1 - \varepsilon < y < y_1 + \varepsilon \mid H_0 \,]$$

In the limit, as $\varepsilon \to 0$:

$$p(y_1 \mid H_1) > p(y_1 \mid H_0)$$

which suggests that we should <u>reject H_0</u> as being "less likely" than H_1.

When $n > 1$ we generalize this idea by using $p(y_1, y_2, \ldots, y_n)$ — i.e. the <u>Likelihood Function</u>.

We're going to consider <u>three</u> general testing principles that use this intuition. They differ in terms of the point at which we evaluate the L.F. — at H_0, at H_1, or at both.

History —

* Likelihood Ratio Test (Both)
  (Neyman & Pearson ; 1928-1934)
* Wald Test (H_1)
  (Wald ; 1943)
* Lagrange Multiplier Test (H_0)
  (Silvey ; 1959)
  (Variation of Rao's "Score Test"; 1948)

First, we'll deal with the LRT —

<u>Lemma (Neyman - Pearson)</u>:

Suppose we have $\{y_1, \ldots, y_n\}$ drawn from $p(y_1, \ldots, y_n ; \theta)$.

Let
$$\lambda = \lambda(y_1, \ldots, y_n ; \theta_0, \theta_1) = \frac{p(y_1, \ldots, y_n ; \theta_0)}{p(y_1, \ldots, y_n ; \theta_1)} = \frac{L(\theta_0)}{L(\theta_1)} \qquad (= T_n)$$

& let $C = \{ \lambda : \lambda \le k \}$

where 'k' is such that

$$\Pr.[\lambda \in C \mid \theta = \theta_0] = \alpha ,$$

& α = chosen size for test.

Then C is the best critical region for testing $\underline{H_0 : \theta = \theta_0}$ against $\underline{H_1 : \theta = \theta_1}$, in the sense that the associated test is Most Powerful.

N.B.

i) No guarantee that such a test exists (it will if $p(y_1, --, y_n)$ is a <u>density</u>)

ii) 'Simple' null & alternative — no parameters to estimate. (Very special case.)

iii) Idea – put points into critical region until it reaches size of α.

iv) To max. power, points that are more likely under H_1 than under H_0 should be put into C.

v) If a point $(y_1, ---, y_n)$ is more likely under H_1 than under H_0 then this increases denominator of λ & makes λ <u>smaller</u>. Hence, reject if $\lambda < k$.

Practical Issue –

To apply test we need to know k, such that $P_r.[\lambda < k \mid \theta = \theta_0] = \alpha$.

This requires knowledge of <u>distribution</u> of the test statistic, λ. This may be tricky!

<u>Example</u>: Binomial Dist'n.

Recall : $L(p) = {}^nC_x \, p^x (1-p)^{n-x}$

& $\tilde{p} = x/n$.

$H_0 : p = 1/3$ vs. $H_1 : p = 2/3$

$$\lambda = \frac{{}^nC_x (1/3)^x (2/3)^{n-x}}{{}^nC_x (2/3)^x (1/3)^{n-x}}$$

$$= \left(\tfrac{1}{2}\right)^x 2^{n-x}$$

$$= 2^{n-2x}.$$

So, calculate λ & reject H_0 if $\lambda < k$, where k is chosen so that

$P_r[\lambda < k \mid p = 1/3] = \alpha$. How??

Now, usually in practice we <u>don't</u> have <u>simple</u> H_0 & H_1. Usually $\underset{\sim}{\theta}$ is vector.

For example –

$$y_i \sim N(\mu, \sigma^2)$$

$$H_0 : \mu = 0 \quad \underline{vs.} \quad H_1 : \mu > 0 \qquad (\underline{etc.})$$

Generalized Likelihood Ratio Test –

Let $\tilde{\lambda} = L(\tilde{\theta}_0) / L(\tilde{\theta}_1)$

& reject H_0 if $\tilde{\lambda} < k$, where k s.t.

$$\Pr.[\tilde{\lambda} < k \mid H_0 \text{ true}] = \alpha.$$

We are now evaluating L.F. twice – once under H_0 & once under H_1.

Often this LRT has good power properties, but need not be (U)MP. Sometimes is.

Another point to note is that while dist'n. of $\tilde{\lambda}$ may be unclear, we may be able to resolve issue after suitable transformation.

<u>Example</u>: $\quad y = X\beta + \varepsilon \quad ; \varepsilon \sim N[0, \sigma^2 I]$

$$H_0 : R\beta = q \quad \underline{vs.} \quad H_1 : R\beta \neq q.$$

Recall –

$$L(\beta, \sigma^2) = (\sigma^2 2\pi)^{-n/2} \exp\left[-\tfrac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right]$$

Leads to the MLE's: $\begin{cases} \tilde{\beta}_1 = (X'X)^{-1}X'y \\ \tilde{\sigma}_1^2 = \frac{1}{n}(y - X\tilde{\beta}_1)'(y - X\tilde{\beta}_1) \end{cases}$

So, $L(\tilde{\beta}_1, \tilde{\sigma}_1^2) = (2\pi\tilde{\sigma}_1^2)^{-n/2} \exp(-n/2)$.

Now, imposing the restrictions, we get the restricted MLE's —

$$\tilde{\beta}_0 = \tilde{\beta}_1 + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(q - R\tilde{\beta}_1)$$

$$\& \quad \tilde{\sigma}_0^2 = 1/n \, (y - X\tilde{\beta}_0)'(y - X\tilde{\beta}_0).$$

So: $L(\tilde{\beta}_0, \tilde{\sigma}_0^2) = (2\pi\tilde{\sigma}_0^2)^{-n/2} \exp(-n/2)$.

Now, the Likelihood Ratio is:

$$\tilde{\lambda} = L(\tilde{\beta}_0, \tilde{\sigma}_0^2) / L(\tilde{\beta}_1, \tilde{\sigma}_1^2)$$

$$= (\tilde{\sigma}_0^2 / \tilde{\sigma}_1^2)^{-n/2}$$

$$= \left[\frac{\frac{1}{n}(y - X\tilde{\beta}_0)'(y - X\tilde{\beta}_0)}{\frac{1}{n}(y - X\tilde{\beta}_1)'(y - X\tilde{\beta}_1)} \right]^{-n/2}$$

$$= (SSR_0 / SSR_1)^{-n/2} .$$

The distribution of $\tilde{\lambda}$ is unknown. So how do we apply the LRT?

Consider the monotonic (decreasing) fctn. of $\tilde{\lambda}$ —

$$\tilde{\lambda}^* = (\tilde{\lambda}^{-2/n} - 1)(n-k)/J \qquad (J = \#\text{ restrictions})$$

$$[\partial \tilde{\lambda}^* / \partial \tilde{\lambda} = (\tfrac{n-k}{J})(-\tfrac{2}{n})\tilde{\lambda}^{-2/n - 1} < 0]$$

So, rejecting H_0 if $\tilde{\lambda} < k$ is equivalent to rejecting H_0 if $\tilde{\lambda}^* > k^*$;

& $$\tilde{\lambda}^* = \frac{(SSR_0 - SSR_1)/J}{SSR_1/(n-k)} .$$

Note familiarity of $\tilde{\lambda}^*$ — it is F-distributed if H_0 true! So:

(i) Choose α.

(ii) Get k^* from table for $F_{J, n-k}$

(iii) Reject H_0 if $\tilde{\lambda}^* > k^*$.

The usual F-test for linear restrictions (& t-test) is <u>equivalent to the LRT</u>!

(In this case it is a UMP Test.)

However, in other problems, may be impossible to get dist'n. for $\tilde{\lambda}$ in finite samples, even after transformation.

There is an extremely important class of testing problems where we can always apply LRT with <u>asymptotic</u> validity — "<u>Nested</u>" Hypotheses.

Suppose H_0 is nested within H_1. That is, H_0 involves imposing restrictions on H_1.

[e.g. : $\varepsilon_t = \rho \varepsilon_{t-1} + u_t$

$H_0 : \rho = 0$ vs. $H_A : \rho \neq 0$.]

In this case, clearly:

$$\tilde{\lambda} = [L(\tilde{\theta}_0) / L(\hat{\theta}_1)] \in (0, 1]$$

Let m = # independent restrictions to get H_0 from H_1.

Then –

If the null & alternative hypotheses are nested, & the usual regularity conditions are satisfied,

$$LRT = -2 \log \tilde{\lambda} \xrightarrow{d} \chi^2_{(m)}$$

if H_0 is true.

N.B.:

* This last result is only for "nested" case.
* Only an asymptotic result.
* LRT is a "consistent" test.
* Power may be low in finite samples.
* Computational point – need to evaluate $L(\theta)$ twice – $L(\tilde{\theta}_0)$ & $L(\hat{\theta}_1)$.

  [This is usually very simple in practice.]

Example — Probit model :

LIKELIHOOD RATIO TEST EXAMPLE – PROBIT MODEL

The data set on voting behaviour is from Pindyck and Rubinfeld, *Econometric Models and Economic Forecasts*, 1998, Fourth Edition, Table 11.8, p. 332.

The variables are:

YESVM = dummy variable equal to 1 if individual voted yes in the election; 0 if they voted no.

PUB12 = 1 if 1 or 2 children in public school; = 0 otherwise

PUB34 = 1 if 3 or 4 children in public school; = 0 otherwise

PUB5 = 1 if 5 or more children in public school; = 0 otherwise

PRIV = 1 if 1 or more children in private school; = 0 otherwise

YEARS = number of years living in the community

SCHOOL = 1 if individual is employed as a teacher; = 0 otherwise

LOGINC= logarithm of annual household income (in dollars)

PTCON = logarithm of property taxes (in dollars) paid per year

Pr.(YESVM = 1) = F(PUB12, PUB34, PUB5, PRIV, YEARS, SCHOOL, LOGINC, PTCON) + ERROR

Dependent Variable: YESVM
Method: ML - Binary Probit (Quadratic hill climbing)

Sample: 1 95
Included observations: 95
Convergence achieved after 6 iterations
QML (Huber/White) standard errors & covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.958854	4.468545	-0.662151	0.5079
PUB12	0.368347	0.388089	0.949130	0.3426
PUB34	0.691512	0.438233	1.577954	0.1146
PUB5	0.295633	0.774494	0.381712	0.7027
PRIV	-0.211105	0.452844	-0.466176	0.6411
SCHOOL	1.582916	0.481557	3.287077	0.0010
YEARS	-0.015767	0.017930	-0.879372	0.3792
LOGINC	1.314272	0.440957	2.980501	0.0029
PTCON	-1.464083	0.570259	-2.567401	0.0102

Mean dependent var	0.621053	S.D. dependent var	0.487699
S.E. of regression	0.464050	Akaike info criterion	1.308242
Sum squared resid	18.51941	Schwarz criterion	1.550189
Log likelihood	-53.14151	Hannan-Quinn criter.	1.406007
Restr. log likelihood	-63.03691	Avg. log likelihood	-0.559384
LR statistic (8 df)	19.79080	McFadden R-squared	0.156978
Probability(LR stat)	0.011157		

Obs with Dep=0	36	Total obs	95
Obs with Dep=1	59		

Dependent Variable: YESVM
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 02/10/06 Time: 08:55
Sample: 1 95
Included observations: 95
Convergence achieved after 6 iterations
QML (Huber/White) standard errors & covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-5.133959	4.107513	-1.249895	0.2113
SCHOOL	1.793063	0.424241	4.226515	0.0000
LOGINC	1.327226	0.426446	3.112292	0.0019
PTCON	-1.137838	0.525152	-2.166681	0.0303

Mean dependent var	0.621053	S.D. dependent var	0.487699
S.E. of regression	0.459992	Akaike info criterion	1.243574
Sum squared resid	19.25497	Schwarz criterion	1.351105
Log likelihood	-55.06975	Hannan-Quinn criter.	1.287025
Restr. log likelihood	-63.03691	Avg. log likelihood	-0.579682
LR statistic (3 df)	15.93432	McFadden R-squared	0.126389
Probability(LR stat)	0.001170		

Obs with Dep=0	36	Total obs	95
Obs with Dep=1	59		

Test joint hypothesis that all coefficients of the omitted 5 regressors are zero:

LRT = 2[(-53.14151) – (-55.06975)] = 3.856 $\longrightarrow \chi^2(5)$
(5% crit. Value = 11.07)

The Wald Test :

This test is also designed for testing "nested" hypotheses, but requires only $\tilde{\theta}_1$ to be calculated. Useful if calculation of $\tilde{\theta}_0$ is computationally difficult.

To motivate Wald Test, consider linear restrictions :

$$H_0 : \underset{(J\times k)(k\times 1)}{R\,\underset{\sim}{\theta}} = \underset{(J\times 1)}{\underset{\sim}{r}} \quad \text{vs.} \quad H_1 : R\underset{\sim}{\theta} \neq \underset{\sim}{r}$$

If H_0 is true, expect $R\tilde{\theta}_1 \approx r$, at least as $n \rightarrow \infty$.

We know that $\sqrt{n}\,(\tilde{\theta}_1 - \theta) \xrightarrow{d} N[0, IA(\theta)^{-1}]$

So:

$$\sqrt{n}\,R(\tilde{\theta}_1 - \theta) \xrightarrow{d} N[0, RIA^{-1}R']$$

and if H_0 is true $(R\theta = r)$ —

$$\sqrt{n}\,(R\tilde{\theta}_1 - r) \xrightarrow{d} N[0, RIA^{-1}R'].$$

Immediately,

$$\sqrt{n}\,(R\tilde{\theta}_1 - r)'\,[R\,IA^{-1}R']^{-1}\sqrt{n}\,(R\tilde{\theta}_1 - r)$$

$$= n\,(R\tilde{\theta}_1 - r)'\,[R\,IA^{-1}R']^{-1}\,(R\tilde{\theta}_1 - r)$$

$$\xrightarrow{d} \chi^2_{(J)} \qquad (\text{If } H_0 \text{ is true.})$$

Problem: $IA = IA(\theta)$ & is unobservable.

Just replace it with any consistent estimator, & asy. dist'n. unchanged.

Estimate $IA(\theta)$ by $\frac{1}{n} I^*(\tilde{\theta}_1)$, where

$$\text{plim}\,[\tfrac{1}{n} I^*(\tilde{\theta}_1)] = IA(\theta).$$

So, replace $\frac{1}{n} IA^{-1}$ by $I^*(\tilde{\theta}_1)^{-1}$:

Wald test statistic is —

$$\omega = (R\tilde{\theta}_1 - r)'\,[R\,I^*(\tilde{\theta}_1)^{-1}R']^{-1}\,(R\tilde{\theta}_1 - r)$$

$$\xrightarrow{d} \chi^2_{(J)} \qquad (\text{if } H_0 \text{ is true})$$

Test: Reject H_0 if $\omega > \chi^2_{(J)}(\text{crit})$ for chosen significance level.

(Makes sense — if H_0 true, expect that $R\tilde{\theta}_1 \approx r$, & so ω will be small.)

* Although Wald test statistic has same asymptotic dist'n. as LRT, usually differs in finite samples.

* Also, dist'ns. of Wald & LRT may differ if H_0 false $\Rightarrow$ different powers.

<u>Example</u>:

$$y = X\beta + \varepsilon \;; \quad \varepsilon \sim N[0, \sigma^2 I]$$

$$H_0 : R\beta = r \quad \text{vs.} \quad H_1 : R\beta \neq r.$$

With Normal errors —

$$I(\beta, \sigma^2) = \begin{bmatrix} (X'X)/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{bmatrix}$$

$$IA = \text{plim}\,[\tfrac{1}{n} I] = \begin{bmatrix} Q/\sigma^2 & 0 \\ 0 & 1/2\sigma^4 \end{bmatrix}$$

where $Q = \underset{n\to\infty}{\text{Limit}}\,[\tfrac{1}{n} X'X]$.

Note that

$$IA^{-1} = \begin{bmatrix} \sigma^2 Q^{-1} & 0 \\ 0 & 2\sigma^4 \end{bmatrix}.$$

The "block" relating to the 'β' part of the model is $\sigma^2 Q^{-1}$. For this block, $I^* = (X'X/\tilde{\sigma}_1^2)$, and

$$I^{*-1} = \tilde{\sigma}_1^2 (X'X)^{-1}.$$

So, to test H_0, use the statistic:

$$W = (R\tilde{\beta}_1 - r)'[R\tilde{\sigma}_1^2 (X'X)^{-1}R']^{-1}(R\tilde{\beta}_1 - r)$$

$$= (R\tilde{\beta}_1 - r)'[R(X'X)^{-1}R']^{-1}(R\tilde{\beta}_1 - r)/\tilde{\sigma}_1^2$$

Note:

(i) W depends only on unrestricted MLE.

(ii) Can replace $\tilde{\sigma}_1^2$ with any other consistent estimator of σ^2 — e.g., s^2.

(iii) Then,

$$W^* = \left(\frac{1}{J}\right)W = \frac{(R\tilde{\beta}_1 - r)'[R(X'X)^{-1}R']^{-1}(R\tilde{\beta}_1 - r)/J}{s^2}$$

So, for this problem, the Wald test statistic proportional to the usual F-statistic. $W^* \sim F_{J,n-k}$ in finite samples. Because $J>0$, the Wald test & F-test are equivalent for this case.

Non-Linear Restrictions:

If we want to test non-linear restrictions, we can use the LRT. We can also use the Wald test, but care is needed!

Suppose we have J differentiable independent restrictions:

$$r_j(\theta) = 0 \quad ; \quad j = 1, 2, \ldots, J$$

Taking a Taylor's series approximation:

$$\underset{(1\times 1)}{r_j(\tilde{\theta})} \approx \underset{(1\times 1)}{r_j(\theta)} + \underset{(1\times k)}{(\tilde{\theta}-\theta)'}\underset{(k\times 1)}{\left[\frac{\partial r_j(\theta)}{\partial \theta}\right]}$$

27.

Now, if H_0 is true ($r_j(\theta) = 0$)

then: $r_j(\tilde{\theta}) \approx (\tilde{\theta} - \theta)' \left[\frac{\partial r_j(\theta)}{\partial \theta} \right]$

Now, let

$$\underset{(k \times J)}{\tilde{R}} = \left[\left.\frac{\partial r_1(\theta)}{\partial \theta}\right|_{\tilde{\theta}}, \ldots, \left.\frac{\partial r_J(\theta)}{\partial \theta}\right|_{\tilde{\theta}} \right]$$

& let $\underset{(J \times 1)}{r(\tilde{\theta})} = [r_1(\tilde{\theta}), \ldots, r_J(\tilde{\theta})]'$

Then, if we use $\tilde{\theta}_1$ for $\tilde{\theta}$:

$$W = r(\tilde{\theta}_1)' [\tilde{R}' I^*(\tilde{\theta}_1)^{-1} \tilde{R}]^{-1} r(\tilde{\theta}_1) \xrightarrow{d} \chi^2_{(J)} \quad (\text{if } H_0 \text{ is true})$$

Problem (Gregory & Veall, 1985) -

Unlike the LRT, the Wald test statistic is not invariant to the way restrictions are written:

$$\begin{cases} H_0 : \beta_2 = 1 - (\beta_0/\beta_1) \\ H_0' : \beta_1\beta_2 = \beta_1 - \beta_0 \end{cases}$$

TABLE I

NUMBER OF REJECTIONS AT THE 5 PER CENT LEVEL IN 1000 TRIALS USING WALD TESTS OF SIMPLE NONLINEAR RESTRICTIONS

Case	Form of Restriction	$N=20$	$N=30$	$N=50$	$N=100$	$N=500$
Null hypothesis true:						
$\beta_1=10.0,\ \beta_2=0.1$	A	293	253	206	142	78
	B	65	62	63	64	39
$\beta_1=5.0,\ \beta_2=0.2$	A	201	152	119	108	77
	B	64	58	57	48	56
$\beta_1=2.0,\ \beta_2=0.5$	A	86	89	78	58	43
	B	61	53	71	64	40
$\beta_1=1.0,\ \beta_2=1.0$	A	53	45	53	43	44
	B	69	47	65	46	47
Null hypothesis false:						
$\beta_1=1.5,\ \beta_2=1.0$	A	443	554	833	985	1000
	B	278	399	728	971	1000
$\beta_1=1.0,\ \beta_2=0.5$	A	65	196	601	992	1000
	B	584	775	943	998	1000

(with one degree of freedom because there is only one restriction). Using (4), the two alternative Wald test statistics can be calculated as:

(5) $$W^A = (\hat{\beta}_1\hat{\beta}_2 - 1)^2/(\hat{\beta}_2^2 v_{11} + 2v_{12} + v_{22}/\hat{\beta}_2^2)$$

and

(6) $$W^B = (\hat{\beta}_1\hat{\beta}_2 - 1)^2/(\hat{\beta}_2^2 v_{11} + 2\hat{\beta}_1\hat{\beta}_2 v_{12} + \hat{\beta}_1^2 v_{22}),$$

<u>The Lagrange Multiplier Test</u>:

Again, nested hypotheses – base test only on the <u>restricted</u> MLE for parameters. (Especially useful if imposing the restrictions makes MLE simpler.)

Starting point —

Ignoring any restrictions, obviously

$$[\partial \log L_1(\theta) / \partial\theta]\big|_{\tilde{\theta}_1} = 0$$

i.e. $D \log L_1(\tilde{\theta}_1) = 0$, for simplicity.

If H_0 is true, expect also that

$$D \log L_1(\tilde{\theta}_0) \simeq 0 .$$

This suggests a test statistic of the form:

$$LM = D\log L_1(\tilde{\theta}_0)' \, I^*(\tilde{\theta}_0)^{-1} \, D\log L_1(\tilde{\theta}_0)$$

* $LM \xrightarrow{d} \chi^2_{(J)}$, <u>if H_0 true</u>.

<u>Example</u>:

$$y = X\beta + \varepsilon \; ; \; \varepsilon \sim N[0, \sigma^2 I]$$

$$H_0 : R\beta = r \quad \underline{vs.} \quad H_1 : R\beta \neq r .$$

$$\log L = -\tfrac{1}{2}\log 2\pi - \tfrac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y'y + \beta'X'X\beta - 2\beta'X'y)$$

$$(\partial \log L / \partial\beta) = -\tfrac{1}{2\sigma^2}(2X'X\beta - 2X'y) = -\tfrac{1}{\sigma^2}(X'X\beta - X'y)$$

(Just focus on β — see H_0, H_1.)

Now, as we saw earlier, $I^{*-1} = \sigma^2(X'X)^{-1}$.

This can be estimated consistently by $\tilde{\sigma}_0^2(X'X)^{-1}$, where $\tilde{\sigma}_0^2 = \frac{1}{n}(y - X\tilde{\beta}_0)'(y - X\tilde{\beta}_0)$

and $\tilde{\beta}_0$ = RLS estimator.

So:

$$LM = \left(-\frac{1}{\tilde{\sigma}_0^2}\right)(X'X\tilde{\beta}_0 - X'y)'\tilde{\sigma}_0^2(X'X)^{-1} \cdot \left(-\frac{1}{\tilde{\sigma}_0^2}\right)(X'X\tilde{\beta}_0 - X'y)$$

$$= \frac{(X'X\tilde{\beta}_0 - X'y)'(X'X)^{-1}(X'X\tilde{\beta}_0 - X'y)}{\tilde{\sigma}_0^2} .$$

The LM test statistic is expressed fully in terms of <u>restricted</u> MLE's. However, for this problem we can re-arrange LM —

Recall:

$$\tilde{\beta}_0 = \tilde{\beta}_1 + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\tilde{\beta}_1)$$

where $\tilde{\beta}_1 = (X'X)^{-1}X'y$.

So,

$$X'X\tilde{\beta}_0 = X'X\tilde{\beta}_1 + (X'X)(X'X)^{-1}R' \cdot [R(X'X)^{-1}R']^{-1}(r - R\tilde{\beta}_1)$$

But $X'X\tilde{\beta}_1 = (X'X)(X'X)^{-1}X'y = X'y$.

So,

$$X'X\tilde{\beta}_0 = X'y + R'[R(X'X)^{-1}R']^{-1}(r - R\tilde{\beta}_1)$$

& so

$$LM = \frac{(r - R\tilde{\beta}_1)'[R(X'X)^{-1}R']^{-1}(r - R\tilde{\beta}_1)}{\tilde{\sigma}_0^2}$$

If H_0 is true, we can replace $\tilde{\sigma}_0^2$ with any other consistent estimator of σ^2, such as $s^2 = (y - X\tilde{\beta}_1)'(y - X\tilde{\beta}_1)/(n-k)$.

Then,

$$LM^* = \left(\frac{\tilde{\sigma}_0^2}{s^2}\right)\left(\frac{LM}{J}\right) = F$$

$$= \frac{(r - R\tilde{\beta}_1)'[R(X'X)^{-1}R']^{-1}(r - R\tilde{\beta}_1)/J}{s^2}$$

$$\sim F_{J,\,n-k} \quad \text{(in finite samples)}$$

So, for this problem, the LM Test is equivalent to the usual F-test.

<u>Summary</u>:

(i) $LRT, W, LM \xrightarrow{d} \chi^2_{(J)}$ for nested hypotheses.

(ii) Properties differ in finite samples.

(iii) Numerical values of LRT, W & LM test statistics generally differ.

(iv) Don't use Wald test with non-linear restrictions.

(v) Choice largely one of computational convenience.

Some further examples of all tests —

Example: Exponential dist'n.

$$p(y_i \mid \theta) = \tfrac{1}{\theta}\, e^{-y_i/\theta} \qquad ; y_i > 0$$

$(E(y_i) = \theta.)$

$$L(\theta \mid \underset{\sim}{y}) = \prod_i p(y_i \mid \theta) = \theta^{-n} \exp\left[-\tfrac{1}{\theta} \sum_i y_i\right]$$

$$\log L = -n \log\theta - \tfrac{1}{\theta} \sum_i y_i$$

$$(\partial \log L / \partial\theta) = -n/\theta + \tfrac{1}{\theta^2} \sum_i y_i = 0$$

$$\Rightarrow \tilde{\theta} = \tfrac{1}{n} \sum_i y_i = \bar{y}$$

$$(\partial^2 \log L / \partial\theta^2) = (n/\theta^2) - (2/\theta^3) \sum_i y_i$$

$$= n/\theta^2 \left[1 - 2\bar{y}/\theta\right]$$

$(< 0, \text{ if } \tilde{\theta} = \bar{y}.)$

So, $I(\theta) = -E\left[\frac{\partial^2 \log L}{\partial\theta^2}\right]$

$$= -n/\theta^2 \left[1 - \tfrac{2}{\theta} E(\bar{y})\right]$$

$$= -n/\theta^2 \left[1 - \frac{2\theta}{\theta}\right]$$

$$= n/\theta^2$$

$$IA(\theta) = \lim_{n\to\infty} \left(\tfrac{1}{n} I(\theta)\right) = 1/\theta^2$$

Consider — $H_0: \theta = 1$ vs. $H_1: \theta \neq 1$.

(a) $LRT = -2 \log \lambda$

$$= 2\left[\log \tilde{L}_1 - \log \tilde{L}_0\right]$$

$$= 2\left[-n \log \tilde{\theta}_1 - \tfrac{1}{\tilde{\theta}_1} \sum_i y_i + n \log(1) + (1) \sum_i y_i\right]$$

$$= 2n(\bar{y} - \log \bar{y} - 1)$$

$$\xrightarrow{d} \chi^2_{(1)}$$

(Does statistic make sense?)

(b) $W = (R\tilde{\theta}_1 - r)'\left[R\, I^*(\tilde{\theta}_1)^{-1} R'\right]^{-1} (R\tilde{\theta}_1 - r)$

The restriction is $\theta = 1$, so:

$R = r = 1$; $\tilde{\theta}_1 = \bar{y}$; $\tilde{\theta}_0 = 1$.

So, $I^*(\tilde{\theta}_1) = (n/\hat{\theta}_1^2) = (n/\bar{y}^2)$

($\text{plim}(\tfrac{1}{n} I^*) = \text{plim}(1/\bar{y}^2) = 1/\theta^2 = IA$.

And: $W = (\bar{y} - 1)\left[1 \cdot (n/\bar{y}^2)^{-1} \cdot 1\right]^{-1} (\bar{y} - 1)$

$$= n\left[(\bar{y} - 1)/\bar{y}\right]^2 \xrightarrow{d} \chi^2_{(1)}$$

(c) $LM = D\log L_1(\tilde{\theta}_0)' I^*(\tilde{\theta}_0)^{-1} D\log L_1(\tilde{\theta}_0)$

$= (-n/1 + 1 \cdot \sum_i y_i)(1/n)(-n/1 + 1 \cdot \sum_i y_i)$

$= n[(\bar{y}-1)]^2$

$\xrightarrow{d} \chi^2_{(1)}$.

In each case, reject H_0 if test statistic exceeds critical $\chi^2_{(1)}$ value for chosen test size.

Example: (Binomial dist'n.)

$L = {}^nC_x\, p^x (1-p)^{n-x}$

$\log L = \text{const.} + x\log p + (n-x)\log(1-p)$

$(\partial \log L/\partial p) = x/p - \frac{n-x}{1-p}$

$\Rightarrow \tilde{p} = x/n$.

$(\partial^2 \log L/\partial p^2) = -x/p^2 - \frac{n-x}{(1-p)^2}$

Consider: $H_0: p = 1/2$ vs $H_A: p \neq 1/2$

(a) $LRT = -2\log[\tilde{L}_0/\tilde{L}_1]$

$= 2[\log \tilde{L}_1 - \log \tilde{L}_0]$

so, $LRT = 2[\text{const.} + x\log(x/n) + (n-x)\log(1-x/n) - \text{const.} - x\log(1/2) - (n-x)\log(1/2)]$

$= 2[x\log(x/n) + (n-x)\log(1-x/n) - n\log(1/2)]$

$\xrightarrow{d} \chi^2_{(1)}$; if H_0 is true.

(b) Wald Test:

$W = (\tilde{p} - 1/2)\, I^*(\tilde{p})\, (\tilde{p} - 1/2)$

& $I = -E[\partial^2 \log L/\partial p^2]$

$= -\left[-\frac{E(x)}{p^2} - \frac{n - E(x)}{(1-p)^2}\right]$

$= \left[\left(\frac{np}{p^2}\right) + \frac{n(1-p)}{(1-p)^2}\right] = \left[n/p + \frac{n}{1-p}\right]$

$IA = \lim_{n\to\infty}[1/n\, I] = [1/p + \frac{1}{1-p}]$

$= 1/[p(1-p)]$

& So $I^* = \frac{n}{p(1-p)}$ $\quad$ ($\text{Lim.}[\frac{1}{n} I^*] = IA$)

So, $W = n(\tilde{p} - \tfrac{1}{2})^2 / [\tilde{p}(1-\tilde{p})] \xrightarrow{d} \chi^2_{(1)}$

where $\tilde{p} = x/n$.

(c) $LM = D\log L_1(\tilde{p}_0)' \, I^*(\tilde{p}_0)^{-1} \, D\log L_1(\tilde{p}_0)$

$D\log L_1(\tilde{p}_0) = \left(\frac{x}{1/2}\right) - \left(\frac{n-x}{1-1/2}\right)$

$= 2(x-(n-x)) = 2(2x-n)$

& $I^*(\tilde{p}_0) = n/[\tfrac{1}{2}(1-\tfrac{1}{2})] = 4n$

So, $LM = [2(2x-n)]^2/4n$

$= \left[\frac{(2x-n)^2}{n}\right] \xrightarrow{d} \chi^2_{(1)}$.

Suppose that $n = 100$ & $x = 40$.

Then: $\left\{ \begin{array}{l} LRT = 4.027 \\ W = 4.167 \\ LM = 4.000 \end{array} \right.$

(5% critical value for $\chi^2_{(1)} = 3.8415$)

In each case, just reject H_0.