Lecture 4

MLE

(revised, 2010)

Read
- Wooldridge ch 13
- Green ch 17, 18
- Verbeek ch 6

ec 671

Lee

Maximum Likelihood Estimation (MLE)

MLE of θ is obtained by maximizing the likelihood functi[on]

$$\hat{\theta} = \arg\max \mathcal{L}(y; x; \theta)$$

Likelihood function is a joint probability function of the sample data.

$$\mathcal{L} = f(y_1, y_2, \ldots, y_N)$$
$$= f(y_1) f(y_2) \cdots f(y_N) \quad \text{if random samples are independent}$$
$$= \prod_{i=1}^{N} f(y_i)$$

Note when we use $f(y_i|X_i)$, it's conditional MLE (conditional on X_i), as we do not consider the joint dist of (y_i and X_i)

$$\mathcal{L} = \prod_{i=1}^{N} f(y_i | X_i; \theta)$$

Note We take log

$$\log \mathcal{L} = \log \prod_{i=1}^{N} f(y_i) = \sum_{i=1}^{N} \log f(y_i)$$

$\hat{\theta}$ is equivalently obtained from $\log \mathcal{L}$.

why take log?

i) Easy to handle ($\sum_{i=1}^{N} \log f(y_i)$ vs $\prod_{i=1}^{N} f(y_i)$)

ii) $\mathcal{L}$ can be a very small value when N is big; $(\text{prob})^{1000} \simeq 0.00000\ldots$

We maximize $\mathcal{L}$ or $\log\mathcal{L}$ to obtain $\hat{\theta}$.

$$\frac{\partial \log \mathcal{L}}{\partial \theta} = 0 \quad \Rightarrow \quad \hat{\theta} = \ldots \text{(solution)}$$

i) We may obtain closed from solutions for some special cases (linear models..)

ii) In general, we use "numerical optimizations" using computers; it's enough to derive $\log\mathcal{L}$.

To define the likelihood function, we need

i) correctly specified model

ii) correctly specified density function, $f(y_i|x_i)$

then, MLE is asymptotically the most efficient.

MLE is an M-estimator (see lecture 8.)

Ex 1) $y_i \sim N(\mu, \sigma^2)$ independent sample.

Find the MLE of μ and σ^2.

$\Rightarrow$ that is, $\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$. Find $\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix}$.

$$f(y_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2\right] \quad \text{using normal density}$$

$$\mathcal{L} = \prod_{i=1}^{N} f(y_i) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2\right] = \prod_{i=1}^{N} \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} \exp[\cdots]$$

$$* \left(\log\mathcal{L} = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i-\mu}{\sigma}\right)^2 \right. \quad \text{since } \log e^x = x$$

then, we can find $\hat{\theta}$ by numerical optimizations using $\log\mathcal{L}$.

In this example, we can find closed form solutions, by solving $\frac{\partial \log L}{\partial \theta} = 0$; score of the log likelihood

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2} \Sigma (y_i - \mu) \Rightarrow 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \hat{\mu}) = 0 \quad \Rightarrow \quad \Sigma y_i - n\hat{\mu} = 0$$

$$\Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n} y_i}{N} = \bar{Y} \text{ ; same as sample mean!} \quad (*)$$

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - \mu)^2 \Rightarrow 0$$

$$\Rightarrow -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^{n} (y_i - \hat{\mu})^2 = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad (**)$$

Note $\hat{\sigma}^2$ is biased; $E(\hat{\sigma}^2) \neq \sigma^2$

$s^2 = \frac{1}{n-1} \Sigma (y_i - \bar{y})^2$ is unbiased: $E(s^2) = \sigma^2$.

Thus

$$E(\hat{\sigma}^2) = E\left[\frac{n-1}{n} \overbrace{\frac{1}{n-1} \Sigma (y_i - \bar{y})^2}^{s^2}\right]$$

$$= \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

But, $\hat{\sigma}^2$ is more efficient than s^2.

Thus (*) and (**) give the MLE of θ

$$\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix}$$

Note One may check the S.O.C. to see $\log \mathcal{L}$ is concave. $\frac{\partial^2 \log L}{\partial \theta^2} < 0$ i.e. $\begin{cases} \frac{\partial^2 \log L}{\partial \mu^2} < 0 \\ \frac{\partial^2 \log L}{\partial \sigma^2} < 0 \end{cases}$

Note In this particular example

$$\frac{\partial^2 \log L}{\partial \mu \partial \sigma^2} = 0$$

($\hat{\mu}$ and $\hat{\sigma}^2$ are independent!)

ex2) $y_i = \alpha + \beta X_i + e_i$, $e_i \sim N(0, \sigma^2)$

Find $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2)'$.

$f(y_i) = f(e_i)$ with e_i being replaced with
$e_i = y_i - \alpha - \beta X_i$.

since $\left|\frac{de_i}{dy_i}\right| = 1$. Note $\left[\begin{array}{l} f(y_i) = f(x_i) \left|\frac{dx_i}{dy_i}\right| \\ \text{function of r.v.} \end{array} \right.$

$$f(y_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y_i - \alpha - \beta X_i}{\sigma}\right)^2\right]$$

$$\mathcal{L} = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y_i - \alpha - \beta X_i}{\sigma}\right)^2\right]$$

$$\log \mathcal{L} = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \alpha - \beta X_i}{\sigma}\right)^2$$

thus, we max the $\log \mathcal{L}$ to obtain $\hat{\theta}$.

Note This is an example for which closed form solutions exist.

$$\frac{\partial \log L}{\partial \hat{\alpha}} = \frac{1}{\hat{\sigma}^2} \Sigma (y_i - \hat{\alpha} - \hat{\beta} X_i) = 0$$

$$\frac{\partial \log L}{\partial \hat{\beta}} = \frac{1}{\hat{\sigma}^2} \Sigma (y_i - \hat{\alpha} - \hat{\beta} X_i) X_i = 0$$

$$\frac{\partial \log L}{\partial \hat{\sigma}^2} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \Sigma (y_i - \hat{\alpha} - \hat{\beta} X_i)^2 = 0$$

from these, we obtain (try!)

$$\left.\begin{aligned}\hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \\ \hat{\beta} &= \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2}\end{aligned}\right\}\quad \text{same as OLS!}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

($\hat{\sigma}^2$ is again biased.)

Ex3) $f(x_i) = \dfrac{e^{-\lambda}\lambda^{x_i}}{x_i!}$ Poisson dist.

$$\mathcal{L} = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$\lg\mathcal{L} = \sum_{i=1}^{n} \lg\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right)$$

$$= \sum_{i=1}^{n}\Big(\underbrace{\lg e^{-\lambda}}_{-\lambda} + \underbrace{\lg\lambda^{x_i}}_{x_i\lg\lambda} - \lg x_i!\Big)$$

$$= -n\lambda + (\lg\lambda)\sum_{i=1}^{n}x_i - \sum_{i=1}^{n}(\lg x_i!)$$

$$\simeq -n\lambda + (\lg\lambda)\sum_{i=1}^{n}x_i \quad \text{discarding the constant term}$$

Also, there is a closed form solution

$$\frac{\partial\lg\mathcal{L}}{\partial\lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n}x_i = 0$$

$$\Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x}\ ; \quad \text{sample mean}$$

g) If we have 10 obs : 5, 0, 1, 1, 0, 3, 2, 3, 4, 1

$$\hat{\lambda} = \frac{1}{n}\Sigma x_i = 2$$

$$\lg\mathcal{L}^* = -10(2) + (\lg 2)(\underbrace{20}_{\Sigma x_i}) - 12.242 = -26.22$$

Ex 4) Gamma dist.

$$f(y_i, X_i : \beta, \alpha) = \frac{(\beta + X_i)^{-\alpha}}{\Gamma(\alpha)} \, y_i^{\alpha-1} \, e^{-y_i/(\beta+X_i)}$$

<u>Note</u> $f(x) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$: gamma density

$$\mathcal{L} = \prod_{i=1}^{n} \frac{(\beta+X_i)^{-\alpha}}{\Gamma(\alpha)} y_i^{\alpha-1} e^{-y_i/(\beta+X_i)}$$

$$\log \mathcal{L} = \sum_{i=1}^{n} \frac{(\beta+X_i)^{-\alpha}}{\Gamma(\alpha)} y_i^{\alpha-1} e^{-y_i/(\beta+X_i)}$$

$$= -\alpha \sum_i \log(\beta+X_i) - n \log \Gamma(\alpha) + (\alpha-1)\sum_i \log y_i$$

$$- \sum_i \frac{y_i}{\beta+X_i}$$

$\theta = (\alpha, \beta)'$

$$\frac{\partial \log \mathcal{L}}{\partial \alpha} = -\sum_i \log(\beta+X_i) - n\,\psi(\alpha) + \sum_i \log y_i \Rightarrow 0$$

where $\psi(\alpha) = \frac{d \log \Gamma(\alpha)}{d\alpha}$

$$\frac{\partial \log \mathcal{L}}{\partial \beta} = -\alpha \sum_i \left(\frac{1}{\beta+X_i}\right) + \sum_i \frac{y_i}{(\beta+X_i)^2} \Rightarrow 0$$

From these, it seems that it's hard to obtain closed form solutions for $\hat{\alpha}$ and $\hat{\beta}$.
($\hat{\alpha}$ is a function of $\hat{\beta}$, and vice versa)

Thus, it's sufficient to define $\log \mathcal{L}$ and use numerical optimization

Greene p158
Table 4.2
"Lindop"

7

```
/*==================================================
Section 4.9.4.  Example of Various Test Procedures.
*/==================================================
Read ; Nobs = 20 ; Nvar = 3 ; Names = I,Y,E$
     1          20.5      12
     2          31.5      16
     3          47.7      18
     4          26.2      16
     5          44.0      12
     6          8.28      12
     7          30.8      16
     8          17.2      12
     9          19.9      10
    10          9.96      12
    11          55.8      16
    12          25.2      20
    13          29.0      12
    14          85.5      16
    15          15.1      10
    16          28.5      18
    17          21.4      16
    18          17.7      20
    19          6.42      12
    20          84.9      16
Sample;1-20$
?
? Just change name to be consistent with text
?
Create;x=e$
?
? Unrestricted maximum likelihood estimation.
?
Maxize ; fcn = -r*log(beta+x)-log(gma(r))-y/(beta+x)+(r-1)*log(y)      ⟸ logL
       ; start=-5,1
       ; labels=beta,r$
/*
              +----------------------------------------------+
              | User Defined Optimization                    |
              | Maximum Likelihood Estimates                 |
              | Dependent variable               Function    |
              | Weighting variable                    ONE    |
              | Number of observations                 20    |
              | Iterations completed                    4    |
              | Log likelihood function         -82.91605    |
              +----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 BETA     -4.718503621      3.6568024       -1.290   .1969
 R         3.150896345      1.2398481        2.541   .0110
*/
```

7-1

```
* refer to table 4.2, page 158 of Greene's Econometric Analysis, edition 4
* the corresponding table of edition 5 is table 17.1 on page 491
* first import data

use http://fmwww.bc.edu/ec-p/data/Greene2000/TBL4-1, clear

* note that the variable e is actually the independent variable x
* assume the conditional distribution is gamma
* the coefficient of e is restricted to be one.
* this is a two-paremeter restricted MLE model


capture program drop mle_gamma
program define mle_gamma
version 8.0
args lnf theta1 theta2
quietly replace `lnf'=-`theta2'*ln(`theta1')-lngamma(`theta2')+(`theta2'-1)*ln($ML_y1)-$ML_y1/`theta1'
end

constraint define 1 e=1
ml model lf mle_gamma (y=e) (), constraints(1)
ml maximize


/*
. capture program drop mle_gamma

. program define mle_gamma
  1. version 8.0
  2. args lnf theta1 theta2
  3. quietly replace `lnf'=-`theta2'*ln(`theta1')-lngamma(`theta2')+(`theta2'-1)*ln($ML_y1)-$ML_y1/`theta1'
  4. end

.
. constraint define 1 e=1

. ml model lf mle_gamma (y=e) (), constraints(1)

. ml maximize

initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood = -1287.8497
rescale:       log likelihood = -109.84554
rescale eq:    log likelihood = -87.102313
Iteration 0:   log likelihood = -118.44291  (not concave)
Iteration 1:   log likelihood = -86.296196
Iteration 2:   log likelihood = -83.735966
```

log L

7-2


```
Iteration 3:   log likelihood = -83.088558
Iteration 4:   log likelihood = -82.924748
Iteration 5:   log likelihood = -82.916083
Iteration 6:   log likelihood = -82.916049
Iteration 7:   log likelihood = -82.916049

                                                  Number of obs   =         20
                                                  Wald chi2(0)    =          .
Log likelihood = -82.916049                       Prob > chi2     =          .

 ( 1)  [eq1]e = 1
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
           e |          1          .        .       .            .           .
       _cons |  -4.718502   2.345027    -2.01   0.044     -9.31467   -.1223335
-------------+----------------------------------------------------------------
eq2          |
       _cons |   3.150896   .7942619     3.97   0.000     1.594171    4.707621
------------------------------------------------------------------------------

.
end of do-file

*/
```

Ex5) Bernoulli dist

$$p = Prob(y_i = 1), \quad 1-p = Prob(y_i = 0)$$

$$\Rightarrow \quad p(y_i) = p^{y_i}(1-p)^{1-y_i}, \quad y_i = 0, 1$$

$$\max_p \mathcal{L} = \prod_{i=1}^{n} p(y_i) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

$$\ln_p \mathcal{L} = \sum_{i=1}^{n} \left[y_i \ln p + (1-y_i)\ln(1-p) \right]$$

$$\frac{\partial \ln L}{\partial p} = \sum_{i=1}^{n} \left(\frac{y_i}{p} - \frac{1-y_i}{1-p} \right) \Rightarrow 0$$

$$\Rightarrow \quad \hat{p} = \frac{1}{n}\sum_{i}^{n} y_i \quad = \text{proportion of } (y_i = 1).$$

Ex6) Probit model

$$p = Prob(y_i = 1), \quad 1-p = Prob(y_i = 0)$$

$$y_i = \begin{cases} 1 & \text{if } y_i^* = X_i\beta + e_i > 0 \\ 0 & \text{if } y_i^* = X_i\beta + e_i \le 0 \end{cases}$$

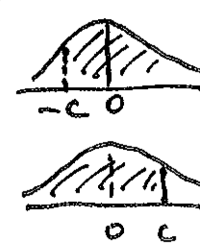
thus

$$\begin{aligned} p(y_i = 1) &= p(y_i^* > 0) = p(X_i\beta + e_i > 0) \\ &= p(e_i > -X_i\beta) \\ &= p(e_i < X_i\beta) \\ &= \Phi(X_i\beta) \quad \text{cdf of std. normal} \end{aligned}$$

$P(z > -c)$

$\|$

$P(z < c)$



$$p(y_i = 0) = 1 - \Phi(X_i\beta)$$

thus

$$\max_\beta \mathcal{L} = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i} = \prod_{i=1}^{n} \Phi(X_i\beta)^{y_i}\left[1 - \Phi(X_i\beta)\right]^{1-y_i}$$

… Now, max over β (instead of p)

$$\max_{\beta} \log L = \sum_{i=1}^{n} \left[y_i \cdot \log \Phi(X_i\beta) + (1-y_i)\cdot \log[1-\Phi(X_i\beta)] \right]$$

$$\frac{\partial \log L}{\partial \beta} = y_i \cdot \frac{\phi(X_i\beta)}{\Phi(X_i\beta)} \cdot X_i - (1-y_i)\cdot \frac{\phi(X_i\beta)}{1-\Phi(X_i\beta)} \cdot X_i \Rightarrow 0$$

where $\phi(X_i\beta)$ is the pdf of the std. normal dist.
($\Phi' = \phi$: differentiation of cdf = pdf)
$F(x)' = f(x)$

$$\frac{\partial \log L}{\partial \beta} = \frac{\phi(X_i\beta)\, X_i'(y_i - \Phi(X_i\beta))}{\Phi(X_i\beta)(1-\Phi(X_i\beta))}$$ after some algebra

Ex 7) Poisson regression

$$f(y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad \text{where } \lambda_i = \exp(X_i\beta), \; y = 0, 1, 2, \cdots$$

$(\lambda \to \beta)$

That is, $E(y_i | X_i) = \exp(X_i\beta)$.

Note this is a log-linear model

$$y_i = \exp(X_i\beta) \rightarrow \log y_i = X_i\beta$$

As in Ex 3)

$$\log \mathcal{L} = \sum_{i=1}^{n} \log \left(\frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \right)$$

$$= \sum_{i=1}^{n} \left(\log e^{-\lambda_i} + \log \lambda_i^{y_i} - \log y_i! \right)$$

$$= \sum_{i=1}^{n} \left[-\lambda_i + y_i \log \lambda_i - \log y_i! \right)$$

$$= \sum_{i=1}^{n} \left(-e^{X_i\beta} + y_i \underbrace{\log e^{X_i\beta}}_{= X_i\beta} - \underbrace{\log y_i!}_{\text{can be ignored}} \right)$$

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \left[-X_i e^{X_i\beta} + y_i \cdot X_i \right] \Rightarrow 0$$

"no closed form solution" for β

```
* exercise 5.3 of Greene's textbook, edition 4, page 208-209
* input the data

use 53_poisson.dta, clear

capture program drop mle_poisson

program define mle_poisson
version 8.0
args lnf theta1
quietly replace `lnf'=-exp(`theta1')+$ML_y1*(`theta1')-lngamma($ML_y1+1)
end

ml model lf mle_poisson (y=x)
ml maximize
```

using $n! = \Gamma(n+1)$, $\log(y_i!) = \log\Gamma(y_i+1)$

```
capture program drop mle_poisson1

program define mle_poisson1
version 8.0
args lnf theta1
quietly replace `lnf'=-exp(`theta1')+$ML_y1*(`theta1')
end
```

(w/o $\log y_i!$)

```
ml model lf mle_poisson1 (y=x)
ml maximize

/*

y       x
6       1.5
7       1.8
4       1.8
10      2
10      1.3
6       1.6
4       1.2
7       1.9
2       1.8
3       1
6       1.4
5       .5
3       .8
3       1.1
4       .7
```

```
. do "C:\Documents and Settings\jlee\My Documents\EC671\53_poisson.do"

.
. * exercise 5.3 of Greene's textbook, edition 4, page 208-209
. * input the data
.
. use 53_poisson.dta, clear

.
. capture program drop mle_poisson

. program define mle_poisson
  1. version 8.0
  2. args lnf theta1
  3. quietly replace `lnf'=-exp(`theta1')+$ML_y1*(`theta1')-lngamma($ML_y1+1)
  4. end

.
. ml model lf mle_poisson (y=x)

. ml maximize

initial:       log likelihood = -102.38698
alternative:   log likelihood = -72.117799
rescale:       log likelihood = -38.222822
Iteration 0:   log likelihood = -38.222822
Iteration 1:   log likelihood = -32.137797
Iteration 2:   log likelihood = -32.066088
Iteration 3:   log likelihood = -32.066011
Iteration 4:   log likelihood = -32.066011

                                                  Number of obs   =         15
                                                  Wald chi2(1)    =       2.72
Log likelihood = -32.066011                       Prob > chi2     =     0.0993

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .4248897   .2578109     1.65   0.099    -.0804104    .9301898
       _cons |   1.078078   .3886433     2.77   0.006     .3163506    1.839805
------------------------------------------------------------------------------

. capture program drop mle_poisson1
```

```
.
. program define mle_poisson1
  1. version 8.0
  2. args lnf theta1
  3. quietly replace `lnf'=-exp(`theta1')+$ML_y1*(`theta1')
  4. end

.
. ml model lf mle_poisson1 (y=x)

. ml maximize

initial:       log likelihood =        -15
alternative:   log likelihood =  15.269181
rescale:       log likelihood =  49.164159
Iteration 0:   log likelihood =  49.164159
Iteration 1:   log likelihood =  55.254407
Iteration 2:   log likelihood =  55.320936
Iteration 3:   log likelihood =   55.32097
Iteration 4:   log likelihood =   55.32097

                                                  Number of obs   =         15
                                                  Wald chi2(1)    =       2.72
Log likelihood =   55.32097                       Prob > chi2     =     0.0993

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .4248897   .2578109     1.65   0.099    -.0804104    .9301898
       _cons |   1.078078   .3886433     2.77   0.006     .3163506    1.839805
------------------------------------------------------------------------------

.
end of do-file

*/
```

Properties of MLE ($\hat{\theta}$)

Under general conditions, the MLE is

(i) consistent

(ii) asymptotically efficient

(iii) asymptotically normal

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N\left[0, \lim_{n\to\infty}\left(\frac{1}{n}I\right)^{-1}\right]$$

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

; information matrix

i.e. $Var(\hat{\theta}) = I(\hat{\theta})^{-1}$

Result 1 $\quad E\dfrac{\partial \ln L(y,\theta)}{\partial\theta} = 0 \Rightarrow$ $\boxed{\text{Let } \frac{\partial \ln L(y,\theta)}{\partial\theta} = S(\theta) \text{ "score vector"}}$

proof) Let $L(y,\theta)$ = joint density of y

i.e. $\int L(y,\theta)\,dy = 1$

differentiate w.r.t. θ

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

$$\Rightarrow \int \frac{\partial \ln L(y,\theta)}{\partial\theta}\cdot L(y,\theta)\,dy = 0$$

since $\dfrac{\partial \ln L(y,\theta)}{\partial\theta} = \dfrac{1}{L(y,\theta)}\cdot\dfrac{\partial L(y,\theta)}{\partial\theta}$

$$\Rightarrow E\frac{\partial \ln L(y,\theta)}{\partial\theta} = 0$$

"On average, it is evaluated at true value of θ."

Note this is a regularity condition.

Result 2 $\quad Var\left(\frac{\partial \ln L}{\partial \theta}\right) \equiv E\left(\frac{\partial \ln L}{\partial \theta}\right)\left(\frac{\partial \ln L}{\partial \theta'}\right) = -E\left(\frac{\partial^2 \ln L}{\partial \theta \partial \theta'}\right)$

"(conditional) (same) information matrix equality"

proof) Result 1 indicates

$$E\frac{\partial \ln L}{\partial \theta} = 0 \quad \text{ie} \quad \int \frac{\partial \ln L}{\partial \theta} L\, dy = 0$$

(L is a density)

Differentiate $\int \frac{\partial \ln L}{\partial \theta} L\, dy$ wrt θ'.

$$\int \left[\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} L + \frac{\partial \ln L}{\partial \theta} \underbrace{\frac{\partial L}{\partial \theta'}}\right] dy = 0$$

Write $\frac{\partial L}{\partial \theta'} = \frac{\partial \ln L}{\partial \theta'} L$ (since $\frac{\partial \ln L}{\partial \theta'} = \frac{1}{L}\frac{\partial L}{\partial \theta'}$)

$$\Rightarrow \int \left(\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} + \frac{\partial \ln L}{\partial \theta}\frac{\partial \ln L}{\partial \theta'}\right) L\, dy = 0$$

$$\Rightarrow E\left[\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} + \frac{\partial \ln L}{\partial \theta}\frac{\partial \ln L}{\partial \theta'}\right] = 0$$

(like $\int x f(x)\,dx = E(x)$)

$$\Rightarrow -E\left(\frac{\partial^2 \ln L}{\partial \theta \partial \theta'}\right) = E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)\left(\frac{\partial \ln L}{\partial \theta'}\right)\right]$$

Note

i) $\frac{\partial^2 \ln L}{\partial \theta \partial \theta'}$ = Hessian matrix

ii) $-E[\text{Hessian}] = \mathcal{I}$ = information matrix

Result 3 Cramer-Rao Lower bound

$Var(\hat{\theta}) = \mathcal{I}(\hat{\theta})^{-1}$ achieves the least variance.

It's called the Cramer-Rao lower bound.

Let $\tilde{\theta} = g(y)$ be any other unbiased estimator.

"$Var(\tilde{\theta}) - \mathcal{I}(\hat{\theta})^{-1}$ is psd."

Result 4 MLE is consistent

$$\frac{1}{n}\ln L(y,\theta) = \frac{1}{n}\sum_n \ln f(y_i,\theta) \rightarrow E \ln f(y,\theta)$$

by law of large number

$$\Rightarrow \frac{\partial E \ln f(y,\theta)}{\partial \theta} = 0$$

Back to Ex1)

$$\frac{\partial^2 \ln L}{\partial \mu \partial \mu'} = -\frac{n}{\sigma^2}, \quad \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\Sigma(y_i-\mu)^2$$

$$\frac{\partial^2 \ln L}{\partial \mu \partial \sigma^2} = -\frac{1}{\sigma^4}\Sigma(y_i-\mu)$$: Expected value of this $= 0$

$$\mathcal{I}^{-1} = \left[-E\left(\frac{\partial^2 \ln L}{\partial \theta \partial \theta'}\right)\right]^{-1} = \left(-E\begin{bmatrix} \frac{-n}{\sigma^2} & \frac{1}{\sigma^4}\Sigma(y_i-\mu) \\ & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\Sigma(y_i-\mu)^2 \end{bmatrix}\right)^{-1} = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{+n}{2\sigma^4} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}$$

since $E\left(\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \sigma^2}\right) = E\left[\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\Sigma(y_i-\mu)^2\right]$

$$= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}(n\sigma^2) = \frac{-n}{2\sigma^4}$$

That is,

$$\left.\begin{aligned} Var(\hat{\mu}) &= \sigma^2/n \\ Var(\hat{\sigma}^2) &= \frac{2\sigma^4}{n} \end{aligned}\right\} \text{ these are C-R bound.}$$

$$Cov(\hat{\mu}, \hat{\sigma}^2) = 0$$

Back to Ex3)

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = \frac{-1}{\lambda^2}\Sigma X_i$$ Poisson

Hessian

$$\mathcal{I} = -E\frac{\partial^2 \ln L}{\partial \lambda^2} = -E\left[\frac{-1}{\lambda^2}\Sigma X_i\right]$$

$$= \frac{1}{\lambda^2} n\lambda = \frac{n}{\lambda}$$ since $E(X_i) = \lambda$ Poisson (also $Var(X_i) = \lambda$)

$\mathcal{I}^{-1} = \frac{\lambda}{n}$: this is CR.

$$\Rightarrow Var(\hat{\lambda}) = \frac{\lambda}{n} = \mathcal{I}^{-1}(\hat{\lambda})$$

Estimates of Asymptotic Variance

Let
$$A = -E\left[\frac{\partial^2 \ln L}{\partial\theta\partial\theta'}\right] = -E(\text{Hessian}) = \text{inf. matrix}$$

$$B = E\left[\frac{\partial \ln L}{\partial\theta}\frac{\partial \ln L}{\partial\theta'}\right] = E(S(\theta)\,S(\theta)')$$

where $S(\theta) = \frac{\partial \ln L}{\partial\theta}$ = score vector

$$C = \frac{\partial^2 \ln L}{\partial\theta\partial\theta'} = \text{Hessian}$$

Estimates of variances (any of these: A^{-1} is preferred)

$$\hat{A}^{-1} = \left[\sum_{i=1}^{n}\left(-E(H(y_i, x_i, \hat{\theta}))\right)\right]^{-1} \text{ using inf. matrix}$$

$$\hat{B}^{-1} = \left[\sum_{i=1}^{n} s_i(\hat{\theta})\, s_i(\hat{\theta})'\right]^{-1} \text{ using score vectors}$$

$$\hat{C}^{-1} = \left[-\sum_{i=1}^{n}\frac{\partial^2 \ln L(y_i, x_i, \hat{\theta})}{\partial\hat{\theta}\partial\hat{\theta}'}\right]^{-1} \text{ using Hessian matrix}$$

<u>Note</u>

- $\hat{A}^{-1}$ is preferred, but it may be troublesome to evaluate expectations.
- $\hat{B}^{-1}$ is p.s.d., but may behave poorly in small samples (BHHH uses this.)
- $\hat{C}^{-1}$ is easy to use, but the inverse may not exist.

Linearized MLE

Let $\hat{\theta}$ be any consistent estimator of θ.

Define the linearized MLE as

$$\tilde{\theta} = \hat{\theta} - \left(\frac{\partial^2 \ln L}{\partial\theta\,\partial\theta'}\right)^{-1}\left(\frac{\partial \ln L}{\partial\theta}\right)$$

where derivatives are evaluated at $\theta = \hat{\theta}$.

then $\tilde{\theta}$ is consistent and asymptotically efficient.

<u>Note</u> this is a mean value theorem. (Taylor expansion)

$$f(x) = f(a) + f'(b)(x-a)$$

where $a < b < x$

thus,

$$\left.\frac{\partial \ln L}{\partial\theta}\right|_{\theta=\tilde{\theta}} = \left.\frac{\partial \ln L}{\partial\theta}\right|_{\theta=\hat{\theta}} + \frac{\partial^2 \ln L}{\partial\theta\,\partial\theta'}(\tilde{\theta}-\hat{\theta}) = 0$$

(FOC = 0)

$$\Rightarrow \tilde{\theta} - \hat{\theta} = -\left(\frac{\partial^2 \ln L}{\partial\theta\,\partial\theta'}\right)^{-1}\left(\left.\frac{\partial \ln L}{\partial\theta}\right|_{\theta=\hat{\theta}}\right)$$

<u>Note</u> this idea is extended in numerical optimizations.

① Newton-Raphson scheme

$$\tilde{\theta} = \hat{\theta} - \left(\frac{\partial^2 \ln L}{\partial\theta\,\partial\theta'}\right)^{-1}\left(\frac{\partial \ln L}{\partial\theta}\right)$$ using the Hessian matrix

② Method of scoring

$$\tilde{\theta} = \hat{\theta} - \left[-I(\hat{\theta})\right]^{-1}\left(\frac{\partial \ln L}{\partial\theta}\right)$$ using inf. matrix

(simpler after taking expectations of Hessian)

Hypothesis Testing

1) LR (Likelihood Ratio) test

$$LR = 2[\ln L(\hat{\theta}) - \ln L(\tilde{\theta})] \sim \chi^2_m$$

2) Wald test

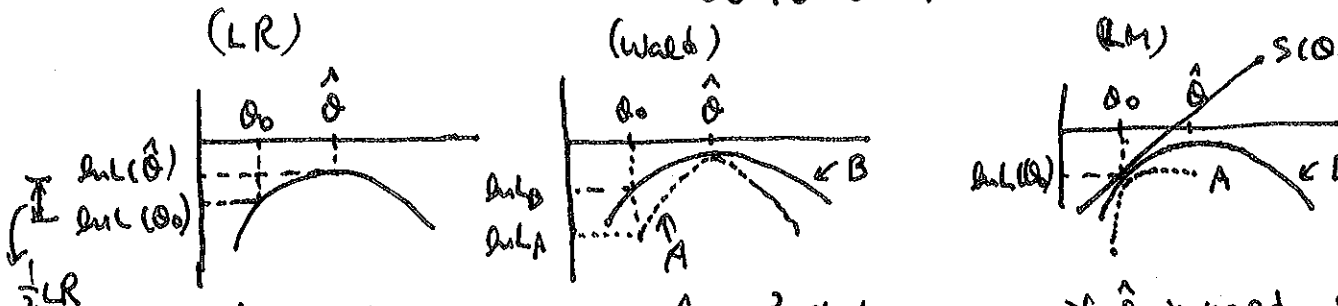
$$Wald = (R\hat{\theta} - r)'[R\,var(\hat{\theta})R']^{-1}(R\hat{\theta} - r) \sim \chi^2_m$$

where $var(\hat{\theta}) = [-H(\hat{\theta})]^{-1}$ for instance (stata uses this.); $\hat{A}^{-1}$, $\hat{B}^{-1}$ or $\hat{C}^{-1}$ can be used.

3) LM test

$$LM = s(\tilde{\theta})' I(\tilde{\theta})^{-1} s(\tilde{\theta}) \sim \chi^2_m$$

where $s(\tilde{\theta}) = \frac{\partial \ln L}{\partial \theta}\Big|_{\theta = \tilde{\theta}}$, $I(\tilde{\theta}) = -E\left[\frac{\partial^2 \ln L}{\partial \theta \partial \theta'}\right]\Big|_{\theta=\tilde{\theta}}$



As $(\hat{\theta} - \theta_0)$ gets larger $\frac{1}{2}LR$ gets larger.

When $(\hat{\theta} - \theta_0)^2$ gets larger, so is $\ln L(\hat{\theta}) - \ln L(\theta_0)$, which depends on the curvature

If $\hat{\theta}$ is used, $s(\hat{\theta}) = 0$. When $\tilde{\theta}$ is used, we wish to have $s(\tilde{\theta})$ as close to zero. Curvature is also used.

- LR : both $\hat{\theta}$ and $\tilde{\theta}$
- Wald : $\hat{\theta}$ only
- LM : $\tilde{\theta}$ only

Ex) $y_1, y_2, \cdots y_n$: random sample from $N(\mu, \sigma^2)$.

$H_0: \mu = 0$. Assume σ^2 is known.

$$\ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_i (y_i - \mu)^2$$

<u>LR test</u>

$$\ln L_u = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\Sigma(y_i - \bar{y})^2$$

since MLE of $\mu = \bar{y}$ (σ^2 is known.)

$$\ln L_R = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\Sigma(y_i - 0)^2$$

since $H_0: \mu = 0$ is imposed.

$$\underline{\underline{LR}} = 2(\ln L_u - \ln L_R) = 2\left[-\frac{1}{2\sigma^2}\Sigma(y_i - \bar{y})^2 + \frac{1}{2\sigma^2}\Sigma y_i^2\right]$$

$$= \frac{n\bar{y}}{\sigma^2} = \left(\frac{\bar{y}}{\sigma/\sqrt{n}}\right)^2 \sim \chi_1^2$$

<u>Wald test</u>

$g(\theta) = \mu$. $g(\hat{\theta}) = \hat{\mu} - 0 = \bar{y}$

$var(\hat{\mu}) = \frac{\sigma^2}{n}$

$$\underline{\underline{\text{Wald}}} = g(\hat{\theta})' \, var(\hat{\mu})^{-1} g(\hat{\theta})$$

$$= \bar{y}'\left(\frac{\sigma^2}{n}\right)^{-1}\bar{y} = \left(\frac{\bar{y}}{\sigma/\sqrt{n}}\right)^2 \sim \chi_1^2$$

<u>LM test</u> (Score test)

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\Sigma(y_i - \mu) \Rightarrow \left.\frac{\partial \ln L}{\partial \mu}\right|_{\theta=\theta_0} = \frac{1}{\sigma^2}\Sigma(y_i - 0)$$

$$= n\bar{y}/\sigma^2$$

$$\tilde{\mathcal{I}} = \frac{n}{\sigma^2} \Leftarrow \mathcal{I}|_{\theta=\tilde{\theta}}$$

$$\underline{\underline{LM}} = \left(\frac{n\bar{y}}{\sigma^2}\right)'\left(\frac{\sigma^2}{n}\right)\left(\frac{n\bar{y}}{\sigma^2}\right) = \left(\frac{\bar{y}}{\sigma/\sqrt{n}}\right)^2 \sim \chi_1^2$$

MLE exercise

1. (Greene, Ex 4.17 and 4.18, p. 162) Suppose that x has the Weibull distribution,

$$f(x) = \alpha\beta x^{\beta-1}\exp(-\alpha x^{\beta}),\ \ x \ge 0,\ \alpha,\ \beta > 0.$$

(i) Obtain the log-likelihood function for a random sample of n observations.

(ii) Find the score vectors and the Hessian matrix.

(iii) Modifying the example STATA program, obtain the MLE of α and β for the following data:

1.0343 .49254 1.2742 1.4019 .32556 .29965 .26423
1.0878 1.9461 .47615 3.6454 .15344 1.2357 .96381
.33453 1.1227 2.0296 1.2797 .96080 2.0070

2. Verbeek, Ex 6.1, p. 186 (a) – (g). (next page)

3. Find (i) the score vectors, (ii) the Hessian matrix, and (iii) Information matrix (also, variance of the MLE and CR bound) of each of the followings.

(a) Ex 2: regression, $y_i = \alpha + \beta X_1 + e_i$ .. page 4 of this lecture note

(b) Ex 3: Poisson distribution .. page 5

(c) Ex 5, Bernoulli distribution (note: $E(y_i) = p$) .. page 8

4. Find the LR, Wald and LM statistics testing the hypothesis, $\lambda = 3$, for a random sample of $x_1,..,x_{10}$ (given at the bottom of page 5, where $\Sigma x_i = 20$ and $\hat{\lambda} = 2$ (MLE estimate)), when the density function is given as

$$f(x_i) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}.$$

18

Exercises (Verbeek)

Exercise 6.1 (The Normal Linear Regression Model)

Consider the following linear regression model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i,$$

where $\beta = (\beta_1, \beta_2)'$ is a vector of unknown parameters, and x_i is a one-dimensional observable variable. We have a sample of $i = 1, \ldots, N$ independent observations and assume that the error terms ε_i are $NID(0, \sigma^2)$, independent of all x_i. The density function of y_i (for a given x_i) is then given by

$$f(y_i|\beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\sigma^2}\right\}.$$

a. Give an expression for the loglikelihood contribution of observation i, $\log L_i(\beta, \sigma^2)$. Explain why the loglikelihood function of the entire sample is given by

$$\log L(\beta, \sigma^2) = \sum_{i=1}^{N} \log L_i(\beta, \sigma^2).$$

b. Determine expressions for the two elements in $\partial \log L_i(\beta, \sigma^2)/\partial\beta$ and show that both have expectation zero for the true parameter values.

c. Derive an expression for $\partial \log L_i(\beta, \sigma^2)/\partial\sigma^2$ and show that it also has expectation zero for the true parameter values.

Suppose that x_i is a dummy variable equal to 1 for males and 0 for females, such that $x_i = 1$ for $i = 1, \ldots, N_1$ (the first N_1 observations) and $x_i = 0$ for $i = N_1 + 1, \ldots, N$.

d. Derive the first order conditions for maximum likelihood. Show that the maximum likelihood estimators for β are given by

$$\hat\beta_1 = \frac{1}{N - N_1} \sum_{i=N_1+1}^{N} y_i, \quad \hat\beta_2 = \frac{1}{N_1} \sum_{i=1}^{N_1} y_i - \hat\beta_1.$$

What is the interpretation of these two estimators? What is the interpretation of the true parameter values β_1 and β_2?

e. Show that

$$\partial^2 \log L_i(\beta, \sigma^2)/\partial\beta\partial\sigma^2 = \partial^2 \log L_i(\beta, \sigma^2)/\partial\sigma^2\partial\beta,$$

and show that it has expectation zero. What are the implications of this for the asymptotic covariance matrix of the ML estimator $(\hat\beta_1, \hat\beta_2, \hat\sigma^2)$?

f. Present two ways to estimate the asymptotic covariance matrix of $(\hat\beta_1, \hat\beta_2)'$ and compare the results.

g. Present an alternative way to estimate the asymptotic covariance matrix of $(\hat\beta_1, \hat\beta_2)'$ that allows ε_i to be heteroskedastic.

Note Delta method

Suppose $\hat{\theta}$ has a variance $Var(\hat{\theta}) = \Sigma$.
we wish to find $Var(g(\hat{\theta}))$ where $g(\hat{\theta})$ is a functn of $\hat{\theta}$.

$$Var(g(\hat{\theta})) = \bar{D}\,\Sigma\,\bar{D}'$$

where $\bar{D} = \text{plim}\,\dfrac{\partial g(\theta)}{\partial \theta'}$

eg) $var(s^2) = 2\sigma^4/n$. Find the variance of s.
$s = g(s^2) = \sqrt{s^2}$

let $g(\hat{\theta}) = \sqrt{s^2} = s$. $(\hat{\theta} = s^2,\ g(\hat{\theta}) = s)$

$$\frac{\partial g(\theta)}{\partial (s^2)} = \frac{\partial s}{\partial s^2} = \frac{1}{\partial s^2/\partial s} = \frac{1}{2s} \rightarrow \frac{1}{2\sigma}$$

Alternatively let $s^2 = T$. (plim $\bar{D} \rightarrow D$)

$$\left(\frac{\partial g}{\partial s^2} = \frac{\partial \sqrt{T}}{\partial T} = \frac{1}{2\sqrt{T}} = \frac{1}{2s} \right)$$

$$Var(s^2) = \bar{D}\,\Sigma\,\bar{D}' = \left(\frac{1}{2\sigma}\right)\left(\frac{2\sigma^4}{n}\right)\left(\frac{1}{2\sigma}\right) = \frac{\sigma^2}{2n}$$

eg) $y = X_1\beta_1 + X_2\beta_2 + e$. $H_0: \beta_1\beta_2 = 2$.

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

let $g(\hat{\theta}) = \hat{\beta}_1\hat{\beta}_2$. $\quad D = \dfrac{\partial g(\hat{\theta})}{\partial \hat{\theta}'} = \left(\dfrac{\partial g(\hat{\theta})}{\partial \hat{\beta}_1}, \dfrac{\partial g(\hat{\theta})}{\partial \hat{\beta}_2}\right)$

$= (\hat{\beta}_2, \hat{\beta}_1) \longrightarrow \text{plim}\, D = (\beta_2, \beta_1)$

$$Var(g(\hat{\theta})) = var(\hat{\beta}_1\hat{\beta}_2) = \bar{D}\,\Sigma\,\bar{D}'$$
$$= (\hat{\beta}_2, \hat{\beta}_1)\,\hat{\sigma}^2(X'X)^{-1}\begin{pmatrix}\hat{\beta}_2\\ \hat{\beta}_1\end{pmatrix} = (1\times 1)$$

eg) $H_0: \dfrac{\beta_1}{\beta_2} = 5 \quad g(\hat{\theta}) = \dfrac{\hat{\beta}_1}{\hat{\beta}_2}. \quad D = \left(\dfrac{1}{\hat{\beta}_2}, \dfrac{-\hat{\beta}_1}{\hat{\beta}_2^2}\right)$

$$var\left(\frac{\hat{\beta}_1}{\hat{\beta}_2}\right) = \left(\frac{1}{\hat{\beta}_2}, \frac{-\hat{\beta}_1}{\hat{\beta}_2^2}\right)\hat{\sigma}^2(X'X)^{-1}\begin{pmatrix}\dfrac{1}{\hat{\beta}_2}\\ \dfrac{-\hat{\beta}_1}{\hat{\beta}_2^2}\end{pmatrix} = (1\times 1)$$

eg) Find $\mathrm{Cov}\begin{pmatrix}\hat\beta_1\hat\beta_2\\ \hat\beta_1/\hat\beta_2\end{pmatrix}$.

$$g(\theta) = \begin{pmatrix}\hat\beta_1\hat\beta_2\\ \hat\beta_1/\hat\beta_2\end{pmatrix}. \quad D = \frac{\partial g(\theta)}{\partial\theta'} = \begin{pmatrix}\partial g_1(\theta)/\partial\theta'\\ \partial g_2(\theta)/\partial\theta'\end{pmatrix}$$

$$= \begin{bmatrix}\hat\beta_2, & \hat\beta_1\\ \frac{1}{\hat\beta_2}, & -\frac{\hat\beta_1}{\hat\beta_2^2}\end{bmatrix} \quad 2\times 2$$

$$\mathrm{cov}\begin{pmatrix}\hat\beta_1\hat\beta_2\\ \hat\beta_1/\hat\beta_2\end{pmatrix} = \begin{bmatrix}\hat\beta_2, & \hat\beta_1\\ \frac{1}{\hat\beta_2}, & -\frac{\hat\beta_1}{\hat\beta_2^2}\end{bmatrix}\hat\sigma^2(X'X)^{-1}\left(\begin{pmatrix}\hat\beta_2\\ \hat\beta_1\end{pmatrix}\begin{pmatrix}\frac{1}{\hat\beta_2}\\ -\frac{\hat\beta_1}{\hat\beta_2^2}\end{pmatrix}\right)$$

then, we can use Wald tests.

<u>Note</u> It is a puzzle that Wald tests are different for basically identical hypothesis (nonlinear)

eg) $H_0: \beta_1\beta_2 = 5$ or $H_0: \beta_1 = \frac{5}{\beta_2}$

<u>Exercise</u> Find the Wald stat. for the hypothesis

$H_0: \beta_1\beta_2 = 1, \quad \beta_2 + \beta_1\beta_3 = 0$ where $\mathrm{Var}(\hat\beta) = \sigma^2(X'X)^{-1}$.

<u>Note</u> Suppose $\sqrt{n}(\hat\theta - \theta) \to N(0, \Sigma)$

$H_0: g(\theta) = 0$

$$\text{Wald} = n\, g(\hat\theta)'\,\hat\Omega^{-1} g(\hat\theta) \sim \chi^2_m$$

where $\Omega = D\Sigma D'$, $D = \frac{\partial g(\theta)}{\partial\theta'}$

Alternative form of the LR test $\ni nR^2$

$$\ln L = \sum_i \ln f(y_i, \theta)$$

i) $$\frac{\partial \ln L}{\partial \theta} = \sum_i \frac{\partial \ln f(y_i,\theta)}{\partial \theta} \overset{\text{let}}{=} W' \underset{\sim}{i}$$

where $\underset{n\times K}{W}$ is defined by $W_{ij} = \dfrac{\partial \ln f(y_i,\theta)}{\partial \theta_j}$

$\underset{\sim}{i} = (1, \cdots, 1)'$: $n\times 1$

$$\Rightarrow \frac{\partial \ln L}{\partial \tilde{\theta}} = \tilde{W}' \underset{\sim}{i} \quad \text{which is } W \text{ evaluated at } \tilde{\theta}$$

ii) $$\mathcal{I} = -E\frac{\partial^2 \ln L}{\partial\theta\partial\theta'} = E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)\left(\frac{\partial \ln L}{\partial \theta'}\right)'\right]$$

$$\Rightarrow \tilde{\mathcal{I}}(\tilde{\theta}) = \sum_i \left(\frac{\partial \ln f_i}{\partial \tilde{\theta}}\right)\left(\frac{\partial \ln f_i}{\partial \tilde{\theta}'}\right)' = \tilde{W}'\tilde{W}$$

then

$$LM = \left(\frac{\partial \ln L}{\partial \tilde{\theta}}\right)' \mathcal{I}(\tilde{\theta})^{-1}\left(\frac{\partial \ln L}{\partial \tilde{\theta}}\right) = \underset{\sim}{i}'\tilde{W}(\tilde{W}'\tilde{W})^{-1}\tilde{W}'\underset{\sim}{i} \sim \chi^2_m$$

Here, $$\underset{\sim}{i}'\tilde{W}(\tilde{W}'\tilde{W})^{-1}\tilde{W}'\underset{\sim}{i} = n \cdot \frac{\underset{\sim}{i}'\tilde{W}(\tilde{W}'\tilde{W})^{-1}\tilde{W}'\underset{\sim}{i}}{\underset{\sim}{i}'\underset{\sim}{i}}$$

where $\underset{\sim}{i}'\underset{\sim}{i} = n$

$= nR^2$ where R^2 is obtained from the regression of $\underset{\sim}{i}$ on $\tilde{W}$.

Note $$R^2 = \frac{\hat{y}'\hat{y}}{y'M_i y} = \frac{\hat{y}'\hat{y}}{y'y} \quad \text{if there is no constant (uncentered } R^2)$$

where $\hat{y}'\hat{y} = \hat{\beta}'X'X\hat{\beta} = y'X(X'X)^{-1}(X'X)(X'X)^{-1}X'y$
$= y'X(X'X)^{-1}X'y$

choose $y = \underset{\sim}{i}$, $X = \tilde{W}$.

22

stata examples (OLS)

Statistics/Data Analysis

```
      log:  C:\Documents and Settings\jli2\My Documents\Log\MLE_log\Linear_414.smcl
 log type:  smcl
opened on:  29 Sep 2004, 14:54:40

1 . * linear regression model
2 . * Wooldridge textbook, page 80, exercise 4.14
3 . use http://fmwww.bc.edu/ec-p/data/wooldridge/ATTEND

4 . * use "canned command--regress" to run regression
5 . regress stndfnl atndrte frosh soph

      Source |       SS       df       MS              Number of obs =     680
-------------+------------------------------           F(  3,   676) =    6.74
       Model |  19.3023743     3  6.43412478           Prob > F      =  0.0002
    Residual |  645.461067   676  .954824063           R-squared     =  0.0290
-------------+------------------------------           Adj R-squared =  0.0247
       Total |  664.763441   679   .97903305           Root MSE      =  .97715

------------------------------------------------------------------------------
     stndfnl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     atndrte |   .0081634   .0022031     3.71   0.000     .0038376    .0124892
       frosh |  -.2898943   .1157244    -2.51   0.012    -.5171167   -.0626719
        soph |  -.1184455   .0990267    -1.20   0.232    -.3128824    .0759913
       _cons |  -.5017308   .1963139    -2.56   0.011    -.8871892   -.1162724
------------------------------------------------------------------------------
```

*

```
6 . * use self-defined maximum likelihood estimator
7 . program define linear_regress
  1. args lnf theta1 theta2
  2. quietly replace `lnf'=-0.5*ln(`theta2')-($ML_y1-`theta1')^2/(2*`theta2')
  3. end

8 . ml model lf linear_regress ( stndfnl= atndrte frosh soph) ()

9 . ml maximize

initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood = -579.52351
rescale:       log likelihood = -579.52351
rescale eq:    log likelihood = -332.38258
Iteration 0:   log likelihood = -332.38258
Iteration 1:   log likelihood = -322.63624
Iteration 2:   log likelihood = -322.27735
Iteration 3:   log likelihood = -322.27651
Iteration 4:   log likelihood = -322.27651

                                                  Number of obs   =        680
                                                  Wald chi2(3)    =      20.34
Log likelihood = -322.27651                       Prob > chi2     =     0.0001

------------------------------------------------------------------------------
     stndfnl |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
eq1          |
     atndrte |   .0081634   .0021966     3.72   0.000     .0038581    .0124687
       frosh |  -.2898943   .1153835    -2.51   0.012    -.5160418   -.0637468
        soph |  -.1184455    .098735    -1.20   0.230    -.3119626    .0750715
       _cons |  -.5017308   .1957357    -2.56   0.010    -.8853657   -.1180959
-------------+----------------------------------------------------------------
eq2          |
       _cons |   .9492075    .051478    18.44   0.000     .8483124    1.050103
------------------------------------------------------------------------------
```

```
10 . * in the same fashion, you can try exercise 4.13 on page 80
11 . log close
       log:  C:\Documents and Settings\jli2\My Documents\Log\MLE_log\Linear_414.smcl
  log type:  smcl
 closed on:  29 Sep 2004, 15:01:56
```

Tobit


```
      log:  C:\Documents and Settings\jli2\My Documents\Log\MLE_log\tobit_165.smcl
 log type:  smcl
opened on:  29 Sep 2004, 20:09:46

1 . * tobit model, see wooldridge page 546. exercise 16.5
2 . use http://fmwww.bc.edu/ec-p/data/wooldridge/FRINGE

3 . * run OLS regression
4 . regress hrbens exper age educ tenure married male white nrtheast nrthcen south union
```

Source	SS	df	MS
Model	101.132288	11	9.19384436
Residual	170.839786	604	.282847328
Total	271.972074	615	.442231015

Number of obs	=	616
F(11, 604)	=	32.50
Prob > F	=	0.0000
R-squared	=	0.3718
Adj R-squared	=	0.3604
Root MSE	=	.53183

hrbens	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
exper	.0029862	.0043435	0.69	0.492	-.005544	.0115164
age	-.0022495	.0041162	-0.55	0.585	-.0103333	.0058343
educ	.082204	.0083783	9.81	0.000	.0657498	.0986582
tenure	.0281931	.0035481	7.95	0.000	.021225	.0351612
married	.0899016	.0510187	1.76	0.079	-.010294	.1900971
male	.251898	.0523598	4.81	0.000	.1490686	.3547274
white	.098923	.0746602	1.32	0.186	-.0477021	.2455481
nrtheast	-.0834306	.0737578	-1.13	0.258	-.2282836	.0614223
nrthcen	-.0492621	.0678666	-0.73	0.468	-.1825451	.084021
south	-.0284978	.0673714	-0.42	0.672	-.1608084	.1038129
union	.3768401	.0499022	7.55	0.000	.2788372	.4748429
_cons	-.6999244	.1772515	-3.95	0.000	-1.048028	-.3518203

```
5 . * run canned "tobit" model
6 . tobit hrbens exper age educ tenure married male white nrtheast nrthcen south union, ll
```

Tobit estimates

Number of obs	=	616
LR chi2(11)	=	283.86
Prob > chi2	=	0.0000
Pseudo R2	=	0.2145

Log likelihood = -519.66616

hrbens	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
exper	.0040631	.0046627	0.87	0.384	-.0050939	.0132201
age	-.0025859	.0044362	-0.58	0.560	-.0112981	.0061263
educ	.0869168	.0088168	9.86	0.000	.0696015	.1042321
tenure	.0287099	.0037237	7.71	0.000	.021397	.0360227
married	.1027574	.0538339	1.91	0.057	-.0029666	.2084814
male	.2556765	.0551672	4.63	0.000	.1473341	.364019
white	.0994408	.078604	1.27	0.206	-.054929	.2538105
nrtheast	-.0778461	.0775035	-1.00	0.316	-.2300547	.0743625
nrthcen	-.0489422	.0713965	-0.69	0.493	-.1891572	.0912729
south	-.0246854	.0709243	-0.35	0.728	-.1639731	.1146022
union	.4033519	.0522697	7.72	0.000	.3006999	.5060039
_cons	-.8137158	.1880725	-4.33	0.000	-1.18307	-.4443616
_se	.5551027	.0165773	(Ancillary parameter)			

```
Obs. summary:        41   left-censored observations at hrbens<=0
                    575      uncensored observations
```

```
 7 . * the difference between tobit and OLS is small, why?
 8 . * user-defined MLE procedure
 9 . program drop _all

10 . program define tobit_165
   1. args lnf theta1 theta2
   2. quietly replace `lnf'=ln(norm(-`theta1'/sqrt(`theta2'))) if $ML_y1==0
   3. quietly replace `lnf'=-0.5*ln(`theta2')-($ML_y1-`theta1')^2/(2*`theta2') if $ML_y1!=0
   4. end

11 . ml model lf tobit_165 (hrbens = exper age educ tenure married male white nrtheast nrthcen sout

12 . ml maximize

initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood = -224.43978
rescale:       log likelihood = -193.47444
rescale eq:    log likelihood = -140.95815
Iteration 0:   log likelihood = -140.95815
Iteration 1:   log likelihood = -40.265962
Iteration 2:   log likelihood =   2.516603  (not concave)
Iteration 3:   log likelihood =  6.0557566  (not concave)
Iteration 4:   log likelihood =  7.9079098  (not concave)
Iteration 5:   log likelihood =  8.3157337
Iteration 6:   log likelihood =  8.6986728
Iteration 7:   log likelihood =  8.7234839
Iteration 8:   log likelihood =  8.7234948
Iteration 9:   log likelihood =  8.7234948

                                                  Number of obs   =        616
                                                  Wald chi2(11)   =     355.32
Log likelihood =  8.7234948                       Prob > chi2     =     0.0000
```

hrbens	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
eq1						
exper	.0040631	.0046628	0.87	0.384	-.0050758	.013202
age	-.0025859	.0044363	-0.58	0.560	-.0112809	.0061091
educ	.0869168	.008817	9.86	0.000	.0696357	.1041979
tenure	.0287099	.0037238	7.71	0.000	.0214114	.0360083
married	.1027574	.0538353	1.91	0.056	-.002758	.2082727
male	.2556765	.0551687	4.63	0.000	.1475479	.3638051
white	.0994408	.0786061	1.27	0.206	-.0546243	.2535058
nrtheast	-.0778461	.0775056	-1.00	0.315	-.2297543	.0740621
nrthcen	-.0489422	.0713984	-0.69	0.493	-.1888805	.0909961
south	-.0246854	.0709261	-0.35	0.728	-.1636981	.1143272
union	.4033519	.0522711	7.72	0.000	.3009025	.5058013
_cons	-.8137158	.1880775	-4.33	0.000	-1.182341	-.4450907
eq2						
_cons	.308139	.0184054	16.74	0.000	.2720651	.344213

```
13 . * we get the same result as that of canned command
14 . log close
       log:  C:\Documents and Settings\jli2\My Documents\Log\MLE_log\tobit_165.smcl
  log type:  smcl
 closed on:  29 Sep 2004, 20:13:50
```

Ordered probit

Statistics/Data Analysis

```
       log:  C:\Documents and Settings\jli2\My Documents\Log\MLE_log\order_probit.smcl
  log type:  smcl
 opened on:  29 Sep 2004, 21:07:34

1 . * order probit model
2 . * we use the data from stata official website
3 . use http://www.stata-press.com/data/r8/fullauto
  (Automobile Models)

4 . * we need to look at how the variable "rep77" is labeled.
5 . des

 Contains data from http://www.stata-press.com/data/r8/fullauto.dta
   obs:            74                          Automobile Models
  vars:            18                          3 Sep 2002 12:25
  size:         3,700 (99.6% of memory free)
-------------------------------------------------------------------------------
               storage  display     value
 variable name   type   format      label      variable label
-------------------------------------------------------------------------------
 make            int    %8.0g       make       Make
 model           int    %8.0g       model      Model
 price           int    %8.0g                  Price
 mpg             int    %8.0g                  Mileage (mpg)
 rep78           int    %9.0g       repair     Repair Record 1978
 rep77           int    %9.0g       repair     Repair Record 1977
 hdroom          float  %6.1f                  Headroom (in.)
 rseat           float  %6.1f                  Rear Seat (in.)
 trunk           int    %8.0g                  Trunk space (cu. ft.)
 weight          int    %8.0g                  Weight (lbs.)
 length          int    %8.0g                  Length (in.)
 turn            int    %8.0g                  Turn Circle (ft.)
 displ           int    %8.0g                  Displacement (cu. in.)
 gratio          float  %6.2f                  Gear Ratio
 order           int    %8.0g                  Original order
 foreign         int    %8.0g       foreign    Foreign
 wgtd            float  %9.0g
 wgtf            float  %9.0g
-------------------------------------------------------------------------------
 Sorted by:  make  model

6 . label list repair
 repair:
           1 Poor
           2 Fair
           3 Average
           4 Good
           5 Excellent

7 . * run the canned "oprobit" routine
8 . oprobit rep77  foreign length mpg

 Iteration 0:   log likelihood = -89.895098
 Iteration 1:   log likelihood = -78.141221
 Iteration 2:   log likelihood = -78.020314
 Iteration 3:   log likelihood = -78.020025

 Ordered probit estimates                          Number of obs   =         66
                                                   LR chi2(3)      =      23.75
                                                   Prob > chi2     =     0.0000
 Log likelihood = -78.020025                       Pseudo R2       =     0.1321
```

rep77	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
foreign	1.704861	.4246786	4.01	0.000	.8725057	2.537215
length	.0468675	.012648	3.71	0.000	.022078	.0716571
mpg	.1304559	.0378627	3.45	0.001	.0562464	.2046654
_cut1	10.1589	3.076749	(Ancillary parameters)			
_cut2	11.21003	3.107522				
_cut3	12.54561	3.155228				
_cut4	13.98059	3.218786				

```
 9 . * user-defined MLE procedure
10 . program drop _all

11 . program define orderprobit_stata
  1. args lnf theta1 theta2 theta3 theta4 theta5
  2. quietly replace `lnf'=ln(norm(`theta2'-`theta1')) if $ML_y1==1
  3. quietly replace `lnf'=ln(norm(`theta3'-`theta1')-norm(`theta2'-`theta1')) if $ML_y1==2
  4. quietly replace `lnf'=ln(norm(`theta4'-`theta1')-norm(`theta3'-`theta1')) if $ML_y1==3
  5. quietly replace `lnf'=ln(norm(`theta5'-`theta1')-norm(`theta4'-`theta1')) if $ML_y1==4
  6. quietly replace `lnf'=ln(1-norm(`theta5'-`theta1')) if $ML_y1==5
  7. end

12 . ml model lf orderprobit_stata ( rep77= foreign length mpg,noconstant) () () () ()

13 . ml maximize

initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood =  -452.7671
rescale:       log likelihood = -191.63415
rescale eq:    log likelihood = -180.53523
Iteration 0:   log likelihood = -180.53523
Iteration 1:   log likelihood = -101.92816
Iteration 2:   log likelihood =  -79.26036
Iteration 3:   log likelihood = -78.106616
Iteration 4:   log likelihood = -78.020063
Iteration 5:   log likelihood = -78.020025
Iteration 6:   log likelihood = -78.020025

                                                  Number of obs   =         66
                                                  Wald chi2(3)    =      21.93
Log likelihood = -78.020025                       Prob > chi2     =     0.0001
```

rep77	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
eq1						
foreign	1.704861	.4246799	4.01	0.000	.8725037	2.537218
length	.0468676	.012648	3.71	0.000	.0220779	.0716572
mpg	.130456	.0378628	3.45	0.001	.0562463	.2046657
eq2						
_cons	10.15891	3.076759	3.30	0.001	4.128573	16.18925
eq3						
_cons	11.21004	3.107531	3.61	0.000	5.119386	17.30068
eq4						
_cons	12.54562	3.155238	3.98	0.000	6.361463	18.72977
eq5						
_cons	13.9806	3.218798	4.34	0.000	7.671871	20.28933

28

```
14 . * remark 1: we need the option "noconstant"
15 . * remark 2: this is a five-equation model
16 . log close
       log:  C:\Documents and Settings\jli2\My Documents\Log\MLE_log\order_probit.smcl
  log type:  smcl
 closed on:  29 Sep 2004, 21:12:39
```

Neg-binomial

Statistics/Data Analysis

```
      log:  C:\Documents and Settings\jli2\My Documents\Log\MLE_log\negative_binomial.smcl
 log type:  smcl
opened on:  29 Sep 2004, 22:08:30

1 . * count data_negative binomial model
2 . * we use the data from stata official website
3 . use http://www.stata-press.com/data/r8/rod93

4 . * to generate relevant variables
5 . quietly tab cohort, gen(coh)

6 . * first run the canned "nbreg" routine
7 . nbreg  deaths coh2 coh3

  Fitting Poisson model:

  Iteration 0:   log likelihood =  -325.7993
  Iteration 1:   log likelihood =  -325.7993

  Fitting constant-only model:

  Iteration 0:   log likelihood = -114.33669
  Iteration 1:   log likelihood = -110.33038
  Iteration 2:   log likelihood = -108.56521
  Iteration 3:   log likelihood = -108.56018
  Iteration 4:   log likelihood = -108.56018

  Fitting full model:

  Iteration 0:   log likelihood = -108.48867
  Iteration 1:   log likelihood = -108.48841
  Iteration 2:   log likelihood = -108.48841

  Negative binomial regression                      Number of obs   =         21
                                                    LR chi2(2)      =       0.14
                                                    Prob > chi2     =     0.9307
  Log likelihood = -108.48841                       Pseudo R2       =     0.0007
```

deaths	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
coh2	.0591305	.2978419	0.20	0.843	-.5246289	.64289
coh3	-.0538792	.2981621	-0.18	0.857	-.6382662	.5305077
_cons	4.435906	.2107213	21.05	0.000	4.0229	4.848912
/lnalpha	-1.207379	.3108622			-1.816657	-.5980999
alpha	.29898	.0929416			.1625683	.5498555

```
  Likelihood-ratio test of alpha=0:  chibar2(01) =  434.62 Prob>=chibar2 = 0.000

8 . * user-define MLE procedure
9 . program drop _all
```

```
10 . program define neg_bin_stata
  1. args lnf theta1 theta2
  2. quietly replace `lnf'=lngamma((1/`theta2')+$ML_y1)-lngamma($ML_y1+1)-lngamma(1/`theta2')+(1
> )))
  3. end

11 . ml model lf neg_bin_stata ( deaths= coh2 coh3) ()

12 . ml maximize

initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood = -1347.5967
rescale:       log likelihood = -130.77236
rescale eq:    log likelihood = -114.30781
Iteration 0:   log likelihood = -114.30781
Iteration 1:   log likelihood = -109.97715
Iteration 2:   log likelihood = -108.54119
Iteration 3:   log likelihood = -108.48846
Iteration 4:   log likelihood = -108.48841
Iteration 5:   log likelihood = -108.48841

                                                  Number of obs   =         21
                                                  Wald chi2(2)    =       0.14
Log likelihood = -108.48841                       Prob > chi2     =     0.9306
```

deaths	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
eq1						
coh2	.0591305	.2978419	0.20	0.843	-.524629	.64289
coh3	-.0538792	.2981621	-0.18	0.857	-.6382662	.5305078
_cons	4.435906	.2107214	21.05	0.000	4.0229	4.848912
eq2						
_cons	.2989801	.0929416	3.22	0.001	.1168179	.4811423

```
13 . * we get the same result
14 . * the unconditional pdf is given in the description of nbreg command in stata manual
15 . log close
       log:  C:\Documents and Settings\jli2\My Documents\Log\MLE_log\negative_binomial.smcl
  log type:  smcl
 closed on:  29 Sep 2004, 22:11:25
```

Logit

Statistics/Data Analysis

```
      log:  C:\Documents and Settings\jli2\My Documents\Log\MLE_log\logit_158.smcl
 log type:  smcl
opened on:  29 Sep 2004, 18:58:24

 1 . use http://fmwww.bc.edu/ec-p/data/wooldridge/BWGHT

 2 . * generate dummy variable "smokes"=1 if smoking during pregnance
 3 . gen smokes =  cigs>0

 4 . * run the canned routin "logit"
 5 . logit smokes motheduc white lfaminc

   Iteration 0:   log likelihood = -593.10529
   Iteration 1:   log likelihood =  -551.1645
   Iteration 2:   log likelihood = -548.74295
   Iteration 3:   log likelihood = -548.73367
   Iteration 4:   log likelihood = -548.73367

   Logit estimates                                   Number of obs   =       1387
                                                     LR chi2(3)      =      88.74
                                                     Prob > chi2     =     0.0000
   Log likelihood = -548.73367                       Pseudo R2       =     0.0748
```

smokes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
motheduc	-.2518731	.0372045	-6.77	0.000	-.3247925	-.1789536
white	.3438842	.2002178	1.72	0.086	-.0485355	.7363038
lfaminc	-.2962299	.0866347	-3.42	0.001	-.4660309	-.126429
_cons	2.012535	.4474219	4.50	0.000	1.135604	2.889466

```
 6 . * user-defined MLE procedure
 7 . program define logit_158
   logit_158 already defined
   r(110);

 8 . program drop _all

 9 . program define logit_158
     1. args lnf theta
     2. quietly replace `lnf' =ln(exp(`theta')/(1+exp(`theta'))) if $ML_y1==1
     3. quietly replace `lnf' =ln(1-exp(`theta')/(1+exp(`theta'))) if $ML_y1==0
     4. end

10 . ml model lf logit_158 ( smokes= motheduc white lfaminc)

11 . ml maximize

   initial:       log likelihood = -961.39514
   alternative:   log likelihood = -763.54478
   rescale:       log likelihood = -600.04915
   Iteration 0:   log likelihood = -600.04915
   Iteration 1:   log likelihood =  -556.4819
   Iteration 2:   log likelihood = -548.75296
   Iteration 3:   log likelihood = -548.73367
   Iteration 4:   log likelihood = -548.73367

                                                     Number of obs   =       1387
                                                     Wald chi2(3)    =      78.42
   Log likelihood = -548.73367                       Prob > chi2     =     0.0000
```

smokes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
motheduc	-.2518731	.0372045	-6.77	0.000	-.3247925	-.1789536
white	.3438842	.2002178	1.72	0.086	-.0485355	.7363038
lfaminc	-.2962299	.0866347	-3.42	0.001	-.4660309	-.126429
_cons	2.012535	.4474219	4.50	0.000	1.135604	2.889466

```
12 . log close
       log:  C:\Documents and Settings\jli2\My Documents\Log\MLE_log\logit_158.smcl
  log type:  smcl
 closed on:  29 Sep 2004, 19:01:44
```

```
* Poisson regression (Ex. 5.3, Greene, p. 208)

input id y x
     1         6     1.5
     2         7     1.8
     3         4     1.8
     4         10    2.0
     5         10    1.3
     6         6     1.6
     7         4     1.2
     8         7     1.9
     9         2     1.8
    10         3     1.0
    11         6     1.4
    12         5     0.5
    13         3     0.8
    14         3     1.1
    15         4     0.7
end

list

* Poisson regression;

poisson y x

* Poisson MLE

*clear
*insheet using poisson_data.txt

log using poisson_output.log, replace

/* this is the "canned" routine that estimates the poisson regression */

poisson y x

/* this maximizes lnL directly, using logged factorial of y */

program define poisreg1
 args lnf theta
 quietly replace `lnf' = -exp(`theta') + $ML_y1*(`theta') - lnfact($ML_y1)
end

ml model lf poisreg1 (y=x)
ml maximize
```

```
clear
set mem 200m
set more off
*cd "c:\upcd1\work\stata_prg" /* please change to your own working directory */

log using mle_probit_ex.log, replace

webuse auto.dta, clear
 /* to use data directly from STATA official website */
save auto.dta, replace
 /* to save the data set you just netted from the website to your working directory */

webuse weib.dta,clear
 /* to use data directly from STATA official website */
save weib.dta, replace
 /* to save the data set you just netted from the website to your working directory */

**************************
* Overview of ml *
**************************

**********************************************
* one equation one dependent variable *
**********************************************

use auto.dta, clear
capture program drop myprobit

program myprobit
 /* to use maximum likelihood method to estimate the probit model */
 version 8.0
 args lnf theta1
 quietly replace `lnf' = ln(norm(`theta1')) if $ML_y1==1
 quietly replace `lnf' = ln(norm(-`theta1')) if $ML_y1==0
end

ml model lf myprobit (foreign = mpg weight)
  /* to estimate the basic probit model using method lf */
ml check
  /* to verify that the log-likelihood evaluator you have written seems to work */
ml search
  /* to find a better initial value so that the convergence of log-likelihood takes less time*/
ml maximize
ml graph
  /* to graph the log-likelihood values against the iteration number to verify the process of convergence */

ml model lf myprobit (foreign=mpg weight, nocons)
```

```
  /* to estimate the restricted model without the intercept */
ml maximize

ml model lf myprobit (foreign=)
  /* to estimate the restricted model whose only explantory variable is the constant */
ml maximize

constraint define 1 _b[mpg]==0 /* to define a constraint which drops a coefficient */
ml model lf myprobit (foreign = mpg weight), constraint (1)
  /* to estimate the restricted model using ML*/
ml maximize

ml model lf myprobit (foreign = mpg weight), robust
  /* to specify the robust variance estimator and report it */
ml maximize

ml model lf myprobit (foreign = mpg weight) if mpg<= 25
  /* to specify a subsample to estimate the model*/
ml maximize

ml model lf myprobit (foreign = mpg weight)
ml maximize, gradient hessian trace
  /* to add to the iteration log a report on the current gradient vector,
   current negative Hessian matrix and current parameter vector */

ml model lf myprobit (foreign = mpg weight)
ml maximize,level(99)
  /* to change the significance level from 95 to 99*/


*************************************************
* two equations, one dependent variable *
*************************************************

use auto.dta,clear
capture program drop myreg
  /* to use maximum likelihood method to estimate the linear regression model */
program myreg
version 8.0
args lnf theta1 theta2
  /* to use theta2 to stand for the standard deviation, so this is a two equation model */
quietly replace `lnf' = ln(normd(($ML_y1-`theta1')/`theta2'))-ln(`theta2')
end

ml model lf myreg (mpg =weight displ)/sigma
  /* to use sigma to specify the equation for standard deviation */
ml check
```

37

```
ml search
ml maximize
ml graph

ml model lf myreg (mpg =weight displ) ()
  /* an equivalent specification as the former one which uses sigma */
ml maximize

ml model lf myreg (mpg =weight displ) (sigma:)
  /* another equivalent way to specify the second equation, and the name
   of the second quation has been changed to sigma */
ml maximize

ml model lf myreg (mpg =weight displ) (sigma: weight)
  /* to estimate the conditional heteroskedasticity model, and to
   specify that standard deviation depends on weight and constant */
ml maximize

ml model lf myreg (mpg =weight displ) (sigma: weight, nocons)
  /* to specify the second equation without intercept */
ml maximize

***********************************************
* two equations, two dependent variables *
***********************************************

use weib.dta,clear
capture program drop myweib
  /* to use maximum likelihood method to estimate the Weibull model */

program define myweib
 version 8.0
 args lnf theta1 theta2
 tempvar p M R
   /* to make programming easier by introducing temporary variabls*/
 quietly gen double `p' = exp(`theta2')
   /* temporary variables should be generated as doubles */
 quietly gen double `M' = ($ML_y1*exp(-`theta1'))^`p'
 quietly gen double `R' = ln($ML_y1)-`theta1'
 quietly replace `lnf' = -`M' + $ML_y2*(`theta2'-`theta1'+(`p'-1)*`R')
end

ml model lf myweib (studytime died =drug2 drug3 age) ()
  /* the () corresponds to the Weibull shape parameter s*/
ml check
ml search
ml maximize
```

38

```
ml graph

ml model lf myweib (studytime died =drug2 drug3 age) /sigma
  /* An alternative way to specify the second equation*/
ml maximize

ml model lf myweib (studytime died =drug2 drug3 age) /s
  /* Another equivalent way to specify the second equation*/
ml maximize

*********************************************************************
* another example of two equations, two dependent variables model *
*********************************************************************

use auto.dta,clear
capture program drop myreg1
program myreg1
version 8.0
args lnf theta1 theta2
quietly replace `lnf' = ln(normd(($ML_y1-`theta1')/`theta2'))-ln(`theta2')
end

ml model lf myreg1 (mpg =weight displ) (price=weight)
 /* the second dependent variable is price which depends on weight and constant*/
ml maximize

ml model lf myreg1 (mpg price=weight displ) (weight)
 /* an equivalent specification*/
ml maximize

ml model lf myreg1 (mpg =weight displ) (price=weight, nocons)
 /* the second dependent variable depends only on weight*/
ml maximize

ml model lf myreg1 (mpg price=weight displ) (weight, nocons)
 /* an equivalent specification of the restricted model*/
ml maximize

ml model lf myreg1 (mpg =weight displ) (price=)
 /* the second dependent varible is a constant, in this case
  the model degenerates into one dependent variable model*/
ml maximize

ml model lf myreg1 (mpg =weight displ)/s
  /* an equivalent specification of the one dependent variable model*/
ml maximize
```

```
/*=====================================================
Exercise.  Poisson regression (Ex. 5.3, Greene, p. 208)
*/=====================================================
Read ; Nobs = 15 ; Nvar = 3 ; Names = I,Y,X$
     1        6     1.5
     2        7     1.8
     3        4     1.8
     4        10    2.0
     5        10    1.3
     6        6     1.6
     7        4     1.2
     8        7     1.9
     9        2     1.8
    10        3     1.0
    11        6     1.4
    12        5     0.5
    13        3     0.8
    14        3     1.1
    15        4     0.7
Sample;1-15$
?
? Maximum likelihood estimation.
?
Maxize ; fcn = -exp(a + beta*x) + y * (a + beta*x) - log(Gma(y+1))
       ; start= 1, 0.5
       ; labels= a, beta$

? Or alternatively, using the Poisson regression
Poisson ; Lhs = y
        ; Rhs = one, x $

/*
--> Maxize ; fcn = -exp(a + beta*x) + y * (a + beta*x) - log(Gma(y+1))
    ; start= 1, 0.5
    ; labels= a, beta$
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit from iterations. Exit status=0.

              +----------------------------------------------+
              | User Defined Optimization                    |
              | Maximum Likelihood Estimates                 |
              | Dependent variable               Function    |
              | Weighting variable                    ONE    |
              | Number of observations                 15    |
              | Iterations completed                    5    |
              | Log likelihood function         -32.06601    |
              +----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 A          1.078077604     .52117521        2.069   .0386
```

```
BETA       .4248897068       .33080364        1.284   .1990


--> Poisson ; Lhs = y
   ; Rhs = one, x $

+-----------------------------------------------------------------------+
| Poisson Regression Model - OLS Results                                |
| Ordinary    least squares regression    Weighting variable = none     |
| Dep. var. = Y        Mean=   5.333333333    , S.D.=   2.439750182     |
| Model size: Observations =       15, Parameters =   2, Deg.Fr.=    13 |
| Residuals:  Sum of squares= 68.73970524     , Std.Dev.=       2.29949 |
| Fit:        R-squared=  .175124, Adjusted R-squared =          .11167 |
| Model test: F[  1,      13] =    2.76,    Prob value =          .12056 |
| Diagnostic: Log-L =    -32.7012, Restricted(b=0) Log-L =     -34.1451 |
|             LogAmemiyaPrCrt.=    1.791, Akaike Info. Crt.=      4.627 |
+-----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant  2.371044647      1.8793587        1.262   .2071
 X         2.178153446      1.3111096        1.661   .0967  1.3600000


              +----------------------------------------------+
              | Poisson Regression                           |
              | Maximum Likelihood Estimates                 |
              | Dependent variable                 Y         |
              | Weighting variable               ONE         |
              | Number of observations            15         |
              | Iterations completed               5         |
              | Log likelihood function    -32.06601         |
              | Restricted log likelihood  -33.46887         |
              | Chi-squared                 2.805710         |
              | Degrees of freedom                 1         |
              | Significance level           .9392927E-01    |
              | Chi- squared =    12.22602  RsqP=   .2175    |
              | G  - squared =    12.36334  RsqD=   .1850    |
              +----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant  1.078077619      .38864334        2.774   .0055
 X         .4248896980      .25781092        1.648   .0993  1.3600000

*/
```

```
/*=================================================
 Weibull Model
*/=================================================

Read ; Nobs = 20 ; Nvar = 1 ; Names = X ; byvar$
1.3043  .49254  1.2742  1.4019  .32556  .29965  .26423
1.0878 1.9461  .47615  3.6454  .15344  1.2357  .96381
.33453  1.1227  2.0296 1.2797  .96080  2.0070

?
Maxize ; fcn = log(a) + log(beta) + (beta-1) * log(x) - a * x^(beta)
       ; start= 1.0, .5
     ; TLG=1.D-12
       ; labels=beta,a $


Maxize ; fcn = log(0.88494) + log(beta) + (beta-1) * log(x) - 0.88494*
x^(beta)
       ; start= 1.0
     ; TLG=1.D-12
       ; labels=beta $
/*

--> Maxize ; fcn = log(a) + log(beta) + (beta-1) * log(x) - a * x^beta
    ; start= .5, .5
    ; labels=beta,a$
Note: DFP and BFGS usually take more than 4 or 5
iterations to converge.  If this problem was not
structured for quick convergence, you might want
to examine results closely. If convergence is too
early, tighten convergence with, e.g., ;TLG=1.D-9.
Normal exit from iterations. Exit status=0.

                +-----------------------------------------------+
                | User Defined Optimization                     |
                | Maximum Likelihood Estimates                  |
                | Dependent variable              Function      |
                | Weighting variable                   ONE      |
                | Number of observations                20      |
                | Iterations completed                   4      |
                | Log likelihood function        -20.35084      |
                +-----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 BETA      1.435724189      .26553432         5.407   .0000
 A         .7389347022      .19996900         3.695   .0002
*/
```

```
/*=====================================================
Section 4.9.4.  Example of Various Test Procedures.
*/=====================================================
Read ; Nobs = 20 ; Nvar = 3 ; Names = I,Y,E$
     1          20.5      12
     2          31.5      16
     3          47.7      18
     4          26.2      16
     5          44.0      12
     6          8.28      12
     7          30.8      16
     8          17.2      12
     9          19.9      10
    10          9.96      12
    11          55.8      16
    12          25.2      20
    13          29.0      12
    14          85.5      16
    15          15.1      10
    16          28.5      18
    17          21.4      16
    18          17.7      20
    19          6.42      12
    20          84.9      16
Sample;1-20$
?
? Just change name to be consistent with text
?
Create;x=e$
?
? Unrestricted maximum likelihood estimation.
?
Maxize ; fcn = -r*log(beta+x)-log(gma(r))-y/(beta+x)+(r-1)*log(y)
       ; start=-5,1
       ; labels=beta,r$
/*
              +----------------------------------------------+
              | User Defined Optimization                    |
              | Maximum Likelihood Estimates                 |
              | Dependent variable               Function    |
              | Weighting variable                    ONE    |
              | Number of observations                 20    |
              | Iterations completed                    4    |
              | Log likelihood function         -82.91605    |
              +----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 BETA     -4.718503621      3.6568024       -1.290   .1969
 R         3.150896345      1.2398481        2.541   .0110
*/
```

```
?
? Compute variables that are first and second derivatives
? gb and gr are first derivatives, hbb,hrr,hbr = Hessian
? Also computes log likelihood function
?
Create ; gb=-rml/(betaml+x)+y/(betaml+x)^2
       ; gr=-log(betaml+x)-psi(rml)+log(y)
       ; hbb=rml/(betaml+x)^2-2*y/(betaml+x)^3
       ; hrr=-psp(rml)
       ; hbr=-1/(betaml+x)$
       ; loglik=-rml*log(betaml+x)-log(gma(rml))
              -y/(betaml+x)+(rml-1)*log(y)$
?
? Summing terms produces log likelihood and derivatives.
?
calc;list;lloglu=sum(loglik)
         ;gbu=sum(gb)
         ;gru=sum(gr)
         ;hbbu=sum(hbb)
         ;hrru=sum(hrr)
         ;hbru=sum(hbr)$
         ;hbru=sum(hbr)$
*/
    LLOGLU   = -.82916048583538210D+02
    GBU      = -.16887894027650670D-07
    GRU      =  .54968437801505840D-07
    HBBU     = -.85570382274745960D+00
    HRRU     = -.74591837131888800D+01
    HBRU     = -.22419691609929970D+01
Calculator: Computed   6 scalar results
*/
? Estimators for asymptotic covariance matrix
? 1.  Based on actual Hessian
?
Matrix ; vh=[hbbu/hbru,hrru] ; vh=-1*vh ; list; vh= <vh>$
*/
Matrix VH       has  2 rows and  2 columns.
               1              2
        +----------------------------
       1|  .5499144D+01 -.1652850D+01
       2| -.1652850D+01  .6308517D+00
*/
? 2.  Expected Hessian.  Compute variables and sum
?
Create ; ehbb=rml/(betaml+x)^2
       ; ehrr=psp(rml)
       ; ehbr=1/(betaml+x)$
Calc   ; vehbb=sum(ehbb)
       ; vehrr=sum(ehrr)
       ; vehbr=sum(ehbr)$
Matrix ; list;evh=[vehbb/vehbr,vehrr];evh=<evh>$
/*
Matrix EVH       has  2 rows and  2 columns.
               1              2
        +----------------------------
```

```
       1|  .4900316D+01 -.1472863D+01
       2| -.1472863D+01  .5767540D+00
*/

? 3.  BHHH estimator can be obtained using simple sums
?
Namelist ; G=gb,gr$
Matrix   ; list ; VB = <G'G> $
/*
Matrix VB       has  2 rows and  2 columns.
               1            2
        +--------------------------
       1|  .1337220D+02 -.4321743D+01
       2| -.4321743D+01  .1537223D+01
*/

?----------------------------------------------------------
? Testing procedures for the hypothesis RHO = 1.
?----------------------------------------------------------
? 1.  Form confidence interval
?
Calc ; list ; rholower=r-1.96*sqr(vh(2,2))
            ;rhoupper=r+1.96*sqr(vh(2,2))$
/*
   RHOLOWER=  .15941433617939830D+01
   RHOUPPER=  .47076493284860730D+01
*/
?
? 2.  Likelihood ratio test requires restricted maximum
?     Note it's done by fixing RHO at the start value.
?
Maximize ; fcn=-r*log(beta+x)-log(gma(r))-y/(beta+x)+(r-1)*log(y)
         ; start=-5,1
         ; labels=beta,r
         ; fix=r$
Calc;list; lrtest=-2*(logl-lloglu)$
/*
              +----------------------------------------------+
              | User Defined Optimization                    |
              | Maximum Likelihood Estimates                 |
              | Dependent variable              Function     |
              | Weighting variable                   ONE     |
              | Number of observations                20     |
              | Iterations completed                   2     |
              | Log likelihood function        -88.43626     |
              +----------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 BETA      15.60272448      .24174096E-02 6454.316   .0000
 R         1.000000000    ........(Fixed Parameter)........

   LRTEST  =  .11040428574057930D+02
*/
```

```
? Wald test
? Recompute estimates, then use built-in Wald procedure.
? This uses the BHHH estimator for the VC matrix.
?
Maximize ; fcn=-r*log(beta+x)-log(gma(r))-y/(beta+x)+(r-1)*log(y)
         ; start=-5,1
         ; labels=beta,r$
Wald     ; fn1=r-1$
/*
                +------------------------------------------+
                | WALD procedure. Estimates and standard errors |
                | for nonlinear functions and joint test of    |
                | nonlinear restrictions.                      |
                | Wald Statistic             =     3.00955     |
                | Prob. from Chi-squared[ 1] =      .08278     |
                +------------------------------------------+
+---------+-------------+---------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+-------------+---------------+--------+---------+----------+
 Fncn( 1)  2.150896345      1.2398481        1.735   .0828
*/
?
? Unfortunately, if the test is based on the Hessian, a different
? conclusion is reached.  Using asymptotic results with 20
? observations can lead to this.
?
Calc  ; List ; Waldtest=(rml-1)^2/VH(2,2)$
/*
   WALDTEST=  .73335066911316080D+01
*/
? LM Test.  Compute gradient and Hessian using restricted values.
?
? These maximization results appear above.
?
Maximize ; fcn=-r*log(beta+x)-log(gma(r))-y/(beta+x)+(r-1)*log(y)
         ; start=-5,1
         ; labels=beta,r ; Fix = r $
Calc     ; betaml=b(1);rml=b(2)$
Create   ; gb=-rml/(betaml+x)+y/(betaml+x)^2
         ; gr=-log(betaml+x)-psi(rml)+log(y)  $
Namelist ; G=gb,gr$
Matrix   ; list ; lm=1'G*<G'G>*G'1$
/*
Matrix LM       has  1 rows and  1 columns.
                  1
        +--------------
       1|  .1568679D+02
*/
```

Numerical Optimization Methods

1) Newton-Raphson

$$\hat{\theta}_{j+1} = \hat{\theta}_j - \Big(\underbrace{\sum_{i=1}^{n} H_i(\theta)}_{\text{Hessian}}\Big)^{-1}\Big(\underbrace{\sum_{i=1}^{n} s_i(\theta)}_{\text{score}}\Big)$$

: H^{-1} may not exist.

stop if $\left(\sum s_i(\theta)\right)'\left(\sum H_i(\theta)\right)^{-1}\left(\sum s_i(\theta)\right) \leq \varepsilon$ (small values)

ii) BHHH

$$\hat{\theta}_{j+1} = \hat{\theta}_j - \underbrace{r}_{\text{step size}}\underbrace{\Big(\sum \underbrace{s_i(\theta)s_i(\theta)'}_{\text{product of scores}}\Big)^{-1}\left(\sum s_i(\theta)\right)}_{\text{step direction}}$$

: easy, no need to use second derivatives (only score vectors)

r = coeff of the LS of 1 on $s_i(\hat{\theta})$.

iii) Generalized Gauss-Newton

$$\hat{\theta}_{j+1} = \hat{\theta}_j - r\Big(\underbrace{\sum_{i}^{n} A_i(\theta)}_{\text{information}}\Big)^{-1}\Big(\sum_{i}^{n} s_i(\theta)\Big)$$

: simpler

```
output file = mle_NR_poisson.out reset;

format /rd/ml 8,4;

let x[10,1] = 5 0 1 1 0 3 2 3 4 1;
x';
sum = sumc(x);
"Sum = " sum;
n = rows(x);
"N   = " n;

sw = 0;
beta0 = 1;
old_logl = - 100000000000;
i = 1;

"";
"i score hessian beta0 beta1 new_logl"; "";

do while i <= 1000;

   score = -n + (1/beta0)*sum;
   hessian = -(1/beta0^2)*sum;
   beta1 = beta0 - (1/hessian)*score;

   new_logl = -n*beta1 + ln(beta1)*sum;

   i score hessian beta0 beta1 new_logl;

   if new_logl - old_logl < 0.00001; goto next;  endif;

   beta0 = beta1;
   old_logl = new_logl;

   i = i + 1;
 endo;

 next:
 "";
 "Optimal estimate:";
 "#   score hessian beta0 beta1 new_logl"; "";
  i   score hessian beta0 beta1 new_logl;
```

$$f(x_i) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$\log L = -n\lambda + (\log\lambda)\sum_{i=1}^{n} x_i$$

$$\frac{\partial \ln L}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i$$

$$\tilde{\theta} = \hat{\theta} - \left(\frac{\partial^2 \ln L}{\partial \lambda^2}\right)^{-1}\left(\frac{\partial \ln L}{\partial \lambda}\right)$$

$$\mathcal{I} = -E\,\frac{\partial^2 \ln L}{\partial \lambda^2}$$

$$\mathcal{I}^{-1} = \mathrm{Var}(\hat{\lambda}) =$$

```
 5.0000   0.0000   1.0000   1.0000   0.0000   3.0000   2.0000   3.0000
 4.0000   1.0000
Sum =  20.0000
N   =  10.0000

i score hessian beta0 beta1 new_logl

  1.0000  10.0000 -20.0000   1.0000   1.5000  -6.8907
  2.0000   3.3333  -8.8889   1.5000   1.8750  -6.1778
  3.0000   0.6667  -5.6889   1.8750   1.9922  -6.1372
  4.0000   0.0392  -5.0393   1.9922   2.0000  -6.1371
  5.0000   0.0002  -5.0002   2.0000   2.0000  -6.1371

Optimal estimate:
#   score hessian beta0 beta1 new_logl

  5.0000   0.0002  -5.0002   2.0000   2.0000  -6.1371
```

QMLE (Quasi-MLE) also called pseudo-MLE

MLE that maximizes a likelihood function different from the model's likelihood function.
(misspecified model)

Eg) Poisson regression

$$f(y|x) = \frac{e^{-\mu(x)} \mu(x)^y}{y!} \quad y = 0, 1, 2, \cdots$$

where $\mu(x) = m(x, \beta) \overset{(let)}{=} \exp(X\beta)$

we assume that $\mu(x) = \exp(X\beta)$

$\left(\text{thus, } \frac{\partial E(y|x)}{\partial x_j} = \exp(X\beta) \cdot \beta_j \right)$

Then
$$\mathcal{L}_i = y_i \log [m(x_i, \beta)] - m(x_i, \beta)$$

thus
$\max \sum_{i=1}^{n} \mathcal{L}_i$ is a quasi-MLE,
since $\mu(x) = \exp(X\beta)$ is an assumption
and this may be mis-specified.

Note $E(y_i|X_i) = Var(y_i|X_i)$ in the Poisson model.
But the above MLE does not utilize this information. $Var(y_i|X_i)$ can be anything. (see Wooldridge, p.649)

Eg) Probit Model

MLE $\mathcal{L}_i = y_i \log \Phi(X_i\beta) + (1-y_i)(1 - \log \Phi(X_i\beta))$
the normality assumption may be invalid.

$\Rightarrow$ Q-MLE uses a robust "sandwich" estimator for the variance.

$$Var(\hat{\beta}_{QMLE}) = \hat{A}^{-1} \hat{B} \hat{A}^{-1}$$

where $\hat{A}^{-1} = \widehat{E(-H(\beta))}^{-1} = \left(-\sum_{i=1}^{n} H_i(\hat{\theta})\right]^{-1}$

$$\hat{B} = \left[\sum s_i(\hat{\theta}) s_i(\hat{\theta})'\right]$$

where $s_i(\hat{\theta})$ is the score vector.

$\hat{B} = Var(s_i(\hat{\theta}))$.

Note If the model is correctly specified

$$\text{plim} \frac{1}{n}\hat{B} = \text{plim} \frac{1}{n}(-\hat{H})$$

then, using $\hat{A}^{-1}$ or $\hat{B}^{-1}$ would not make a difference.

Other examples for QMLE

1) Poisson exponential regression, as discussed: QMLE

$m(X_i, \beta) = \exp(X_i\beta)$: the assumption $Var(y|x) = \sigma^2 E(y|x)$ is NOT used.

2) Probit MLE

If normality assumption is not satisfied, use the sandwich (co)variance.

3) Fractional logit regression (useful!)

y is restricted to the unit interval

$[0, 1]$ or $[a, b] \rightarrow 0 \leq \frac{y-a}{b-a} \leq 1$

eg) "fractions" : 'graduation rates' "401 K pension plans" Papke & Wooldridge (1996, JAE)
'charity ratio'
'equity ratio'
:
'percentage'

- Someone uses a Tobit model (censored regression model)
  · This makes sense if there are MANY (censored) obs with zero or upper-bound values. But in my view, using a Tobit model simply because the data values are restricted bet 0 & 1 may be wrong.

- Alternative options (not good)
  ① 'log-odd' transformation and use it as a dep. var
  : $\log\left(\frac{y}{1-y}\right)$ if $0 < y < 1$ (not 0 or 1)
  $\rightarrow \beta$ is difficult to interpret
  ② $E(Y|X) = \frac{\exp(X\beta)}{1+\exp(X\beta)}$ = logistic

  A nonlinear LS can be used.
  $\rightarrow Var(y|x) = \sigma^2$ is hard to hold

- QMLE can be used
  use the usual MLE of probit/logit models
  But y_i = fraction (not 0, 1) is used.

$$\log \mathcal{L}_i = y_i \Phi(X_i\beta) + (1-y_i)(1-\Phi(X_i\beta))$$
(not 0,1)

Short Quiz on MLE

ec671

1. Consider a Poisson dist.

$$f(y_i) = \frac{e^{-\lambda} \lambda^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \ldots$$

a) Find $\log \mathcal{L}$ and the MLE of λ.

b) Find the score vector, Hessian matrix and information matrix.

c) Find the Cramer-Rao bound.

Now, let $\lambda = \exp(X\beta)$ such that $E(y_i | x_i) = \exp(X_i\beta)$.

d) Find $\log \mathcal{L}$ and the score vector.

2. Consider a regression model, $y_i = \alpha + \beta X_i + e_i$,

$e_i \sim N(0, \sigma^2)$.

a) Find the likelihood function for $\theta = (\alpha, \beta, \sigma^2)'$.

b) Find the MLE of θ.

c) Are the MLE of θ, say $\hat{\theta}_{MLE}$, biased?
That is, $E(\hat{\theta}) = \theta$? for each of $\theta = (\alpha, \beta, \sigma^2)$?

d) Find the Hessian matrix of the MLE estimation.

e) Find $Var(\hat{\theta})$ for each of $\theta = (\alpha, \beta, \sigma^2)$.

3. Consider a probit model, where $e_i \sim N(0, \sigma^2)$

$$y_i = \begin{cases} 1 & \text{if } y_i^* = X_i\beta + e_i > 0 \\ 0 & \text{o/w} \end{cases}$$

a) Find the likelihood function of the probit model.

b) Find the score vector, $\partial \ln L / \partial \beta$.