Asymptotic Distribution Theory:

Partly a summary of results from earlier courses, but also some additional results.

Why? Main justification of MLE's will be their "good" asymptotic properties.

Concerned with behaviour of statistics, such as estimators, as $n \to \infty$.

* Finite-sample results may be intractable.
* Finite-sample moments may not exist.
* May be a convenient approximation.

1. Convergence in Probability -

S_n converges in probability (converges "weakly") to a constant, 'c', if

$$\lim_{n\to\infty} [\Pr.(|S_n - c| < \varepsilon)] = 1 \quad ; \quad \text{for } \varepsilon > 0.$$

Then, write $\operatorname{plim}(S_n) = c$

or $S_n \xrightarrow{p} c$.

If $\hat{\theta}_n \xrightarrow{p} \theta$, then $\hat{\theta}_n$ is <u>weakly consistent</u> for θ.

2. Slutsky's Theorem -

If $S_n \xrightarrow{p} c$ & $g(\cdot)$ is continuous, then $g(S_n) \xrightarrow{p} g(c)$.

e.g., $\tilde{\sigma}^2 \xrightarrow{p} \sigma^2$, so $\tilde{\sigma} \xrightarrow{p} \sigma$.

3. Khintchine's Theorem -
(Weak Law of Large Numbers)

Suppose $\{X_i\}_{i=1}^n$ are drawn from the same dist'n. with mean μ and variance σ^2, & the X_i's are <u>uncorrelated</u>.

Then $\operatorname{plim}(\bar{X}_n) = \mu$.

4. Convergence in Mean Square -

If $E(\hat{\theta}_n) \to \theta$, and $\text{var.}(\hat{\theta}_n) \to 0$ as $n \to \infty$, then $\hat{\theta} \to \theta$ in mean square.

5. Mean Square consistency $\Rightarrow$ weak consistency
To prove this, we use <u>Chebyshev's</u> <u>Inequality</u>: X is random and $g(\cdot) \geq 0$.
Then, $\Pr.[g(X) \geq k] \leq E[g(X)]/k$
(all $k > 0$)

Consider the condition for weak consistency.

$$\Pr.\left[\,|S_n - c| < \varepsilon\,\right] = \Pr.\left[\,(S_n-c)^2 < \varepsilon^2\,\right]$$
$$= 1 - \Pr.\left[\,(S_n-c)^2 \geq \varepsilon^2\,\right]$$

and

$$\Pr\left[(S_n-c)^2 \geq \varepsilon^2\right] \leq E\,(S_n-c)^2/\varepsilon^2 \quad ; \text{ by Chebyshev.}$$

Now, as $n \to \infty$, $E\,(S_n - c)^2 \to 0$ if the statistic is mean-square consistent ; so, if MSC, then

$$\underset{n\to\infty}{\text{Limit}}\ \Pr\left[\,|S_n - c| < \varepsilon\,\right] = 1 - 0 = 1,$$

& we have weak consistency. #

(The converse isn't necessarily true.)

6. Fisher Consistency –

A statistic is "Fisher consistent" if, when calculated from the whole population, it is equal to the true parameter.

e.g. $\bar{x}$ is Fisher consistent for μ, but $\bar{x} + 1/n$ is not. (Both are weakly consistent.)

(No page 4)

7. Fisher consistency $\Rightarrow$ weak consistency

(converse need not be true.)

8. Strong ("Almost Sure") Consistency –

$\hat{\theta}_n$ is "strongly consistent" for θ (or, it converges "almost surely" to θ) if

$$\Pr.\left[\,\underset{n\to\infty}{\text{Limit}}\ \hat{\theta}_n = \theta\,\right] = 1.$$

This requires convergence for every sub-sequence, as well as the overall sequence.

We write $\hat{\theta}_n \xrightarrow{a.s.} \theta$.

(Sometimes say "converges with probability one to θ").

9. Strong convergence $\Rightarrow$ weak convergence.

10. Strong Law of Large Numbers –

This is the counterpart to Khintchine's Theorem : if $\{x_i\}_{i=1}^{n}$ are i.i.d. & $|\mu| < \infty$, then $\bar{x} \xrightarrow{a.s.} \mu$.

11. Convergence in Distribution –

Let $\{x_n\}$ be a sequence of random variables, & let $F_n(x)$ be the c.d.f. of x_n.

Then x_n converges in distribution to the random variable, x, with c.d.f. $F(x)$, if

$$\underset{n\to\infty}{\text{Limit}} \; |F_n(x) - F(x)| = 0.$$

We write, $x_n \xrightarrow{d} x$.

Example: $\{x_i\}_{i=1}^{n} \overset{iid}{\sim} N(\mu, \sigma^2)$

$$\bar{x} = \tfrac{1}{n}\sum_i x_i \; ; \quad s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$$

Then, $t = (\bar{x} - \mu)/(s/\sqrt{n}) \sim t_{n-1}$

Now, $\hat{\sigma}^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2 \xrightarrow{p} \sigma^2$

so $s^2 \xrightarrow{p} \sigma^2$.

$$t = \frac{(\bar{x} - \mu)/(\sigma/\sqrt{n})}{(s/\sigma)}$$

& $\text{plim}\,(s/\sigma) = 1$.

Also, $\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right) = z \sim N(0,1) \rightarrow N(0,1)$

So, $t \xrightarrow{d} (z/1) = z \sim N(0,1)$.

(We've actually used the more general result that if $x_n \xrightarrow{d} x$, & $y_n \xrightarrow{p} c$, then $x_n y_n \xrightarrow{d} cx$.)

12. Lindeberg – Lévy Central Limit Theorem –

Suppose $\{x_i\}_{i=1}^{n}$ are i.i.d. with finite mean & variance, μ & σ^2. Then:

$$\sum_{i=1}^{n} \frac{(x_i - \mu)}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1).$$

[i.e. $\bar{x} \xrightarrow{d} N(\mu, \sigma^2/n)$.]

Proof: The characteristic fctn. always exists,

$$\text{\& } \phi_x(t) = E(e^{itx}) = E\left[1 + itx + \frac{(itx)^2}{2!} + \cdots\right]$$

$$= 1 + it\,E(x) + \frac{(it)^2}{2!}E(x^2) + \cdots$$

$$= 1 + it\mu + \frac{(it)^2}{2!}\mu_2' + \cdots$$

(where $\sigma^2 = \mu_2' - \mu^2$).

Let $y_i = (x_i - \mu)/(\sigma\sqrt{n})$

So, $E(y_i) = 0$; $\text{var.}(y_i) = \frac{1}{n}$

$$\phi_y(t) = E(e^{ity})$$
$$= E\left[1 + ity + \frac{(it)^2}{2!}y^2 + \cdots\right]$$
$$= \left[1 + it(0) + \frac{(it)^2}{2!}E(y^2) + \cdots\right]$$

& $E(y^2) = var(y) + (E(y))^2 = (\frac{1}{n} - 0) = 1/n$.

So, $\phi_y(t) = 1 - t^2/2n + \cdots$. $(i^2 = -1)$.

When we sum several independent r.v.'s, the c.f. of the sum is the product of the individual c.f.'s. (all the same, here).

So, if $z = \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma\sqrt{n}}\right)$ then

$$\phi_z(t) = \prod_i \left[1 - t^2/2n + \cdots\right]$$
$$\approx \left[1 - t^2/2n\right]^n$$

So, $\log \phi_z(t) \approx n \log(1 - t^2/2n)$

& if n large, $|t^2/2n|$ small, &

$$\log(1+x) = x - x^2/2 + x^3/3 - \cdots \quad ; \; |x| < 1$$

So, $\log \phi_z(t) \approx n\{-t^2/2n - \cdots\} \approx t^2/2$

& $\phi_z(t) = e^{-t^2/2} = e^{(it)^2/2}$.

The only dist'n. whose c.f. is of this form is $N(0,1)$!

So, if n is large enough, $z \xrightarrow{d} N(0,1)$.

That is, $\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma\sqrt{n}}\right) = \frac{1}{\sigma\sqrt{n}}[n\bar{x} - n\mu]$

$$= \left(\frac{n(\bar{x} - \mu)}{\sigma/\sqrt{n}}\right)\left(\frac{1}{n}\right) = \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right)$$
$$\xrightarrow{d} N(0,1).$$

& so $\bar{x} \xrightarrow{d} N(\mu, \sigma^2/n)$. #.

Other Central Limit Theorems emerge if we relax the i.i.d. assumption to some degree.

13. Asymptotic Mean & Variance –

The asymptotic moments of x_n are just those of "F", where $x_n \xrightarrow{d} x$, & "F" is the c.d.f. of x.

So, asy. $E(x_n) = \int x f(x)\,dx$

where $f(x) = F'(x)$; etc.

If $\{x_n\}$ has moments for all "n", then the asymptotic moments are obtained by taking limit of finite-sample moments.

e.g. If $E(x_n)$ exists for all n, then

$$\text{asy. } E(x_n) = \underset{n \to \infty}{\text{Limit}}\; E(x_n).$$

Dangerous approach – moments may not exist for all n. Better to look at moments of limiting dist'n, as originally defined.

<u>Example</u> :

$$x \sim F(v_1, v_2)$$

Then, $E(x) = v_2 / (v_2 - 2)$; $v_2 > 0$

Let $y = v_1 x$, then $E(y) = v_1 v_2 / (v_2 - 2)$

Now, $y \xrightarrow{d} \chi^2_{(v_1)}$, & $E[\chi^2_{(v_1)} = v_1]$

i.e. $\text{asy. } E(y) = v_1$

N.B. : In our examples, $v_2 = n - k$, so $n \to \infty$ is the same as $v_2 \to \infty$, & note that

$$\underset{v_2 \to \infty}{\text{Limit}} \left[v_1 v_2 / (v_2 - 2) \right] = v_1 = \text{asy. } E(y).$$

Recall : If we have consistency, then limiting dist'n. is a "spike" – no variance

So, we scale by rate of convergence & ensure we have a non-degenerate limiting dist'n. by considering

$$\sqrt{n}\,(\hat{\theta}_n - \theta) \quad \text{rather than } \hat{\theta}_n \text{ itself.}$$

e.g.:

Suppose $\{x_i\}_{i=1}^{n} \sim \text{iid } N(\mu, \sigma^2)$.

Then $E(\bar{x}) = \mu$ & $\text{var.}(\bar{x}) = \sigma^2/n$

& $\bar{x} \xrightarrow{m.s.} (\mu)$, so $\bar{x} \xrightarrow{p} \mu$.

However, $\sqrt{n}(\bar{x} - \mu) = y$ is such that

$E(y) = 0$ & $\text{var.}(y) = \sigma^2$, so

$$y = \sqrt{n}(\bar{x} - \mu) \sim N(0, \sigma^2) \xrightarrow{d} N(0, \sigma^2).$$

Then, the $\text{asy. } E(y) = 0$, & $\text{asy. var}(y) = \sigma^2$

Finally,

$\text{asy. } E(\bar{x}) = \mu$, & $\text{asy. var.}(\bar{x}) = \sigma^2/n$,

if you wish !

Regression Example —

$$y = X\beta + \varepsilon \quad ; \quad \varepsilon \sim (0, \sigma^2 I)$$

Assume: $\text{plim}\left[\frac{1}{\sqrt{n}} X'\varepsilon\right] = 0$

$\text{plim}\left[\frac{1}{n} X'X\right] = Q$; finite, p.d.s.

OLS : $b = (X'X)^{-1}X'y$

$= \beta + (X'X)^{-1}X'\varepsilon$

$= \beta + \left[\frac{1}{n}X'X\right]^{-1}\left[\frac{1}{\sqrt{n}}X'\varepsilon\right]\left(\frac{1}{\sqrt{n}}\right)$

So, $\text{plim}(b) = \beta + Q^{-1}.0.0 = \beta$

& $b \xrightarrow{p} \beta$.

Also, $V(b) = \sigma^2(X'X)^{-1} = \left(\frac{\sigma^2}{n}\right)\left[\frac{1}{n}X'X\right]^{-1}$

& $V(b) \longrightarrow 0.Q^{-1} = 0$.

So, b is m.s. consistent for β.

Now, $\sqrt{n}(b-\beta) = \sqrt{n}\,(X'X)^{-1}X'\varepsilon$

$= \left[\frac{1}{n}X'X\right]^{-1}\left[\frac{1}{\sqrt{n}}X'\varepsilon\right]$

$\sqrt{n}(b-\beta) \xrightarrow{d}$ Normal, by C.L.T.

In fact,

$\sqrt{n}(b-\beta) \xrightarrow{d} N[0, \sigma^2 Q^{-1}]$.

Why $\sigma^2 Q^{-1}$?

$\left(\frac{1}{n}X'X\right)^{-1} \xrightarrow{p} Q^{-1}$

$\frac{1}{\sqrt{n}}X'\varepsilon \xrightarrow{d} N[0, \sigma^2 Q]$

So, asy. var. $[\sqrt{n}(b-\beta)] = Q^{-1}\sigma^2 Q\, Q^{-1\prime}$
$= \sigma^2 Q^{-1}$

(Q is symmetric)

$$\sqrt{n}(b-\beta) \xrightarrow{d} N[0, \sigma^2 Q^{-1}]$$

unobservable

Recall that, by Khintchine's Theorem,

$\text{plim}(s^2) = \sigma^2$; $s^2 = \frac{1}{n-k}e'e$.

So, a consistent estimator of $\sigma^2 Q^{-1}$ is

$n s^2 (X'X)^{-1}$. [check this]

A consistent estimator of asymptotic covariance matrix of $\sqrt{n}\,(b-\beta)$ is $n s^2 (X'X)^{-1}$, & square roots of diagonal elements give us asymptotic std. errors (of $\sqrt{n}\,(b-\beta)$).

The asymptotic std. errors of b itself are square roots of diagonal elements of $s^2 (X'X)^{-1}$.

[In this particular case the asy. std. errors are same as regular std. errors, but this is not generally the case.]

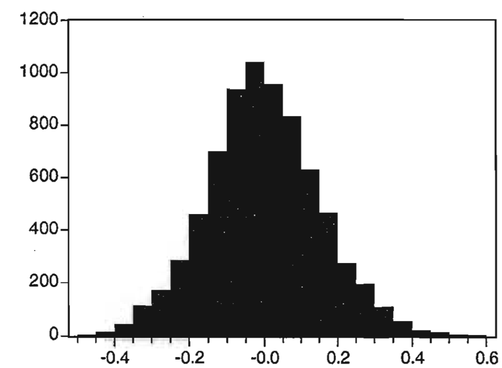
Regression With Independent Logistic Errors:

OLS

Dependent Variable: FOODSHR
Method: **Least Squares**
Date: 01/17/06 Time: 17:14
Sample: 1 7358
Included observations: 7358

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.597703	0.016358	97.67053	0.0000
LTEXP	-0.162177	0.002339	-69.33894	0.0000

R-squared	0.395259	Mean dependent var	0.469982
Adjusted R-squared	0.395177	S.D. dependent var	0.193339
S.E. of regression	0.150360	Akaike info criterion	-0.951291
Sum squared resid	166.3065	Schwarz criterion	-0.949414
Log likelihood	3501.798	F-statistic	4807.889
Durbin-Watson stat	1.365369	Prob(F-statistic)	0.000000



Series: RESID01
Sample 1 7358
Observations 7358

Mean	1.18e-16
Median	-0.005030
Maximum	0.592287
Minimum	-0.489737
Std. Dev.	0.150350
Skewness	0.151758
Kurtosis	3.163682
Jarque-Bera	36.45689
Probability	0.000000

Logistic Distribution

$p(\varepsilon_i) = [\exp(\varepsilon_i / k)] / \{k[1+\exp(\varepsilon_i / k)]\}$; $-\infty < \varepsilon_i < \infty$

$E(\varepsilon_i) = 0$; $Var.(\varepsilon_i) = (k^2\pi^2 / 3)$; Kurtosis > 3

(*Standardized Logistic has k = 1*)

```
@logl ll6
res=(foodshr-c(1)-c(2)*ltexp) /c(3)
ll6=log(@dlogistic(res))-log(c(3))
```

LogL: UNTITLED
Method: **Maximum Likelihood (Marquardt)**
Date: 01/17/06 Time: 17:22
Sample: 1 7358
Included observations: 7358
Evaluation order: By observation
Convergence achieved after 16 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	1.655224	0.017692	93.55964	0.0000
C(2)	-0.170727	0.002563	-66.59971	0.0000
C(3)	0.084928	0.000907	93.63624	0.0000

Log likelihood	3480.245	Akaike info criterion	-0.945160
Avg. log likelihood	0.472988	Schwarz criterion	-0.942346
Number of Coefs.	3	Hannan-Quinn criter.	-0.944193

Gradients

Gradient of objective function at estimated parameters
LogL: UNTITLED
Method: Maximum likelihood
Uses accurate numeric derivatives where necessary

Coefficient	Sum	Mean	Newton Dir.	Method
C(1)	0.410559	5.58E-05	-2.80E-05	- numeric -
C(2)	3.351856	0.000456	4.27E-06	- numeric -
C(3)	-1.510431	-0.000205	-1.55E-06	- numeric -

Final Example: $y_i = \beta X_i + \varepsilon_i$; $\varepsilon_i \sim (0, \sigma^2)$

Let $\hat{\beta} = (\bar{y}/\bar{x})$.

So, $\hat{\beta} = \sum_i y_i / \sum_i x_i = \sum_i (\beta x_i + \varepsilon_i) / \sum_i x_i$

$= \beta + (\sum_i \varepsilon_i / \sum_i x_i)$

$= \beta + \frac{1}{n\bar{x}} \sum_i \varepsilon_i$

So, $E(\hat{\beta}) = 0$

& $var.(\hat{\beta}) = \left(\frac{1}{n\bar{x}}\right)^2 \sum_i \sigma^2 = \frac{\sigma^2}{n\bar{x}^2} \to 0$.

So, $\hat{\beta}$ is m.s. consistent for β.

Now, $(\hat{\beta} - \beta) = \sum_i \left(\frac{\varepsilon_i}{n\bar{x}}\right) = \underbrace{\frac{1}{n}\sum_i z_i}_{=\bar{z}}$; $z_i = \varepsilon_i / \bar{x}$.

By the C.L.T : $\sqrt{n}(\bar{z} - 0) \xrightarrow{d} N[0, var.(z_i)]$

so, $\sqrt{n}(\bar{z} - 0) \xrightarrow{d} N[0, \sigma^2/\bar{x}^2]$

i.e. $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N[0, \sigma^2/\bar{x}^2]$

The a.s.d. $(\hat{\beta})$ is $\sigma/(\sqrt{n}\bar{x})$, & a consistent estimator of σ is

$\hat{\sigma} = \left[\frac{1}{n}\sum_i (y_i - \hat{\beta}x_i)^2\right]^{1/2}$, by Khintchine & Slutsky.

So, a.s.e. $(\hat{\beta}) = \hat{\sigma}/(\sqrt{n}\bar{x})$. #.