The Lindberg-Levy Central Limit Theorem

Let $\{X_1, ..., X_n\}$ be a simple random sample of size n, i.e. a sequence of n i.i.d random variables, drawn from a distribution X with $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2 < \infty$. We are interested in the asymptotic distribution of the sample average $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$.

Central Limit Theorem (CLT) (Lindberg-Levy): The sequence

$$Y_n = n^{1/2}\frac{\overline{X}_n - \mu}{\sigma} = \frac{\sum_{i=1}^{n}(X_i - \mu)}{n^{1/2}\sigma}$$

converges in distribution to $N(0,1)$, i.e. $Y_n \xrightarrow{d} N(0,1)$.

Notes

1. Identical means and variances can be dropped straightforwardly. We need some restrictions on the variance sequence though. In this case, we work with

$$Y_n = \frac{\sum_{i=1}^{n}(X_i - \mu_i)}{\left(\sum_{i=1}^{n}\sigma_i^2\right)^{1/2}}.$$

2. Versions of the Central Limit Theorem with random vectors are also available. Just apply univariate theorems to all linear combinations.

3. The basic requirement is that each term in the sum should make a negligible contribution.

Examples:

1. Estimation of mean μ from a sample of normal random variables: In this case, we estimate μ by $\bar{X}$, and the asymptotic approximation for the distribution of $\bar{X}$ or $(\bar{X} - \mu)$ is **exact**.

2. Consider $n^{1/2}(\hat{\beta} - \beta)$ where $\hat{\beta}$ is the LS estimator.

$$\begin{aligned} n^{1/2}(\hat{\beta} - \beta) &= n^{1/2}(X'X)^{-1}X'\varepsilon \\ &= [X'X/n]^{-1}n^{1/2}[X'\varepsilon/n] \end{aligned}$$

Where $[X'X/n]$ is the sample second moment matrix of the regressors.
$[X'X/n]$ is $O(1)$ or maybe $O_p(1)$depending on assumptions.
Its lim or plim is Q, a KxK p.d. matrix.

Regression Example Cont'd

What about $n^{1/2}[X'\varepsilon/n] = \sqrt{n}(1/n)\sum x_i'\epsilon_i$?

This is $\sqrt{n}$ times a sample mean of $x_i'\epsilon_i$.These have $Ex_i'\epsilon_i = 0$, $Vx_i'\epsilon_i = \sigma^2 Q$ (discuss)

Under the assumption that regressors are well-behaved (i.e., contribution of any particular observation to $[X'\varepsilon/n]$ is negligible), we can apply a Central Limit Theorem and conclude that

$$n^{1/2}(\hat{\beta} - \beta) = [X'X/n]^{-1} n^{1/2}[X'\varepsilon/n] \xrightarrow{D} N(0, \sigma^2 Q^{-1}).$$

Consistent with previous results?

Asymptotic testing

In this section, we will study the "trinity" of asymptotic tests: Likelihood Ratio (LR), Wald, and Lagrange Multiplier (LM) or Score tests

Background: let $\ell(\theta) = \ln L(\theta) = \ln \prod_{i=1}^{n} p(y_i|\theta) = \sum_{i=1}^{n} \ln p(y_i|\theta)$ be the log-likelihood function, where $[y_1, ..., y_N]$ is a random sample from the population $Y \sim p(y|\theta)$ and θ is an unknown vector of parameters. Then we define the Maximum Likelihood (ML) estimator

$$\widehat{\theta} = \arg\max \ell(\theta),$$

the score function of the likelihood function

$$s(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{1}{p(y_i|\theta)} \frac{\partial p(y_i|\theta)}{\partial \theta} \equiv \sum_{i=1}^{n} \frac{p'(y_i|\theta)}{p(y_i|\theta)} = \sum_{i=1}^{n} s_i(\theta),$$

and the information matrix $ni(\theta)$, where

$$i(\theta) = \mathrm{E}\left[\frac{\partial \ln p(y|\theta)}{\partial \theta}\right]^2 = \mathrm{E}\left[\frac{p'(y|\theta)}{p(y|\theta)}\right]^2$$

Notice that

$$\mathrm{E}s_i(\theta) = \int \frac{1}{p(y|\theta)} \frac{\partial p(y|\theta)}{\partial \theta} p(y|\theta)\mathrm{d}y = \frac{\partial}{\partial \theta} \underbrace{\int p(y|\theta)\mathrm{d}y}_{=1} = 0,$$

and hence

$$\mathrm{V}s_i(\theta) = i(\theta)$$

Moreover

$$-\mathrm{E}\left[\frac{\partial^2 \ln p(y|\theta)}{\partial^2 \theta}\right] = -\mathrm{E}\left[\frac{\partial}{\partial \theta} \frac{p'(y|\theta)}{p(y|\theta)}\right] = -\underbrace{\mathrm{E}\left[\frac{p''(y|\theta)}{p(y|\theta)}\right]}_{\frac{\partial^2}{\partial^2 \theta} \int p(y|\theta)\mathrm{d}y = 0} + \underbrace{\mathrm{E}\left[\frac{p'(y|\theta)}{p(y|\theta)}\right]^2}_{i(\theta)}$$

Asymptotic Testing - cont'd

Since the individual scores $s_i(\theta)$ are i.i.d. (*why*?), by CLT

$$n^{-1/2} s_0 \xrightarrow{d} N(0, i_0)$$

where $s_0 = s(\theta_0)$, $i_0 = i(\theta_0)$ and θ_0 is the population value of θ. Moreover, under mild regularity conditions on $\ell(\theta)$, it can be proved that

$$n^{1/2} d_0 \xrightarrow{d} N(0, i_0^{-1})$$

where $d_0 = \widehat{\theta} - \theta_0$ denote the vector of deviations.
Testing: Since $n^{-1/2} s_0$ is asymptotically equivalent to $n^{1/2} i_0 d_0$, we get

$$n^{1/2} d_0 \overset{\text{asy}}{=} n^{-1/2} i_0^{-1} s_0$$

Further, we have that

$$2[\ell(\widehat{\theta}) - \ell(\theta_0)] \overset{\text{asy}}{=} n d_0' i_0 d_0,$$

which can be proved by expanding $\ell(\theta_0)$ around $\widehat{\theta}$ and taking probability limits.

Asymptotic Testing - cont'd

Consider the hypotheses

$$\mathrm{H}_0 : \theta = \theta_0 \quad \mathrm{H}_1 : \theta \neq \theta_0$$

Likelihood Ratio Test: It is based on $LR = L(\theta_0)/L(\widehat{\theta})$. The test statistic

$$-2 \ln LR = 2[\ell(\widehat{\theta}) - \ell(\theta_0)] \xrightarrow{d} \varkappa^2(q),$$

where q is the number of restrictions under H_0.
Wald test: Under H_0 the test statistic

$$W = nd_0' i(\widehat{\theta}) d_0 \xrightarrow{d} \varkappa^2(q),$$

Lagrange Multiplier (or **Score**) **test**: Under H_0 the test statistic

$$LM = n^{-1} s_0' i_0^{-1} s_0 \xrightarrow{d} \varkappa^2(q)$$

Asymptotic Testing - cont'd

Remark: Given that

$$\text{p}\lim i(\widehat{\theta}) = i_0$$

when the null hypothesis is true and that

$$\text{p}\lim(n^{1/2} d_0 - n^{-1/2} i_0^{-1} s_0) = 0,$$

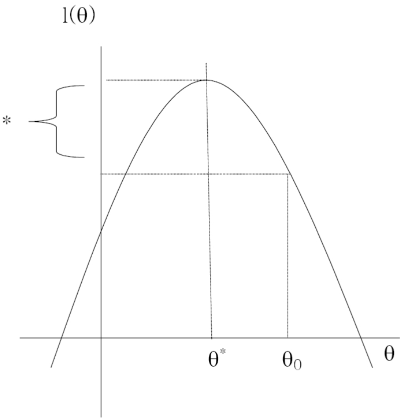
the tests are asymptotically equivalent. Note that the Wald and LM tests are appealing because of their asymptotic equivalence to the LR test, which is an optimal test in the Neyman-Pearson sense.

Discussion:

- What are the computational requirements for these tests?
- Which is best?

For illustrative purposes, θ is one-dimensional.

Here, we look at the change in the log likelihood function $\ell(\theta)$ evaluated at $\hat{\theta}$ and θ_0, $\ell(\hat{\theta})$ and $\ell(\theta_0)$. If the difference between is too large, we reject H_0.



$*$ LR test is based on this difference

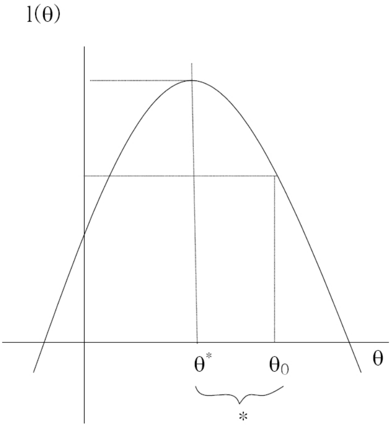
Geometry

Wald Test

Here, we look at the deviation in parameter space.

The difference between $\hat{\theta}$ and θ_0 implies a larger difference between $\ell(\hat{\theta})$ and $\ell(\theta_0)$ for the more curved log likelihood function. Evidence against the hypothesized value θ_0 depends on the curvature of the log likelihood function measured by $ni(\hat{\theta})$.

Hence the test statistic is $n(\hat{\theta} - \theta_0)^2 i(\hat{\theta})$.

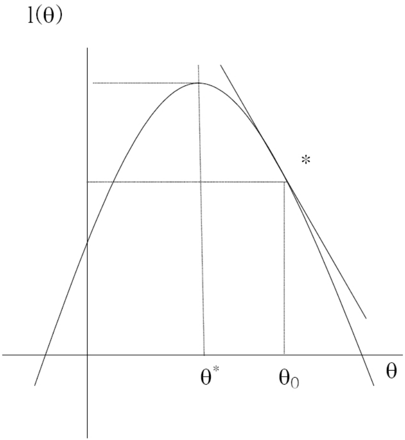


∗ Wald test is based on this difference

Here, we look at the slope of the log likelihood function at the hypothesized value of θ_0.

Since two log likelihood functions can have equal values of s_0 with different distances between $\hat{\theta}$ and θ_0, s_0 must be weighed by the change in slope (i.e. curvature). A bigger change in slope implies less evidence against the hypothesized value θ_0.

Hence the test statistic $n^{-1}s_0^2 i_0^{-1}$.



$*$ Score(LM) test is based on this difference

Why is the score test also called the Lagrange Multipler test?

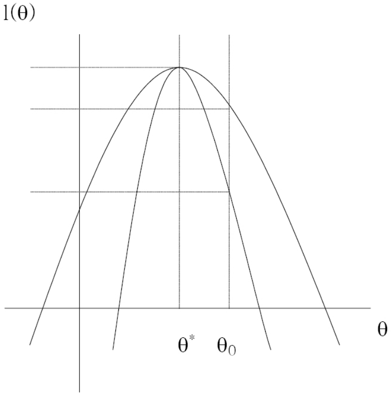
The log likelihood function is maximzied subject to the restriction $\theta = \theta_0$:

$$\max_{\theta} \ell(\theta) - \lambda(\theta - \theta_0).$$

This gives

$$\hat{\theta} = \theta_0 \text{ and } \lambda = s(\theta_0) = \frac{\partial \ell}{\partial \theta_0}.$$

$*$ 2 likelihood functions and a test of $\Theta = \Theta_0$



Asymptotics testing in the Gaussian regression model

- The parameters are $\theta = (\beta', \sigma^2)'$
- The log-likelihood function is

$$\ell(\theta) = -\frac{N}{2}\left(\ln \pi + \ln \sigma^2\right) - \frac{1}{2}\sum_{i=1}^{N}\frac{(Y_i - X_i'\beta)^2}{\sigma^2}$$

- The log-likelihood contribution of the $i-$th observation is

$$\ell_i(\theta) = -\frac{1}{2}\left(\ln \pi + \ln \sigma^2 + (Y_i - X_i'\beta)^2/\sigma^2\right)$$

- The individual scores are given by the first partial derivatives

$$s_i(\theta) = \begin{bmatrix} \partial \ell_i(\theta)/\partial \beta \\ \partial \ell_i(\theta)/\partial \sigma^2 \end{bmatrix} = \begin{bmatrix} X_i(Y_i - X_i'\beta)/\sigma^2 \\ -\frac{1}{2\sigma^2}\left(1 - (Y_i - X_i'\beta)^2/\sigma^2\right) \end{bmatrix}$$

Asymptotics testing in the Gaussian regression model - cont'd

- The score is given by

$$s(\theta) = \sum_{i=1}^{n} s_i(\theta) = \left[\begin{array}{c} X'(Y - X\beta)/\sigma^2 \\ -\frac{1}{2\sigma^2}\left(N - (Y - X\beta)'(Y - X\beta)/\sigma^2\right) \end{array} \right]$$

- The information matrix is $N \times i(\theta)$ where $i(\theta)$ is given by

$$-\mathrm{E}\left[\begin{array}{cc} -X_iX_i'/\sigma^2 & -X_i\varepsilon_i/\sigma^4 \\ -X_i'\varepsilon_i/\sigma^4 & \frac{1}{2\sigma^4}(1 - 2\varepsilon_i^2/\sigma^2) \end{array} \right] = \left[\begin{array}{cc} \mathrm{E}\left(X_iX_i'/\sigma^2\right) & 0_K \\ 0_K' & 1/2\sigma^4 \end{array} \right],$$

which can be estimated by

$$\left[\begin{array}{cc} X'X/N\widehat{\sigma}^2 & 0_K \\ 0_K' & 1/2\widehat{\sigma}^4 \end{array} \right]$$

The Wald test in regression models

If the restrictions $R\beta = r$ are valid, then the quantity $R\widehat{\beta} - r$ should be close to 0. The Wald test statistic is

$$W = (R\widehat{\beta} - r)' \left[\mathrm{Var}(R\widehat{\beta} - r) \right]^{-1} (R\widehat{\beta} - r) \xrightarrow{d} \chi^2(q)$$

where

$$\mathrm{Var}(R\widehat{\beta} - r) = \sigma^2 R(X'X)^{-1} R'$$

- When σ^2 is substituted by its unbiased estimator in the unrestricted model, $W = qF$, where F is the F-test statistic for $H_0 : R\beta = r$ (is this consistent with the distributional results on W and F?).
- The Wald test requires to estimate the unrestricted model. In some cases, this is cumbersome.

The score (LM) test in regression models

LM test are often written as NR^2 in an auxiliary regression. Here we see why. Define the matrix

$$S_0 = [s_{0,1}, ...s_{0,N}]'$$

where

$$s_{0,i} = \frac{\partial \ln p(Y_i|X_i;\theta_0)}{\partial \theta}$$

is the i-th individual score ($i = 1, ..., N$) under the null hypothesis. Than we can write the LM test as

$$LM = 1_N' S_0 (S_0' S_0)^{-1} S_0' 1_N$$

where $1_N = 1 \in R^N$, $s_0 = S_0' 1_N$ and i_0 is estimated by $S_0' S_0 / N$.

- Consider the auxiliary regression model

$$1_N = S_0\gamma + \text{errors}$$

- The OLS estimator of γ and the fitted values of 1_N are respectively given by

$$\widehat{\gamma} = (S_0'S_0)^{-1}S_0'1_N, \quad \widehat{1}_N = S_0(S_0'S_0)^{-1}S_0'1_N$$

- The LM test can be written as

$$LM = \widehat{1}_N'\widehat{1}_N = N\frac{\widehat{1}_N'\widehat{1}_N}{1_N'1_N} = NR^2$$

where R^2 is the determination coefficient of the auxiliary regression.

LM test for omitted variables

- Suppose to run a LM test for $H_0 : \beta_2 = 0$ the unrestricted regression model

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

- Consider the estimated restricted model

$$Y = X_1 b_1 + e$$

- The LM test can be written as NR^2 in the auxiliary regression

$$e = X_1\gamma_1 + X_2\gamma_2 + \text{errors}$$

The LR test in regression models

- For the LR test, we estimate the model both under H_0 and under H_1. The restricted model must be nested in the unrestricted one.
- The LR test statistic is given by

$$LR = 2\left(\ell(\widehat{\beta}, \widehat{\sigma}^2) - \ell(\widehat{\beta}_0, \widehat{\sigma}_0^2)\right) = N \ln\left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}\right)$$

  where $\widehat{\beta}_0$ and $\widehat{\sigma}_0^2$ are, respectively, the ML estimates of β and σ^2 from the restricted model.
- Under H_0, LR is asymptotically distributed as $\chi^2(q)$, where q is the number of restrictions imposed to the restricted model
- The test is insensitive to linear transformations of the models and of the restrictions.