Athens University of Economics and Business
Department of Economics

Postgraduate Program - Master's in Economic Theory
Course: Econometrics II
Prof: Stelios Arvanitis
TA: Alecos Papadopoulos

Semester: Spring 2017-2018

March 22, 2018

Gordin's Central Limit Theorem.

(see Hayashi book, pp. 402-405 and references therein)

A. Statement: Suppose $(y_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic, with autocovariance function γ_k. Suppose also that "**Gordin's condition**" is satisfied. Then:

(i) $E(y_t) = 0$

(ii) $\sum_{k=0}^{\infty} |\gamma_k| < +\infty$

(iii) $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_t \xrightarrow{d} \mathrm{N}\left(0, \sum_{k=-\infty}^{\infty} \gamma_k\right) = \mathrm{N}\left(0, \gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k\right)$

...the second equality since the autocovariance function is symmetric around $k = 0$.

B. Gordin's Condition.

It has three parts.

B.1 GC-1. $E\left(y_t^2\right) < \infty$ **.** The second moment exists and is finite.

B.2 GC-2. Let $f_{t-i} = \sigma(y_{t-i}, i \geq 0)$, a filtration containing the history of the process. Then GC-2 requires

$$E\left(y_t \,|\, f_{t-i}\right) \xrightarrow{m.s} 0, \quad i \longrightarrow \infty$$

...namely, that this conditional expected value **converges in mean-square to zero** (and so also in probability and so also in distribution -see Hayashi ch. 2), as the filtration is "reduced in size". **This condition implies result (i), namely that** $E(y_t) = 0$ **.**

PROOF. Convergence of $E(y_t \mid f_{t-i})$ in mean-square to zero means $\lim_{i\to\infty} E\left(\left[E(y_t \mid f_{t-i})\right]^2\right) = 0$.

This in turn implies

$$\lim_{i\to\infty} E\left(\left|E(y_t \mid f_{t-i})\right|\right) = 0$$

(because convergence in the r-th mean implies convergence in mean of lower-orders).

It follows that $\forall \varepsilon > 0,\ \exists M(\varepsilon):\ 0 \le E\left(\left|E(y_t \mid f_{t-i})\right|\right) < \varepsilon \ ,\ \forall\, i \ge M(\varepsilon)$

Note that, the absolute value function is a convex function:

$$g(x) = |x|, \quad \left|\lambda x_1 + (1-\lambda)x_2\right| \le \lambda|x_1| + (1-\lambda)|x_2| \quad \lambda \in [0,1]$$

When random variables are involved, by Jensen's Inequality this implies $g(E(x)) \le E(g(x))$. In our case, $x \equiv E(y_t \mid f_{t-i})$ and so we obtain

$$0 \le \left|E\left(E(y_t \mid f_{t-i})\right)\right| \le E\left(\left|E(y_t \mid f_{t-i})\right|\right) < \varepsilon$$

But by the Law of Iterated Expectations (the "Tower Property"),

$$E\left(E(y_t \mid f_{t-i})\right) = E(y_t) \le \left|E(y_t)\right|$$

So we have obtained

$$0 \le E(y_t) \le |E(y_t)| < \varepsilon$$

Since ε was arbitrary, $E(y_t)$ is sandwiched to zero. QED.

The intuition here is that as we form the conditional expectation based on "less and less information", eventually our conditional expectation will end up being equal to the "no-specific information" expectation, i.e. the unconditional expected value.

B.3 GC-3.

Preparation. Apply add-and-subtract

$$y_t = y_t - \left[E(y_t|f_{t-1}) - E(y_t|f_{t-1})\right] - \left[E(y_t|f_{t-2}) - E(y_t|f_{t-2})\right] - \\ \dots - \left[E(y_t|f_{t-j}) - E(y_t|f_{t-j})\right]$$

Note that $y_t = E(y_t|f_t)$. Re-arrange also

$$y_t = \left[E(y_t|f_t) - E(y_t|f_{t-1})\right] + \left[E(y_t|f_{t-1}) - E(y_t|f_{t-2})\right] + \left[E(y_t|f_{t-2}) - E(y_t|f_{t-3})\right] \\ + \left[E(y_t|f_{t-j+1}) - E(y_t|f_{t-j+1})\right] + E(y_t|f_{t-j})$$

Note that the last term is single. Compact into

$$y_t = r_{t0} + r_{t1} + \dots + r_{t,j-1} + E(y_t|f_{t-j}), \quad r_{t,m} \equiv E(y_t|f_{t-m}) - E(y_t|f_{t-m-1})$$

We can think of this $r_{t,m}$ as the "revision of expectation" as the information set increases from $E(y_t|f_{t-m-1})$ to $E(y_t|f_{t-m})$.

Re-arranging,

$$y_t - r_{t0} - r_{t1} - \dots - r_{t,j-1} = E(y_t|f_{t-j})$$

Due to GC-2, the right-hand-side converges in mean-square to zero so also

$$y_t - r_{t0} - r_{t1} - \ldots - r_{t,j-1} \xrightarrow{m.s} 0 \Rightarrow y_t - \sum_{i=0}^{\infty} r_{t,i} \xrightarrow{m.s} 0$$

This implies that we can write y_t as a so-called "telescoping sum",

$$y_t = \sum_{i=0}^{\infty} r_{t,i}$$

Then, GC-3 requires

$$\sum_{i=0}^{\infty} \left[E\left(r_{t,i}^2\right)\right]^{1/2} < +\infty, \quad \forall t$$

Note that $E\left(r_{t,m}\right) = 0$, so $E\left(r_{t,i}^2\right) = \mathrm{Var}\left(r_{t,i}\right)$ and therefore GC-3 can also be written

$$\sum_{i=0}^{\infty} SD\left(r_{t,i}\right) < +\infty, \quad \forall t$$

namely that the infinite series of **standard deviations** of the forecast revisions converges. Is there any intuition here?

B.4 Discussion.

A necessary condition for GC-3 to hold is that

$$SD\left(r_{t,i}\right) \longrightarrow 0,\ i \to \infty \quad i.e. \quad \lim_{i\to\infty} SD\left[E\left(y_t \mid f_{t-i}\right) - E\left(y_t \mid f_{t-i-1}\right)\right] = 0$$

(although convergence of the sequence to zero is not necessarily monotonic).

GC-3 tells us that as we move further away from *t*, the variability in the expectation revision due to increasing information from a period to the next eventually should tend to zero (again, as a necessary but not sufficient condition for GC-3). This could be expected from GC-2, since both conditional expectations converge to the same constant as $i \to \infty$ and so their difference tends to the constant zero and so has zero variance at the limit. Informally, eventually the lower is the overall level of information the *less* is the variability in the change of the conditional expectation.

But also, this may appear counter-intuitive: after all, when we have little information, every added piece of information should "count for more" (and so lead to larger revisions / larger variances), compared to when we have a lot of information already (larger sigma algebras) as we move closer to the *t*-instance.

But this "intuitive" objection is misguided, *because it does not take into account the fact that the sequence of expectation revisions is correlated.* This means that past variability feeds to a degree into current variability, and so as we move closer to *t* we "accumulate variance" on y_t.

And what GC-3 does is to put exactly an overall restriction on the "degree of dependence" that we can accommodate and still obtain a Central Limit Theorem with the "usual" structure (i.e. a Normal distribution with a finite variance). And this necessarily translates initially into restrictions on the variances/standard deviations of the components of the telescoping sum.

C. An example to show the need for restrictions on correlation.

Consider a sequence of zero-mean, identical-variance, *equicorrelated* random variables $\mathrm{Cov}(x_t, x_{t-k}) = \mathrm{Cov}(x_t, x_{t-m}) = \rho, \ \forall k, m$. So correlation remains of equal strength no matter how "far apart" are the variables in the sequence. Consider the variance of their sum,

$$\mathrm{Var}\left(\sum_{i=1}^{n} x_i\right) = n\,\mathrm{Var}(x_0) + n(n-1)\rho$$

If we try to scale the sum by $1/\sqrt{n}$ *only* we will get

$$\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i\right)=\operatorname{Var}\left(x_0\right)+\left(n-1\right)\rho \longrightarrow \infty$$

To obtain a finite variance we need to scale by $1/n$ and then

$$\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n}\operatorname{Var}\left(x_0\right)+\left(1-1/n\right)\rho \longrightarrow \rho$$

This tells us that even the sample mean of equicorrelated random variables converges to a random variable. We do not discuss which may be the distribution here, the above was just to show that we need to impose restrictions on the correlation strength in order to obtain an applicable CLT.

Philosophical musings: We see that we need to impose restriction on the ***memory*** of a process, in order for the tools we have available to be put to use. The question is: are we imposing an artificial property on the real-world phenomena, that will lead to misleading results?

Well, both historical experience and intuition say that, as the distance between two entities (whatever these entities might be) increases (either in time or in space), their dependence also weakens and becomes negligible to non-existent. Therefore, the general modeling approach to allow for dependence which nevertheless weakens with distance appears to be a reasonable model of real-world phenomena and relations.

More difficult to defend is the need to impose restrictions on the ***heterogeneity*** of a process. While we can appeal to some universal notion of "inertia" to argue that real-world structures (either physical or social-economic) do not/cannot/don't change often, still, they do change, according to experience. While the concept of weak stationarity allows for higher-order heterogeneity, and while the concept of "integrated processes" and "co-integration" allow for a degree and form of non-stationarity that can be handled, still, even assuming *constant coefficients* in a process may be questioned, *if we have the intention to* *project this process too far in the past or too far in the future.* We should keep this in mind and moderate our ambitions.

Note: such cases in principle can be handled by modeling "slowly varying/evolving" coefficients, but things become way more complex in such a case.

D. An application.

Consider the AR(1) process

$$y_t = ay_{t-1} + u_t, \quad |a| < 1, \; u_t \sim \text{i.i.d. WN}\left(\sigma_u^2\right)$$

We know that GC-1 is satisfied.

For GC-2 we have

$$E\left(y_t \,|\, f_{t-i}\right) = E\left(y_t \,|y_{t-i}, y_{t-i-1}, \ldots\right) = E\left(ay_{t-1} + u_t \,|y_{t-i}, y_{t-i-1}, \ldots\right)$$

$$= aE\left(y_{t-1} |y_{t-i}, y_{t-i-1}, \ldots\right) + E\left(u_t \,|y_{t-i}, y_{t-i-1}, \ldots\right) = \; aE\left(y_{t-1} |y_{t-i}, y_{t-i-1}, \ldots\right)$$

Recursively,

$$E\left(y_t \,|\, f_{t-i}\right) = aE\left(ay_{t-2} + u_{t-1} |y_{t-i}, y_{t-i-1}, \ldots\right) = a^2 E\left(y_{t-2} \,|y_{t-i}, y_{t-i-1}, \ldots\right)$$

etc

$$\ldots E\left(y_t \,|\, f_{t-i}\right) = a^i E\left(y_{t-i} \,|y_{t-i}, y_{t-i-1}, \ldots\right) = a^i y_{t-i}$$

GC-2 requires $\qquad E\left(y_t \,|\, f_{t-i}\right) \xrightarrow{m.s} 0, \quad i \longrightarrow \infty$

which in our case becomes the requirement that

$$a^i y_{t-i} \xrightarrow{m.s} 0, \quad i \longrightarrow \infty \;\Rightarrow\; \lim_{i\to\infty} E\left[\left(a^i y_{t-i}\right)^2\right] = 0 \Rightarrow \lim_{i\to\infty}\left(a^{2i}\cdot E\left(y_{t-i}^2\right)\right) = 0$$

From GC-1 we have that $E\left(y_{t-i}^2\right) < +\infty$ or $E\left(y_{t-i}^2\right) = O(1)$. So

$$\lim_{i\to\infty}\left(a^{2i}\cdot E\left(y_{t-i}^2\right)\right) = O(1)\cdot \lim_{i\to\infty}\left(a^{2i}\right) = 0 \text{ since } |a| < 1.$$

So **GC-2 is also satisfied**.

Turning to the **GC-3** condition, we have for the components of the telescoping sum

$$r_{t,i} \equiv E\left(y_t \,|\, f_{t-i}\right) - E\left(y_t \,|\, f_{t-i-1}\right) = a^i y_{t-i} - a^{i+1} y_{t-i-1}$$

$$= a^i\left(y_{t-i} - a^i y_{t-i-1}\right) = a^i\left(a y_{t-i-1} + u_{t-i} - a^i y_{t-i-1}\right) = a^i u_{t-i}$$

So $\left[E\left(r_{t,i}^2\right)\right]^{1/2} = \left[E\left(a^{2i} u_{t-i}^2\right)\right]^{1/2} = |a|^i \sigma_u$ and GC-3 requires

$$\sum_{i=0}^{\infty} |a|^i \sigma_u < +\infty \Rightarrow \sigma_u \sum_{i=0}^{\infty} |a|^i = \frac{\sigma_u}{1-|a|} < +\infty$$

which holds. So Gordin's Condition is satisfied and the relevant CLT is applicable.

E. Exercises.

1) Calculate the variance of the limiting distribution of

$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_t$ where y_t is AR(1) as in the application above.

2) Derive the general theoretical expression for the variance of the limiting distribution in Gordin's CLT.

3) Show how Gordin's Condition implies absolute summability of the autocovariance function series.

4) Suppose that you were given the process ("AR(1) with drift")

$$y_t = \mu + ay_{t-1} + u_t, \quad |a| < 1, \; u_t \sim \text{i.i.d. WN}\left(\sigma_u^2\right), \quad \mu \in \mathbb{R} \setminus \{0\}$$

Examine whether Gordin's Condition holds. If it does, calculate the variance of the limiting distribution. If it does not, can you apply a transformation to obtain a process for which it holds?

5) Assume that we are given an AR(1) process without drift, but we are told that the process starts at $t = 0$ with y_0 given. Examine Gordin's Condition. If it holds, calculate the variance of the limiting distribution.

--