607
Notions of well-behaved series, in general; well behave time series, in particular

Jon Faust

http://e105.org/e607

September 6, 2016

- ▶ **Readings for this lecture**
  - The course website lists a number of good general time series references

    Both web stuff, e.g., Hamilton's text
  - Hansen's text ch. 12 (through asymptotics; stopping before bootstrap) is good
- ▶ **More advanced on CLTs and WLLNs**
  - For a deeper discussion of time series concepts and their use in in proving WLLNs and CLTs, you can't beat

    McFadden's notes (ch. 4) go

    http://elsa.berkeley.edu/users/mcfadden/e240a%5Fsp01/ch4.pdf
  - And Hal White's book is a classic

    Asymptotic theory for econometricians
- ▶ **Goal**
  - For many of the results in econometrics, we need that a stochastic series is sufficiently well behaved for a WLLN or CLT to apply to the sample mean.
  - This lecture is about what we mean by sufficiently well behaved.
  - We start out in the general case (not specifically time series) and then turn to a bunch of senses of well behaved for time series.

▶ **Background**

- CLTs require a bit more good behavior than WLLNs, so we'll mainly focus on CLTs
- In the general case, we usually end up structuring the good behavior along three features:
  - Limits on the thickness of the tails of the marginal distributions of the individual underlying random variables
  - Limits on how heterogeneous the different marginal distributions are.
  - Limits on the speed of decay of time series dependence: any relation of x_t with x_s must decline sufficiently rapidly as $|t - s|$ increases.

    The distant past cannot have too big of an effect on the current.
- Let me give a bit of intuition for each of these three

▶ **Notation.**

- We'll have y_t, $t = 1, 2, \ldots$
- $Ey_t = \mu$
- $\bar{y}_T$ is the sample mean in a sample of size T and we'll usually leave the T implicit.
- We say that a WLLN applies if,
$$\bar{y} \to_p \mu$$
- We'll say a CLT applies if,
$$\sqrt{T}(\bar{y} - \mu) \to_d N(0, \Omega)$$
  for some finite Ω.

▶ **CLT intuition**

- If a CLT is to apply to $\sqrt{T}(\bar{y} - \mu)$, it must be the case that as more observations are added to the sample mean, unruly behavior of early observations is offset or 'averaged out' leaving a nice Gaussian behavior for the average
- Further, if the sample mean is to be nicely behaved for some large N, it must be the case that for Ns this large, the $(N + 1)^{th}$ observation can't affect the mean very much.

  (Or can only do so with very low probability)
- Otherwise, we'd have to wait for later observations to offset this big effect in order to get to a well-behaved Gaussian value.

▶ **A first example**

- Put an extreme limit on time series dependence: assuming that the y_ts are independent.

- In this case, an upper bound on the variances is a sufficient restriction on tail thickness, and a lower bound on the variances is a sufficient restriction on heterogeneity.
- That is, independence plus

$$0 < \underline{\sigma}^2 < \sigma_t^2 < \bar{\sigma}^2 < \infty$$

is enough for a CLT.

▶ **Aside:: Huh? Why a lower bound on variance?**

- The lower bound on the variances essentially rules out the case where the variance of y_t converges to zero as t increases.
- If the variance converges to zero, later observations cease to have variability that can help average out the variability in earlier observations.

▶ **The Lindberg condition**

- The upper and lower bound on variance just stated sufficient for a CLT in the independent case, but is stronger than needed.
- The Lindberg condition (see McFadden) is a necessary and sufficient condition for a CLT to apply.

▶ **Time series**

- Once we drop the independence assumption, we will continue to need something like tail thickness and heterogeneity restrictions
- We will additionally require that observations farther apart be less related.

  Once again the intuition should be clear: we need enough, but not too much, 'fresh' information content in successive observations
- More subtley: there is an interplay among all three types of restrictions

  for example, the more we limit dependence, the more we can allow thicker tails or heterogeneity
- Thus, there will be no single set of necessary and sufficient conditions, rather there will be many different sets of sufficient conditions.

▶ **Well-behaved time series**

- Before returning to specific CLTs and WLLNS at the end, we turn to a bunch of senses of well-behaved time series that we will appeal to in various contexts.

  Covariance-stationarity, strict stationarity, Markov process, Martingale, Ergodic, Mixing, and a few further ideas

- **To begin. . .**
  - Define general linear model
  - And: Autocovariance, autocorrelation, and sample versions of same
  - And some related bits and drabs
  - Defn: general lin. model

$$y_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$$

  - ε_t, mean 0, variance $0 < \sigma_\varepsilon^2 < \infty$, $E\varepsilon_t\varepsilon_s = 0$ if $s \neq t$.
  - At times will strenthen to ε iid

- **Properties**
  - $y_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$
  - Note we are building up a time series as a simple function (moving average) of things we understand—uncorrelated random variables.
  - If the as decline rapidly enough with k, this time series process is very well behaved

- **Time series**
  - In the iid case, $a_0 = 1$ and all other as are zero.

    That is, all but one a_j are zero.

  - Essentially in time series we learn that most things that are true when all but one a are zero remain true so long as only finitely many a_j are 'large' and the remainder are 'arbitrarily close' to zero.

- **Let's examine the linear model**

- **Mean**
  - $y_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$
  - Mean?

    $Ey_t = 0$ for all t

  - Variance?

$$\text{var}(y_t) = \sigma_\varepsilon^2 \sum_{k=0}^{\infty} a_k^2 \quad \text{for all } t$$

  - So long as the sum converges

- If the sum converges, then infinitely many as must be arbitrarily close to zero and only finitely many are larger in abs. val. than any given bound.

▶ **Aside::** ℓ^2, L^2, **Hilbert spaces, and the frequency domain**

- Thus, any sequence of square summable numbers defines a covariance stationary general linear model.
- By definition, the set of all square summable sequences is what kind of space?

  Hilbert Space.

- Much of time series at this level is just applications of std. Hilbert space results

▶ **Aside::**

- Usually, the Hilbert space of square summable sequences is designated ℓ^2
- And a basic result is that in essentially all respects ℓ^2 is identical to L^2, the space of square integrable functions, L^2.
- The mapping from ℓ^2 to L^2 takes us from the 'time domain' into the 'frequency domain'

  fourier transforms and their inverses take us back and forth

- In the time domain, we talk of autocovariances; in the frequency domain we talk of the spectrum of the time series.
- Why do we go back and forth?

  Because some statements we want to make are easier stated or proven in terms of autocovariances, others are easier stated in terms of properties of the spectrum.

▶ **Define autocovariance**

- Autocovariance at lag k is the covariance of y_t with y_{t-k}.

$$\sigma_k = E(y_t - Ey_t)(y_{t-k} - Ey_{t-k}) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} a_j a_{j-k}$$

- In more general case, autocov. might vary with t.
- For gen. lin. model, each autocovariance is the same for all t.
- If variance is finite ($a \in \ell^2$), σ_k finite and goes to zero as k increases

  See that you can verify this claim

▶ **Sample autocovariance**

- The sample analog of population autocovariance is defined as,

$$\hat{\sigma}_k = (T-k)^{-1} \sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})$$

- Thus, $\hat{\sigma}_0$ is the sample variance.
- Note in any sample, you have 1 more obs. to use in computing $\hat{\sigma}_k$ than in $\hat{\sigma}_{k+1}$.
- This initial condition problem (needing k initial obs. to calculate $\hat{\sigma}(k)$) is a major pain in theory and practice
- But not asymptotically, because $T \approx T-k$ asymptotically (in the relevant sense)

▶ **Define autocorrelation**

- Define autocorrelation:

$$\gamma_k = \sigma_k / \sigma_0^2$$

- Sample version:

$$\hat{\gamma}_k = \hat{\sigma}_k / \hat{\sigma}_0$$

▶ **Define: Covariance-stationary**

- Covariance-stationary. A time series y_t is covariace-stationary iff,
  i) $E[y_t] = \mu$ for all t.
  ii) $\text{var}(y_t) \equiv \sigma_0 \equiv \sigma_y^2 < \infty$ for all t.
  iii) $\text{cov}(y_t, y_{t-k}) = \sigma_k$ for all t, k.
  (note: we are ignoring other deterministic elements than μ)

▶ **Language**

- cov-stat. also called weakly stationary.

▶ **Intimate relations**

- There are intimate relations b/t cov. stat. processes and the gen. lin. model
- The gen. lin. model is covariance stationary so long as the variance is finite ($\sum a^2 < \infty$)
- Moreover, every cov. stat. process can be written in the form of the gen. lin. model

(Wold Decomposition Thm.)

▶ **Wold decomp. thm.**

- (Ignoring any linearly deterministic part), every covariance stationary process has a representation:

$$y_t = \sum_{i=0}^{P} \theta_i \varepsilon_{t-i}$$

with...

- $E\varepsilon_t = 0 \quad E\varepsilon_t^2 = \sigma_\varepsilon^2$
- $\text{cov}(\varepsilon_{t+i}|y_t) = 0$ for all $i > 0$.
- The order of the process, P, may be infinite.
- With an additional restriction called 'invertibility' (to be discussed later) the representation is unique.

▶ **Note on sufficiently well behaved**

- All three aspects of good behavior are imposed by the assumption of cov. stationarity

  tail thickness is limited by constant variance; there is no heterogeneity at all in the first two moments; and dependence dies out. In any given situation we often need even stricter restrictions, but covariance stationarity is a good start.

▶ **Aside:: Decay of Dependence:**

- Using stuff we will see in the lecture on the math of covariance stationary processes, you can show that every covaraiance stationary process has an autocovariance function the decays exponentially.

▶ **2nd definition well-behaved: strict stationarity**

- Defn: strict stationarity.
- For any $k \geq 0$, the distribution of $y_t, \ldots, y_{t+k}$ is the same regardless of the choice of t.
- The distributions of any finite stretch of outcomes is the same, independent of where in time that stretch resides

  This formalizes 'process looks same' at any point in time

▶ **Relation weak and strict**

- Does strict imply weak?

  Strict didn't require finite variance. An iid variable with infinite variance is strictly stationary but not cov. stat.

- Does weak imply strict?

  Obviously not: weak only places restrictions on the first two moments, strict is about the entire distribution

▶ **A usage note**

- Stationary is often used without a modifier.
- A hardcore theorist means strictly stationary
- An applied person probably means weakly stat.
- In this class, I'll generally mean weakly stat.

▶ **Note on good behavior for strict stat.**

- Strict stationarity obvsiously puts very strong restrictions on heterogeneity of the distributions observations:

  any two spans of k observations must have *identical* distributions

- Strict stationarity does not, however, imply anything about tail thickness of individual distributions

▶ **Two additional concepts**

- Mixing
- Ergodic

▶ **Mixing**

- Strict stationarity requires all properties remain the same through time
- Covariance stationarity requires that the second moments remain the same through time
- Mixing conditions give us a fairly flexible way to formalize the the rate at which dependence decays.
- We will not emphasize mixing much, but it is a good concept to understand
- I sketch here for completeness

▶ **Mixing**

- Suppose we pick a measure, m, for 'how far from indepenedent' two random variables are.

  thus, for any t and k, $M(x_t, x_{t-k})$ is a measure of the dependence between the two.

- Just assume we pick a meaningful measure here.
- For any heterogenous time series, we could then compute for a given lag k, the supremum of this measure across all t:

$$M(k) = \sup_t m(x_t, x_{t-k})$$

- That is, two observations that come k periods apart apart are never more dependent than $M(k)$.
- Mixing conditions require that $M(k)$ fall at a certain rate. E.g., we could assume,

$$M(k) < 1/k^2$$

or

$$\sum M(k) < \infty$$

or,. . . , well you get the picture.

- For actual mixing conditions and details see McFadden's notes
- The faster dependence dies off, the few restrictions of other sorts we'll need for a CLT to apply.

▶ **Aside:: Mixing, the metric**

- What is this metric on 'how far two random variables are from independent'?

  well, here's a sketch.

- We have rv's x_1 and x_2 and want a measure of how dependent they are
- Suppose V_1 is an event for x_1 and V_2 for x_2

  e.g., V_1 is $x_1 > 4$ and V_2 is $2 < x_2 < 3$

- If the two rv's are independent, then by definition

$$p(V_1 \& V_2) = p(V_1)p(V_2)$$

  And this is true for all pairs of indep. events

▶ **Aside::**

- One might take $|p(V_1 \& V_2) - p(V_1)p(V_2)|$ as a measure of how far the two are from independence.

  but this will depend on the choice of events

- Thus, let's take the sup. of this expression across all pairs of events V_1 and V_2

▶ **Aside::**

- For a time series, w_t, we can define

$$M(k) = \sup |p(V_t \& V_{t-k}) - p(V_t)p(V_{t-k})|$$

where V_t is an event for w_t, V_{t-k} is an event for w_{t-k} and the sup is over all t and all such events.

- $M(s)$ is the biggest mistake you could make in terms of probabilities of events in treating events k periods apart as if they were independent.
- With a metric like this, we can then require this dependence to fall fall at a certain rate,

▶ **Note:**

- This was just a sketch of some key ideas on mixing
- Mainly intended to highlight the fact that in time series to get our WLLNs and CLTs to go, it will always be crucial that we assume that events sufficiently far apart are nearly independent

▶ **Ergodic**

- With iid data, we know that if we get one sample that is long enough, we can learn most everything there is to know about the distribution the observations are drawn from
- This needn't be the case in time series.

  If an abnormally high value for x_t tends to be associated with abnormally high values for x_{t+k} for all $k > 0$, then our 'bad draw' for x_t will cause the whole sample to be unrepresentative.

▶ **Nonergodic: simple example**

- Think of a Markov Chain where the process jumps between among values $A_1, \ldots, A_N$.
- Suppose A_3 is an absorbing state so that if you land there, you never leave.
- So long as we visit this state with positive probability, every sample ultimately becomes a constant series of A_3 observations.

  Information ceases to build up as more observations come in.

▶ **Ergodic**

- I will not formally define ergodic (which requires more technical stuff than I want to get into in this class).
- The idea, however, is pretty straightforward:

  A process is ergodic if every sufficiently long sample fully reveals the properties of the process.

▶ **Ergodic for the mean**

- You will sometimes see the expression 'ergodic for the mean' or higher moments
- This means that sample moments converge to the stated population moments for this process

▶ **Some intuition**

- Suppose we have M independent samples of length T from some time series process

- We can put these in an $T \times M$ rectangular matrix.

  The columns of this matrix are mutually independent; the rows in any column are not independent.

- You should be comfortable with the idea that the sample mean of elements on, say, the m^{th} row will be consistent for the mean as M gets large.

  If the process has a mean, then this will hold regardless of the dependence in the process.

- We say that the process is erogodic (for the mean) if the sample mean of any column is also consistent for the mean as T gets large.

▶ **Aside:: Numerical work**

- Suppose we want to learn features of a process such as unconditional moments
- In time series, this can be a pain
- If
  - The process is ergodic
  - And if we can generate data from the process

    Then numerical work is easy

▶ **Aside::**

- We generate one very long sample and treat sample moments as good estimates of population moments

  (Or sometimes more effic. to generate several long-ish samples and average all of them)

- end aside

▶ **A couple more terms**

- More terms for structuring dependence
- Martingale, martingale difference sequence
- Markov process, markov chain

▶ **Martingale**

- x_t follows a Martingale process iff

$$E[x_t|x_{t-1},\ldots] = E[x_t|x_{t-1}] = x_{t-1}$$

Your best expectation of x tomorrow is x today. This is true whether your information includes only x today or the entire prior history of x.

▶ **Martingale difference sequence**

- A Martingale difference sequnce (MDS) is the first differnce of a Martingale
- If x_t follows a Martingale, then $z_t \equiv x_t - x_{t-1}$ is an MDS
- If z_t is an MDS, what is $E[z_t|z_{t-1}, z_{t-2}, \ldots]$?

  Zero. In words, Ez_t is the expected change in a Martingale. This is zero.

▶ **MDS**

- If z_t is an MDS, what is the covariance of z_t and z_{t-k}?

  Zero.

- Thus, an MDS is a very general notion of a 'not serially correlated' process

▶ **Markov process**

- x_t follows a Markov process if,

$$F(x_t|x_{t-1}, \ldots) = F(x_t|x_{t-1})$$

  where $F(|)$ is the conditional distribution of.

- With a Martingale, everything the history of x through time $t-1$ reveals about the mean of x_t is captured by x_{t-1}
- With a Markov, everything the history of x through time $t-1$ reveals about any aspect of the distribution of x_t is captured by x_t

▶ **Markov chain**

- If x can only take on a discrete set of values, then rather than saying Markov process we sometimes say Markov chain
- Markov chains can be analyzed using some particularly simple tools and are often useful for working examples

▶ **Higher order Markov**

- If

$$F(x_t|x_{t-1}, \ldots) = F(x_t|x_{t-1}, \ldots, x_{t-p})$$

  then we say x follows a p^{th}-order Markov

▶ **That's it for senses of well-behaved**

- **Parting shot**
  - McFadden's notes have nice tables using these ideas in WLLNs and CLTs

    McFadden's ch. 4 go

    http://elsa.berkeley.edu/users/mcfadden/e240a%5Fsp01/ch4.pdf

- **Note:**
  - In the tables we are taking the mean of y_k, $k = 1, \ldots, n$, so the sample is of size n, and
  - X_n is the mean of the ys—the thing a WLLN would apply to.
  - Z_n is $\sqrt{n}X_n$—the thing a CLT would apply to.
  - Finally, McFadden uses what for some purposes is a better notation for convergence in distribution than I am using
  - I am writing,

    $$Z_n \to_d N(0, \sigma^2)$$

    in words, Z_n converges in distribution to $N(0, \sigma^2)$
  - McFadden writes,

    $$Z_n \to_d Z_0 \sim N(0, \sigma^2)$$

    In words, Z_n converges to some random variable Z_0 that is distributed $N(0, \sigma^2)$.
  - If you are interested in theory, you might ponder why my notation might be viewed as a bit unseemly and McFadden's might be preferred.

- **WLLNs**

FIGURE 4.4. LAWS OF LARGE NUMBERS FOR $X_n = n^{-1}\sum_{k=1}^{n} Y_k$

WEAK LAWS (WLLN)

1 (Khinchine) If the Y_k are i.i.d., and $\mathbf{E}\,Y_k = \mu$, then $X_n \to_p \mu$

2 (Chebyshev) If the Y_k are uncorrelated with $\mathbf{E}\,Y_k = \mu$ and $\mathbf{E}(Y_k - \mu)^2 = \sigma_k^2$ satisfying

$$\sum_{k=1}^{\infty} \sigma_k^2/k^2 < +\infty, \text{ then } X_n \to_p \mu$$

3 If the Y_k have $\mathbf{E}\,Y_k = \mu$, $\mathbf{E}(Y_k-\mu)^2 \equiv \sigma_k^2$, and $|\mathbf{E}(Y_k-\mu)(Y_m-\mu)| \le \rho_{km}\sigma_k\sigma_m$ with

$$\sum_{k=1}^{\infty} \sigma_k^2/k^{3/2} < +\infty \text{ and } \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}\sum_{m=1}^{n} \rho_{km} < +\infty, \text{ then } X_n \to_p \mu$$

STRONG LAWS (SLLN)

1 (Kolmogorov I) If the Y_k are i.i.d., and $\mathbf{E}\,Y_k = \mu$, then $X_n \to_{as} \mu$

2 (Kolmogorov II) If the Y_k are independent, with $\mathbf{E}\,Y_k = \mu$, and $\mathbf{E}(Y_k-\mu)^2 = \sigma_k^2$ satisfying

$$\sum_{k=1}^{\infty} \sigma_k^2/k^2 < +\infty, \text{ then } X_n \to_{as} \mu$$

3 (Martingale) Y_k adapted to $\sigma(\ldots,Y_{k-1},Y_k)$ is a martingale difference sequence, $\mathbf{E}Y_t^2 = \sigma_t^2$, and $\sum_{k=1}^{\infty} \sigma_k^2/k^2 < +\infty$, then $X_n \to_{as} 0$

4 (Serfling) If the Y_k have $\mathbf{E}\,Y_k = \mu$, $\mathbf{E}(Y_k-\mu)^2 = \sigma_k^2$, and $|\mathbf{E}(Y_k-\mu)(Y_m-\mu)| \le \rho_{|k-m|}\sigma_k\sigma_m$,

with $\sum_{k=1}^{\infty} (\log k)^2\sigma_k^2/k^2 < +\infty$ and $\sum_{k=1}^{\infty} \rho_{|k-m|} < +\infty$, then $X_n \to_{as} \mu$

- **CLTs**

FIGURE 4.5. CENTRAL LIMIT THEOREMS FOR $Z_n = n^{-1/2}\sum_{i=1}^{n} Y_i$

1 (Lindeberg-Levy) Y_k i.i.d., $\mathbf{E}Y_k = 0$, $\mathbf{E}Y_k^2 = \sigma^2$ positive and finite $\Longrightarrow Z_n \to_d Z_o \sim N(0,\sigma^2)$

2 (Lindeberg-Feller) If Y_k independent, $\mathbf{E}Y_k = 0$, $\mathbf{E}Y_k^2 = \sigma_k^2 \in (0,+\infty)$, $c_n^2 = \sum_{k=1}^{n} \sigma_k^2$, then $c_n^2 \to +\infty$, $\lim_{n\to\infty} \max_{1\le k\le n} \sigma_k/c_n = 0$, and $U_n = \sum_{k=1}^{n} Y_k/c_n \to_d U_o \sim N(0,1)$ if and only if the *Lindeberg condition* holds: for each $\varepsilon > 0$, $\sum_{k=1}^{n} \mathbf{E}\,Y_k^2\cdot\mathbf{1}(|Y_k| > \varepsilon c_n)/c_n^2 \to 0$

3 If Y_k independent, $\mathbf{E}Y_k = 0$, $\mathbf{E}Y_k^2 = \sigma_k^2 \in (0,+\infty)$, $c_n^2 = \sum_{k=1}^{n} \sigma_k^2$ have $c_n^2 \to +\infty$ and $\lim_{n\to\infty} \max_{1\le k\le n} \sigma_k/c_n = 0$, then each of the following conditions is sufficient for the Lindeberg condition:

(i) For some $r > 2$, $\sum_{k=1}^{n} \mathbf{E}\,|Y_k|^r/c_n^r \to 0$.

(ii) (Liapunov) For some $r > 2$, $\mathbf{E}\,|Y_k/\sigma_k|^r$ is bounded uniformly for all n.

(iii) For some $r > 2$, $\mathbf{E}\,|Y_k|^r$ is bounded, and c_k^2/k is bounded positive, uniformly for all k.

4 Y_k a martingale difference sequence adapted to $\sigma(\ldots,Y_{k-1},Y_k)$ with $|Y_k| < M$ for all t and $\mathbf{E}Y_k^2 = \sigma_k^2$ satisfying $n^{-1}\sum_{k=1}^{n} \sigma_k^2 \to \sigma_o^2 > 0 \Longrightarrow Z_n \to_d Z_o \sim N(0,\sigma_o^2)$

5 (Ibragimov-Linnik) Y_k stationary and strong mixing with $\mathbf{E}\,Y_k = 0$, $\mathbf{E}\,Y_k^2 = \sigma^2 \in (0,+\infty)$, $\mathbf{E}Y_{k+s}Y_k = \sigma^2\rho_s$, and for some $r > 2$, $\mathbf{E}|Y_n|^r < +\infty$ and $\sum_{k=1}^{\infty} \alpha(k)^{1-2/r} < +\infty \Longrightarrow \sum_{s=1}^{\infty} |\rho_s| < +\infty$ and $Z_n \to_d Z_o \sim N(0,\sigma^2(1+2\sum_{s=1}^{\infty} \rho_s))$

- **Consider McFadden's CLTs 4 and 5**

- Let's see how they bound heterogeneity above and below and limit dependence
- The MDS (martingale difference sequence) CLT bounds the variance of the r.v. from below by assuming

$$n^{-1} \sum \sigma_n^2 \to \sigma_0^2$$

  If the variances were converging to zero, for example, this won't hold.

- We have a very strong bound on heterogeneity from above by assuming that the r.v. has bounded support of $+/-M$.
- Under this very strong upper bound on the heterogeneity, the bounds that the MDS assumption place on dependence are sufficient for the CLT.

  Remember the MDS assumptions implies $Ey_n|y_{n-} = 0$.

- Thus, with the finite support assumption, we need few further limits on differences in the distributions of the ys.
- We are seldom willing to assume bounded support, so this CLT is not of great practical use in macro

► **Ibragimov-Linnick (CLT 5)**

- In Ibragimov-Linnick we have a nice example of the tradeoff between limits on the thickness of the tails of the distributions of the ys and the rate at which dependence vanishes.
- First note that y_n is strictly stationary, so there is no heterogeneity in the distribution of y_t and y_s.
- We assume $E|y_n|^r$ is finite for some $r > 2$.

  That is, we constrain the tail thickness by assuming that we have at least slightly more than 2 finite moments.

- Without getting into the details of how we are measuring dependence, we are requiring that in the relevant notion, dependence at lag k is bounded above by a sequence, $\alpha(k)$, and those αs decline to zero with k fast enough that,

$$\sum \alpha(k)^{1-2/r} < \infty$$

  where the r is the same r as in the moment restriction. Ponder for a moment: bigger r means that the αs can decay more slowly.

- As r rises from 2 in the moment condition, we allow the dependence to decay more slowly.

  trade off between tail thickness and decay of dependence.

- This should make sense, but you might want to ponder it a bit.

- **Practical relevance**
  - Ultimately, we are not mainly concerned with what is the minimum required assumptions for a valid asymptotic approximation to exist
  - We care about what conditions are needed for the asymptotic approximation to provide a good approximation to the exact distribution in relevant cases and sample sizes.
  - These mixing conditions don't have any formal implication in any finite sample size, so they are not directly of practical relevance.
  - But the theory results have (in practice) provided many important insights allowing us to structure our understanding of things that have practical relevance.