Lecture 4

Central Limit Theorems

When the number of observations goes to infinity, many estimators practically used in time series analysis turn out to be asymptotically normally distributed. For instance, in Lecture 6 we will demonstrate the asymptotic normality of the Yule-Walker estimator, that is an estimator of the coefficients of an $AR(p)$. The tool that will be used to prove such an asymptotic normality is the central limit theorem for martingale difference series that we will develop in this lecture.

Classically, the central limit theorem states that the mean of independent, identically distributed random variables with finite variance has a Gaussian asymptotic distribution. More precisely, the statement is the following:

Theorem 1. *Let $(X_i)_{i\geq 1}$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = \mu < \infty$ and $Var[X_i] = \sigma^2 < \infty$, for all i. Then the sequence $\left(\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)\right)_n$ weakly converges to a Gaussian distribution $\mathcal{N}(0, \sigma^2)$.*

The assumption of identical distributions in Theorem 1 can be relaxed by including different requirements such as the following Lindeberg condition.

Theorem 2 (Lindeberg Central Limit Theorem)**.** *Let $(X_i)_{i\geq 1}$ be a sequence of independent random variables with $\mathbb{E}[X_i] = \mu_i < \infty$ and $Var[X_i] = \sigma_i^2 < \infty$, for all i. Define*

$$\begin{aligned} Y_i &= X_i - \mu_i, \quad \forall i; \\ S_n &= \sum_{i=1}^n (X_i - \mu_i); \\ v_n^2 &= Var(S_n) = \sum_{i=1}^n \sigma_i^2. \end{aligned}$$

If the Lindeberg condition:

$$\frac{1}{v_n^2}\sum_{i=1}^n \mathbb{E}\left[Y_i^2 \mathbf{1}_{|Y_i|>\varepsilon v_n}\right] \to 0, \quad \forall \varepsilon > 0$$

is satisfied, then the sequence $\left(\frac{S_n}{v_n}\right)_n$ weakly converges to a Gaussian distribution $\mathcal{N}(0, 1)$.

A natural question is then how to extend the central limit theorem to dependent variables. Observe that if the only assumption that we have is that the X_i's are uncorrelated, then generally the central limit theorem does not hold, that is $\left(\frac{S_n}{v_n}\right)_n$ does not weakly converge to a Gaussian distribution. To see that, take e.g. a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables $(\varepsilon_i)_{i\geq 1}$ and a centered random variable Y independent of $(\varepsilon_i)_{i\geq 1}$ with $E[Y^2]=1$. Set $X_i = Y\varepsilon_i$; then,

$$\frac{S_n}{v_n} = \frac{1}{\sqrt{n}} Y(\varepsilon_1 + \cdots + \varepsilon_n) \stackrel{d}{=} YN, \quad N \sim \mathcal{N}(0,1)$$

that is not Gaussian for an arbitrary Y.

However, for time series analysis, the need for a central limit theorem will often arise in a situation where we are summing dependent random variables. This is the case, for instance, if we want to establish the asymptotic normality of the empirical mean of a stationary process. Given a weakly stationary time series $(X_t)_{t\in\mathbb{Z}}$ with mean μ and autocovariance function c we have already seen that if $\sum_{h\in\mathbb{Z}} |c(h)| < \infty$ then $n\mathrm{Var}(\widehat{\mu}_n) \to \sum_{h\in\mathbb{Z}} c(h)$. In any case,

$$n\mathrm{Var}(\widehat{\mu}_n) \leq \sum_{h\in\mathbb{Z}} |c(h)|.$$

Below, we will give sufficient conditions for the sequence $(\sqrt{n}(\widehat{\mu}_n - \mu))_n$ to be asymptotically normally distributed with mean zero and variance $\sum_{h\in\mathbb{Z}} c(h)$.

Such conditions are of two types: martingale and mixing. We will prove a central limit theorem for martingale difference series but before that let us see some examples of results that one can obtain under mixing conditions.

Examples of CLT under mixing conditions

Definition 1. *A time series $X = (X_t)_{t\in\mathbb{Z}}$ is called* m-dependent *if the random vectors $(\ldots, X_{t-1}, X_t)$ and $(X_{t+m+1}, X_{t+m+2}, \ldots)$ are independent for every $t \in \mathbb{Z}$.*

Theorem 3. *Let $X = (X_t)_{t\in\mathbb{Z}}$ be a strictly stationary, m-dependent time series with mean zero and finite variance. Then the sequence $\sqrt{n}\widehat{\mu}_n$ converges in distribution to a normal distribution with mean 0 and variance $\sum_{h\in\mathbb{Z}} c(h)$.*

Proof. See e.g. [1]. □

Remark 1. *The property of m-dependence is a natural generalization of independence. Observations of an m-dependent process are independent provided they are separated in time by more than m time units. When $m = 0$, m-dependence coincides with independence. For instance, as we already pointed out in Lecture 2, $MA(q)$ processes are m-dependent with $m = q$.*

Theorem 4. *Let X be a linear process of the form*

$$X_t = \mu + \sum_{j\in\mathbb{Z}} \psi_j Z_{t-j},$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random variables with $\mathbb{E}[Z_0] = 0$, $\mathbb{E}[Z_0^2] = \sigma^2 < \infty$ and $\sum_{j\in\mathbb{Z}} |\psi_j| < \infty$. Then the sequence $\sqrt{n}(\widehat{\mu}_n - \mu)$ converges in distribution to a Gaussian distribution with mean zero and variance $v = \sigma^2\left(\sum_{j\in\mathbb{Z}} \psi_j\right)^2$.

Proof. See e.g. [1]. □

For central limit theorems under α-mixing conditions see, e.g. [1], Theorem 4.7.

Central Limit Theorem for martingale difference series

Reminder

Definition 2. *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space,* $\mathcal{H} \subset \mathscr{F}$ *be a sub-*σ*-algebra of* $\mathscr{F}$ *and* $X : \Omega \to \mathbb{R}$ *be a real random variable with* $\mathbb{E}[|X|] < \infty$*. The* conditional expectation *of* X *given* $\mathcal{H}$*, denoted as* $\mathbb{E}[X|\mathcal{H}]$*, is the a.s. unique random variable which satisfies the following two conditions:*

1. $\mathbb{E}[X|\mathcal{H}]$ *is* $\mathcal{H}$*-measurable.*

2. *For any* $H \in \mathcal{H}$*,*
$$\int_H \mathbb{E}[X|\mathcal{H}] d\mathbb{P} = \int_H X d\mathbb{P}.$$

If $\mathcal{H} = \sigma(Z)$ for a certain random variable Z, we will write $\mathbb{E}[X|Z]$ instead of $\mathbb{E}[X|\sigma(Z)]$. Recall that the conditional expectation enjoys the following properties:

Proposition 1. *Let* X, X_1, X_2 *be integrable random variables and* $\alpha, \beta \in \mathbb{R}$*. Then,*

- $\mathbb{E}[\alpha X_1 + \beta X_2|\mathcal{H}] = \alpha\mathbb{E}[X_1|\mathcal{H}] + \beta\mathbb{E}[X_2|\mathcal{H}]$ *a.s.*
- *If* $X \geq 0$ *a.s., then* $\mathbb{E}[X|\mathcal{H}] \geq 0$ *a.s.*
- $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]] = \mathbb{E}[X]$*.*
- *If* $\mathcal{H}' \subset \mathcal{H}$*,* $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{H}'] = \mathbb{E}[X|\mathcal{H}']$ *a.s.*
- *If* Z *is* $\mathcal{H}$*-measurable,* $\mathbb{E}[ZX|\mathcal{H}] = Z\mathbb{E}[X|\mathcal{H}]$ *a.s. (Here require that* $X \in L_p$ *and* $Z \in L_q$ *for* $1 \leq p \leq \infty$ *and* $p^{-1} + q^{-1} = 1$*.)*
- *If* X *is independent of* $\mathcal{H}$*,* $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$ *a.s.*

Definition 3. *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space. A* filtration $(\mathscr{F}_t)_{t\in T}$ *is a nondecreasing collection of sub-*σ*-algebra of* $\mathscr{F}$*:* $\mathscr{F}_s \subset \mathscr{F}_t \subset \mathscr{F}$ *for any* $s \leq t$*,* $s, t \in T$*.*

Definition 4. *Let* $(\mathscr{F}_t)_{t\in T}$ *be a filtration. A process* $(X_t)_{t\in T}$ *is a* $(\mathscr{F}_t)$*-martingale if:*

- X_t *is* $\mathscr{F}_t$*-measurable, for every* t*.*
- $\mathbb{E}[|X_t|] < \infty$*, for every* t*.*
- $\mathbb{E}[X_t|\mathscr{F}_s] = X_s$ *a.s. for any* $s < t$*.*

The martingale central limit theorem applies to the special time series X for which the partial sums $(\sum_{t=1}^n X_t)_n$ are a martingale. This can be rephrased by saying that X satisfies the following definition.

Definition 5. *A* martingale difference series *relative to a given filtration* $(\mathscr{F}_t)_{t\in T}$ *is a time series* $X = (X_t)_{t\in T}$ *such that, for every* t *:*

1. X_t *is* $\mathscr{F}_t$*-measurable.*

2. $\mathbb{E}[|X_t|] < \infty$, *for every* t.

3. $\mathbb{E}[X_t|\mathscr{F}_{t-1}] = 0$, *a.s.*

We recall also this classical proposition that will be useful in the following.

Proposition 2. *If* $(X_n)_n$ *is a sequence of random variables bounded in* L_1 *and converging to zero in probability, then* $\mathbb{E}[X_n] \to 0$ *as* $n \to \infty$.

Statement and proof of the martingale CLT

Theorem 5. *Let* $(X_t)_{t\in\mathbb{N}}$ *be a martingale difference series relative to the filtration* $(\mathscr{F}_t)_{t\in\mathbb{N}}$ *that satisfies the following conditions:*

(H1) There exists a positive constant v *such that*

$$n^{-1}\sum_{t=1}^{n}\mathbb{E}[X_t^2|\mathscr{F}_{t-1}] \xrightarrow[n\to\infty]{\mathbb{P}} v.$$

(H2) For all $\varepsilon > 0$

$$n^{-1}\sum_{t=1}^{n}\mathbb{E}[X_t^2\mathbf{1}_{|X_t|>\varepsilon\sqrt{n}}|\mathscr{F}_{t-1}] \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Then, $\sqrt{n}\widehat{\mu}_n$ *converges in distribution to a normal distribution with mean* 0 *and variance* v.

Remark 2. *The martingale central limit theorem can be seen as another type of generalization of the ordinary central limit theorem. The classical central limit theorem shows the asymptotic normality of the partial sums of i.i.d. random variables* X_i*, provided that the* X_i *are centered and not too big (finite variance suffices). The martingale central limit theorem states that there is asymptotic normality for the partial sums of random variables* X_i *provided that the first moment given the past is zero and the second moment given the past is not too big.*

Proof. Consider the events $A_t = \left\{\frac{1}{n}\sum_{j=1}^{t}\mathbb{E}[X_j^2|\mathscr{F}_{j-1}] \le 2v\right\}$. In particular, $A_n \subset \cdots \subset A_2 \subset A_1$ for any n and $\cap_{t=1}^{n} A_t = A_n$. Moreover, by Assumption (H1), $\mathbb{P}(A_n) \to 1$ as $n \to \infty$. Set

$$X_{n,t} = \frac{1}{\sqrt{n}}X_t\mathbf{1}_{A_t}$$

and observe that $X_{n,t}(\omega) = \frac{1}{\sqrt{n}}X_t(\omega)$ for any $\omega \in A_n$. Because $\mathbb{P}(A_n) \to 1$, in order to prove that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t \xrightarrow[n\to\infty]{d} \mathcal{N}(0,v)$ it is enough to prove that $\sum_{t=1}^{n} X_{n,t} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,v)$ which is equivalent to prove that

$$\mathbb{E}\left[e^{iu\sum_{t=1}^{n} X_{n,t}}\right] \xrightarrow[n\to\infty]{} e^{-\frac{u^2 v}{2}}. \tag{1}$$

Therefore, from now on, we will focus on demonstrating (1).

We begin by observing that the random variables $X_{n,t}$ are martingale difference series relative to the filtration $(\mathscr{F}_t)$ (because $A_t \in \mathscr{F}_{t-1}$) that satisfy the following properties:

$$\sum_{t=1}^{n} \mathbb{E}[X_{n,t}^2 | \mathscr{F}_{t-1}] \leq 2v; \tag{2}$$

$$\sum_{t=1}^{n} \mathbb{E}[X_{n,t}^2 | \mathscr{F}_{t-1}] \xrightarrow[n\to\infty]{\mathbb{P}} v; \tag{3}$$

$$\sum_{t=1}^{n} \mathbb{E}[X_{n,t}^2 \mathbf{1}_{|X_{n,t}|>\varepsilon} | \mathscr{F}_{t-1}] \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad \forall \varepsilon > 0. \tag{4}$$

Furthermore, we shall use the following facts:

- $|e^{ix} - 1 - ix| \leq \frac{x^2}{2}$ for every $x \in \mathbb{R}$.
- Define $R : \mathbb{R} \to \mathbb{C}$ as $e^{ix} - 1 - ix + \frac{x^2}{2} = x^2 R(x)$. Then $|R(x)| \leq 1$ and $R(x) \to 0$ as $x \to 0$.
- If $\delta, \varepsilon > 0$ are chosen such that $|R(x)| < \varepsilon$ if $|x| \leq \delta$, then $|e^{ix} - 1 - ix + x^2/2| \leq x^2 \mathbf{1}_{|x|>\delta} + \varepsilon x^2$.

For fixed $u \in \mathbb{R}$ define

$$R_{n,t} = \mathbb{E}[e^{iuX_{n,t}} - 1 - iuX_{n,t} | \mathscr{F}_{t-1}].$$

It follows that $|R_{n,t}| \leq \frac{u^2}{2} \mathbb{E}[X_{n,t}^2 | \mathscr{F}_{t-1}]$ and therefore, using the definition of A_t:

$$\sum_{t=1}^{n} |R_{n,t}| \leq \frac{u^2}{2} \sum_{t=1}^{n} \mathbb{E}[X_{n,t}^2 | \mathscr{F}_{t-1}] \leq u^2 v; \tag{5}$$

$$\max_{1\leq t\leq n} |R_{n,t}| \leq \frac{u^2}{2} \left(\sum_{t=1}^{n} \mathbb{E}[X_{n,t}^2 \mathbf{1}_{|X_{n,t}|>\delta} | \mathscr{F}_{t-1}] + \delta^2 \right), \quad \forall \delta > 0; \tag{6}$$

$$\sum_{t=1}^{n} \left| R_{n,t} + \frac{u^2}{2} \mathbb{E}[X_{n,t}^2 | \mathscr{F}_{t-1}] \right| \leq u^2 \sum_{t=1}^{n} \left(\mathbb{E}[X_{n,t}^2 \mathbf{1}_{|X_{n,t}|>\delta} | \mathscr{F}_{t-1}] + \varepsilon \mathbb{E}[X_{n,t}^2 | \mathscr{F}_{t-1}] \right). \tag{7}$$

From (6) and (4) it follows that the sequence $(\max_{1\leq t\leq n} |R_{n,t}|)_n$ tends to zero in probability whereas, from (7), (2) and (4) we deduce that the sequence $(\sum_{t=1}^{n} R_{n,t})_n$ tends in probability to $-\frac{u^2}{2} v$.

Consider the function $S : \mathbb{R} \to \mathbb{R}$ defined by $\log(1-x) = -x + xS(x)$ and observe that it satisfies $S(x) \to 0$ as $x \to 0$. It follows that $\max_{1\leq t\leq n} |S(R_{n,t})| \xrightarrow[n\to\infty]{\mathbb{P}} 0$ and that (using (5)) $(\sum_{t=1}^{n} R_{n,t} S(R_{n,t}))_n$ converges to zero in probability. Thus,

$$\prod_{t=1}^{n} (1 - R_{n,t}) = e^{\sum_{t=1}^{n} \log(1-R_{n,t})} = e^{-\sum_{t=1}^{n} R_{n,t} + \sum_{t=1}^{n} R_{n,t} S(R_{n,t})} \xrightarrow[n\to\infty]{\mathbb{P}} e^{u^2 v/2}. \tag{8}$$

Furthermore, from (5) it follows

$$\prod_{t=1}^{k} |1 - R_{n,t}| \leq \exp\left(\sum_{t=1}^{n} |R_{n,t}|\right) \leq e^{u^2 v}, \quad \forall k \leq n. \tag{9}$$

As a consequence of (8) and (9) we obtain:

$$\mathbb{E}\left|\prod_{t=1}^{n}(1-R_{n,t})-e^{\frac{u^2v}{2}}\right| \xrightarrow[n\to\infty]{} 0. \tag{10}$$

Moreover, using that $\mathbb{E}[X_{n,t}|\mathscr{F}_{t-1}]=0$ we can also deduce that, for every t,

$$\begin{aligned}\mathbb{E}[e^{iuX_{n,t}}(1-R_{n,t})|\mathscr{F}_{t-1}] &= (1-R_{n,t})\mathbb{E}\big[e^{iuX_{n,t}}-1-iuX_{n,t}+1|\mathscr{F}_{t-1}\big]\\ &= (1-R_{n,t})(R_{n,t}+1)=1-R_{n,t}^2.\end{aligned} \tag{11}$$

Therefore, conditioning on $\mathscr{F}_{n-1}$ and using (11), we get

$$\begin{aligned}\mathbb{E}\left[\prod_{t=1}^{n}e^{iuX_{n,t}}(1-R_{n,t})\right] &= \mathbb{E}\left[\mathbb{E}\left[\prod_{t=1}^{n}e^{iuX_{n,t}}(1-R_{n,t})|\mathscr{F}_{n-1}\right]\right]\\ &= \mathbb{E}\left[\prod_{t=1}^{n-1}e^{iuX_{n,t}}(1-R_{n,t})\mathbb{E}[e^{iuX_{n,n}}(1-R_{n,n})|\mathscr{F}_{n-1}]\right]\\ &= \mathbb{E}\left[\prod_{t=1}^{n-1}e^{iuX_{n,t}}(1-R_{n,t})\right]-\mathbb{E}\left[\prod_{t=1}^{n-1}e^{iuX_{n,t}}(1-R_{n,t})R_{n,n}^2\right].\end{aligned}$$

By repeating this argument, we can conclude that

$$\begin{aligned}\mathbb{E}\left[\prod_{t=1}^{n}e^{iuX_{n,t}}(1-R_{n,t})\right] &= \mathbb{E}\left[\mathbb{E}\left[\prod_{t=1}^{n-1}e^{iuX_{n,t}}(1-R_{n,t})|\mathscr{F}_{n-2}\right]\right]-\mathbb{E}\left[\prod_{t=1}^{n-1}e^{iuX_{n,t}}(1-R_{n,t})R_{n,n}^2\right]\\ &= \dots\\ &= \mathbb{E}\big[e^{iuX_{n,1}}(1-R_{n,1})\big]-\mathbb{E}\left[\sum_{k=2}^{n}\prod_{t=1}^{k-1}e^{iuX_{n,t}}(1-R_{n,t})R_{n,k}^2\right].\end{aligned}$$

By means of (9) and (11), joined with $\mathbb{E}[e^{iuX_{n,1}}(1-R_{n,1})]=1-\mathbb{E}[R_{n,1}^2]$ (from (11)), it follows:

$$\left|\mathbb{E}\left[\prod_{t=1}^{n}e^{iuX_{n,t}}(1-R_{n,t})\right]-1\right| = \left|-\sum_{k=1}^{n}\mathbb{E}\left[\prod_{t=1}^{k-1}e^{iuX_{n,t}}(1-R_{n,t})R_{n,k}^2\right]\right| \le e^{u^2v}\mathbb{E}\left[\sum_{t=1}^{n}R_{n,t}^2\right].$$

We can therefore conclude (using Proposition 2)

$$\left|\mathbb{E}\left[\prod_{t=1}^{n}e^{iuX_{n,t}}(1-R_{n,t})\right]-1\right| \xrightarrow[n\to\infty]{} 0, \tag{12}$$

because $\sum_{t=1}^{n}|R_{n,t}|\le u^2v$ and $\max_{1\le t\le n}|R_{n,t}|$ tends to zero in probability.

Finally, combining (10) and (12), we get

$$\left|\mathbb{E}\left[\prod_{t=1}^{n}e^{iuX_{n,t}}e^{u^2v/2}\right]-1\right| = \left|\mathbb{E}\left[e^{iu\sum_{t=1}^{n}X_{n,t}}\left(e^{\frac{u^2v}{2}}-\prod_{t=1}^{n}(1-R_{n,t})\right)\right]\right|+o(1)\to 0,$$

i.e. (1) is established, as desired. □

References

[1] van der Vaart, A.: http://www.math.leidenuniv.nl/ avdvaart/timeseries/dictaat.pdf