

Time Series Analysis:

Introduction to time series and forecasting

Handout 0: Preliminaries

Laura Mayoral

IAE and Barcelona GSE

- This document reviews three basic points that will be used during this course.
- Some basic probability concepts needed to define random variables
- Converge of random variables
- Definition of estimators and basic properties.
- In case you are not familiar with this material, you can review it here or in any statistics book you like.

Basic probability concepts

Definition 1 *Algebra.* Let Ω be a set of points ω . A system \mathcal{A} of subsets of Ω is called an algebra if

a) $\Omega \subset \mathcal{A}$

b) If $A, B \subset \mathcal{A} \Rightarrow A \cup B \subset \mathcal{A}$ and $A \cap B \subset \mathcal{A}$

c) $A \subset \mathcal{A} \Rightarrow A^c \subset \mathcal{A}$

Definition 2 *σ -Algebra.* A system \mathcal{F} of subsets of Ω is a σ -algebra if it is an algebra and for $A_n \in \mathcal{F}$, $n=1,2,\dots$ then

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F} \text{ and } \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}.$$

Definition 3 *Probability measure.* Let \mathcal{F} be a σ -algebra of subsets of Ω . A set function $P = P(A)$, $A \in \mathcal{A}$ taking values in $[0, 1]$ is called a probability measure if

a) $P(\Omega) = 1$,

b) For all pairwise disjoint subsets, $A_1, A_2, \dots \in \mathcal{A}$, it holds that

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Probability measures have, among others, the following properties:

a) If \emptyset is the empty set, $P(\emptyset) = 0$

b) If $A, B \in \mathcal{A}$ then, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

c) If $A, B \in \mathcal{A}$ and $B \subseteq A$, then $P(B) \leq P(A)$.

Definition 4 *Measurable space.* The space Ω together with a σ -algebra \mathcal{F} of its subsets is a measurable space, denoted by (Ω, \mathcal{F}) .

Definition 5 *Probability space.* An ordered triple (Ω, \mathcal{F}, P) where Ω is a set of points ω , \mathcal{F} is a σ -algebra of subsets of Ω and P is a probability measure is called a probability space or a Probability model.

Definition 6 *Random variable.* Let (Ω, \mathcal{F}, P) be a probability space and $(\mathbb{R}, \mathcal{F}')$ a measurable space, where \mathbb{R} denotes the set of real numbers. A function $X: \Omega \rightarrow \mathbb{R}$, is a real-valued random variable if

$$\{\omega : X(\omega) \leq r\} \in \mathcal{F}, \text{ for all } r \in \mathbb{R}$$

An interpretation of this is that the pre-image of the “well-behaved” subsets of X (the elements of Ω) are events and, hence, are assigned a probability by P .

Definition 7 *The function $F_X(x) = P(\omega : X(\omega) \leq x)$, $x \in \mathbb{R}$ is called the distribution function of X .*

Review of asymptotic distribution theory

- In many occasions we will be interested in describing the distribution of statistics involving random variables.
- Unfortunately, most of the times it will be impossible to derive the exact distribution because it is too complicated.
- Sometimes we can approximate the distribution under the assumption that the sample size T is very large.
- This is called [asymptotic theory](#).
- We now present some basic asymptotic results that will be used in subsequent lectures.

Converge of random variables

Let $\{a_t\}$ be a sequence of strictly positive real numbers and let $\{X_n, n=1, 2, \dots\}$ be a sequence of random variables all defined in the same probability space.

Definition 8 (*Convergence in probability to zero*) X_n converges in probability to zero, written $X_n = o_p(1)$ or $X_n \xrightarrow{p} 0$, if for every $\varepsilon > 0$

$$P(|X_n| > \varepsilon) \rightarrow 0 \text{ as } n \rightarrow \infty$$

Definition 9 (*Boundedness in probability*) The sequence $\{X_n\}$ is bounded in probability, denoted as $X_n = O_p(1)$, if for every $\varepsilon > 0$ there exists $\delta(\varepsilon) \in (0, \infty)$ such that

$$P(|X_n| > \delta(\varepsilon)) < \varepsilon \text{ for all } n$$

Clearly if $X_t = o_p(1)$, then $X_t = O_p(1)$.

Definition 10 (Convergence in probability)

X_n converges in probability to X iff $X_n - X = o_p(1)$.

Definition 11 (Convergence in r^{th} mean, $r \geq 0$). The sequence of random variables $\{X_n\}$ converges in r^{th} mean to X , denoted by $X_n \xrightarrow{r} X$, if $E(|X_n - X|^r) \rightarrow 0$ as $n \rightarrow \infty$.

If $r = 2$, then X_n is said to converge to X in mean square.

Definition 12 (Convergence in distribution) The sequence of random variables $\{X_n\}$ converges in distribution to X , written as $X_n \xrightarrow{d} X$ or $X_n \Rightarrow X$, if $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ for all $x \in C$, where C is the set of continuity points of the distribution function $F_X(\cdot)$ of X .

Definition 13 (Almost sure convergence). The sequence of random variables $\{X_n\}$ converges almost surely or with probability 1 to X , written as $X_n \xrightarrow{a.s.} X$, if

$$P\left(\left\{\omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

Proposition 1 (Relation among convergence concepts).

i) if $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$; (the converse is not true in general)

ii) if $X_n \xrightarrow{r} X \Rightarrow X_n \xrightarrow{s} X \Rightarrow X_n \xrightarrow{p} X$, with $r > s \geq 1$.

iii) if X is a constant $X_n \xrightarrow{p} X \Leftrightarrow X_n \xrightarrow{d} X$.

We now present some useful results that will be of much help in order to compute asymptotic distributions.

Definition 14 (*Algebra of probability orders*)

- i) $X_n = o_p(a_n)$ iff $a_n^{-1}X_n = o_p(1)$,
- ii) $X_n = O_p(a_n)$ iff $a_n^{-1}X_n = O_p(1)$.

Proposition 2 *If X_n and Y_n , $n=1, 2, \dots$ are random variables defined on the same probability space and $a_n > 0$, $b_n > 0$, $n = 1, 2, \dots$, then*

- i) *if $X_n = o_p(a_n)$ and $Y_n = o_p(b_n)$, then $X_n Y_n = o_p(a_n b_n)$; $X_n + Y_n = o_p(\max(a_n, b_n))$; $|X_n|^r = o_p(a_n^r)$, for $r > 0$*
- ii) *if $X_n = o_p(a_n)$ and $Y_n = O_p(b_n)$, then $X_n Y_n = o_p(a_n b_n)$*
- iii) *the statement (i) is valid if $o_p(\cdot)$ is replaced everywhere by $O_p(\cdot)$*

Proposition 3 (*The Cramer-Wold device*) Let $\{X_n\}$ be a sequence of random k -vectors. Then $X_n \xrightarrow{d} X$ if and only if $\lambda' X_n \xrightarrow{d} \lambda' X$ for all $\lambda \in \mathbb{R}^k$.

Proposition 4 If $\{X_n\}$ and $\{Y_n\}$ are two sequences of random k -vectors such that $X_n - Y_n = o_p(1)$ and $X_n \xrightarrow{d} X$, then $Y_n \xrightarrow{d} X$.

Proposition 5 If $\{X_n\}$ is a sequence of random k -vectors such that $X_n \xrightarrow{d} X$ and if $h : \mathbb{R}^k \rightarrow \mathbb{R}$ is a continuous mapping then $h(X_n) \xrightarrow{d} h(X)$.

Proposition 6 (*Slutzky's Theorem*) Let $\{X_n\}, \{Y_n\}$ be sequences of random k -vectors such that $X_n \xrightarrow{p} c$ and $Y_n \xrightarrow{d} Y$. Then, $X_n + Y_n \Rightarrow c + Y$ and $X_n Y_n \Rightarrow cY$.

Proposition 7 (*The 'delta method'*) Let $\{X_n\}$ be a sequence of k -dimensional random vectors, such that $X_n \xrightarrow{p} c$, and $\sqrt{T} (X_n - c) \xrightarrow{d} z$. Let $a(\cdot) : \mathbb{R}^k \rightarrow \mathbb{R}^r$ be a function with continuous derivatives evaluated at c

$$A(c) = \frac{\partial a(c)}{\partial c'}$$

then

$$\sqrt{T} (a(X_n) - a(c)) \xrightarrow{d} A(c) z$$

Proposition 8 If $\{X_n\}$ is a sequence of random variables and $E(X_n) \rightarrow a$, where a is a constant and $Var(X_n) \rightarrow 0$, then X_n converges in mean square to a .

Estimators and basic properties

- An estimator is a function of the observable data that is used to estimate an unknown population parameter. In general, many different estimators are possible for any given parameter.
- An estimate is the result from the actual application of the function to a particular set of data.
- Let $\hat{\theta}_T$ be an estimator of the population parameter θ . Then, $\hat{\theta}_T$ is a function that maps each sample S to its sample estimate $\hat{\theta}_T(S)$. The sequence $\{\hat{\theta}_T\}$ is an example of a sequence of random variables, so the concepts introduced above are applicable to $\{\hat{\theta}_T\}$.

- Some desirable properties of $\hat{\theta}_T$ are the following.
- *Consistency*: $\hat{\theta}_T$ is consistent if $\hat{\theta}_T \xrightarrow{p} \theta$ as $T \rightarrow \infty$.
- *Unbiasedness*: $\hat{\theta}_T$ is unbiased if $E(\hat{\theta}_T) = \theta$ and is asymptotically unbiased if $\lim_{T \rightarrow \infty} E(\hat{\theta}_T) = \theta$.
- *Efficiency*: The unbiased estimator $\hat{\theta}_T$ is unbiased if it has the lowest possible variance among all unbiased estimators.
- *Asymptotic Normality*. The consistent estimator $\hat{\theta}_T$ is asymptotically normal around the true parameter θ if $\sqrt{T}(\hat{\theta}_T - \theta) \xrightarrow{d} N(0, V)$, where V is called the asymptotic variance of $\hat{\theta}_T$.

Example 1 Consider the OLS estimator of β in $y_t = x_t\beta + \varepsilon_t$. We will show below that under appropriate assumptions, $\hat{\beta}$ is consistent and asymptotically normal (as in cross-sectional data). To do so, we need to refresh first: the law of large numbers and the central limit theorem.