1. Introduction

March 31, 2012

Junhui Qian

1 Introduction

A time series is a time-indexed sequence of random variables. We may also call (X_t) a stochastic process or simply "process". [1]

We write a generic time series as $X = (X_t, t \in \mathcal{T})$, where $\mathcal{T}$ is an index set. The set $\mathcal{T}$ may be discrete, in which case X is called a discrete-time process. And $\mathcal{T}$ may be a continuous time interval, in which case X is called a continuous-time process. In this course we will focus on the discrete-time processes.

To fix ideas, we denote the probability triple as $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the sample space, $\mathcal{F}$ is the space of events (σ-algebra of Ω), and $\mathbb{P}$ is the probability measure.

A time series X is a mapping from the product space of $\Omega \times \mathcal{T}$ to $\mathbb{R}$. We may write this mapping as $X_t = X_t(\omega) = X(\omega, t)$. It is obvious that $X(t, \cdot)$ is a r.v. and $X(\cdot, \omega)$ is called a sample path.

A filtration is a non-decreasing indexed sequence of σ-fields $(\mathcal{F}_t)$. $(\mathcal{F}_s) \subset (\mathcal{F}_t)$ if $s < t$. A filtration is an ever-finer sequence of information sets.

σ-field generated by a r.v. $X - \sigma(X)$ is defined as

$$\sigma(X) = \{X^{-1}(B) | B \in \mathbb{B}(\mathbb{R})\},$$

where $\mathbb{B}(\mathbb{R})$ denotes the Borel σ-field of $\mathbb{R}$. Roughly speaking, $\sigma(X)$ is the set of all information we may know through observation of X.

The natural (or standard) filtration of a time series $X = (X_t)$ is given by $\mathcal{F}_t = \sigma((X_s)_{s \leq t})$, which is roughly information contained in X up to time t.

[1] In some textbooks, a time series is considered a realization of the underlying stochastic process. We do not differentiate these two terms.

2 Ergodicity, Stationarity, and Mixing Properties

2.1 Ergodicity

As in other areas of econometrics, the objectives of time series analysis consist in inference and forecast. To have some chance of success, we need to focus on series with some nice properties. One important property of time series is ergodicity, which dictates that the sample moments of a series converges to the corresponding population moments, if the latter exist.

Definition $X = (X_t)$ is "mean ergodic" if

$$\lim_{T\to\infty} \mathbb{E}\left(\frac{1}{T}\sum_{t=1}^{T} X_t - \mu\right)^2 = 0,$$

and "variance ergodic" if

$$\lim_{T\to\infty} \mathbb{E}\left(\frac{1}{T}\sum_{t=1}^{T} (X_t - \mu)^2 - \sigma^2\right)^2 = 0.$$

Unfortunately, the ergodicity properties cannot be verified and hence have to be assumed.

In order to be ergodic, a stochastic process must be in some kind of equilibrium. Stationarity characterizes one such kind of equilibrium.

2.2 Stationarity

Definition. A time series $X = (X_t)$ is (strictly) stationary if the joint distribution of $(X_{t_1+h}, ..., X_{t_k+h})$ is independent of h for every $t_1, ..., t_k \in \mathcal{T}$.

Definition. A time series $X = (X_t)$ is weakly stationary if the first two moments of $(X_{t_1+h}, ..., X_{t_k+h})$ exist and are independent of h for every $t_1, ..., t_k \in \mathcal{T}$.

Weak stationarity can be checked. We define autocovariance function and auto-correlation function (ACF) at lag h as

$$\gamma_X(h) = \mathrm{cov}(X_t, X_{t+h}), \quad \rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}.$$

It is obvious that for both stationary and weakly stationary processes, γ_X and ρ_X are well defined.

An i.i.d. time series is a special case of strictly stationary process.

White noise is a special case of weakly stationary process. X_t is called a white noise if X_t is zero-mean and

$$\gamma_X(h) = \begin{cases} \sigma^2, \text{ if } h = 0, \\ 0, \text{ if } h \neq 0. \end{cases}$$

If, in addition, X_t is i.i.d., it is called a strong white noise.

White noise is an important building block for constructing other series. Often we may think of white noise as innovations or shocks to the economic system. These shocks may be nonindependent, but must be uncorrelated. Indeed, an important class of white noises are ARCH/GARCH processes, which are serially uncorrelated but nonindependent.

3 Data Preprocessing

Most analytical and technical machineries in time series are developed for stationary processes. However, stationarity is a restrictive assumption for most time series we see in practice. Two most common violations of stationarity are trend/cycle and seasonality.

Trend/cycle is long term (low frequency) variation over time, and seasonality is regular cyclic movement typically seen in monthly and quarterly data. Many macroeconomic time series such as GDP and CPI, exhibit both trend/cycle and seasonal movements.

To apply machineries that are developed for stationary time series, we must preprocess data so that the stationarity assumption is more plausible for the transformed data.

3.1 Least Square Detrending

We may assume that a time series X_t can be decomposed into the sum of a deterministic time trend and a stationary process,

$$X_t = d_t + w_t,$$

where $d_t = \sum_{i=0}^{m} \beta_i t^i$ is deterministic and w_t is a zero-mean stationary process. If this assumption is plausible, we may subtract the time trend from X_t. This practice is called detrending.

Of course d_t is not observable and hence has to be estimated. We may fix an integer for m, the degree of time polynomial, and estimate $\beta_0, \beta_1, ..., \beta_m$ by minimizing

$$\sum_{i=0}^{m} \left(X_t - \beta_i t^i\right)^2.$$

This is an ordinary least square problem.

3.2 Log Difference

Many macroeconomic and financial time series exhibit exponential trend. For this type of series, taking log difference may yield plausibly stationary processes. The log difference is defined by

$$r_t = \log(X_t/X_{t-1}).$$

If X_t/X_{t-1} is close to one, then the above expression may be approximated by $X_t/X_{t-1} - 1$.

3.3 Seasonal Adjustment

Seasonal adjustment may be done by compare year-on-year changes. For example, it is common practice to look at the GDP growth rates of the same quarter, $\log(\mathrm{GDP}_t/\mathrm{GDP}_{t-4})$.

Another simple way to adjust seasonality is to take averages. For example, we may calculate the seasonal adjusted GDP as,

$$\mathrm{GDP}_{sa,t} = (\mathrm{GDP}_t + \mathrm{GDP}_{t-1} + \mathrm{GDP}_{t-2} + \mathrm{GDP}_{t-3})/4.$$

There are more advanced ways to eliminate seasonal movement. See X-12-ARIMA and Tramo/Seats.

3.4 Filtering

A general way to transform a nonstationary series into a stationary one is by linear filtering. Let L denote lag operator. A linear filter Φ is defined by $Y_t = \Phi(L)X_t \equiv \left(\sum_{i\in\mathbb{Z}} \phi_i L^i\right) X_t$. If $\phi_i = 0$ for all $i < 0$, then the filter is called "causal".

Special cases of linear causal filters are difference operator $\nabla X_t = (1 - L)X_t$, seasonal difference $\nabla^k X_t = (1 - L^k)X_t$, etc.

4 Pre-Analysis

4.1 Descriptive Statistics

To summarize distributional information in a time series data, we calculate statistics such as sample mean, variance, skewness, kurtosis, median (50% quantile) and other quantiles.

4.2 Sample Mean

We usually use sample mean, $\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$, to estimate the corresponding population mean μ. Under general conditions, $\bar{X}$ is a consistent estimator of μ, and $\sqrt{T}(\bar{X} - \mu) \to N(0, V)$, where $V = \sum_k \gamma(k)$.

To test hypotheses on the mean, however, we need reliable estimator for V. The obvious choice $\hat{V} = \sum_k \hat{\gamma}(k)$ is inaccurate since the quality of $\hat{\gamma}(k)$ deteorietes as $k \to T$, since

$$\hat{\gamma}(k) = \frac{1}{T-k} \sum_{t=k+1}^{T} (X_t - \bar{X})(X_{t-k} - \bar{X}).$$

Instead, we may use the method of "batched means" or "blockwise bootstrap".

In the following, since the population mean can be reliably estimated, we only consider zero-mean stationary time series.

4.3 ACF and PACF

A powerful tool for identifying time series models is autocorrelation function (ACF). It is defined for weakly stationary processes, say, X_t,

$$\rho_k = \frac{\text{cov}(X_t, X_{t-k})}{\text{var}(X_t)} = \frac{\gamma_k}{\gamma_0}. \tag{1}$$

The sample counterpart is

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T} (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^{T} (X_t - \bar{X})^2}, 0 \leq k < T-1, \tag{2}$$

where $\bar{X} = \frac{1}{T} \sum_t X_t$.

The ACF defined in (1) is meaningless for nonstationary time series. Nonetheless, the sample ACF is routinely examined for apparently nonstationary time series in empirical studies. Figure 1 examines the sample ACF's of six time series. The three categories (stationary, unit root, and seasonal time series), each of which corresponds to each row, exhibit stark differences in the form of the sample ACF.

ACF cannot take it into account that X_{t-k} might be correlated with X_t through intermediate observations, say, X_{t-l}, $0 < l < k$. PACF (Partial ACF) examines the direct correlation between X_{t-k} and X_t, excluding the intermediary effect of X_{t-l}, $0 < l < k$. This is done through multiple linear regression of X_t on X_{t-k}, controlling for intermedi-

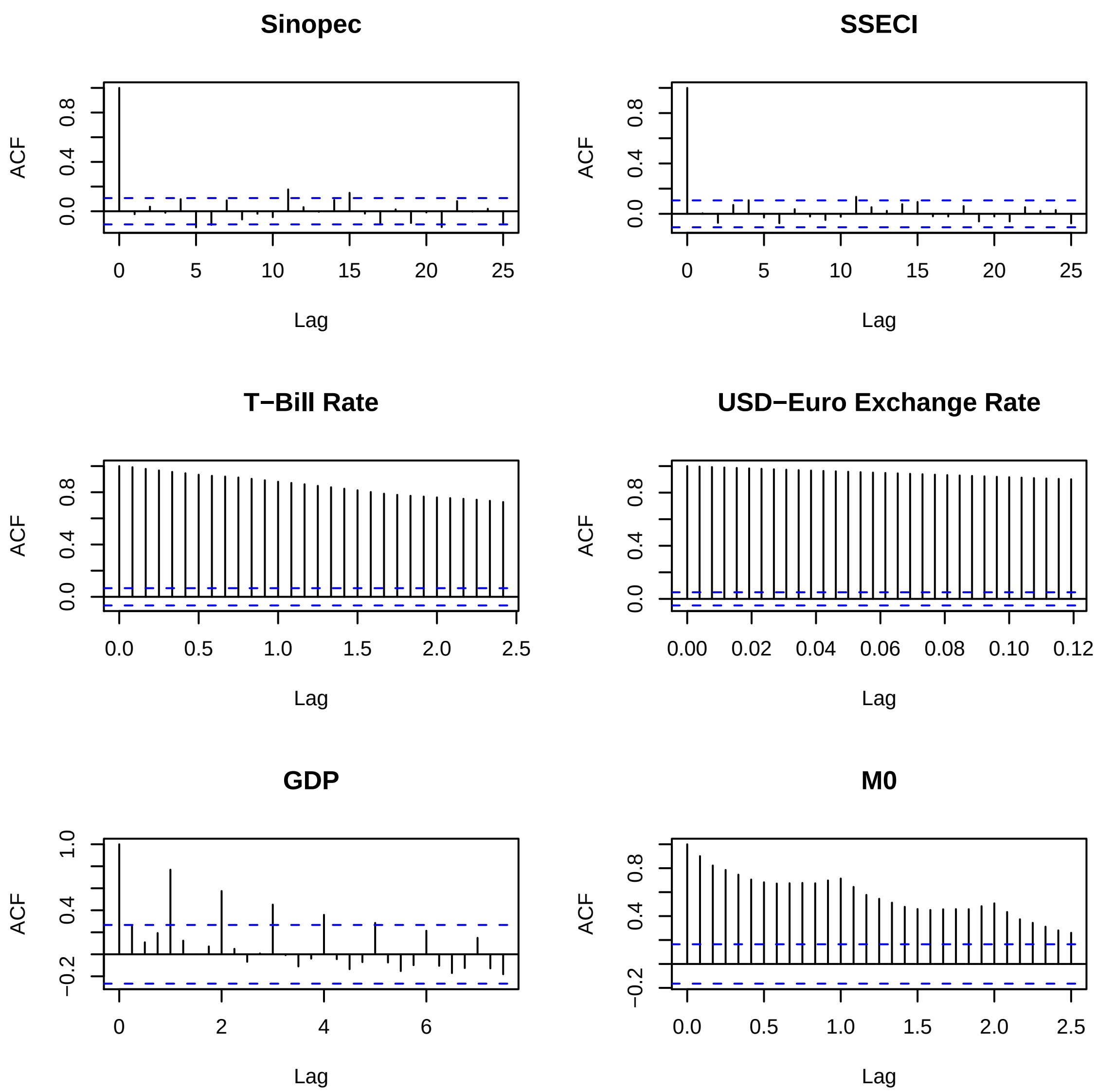

Figure 1: Sample ACF's for the data mentioned above. For the first row, each lag on X-axis corresponds to one day. For others, each unit of lag corresponds to one year. The ACF of M0 is calculated on detrended data.

aries. Specifically, to calculate k-order PACF, we write

$$X_t = \phi_{k0} + \phi_{k1} X_{t-1} + \cdots + \phi_{kk} X_{t-k} + \epsilon_t, \tag{3}$$

where ϕ_{ki}, $i = 0, ..., k$, are regression coefficients. We define k-order PACF as

$$\text{PACF}(k) = \phi_{kk}. \tag{4}$$

Figure 2 shows the PACF's of the six time series examined above. We will be better equipped to read these diagrams as we go on. For now, we start from the basics. In the rest of this chapter, we introduce basic linear models of time series in the order of white noise, autoregressive (AR) models, moving average (MA) models, ARMA, ARIMA, and seasonal models.

4.4 Nonparametric Statistics

We may estimate density functions directly using kernel smoothing. For a stationary time series of vectors $(X_t \in \mathbb{R}^d)$, the joint distribution may be obtained by

$$\hat{f}(x) = \frac{1}{Th^d} \sum_{t=1}^{T} K\left(\frac{x - X_t}{h}\right), \; x \in \mathbb{R}^d,$$

where $K(\cdot)$ is a bounded symmetric density such that

$$\lim_{\|u\| \to \infty} \|u\|^d K(u) = 0, \quad \text{and} \quad \int_{\mathbb{R}^d} \|u\|^2 K(u) du < \infty,$$

where $\|\cdot\|$ is any norm over $\mathbb{R}^d$.

To see how $Y = (Y_t \in \mathbb{R})$ may be dependent on $X = (X_t \in \mathbb{R}^d)$, we may estimate the following nonparametric regression,

$$Y_t = r(X_t) + u_t,$$

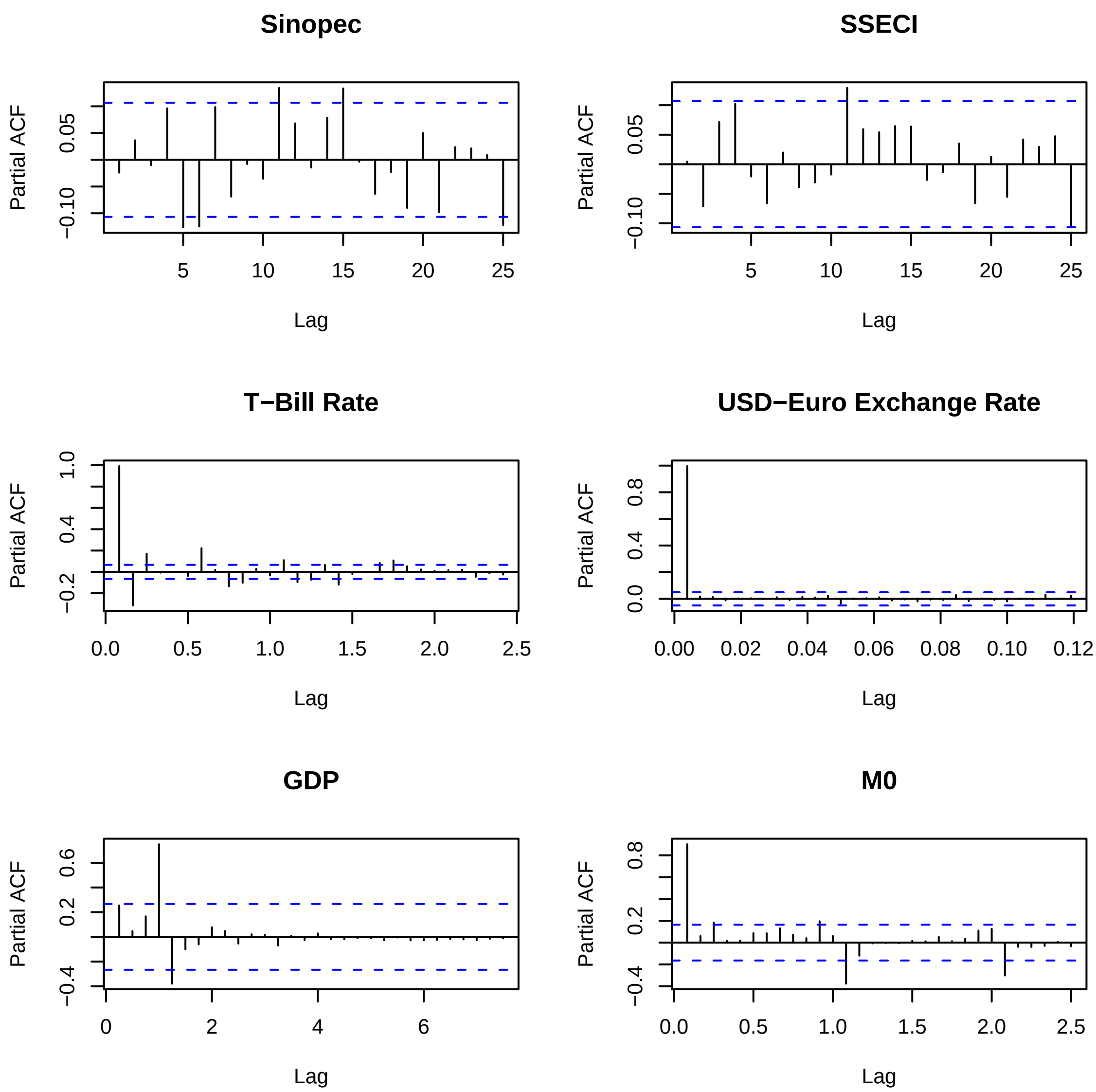

Figure 2: Sample PACF's. For the first row, each lag on X-axis corresponds to one day. For others, each unit of lag corresponds to one year. The ACF of M0 is calculated on detrended data.

where $r(\cdot)$ is a smooth function. A popular estimator of $r(\cdot)$ is given by

$$\hat{r}(x) = \frac{\sum_{t=1}^{T} Y_t K\left(\frac{x-X_t}{h}\right)}{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right)}, \quad x \in \mathbb{R}^d.$$

This estimator is called Nadaraya-Watson or locally constant estimator.

Under suitable conditions, both $\hat{f}(x)$ and $\hat{r}(x)$ behave satisfactorily as $T \to \infty$. $\hat{f}(x)$ and $\hat{r}(x)$ usually serve as preliminary analysis. We refer to Bosq (1998) for an in-depth study of nonparametric statistics for stochastic processes.

5 Important Definitions and Theorems

5.1 Mixing Properties

Mixing is a measure of how dependent between two information sets. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{B}$ and $\mathcal{C}$ be two sub σ-field of $\mathcal{F}$. We define:

- $\alpha = \alpha(\mathcal{B}, \mathcal{C}) = \sup_{B \in \mathcal{B}, C \in \mathcal{C}} |\mathbb{P}(B \bigcap C) - \mathbb{P}(B)\mathbb{P}(C)|$,
- $\phi = \phi(\mathcal{B}, \mathcal{C}) = \sup_{B \in \mathcal{B}, C \in \mathcal{C}, \mathbb{P}(B)>0} |\mathbb{P}(C) - \mathbb{P}(C|B)|$,
- $\beta = \beta(\mathcal{B}, \mathcal{C}) = \mathbb{E} \sup_{C \in \mathcal{C}} |P(C) - P(C|\mathcal{B})|$,
- $\psi = \psi(\mathcal{B}, \mathcal{C}) = \sup_{B \in \mathcal{B}, C \in \mathcal{C}, \mathbb{P}(B), \mathbb{P}(C)>0,} \left|\frac{\mathbb{P}(B \bigcap C)}{\mathbb{P}(B)\mathbb{P}(C)} - 1\right|$,
- $\rho = \rho(\mathcal{B}, \mathcal{C}) = \sup_{X \in L^2(\mathcal{B}), X \in L^2(\mathcal{C})} |\mathrm{corr}(X, Y)|$.

These coefficients satisfy the following inequalities:

$$2\alpha \leq \beta \leq \phi \leq \psi \quad \text{and} \quad 4\alpha \leq \rho \leq 2\phi^{1/2} \tag{5}$$

Definition. A process $X = (X_t)$ is α-mixing (or strongly mixing) if

$$\alpha_k = \sup_t \alpha(\sigma(X_s, s \leq t), \sigma(X_s, s \geq t+k)) \to 0, \text{ as } k \to \infty,$$

where the sup may be omitted if X is stationary. Similarly we may define β-mixing, ψ-mixing, ϕ-mixing, and ρ-mixing.

The inequalities in (5) implies that ψ-mixing $\Rightarrow$ ϕ-mixing $\Rightarrow$ β-mixing $\Rightarrow$ α-mixing, and ϕ-mixing $\Rightarrow$ ρ-mixing $\Rightarrow$ α-mixing. Hence α-mixing is the weakest among the five and ψ-mixing the strongest. ψ-mixing, ϕ-mixing, and β-mixing are often too restrictive for applications.

If X is a Gaussian stationary ϕ-mixing process, then it is m-dependent, ie, for some m, $\sigma(X_s, s \le t)$ and $\sigma(X_s, s \le t+k)$ are independent for $k > m$. And for a Gaussian process, since $\rho_k \le 2\pi\alpha_k$, α-mixing and ρ-mixing are equivalent.

Finally, we state two useful inequalities. Let X and Y be two real valued variables such that $X \in L^q(\mathbb{P})$, and $Y \in L^r(\mathbb{P})$ where $q > 1$, $r > 1$, and $\frac{1}{q} + \frac{1}{r} = 1 - \frac{1}{p}$, then

$$|\text{cov}(X,Y)| \le 2p(2\alpha)^{1/p}\|X\|_q\|Y\|_r, \tag{6}$$

where $\alpha = \alpha(\sigma(X), \sigma(Y))$. If in particular, $X \in L^\infty(\mathbb{P})$, and $Y \in L^\infty(\mathbb{P})$, then

$$|\text{cov}(X,Y)| \le 4\alpha\|X\|_\infty\|Y\|_\infty. \tag{7}$$

5.2 Convergence

In this subsection, for more generality and for consistency with standard texts on the subject, we use subscript n instead of t.

Definition: Modes of Convergence. Let (X_n) be sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let X be another random variable on the same space. We define

- $X_n \to X$ pointwise if $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$
- $X_n \to X$ almost surely (a.s.) $(X_n \to_{a.s.} X)$ if $\{\omega : X_n(\omega) \nrightarrow X(\omega)\}$ is a null set.

- $X_n \to X$ in probability ($X_n \to_p X$) if, for all $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) \to 0$.
- $X_n \to X$ in L^1 (or in mean) if $\mathbb{E}|X_n - X| \to 0$.
- $X_n \to X$ in L^2 (or in mean square) if $\mathbb{E}|X_n - X|^2 \to 0$.
- $X_n \to X$ in distribution ($X_n \to_d X$) if $F_{X_n}(x) \to F_X(x)$ for every continuity point of F_X.

We can establish that pointwise convergence implies almost sure convergence; that L^2 convergence implies L^1 convergence; that almost sure convergence and L^1 convergence imply convergence in probability; and that convergence in probability implies convergence in distribution. The inverse directions do not hold in general. However, Lévy's dominated convergence theorem (see below) provides a sufficient condition for almost sure convergence or convergence in probability to imply L^1 convergence.

L^1 and L^2 convergences are special cases of L^r convergence, which is defined as $\mathbb{E}|X_n - X|^r \to 0$ as $n \to \infty$. L^r denotes a space $\{X \sim r.v. | \mathbb{E}|X|^r < \infty\}$ with a metric $\|X - Y\|_r = (\mathbb{E}|X - Y|^r)^{1/r}$. Convergence in L^r means that the distance between X_n and X tends to 0.

Convergence in distribution is also called weak convergence, in the sense that it is equivalent to $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for every bounded and uniformly continuous f. Convergence in distribution may also be defined by convergence in characteristic functions, $\phi_n(\tau) \to \phi(\tau)$ for all $\tau \in \mathbb{R}$, where ϕ_n and ϕ are characteristics functions for X_n and X, respectively. And we have

Continuous Mapping Theorem. Let f be a continuous function. Then if $X_n \to_d X$, then $f(X_n) \to_d f(X)$. The conclusion also holds for convergence a.s. and in probability.

We also have,

Slutsky Theorem. If $X_n \to_d X$ and $Y_n \to_d c$ for a constant c, then

(a) $X_n + Y_n \to_d X + c$,

(b) $X_n Y_n \to_d cX$,

(c) $X_n/Y_n \to_d X/c$ with $c \neq 0$.

Note that it is not necessary for a sequence of random variables to be defined on the same probability space to converge in distribution to a random variable X which is also not necessarily on the same probability space as X_n. However, we do have,

Skorokhod Representation Theorem. Let (X_n) be a sequence of random variables (not necessarily all defined on the same probability space), and assume that $X_n \to X$ (with X not necessarily on the same probability space as the X_n). Let F_n be the distribution function of X_n and let F be the distribution function of X. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables Y_n and Y, all defined on $(\Omega, \mathcal{F}, \mathbb{P})$, such that Y has distribution function F and each Y_n has distribution function F_n, and $Y_n \to_{a.s.} Y$ as $n \to \infty$.

Lévy's Dominated Convergence Theorem. If a sequence of random variables $(X_n, n = 1, 2, ...)$ satisfies $X_n \to_{a.s.} X$, $|X_n| < Y$, where Y is a random variable with $\mathbb{E}Y < \infty$, then it follows that $\mathbb{E}|X| < \infty$, $\mathbb{E}X_n \to \mathbb{E}X$, and $\mathbb{E}|X_n - X| \to 0$.

The stronger result which only need $X_n \to_p X$ is also true. This theorem provides a sufficient condition for the almost sure convergence (or convergence in probability) to imply L1-convergence. The condition $|X_n| < Y$ and $\mathbb{E}Y < \infty$ could be relaxed. Instead, the sequence (X_n) should be uniformly integrable. The theorem is a special case of Lebesgue's dominated convergence theorem in measure theory.

Monotone Convergence Theorem. Let (X_n) be a non-decreasing sequence of nonnegative random variables. If $X_n \to X$ pointwise, then $X_n \to X$ in L^1.

5.3 Law of Large Numbers

5.3.1 Classical LLN

5.3.2 LLN for Dependent Processes

Let X_t be a weakly stationary process with mean μ and covariances γ_k which satisfy

$$\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty,$$

then we have

$$\bar{X}_T \equiv \frac{1}{T}\sum_{t=1}^{T} X_t \to_{L^2} \mu.$$

In words, X_t is mean ergodic if autocovariances decline fast enough.

To show this, we write

$$\begin{aligned}
\mathbb{E}\left(\bar{X}_T - \mu\right)^2 &= \frac{1}{T^2}\mathbb{E}\left(\sum_t X_t - \mu\right)^2 &\quad (8)\\
&= \frac{1}{T}\left\{\gamma_0 + \sum_{k=1}^{T-1}(T-k)/T(2\gamma_k)\right\} &\quad (9)\\
&\le \frac{1}{T}\left\{|\gamma_0| + \sum_{k=1}^{T-1}(T-k)/T(2|\gamma_k|)\right\}\\
&\le \frac{1}{T}\left\{|\gamma_0| + \sum_{k=1}^{T-1}(2|\gamma_k|)\right\}\\
&\to 0.
\end{aligned}$$

Interestingly, we have

$$T\mathbb{E}\left(\bar{X}_T - \mu\right)^2 = \sum_{k=-\infty}^{\infty} \gamma_k.$$

Note that in (**??**), autocovariances with large k are negligible and those with small k are given an approximately unit weight.

5.4 Central Limit Theorems

5.4.1 Classical CLT

Lindeberg-Lévy. Let (X_i) be a sequence of i.i.d. random variables such that $\mathbb{E}X_i = \mu$ and $\mathrm{var}(X_i) = \sigma^2 < \infty$, then

$$\sqrt{n}(\bar{X} - \mu) \to_d N(0, \sigma^2).$$

Lindeberg-Feller. Let (X_i) be a sequence of independent random variables such that $\mathbb{E}X_i = \mu_i$, $\mathrm{var}(X_i) = \sigma_i^2 < \infty$. Define $s_n^2 = \sum_{i=1}^n \sigma_i^2$ and assume $\lim_{n\to\infty} \frac{1}{n} s_n^2 = \sigma^2$. If no σ_i^2 dominates, and

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i - \mu_i)^2 I_{|X_i - \mu_i| > \epsilon\sqrt{n}} \to 0 \quad \text{for every} \quad \epsilon > 0,$$

then

$$n^{-1/2}\sum_i (X_i - \mu_i) \to_d N(0, \sigma^2).$$

Liapounov. Let $(X_{n,i})$ be a independent double array such that $\mathbb{E}|X_{n,i}|^{2+\delta} < \infty$ for some $\delta > 0$. If

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}|X_i - \mu_i|^{2+\delta} = 0,$$

then

$$\frac{\sum_{i=1}^n (X_i - \mu_i)}{s_n} \to_d N(0, 1).$$

5.4.2 CLT for Dependent Processes

A time series is m-dependent if $(..., X_{t-1}, X_t)$ and $(X_{t+m+1}, X_{t+m+2}, ...)$ are independent for all t. For example, moving average processes are m-dependent. And it is clear that 0-dependence is equivalent to independence.

CLT for m-Dependence. Let (X_t) be strictly stationary, m-dependent time series with mean μ and a finite variance. Then we have

$$\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T}(X_t-\mu)\right)\to_d N(0,V),$$

where $V=\sum_{k=-m}^{m}\gamma_X(k)$.

CLT for Linear Processes. Define $X_t=\mu+\sum_{k=-\infty}^{\infty}\varphi_k\varepsilon_{t-k}$, where (ε_t) is strong white noise with variance σ^2. If $\sum_k|\varphi_k|<\infty$, then

$$\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T}(X_t-\mu)\right)\to_d N(0,V),$$

where $V=\sigma^2\left(\sum_k\varphi_k\right)^2$.

We extend the strong mixing coefficient function $\alpha(k)$ to $[0,\infty)$ by setting $\alpha(k+h)=\alpha(k)$ for $h\in(0,1)$ for all k. The generalized α function is thus a monotone step-function decreasing from $1/2$ to zero at infinity if the underlying process is α-mixing.

We define the inverse function of $\alpha(k)$ as

$$\alpha^{-1}(u)=\inf\{x\geq 0:\alpha(x)\leq u\}=\sum_{k=0}^{\infty}I_{u<\alpha(k)}.$$

CLT for α-mixing Processes. If X_t is a strictly stationary time series with mean zero such that $\int_0^1\alpha^{-1}(u)F_{|X_0|}^{-1}(1-u)^2du<\infty$, then the series $V=\sum_k\gamma_X(k)$ converges and $\sqrt{T}\bar{X}\to_d N(0,V)$.

We define a martingale difference sequence (MDS) as a time series X_t such that X_t is $\mathcal{F}_t$-measurable and $\mathbb{E}(X_t|\mathcal{F}_{t-1})=0$, where $(\mathcal{F}_t)$ is a filtration.

CLT for MDS Let (X_t) be a martingale difference sequence with respect to the filtration $(\mathcal{F}_t)$. If

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(X_t^2|\mathcal{F}_{t-1}) &\to_p V,\\ \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(X_t^2\cdot I\{|X_t|>\epsilon\sqrt{T}\}|\mathcal{F}_{t-1}) &\to_p 0,\end{aligned}$$

then $\sqrt{T}\bar{X}\to_d N(0,V)$.

D. Bosq (1998), Nonparametric statistics for stochastic processes, Springer.