

LECTURE NOTES AND EXERCISES
STK4060 – TIME SERIES, AUTUMN 2025
September 15, 2025

EMIL A. STOLTENBERG

ABSTRACT. These are lecture notes and exercises for the times series course STK4060 (Master level) and STK9060 (PhD level) given at the Department of Mathematics, University of Oslo, for the Autumn semester 2025. The notes rely heavily on [Shumway and Stoffer \(2025\)](#), and on [Brockwell and Davis \(1991\)](#). Many thanks to my colleagues [Adam Lee](#) for letting me borrow freely from his Barcelona notes, and to Nils Lid Hjort for letting me reuse parts of his timeseries [notes from the Spring of 2022](#).

CONTENTS

1. Time series data	2
1.1. The distribution of a process	3
1.2. Measures of dependence and stationarity	4
2. Stationary ARMA-processes	8
3. Forecasting	12
3.1. L^2 and the Projection theorem	12
3.2. Linear prediction in L^2	13
4. Estimation and Large sample theory	14
References	19

1. TIME SERIES DATA

A *stochastic process* is a collection $X = \{X_t : t \in T\}$ of random variables defined on a common probability space $(\Omega, \mathcal{F}, \text{Pr})$. When the index set T is $[0, \infty)$, $[0, 1]$ or the like, we refer to X as a continuous time process. When the index set is discrete, we have a discrete time process. In time series analysis (at least how we deal with it in this course) the index set T will always be a discrete set, typically $\mathbb{N} = \{1, 2, \dots\}$, $\mathbb{N} \cup \{0\} = \{0, 1, 2, \dots\}$, or $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$.

For fixed $\omega \in \Omega$, we refer to

$$t \mapsto X_t(\omega),$$

as a *sample path* of X ; while for a fixed $t \in T$, $\omega \mapsto X_t(\omega)$ is a random variable. The actual data we find in the data files on our computer, are the realisations of a sample path of X observed at a given set of time points $t_1, \dots, t_n \in T$. sample path

A basic building block for many of the processes we work with in this course is the white noise process W .

Definition 1. The process $W = \{W_t : t \in T \subset \mathbb{Z}\}$ is white noise with variance σ^2 if $\text{E } W_t = 0$ for all t and white noise

$$\text{Cov}(X_{t+h}, X_t) = \text{E } X_{t+h} X_t = \begin{cases} \sigma^2, & h = 0, \\ 0, & h \neq 0. \end{cases}$$

We write $W \sim \text{WN}(0, \sigma^2)$.

Exercise 1.1 Some properties of white noise processes.

(a) Show that an i.i.d. process of mean zero random variables with finite variance is white noise.

(b) Let $\{Z_t : t \in \mathbb{Z}\}$ be i.i.d. standard normal random variables, and define

$$W_t = Z_t Z_{t-1}, \quad t \in \mathbb{Z}.$$

Show $W = \{W_t : t \in \mathbb{Z}\}$ is a white noise process, but that it is not independent.

(c) Construct your own example of a white noise process that is not i.i.d.

Here are examples of some other processes that time and again this course.

Exercise 1.2 If we smooth white noise we obtain a moving average process. Let $W \sim \text{WN}(0, \sigma^2)$ and define the process MA(1)

$$X_t = W_t + \theta W_{t-1}.$$

This is a moving average process of order 1.

(a) Show that $\text{E } X_t = 0$ for all t and that

$$\text{Cov}(X_{t+h}, X_t) = \begin{cases} \sigma^2(1 + \theta^2), & h = 0, \\ \sigma^2\theta, & |h| = 1, \\ 0, & |h| \geq 2. \end{cases}$$

(b) Define $Y_t = \mu + X_t$, and suppose that we observe $Y_1, \dots, Y_n$ and take the average $\bar{Y}_n = n^{-1} \sum_{t=1}^n Y_t$ as an estimate of μ . Show that $\bar{Y}_n$ is unbiased and find an expression for the variance of this estimator.

Exercise 1.3 Let $W = \{W_t : t \geq 0\}$ be i.i.d. $N(0, \sigma^2)$ and for $t \geq 1$ define AR(1)

$$X_t = \phi X_{t-1} + W_t,$$

with $X_0 = W_0$, an autoregressive process of order 1.

(a) Let $\phi = 0.8$ and $\sigma^2 = 1$. For $t = 0, \dots, 500$ make a plot of a few realisations of this time series.

(b) In your simulations, try some other values of ϕ , and observe how the process changes behaviour.

(c) Deduce that $X_t = \sum_{j=0}^t \phi^j W_{t-j}$. Show that $E X_t = 0$ and that for $h \geq 0$

$$\text{Cov}(X_{t+h}, X_t) = \sigma^2 \phi^h \frac{1 - \phi^{2(t+1)}}{1 - \phi^2}.$$

Exercise 1.4 Let $W = \{W_t: t \geq 0\}$ be white noise and define the random walk with drift model

$$X_t = \mu + X_{t-1} + W_t, \quad t \geq 1$$

with $X_0 = x_0$, say.

(a) Show that $X_t = x_0 + \mu t + \sum_{j=1}^t W_j$. For some values of x_0 and μ , plot a few sample paths of this process along with the line $E X_t = x_0 + \mu t$.

(b) Show that $\text{Cov}(X_s, X_t) = \min(s, t)\sigma^2$.

1.1. The distribution of a process. Consider a random variable X defined on a probability space $(\Omega, \mathcal{F}, \Pr)$, i.e., a measurable function $X: \Omega \rightarrow \mathbb{R}$, where $\mathbb{R}$ is equipped with the Borel- σ -algebra $\mathcal{B}$. The random variable X induces a probability measure $P_X = \Pr X^{-1}$ on $(\mathbb{R}, \mathcal{B})$, called the distribution of X . We write $X \sim Y$ if X and Y have the same distribution. The cumulative distribution function (c.d.f.) of X is $F_X(x) = P_X(-\infty, x] = \Pr(X \leq x)$. The characteristic function of a random variable X is

$$\varphi(t) = E \exp(itX) = E \{\cos(tX) + i \sin(tX)\}.$$

Both c.d.f.s and c.d.s generalise to higher dimensions: If $X = (X_1, \dots, X_k)^t$ is a random vector, its c.d.f. is $F(x_1, \dots, x_k) = \Pr(X_1 \leq x_1, \dots, X_k \leq x_k)$ and its c.f. is $\varphi(t) = \varphi(t_1, \dots, t_k) = E \exp(it^t X) = E \exp(it_1 X_1 + \dots + it_k X_k)$. To prove that the c.d.f. and the c.f. of a random variable determine its distribution, one may appeal to Dynkin's lemma (see any probability textbook, e.g., Williams (1991, Lemma 1.6(b), p. 19)) which says that:

Lemma 1. *If two probability measures agree on a collection of sets closed under finite intersections (i.e., a π -system), then they agree on the σ -algebra generated by this collection of sets.*

Exercise 1.5 The c.d.f. F and the c.f. φ characterize distributions. We here prove things for random variables, but they generalise to random vectors.

(a) Let F_X and F_Y be the c.d.f.s of the random variable X and Y , respectively, and suppose that $F_X = F_Y$. Show that the collection of sets

$$\Pi = \{(-\infty, x]: x \in \mathbb{R}\},$$

is closed under finite intersections, and that it generates the Borel- σ -algebra on $\mathbb{R}$. Conclude that $X \sim Y$.

(b) For a random variable with c.d.f. F and c.f. φ we have the *inversion formula*

$$F(b) - F(a) = \lim_{\sigma \rightarrow 0} \frac{1}{2\pi} \int \frac{\exp(-itb) - \exp(-ita)}{-it} \varphi(t) \exp(-\frac{1}{2}t^2\sigma^2) dt,$$

valid for all continuity points $a < b$ of F . Let X and Y be random variables with characteristic functions φ_X and φ_Y , respectively. Show that if $\varphi_X(t) = \varphi_Y(t)$ for all $t \in \mathbb{R}$, then $X \sim Y$.

random walk

Dynkin's lemma

(c) As an extra, show that if φ is integrable, i.e. $\int |\varphi(t)| dt < \infty$, then X has density f with respect to Lebesgue measure (so $F(x) = \int_{-\infty}^x f(y) dy$) given by

$$f(x) = \frac{1}{2\pi} \int \exp(-itx) \varphi(t) dt.$$

Finite dimensional distributions determine the distribution of a process. In particular, if $X = \{X_t : t \in T\}$ and $Y = \{Y_t : t \in T\}$ are two stochastic processes, then X and Y have the same distribution, we write $X \sim Y$, if and only if

$$(X_{t_1}, \dots, X_{t_k}) \sim (Y_{t_1}, \dots, Y_{t_k}),$$

for all finite collections $t_1, \dots, t_k \in T$, $k \geq 1$. See, for example, [Kallenberg \(2002, Prop. 3.2, p. 48\)](#). A question that arises is when a collection of finite dimensional distribution corresponds to a process. For an index set $T \subset \mathbb{R}$, let

$$\tau = \{(t_1, \dots, t_k) : t_1 < \dots < t_k \in T, k \geq 1\}.$$

Then $\{F_t : t \in \tau\}$ is a collection of finite dimensional distributions. Write also t_{-i} for the vectors where the i th element is deleted, e.g., if $t = (t_1, t_2, t_3)$, then $t_{-2} = (t_1, t_3)$.

Theorem 2. (*Kolmogorov, Daniell*). $\{F_t : t \in \tau\}$ are the finite dimensional distributions of some process if and only if

$$\lim_{x_i \rightarrow \infty} F_t(x) = F_{t_{-i}}(x_{-i}). \quad (1)$$

Proof. See for example [Billingsley \(1995\)](#). □

Write $\{\phi_t(\xi) : t \in \tau\}$ for a collection of finite dimensional c.f.s, where $\phi_t(\xi) = \int \exp(i\xi^t x) dF_t(x)$. A condition equivalent to (1) is then that

$$\lim_{\xi_i \rightarrow 0} \phi_t(\xi) = \phi_{t_{-i}}(\xi_{-i}).$$

Exercise 1.6 Let $W_1, W_2, \dots$ be i.i.d. $N(0, 1)$, and define the process $X = \{X_t : t \in \mathbb{N} \cup \{0\}\}$ by $X_t = \sum_{j=1}^t W_j$ for $t \geq 1$, and $X_0 = 0$.

(a) Show that the process X just defined is so that (i) $X_0 = 0$; (ii) it has independent increments; and (iii) $X_t - X_s \sim N(0, t - s)$ for all $t > s \geq 0$.

(b) Suppose that we were only given (i), (ii), and (iii). Show that these describe the finite dimensional distributions of a process.

(c) Assume that we retain (i) and (ii), but replace (iii) with (iii)', requiring that $X_t - X_s \sim N(0, (t - s)^2)$, say. Show that there is no process corresponding to these finite dimensional distributions.

1.2. Measures of dependence and stationarity.

Definition 2. The autocovariance function of the process $\{X : X_t \in T\}$

autocovariance

$$\gamma(s, t) = \text{Cov}(X_s, X_t) = E(X_t - E X_t)(X_s - E X_s),$$

provided $E X_t^2 < \infty$.

Definition 3. The autocorrelation function of the process $\{X : X_t \in T\}$

autocorrelation

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}},$$

provided $E X_t^2 < \infty$.

We have already computed some autocovariance functions in the exercises above. A property making a time series amenable to analysis is stationarity.

stationarity

Definition 4. The process $\{X_t: t \in \mathbb{Z}\}$ is *stationary* if $E X_t^2 < \infty$ and $E X_t = \mu$ for all t , and

$$\gamma(s, t) = \gamma(s + h, t + h), \quad \text{for all } s, t, h \in \mathbb{Z}.$$

This form of stationarity is also often referred to as weak stationarity, covariance stationarity, as well as other things.

strict stationarity

Definition 5. The time series $\{X_t: t \in \mathbb{Z}\}$ is *strictly stationary* if

$$(X_{t_1}, \dots, X_{t_k}) \sim (X_{t_1+h}, \dots, X_{t_k+h})$$

for all $k \geq 1$ and all $t_1, \dots, t_k, h \in \mathbb{Z}$.

The autocovariance function $\gamma(s, t)$ of a stationary time series only depends on s and t through $|t - s|$, i.e.,

$$\gamma(t + h, t) = \text{Cov}(X_{t+h}, X_t) = \text{Cov}(X_h, X_0) = \gamma(t, 0).$$

This motivates the following definition.

Definition 6. The autocovariance function of a stationary time series $\{X_t: t \in \mathbb{Z}\}$ is

$$\gamma(h) = \text{Cov}(X_h, X_0), \quad \text{for } h \in \mathbb{Z}.$$

Notice that the autocovariance function of a stationary time series is symmetric around zero,

$$\gamma(h) = \gamma(-h), \quad \text{for all } h \in \mathbb{Z}.$$

Exercise 1.7 In this exercise we look at the relation between stationarity and strict stationarity. Let $X = \{X_t: t \in \mathbb{Z}\}$ be a times series.

- (a) Suppose that X is strictly stationary such that $E X_t^2 < \infty$. Show that X is stationary.
- (b) Construct a strictly stationary process that is not stationary.
- (c) Exhibit a stationary process that is not strictly stationary.
- (d) Show that if X is a stationary Gaussian process, then X is strictly stationary.

Exercise 1.8 Is the process (strictly) stationary?

- (a) Let A and B be independent mean zero random variables with unit variance, and consider the process $X_t = A \cos(\theta t) + B \sin(\theta t)$ for $\theta \in [-\pi, \pi]$ and $t \in \mathbb{Z}$. Is X_t stationary?
- (b) Consider the process $X_t = \phi X_{t-1} + W_t$ for $t \geq 1$ and $X_0 = W_0$ for i.i.d. random variables $W_0, W_1, \dots$ with $E W_t = 0$ and $E W_t^2 = \sigma^2$. Show that this process is *not* stationary.

Exercise 1.9 Suppose that $Y_1, \dots, Y_n$ is a sample from a stationary time series with $E Y_t = \mu$ for all $t \in \mathbb{Z}$ and

$$\gamma(h) = \text{Cov}(Y_{t+h}, Y_t) = \tau^2 \rho(|h|),$$

where $\rho(h)$ is the autocorrelation function. An estimator for μ is $\bar{Y}_n = n^{-1} \sum_{t=1}^n Y_t$, and had the $Y_1, \dots, Y_n$ been i.i.d. with variance τ^2 , we would have had $E \bar{Y}_n = \mu$ and $\text{Var}(\bar{Y}_n) = \gamma(0)/n = \tau^2/n$.

(a) They are, however, not i.i.d.. Show that

$$\text{Var}(\bar{Y}_n) = \frac{\tau^2}{n} \left\{ 1 + 2 \sum_{h=1}^n (1 - h/n) \rho(h) \right\} = \frac{\tau^2}{n} \sum_{h=-n}^n (1 - |h|/n) \rho(|h|).$$

(b) With $\rho(|h|) = \phi^{|h|}$ for $|\phi| < 1$, show that

$$\text{Var}(\bar{Y}_n) = \frac{\tau^2}{n} \frac{1 + \phi}{1 - \phi} + o(1/n).$$

With a positive correlation we see that the variance of $\bar{Y}_n$ becomes bigger than under independence.

Exercise 1.10 Let $(X_n)_{n \geq 1}$ and X be random variables.

(a) Show that $X_n \rightarrow X$ almost surely if and only if the event

$$\cap_{n \geq 1} \cup_{k \geq n} \{\|X_k - X\| \geq \varepsilon\},$$

has probability zero for all $\varepsilon > 0$.

(b) Show that if $X_n \rightarrow X$ almost surely, then $X_n \rightarrow X$ in probability.

(c) Show that if $X_n \rightarrow X$ in probability, then $X_n \rightarrow X$ in distribution.

(d) Show that if $E(X_n - X)^p \rightarrow 0$, then $X_n \rightarrow X$ in probability.

Exercise 1.11 Let $A_1, A_2, \dots$ be a sequence of events. Consider $\cap_{n \geq 1} \cup_{m \geq n} A_m$, the full-sequence event corresponding to the A_n occurring infinitely often.

(a) Show that if $\sum_{n=1}^{\infty} \Pr(A_n) < \infty$, then $\Pr(\cap_{n \geq 1} \cup_{m \geq n} A_m) = 0$. This is the Borel–Cantelli lemma

Borel–Cantelli

(b) Assume in addition that the $A_1, A_2, \dots$ are independent events. Show that if $\sum_{n=1}^{\infty} \Pr(A_n)$ is divergent, then $\Pr(\cap_{n \geq 1} \cup_{m \geq n} A_m) = 1$. To show this, you may (prove and) use the inequality $1 + x \leq \exp(x)$, valid for all $x \in \mathbb{R}$.

(c) Let $(X_n)_{n \geq 1}$ be a sequence of random variables such that $X_n \rightarrow_p X$. Use the Borel–Cantelli lemma to find a subsequence $(X_{n_k})_{k \geq 1}$ such that $X_{n_k} \rightarrow X$ almost surely.

Exercise 1.12 Let (x_j) be a sequence of real numbers.

(a) Show that if (x_j) is absolutely summable, i.e., $\sum_j |x_j|$ is convergent, then $\sum_j x_j$ is convergent.

(b) Show that if (x_j) is absolutely summable, then $\sum_j x_j^2$ is convergent. Exhibit an example showing that the converse is false.

Exercise 1.13 Increasing sequences, and convergence of sums.

(a) Let $0 \leq Y_1 \leq Y_2 \leq Y_3 \leq \dots$ be random variables such that $Y_n \rightarrow_p Y$. Show that $Y_n \rightarrow Y$ almost surely.

(b) Let $X_1, X_2, \dots$ be random variable such that $\sum_{j=1}^{\infty} E|X_j|$ is convergent. Show that $\sum_{j=1}^{\infty} X_j$ is almost surely convergent.

(c) Let $X_1, X_2, \dots$ be independent random variables. The Kolmogorov three series theorem says that $\sum_{j=1}^{\infty} X_j$ is convergent if for some $A > 0$, and only if for any $A > 0$, the three series

Kolmogorov's
three series
theorem

$$\sum_{j=1}^{\infty} \Pr(|X_j| > A), \quad \sum_{j=1}^{\infty} E X_j I_{|X_j| \leq A}, \quad \text{Var}(X_j I_{|X_j| \leq A}),$$

are convergent. Verify that the condition from (b), namely $\sum_{j=1}^{\infty} E|X_j|$ is convergent, implies convergence of the three series of the three series theorem.

(d) (xx prove the three series theorem. need some more additional results xx)

(e) Let $(\xi_n)_{n \geq 1}$ random variables such $\Pr(\xi_n = 1) = \Pr(\xi_n = -1) = 1/2$ for all n , and set $X_n = \xi_n/n$ for $n = 1, 2, \dots$. Show that $(X_n)_{n \geq 1}$ satisfies ??&??&?? of the three series theorem, but $\sum_{j=1}^{\infty} E|X_j| = \sum_{j=1}^{\infty} (1/j)$ is an harmonic series and diverges.

(f)

Exercise 1.14 Let $W = \{W_t \in t \in \mathbb{Z}\}$ be white noise, and define the process

$$X_t = \phi X_{t-1} + W_t, \quad t \in \mathbb{Z}. \quad (2)$$

This is an autoregressive process of order 1, assuming X_t is stationary. We give a proper definition of such processes below.

(a) ‘Substitute backwards’ to show that $X_t = \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j W_{t-j}$.

(b) Assuming that the AR(1)-process is stationary and that $|\phi| < 1$, show that $E(X_t - \sum_{j=0}^{k-1} \phi^j W_{t-j})^2 \rightarrow 0$ as $k \rightarrow \infty$. We are thus justified in writing

$$X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}, \quad (3)$$

with this equality being true in the mean square sense.

causal solution

(c) Show that the equality (3) is true almost surely as well. The solution in (3) is said to be *causal* (or future-independent), as X_t only depends on W_s for $s \leq t$.

(d) Show that the process in (3) satisfies equation (2), and that X_t defined by (3) is indeed stationary, in particular,

$$\gamma(h) = \sigma^2 \frac{\phi^{|h|}}{1 - \phi^2}.$$

Contrast this with what you found in Ex. 1.8(b), and we see that the modelling choice $t \in \mathbb{Z}$ versus $t \in \mathbb{N}$ plays a role in determining stationarity.

Exercise 1.15 Consider the regression model

$$Y_i = \beta X_i + \varepsilon_i, \quad i = 1, \dots, n$$

where $(X_1, \varepsilon_1), \dots, (X_n, \varepsilon_n)$ are i.i.d. pairs, X_i is independent of ε_i for each i ; $E\varepsilon_1 = 0$ and $E\varepsilon_1^2 = \sigma^2$; while $EX_1 = \mu_X$ and $\text{Var } X_1 = \sigma_X^2$.

(a) Show that with $\hat{\beta}_n$ the least squares estimator, we can write $\sqrt{n}(\hat{\beta}_n - \beta) = n^{-1/2} \sum_{i=1}^n X_i \varepsilon_i / \{n^{-1} \sum_{i=1}^n X_i^2\}$.

(b) Show that

$$\sqrt{n}(\hat{\beta}_n - \beta) \rightarrow_d N\left(0, \frac{\sigma^2}{\sigma_X^2 + \mu_X^2}\right).$$

(c) Things are more complicated when the covariates are fixed, i.e., when the model is $Y_i = \beta x_i + \varepsilon_i$, with $x_1, \dots, x_n$ fixed numbers. The reason being that the summands in $\sum_{i=1}^n x_i \varepsilon_i$ are no longer identically distributed. One must then use the Lindeberg CLT (see for example Theorem 5 in the lecture notes [Stoltenberg \(2025\)](#)). Write $s_n^2 = \sum_{i=1}^n x_i^2$, and let $\hat{\beta}_n = \sum_{i=1}^n x_i Y_i / s_n^2$ be the least squares estimator. Show that if $\max_{j \leq n} x_j^2 / \sum_{i=1}^n x_i^2 \rightarrow 0$, then $s_n(\hat{\beta}_n - \beta) \rightarrow_d N(0, \sigma^2)$.

Exercise 1.16 Let $X = \{X_t : t \in \mathbb{Z}\}$ be the causal AR(1)-process $X_t = \phi X_{t-1} + W_t$, where $(W_t)_{t \in \mathbb{Z}}$ are i.i.d. $N(0, \sigma^2)$. By Ex. 1.14 it has the representation $X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$. Suppose that we observe $X_1, X_2, \dots, X_n$. The least squares estimator for ϕ is $\hat{\phi}_n = \sum_{t=1}^n X_t X_{t-1} / \{\sum_{t=1}^n X_{t-1}^2\}$.

(a) Show that we may write

$$\sqrt{n}(\hat{\phi}_n - \phi) = \frac{n^{-1/2} \sum_{t=1}^n X_{t-1} W_t}{n^{-1} \sum_{t=1}^n X_{t-1}^2}.$$

(b) To prove to the sequence above is approximately normal, we can use a martingale CLT, see for example Theorem 9 in [Stoltenberg \(2025\)](#). Let

$$\mathcal{F}_t = \sigma(\dots, W_{-1}, W_0, W_1, \dots, W_t),$$

and verify that $X_{t-1} W_t$ is $\mathcal{F}_t$ -measurable for each t ; that $E(X_{t-1} W_t \mid \mathcal{F}_{t-1}) = 0$, and that $E(X_{t-1}^2 W_t^2) < \infty$ for each t . These three conditions ensure that [Theorem 9](#) is applicable.

(c) Show that $X_t \sim N(0, \sigma^2/(1 - \phi^2))$ for each $t \in \mathbb{Z}$. This, of course, is due to the normality assumption. Deduce that for $h \geq 0$

$$\text{Cov}(X_{t+h}^2, X_t^2) = \sigma^4 \frac{2\phi^{2h}}{(1 - \phi^2)^2}.$$

(d) It is a bit tedious, but show that

$$\text{Var}\left(\frac{1}{n} \sum_{t=1}^n X_{t-1}^2\right) = \frac{1}{n} \frac{2\sigma^4}{(1 - \phi^2)^2} + \frac{4\sigma^4}{n^2} \frac{\phi^2}{(1 - \phi^2)^3} \left(n - \frac{1 - \phi^{2n}}{1 - \phi^2}\right).$$

(e) We first check Condition (i) of Theorem 9. Show that

$$\frac{1}{n} \sum_{t=1}^n E(X_{t-1}^2 W_t^2 \mid \mathcal{F}_{t-1}) = \frac{\sigma^2}{n} \sum_{t=1}^n X_{t-1}^2 \rightarrow_p \frac{\sigma^4}{1 - \phi^2}.$$

Here the law of large numbers is not immediately applicable as the summands are not independent, but Chebyshev's inequality can be used.

(f) We must check the Lindeberg condition, namely that for any $\varepsilon > 0$,

$$\frac{1}{n} \sum_{t=1}^n E\{(X_{t-1} W_t)^2 I_{|X_{t-1} W_t| \geq \sqrt{n}\varepsilon} \mid \mathcal{F}_{t-1}\} \rightarrow_p 0. \quad (4)$$

One may try to handle this directly. It is more convenient, I think, to go via the Lyapunov condition. In general, let $(\xi_{n,i}, \mathcal{G}_{n,i})_{1 \leq i \leq n, n \geq 1}$ be a square integrable martingale difference array. Assume that this array satisfies the (conditional) Lyapunov condition, i.e., that for some $\delta > 0$, $\sum_{i=1}^n E(|\xi_{n,i}|^{2+\delta} \mid \mathcal{G}_{n,i-1}) \rightarrow_p 0$, and show that this implies the Lindeberg condition $\sum_{i=1}^n E(\xi_{n,i}^2 I_{|\xi_{n,i}| \geq \varepsilon} \mid \mathcal{G}_{n,i-1}) \rightarrow_p 0$, for any $\varepsilon > 0$. Deduce that (4) is indeed satisfied.

(g) Deduce from the above that $\sqrt{n}(\hat{\phi}_n - \phi) \rightarrow_d N(0, (1 - \phi^2))$.

(h) Attempt to extend this convergence in distribution result to the more general case where $W_t \sim \text{WN}(0, \sigma^2)$, perhaps starting with the i.i.d. case.

2. STATIONARY ARMA-PROCESSES

Autoregressive moving average processes consists of an autoregressive part and a moving average part.

Definition 7. The process $X = \{X_t : t \in \mathbb{Z}\}$ is an ARMA(p, q) process if X is stationary and for every $t \in \mathbb{Z}$

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}, \quad (5)$$

where $W \sim \text{WN}(0, \sigma^2)$ is white noise. We say that X is an ARMA(p, q) with mean μ if $X_t - \mu$ is ARMA(p, q).

backshift
operator

Introduce the backshift operator (or lag operator) B , defined by $B^j X_t = X_{t-j}$ for $j \in \mathbb{Z}$; and also the AR-operator (or AR lag polynomial)

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p,$$

and the MA-operator (or MA lag polynomial)

$$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

The equation in (5) may then be written

$$\phi(B)X_t = \theta(B)W_t.$$

AR(p)

If $\theta(B) = 1$ we obtain an AR(p)-process $\phi(B)X_t = W_t$; and if $\phi(B) = 1$ we obtain a MA(q)-process $X_t = \theta(B)W_t$.

MA(q)

If $a(B) = a_0 + a_1 B + \cdots + a_k B^k$ some lag polynomial, then $\phi(B)X_t = \theta(B)W_t$ and $a(B)\phi(B)X_t = a(B)\theta(B)W_t$ describe the same ARMA-process. We avoid this parameter redundancy by assuming that the lag polynomials are always in their simplest form, that is, they have no common factors.

parameter
redundancy

Exercise 2.1 Consider the processes AR(1) process $X_t = \phi X_{t-1} + W_t$ and the (seemingly) ARMA(2, 1)-process $Y_t = (\phi - \theta)Y_{t-1} + \phi\theta Y_{t-2} + W_t + \theta W_{t-1}$. Show that these are indeed the same process. Perhaps write also a little R-script where you see that the sample paths overlap.

causal ARMA

Definition 8. An ARMA(p, q)-process is causal if there exists a sequence of constants $\psi_0, \psi_1, \dots$ with $\psi_0 = 1$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}, \quad t \in \mathbb{Z}.$$

From Ex. 1.14 we see that the AR(1)-process $X_t = \phi X_{t-1} + W_t$ is causal provided $|\phi| < 1$, in which case $\psi_j = \phi^j$ for $j = 0, 1, 2, \dots$

invertible
ARMA

Definition 9. An ARMA(p, q)-process is invertible if there exists a sequence of constants $\pi_0, \pi_1, \dots$ with $\pi_0 = 1$ such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and

$$W_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t \in \mathbb{Z}.$$

Exercise 2.2 Consider the MA(1) process $X_t = W_t + \theta W_{t-1}$. Show that if $|\theta| < 1$ then this process is invertible, and $W_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j}$ almost surely.

Introduce the AR(p)-polynomial

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p, \quad z \in \mathbb{C},$$

and the MA(q)-polynomial

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q, \quad z \in \mathbb{C}.$$

A process being causal or invertible is connected to the nature of the roots of the $\phi(z)$ and $\theta(z)$.

Exercise 2.3 Consider the AR(1) process $\phi(B)X_t = W_t$, where $\phi(z) = 1 - \phi z$. We know from Ex. 1.14 that if $|\phi| < 1$, then X is causal with $\psi_j = \phi^j$ for $j = 0, 1, 2, \dots$

- (a) Conversely, show that if X is causal, then $|\phi| < 1$.
- (b) Show that $\phi(z) = 1 - \phi z \neq 0$ for all z such that $|z| \leq 1$ if and only if $|\phi| < 1$.
- (c) Conclude that X is causal if and only if the root of $\phi(z) = 1 - \phi z$ lies outside the unit circle.

An AR(1)-process can be expressed as $\phi(B)X_t = (1 - \phi B)X_t = W_t$. Asking for this process to be causal is like asking for an inverse operator $\phi(B)^{-1}$ to exist, that is, an operator such that

$$X_t = \phi(B)^{-1}\phi(B)X_t = \phi(B)^{-1}W_t,$$

so $\phi(B)^{-1}\phi(B) = 1$. As we have seen in Ex. 2.3, when the root of $\phi(z) = 1 - \phi z$ lies outside the unit circle we can write $X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$, and thus

$$\phi(B)^{-1} = (1 - \phi B)^{-1} = \sum_{j=0}^{\infty} \phi^j B^j. \quad (6)$$

We must show that $(1 - \phi B)^{-1}$ is well-defined as an operator from stationary processes to stationary processes.

Exercise 2.4 Let $\phi(B)^{-1}$ be as defined in (6).

- (a) Show that if $Y = \{Y_t : t \in \mathbb{Z}\}$ is stationary, then the process Z defined by $Z_t = \phi(B)^{-1}Y_t = (\sum_{j=0}^{\infty} \phi^j B^j)Y_t$, exists (in the L^2 sense), and that Z is stationary. This is essentially just repeating the arguments from Ex. 1.14.
- (b) Verify that $(1 - \phi B) \sum_{j=0}^{\infty} \phi^j B^j = 1$.

Exercise 2.5 Higher order polynomials can be inverted by factorization. In this exercise we look at the AR(2)-polynomial. The generalisation to higher order polynomials is essentially the proof of Theorem 3 below. The Fundamental theorem of algebra says that for any choice of complex numbers $a_0, a_1, \dots, a_p$ such that $a_p \neq 0$, the equation

$$a_0 + a_1 z + \dots + a_p z^p = 0,$$

has at least one solution $z \in \mathbb{C}$.

- (a) Prove the factorization

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = (1 - \lambda_1 z)(1 - \lambda_2 z),$$

where $\lambda_j = 1/r_j$ in terms of the roots r_1, r_2 of $\phi(z)$ (which need not be distinct).

- (b) If the roots r_1 and r_2 lie outside the unit circle, we have from our effort above that $(1 - \lambda_k B)^{-1} = \sum_{j=0}^{\infty} \lambda_k^j B^j$ for $k = 1, 2$. We may define

$$(1 - \phi_1 B - \phi_2 B^2)^{-1} = (1 - \lambda_1 B)^{-1}(1 - \lambda_2 B)^{-1}.$$

Show that

$$(1 - \phi_1 B - \phi_2 B^2)^{-1} = \frac{\lambda_1}{\lambda_1 - \lambda_2} \sum_{j=0}^{\infty} \lambda_1^j B^j + \frac{\lambda_2}{\lambda_2 - \lambda_1} \sum_{j=0}^{\infty} \lambda_2^j B^j,$$

if r_1 and r_2 are distinct, while $(1 - \phi_1 B - \phi_2 B^2)^{-1} = \sum_{j=0}^{\infty} j \lambda^j B^j$ if there is one (real) root.

(c) Deduce from previous efforts that when the roots of $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$ are outside the unit circle, $(1 - \phi_1 B - \phi_2 B^2)^{-1}$ is a well-defined operator mapping stationary processes to stationary processes. In other words, when $\phi(z) \neq 0$ for all $|z| \leq 1$, then the AR(2)-process $\phi(B)X_t = (1 - \phi_1 B - \phi_2 B^2)X_t = W_t$ is causal.

(d) Show that for an AR(2)-process to be causal, the parameters (ϕ_1, ϕ_2) must be such that

$$\phi_2 + \phi_1 < 1, \quad \phi_2 - \phi_1 < 1, \quad |\phi_2| < 1.$$

We have the following theorem, whose proof is more or less contained in the above exercises (this is Property 3.1 in [Shumway and Stoffer \(2025, p. 97\)](#)).

causal ARMA
iff |roots| > 1

Theorem 3. An ARMA(p, q)-process is causal if and only if $\phi(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$.

Proof. See Appendix B.3 in [Shumway and Stoffer \(2025\)](#) or make your own proof by generalizing the above exercises. $\square$

explosive AR(1)

Exercise 2.6 Let $X_t = \phi X_{t-1} + W_t$ for $t \in \mathbb{Z}$ be an AR(1)-process, and assume that $|\phi| > 1$. In view of the above theorem, this process is clearly not causal. Such models are called *explosive*. Nevertheless, in this exercises we will see that we can obtain a stationary solution.

(a) Show that by manipulating the AR(1) equation and ‘substituting forwards’, we can write $X_t = \phi^{-k} X_{t+k} - \sum_{j=1}^k \phi^{-j} W_{t+j}$.

non-causal
solution

(b) Using the same arguments as in Ex. 1.14, deduce that

$$X_t = - \sum_{j=1}^{\infty} \phi^{-j} W_{t+j}, \quad (7)$$

is the stationary solution to (2) when $|\phi| > 1$. The solution in (7) is *not* causal, as X_t depends on the future innovation terms $W_{t+1}, W_{t+2}, \dots$

(c) Show that X_t also satisfies the causal AR(1) equations

$$X_t = \phi^{-1} X_{t-1} + \widetilde{W}_t,$$

for a suitable white noise process $\widetilde{W}_t \sim \text{WN}(0, \widetilde{\sigma}^2)$. In particular, find an expression for $\widetilde{\sigma}^2$.

invertible
MA(1)

Exercise 2.7 We start by studying the concept of invertibility in the setting of an MA(1)-process. Let $X_t = W_t + \theta W_{t-1}$ for $t \in \mathbb{Z}$ be an MA(1)-process.

(a) Mimic the steps taken in Ex. 1.14 to show that

$$W_t = (-\theta)^k W_{t-k} + \sum_{j=0}^{k-1} (-\theta)^j X_{t-j}.$$

(b) Repeating arguments from previous exercises, show that if $|\theta| < 1$, then

$$W_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j},$$

in L^2 , and indeed also almost surely. Thus, the MA(1)-process (which is an ARMA(0, 1)-process) with $|\theta| < 1$ is invertible, see Def. 9

By subtracting θW_{t-1} on both sides in $X_t = W_t + \theta W_{t-1}$ the MA(1)-process can be written $W_t = -\theta W_{t-1} + X_t$ (which is what you did in Ex. 2.7(a)). Thus,

$$\theta(B)W_t = X_t, \quad \text{with} \quad \theta(z) = 1 + \theta z.$$

We now see that this is very similar to the theory developed for the AR(1)-process, invertibility of the AR-polynomials, etc., in the exercises above. The proof of the following theorem is therefore similar to the proof of Theorem 3.

Theorem 4. *An ARMA(p, q)-process is invertible if and only if $\theta(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$.*

ARMA
invertible iff
|roots| > 1

Proof. Mimic the proof of Theorem 3. □

3. FORECASTING

Forecasting is the problem of predicting the values of $X_{n+1}, X_{n+2}, \dots$ based on having observed $\{X_1, \dots, X_n\}$. We mostly concentrate on the setting where $\{X_1, \dots, X_n\}$ is a sample from a mean zero stationary time series $X = \{X_t : t \in \mathbb{Z}\}$.

3.1. L^2 and the Projection theorem. The time series we are dealing with are defined on a probability space $(\Omega, \mathcal{F}, \text{Pr})$. We let $L^2 = L^2(\Omega, \mathcal{F}, \text{Pr})$ denote all (equivalence classes for a.s. equality of) random variables X such that $\text{E} X^2 < \infty$. If $X, Y \in L^2$ and $a, b \in \mathbb{R}$, then $aX + bY \in L^2$, which is to say that L^2 is a linear space. Introduce the inner product $\langle X, Y \rangle = \text{E} XY$, and the norm $\|X\| = \langle X, X \rangle^{1/2}$. You may show that these actually defines an inner product and a norm respectively. Also, $d(X, Y) = \|X - Y\|$ defines a metric on L^2 . From previous analysis courses we know that L^2 is a complete metric space, meaning that all Cauchy sequences in L^2 converges to an element of L^2 . A Hilbert space is an inner product space that is complete with respect to the metric induced by the norm associated with the inner product. Thus, $L^2(\Omega, \mathcal{F}, \text{Pr})$ is a Hilbert space.

Hilbert space

Theorem 5. (THE PROJECTION THEOREM). *Let $\mathcal{M}$ be a closed subspace of the Hilbert space $\mathcal{H}$, and $y \in \mathcal{H}$. Then there exists a unique element $\hat{y} \in \mathcal{M}$ such that*

Projection
theorem

- (i) $\|y - \hat{y}\| \leq \|y - z\|$ for all $z \in \mathcal{M}$;
- (ii) $\langle y - \hat{y}, z \rangle = 0$ for all $z \in \mathcal{M}$.

The $\hat{y}$ in the Projection theorem is called the orthogonal projection of y onto $\mathcal{M}$. We write this as $\Pi_{\mathcal{M}} y = \hat{y}$, so that $\Pi_{\mathcal{M}}$ is a mapping from $\mathcal{H}$ onto $\mathcal{M}$.

Exercise 3.1 Let $\mathcal{G}$ be a sub- σ -algebra of $\mathcal{F}$, and set $\mathcal{M} = L^2(\Omega, \mathcal{G}, \text{Pr})$. Then $\mathcal{M} \subset L^2(\Omega, \mathcal{F}, \text{Pr})$.

- (a) Show that $\mathcal{M}$ is closed.
- (b) Let $X \in L^2(\Omega, \mathcal{F}, \text{Pr})$, and show that

$$\text{E}(X | \mathcal{G}) = \Pi_{\mathcal{M}} X,$$

almost surely (so equal in L^2). In words, the conditional expectation of X given $\mathcal{G}$ equals the orthogonal projection of X onto the $L^2(\Omega, \mathcal{G}, \text{Pr})$.

3.2. Linear prediction in L^2 . Given that we have observed $X_1, \dots, X_n$ from a time series $\{X_t: t \in \mathbb{Z}\}$ the projection theorem and Ex. 3.1 say that the best prediction, in a mean square sense, we can make of X_t for $t \geq n+1$ is $E(X_{n+1} | X_1, \dots, X_n)$. This conditional expectation is in general very hard to compute. Thus, instead of the projection $E(X_t | X_1, \dots, X_n) = \Pi_{\mathcal{M}(X_1, \dots, X_n)} X_t$, where $\mathcal{M}(X_1, \dots, X_n) = L^2(\Omega, \sigma(X_1, \dots, X_n), \text{Pr})$, we concentrate on projections of X_t onto the closed linear span $\overline{\text{sp}}(1, X_1, \dots, X_n) \subset \mathcal{M}(X_1, \dots, X_n)$.

Exercise 3.2 The closed span $\overline{\text{sp}}(x_t: t \in T)$ of any subset $\{x_t: t \in T\}$ of a Hilbert space $\mathcal{H}$ is defined as the smallest closed subspace of $\mathcal{H}$ which contains x_t for each $t \in T$.

(a) Let $X_1, \dots, X_n$ be in $L^2(\Omega, \mathcal{F}, \text{Pr})$. Show that $\{\alpha_0 + \sum_{i=1}^n \alpha_i X_i: \alpha_i \in \mathbb{R}, i \geq 1\}$ is a closed subspace of L^2 .

(b) Show that $\overline{\text{sp}}(1, X_1, \dots, X_n) = \{\alpha_0 + \sum_{i=1}^n \alpha_i X_i: \alpha_i \in \mathbb{R}, i \geq 1\}$, meaning that for any $Y \in L^2$

$$\Pi_{\overline{\text{sp}}(1, X_1, \dots, X_n)} Y = \alpha_0 + \sum_{i=1}^n \alpha_i X_i,$$

for coefficients $\alpha_0, \alpha_1, \dots, \alpha_n$ that satisfy $\alpha_0 + \sum_{i=1}^n \alpha_i E X_i = E Y$ and

$$\alpha_0 E X_j + \sum_{i=1}^n \alpha_i E X_i X_j = E Y X_j, \quad \text{for } j = 1, \dots, n.$$

best linear
prediction

The projection $\Pi_{\overline{\text{sp}}(1, X_1, \dots, X_n)} Y$ is called the *best linear prediction* for Y given $X_1, \dots, X_n$. It is best in terms of the mse (as a consequence of the projection theorem).

(c) Show that if $Y \in L^2$ has expectation μ , then

$$\Pi_{\overline{\text{sp}}(1, X_1, \dots, X_n)} Y = \mu + \Pi_{\overline{\text{sp}}(X_1, \dots, X_n)} (Y - \mu).$$

In particular, if $\mu = 0$, then $\Pi_{\overline{\text{sp}}(1, X_1, \dots, X_n)} Y = \Pi_{\overline{\text{sp}}(X_1, \dots, X_n)} Y$.

Gram-Schmidt
orthogonalisation

Exercise 3.3 Let $x_1, \dots, x_n$ be linearly independent elements of the Hilbert space $\mathcal{H}$, i.e., $\alpha_1 x_1 + \dots + \alpha_n x_n = 0$ implies that $\alpha_1 = \dots = \alpha_n = 0$.

(a) Define $w_1 = x_1$ and

$$w_k = x_k - \Pi_{\overline{\text{sp}}(w_1, \dots, w_{k-1})} x_k, \quad \text{for } k = 2, \dots, n.$$

Show that $\{e_k = w_k / \|w_k\|: k = 1, \dots, n\}$ is an orthonormal set, i.e., $\langle e_k, e_j \rangle = 0$ for $j \neq k$ and $\langle e_k, e_k \rangle = 1$; and that $\overline{\text{sp}}(x_1, \dots, x_n) = \overline{\text{sp}}(e_1, \dots, e_n)$.

(b) Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Show that $(X, (Y - \rho X) / \sqrt{1 - \rho^2})$ is an orthonormal set.

(c) (xx something more xx)

Exercise 3.4 Let X and ε be independent standard normal random variables, and let $Y = f(X) + \varepsilon$ with $f(X) = I(X \geq 0)$. In terms of the mse, the best predictor of Y in terms of X is $E(Y | X) = f(X)$ which has mean squared error $E(Y - f(X))^2 = E\varepsilon^2 = 1$.

(a) Determine the coefficients in $\Pi_{\overline{\text{sp}}(1, X)} Y = \beta_0 + \beta_1 X$, i.e., find the best linear predictor.

(b) Show that the mean squared error of the best linear predictor is

$$E(Y - \Pi_{\overline{\text{sp}}(1,X)} Y)^2 = 1 + 1/4 - \phi(0)^2 \approx 1.090845,$$

which is, perhaps, somewhat surprising.

Exercise 3.5 As we shall see, the ACF of an MA(q)-process is zero for $|h| > q$. We want a measure of autocorrelation for AR-processes that behaves in this fashion, meaning that it tells us something about the order of dependence. This is the motivation for the partial autocorrelation function (PACF).

(a) Let $X_t = \theta(B)W_t$ be an MA(q)-process. Show that

$$\rho(h) = \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1 + \theta_1^2 + \dots + \theta_q^2}, \quad \text{for } 1 \leq h \leq q,$$

with $\rho(h) = 0$ for $h > q$, and $\rho(0) = 1$. Since $\gamma(h) = \gamma(-h)$ and $\rho(h) = \gamma(h)/\gamma(0)$ this fully described the ACF of the MA(q).

(b) Let $X = \{X_t : t \in \mathbb{Z}\}$ be a stationary process with. The PACF of X is defined by $\alpha(1) = \rho(1)$ and

$$\alpha(h) = \text{Cov}(X_h - \Pi_{\overline{\text{sp}}(1, X_1, \dots, X_{h-1})} X_h, X_0 - \Pi_{\overline{\text{sp}}(1, X_1, \dots, X_{h-1})} X_0), \quad \text{for } h \geq 2.$$

In words, the PACF is (for $h \geq 2$) is the correlation of two residuals obtained by regressing X_0 and X_h on the observations in between. Let $X_t = \theta X_{t-1} + W_t$ be a causal AR(1) process. Show that

$$\Pi_{\overline{\text{sp}}(X_1, \dots, X_{h-1})} X_0 = \phi X_1, \quad \text{and} \quad \Pi_{\overline{\text{sp}}(X_1, \dots, X_{h-1})} X_h = \phi X_{h-1},$$

and that, consequently, $\alpha(h) = 0$ for all $h \geq 2$.

Exercise 3.6 Consider the causal AR(p)-process $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + W_t$.

(a) Show that for $h - 1 \geq p$

$$\Pi_{\overline{\text{sp}}(X_1, \dots, X_{h-1})} X_h = \sum_{j=1}^p \phi_j X_{h-j}.$$

(b) Show that if $Y \in \overline{\text{sp}}(X_1, \dots, X_{h-1})$, then $Y \in \overline{\text{sp}}(W_t, t \leq h-1)$. Recall that causality gives an infinite series representation of X_t . Deduce that $\alpha(h) = 0$ for all $h \geq p+1$.

4. ESTIMATION AND LARGE SAMPLE THEORY

Exercise 4.1 Let X_n be a sequence for which we want to show weak convergence, but that it is difficult to do this directly. Instead we prove start with the approximations $Y_{m,n}$ and prove weak convergence for these. What are sufficient conditions for this to imply that weak convergence of X_n ?

(a) Suppose that

- (i) $Y_{m,n} \rightarrow_d Y_m$ as $n \rightarrow \infty$, for each m ;
- (ii) $Y_m \rightarrow_d Y$; and that
- (iii) $\lim_m \limsup_n \Pr(|X_n - Y_{m,n}| \geq \varepsilon) = 0$ for each $\varepsilon > 0$.

Then $X_n \rightarrow_d Y$. One way of proving this is to take a closed set F , define $F^\varepsilon = \{x : d(x, F) \leq \varepsilon\}$, where $d(x, F) = \inf\{|x - y| : y \in F\}$, and show that

$$\limsup_n \Pr(X_n \in F) \leq \Pr(Y \in F^\varepsilon).$$

Please fill in the details and finish up the proof.

(b) Show that if $E|X_n - Y_{m,n}|^2 \rightarrow 0$ as $m, n \rightarrow 0$, then (iii) holds. Attempt also to give a proof of the lemma in (a) by way of characteristic functions, i.e., showing that $E \exp(itX_n) \rightarrow E \exp(itY)$.

Exercise 4.2 Let $X = \{X_t : t \in \mathbb{Z}\}$ be a stationary sequence with mean μ and autocovariance function γ . Suppose we observe a sample $X_1, \dots, X_n$. From Ex. 1.9 we know that

$$\text{Var}(\bar{X}_n) = \frac{1}{n} \sum_{h=-n}^n (1 - |h|/n) \gamma(h) = \frac{1}{n} \sum_{h=-(n-1)}^{n-1} (1 - |h|/n) \gamma(h)$$

(a) Show that if $\gamma(n) \rightarrow 0$ as $n \rightarrow \infty$, then $\text{Var}(\bar{X}_n) \rightarrow 0$, and consequently, $\bar{X}_n$ converges in probability to μ .

(b) Suppose also that $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$. Show that this implies that

$$n \text{Var}(\bar{X}_n) \rightarrow \sum_{h=-\infty}^{\infty} \gamma(h).$$

Exercise 4.3 (A CLT FOR M -DEPENDENT DATA). A sequence $X = \{X_t : t \in \mathbb{Z}\}$ is M -dependent if for every t

$$\{X_s : s \leq t\} \perp \{X_s : s \geq t + M + 1\},$$

where the symbol $\perp$ means independence. Thus, if $M = 0$, then the sequence is simply independent, if $M = 1$, then $(\dots, X_{t-1}, X_t)$ and $(X_{t+2}, X_{t+3}, \dots)$ are independent, and so on.

(a) Let $X = \{X_t : t \in \mathbb{Z}\}$ be a strictly stationary M -dependent sequence with mean zero, from which we observe $X_1, \dots, X_n$, and let $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ be the empirical mean of these. For $m > 2M$ show that we can write

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^{\lfloor n/m \rfloor} U_j + \frac{1}{n} \sum_{j=1}^{\lfloor n/m \rfloor} V_j + r_{m,n},$$

in terms of

$$U_j = \sum_{i=(j-1)m+1}^{jm-M} X_i, \quad \text{and} \quad V_j = \sum_{i=jm-M+1}^{jm} X_i, \quad \text{for } j = 1, \dots, \lfloor n/m \rfloor \quad (8)$$

and a remainder

$$r_{m,n} = \frac{1}{n} \sum_{i=\lfloor n/m \rfloor m + 1}^n X_i, \quad (9)$$

taking this sum to be zero if $n = \lfloor n/m \rfloor m$. Here $\lfloor x \rfloor = \max\{z \in \mathbb{Z} : z \leq x\}$, and we have $\lfloor x \rfloor = x - \{x\}$ where $\{x\}$ is the fractional part, and $0 \leq \{x\} < 1$.

To get a sense of the splitting going on here, assume that $n = 11$, and $M = 2$, so $5 = m > 2M = 4$, say. Then we can split the $n = 10$ observations in $\lfloor n/m \rfloor = \lfloor 11/5 \rfloor = 2$ blocks

$$\underbrace{X_1, X_2, X_3, X_4, X_5}_{\text{block 1}}, \underbrace{X_6, X_7, X_8, X_9, X_{10}}_{\text{block 2}}, \underbrace{X_{11}}_{\text{remainder}}.$$

Now, take away the $M = 2$ last observations in each block

$$\underbrace{X_1, X_2, X_3, \cancel{X_4}, \cancel{X_5}}_{\text{block 1}}, \underbrace{X_6, X_7, X_8, \cancel{X_9}, \cancel{X_{10}}}_{\text{block 2}}, \underbrace{X_{11}}_{\text{remainder}}.$$

By the $M = 2$ dependence, (X_1, X_2, X_3) and (X_6, X_7, X_8) are independent, and

$$\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n}(U_1 + U_2) + \frac{1}{n}(V_1 + V_2) + \frac{X_{11}}{n},$$

for $U_1 = X_1 + X_2 + X_3$, $U_2 = X_6 + X_7 + X_8$, $V_1 = X_4 + X_5$, and $V_2 = X_9 + X_{10}$. In particular, U_1 and U_2 are independent, in fact i.i.d. due to the strict stationarity; and so are V_1 and V_2 .

(b) To get a CLT for $\sqrt{n}\bar{X}_n$ we assume that $\sum_{h=-M}^M \gamma(h) > 0$. Let $U_1, \dots, U_{\lfloor n/m \rfloor}$ be as defined in (8). Show that $n^{-1/2} \sum_{j=1}^{\lfloor n/m \rfloor} U_j \rightarrow_d Y_m$ with $Y_m \sim N(0, \sigma_m^2)$, as $n \rightarrow \infty$ for each m , where

$$\sigma_m^2 = \frac{1}{m} \text{Var}(U_1) = \left(1 - \frac{M}{m}\right) \sum_{h=-(m-M)}^{m-M} \left(1 - \frac{|h|}{m-M}\right) \gamma(h).$$

(c) Establish that

$$\sigma_m^2 \rightarrow \sum_{h=-M}^M \gamma(h), \quad \text{as } m \rightarrow \infty,$$

and show that this entails that $Y_m \rightarrow_d Y$ with $Y \sim N(0, \sum_{h=-M}^M \gamma(h))$.

(d) In view of Ex. 4.1, to conclude that $\sqrt{n}\bar{X}_n \rightarrow_d Y$, it now remains to show that

$$\lim_m \limsup_n \Pr \left(\left| \sqrt{n}\bar{X}_n - \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor n/m \rfloor} U_j \right| \geq \varepsilon \right) = 0$$

for each $\varepsilon > 0$. Establish that this is the case, and the CLT for strictly stationary M -dependent data is proven.

We would like to loosen the strict stationary assumption imposed in Ex. 4.3, and instead simply assume that $X = \{X_t : t \in \mathbb{Z}\}$ is stationary with mean zero. In Ex. 4.3 the strict stationarity of X entails that $U_1, \dots, U_{\lfloor n/m \rfloor}$ are i.i.d., and we can use the CLT for i.i.d. data that $n^{-1/2} \sum_{j=1}^{\lfloor n/m \rfloor} U_j \rightarrow_d N(0, \sigma_m^2)$. When X is stationary and M -dependent (but not strictly stationary), the $U_1, \dots, U_{\lfloor n/m \rfloor}$ are independent, but not identically distributed, and we would need a Lindeberg type CLT (see for example Theorem 5 in [Stoltenberg \(2025\)](#)).

Let $X = \{X_t : t \in \mathbb{Z}\}$ be a stationary process M -dependent process with mean zero and autocovariance function γ such that $\sum_{h=-M}^M \gamma(h) > 0$, and set $k_n = \lfloor n/m \rfloor$. The stationarity of X ensures that $\mathbb{E} U_1^2 = \dots = \mathbb{E} U_{\lfloor n/m \rfloor}^2$ for all m and n . Thus, to get the Lindeberg CLT to go through, we only need to verify that the appropriate Lindeberg condition is satisfied, that is

$$\frac{1}{\lfloor n/m \rfloor} \sum_{j=1}^{\lfloor n/m \rfloor} \mathbb{E} U_j^2 I\{|U_j| \geq \sqrt{\lfloor n/m \rfloor} \varepsilon\} \rightarrow 0, \quad \text{as } n \rightarrow \infty, \quad (10)$$

for each $\varepsilon > 0$ and each $m > 2M$. Thus, if (10) holds, then $n^{-1/2} \sum_{j=1}^{\lfloor n/m \rfloor} U_j \rightarrow_d N(0, \sigma_m^2)$, with σ_m^2 as in the exercise above. But (10) seems like an awkward condition (or?), and can be difficult to check. Consider the condition

$$\mathbb{E} |X_t|^3 = \kappa, \quad \text{for all } t \in \mathbb{Z}, \quad (11)$$

for some finite κ , that is, a third moment stationary condition. Under this condition, the triangle inequality (Minkowski's inequality) yields

$$(\mathbb{E} |\sum_{i=a}^b X_i|^3)^{1/3} \leq (b-a+1)\kappa^{1/3}.$$

for any $a < b$ in $\mathbb{Z}$. In particular, if the summands in $U_j = \sum_{i=(j-1)m+1}^{jm-M} X_i$ satisfy (11), then

$$\mathbb{E} U_j^3 \leq (m-M)^3 \kappa, \quad \text{for } j = 1, \dots, \lfloor n/m \rfloor. \quad (12)$$

Exercise 4.4 Let $X = \{X_t : t \in \mathbb{Z}\}$ be a stationary M -dependent mean zero sequence satisfying (11).

(a) Let $a_j < b_j$ for $j = 1, \dots, k$ be such that $b_j - a_j = h$ for all j , and form $A_{k,j} = k^{-1/2} \sum_{i=a_j}^{b_j} X_i$ for $j = 1, \dots, k$. Show that for any $\varepsilon > 0$

$$\sum_{j=1}^k \mathbb{E} A_{k,j}^2 I\{|A_{k,j}| \geq \varepsilon\} \rightarrow 0, \quad \text{as } k \rightarrow \infty.$$

(b) Tweak the proof from Ex. 4.3 where necessary to show that

$$\sqrt{n} \bar{X}_n \rightarrow_d N(0, \sum_{h=-M}^M \gamma(h)),$$

as $n \rightarrow \infty$, where, as usual $\gamma(h) = \mathbb{E} X_{t+h} X_t$ is the autocovariance function of X . We now have a CLT for stationary M -dependent data under the third moment condition of (11).

(c) (xx an application xx)

If $X_t = \phi X_{t-1} + W_t$ is an AR(1) process with $|\phi| < 1$, then it has the stationary solution $X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$. On the other hand, if $|\phi| > 1$, it has stationary solution $X_t = -\sum_{j=1}^{\infty} \phi^{-j} W_{t+j}$. In fact, it can be shown that any ARMA(p, q) process $\phi(B)X_t = \theta(B)W_t$ such that $\phi(z) \neq 0$ for all $|z| = 1$ has a unique stationary solution of the form $X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$ with coefficients determined by $\theta(z)\phi(z)^{-1} = \sum_{j=-\infty}^{\infty} \psi_j z^j$ for $r^{-1} < |z| < r$ for some $r > 1$ (see Brockwell and Davis (1991, Theorem 3.1.3, p. 88)). To get the stationary and causal AR(1)-solution for example, set $\psi_j = 0$ for $j < 0$ and $\psi_j = \phi^j$ for $j \geq 0$.

linear process

Definition 10. A linear process $X = \{X_t : t \in \mathbb{Z}\}$ is defined as

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}, \quad \text{with } \sum_{j=-\infty}^{\infty} |\psi_j| < \infty,$$

where $W \sim \text{WN}(0, \sigma^2)$ is white noise.

Exercise 4.5 Let X be a linear process.

(a) Show that X_t exists in L^2 .

(b) Show that the autocovariance function of X is

$$\gamma(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_{j+h} \psi_j, \quad h \in \mathbb{Z}.$$

(c) Show that $\sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 (\sum_{j=-\infty}^{\infty} \psi_j)^2$.

(d) Let $X_1, \dots, X_n$ be a sample from X . Show that $\bar{X}_n \rightarrow_p \mu$.

Exercise 4.6 Let $X = \{X_t : t \in \mathbb{Z}\}$ be the (mean zero) linear process $X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$ where $W = \{W_t : t \in \mathbb{Z}\}$ are i.i.d. mean zero with variance σ^2 .

(a) For $k \geq 1$ define the process $X^m = \{X_t^k : t \in \mathbb{Z}\}$ by $X_t^k = \sum_{j=-k}^k \psi_j W_{t-j}$. Let $M = M_k = 2k$, and deduce that X^k is a strictly stationary M -dependent process.

(b) Show that the autocovariance function of X^k is

$$\gamma_k(h) = \sigma^2 \sum_{j=-k}^{k-h} \psi_{j+h} \psi_j, \quad \text{for } 0 \leq h \leq 2k,$$

and we note that by the stationarity $\gamma_k(-h) = \gamma_k(h)$. Deduce also that

$$\sum_{h=-2k}^{2k} \gamma_k(h) = \sigma^2 \left(\sum_{j=-k}^k \psi_j \right)^2.$$

Provided $\sum_{j=-k}^k \psi_j \neq 0$, it now follows from the CLT for M -dependent data in Ex. 4.3 that $n^{-1/2} \sum_{i=1}^n X_i^k \rightarrow_d N(0, \sigma^2 (\sum_{j=-k}^k \psi_j)^2)$.

(c) Use the two-steps weak convergence lemma in Ex. 4.1 to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow_d N(0, \sigma^2 (\sum_{j=-\infty}^{\infty} \psi_j)^2)$$

provided $\sum_{j=-\infty}^{\infty} \psi_j \neq 0$.

Exercise 4.7 In Ex. 1.16 we found the limiting distribution of the least squares estimator of the regression coefficient in a causal AR(1)-process by means of martingale central limit theorem. In this exercise we derive the same result by other means. Consider the AR(1)-process $X_t = \phi X_{t-1} + W_t$, $t \in \mathbb{Z}$ for i.i.d. white noise $W \sim WN(0, \sigma^2)$ with $|\phi| < 1$. With $\hat{\phi}_n$ the least squares estimator we have that $\sqrt{n}(\hat{\phi}_n - \phi) = n^{-1/2} \sum_{t=1}^n X_{t-1} W_t / \{n^{-1} \sum_{t=1}^n X_{t-1}^2\}$, and we know from Ex. 1.16 that $n^{-1} \sum_{t=1}^n X_{t-1}^2 \rightarrow_p \sigma^2 / (1 - \phi^2)$. Thus, by the Cramér–Slutsky rules it suffices to show that $n^{-1/2} \sum_{t=1}^n X_{t-1} W_t$ converges in distribution to a normal limit.

(a) For $m > 1$, define $A_t^m = W_t \sum_{j=0}^{m-2} \phi^j W_{t-1-j}$ for $t \in \mathbb{Z}$. Show that $A^m = \{A_t^m : t \in \mathbb{Z}\}$ is a strictly stationary m -dependent sequence with mean zero, and autocovariance function

$$\gamma_{A^m}(h) = E A_{t+h}^m A_t^m = \sigma^4 \frac{1 - \phi^{2(m-1)}}{1 - \phi^2}, \quad \text{for } h = 0,$$

with $\gamma_{A^m}(h) = 0$ for $|h| > 0$.

(b) Deduce that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n A_t^m \rightarrow_d N(0, \sigma^4 \frac{1 - \phi^{2(m-1)}}{1 - \phi^2}), \quad \text{as } n \rightarrow \infty,$$

and then apply the two-steps weak convergence lemma to confirm that

$$n^{-1} \sum_{t=1}^n X_{t-1} W_t \rightarrow_d N(0, \sigma^4 / (1 - \phi^2)),$$

and we get $\sqrt{n}(\hat{\phi}_n - \phi) \rightarrow_d N(0, 1 - \phi^2)$, by the Cramér–Slutsky rules, as we knew already.

(c) Attempt to generalise the above to causal AR(p)-processes $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + W_t$, where the least square estimator $\hat{\phi}_n$ of $(\phi_1, \dots, \phi_p)^t$ is

$$\hat{\phi}_n = \left(\sum_{t=p+1}^n Z_{t-1} Z_{t-1}^t \right)^{-1} \sum_{t=p+1}^n Z_{t-1} Z_t,$$

in terms of $Z_t = (X_t, \dots, X_{t-p+1})^t$.

Exercise 4.8 Let $X = \{X_t : t \in \mathbb{Z}\}$ be the linear process $X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$ for i.i.d. mean zero noise terms W with variance σ^2 . Consider the estimator of $\gamma(h)$ given by

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

Recall from Ex. 4.5 that $\bar{X}_n \rightarrow_p \mu$.

(a) Show that we can write

$$\hat{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \mu)(X_t - \mu) + o_p(n^{-1/2}),$$

as $n \rightarrow \infty$.

(b) (xx rewrite and check xx) Define $Y_t = X_t - \mu$ and $Y_t^m = \sum_{j=-m}^m \psi_j W_{t-j}$ for its truncated version, and let

$$\gamma_m(h) = \mathbb{E} Y_{t+h}^m Y_t^m = \sigma^2 \sum_{j=-m}^{m-h} \psi_{j+h} \psi_j, \quad 0 \leq h \leq 2m.$$

Deduce from the CLT for strictly stationary M -dependent data that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n-h} (Y_{t+h}^m Y_t^m - \gamma_m(h)) \quad (13)$$

has a mean zero normal limit as $n \rightarrow \infty$, for each m . Use the two-step lemma and deduce a clt for $\hat{\gamma}_n(h)$.

REFERENCES

- Billingsley, P. (1995). *Probability and Measure. Third Edition*. John Wiley & Sons, New York.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods. Second Edition*. Springer, New York.
- Kallenberg, O. (2002). *Foundations of Modern Probability. Second Edition*. Springer, Berlin.
- Shumway, R. H. and Stoffer, D. S. (2025). *Time Series Analysis and Its Applications: With R Examples. Fifth Edition*. Springer, New York.
- Stoltenberg, E. A. (2025). The Martingale CLT. [Lecture notes STK4090/9090 Spring 2025](#).
- Williams, D. (1991). *Probability with Martingales*. Cambridge University Press, Cambridge.