LECTURE NOTES AND EXERCISES
STK4060 – TIME SERIES, AUTUMN 2025

November 20, 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; and to Sebastian Grau Nielsen and Thomas Engl for pointing out mistakes and proposing improvements.

Contents

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 $(\Omega, \mathcal{F}, \Pr)$. 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 $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{N} \cup \{0\} = \{0, 1, 2, \ldots\}$, or $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

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. sample path
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 $t_1, \ldots, t_n \in T$.

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 white noise
if $\mathrm{E}\, W_t = 0$ for all t and

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

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

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 $W \sim \mathrm{WN}(0, \sigma^2)$ 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 $\mathrm{E}\, X_t = 0$ for all t and that

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

(b) Define $Y_t = \mu + X_t$, and suppose that we observe $Y_1, \ldots, Y_n$ 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. $\mathrm{N}(0, \sigma^2)$ 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 $\mathrm{E}\, X_t = 0$ and that for $h \geq 0$

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

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

random walk

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

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 $\mathrm{E}\, X_t = x_0 + \mu t$.

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

1.1. **The distribution of a process.** Consider a random variable X defined on a probability space $(\Omega, \mathcal{F}, \Pr)$, i.e., a measureable function $X : \Omega \to \mathbb{R}$, 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

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

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

Dynkin's lemma

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*

$$F(b) - F(a) = \lim_{\sigma \to 0} \frac{1}{2\pi} \int \frac{\exp(-itb) - \exp(-ita)}{-it} \varphi(t) \exp(-\tfrac{1}{2} t^2 \sigma^2)\, \mathrm{d}t, \qquad (1)$$

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 $X \sim Y$.

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

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

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

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

for all finite collections $t_1, \ldots, t_k \in T$, $k \geq 1$. 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

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

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

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

$$\lim_{x_i \to \infty} F_t(x) = F_{t_{-i}}(x_{-i}). \tag{2}$$

Proof. See for example Billingsley (1995). □

Write $\{\phi_t(\xi) \colon t \in \tau\}$ for a collection of finite dimensional c.f.s, where $\phi_t(\xi) = \int \exp(i\xi^{\mathrm{t}} x)\,\mathrm{d}F_t(x)$. A condition equivalent to (2) is then that

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

Exercise 1.6 Let $W_1, W_2, \ldots$ be i.i.d. $\mathrm{N}(0,1)$, and define the process $X = \{X_t \colon t \in \mathbb{N} \cup \{0\}\}$ 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) $X_t - X_s \sim \mathrm{N}(0, t-s)$ 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 $X_t - X_s \sim \mathrm{N}(0, (t-s)^2)$, 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 $\{X \colon X_t \in T\}$ autocovariance

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

provided $\mathrm{E}\,X_t^2 < \infty$.

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

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

provided $\mathrm{E}\,X_t^2 < \infty$.

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

stationarity

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

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

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

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

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

This motivates the following definition.

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

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

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 $X = \{X_t \colon t \in \mathbb{Z}\}$ be a times series.

(a) Suppose that X is strictly stationary such that $\mathrm{E}\, X_t^2 < \infty$. 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 $W_0, W_1, \ldots$ with $\mathrm{E}\, W_t = 0$ and $\mathrm{E}\, W_t^2 = \sigma^2$. Show that this process is *not* stationary.

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

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

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 $Y_1, \ldots, Y_n$ been i.i.d. with variance τ^2, we would have had $\mathrm{E}\, \bar{Y}_n = \mu$ and $\mathrm{Var}(\bar{Y}_n) = \gamma(0)/n = \tau^2/n$.

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

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

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

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

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 $X_n \to X$ 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 $X_n \to X$ almost surely, then $X_n \to X$ in probability.

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

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

Exercise 1.11 Let $A_1, A_2, \ldots$ 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 the Borel–Cantelli lemma Borel–Cantelli

(b) Assume in addition that the $A_1, A_2, \ldots$ 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 $X_n \to_p X$. Use the Borel–Cantelli lemma to find a subsequence $(X_{n_k})_{k\geq 1}$ such that $X_{n_k} \to X$ 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 $0 \leq Y_1 \leq Y_2 \leq Y_3 \leq \cdots$ be random variables such that $Y_n \to_p Y$. Show that $Y_n \to Y$ almost surely.

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

(c) Let $X_1, X_2, \ldots$ 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

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

are convergent. Verify that the condition from (b), namely $\sum_{j=1}^{\infty} \mathrm{E}\,|X_j|$ 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 $n = 1, 2, \ldots$. Show that $(X_n)_{n\geq 1}$ satisfies **??**&**??**&**??** of the three series theorem, but $\sum_{j=1}^{\infty} \mathrm{E}\,|X_j| = \sum_{j=1}^{\infty}(1/j)$ 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

$$X_t = \phi X_{t-1} + W_t, \quad t \in \mathbb{Z}. \tag{3}$$

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 $\mathrm{E}\,(X_t - \sum_{j=0}^{k-1} \phi^j W_{t-j})^2 \to 0$ as $k \to \infty$. We are thus justified in writing

$$X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}, \tag{4}$$

with this equality being true in the mean square sense.

(c) Show that the equality (4) is true almost surely as well. The solution in (4) is said to be *causal* (or future-independent), as X_t only depends on W_s for $s \leq t$.

causal solution

(d) Show that the process in (4) satisfies equation (3), and that X_t defined by (4) 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

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

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

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

(b) Show that

$$\sqrt{n}(\widehat{\beta}_n - \beta) \to_d \mathrm{N}\big(0, \frac{\sigma^2}{\sigma_X^2 + \mu_X^2}\big).$$

(c) Things are more complicated when the covariates are fixed, i.e., when the model is $Y_i = \beta x_i + \varepsilon_i$, with $x_1, \ldots, x_n$ 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 $\widehat{\beta}_n = \sum_{i=1}^{n} x_i Y_i / s_n^2$ be the least squares estimator. Show that if $\max_{j\leq n} x_j^2 / \sum_{i=1}^{n} x_i^2 \to 0$, then $s_n(\widehat{\beta}_n - \beta) \to_d \mathrm{N}(0, \sigma^2)$.

Exercise 1.16 Let $X = \{X_t \colon t \in \mathbb{Z}\}$ 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. $\mathrm{N}(0, \sigma^2)$. By Ex. 1.14 it has the representation $X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$. Suppose that we observe $X_1, X_2, \ldots, X_n$. The least squares estimator for ϕ is $\widehat{\phi}_n = \sum_{t=1}^{n} X_t X_{t-1}/\{\sum_{t=1}^{n} X_{t-1}^2\}$.

(a) Show that we may write

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

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

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

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

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

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

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

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

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

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

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 $\varepsilon > 0$,

$$\frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\,\big\{(X_{t-1}W_t)^2 I_{|X_{t-1}W_t| \geq \sqrt{n}\varepsilon} \,|\, \mathcal{F}_{t-1}\big\} \to_p 0. \tag{5}$$

One may try to handle this directly. It is more convenient, I think, to go via the Lyapunov conditon. 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 conditon, i.e., that for some $\delta > 0$, $\sum_{i=1}^{n} \mathrm{E}\,\big(|\xi_{n,i}|^{2+\delta} \,|\, \mathcal{G}_{n,i-1}\big) \to_p 0$, and show that this implies the Lindeberg condition $\sum_{i=1}^{n} \mathrm{E}\,\big(\xi_{n,i}^2 I_{|\xi_{n,i}|\geq\varepsilon} \,|\, \mathcal{G}_{n,i-1}\big) \to_p 0$, for any $\varepsilon > 0$. Deduce that (5) is indeed satisfied.

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

(h) Attempt to extend this convergence in distribution result to the more general case where $W_t \sim \mathrm{WN}(0, \sigma^2)$, 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 $t \in \mathbb{Z}$ ARMA(p,q)

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}, \tag{6}$$

where $W \sim \mathrm{WN}(0,\sigma^2)$ 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 (6) may then be written

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

AR(p)

MA(q)

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 $X_t = \theta(B)W_t$.

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 $\psi_0, \psi_1, \ldots$ 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 $j = 0, 1, 2, \ldots$.

invertible ARMA

Definition 9. An ARMA(p,q)-process is invertible if there exists a sequence of constants $\pi_0, \pi_1, \ldots$ 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 $j = 0, 1, 2, \ldots$.

(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

$$\phi(B)^{-1} = (1 - \phi B)^{-1} = \sum_{j=0}^{\infty} \phi^j B^j. \tag{7}$$

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 (7).

(a) Show that if $Y = \{Y_t \colon t \in \mathbb{Z}\}$ is stationary, then the process Z defined by $Z_t = \phi(B)^{-1}Y_t = \left(\sum_{j=0}^{\infty} \phi^j B^j\right)Y_t$, 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 $a_0, a_1, \ldots, a_p$ such that $a_p \neq 0$, the equation

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

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 efforst 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 $|\text{roots}| > 1$

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. □

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

$$X_t = -\sum_{j=1}^{\infty} \phi^{-j} W_{t+j}, \tag{8}$$

is the stationary solution to (3) when $|\phi| > 1$. The solution in (8) is *not* causal, as X_t depends on the future innovation terms $W_{t+1}, W_{t+2}, \ldots$.

(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 $\widetilde{W}_t \sim \mathrm{WN}(0, \widetilde{\sigma}^2)$. 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 $|\text{roots}| > 1$

Proof. Mimic the proof of Theorem 3. □

3. Forecasting

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

3.1. **L^2 and the Projection theorem.** The time series we are dealing with are defined on a probability space $(\Omega, \mathcal{F}, \Pr)$. We let $L^2 = L^2(\Omega, \mathcal{F}, \Pr)$ denote all (equivalence classes for a.s. equality of) random variables X such that $\mathrm{E}\, X^2 < \infty$. 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 $\langle X, Y \rangle = \mathrm{E}\, XY$, 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, $L^2(\Omega, \mathcal{F}, \Pr)$ 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 an unique element $\widehat{y} \in \mathcal{M}$ such that*

Projection theorem

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

The $\widehat{y}$ in the Projection theorem is called the orthogonal projection of y onto $\mathcal{M}$. We write this as $\Pi_{\mathcal{M}} y = \widehat{y}$, 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 $\mathcal{M} = L^2(\Omega, \mathcal{G}, \Pr)$. Then $\mathcal{M} \subset L^2(\Omega, \mathcal{F}, \Pr)$.

(a) Show that $\mathcal{M}$ is closed.

(b) Let $X \in L^2(\Omega, \mathcal{F}, \Pr)$, and show that

$$\mathrm{E}\,(X \mid \mathcal{G}) = \Pi_{\mathcal{M}} X,$$

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 $L^2(\Omega, \mathcal{G}, \Pr)$.

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

Exercise 3.2 The closed span $\bar{\mathrm{sp}}(x_t \colon t \in T)$ of any subset $\{x_t \colon t \in T\}$ 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 $L^2(\Omega, \mathcal{F}, \mathrm{Pr})$. Show that $\{\alpha_0 + \sum_{i=1}^n \alpha_i X_i \colon \alpha_i \in \mathbb{R}, i \geq 1\}$ is a closed subspace of L^2.

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

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

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

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

best linear prediction

The projection $\Pi_{\overline{\mathrm{sp}}(1,X_1,\dots,X_n)} Y$ 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

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

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

Gram–Schmidt orthogonalisation

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

(a) Define $w_1 = x_1$ and

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

Show that $\{e_k = w_k / \|w_k\| \colon k = 1, \dots, n\}$ 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 $\overline{\mathrm{sp}}(x_1, \dots, x_n) = \overline{\mathrm{sp}}(e_1, \dots, e_n)$.

(b) Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}_2\big(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\big)$$

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 $\mathrm{E}\,(Y \,|\, X) = f(X)$ which has mean squared error $\mathrm{E}\,(Y - f(X))^2 = \mathrm{E}\,\varepsilon^2 = 1$.

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

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

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

which is, perhaps, somwhat 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

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

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

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

In words, the PACF is (for $h \ge 2$) 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

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

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

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

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

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

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

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

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

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

Please fill in the details and finish up the proof.

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

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

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

(a) Show that if $\gamma(n) \to 0$ as $n \to \infty$, then $\mathrm{Var}(\bar{X}_n) \to 0$, 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

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

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

M-dependence

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

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

(a) Let $X = \{X_t \colon t \in \mathbb{Z}\}$ be a strictly stationary M-dependent sequence with mean zero, from which we observe $X_1, \ldots, X_n$, 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

$$U_j = \sum_{i=(j-1)m+1}^{jm-M} X_i, \quad \text{and} \quad V_j = \sum_{i=jm-M+1}^{jm} X_i, \quad \text{for } j = 1, \ldots, \lfloor n/m \lfloor \tag{9}$$

and a remainder

$$r_{m,n} = \frac{1}{n}\sum_{i=\lfloor n/m \rfloor m+1}^{n} X_i, \tag{10}$$

taking this sum to be zero if $n = \lfloor n/m \rfloor m$. Here $\lfloor x \rfloor = \max\{z \in \mathbb{Z} \colon z \leq x\}$, 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

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

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 $U_1, \ldots, U_{\lfloor n/m \rfloor}$ be as defined in (9). Show that $n^{-1/2}\sum_{j=1}^{\lfloor n/m \rfloor} U_j \to_d Y_m$ with $Y_m \sim \mathrm{N}(0, \sigma_m^2)$, as $n \to \infty$ for each m, where

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

(c) Establish that

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

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

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

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

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 $U_1, \ldots, U_{\lfloor n/m \rfloor}$ are i.i.d., and we can use the CLT for i.i.d. data to conclude that $n^{-1/2}\sum_{j=1}^{\lfloor n/m \rfloor} U_j \to_d \mathrm{N}(0, \sigma_m^2)$. When X is stationary and M-dependent (but not strictly stationary), the $U_1, \ldots, U_{\lfloor n/m \rfloor}$ 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 $\sum_{h=-M}^{M}\gamma(h) > 0$. The stationarity of X ensures that $\mathrm{E}\,U_1^2 = \cdots = \mathrm{E}\,U_{\lfloor n/m \rfloor}^2 = 0$ 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

$$\frac{1}{\lfloor n/m \rfloor}\sum_{j=1}^{\lfloor n/m \rfloor}\mathrm{E}\,U_j^2 I\{|U_j| \geq \sqrt{\lfloor n/m \rfloor}\varepsilon\} \to 0, \quad \text{as } n \to \infty, \tag{11}$$

for each $\varepsilon > 0$ and each $m > 2M$. Thus, if (11) holds, then $n^{-1/2}\sum_{j=1}^{\lfloor n/m \rfloor} U_j \to_d \mathrm{N}(0, \sigma_m^2)$, with σ_m^2 as in the exercise above. But (11) seems like an awkward condition (or?), and can be difficult to check. Consider the condition

$$\mathrm{E}\,|X_t|^3 = \kappa, \quad \text{for all } t \in \mathbb{Z}, \tag{12}$$

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

$$\big(\mathrm{E}\,|\sum_{i=a}^{b} X_i|^3\big)^{1/3} \le (b-a+1)\kappa^{1/3}.$$

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 (12), then

$$\mathrm{E}\,U_j^3 \le (m-M)^3\kappa, \quad \text{for } j = 1,\dots,\lfloor n/m\rfloor. \tag{13}$$

Exercise 4.4 Let $X = \{X_t\colon t\in\mathbb{Z}\}$ be a stationary M-dependent mean zero sequence satisfying (12).

(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 $\varepsilon > 0$

$$\sum_{j=1}^{k} \mathrm{E}\,A_{k,j}^2 I\{|A_{k,j}| \ge \varepsilon\} \to 0, \quad \text{as } k\to\infty.$$

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

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

as $n\to\infty$, where, as usual $\gamma(h) = \mathrm{E}\,X_{t+h}X_t$ is the autocovariance function of X. We now have a CLT for stationary M-dependent data under the third moment condition of (12).

(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 an 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 $j \ge 0$.

linear process

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

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

where $W\sim \mathrm{WN}(0,\sigma^2)$ 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 $\sum_{h=-\infty}^{\infty}\gamma(h) = \sigma^2\big(\sum_{j=-\infty}^{\infty}\psi_j\big)^2$.

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

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

(a) For $k \geq 1$ define the process $X^k = \{X_t^k \colon t \in \mathbb{Z}\}$ 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

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

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

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

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

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 $W \sim \mathrm{WN}(0, \sigma^2)$ with $|\phi| < 1$. With $\widehat{\phi}_n$ the least squares estimator we have that $\sqrt{n}(\widehat{\phi}_n - \phi) = n^{-1/2} \sum_{t=1}^n X_{t-1} W_t / \{n^{-1} \sum_{t=1}^n X_{t-1}^2\}$, and we know from Ex. 1.16 that $n^{-1} \sum_{t=1}^n X_{t-1}^2 \to_p \sigma^2/(1-\phi^2)$. 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 $A^m = \{A_t^m \colon t \in \mathbb{Z}\}$ is a strictly stationary m-dependent sequence with mean zero, and autocovariance function

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

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

(b) Deduce that

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

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

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

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

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

$$\widehat{\phi}_n = \Big(\sum_{t=p+1}^{n} Z_{t-1} Z_{t-1}^{\mathrm{t}} \Big)^{-1} \sum_{t=p+1}^{n} Z_{t-1} X_t,$$

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

Exercise 4.8 Let $X = \{X_t \colon t \in \mathbb{Z}\}$ 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

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

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

(a) Show that we can write

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

as $n \to \infty$.

(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

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

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

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n-h} (Y_{t+h}^m Y_t^m - \gamma_m(h)) \tag{14}$$

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

5. Spectral analysis

time domain

So far in these notes we have analysed stationary time series data as linear combinations of white noise, i.e., $X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$. This is called *time domain* analysis as the properties of the time series (the autocovariance function) is analysed at different time intervals, lags.

frequency domain

We can also study the time series in what is known as the *frequency domain*, that is, instead of representing X_t as a linear combination of the past and future white white noise, we can represent X_t as a sum of sinusoidal components with uncorrelated random coefficients. This approach is called spectral analysis, and plays a similar role for stochastic processes as Fourier analysis does for deterministic functions.

As motivation, we start with a simple case. Consider the process $X = \{X_t \colon t \in \mathbb{Z}\}$ defined by

$$X_t = A \cos(2\pi\omega t + \phi), \quad t \in \mathbb{Z}. \tag{15}$$

Here ω is a *frequency index* defined as the cycles per unit of time; A is the *amplitude* or height of the function; while ϕ is the *phase*, which determines the start point of the cosine function. Randomness can be introduced by letting A and ϕ be random variables. To get a feel for the parameters ω, A, and ϕ, you can play around with a version of this script. In particular, observe what happens (to the observed values) when the frequency ω is outside of $[-1/2, 1/2]$ (see also the discussion in Shumway and Stoffer (2025, p. 178)).

```
t <- seq(-10,10,0.01)
A <- 1 # amplitued
phi <- 0.33 # phase
omega <- 1/10 # frequency
x <- A*cos(2*pi*omega*t + phi)
postscript("spectral_plot1.eps")
par(mfrow=c(1,1))
plot(t,x,type="l",lwd=2,frame.plot=FALSE)
tobs <- seq(min(t),max(t),1)
xobs <- x[t%in%tobs]
points(tobs,xobs,col="magenta",pch=1)
abline(A,0,col="cornflowerblue",lwd=2);abline(-A,0,col="cornflowerblue",lwd=2)
dev.off()
```

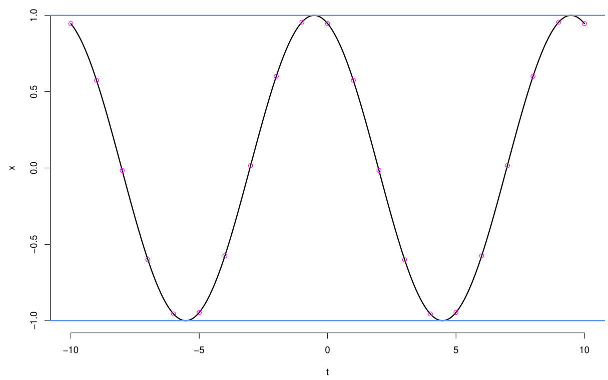

FIGURE 1. The plot generated by the script above. Light blue lines indicate the amplitude A. The magenta dots indicates the X_t for $t \in \mathbb{Z}$, i.e., the observed values.

Exercise 5.1 Let X be the process defined above in (15).

(a) Use the trigonometric identity

$$\cos(x \pm y) = \cos(x)\cos(y) \mp \sin(x)\sin(y), \tag{16}$$

to show that X can be written

$$X_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t).$$

with $U_1 = A\cos\phi$ and $U_2 = -A\sin\phi$.

(b) Show that $A = \sqrt{U_1^2 + U_2^2}$ and $\phi = \tan^{-1}(-U_2/U_1)$.

(c) Do a transformation-of-variables analysis, going from (U_1, U_2) to (A, ϕ). Assume that U_1 and U_2 are independent $\mathrm{N}(0, \sigma^2)$. Show that $A^2 \sim \sigma^2\chi_2^2$ and that $\phi \sim \mathrm{unif}(-\frac{1}{2}\pi, \frac{1}{2}\pi)$.

(d) Show also the converse, that is (A, ϕ) are given these distributions, then U_1 and U_2 are i.i.d. $\mathrm{N}(0, \sigma^2)$. It is instructive to verify this via simulations.

(e) For a given frequency parameter ω, and with U_1, U_2 independent mean zero with variance σ^2, show that $X_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t)$ has autocovariance function $\gamma(h) = \sigma^2 \cos(2\pi\omega h)$, and is hence stationary.

Exercise 5.2 Consider the time series $X = \{X_t \colon t \in \mathbb{Z}\}$ defined by

$$X_t = \sum_{k=1}^{q} \{U_{k,1} \cos(2\pi\omega_k t) + U_{k,2} \sin(2\pi\omega_k t)\}, \quad t \in \mathbb{Z}$$

for independent mean zero random variables $U_{k,1}, U_{k,2}$, $k = 1, \dots, q$ with variances $\operatorname{Var} U_{k,1} = \operatorname{Var} U_{k,2} = \sigma_k^2$, and frequency parameters $\omega_1, \dots, \omega_q$.

(a) For some small q, say $q = 3$ or $q = 4$, simulate a few realisations a process of the this type.

(b) Show that the autocovariance function of X is

$$\gamma(h) = \sum_{k=1}^{q} \sigma_k^2 \cos(2\pi\omega_k h), \quad h \in \mathbb{Z}.$$

In particular, the variance of X_t is $\sum_{k=1}^{q} \sigma_k^2$, the sum of the individual variance associated with the pair $(U_{k,1}, U_{k,2})$.

(c) Recall that from Euler's formula $\exp(ix) = \cos(x) + i\sin(x)$ we have

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2}, \quad \text{and} \quad \sin(x) = \frac{e^{ix} - e^{-ix}}{2i}.$$

Show that X_t can be expressed as

$$X_t = \sum_{k=1}^{q} \{\tfrac{1}{2}(U_{k,1} - iU_{k,2}) \exp(i2\pi\omega_k t) + \tfrac{1}{2}(U_{k,1} + iU_{k,2}) \exp(-i2\pi\omega_k t)\}.$$

(d) When $t \in \mathbb{Z}$, as we basically assume all the time, we are really only interested in frequencies $\omega \in [-1/2, 1/2]$. Assume that

$$-\tfrac{1}{2} < \omega_1 < \omega_2 < \cdots < \omega_q = \tfrac{1}{2},$$

are so that $\omega_j = -\omega_{q-j}$ for $j = 1, \dots, q - j$. Show that we then may write the process X as

spectral representation

$$X_t = \sum_{j=1}^{q} U_j \exp(i2\pi\omega_j t), \tag{17}$$

where

$$U_j = \tfrac{1}{2}\{(U_{j,1} + U_{q-j,1}) + i(U_{q-j,2} - U_{j,2})\}, \quad \text{for } j = 1, \dots, q - 1,$$

and $U_q = U_{q,1}$. The expression for X_t given in (17) is the *spectral representation* of the process X.

(e) Show that the $U_1, \dots, U_q$ are mean zero; uncorrelated, i.e. orthogonal $\operatorname{E} U_j \bar{U}_k = 0$ for $j \neq k$; and identify the variances, say $\tau_j^2 = \operatorname{Var} U_j = \operatorname{E} U_j \bar{U}_j$ for $j = 1, \dots, q$. Here $\bar{U}_j$ denotes the complex conjugate of U_j.

(f) Show that the autocovariance function of X is

$$\operatorname{E} X_{t+h} \bar{X}_t = \sum_{j=1}^{q} \tau_j^2 \exp(i2\pi\omega_j h) = \sum_{j=1}^{q} \sigma_j^2 \cos(2\pi\omega_j h) = \gamma(h),$$

as it should.

spectral rep. of autocovariance

(g) Verify also that we can express γ as the Riemann–Stieltjes integral

$$\gamma(h) = \int_{-1/2}^{1/2} \exp(i2\pi\omega h)\,\mathrm{d}F(\omega), \tag{18}$$

with $F(\omega) = \sum_{j:\omega_j \le \omega} \tau_j^2$. This function F is the *spectral distribution function* of the process X.

spectral distribution

The equations in (17) and (18) are fundamental to spectral analysis; they are the spectral representation of the process X itself, and the corresponding spectral representation of its autocovariance function, respectively. Notice that all the mass of F to the interval $(-1/2, 1/2]$, and that the mass assigned to $\omega_j \in (-1/2, 1/2]$ is the variance of the coefficient U_j in (17), in fact $F(\omega) = \sum_{j:\omega_j \le \omega} \mathrm{E}\,|U_j|^2$.

The example in Ex. 5.1 is not special. In fact, *every* mean zero stationary process X has a representation that generalises (17), namely

$$X_t = \int_{-1/2}^{1/2} \exp(i2\pi\omega t)\,\mathrm{d}Z(\omega),$$

where $Z = \{Z(\omega) \colon \omega \in [-1/2, 1/2]\}$ is a certain orthogonal-increment process. A proper definition of such process and the integral above is given shortly. The autocovariance function of such process can then be expressed as

$$\gamma(h) = \int_{-1/2}^{1/2} \exp(i2\pi\omega h)\,\mathrm{d}F(\omega),$$

for a distribution function F such that $F(\omega) = 0$ for $\omega \le -1/2$ and $F(\omega) = \gamma(0) = \mathrm{E}\,|X_t|^2$ for $\omega \ge 1/2$.

The spectral representation of $\gamma(h)$ is easier to establish than that of X_t, since it does not require the notion of stochastic integration. We therefore start with the spectral representation of the autocovariance.

5.1. **The spectral distribution.** A function $K \colon \mathbb{Z} \to \mathbb{C}$ is non-negative definite if for all $n \ge 1$

non-negative definite

$$\sum_{s=1}^{n} \sum_{t=1}^{n} a_s K(s-t) \bar{a}_t \ge 0,$$

for all $a_1, \dots, a_n \in \mathbb{C}$. If $X = \{X_t \colon t \in \mathbb{Z}\}$ is a mean zero stationary time series (possibly complex valued), its autocovariance function is

$$\gamma(h) = \mathrm{E}\, X_{t+h} \bar{X}_t.$$

Since (as you may check)

$$\sum_{s=1}^{n} \sum_{t=1}^{n} a_s \gamma(s-t) \bar{a}_t = \mathrm{E}\,\Big|\sum_{t=1}^{n} a_t X_t\Big|^2 \ge 0,$$

the autocovariance function of a mean zero stationary time series is non-negative definite.

Exercise 5.3 Let γ be the autocovariance function of a (possibly complex valued) stationary process. Show that $\gamma(0) \ge 0$, $|\gamma(h)| \le \gamma(0)$, and $\gamma(h) = \overline{\gamma(-h)}$. Notice that if γ is real, then this last property says that γ is even, i.e., $\gamma(h) = \gamma(-h)$

Theorem 6. *A function $K \colon \mathbb{Z} \to \mathbb{C}$ is the autocovariance function of a (possibly complex valued) stationary time series if and only if $K(h) = \overline{K(-h)}$ and K is non-negative definite.*

Proof. Sketch. One direction is proved above. For the converse, we assume that K is non-negative definite (and concentrate on the real case). We need to show that that K is the covariance function of some process. Let $X = \{X_t \colon t \colon \mathbb{Z}\}$ be a time series such that $\mathrm{E}\exp(i\sum_{i=1}^k t_i X_{t_i}) = \exp(-\frac{1}{2}\sum_{i=1}^k \sum_{j=1}^k t_i K(t_i - t_j) t_j)$ for any $t_1, \ldots, t_k \in \mathbb{Z}$. Apply the Kolmogorov consistency theorem (see above) to ensure that such a process exists. □

Theorem 7. (Herglotz). *A function $\gamma \colon \mathbb{Z} \to \mathbb{C}$ is non-negative definite if and only if it can be expressed as*

$$\gamma(h) = \int_{-1/2}^{1/2} \exp(i2\pi\omega h)\,\mathrm{d}F(\omega),$$

where F is a right-continuous, non-decreasing, bounded function on $[-1/2, 1/2]$ such that $F(\omega) = F(1/2) = 0$ for all $\omega \le -1/2$ and $F(\omega) = F(1/2) = \gamma(0)$ for all $\omega \ge 1/2$.

Proof. Sketched in class. □

Exercise 5.4 As an exercise, finish up the proof of Theorem 7 and fill in the details. The proof uses Helly's selection theorem: If $(G_n)_{n\ge1}$ is a sequence of increasing functions $G_n \colon \mathbb{R} \to \mathbb{R}$ that are uniformly bounded, e.g., $a \le G_n(x) \le b$, then there is a subsequence $(G_{n_k})_{k\ge1}$ that converges pointwise to a distribution function.

spectral density

The function F in Theorem 7 is the spectral distribution function of γ. If there is a non-negative function f such that

$$F(\omega) = \int_{-1/2}^{\omega} f(v)\,\mathrm{d}v, \quad \text{for } \omega \in [-1/2, 1/2],$$

then f is called the *spectral density* of γ.

The two above theorems combine to give the following corollary

Corollary 8. *A function $\gamma \colon \mathbb{Z} \to \mathbb{C}$ is the autocovariance function of a stationary process $X = \{X_t \colon t \in \mathbb{Z}\}$ if and only if*

(i) *$\gamma(h) = \int_{-1/2}^{1/2} \exp(i2\pi\omega h)\,\mathrm{d}F(\omega)$ for all $h \in \mathbb{Z}$, with F as described in Theorem 7; or*

(ii) *γ is non-negative definite.*

Exercise 5.5 Prove Corollary 8.

From the above corollary we see that a spectral distribution function F determines the autocovariance function γ. Importantly, the converse is also true, F is uniquely determined by γ. In light of the next theorem, this is easily see to be the case when γ is absolutely summable. The general case (i.e., no density) is harder to prove (see Ex. [xx make one xx]).

The next theorem is similar to a theorem a theorem pertaining to characteristic functions that we proved in Ex. **8(c)**.

Theorem 9. *If the autocovariance function $\gamma(h)$ of a stationary process is absolutely summable, i.e., $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then the spectral density f exists and is given by*

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h)\exp(-i2\pi h\omega), \quad -1/2 \le \omega \le 1/2.$$

Proof. Proved in class. Use the dominated convergence theorem to show that $\gamma(h) = \int_{-1/2}^{1/2} \exp(i2\pi h\omega) f(\omega)\,\mathrm{d}\omega$. $\square$

5.2. **Spectral distribution of ARMA processes.** Let $W = \{W_t : t \in \mathbb{Z}\}$ be white noise, i.e., mean zero random variables with autocovariance $\gamma_W(h) = 0$ if $h \neq 0$ and $\gamma_W(0) = \sigma^2$. Then $\sum_{h=-\infty}^{\infty} |\gamma_W(h)| = \gamma_W(0) = \sigma^2$, and the spectral density is $f_W(\omega) = \sigma^2$ for $-1/2 \leq \omega \leq 1/2$, zero outside.

The result of the next exercise was proven in class.

Exercise 5.6 Let $Y = \{Y_t : t \in \mathbb{Z}\}$ be a mean zero stationary process with autocovariance function γ_Y, and spectral distribution F_Y. Define

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}, \quad \text{where} \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

(a) Show that X is mean zero and stationary, with autocovariance functions

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h - j + k), \quad h \in \mathbb{Z}.$$

(b) Show that the spectral distribution of X is

$$F_X(\omega) = \int_{-1/2}^{\omega} \exp(i2\pi hv) \Big| \sum_{j=-\infty}^{\infty} \exp(-i2\pi jv) \Big|^2 \mathrm{d}F_y(v), \quad -1/2 < \omega < 1/2,$$

with $F_X(\omega) = 0$ for $\omega \leq -1/2$ and $F_X(\omega) = \gamma_X(0)$ for $\omega \geq 1/2$.

(c) Assume that Y has spectral density f_Y. Show that X has spectral density

$$f_X(\omega) = |\psi(\exp(-i2\pi\omega))|^2 f_Y(\omega),$$

where $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$.

Exercise 5.7 Let $X = \{X_t : t \in \mathbb{Z}\}$ be a causal ARMA(p, q) process, i.e., $\phi(B)X_t = \theta(B)W_t$, for a white noise process W. See Theorem 3 for conditions being equivalent to causality. spectral density of ARMA(p,q)

(a) Show that X has spectral density

$$f_X(\omega) = \frac{|\theta(\exp(-i2\pi\omega))|^2}{|\phi(\exp(-i2\pi\omega))|^2} \sigma^2, \quad -1/2 \leq \omega \leq 1/2,$$

and zero outside.

(b) Let X be a causal AR(1) process $X_t = \phi X_{t-1} + W_t$. Show that its spectral density is

$$f_1(\omega) = \frac{\sigma^2}{1 - 2\phi\cos(2\pi\omega) + \phi^2}, \quad \text{for } -1/2 \leq \omega \leq 1/2,$$

and zero outside. Deduce that

$$\int_{-\infty}^{1/2} f_1(\omega)\,\mathrm{d}\omega = \frac{\sigma^2}{1 - \phi^2}.$$

(c) Plot the spectral densities of an AR(1) process with $0 < \phi < 1$ and another with $-1 < \phi < 0$. Plot also one realisation of each of these processes. Attempt an interpretation of the spectral densities. In Figure 2 I have made two such plots of the spectral densities, with $\phi = 0.456$ and $\phi = -0.456$, with $\sigma^2 = 1$ in for both processes. Realisations of these processes are given in Figure 3.

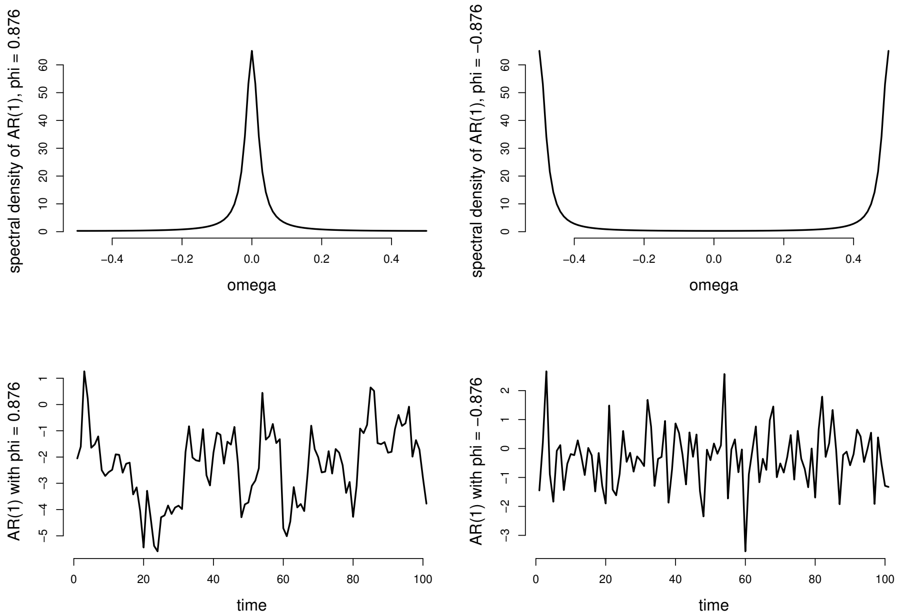

FIGURE 2. Top row: Spectral densities of AR(1) processes with positive and negative ϕ coefficients, and white noise with variance $\sigma^2 = 1$ in both. Bottom row: One realisation of each of these processes.

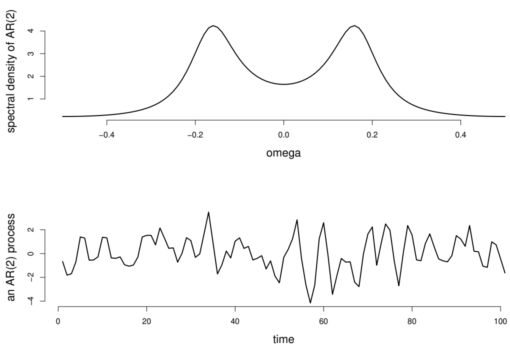

FIGURE 3. The spectral density and a realisation of an AR(2) process with $\phi_1 = 0.66$, $\phi_2 = -0.44$ and $\sigma^2 = 1$.

(d) Show that the spectral density of a (causal) AR(2) process is

$$f_2(\omega) = \frac{\sigma^2}{\{1 - \phi_1 \cos(2\pi\omega) - \phi_2 \cos(4\pi\omega)\}^2 + \{\phi_1 \sin(2\pi\omega) + \phi_2 \sin(4\pi\omega)\}^2},$$

for $-1/2 \leq \omega \leq 1/2$ and zero outside. Make a plot of the spectral density, perhaps also a realisation of the process, and attempt an interpretation. An example is given in Figure 3.

We are primarily interested in causal processes. There is, however, a theorem stating that if $\phi(z) \neq 0$ for all $|z| = 1$, then an ARMA(p, q) process has an unique stationary solution $X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$ with absolutely summable coefficients (see e.g. Brockwell and Davis (1991, Theorem 3.1.3, p. 88)). It is clear from you efforts in Ex. 5.7(a) that such a process has a spectral density of the same form as found in that exercise.

5.3. **Estimation, discrete Fourier transform and the periodogram.** Given that it exists, the spectral density of a stationary time series $X = \{X_t \colon t \in \mathbb{Z}\}$ is

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-i2\pi h\omega), \quad -1/2 \leq \omega \leq 1/2.$$

How may we estimate the spectral density? We will take a detour of sorts via complex vectors $x = (x_1, \dots, x_n) \in \mathbb{C}^n$. If $x, y \in \mathbb{C}^n$, their inner product is $\langle x, y\rangle = \sum_{i=1}^{n} x_i \bar{y}_i$. Soon enough, the x will be a (real valued) realisation of a sample $X_1, \dots, X_n$ from a stationary time series $X = \{X_t \colon t \in \mathbb{Z}\}$.

For a vector $x = (x_1, \dots, x_n)$, define its *fundamental frequencies*

$$F_n = \{j \in \mathbb{Z}\colon -1/2 < j/n < 1/2\},$$

and set

$$e_j = \frac{1}{\sqrt{n}} \begin{pmatrix} \exp(i2\pi j/n) \\ \exp(i2\pi 2j/n) \\ \vdots \\ \exp(i2\pi(n-1)j/n) \\ \exp(i2\pi nj/n) \end{pmatrix}, \quad \text{for } j \in F_n.$$

Exercise 5.8 We here show that $\{e_j,\ j \in F_n\}$ is an orthonormal basis for $\mathbb{C}^n$, and derive the discrete Fourier transform (DFT).

(a) Show that $\{e_j,\ j \in F_n\}$ is an orthonormal set.

(b) Show that $x = \sum_{j\in F_n} d(j/n) e_j$ for any $x \in \mathbb{C}^n$, that is

$$x_t = \frac{1}{\sqrt{n}} \sum_{j\in F_n} d(j/n) \exp(i2\pi tj/n), \quad t = 1, \dots, n,$$

where $d(j/n)$ for $j \in F_n$ is the *discrete Fourier transform* (DFT) discrete Fourier transform

$$d(j/n) = \langle x, e_j\rangle = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t \exp(-i2\pi tj/n).$$

(c) The *periodogram* $I(j/n)$ is defined as the squared modulus of the DFT, periodogram

$$I(j/n) = |d(j/n)|^2,$$

for $j = 0, 1, 2, \dots, n-1$. Recall that the modulus of the complex number $z = a + ib$ is $|z| = \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}$. For any complex number z, different from 1, you may show that

$$\sum_{k=0}^{n} z^k = 1 + z + z^2 + \cdots + z^n = \frac{1 - z^{n+1}}{1 - z}.$$

Use this to show that

$$\sum_{t=1}^{n} \exp(-i2\pi tj/n) = 0.$$

Deduce that for any constant a, the DFT can be expressed as

$$d(j/n) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (x_t - a) \exp(-i2\pi tj/n), \quad j \in F_n.$$

We are particularly interested in the case where $a = \bar{x}_n = n^{-1}\sum_{i=1}^n x_i$, or $a = \mu$, where μ is the mean of a stationary time series.

(d) Show that (we here assume $x \in \mathbb{R}^n$ to ease the notation) $I(0) = n\bar{x}_n^2$, and for $j \in F_n \setminus \{0\}$

$$
\begin{aligned}
I(j/n) &= \frac{1}{n}\sum_{t=1}^n\sum_{s=1}^n (x_t - \bar{x}_n)(x_s - \bar{x}_n)\exp(-i2\pi(t-s)j/n) \\
&= \sum_{h=-(n-1)}^{n-1} \widehat{\gamma}(h)\exp(-i2\pi hj/n),
\end{aligned}
$$

where

$$
\widehat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x}_n)(x_t - \bar{x}_n).
$$

This certainly hints at $I(j/n)$ being a decent estimator of the spectral density (when x is the realistion of a stationary time series).

Let $X = \{X_t \colon t \in \mathbb{Z}\}$ be a stationary times series with mean μ and absolutely summable autocovariance function $\gamma(h)$. To use the periodogram as an estimator of the spectral density, the domain of the periodogram must be extended to $[-1/2, 1/2]$. To this end, for $n \geq 1$ and $\omega \in [0, 1/2]$, define

$$
g(n,\omega) = \sum_{j \in F_n} \frac{j}{n} I\{j/n - 1/(2n) < \omega \leq j/n + 1/(2n)\},
$$

and for $\omega \in [-1/2, 0)$, set $g(n,\omega) = g(n,-\omega)$. We know define the extension (but use the same symbol)

$$
I(\omega) = I(g(n,\omega)), \quad -1/2 \leq \omega \leq 1/2.
$$

You may verify that the periodogram is an even function, $I(\omega) = I(-\omega)$; and that when $\omega = j/n$ for $j \in F_n$, then $I(\omega)$ agrees with the definition of the periodogram given in Ex. 5.8. Also, $g(n,\omega) \to \omega$ as $n \to \infty$.

Exercise 5.9 Let $X = \{X_t \colon t \in \mathbb{Z}\}$ be a stationary time series with mean μ, absolutely summable autocovariance function $\gamma(h)$, and spectral density f. Let $I_n(\omega)$ be the periodogram of the sample $X_1, \ldots, X_n$. Verify that we may write

$$
I_n(\omega) = \sum_{h=-(n-1)}^{n-1} \frac{1}{n}\sum_{t=1}^{n-|h|}(X_{t+|h|} - \mu)(X_t - \mu)\exp(-i2\pi h g(n,\omega)),
$$

for $\omega \in [-1/2, 0) \cup (0, 1/2]$.

(a) Show that

$$
\mathrm{E}\, I_n(0) - n\mu^2 \to f(0),
$$

and that

$$
\mathrm{E}\, I_n(\omega) \to f(\omega), \quad \text{for } \omega \in [-1/2, 0) \cup (0, 1/2].
$$

(b) Show that when the mean $\mu = 0$, then

$$
\sup_{-1/2 \leq \omega \leq 1/2} |\mathrm{E}\, I_n(\omega) - f(\omega)| \to 0,
$$

that is, we have uniform convergence.

Notice, however, that the above exercise does not imply that the periodogram is a consistent estimator of the spectral density (i.e., convergence in probability is not implied).

5.4. **Convergence in distribution of the periodogram.** The DFT of sample $X_1, \dots, X_n$ can be expressed as expressed as

$$d(j/n) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \exp(-i2\pi tj/n) = \frac{1}{\sqrt{2}} \big(\alpha_n(j/n) - i\beta_n(j/n)\big),$$

where

$$\alpha(j/n) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} X_t \cos(2\pi tj/n) \quad \text{and} \quad \beta(j/n) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} X_t \sin(2\pi tj/n).$$

For reasons that will become clear, we will only need these for $j = 1, \dots, q$, with $q = \lfloor (n-1)/2 \rfloor$. Notice that the periodogram of $X_1, \dots, X_n$ can then be expressed as

$$I(j/k) = \frac{1}{2} \big\{ \alpha(j/n)^2 + \beta(j/n)^2 \big\}, \quad j = 1, \dots, q.$$

We may also introduce the vectors

$$c_j = \sqrt{\frac{2}{n}} \begin{pmatrix} \cos(2\pi j/n) \\ \cos(2\pi 2j/n) \\ \vdots \\ \cos(2\pi nj/n) \end{pmatrix}, \quad \text{and} \quad s_j = \sqrt{\frac{2}{n}} \begin{pmatrix} \sin(2\pi j/n) \\ \sin(2\pi 2j/n) \\ \vdots \\ \sin(2\pi nj/n) \end{pmatrix},$$

for $j = 1, \dots, q$.

Exercise 5.10 Let $X = (X_1, \dots, X_n)$, and $q = \lfloor (n-1)/2 \rfloor$ as above.

(a) Verify that $\{c_1, s_1, \dots, c_q, s_q\}$ is an orthonormal set, that is, $\langle c_j, c_j \rangle = 1$, and $\langle c_j, c_k \rangle = 0$ for $j \neq k$, $\langle c_j, s_k \rangle = 0$ for an j, k, and so on. Here $\langle x, y \rangle = \sum_{i=1}^{q} x_i y_i$. Use this to show that the random variables $\alpha(j/n), \beta(j/n)$ for $j = 1, \dots, q$ are mean zero and uncorrelated.

(b) Suppose $X_1, \dots, X_n$ are i.i.d. $\mathrm{N}(0, \sigma^2)$. Deduce from the above that the periodogram of these random variables are

$$I(j/n) \sim \sigma^2 U, \quad U \sim \mathrm{Expo}(1),$$

i.e. U has c.d.f. $\Pr(U \leq u) = 1 - \exp(-u)$ for $u \geq 0$, and $\Pr(U \leq u) = 0$ for $u < 0$. In fact, establish that for n big enough (to ensure that for $0 < \omega_1 < \omega_2 < \cdots < \omega_m < 1/2$, none of the $g(n, \omega_i)$ for $i = 1, \dots, m$ are equal), we have that

$$(I(\omega_1), \dots, I(\omega_m)) \sim (\sigma^2 U_1, \dots, \sigma^2 U_m),$$

where $U_1, \dots, U_m$ are i.i.d. Expo(1). and we write $I(\omega_i) = I(g(n, \omega_1))$.

In the next exercise we establish an asymptotic version of this result. Keep also in find that when $X_1, \dots, X_n$ are i.i.d. $\mathrm{N}(0, \sigma^2)$ as here, or, more generally white noise with variance σ^2, then it has spectral density σ^2.

Exercise 5.11 Let $X_1, \dots, X_n$ be i.i.d. random variables with variance σ^2, and write I_n for its periodogram. For $0 < \omega_1 < \omega_2 < \ldots < \omega_m < 1/2$, we are to show that

$$(I_n(\omega_1), \dots, I_n(\omega_m)) \to_d (\sigma^2 U_1, \dots, \sigma^2 U_m), \tag{19}$$

for $U_1, \dots, U_m$ i.i.d. Expo(1).

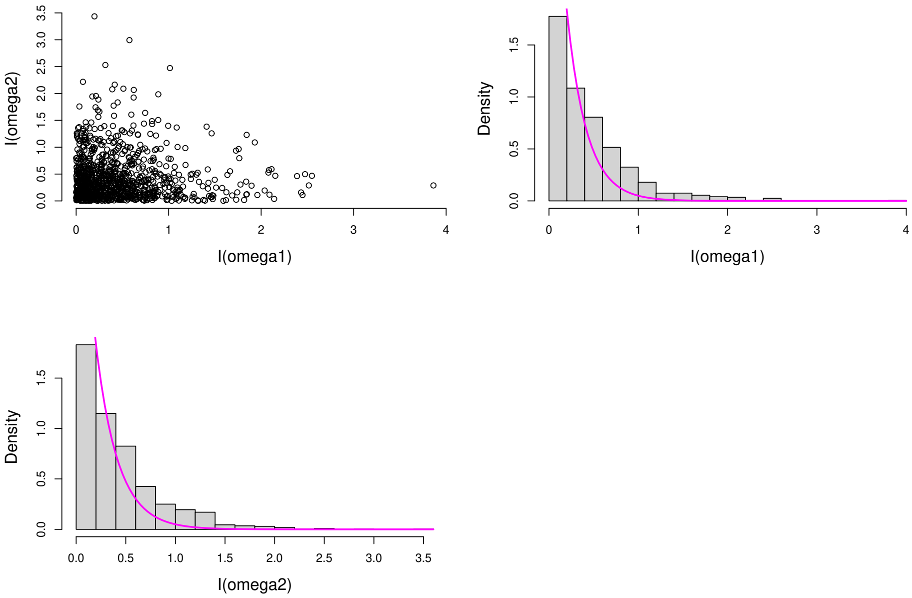

Figure 4. The simulations of Ex. 5.11**(c)**. In the histograms the exact densities are indicated in magenta.

(a) Write $X^n = (X_1, \dots, X_n)$ and $\alpha(g(n,\omega)) = \langle X^n, c_{g(n,\omega)n}\rangle$ and $\beta(g(n,\omega)) = \langle X^n, s_{g(n,\omega)n}\rangle$ be as defined above. Explain that the claim in (19) will follow if we show the joint convergence

$$\big(\alpha(g(n,\omega_1)), \beta(g(n,\omega_1)), \dots, \alpha(g(n,\omega_m)), \beta(g(n,\omega_m))\big)^{\mathrm{t}} \to_d \mathrm{N}_{2m}(0, \sigma^2 I_{2m}),$$

and that in order to prove this joint convergence is it essentially enough to show that $\alpha(g(n,\omega)) \to_d \mathrm{N}(0,\sigma^2)$.

(b) Show that $\alpha(g(n,\omega)) \to_d \mathrm{N}(0,\sigma^2)$. To show this it suffices to show that $\alpha(g(n,\omega))$ has the variance it should, and that it satisfies the Lindeberg condition. See, for example, Theorem 5 in Stoltenberg (2025).

(c) Illustrate the result of this exercise via simulations. In Figure 4 I have plotted 1000 simulations of $(I_n(\omega_1), I_n(\omega_2))$ for $n = 100$, and $\omega_1 = 0.12$ and $\omega_2 = 0.25$. The X_t were taken as i.i.d. $\mathrm{Gamma}(2,3) - 2/3$. Try out some other distributions as well.

The lemma of the next exercise is the key to the convergence in distribution result for the periodogram of linear processes.

Exercise 5.12 Let $X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$ for $t \in \mathbb{Z}$, where $Z = \{Z_t \colon t \in \mathbb{Z}\}$ is i.i.d. white noise with variance σ^2. In Ex. 5.6**(c)** we saw that the spectral density of X is of the form $f_X(\omega) = |\psi(\exp(-i2\pi\omega)|^2 f_W(\omega)$ for $-1/2 \leq \omega \leq 1/2$, in terms of the spectral density f_W of W, which is $f_W(\omega) = \sigma^2$ since W is white noise. Since the periodogram is a sample version of the spectral density, it might not be surprising that a similar relationship holds between the peridodograms of $X_1, \dots, X_n$ and $W_1, \dots, W_n$. In fact, with $I_{n,X}$ and $I_{n,W}$ the periodograms of

$X_1, \dots, X_n$ and $W_1, \dots, W_n$, respectively,

$$I_{n,X}(k/n) = |\psi(\exp(-i2\pi k/n))|^2 I_{n,W}(k/n) + r_n(k/n), \tag{20}$$

for $k \in F_n, k \geq 0$, where $\max_k \mathrm{E}\,|r_n(k/n)|^2 \to 0$. This is what we set out to show in this exercise.

(a) Let d_X and d_W be the DFTs of $X_1, \dots, X_n$ and $W_1, \dots, W_n$, respectively. Show that for any $k \in F_n, k \geq 0$,

$$\begin{aligned} d_X(k/n) &= \frac{1}{\sqrt{n}} \sum_{j=-\infty}^{\infty} \psi_j \exp(-i2\pi jk/n) \big\{ \sqrt{n} d_W(k/n) + U_{n,j} \big\} \\ &= \psi(\exp(-i2\pi k/n) d_W(k/n) + Y_n(k/n), \end{aligned}$$

where

$$U_{n,j} = \sum_{t=1-j}^{n-j} W_t \exp(-i2\pi tk/n) - \sum_{t=1}^{n} W_t \exp(-i2\pi tk/n),$$

and

$$Y_n(k/n) = \frac{1}{\sqrt{n}} \sum_{j=-\infty}^{\infty} \psi_j \exp(-i2\pi jk/n) U_{n,j}.$$

(b) Verify that for $|j| < n$, the random variable $U_{n,j}$ is a sum of $2|j|$ independent random variables, while when $|j| \geq n$, it is a sum of $2n$ random variables. Deduce the bound

$$\mathrm{E}\,|U_{n,j}|^2 \leq 2\sigma^2 \min(|j|, n).$$

(c) Show that

$$\mathrm{E}\,|Y_n(k/n)|^2 \leq 2\sigma^2 \left(n^{-1/2} \sum_{j=-\infty}^{\infty} |\psi_j| \min(|j|, n)^{1/2} \right)^2.$$

For any $m \geq 1$, show that

$$\limsup_{n\to\infty} n^{-1/2} \sum_{j=-\infty}^{\infty} |\psi_j| \min(|j|, n)^{1/2} \leq \sum_{|j|>m} |\psi_j|,$$

and conclude that $\max_k \mathrm{E}\,|Y_n(k/n)|^2 \to 0$, where the max is over $k \in F_n, k \geq 0$.

(d) Verify that since $I_{n,X}(k/n) = |d_X(k/n)|^2 = d_X(k/n) d_X(-k/n)$, the remainder term in (20) is

$$r_n(\omega_k) = \psi(e^{-i2\pi\omega_k}) d_W(\omega_k) Y_n(-\omega_k) + \psi(e^{-i2\pi\omega_k}) d_W(-\omega_k) Y_n(\omega_k) + |Y_n(\omega_k)|^2,$$

where $\omega_k = k/n$. Deduce from this that

$$\max_{k\in F_n, k\geq 0} \mathrm{E}\,|r_n(k/n)|^2 \to 0,$$

and check that this implies $\sup_{0\leq\omega\leq 1/2} \mathrm{E}\,|r_n(g(n,\omega))|^2 \to 0$, and $r_n(g(n,\omega)) \to_p 0$. This finishes the proof of (20).

Exercise 5.13 Let $X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$ for i.i.d. white noise W and coefficients such that $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Let $I_{n,X}$ and $I_{n,W}$ be the periodograms of $X_1, \dots, X_n$ and $W_1, \dots, W_n$, respectively. Let f be the spectral density of X, and assume that $f(\omega) > 0$ for all $\omega \in [-1/2, 1/2]$.

(a) Deduce from the above exercise that for any $\omega \in [0, 1/2]$,

$$I_{n,X}(\omega) = f(g(n,\omega)) I_{n,W}(\omega)/\sigma^2 + r_n(g(n,\omega)).$$

(b) Now, deduce from previous efforts that for $0 < \omega_1 < \omega_2 < \ldots < \omega_m < 1/2$,

$$\big(I_{n,X}(\omega_1), \ldots, I_{n,X}(\omega_m)\big) \to_d (f(\omega_1)U_1, \ldots, f(\omega_m)U_m), \tag{21}$$

where $U_1, \ldots, U_m$ are i.i.d. Expo(1) random variables.

(c) For an causal AR(1) time series $X_t = \phi X_{t-1} + W_t$, simulate 1000 realisations of $(I_{n,X}(\omega_1), I_{n,X}(\omega_2))$ and make plots such as those in Figure 4, illustrating the asymptotic result.

5.5. **Whittle likelihood.** The joint density of the left hand side in (21), i.e., the limiting distribution, is

$$\prod_{j=1}^{m} (1/f(\omega_j)) \exp\{-u_j/f(\omega_j)\}. \tag{22}$$

Whittle log-likelihood

The form of this density motivates the approximate log-likelihood function

$$\ell_n^w = -\sum_{j=1}^{\lfloor (n-1)/2 \rfloor} \big\{ \log f(j/n) + \frac{|d_X(j/n)|^2}{f(j/n)} \big\}, \tag{23}$$

called the Whittle log-likelihood (after Peter Whittle who introduced it in his PhD at Uppsala universitet in 1951). When the spectral density f is parametrised, i.e., $f(\cdot, \theta)$, so that

$$\ell_n^w(\theta) = -\sum_{j=1}^{\lfloor (n-1)/2 \rfloor} \big\{ \log f(j/n, \theta) + \frac{|d_X(j/n)|^2}{f(j/n, \theta)} \big\},$$

this is approximately a likelihood function as those we are used to. One major difference between (23) and (22) is that in (22) m is fixed, while in (23) $m = m_n = \lfloor (n-1)/2 \rfloor$. It is harder to show that (23) approximates the limiting density of the (growing) vector $(I_{n,X}(1/n), \ldots, I_{n,X}(m_n/n))$ than showing that (22) approximates the limiting density of $((I_{n,X}(\omega_1), \ldots, I_{n,X}(\omega_m))$ for fixed m, as we indeed have shown above. That being said, one can estimate parameters however ones likes, the arbiter is not how an estimator is derived, but how well it performs.

Exercise 5.14 (xx the oblig exercise can be included here xx)

Exercise 5.15 Simulate data from a causal AR(2) process.

(a) Use the Whittle likelihood to estimate the parameters ϕ_1 and ϕ_2.

(b) Also estimate the parameters using the Yule–Walker equations (essentially least squares).

(c) Perform a simulation study to investigate the (possible) difference in performance of the two estimators.

5.6. **The spectral representation theorem.** From a statistical perspective, the spectral distribution is more important than the spectral representation. This is because the spectral distribution is a natural estimand. [xx write this out for completeness at some point, but not part of curriculum this semester xx]

5.7. **Unit roots.** Recall that an ARMA process $\phi(B)X = \theta(B)W$ is causal if and only if $\phi(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$. Above we have also quoted a theorem saying that $\phi(B)X = \theta(B)W$ has a unique stationary solution of the form $X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$ provided $\psi(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| = 1$ (Brockwell and Davis, 1991, p. 88).

The AR(1)-process (once again) $X_t = \phi X_{t-1} + W_t$ serves as a nice example. We known that if $|\phi| < 1$ then the process is stationary and causal, and that if $|\phi| > 1$ then a stationary (but not causal) solution exists, see Ex. 2.6. Given a sample $X_1, \ldots, X_n$ from a AR(1) process it is therefore of interest to test whether $\phi(z) = 1 - \phi z$ has a unit root, equivalently, whether $\phi = 1$.

Exercise 5.16 The details of the following exercise is not part of the curriculum in stk4060. But the student ought to know the statement of Donsker's theorem. For some of the exercises below we need the concept of a sequence of stochastic processes converging to a limiting stochastic process. In particular, we will encounter situations where a sequence $(\xi_n)_{n\geq 1}$ of stochastic processes on $[0,1]$ converges weakly to a Brownian motion B on $[0,1]$.

(a) Go to R and run the following little code,

```
n <- 10^4
ww <- 2*rbinom(n,1,1/2) - 1
xi <- cumsum(ww)/sqrt(n)
plot(1:n,xi,type="l",lwd=1,frame.plot=FALSE)
```

You might also try this script with

```
ww <- 2*rbinom(n,1,1/2) - 1
```

and observe that you get more or less the same thing. From the plots made in this script we sort of see that the sequence of processes $(\xi_n)_{n\geq 1}$ where

$$\xi_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} W_i, \quad \text{for } t \in [0,1], \tag{24}$$

and $W_1, W_2, \ldots$ are i.i.d. mean zero random variables with unit variance, converges in distribution (weakly) to a Brownian motion on $[0,1]$.

(b) To prove that a sequence of processes $(\xi_n)_{n\geq 1}$ converges in distribution to some limit process ξ it suffices to establish finite-dimensional convergence and tightness. Finite-dimensional convergence means that

$$(\xi_n(t_1), \ldots, \xi_n(t_k)) \to_d (\xi(t_1), \ldots, \xi(t_k)), \tag{25}$$

for all $0 < t_1 < t_2 < \cdots < t_k < 1$ and all k; while tightness means that for any $\varepsilon > 0$ we can find a compact set K such that

$$\Pr(\xi_n \in K) > 1 - \varepsilon, \quad \text{for all } n. \tag{26}$$

Why is this enough? First, recall from Ex. 1.5**(c)** (where a theorem in Kallenberg (2002) is cited) that the processes X and Y have the same distribution if all their finite-dimensional distributions agree. Second, by Prokhorov's theorem (e.g., Billingsley, 1968, Theorem 6.1, p. 37), if a sequence $(\xi_n)_{n\geq 1}$ is tight then it has a weakly convergent subsequence. Deduce from these two facts that (26) and (25) entail that $\xi_n \to_d \xi$, as a process.

(c) The compactness of the set K in (26) requires that we are precises about the function space we are working on, that is, the space of which K is a compact subset. There are two space that appear most often: (i) The space of continuous function on $[0,1]$ equipped with the uniform metric $\sup_{t\in[0,1]} |x(t) - y(t)|$. We denoted this space $C[0,1]$. And (ii) the space of càdlàg function (right continuous with left hand limits) in $[0,1]$ equipped with the Skorokhod metric (see Billingsley, 1968, Ch. 3). This space is denoted $D[0,1]$. Both $D[0,1]$ and $C[0,1]$ are complete and separable metric spaces (so-called Polish space), meaning that all Cauchy sequences converge

$C[0,1]$

$D[0,1]$

(completeness) and that they have a countable dense subset (separable). Attempt to prove this fact for $C[0,1]$.

(d) In $D[0,1]$ and i $C[0,1]$ there is a useful criterion for establishing tightness of a sequence $(\xi_n)_{n\geq 1}$. In $D[0,1]$, of for $t_1 \leq t \leq t_2$, $\gamma \geq 0$ and $\alpha \geq 1/2$ we have that

$$\mathrm{E}\,|\xi_n(t) - \xi_n(t_1)|^{\gamma} |\xi_n(t_2) - \xi_n(t)|^{\gamma} \leq \{F(t_2) - F(t_1)\}^{2\alpha}$$

for a non-decreasing function function F that is continuous on $[0,1]$, then $(\xi_n)_{n\geq 1}$ is tight (Billingsley, 1968, Th. 15.6, p. 128). A sequence in $C[0,1]$ is tight of $(\xi_n()0))_{n\geq 1}$ is tight and $\mathrm{E}\,|\xi_n(t_2) - \xi_n(t_1)|^{\gamma} \leq \{F(t_2) - F(t_1)\}^{\alpha}$ for $\gamma \geq 0$, $\alpha > 1$, and non-decreasing function F that is continuous on $[0,1]$ (Billingsley, 1968, Th. 12.3, p. 95).

(e) Attempt to prove Donsker's theorem: Namely that the sequence $(\xi_n)_{n\geq 1}$ in $D[0,1]$ defined in (24) converges in distribution to a Brownian motion on $[0,1]$. We also mention that the the continuous mapping theorem holds in this case, that is, if $h\colon D[0,1] \to \mathcal{X}$ is continuous and $\xi_n \to_d B$, then $h(\xi_n) \to_d h(B)$ (Billingsley, 1968, Th. 5.1, p. 30).

Exercise 5.17 Let $X_1, \ldots, X_n$ be a sample from the AR(1) process $X_t = \phi X_{t-1} + W_t$, $t \in \mathbb{Z}$ where $\{W_t \colon t \in \mathbb{Z}\}$ is i.i.d. white noise with variance σ^2. Let $\widehat{\phi}_n$ be the least squares estimator for ϕ, so that

$$\widehat{\phi}_n - \phi = \frac{\sum_{t=2}^n X_{t-1} W_t}{\sum_{t=2}^n X_{t-1}^2}. \tag{27}$$

In this exercise we derive the limiting distribution of this quantity under the null hypothesis that $\phi = 1$. At this point it is perhaps worth looking back at Ex. 1.16 and Ex. 4.7 and think about how actively the causal representation $X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$ were used in those proofs. When $\phi = 1$, this representation is not available, so we anticipate another proof and a different limiting distribution.

(a) Show that $X_t = X_0 + \sum_{j=1}^t W_j$ for any $t \geq 0$. Use this to show that the numerator in (27) is

$$\sum_{t=2}^n X_{t-1} W_t = X_0 \sum_{t=2}^n W_t + \tfrac{1}{2}\Big(\sum_{t=1}^n W_t\Big)^2 - \tfrac{1}{2} \sum_{t=1}^n W_t^2,$$

and deduce that

$$\frac{1}{n} \sum_{t=2}^n X_{t-1} W_t \to_d \frac{\sigma^2}{2}(\chi_1^2 - 1),$$

where χ_1^2 is a chi-square(1) random variable, i.e., a Gamma$(\frac{1}{2}, \frac{1}{2})$.

(b) Let B be a Brownian motion on $[0,1]$. By an application of Itô's lemma, show that

$$\int_0^1 B(s)\,\mathrm{d}B(s) \sim \tfrac{1}{2}(\chi_1^2 - 1).$$

i.e., the left and the right side have the same distribution.

(c) We now turn to the numerator. Introduce $\xi_n(t) = n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} W_i$ for $t \in [0,1]$, and show that

$$\frac{1}{n^2} \sum_{t=2}^n X_{t-1}^2 = \frac{1}{n} X_0 + 2X_0 \frac{1}{\sqrt{n}} \int_0^1 \xi_n(s)\,\mathrm{d}s + \int_0^1 \xi_n(s)^2\,\mathrm{d}s \to_d \int_0^1 B(s)^2\,\mathrm{d}s$$

(d) Since the Cramér–Slutsky rules are not applicable to the ratio in (27) (why?), we need to show that

$$\big(n^{-1}\sum_{t=2}^{n} X_{t-1}W_t, n^{-2}\sum_{t=2}^{n} X_{t-1}^2\big) \to_d \big(\int_0^1 B(s)\,\mathrm{d}B(s), \int_0^1 B(s)^2\,\mathrm{d}s\big).$$

Attempt to do so, and conclude that

$$n(\widehat{\phi}_n - 1) \to_d \frac{\int_0^1 B(s)\,\mathrm{d}B(s)}{\int_0^1 B(s)^2\,\mathrm{d}s}.$$

The distribution on the right is a so-called pivot, as it does not depend on unknowns.

(e) Simulate data from a random walk $X_t = X_{t-1} + W_t$. Use the limit distribution above to test null hypothesis of a unit root on the 5 percent level.

Exercise 5.18 Let $X_1, \dots, X_n$ be a sample from the process $X_t = \phi X_{t-1} + V_t$ where V is the linear process $V_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$, for i.i.d. white noise $W = \{W_t : t \in \mathbb{Z}\}$ with variance σ^2, and $0 < \sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Let γ be the autocovariance function of V, so $\gamma(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_{j+h}\psi_j$. Generalise the result of the previous exercise to this more involved setting.

(a) Proceed as in the previous exercise to show that

$$n(\widehat{\phi}_n - 1) \to_d \frac{\int_0^1 B(s)\,\mathrm{d}B(s)}{\int_0^1 B(s)^2\,\mathrm{d}s} + \frac{\sum_{h=1}^{\infty} \gamma(h)}{\sum_{h=-\infty}^{\infty} \gamma(h) \int_0^1 B(s)^2\,\mathrm{d}s}.$$

(b) Propose a consistent estimator, say b_n, such that

$$n(\widehat{\phi}_n - 1) - b_n \to_d \frac{\int_0^1 B(s)\,\mathrm{d}B(s)}{\int_0^1 B(s)^2\,\mathrm{d}s}.$$

(c) Use the above to construct a test of the null hypothesis $\phi = 1$, and perform a small simulation study to check the performance of your test.

Exercise 5.19 The large sample theory underpinning the augmented Dickey–Fuller test is easiest to understand in the AR(2)-case. So we start with $p = 2$, before generalising to any p. Let $X = \{X_t : t \in \mathbb{Z}\}$ be the process $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + W_t$ where W is i.i.d. white noise with variance σ^2. We want to test for unit roots, that is, the null hypothesis

$$\phi(1) = 1 - \phi_1 - \phi_2 = 0,$$

where $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$ is the AR(2)-polynomial.

(a) Verify that we can write

$$X_t = \rho X_{t-1} + \beta \Delta X_{t-1} + W_t,$$

with $\rho = \phi_1 + \phi_2$ and $\beta = -\phi_2$, and $\Delta X_t = X_t - X_{t-1}$. The unit root null hypothesis now reads $\rho = 1$.

(b) Define $\beta(z) = 1 - \beta z$, and let L be the lag-operator $LX_t = X_{t-1}$, so $(1-L)X_t = \Delta X_t$ (now denoted L instead of B to distinguish it from the Brownian motion that will appear). Establish that under the null hypothesis

$$W_t = \phi(L)X_t = \beta(L)(1-L)X_t,$$

so that

$$(1-L)X_t = \beta(L)W_t = \sum_{j=0}^{\infty} \beta^j W_{t-j},$$

provided $\beta(L)$ can be inverted, i.e., that $|\beta| < 1$, which we assume (see Ex. 2.3 and Ex. 2.5). Thus, defining $V_t = \sum_{j=0}^{\infty} \beta^j W_{t-j}$ we can write

$$X_t = X_{t-1} + V_t, \quad t \in \mathbb{Z},$$

which is a process of the type studied (under the null) in Ex. 5.18. Let $\gamma_V(h) = \sigma^2 \beta^{|h|}/(1-\beta^2)$ be the autocovariance function of V. From Ex. 4.6 we know that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} V_t \to_d \mathrm{N}\big(0, \sum_{h=-\infty}^{\infty} \gamma_V(h)\big),$$

and here $\sum_{h=-\infty}^{\infty} \gamma_V(h) = \sigma^2/(1-\beta)^2$.

(c) Suppose we have a sample $X_1, \ldots, X_n$ from X. Let $\theta = (\rho, \beta)^{\mathrm{t}}$, and let $\widehat{\theta}_n$ be the least square estimator. Introduce $Z_t = (X_t, \Delta X_t)^{\mathrm{t}}$, and also the matrix

$$A_n = \begin{pmatrix} n & 0 \\ 0 & \sqrt{n} \end{pmatrix}.$$

Verify that

$$a_n^{-1}(\widehat{\theta}_n - \theta) = \big(a_n^{-1} \sum_{t=3}^{n} Z_{t-1} Z_{t-1}^{\mathrm{t}} a_n^{-1}\big)^{-1} a_n^{-1} \sum_{t=3}^{n} Z_{t-1} W_t$$

and that

$$a_n^{-1} \sum_{t=3}^{n} Z_{t-1} Z_{t-1}^{\mathrm{t}} a_n^{-1} = \begin{pmatrix} n^{-2} \sum_{t=3}^{n} X_{t-1}^2 & n^{-3/2} \sum_{t=3}^{n} X_{t-1} \Delta X_{t-1} \\ n^{-3/2} \sum_{t=3}^{n} X_{t-1} \Delta X_{t-1} & n^{-1} \sum_{t=3}^{n} (\Delta X_{t-1})^2 \end{pmatrix},$$

and, finally, that

$$a_n^{-1} \sum_{t=3}^{n} Z_{t-1} W_t = \sum_{t=3}^{n} \begin{pmatrix} n^{-1} X_{t-1} \\ n^{-1/2} \Delta X_{t-1} \end{pmatrix} W_t.$$

(d) We anticipate that both the numerator and the denominator will converge in distribution (and not in probability), so joint convergence is called for. But to sort things out, let's start term by term. Deduce from Ex. 5.18 that

$$\frac{1}{n^2} \sum_{t=3} X_{t-1}^2 \to_d \big\{ \sum_{h=-\infty}^{\infty} \gamma_V(h) \big\} \int_0^1 B(s)^2 \,\mathrm{d}s.$$

Show also that

$$n^{-3/2} \sum_{t=3}^{n} X_{t-1} \Delta X_{t-1} \to_p 0,$$

and that

$$n^{-1} \sum_{t=3}^{n} (\Delta X_{t-1})^2 \to_p \gamma_V(0) = \frac{\sigma^2}{1-\beta^2}.$$

(e) Now we turn to the numerator. Verify that

$$\frac{1}{n} \sum_{t=3}^{n} X_{t-1} W_t = \frac{1}{n} \sum_{t=3}^{n} X_{t-1} V_t \to_d (1-\beta) \sum_{h=-\infty}^{\infty} \gamma_V(h) \int_0^1 B(s) \,\mathrm{d}B(s).$$

(f) Deduce from the above that

$$n(\widehat{\rho}_n - 1) \to_d \frac{(1-\beta) \int_0^1 B(s) \,\mathrm{d}B(s)}{\int_0^1 B(s)^2 \,\mathrm{d}s}.$$

From this, construct a test for the null hypothesis $\rho = 1$.

6. State space models

In state spaces models we have an equation for the process we observe and an equation for the underlying process. The observation equation is

$$Y_t = A_t X_t + V_t, \quad t \geq 1, \tag{28}$$

while the system or state equation (the underlying process) is

$$X_t = \Phi X_{t-1} + W_t, \quad t \geq 1. \tag{29}$$

Here Y_t is a q dimensional random vector, while X_t is a p dimensional random vector. Thus A_t are $q \times p$ dimensional parameter matrices; Φ is a $p \times p$ parameter vector; and we take $(W_t)_{t\geq 1}$ to consist of i.i.d. $\mathrm{N}_p(0, Q)$ and $(V_t)_{t\geq 1}$ as to consist of i.i.d. $\mathrm{N}_q(0, R)$ with the two sequences independent of each other. Finally, we take $X_0 \sim \mathrm{N}_p(\mu_0, \Sigma_0)$ independent of $(V_t)_{t\geq 1}$ and $(W_t)_{t\geq 1}$. The primary goal, or an intermediate goal on the way to parameter estimation, is the estimation of the state X_t. Let

$$\mathcal{D}_t = \mathcal{D}_0 \vee \sigma(Y_1, \ldots, Y_t), \quad t \geq 0,$$

where $\mathcal{D}_0$ is prior information leading to the distribution of X_0, that is, $X_0 \,|\, \mathcal{D}_0 \sim \mathrm{N}_p(\mu_0, \Sigma_0)$. (The σ-algebra $\mathcal{A} \vee \mathcal{B}$ is the smallest σ-algebra containing the σ-algebras $\mathcal{A}$ and $\mathcal{B}$.) Having observed $Y_1, \ldots, Y_s$, the problem if estimating X_t is called

- forecasting if $s < t$;
- filtering if $s = t$;
- smoothing if $s > t$.

Define

$$\mu_t^s = \mathrm{E}\,(X_t \,|\, \mathcal{D}_s), \quad \text{and} \quad P_t^s = \mathrm{E}\,((X_t - \mu_t^s)^2 \,|\, \mathcal{D}_s). \tag{30}$$

Exercise 6.1 Consider the causal AR(1) process $X = \{X_t \colon t \in \mathbb{Z}\}$, $X_t = \phi X_{t-1} + W_t$, that is observed with noise at times $t \in \{1, 2, \ldots\}$,

$$Y_t = X_t + V_t,$$

Assume that $(W_t)_{t\in\mathbb{Z}}$ and $(V_t)_{t\in\mathbb{N}}$ are independent white noise processes with $W_t \sim \mathrm{N}(0, \sigma_W^2)$ and $V_t \sim \mathrm{N}(0, \sigma_V^2)$.

(a) Find the expectation and autocovariance function of the observation process (Y_t). Show that (Y_t) is stationary, in fact, show that is indeed a stationary ARMA$(1,1)$. Is it causal?

(b) Simulate some data from such a model, and devise a way of estimation ϕ. Check the performance of your estimator in simulations.

Exercise 6.2 For the updating theorems to come we need some knowledge of the multivariate normal distribution. Here is an exercise for when Nils taught stk4060 in 2022. A vector $X = (X_1, \ldots, X_m)$ has the multinormal distribution, with mean ξ and covariance matrix Σ, if its density takes the form

$$f(x) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp\{-\tfrac{1}{2}(x - \xi)^{\mathrm{t}} \Sigma^{-1} (x - \xi)\}, \quad x \in \mathbb{R}^m.$$

We write $X \sim \mathrm{N}_m(\xi, \Sigma)$ to indicate this; note that the distribution is fully specified by giving the ξ and the Σ.

(a) Check that this becomes the classic formula for $\mathrm{N}(\xi, \sigma^2)$ in the one-dimensional case. In the general case, show that $Y = AX$ has distribution $\mathrm{N}_m(A\xi, A\Sigma A^{\mathrm{t}})$, if A is a $m \times m$ matrix. Show that f integrates to 1.

(b) Block X into $X_{(1)}$ and $X_{(2)}$ with dimensions p and q respectively, so $p+q=m$. Write

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

with Σ_{11} of dimension $p \times p$, etc.. Try to show that $X_{(1)} \,|\, (X_{(2)} = x_{(2)})$ is multinormal, in dimension p, with these important formulae for conditional mean and conditional variance:

$$\mathrm{E}(X_{(1)} \,|\, x_{(2)}) = \xi_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x_{(2)} - \xi_{(2)}),$$

and

$$\mathrm{Var}(X_{(1)} \,|\, x_{(2)}) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

In particular, the conditional mean is a linear function of $x_{(2)}$, and the conditional variance is constant.

(c) For the most simple but still interesting case of a normalised binormal distribution, show that if

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \Big),$$

show that $X_2 \,|\, (X_1 = x_1) \sim \mathrm{N}(\rho x_1, 1 - \rho^2)$. Generalise to the case where X_1, X_2 have means ξ_1, ξ_2 and variances σ_1^2 and σ_2^2.

(d) Verify that P_t^s as defined in (30) is $P_t^s = \mathrm{E},(X_t - \mu_t^s)^2$.

Kalman filter

Exercise 6.3 We now derive the Kalman filter. Let the state space model be as specified in (28) and (29). We are to show that

$$X_t \,|\, \mathcal{D}_{t-1} \sim \mathrm{N}_p(X_t^{t-1}, P_t^{t-1}),$$

and

$$X_t \,|\, \mathcal{D}_t \sim \mathrm{N}_p(X_t^t, P_t^t),$$

with

$$X_{t-1}^t = \Phi X_{t-1}^{t-1}, \quad \text{and} \quad P_{t-1}^t = \Phi P_{t-1}^{t-1}\Phi^{\mathrm{t}} + Q.$$

and

$$X_t^t = X_t^{t-1} + K_t(Y_t - A_t X_t^{t-1}), \quad \text{and} \quad P_t^t = (I_p - K_t A_t)P_t^{t-1},$$

with the Kalman gain

$$K_t = P_t^{t-1} A_t^{\mathrm{t}} (A_t P_t^{t-1} A_t^{\mathrm{t}} + R)^{-1}.$$

(a) To prove the above, notice that $X_0 \,|\, \mathcal{D}_0 \sim \mathrm{N}(X_0^0, P_0^0)$.Make the hypothesis $X_{t-1} \,|\, \mathcal{D}_{t-1} \sim \mathrm{N}_p(X_{t-1}^{t-1}, P_{t-1}^{t-1})$, and show that this implies $X_t \,|\, \mathcal{D}_t \sim \mathrm{N}_p(X_t^t, P_t^t)$. To do so, consider the conditional joint distribution of $(X_t, Y_t) \,|\, \mathcal{D}_{t-1}$.

(b) Notice that $X_t \,|\, \mathcal{D}_{t-1} \sim \mathrm{N}_p(X_t^{t-1}, P_t^{t-1})$ is proven along the way.

Exercise 6.4 Look at the smoothing theorem in Shumway and Stoffer (2025, Property 6.2, p. 322). Prove a version of this theorem where you also get the conditional distributions. [xx some more help here xx].

Exercise 6.5 Simulate data from the univariate AR(1) models with noisy observations studied in Ex. 3.6. Make three plots of these data. In each plot plot the observed Y_t and the underlying X_t (in different colours). For each the first plot, add lines indicating the 95 percent one-step forecasting interval; for the second, add 95 percent filtering intervals; and for the thirds, add 95 percent smoothing intervals. In other words, reproduce a version of Figure 6.4 on p. 323 in Shumway and Stoffer (2025).

Exercise 6.6 Consider the model specified in (28) and (29) with $A_t = A$. Write $\theta = (A, \Phi, R, Q, \mu_0, \Sigma_0)$ for the parameters of this model.

(a) Having observed $Y_1, \ldots, Y_n$, show that the log-likelihood function takes the form (ignoring constants)

$$\ell_n(\theta) = -\tfrac{1}{2}\sum_{t=1}^{n}\big\{\log(A_tP_t^{t-1}A_t^{\mathrm{t}}+R)+(Y_t-A_tX_t^{t-1})^{\mathrm{t}}\Sigma_t(\theta)^{-1}(Y_t-A_tX_t^{t-1})\big\}, \quad (31)$$

where $\Sigma_t(\theta) = A_tP_t^{t-1}A_t^{\mathrm{t}} + R$.

(b) Consider the univariate AR(1) models with noisy observation studied in Ex. 3.6.

7. GARCH models

Let $(X_t)_{t\geq 0}$ be the price of a stock. Then

$$r_t = (X_t - X_{t-1})/X_{t-1}, \quad t \geq 1$$

are the returns on the stock. When observing and plotting some return over time, $r_1, r_2, \ldots, r_n$ say, it is often observed that this process goes through periods of higher volatility, then a period of lower volatility. By volatility we mean the variance or conditional variance (given the past) of the return. Generalised autoregressive conditional heteroskedastic is class of models for this clustering of volatility phenomenon. An ARCH(1) model takes the returns as ARCH(1)

$$r_t = \sigma_t\varepsilon_t,$$

where ε_t are i.i.d. mean zero random variables with unit variance; and the volatility is

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2,$$

for $\alpha_0 > 0$ and $\alpha_1 \geq 0$. An ARCH(q) models takes the volatility as

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{q}\alpha_j r_{t-j}^2,$$

Finally, a GARCH(p, q) models the volatility as GARCH(p, q)

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{q}\alpha_j r_{t-j}^2 + \sum_{j=1}^{p}\beta_j\sigma_{t-j}^2.$$

Exercise 7.1 Download daily stock prices for some company, or just take the Equinor data from the course website, and make a returns data set.

(a) Without thinking too much about models, create an estimator for the weakly volatility, an compute this on your data. This gives you a data set $\widehat{\sigma}_w^2$ for week $w = 1, \ldots, m$. Plot $\widehat{\sigma}_w^2$ against $\widehat{\sigma}_{w-1}^2$, try also plotting $\log\widehat{\sigma}_w^2$ against $\log\widehat{\sigma}_{w-1}^2$. Comment on what you observe.

(b) Run a regression of $\log\widehat{\sigma}_w^2$ on $\log\widehat{\sigma}_{w-1}^2$. Try also higher order lags.

Exercise 7.2 Consider the ARCH(1) model

$$r_t = \sigma_t\varepsilon_t, \quad t \geq 1$$

with $\varepsilon_1, \varepsilon_2, \ldots$ standard normal,

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2, \quad t \geq 1$$

where $\alpha_0 > 0, \alpha_1 \geq 1$, with $r_0 \sim \mathrm{N}(0, \alpha_0)$ independent of the ε_t. Define

$$\mathcal{F}_t = \sigma(r_0, r_1, \ldots, r_t).$$

(a) Go to R and simulate returns data from a ARCH(1) model of the type just specified.

(b) Show that $\mathrm{E}\, r_t = 0$ and that the $\mathrm{E}\,(r_t^2 \,|\, \mathcal{F}_{t-1}) = \sigma_t^2$. This latter property is called conditional heteroskedasticity.

(c) Verify also that we may write

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + v_t, \quad \text{with } v_t = \sigma_t^2(\varepsilon_t^2 - 1),$$

and that $\mathrm{E}\, v_t = 0$ and $\mathrm{E}\,(v_t^2 \,|\, \mathcal{F}_{t-1}) = 2\sigma_t^4$, showing that $(v_t)_{t \geq 1}$ is a square integrable martingale difference sequence.

(d) Since the ε_t are assumed standard normal we can write down a likelihood function for (α_0, α_1). Based on the observing $r_1, \ldots, r_n$, verify that this likelihood is

$$\ell_n(\alpha_0, \alpha_1) = -\tfrac{1}{2} \sum_{t=2}^n \log(\alpha_0 + \alpha_1 r_{t-1}^2) - \tfrac{1}{2} \sum_{t=2}^n \frac{r_t^2}{\alpha_0 + \alpha_1 r_{t-1}^2}.$$

Implement this likelihood function is R, and use it to fit an ARCH(1) model to the simualted data from **(a)**. Tru also to fit an ARCH(1) model to the empirical data from Ex. 7.1.

Exercise 7.3 (xx notes xx) Let $(r_t)_{t \geq 1}$ follow an ARCH(1) model as specified in Ex. 7.2. In this exercise we study $Z_n = n^{-1/2} \sum_{t=1}^n r_t$ and $\mathrm{rv}_n = n^{-1} \sum_{t=1}^n r_t^2$, the latter as an estimator of the 'integrated volatility'. We will assume that $0 < \alpha_1 < 1$.

(a) Verify that

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{j=1}^{t-1} \alpha_1^j \prod_{i=1}^{j} \varepsilon_{t-i}^2 + \alpha_1^t r_0^2 \prod_{j=1}^{t-1} \varepsilon_{t-j}^2.$$

From this, show that $\mathrm{E}\, \sigma_t^2 \to \alpha_0/(1 - \alpha_1)$ as $t \to \infty$.

(b)

(c)

References

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons, New York.

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.

Department of Mathematics, University of Oslo
Email address: emilas@math.uio.no