Statistics of Stochastic Processes
(Statistik stochastischer Prozesse)
Notes for the course
in winter semester 2014/15

If you find any mistakes in the script, please send an e-mail to

shevcher@student.hu-berlin.de

Markus Reiß
Humboldt-Universität zu Berlin

Preliminary version, February 12, 2015

Contents

1 Time series

1.1 Stationary processes

Idea: A process is stationary if its law is invariant with respect to time shifts.

1.1 Examples.

- Annual rainfall,
- EUR-USD-exchange rate,
- car accidents,
- heartbeat of a healthy person.

1.2 Counterexamples.

- Tide level at Hamburg harbour,
- stock price of Siemens since 1960,
- population of ladybirds per year.

Taking out trends/cycles this might still yield stationary time series.

1.3 Definition. Let $T \subseteq \mathbb{R}$ with $t,\, s \in T \Rightarrow t+s \in T$ be a time set, mostly $T \in \{\mathbb{N}_0,\, \mathbb{Z},\, \mathbb{R}_0^+,\, \mathbb{R}\}$. A family $(X_t,\, t \in T)$ of random variables on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is a stochastic process. For $T \in \{\mathbb{N}_0,\, \mathbb{Z}\}$ we call X also time series. X is called (strictly) stationary if

$$\forall n \in \mathbb{N},\, t_1, \dots, t_n,\, t \in T : (X_{t_1}, \dots, X_{t_n}) \overset{\mathrm{d}}{=} (X_{t_1+t}, \dots, X_{t_n+t}),$$

$$\text{i.e. } \forall A \in \mathfrak{B}_{\mathbb{R}^n} : \mathbb{P}((X_{t_1}, \dots, X_{t_n}) \in A) = \mathbb{P}((X_{t_1+t}, \dots, X_{t_n+t}) \in A).$$

If X is in L^2, i.e. $\mathbb{E}[X_t^2] < \infty$ for all $t \in T$, then X is called weakly stationary (second order stationary) if the expectation function $t \mapsto \mu(t) := \mathbb{E}[X_t]$ is constant and the covariance function satisfies $\operatorname{Cov}(X_u,\, X_s) = \operatorname{Cov}(X_{u+t},\, X_{s+t})$ for all $u,\, s,\, t \in T$. In that case $t \mapsto c(t) := \operatorname{Cov}(X_s,\, X_{s+t})$ ($s \in T$ arbitrary) is called autocovariance function.

1.4 Example. If $(X_t)_{t \in T}$ are i.i.d., then X is strictly stationary.

1.5 Lemma. *We have: X is L^2 and strictly stationary $\Rightarrow$ X is weakly stationary.*

Proof. Identity in law and L^2-property imply identity of expectations and covariances. □

Problem 1

(a) Find a weakly stationary process that is not strictly stationary.

(b) Prove that for a Gaussian process both notions of stationarity are equivalent.

First statistical problem: Let X be a weakly stationary time series with expectation $\mu = \mathbb{E}[X_t]$. Estimate μ from observations $X_1, \ldots, X_n$.
A natural approach is the empirical mean

$$\hat{\mu}_n := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Note that $\hat{\mu}_n$ is a measurable function of the observations $(X_1, \ldots, X_n)$ and as such a random variable. We call $\hat{\mu}_n$ an estimator. For realisations $x_1, \ldots, x_n$ of $(X_1, \ldots, X_n)$, i.e. $x_k = X_k(\omega_0)$ for some $\omega_0 \in \Omega$, the value (real number) $\hat{\mu}_n(\omega_0) = \frac{1}{n}\sum_{i=1}^{n} x_i$ is called estimated value. Here, we see that $\hat{\mu}_n$ is an unbiased (erwartungstreu) estimator of μ:

$$\mathbb{E}[\hat{\mu}_n] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] \overset{\text{station.}}{=} \mu.$$

1.6 Examples.

(a) If $c(t) = 0$ for $t \neq 0$ (X_t and X_s are uncorrelated for $t \neq s$), then by the weak law of large numbers (LLN) $\hat{\mu}_n \to \mu$ in probability as $n \to \infty$.

(b) Take some $Y \in L^2$ and set $X_i := Y$ for all $i \in \mathbb{N}_0$. Then $(X_i)_{i\in\mathbb{N}_0}$ is weakly stationary ($\mu = \mathbb{E}[Y]$, $c(t) = \operatorname{Cov}(X_i, X_{i+t}) = \operatorname{Var}(Y)$). We see immediately that $\hat{\mu}_n = Y$ does not converge (in probability) to μ, unless $\mathbb{P}(Y = \mu) = 1$.

1.7 Proposition. *If $(X_t,\, t \in \mathbb{Z})$ is weakly stationary with autocovariance function c and mean μ, then we have for $\hat{\mu}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$:*

(a) $\operatorname{Var}(\hat{\mu}_n) \to 0$ *if* $\lim\limits_{n\to\infty} c(n) = 0$, *in particular* $\hat{\mu}_n \to \mu$ *in probability and in* L^2*;*

(b) $n \operatorname{Var}(\hat{\mu}_n) \to \sum\limits_{k=-\infty}^{\infty} c(k)$ *if* $\sum\limits_{k=-\infty}^{\infty} |c(k)| < \infty$.

Proof. (a)

$$\lim_{n\to\infty} c(n) = 0 \Rightarrow \operatorname{Var}(\hat{\mu}_n) = \frac{1}{n^2}\sum_{i,j=1}^{n} \overbrace{\operatorname{Cov}(X_i, X_j)}^{c(i-j)} = \sum_{k=-(n-1)}^{n-1} \frac{n-|k|}{n^2} c(k)$$

$$\leq \frac{1}{n}\sum_{k=-(n-1)}^{n-1} |c(k)| = \frac{2n-1}{n}\left(\frac{1}{2n-1}\sum_{k=-(n-1)}^{n-1} |c(k)|\right) \overset{\text{Césaro mean}}{\longrightarrow} 0.$$

$$\mathbb{E}[(\hat{\mu}_n - \mu)^2] \overset{\hat{\mu}_n \text{unbiased}}{=} \operatorname{Var}(\hat{\mu}_n) \to 0 \iff \hat{\mu}_n \overset{L^2}{\to} \mu \Rightarrow \hat{\mu}_n \overset{\mathbb{P}}{\to} \mu.$$

(b)

$$\sum_{k\in\mathbb{Z}} |c(k)| < \infty \Rightarrow \sup_n (n\operatorname{Var}(\hat{\mu}_n)) \overset{\text{by (a)}}{\leq} \sup_n \sum_{k=-(n-1)}^{n-1} |c(k)| < \infty.$$

Dominated convergence theorem (DCT):

$$\lim_{n\to\infty} (n\operatorname{Var}(\hat{\mu}_n)) \overset{\text{by (a)}}{=} \lim_{n\to\infty} \sum_{k=-(n-1)}^{n-1} \underbrace{\left(1-\frac{|k|}{n}\right)}_{\to 1} c(k) = \sum_{k\in\mathbb{Z}} c(k).$$

□

1.8 Remarks. Part (a) shows in particular that $\hat{\mu}_n$ is a consistent estimator: $\hat{\mu}_n \overset{\mathbb{P}}{\to} \mu$. Part (b) shows that the rate of convergence is $\frac{1}{\sqrt{n}}$: $\sqrt{n}(\hat{\mu}_n - \mu)$ is bounded in L^2 (and then also in probability).
If $\sum_{k\in\mathbb{Z}} |c(k)|$ is finite, the time series is said to have short range dependence, otherwise it is called long range dependent.

Question: Do we even have $\hat{\mu}_n \overset{\text{a.s.}}{\to} \mu$? What if X is strictly stationary, but $X_t \in L^1 \backslash L^2$? (cf. strong LLN)

Tool: Birkhoff's ergodic theorem (T left shift on sequence space, J T-ivariant σ-algebra):

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X \circ T^i \overset{\text{a.s.}, L^1}{\longrightarrow} \mathbb{E}[X|J].$$

If T (respectively (X_t)) is ergodic, i.e. J is trivial, then $\mathbb{E}[X|J] \overset{\text{a.s.}}{=} \mathbb{E}[X] = \mu$.

Problem 2: Let $(X_n, n \in \mathbb{N}_0)$ be a strictly stationary process. Construct another strictly stationary process $(\tilde{X}_m, m \in \mathbb{Z})$ such that $(\tilde{X}_{m+n}, n \in \mathbb{N}_0) \overset{\text{d}}{=} (X_n, n \in \mathbb{N}_0)$ for all $m \in \mathbb{Z}$. $\tilde{X}$ is the canonical extension of X from $\mathbb{N}_0$ to $\mathbb{Z}$.

Problem 3: Consider a weakly stationary process $(X_t, t \in \mathbb{R})$ such that $(t, \omega) \mapsto X_t(\omega)$ is $\mathfrak{B}_{\mathbb{R}} \otimes \mathscr{F}$-measurable (i.e. X is a measurable process). Construct an estimator $\hat{\mu}_T$ of $\mu = \mathbb{E}[X_t]$ based on observing $(X_t, t \in [0,T])$ (analogous to $\hat{\mu}_n$). Study its mean and asymptotic variance under suitable conditions for c.

For statistical inference, e.g. confidence intervals, an (asymptotic) distribution of $\sqrt{n}(\hat{\mu}_n - \mu)$ in the previous proposition would be desirable.

Conjecture: $\sqrt{n}(\hat{\mu}_n - \mu) \to \mathrm{N}(0, \sum_{k\in\mathbb{Z}} c(k))$ under suitable conditions.

Even if we had such a result, a priori we do not know the asymptotic variance $\sum_{k\in\mathbb{Z}} c(k)$ and we need to estimate it. Alternative approach is a resampling/ bootstrap approach.

1.9 Lemma. *The autocovariance function $c : \mathbb{Z} \to \mathbb{R}$ of a weakly stationary process $(X_t,\, t \in \mathbb{Z})$ satisfies:*

(a) c is symmetric: $c(-k) = c(k)$, $k \in \mathbb{Z}$,

(b) $c(0) \geq 0$ and $|c(k)| \leq c(0)$,

(c) c is positive semi-definite:

$$\forall m \in \mathbb{N},\, a_1, \dots, a_m \in \mathbb{R} : \sum_{i,j=1}^{m} a_i a_j c(i-j) \geq 0.$$

Proof. (a) $\operatorname{Cov}(X_s,\, X_t) = \operatorname{Cov}(X_t,\, X_s)$,

(b) $c(0) = \operatorname{Var}(X_t) \geq 0$,
$c(k)^2 = \operatorname{Cov}(X_k,\, X_0)^2 \overset{\text{Cauchy-Schwarz}}{\leq} \operatorname{Var}(X_k) \operatorname{Var}(X_0) \overset{\text{station.}}{=} c(0)^2$,

(c) $\sum_{i,j=1}^{m} a_i a_j c(i-j) = \operatorname{Var}(\sum_{i=1}^{m} a_i X_i) \geq 0$.

□

1.10 Definition. The 'canonical' estimator $\hat{c}(k)$ of the autocovariance function at lag k from observing $X_1, \dots, X_n$, $n \geq k$, is given by

$$\hat{c}(k) = \frac{1}{n} \sum_{l=1}^{n-k} (X_l - \hat{\mu}_n)(X_{l+k} - \hat{\mu}_n).$$

Set $\hat{c}(-k) := \hat{c}(k)$. The empirical autocovariance matrix is then

$$\hat{C}_n := \begin{pmatrix} \hat{c}(0) & \hat{c}(1) & \dots & \hat{c}(n-1) \\ \hat{c}(1) & \hat{c}(0) & \dots & \hat{c}(n-2) \\ \vdots & \ddots & \ddots & \vdots \\ \hat{c}(n-1) & \dots & \hat{c}(1) & \hat{c}(0) \end{pmatrix}.$$

Problem 4:

(a) Verify the bias-variance decomposition for an estimator $\hat{\vartheta}$ of $\vartheta \in \mathbb{R}$ with $\mathbb{E}[\hat{\vartheta}^2] < \infty$:

$$\mathbb{E}[(\hat{\vartheta} - \vartheta)^2] = \underbrace{(\mathbb{E}[\hat{\vartheta}] - \vartheta)^2}_{\text{Bias}^2} + \operatorname{Var}(\hat{\vartheta}).$$

(b) Let $Y_1, \dots, Y_n \overset{\text{i.i.d}}{\sim} \mathrm{N}(\mu,\, \sigma^2)$ and $\hat{\sigma}^2_\alpha = \frac{\alpha}{n-1} \sum_{i=1}^{n} (Y_i - \hat{\mu}_n)^2$, $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$, $\alpha > 0$. Show that $\hat{\sigma}^2_\alpha$ is unbiased iff $\alpha = 1$ and determine $\alpha = \alpha_{\text{opt}} > 0$ such that $\mathbb{E}[(\hat{\sigma}^2_\alpha - \sigma^2)^2]$ is minimal. How would you choose α in practice?

1.11 Lemma. *$\hat{C}_n$ (or $\hat{c}$ on $\{-n+1, \dots, n-1\}$) is positive semi-definite:*

$$\forall a_1, \dots, a_n \in \mathbb{R} : \sum_{i,j=1}^{n} a_i a_j \hat{c}(i-j) \geq 0.$$

1.12 Remark. For this it is essential that the prefactor before the sum in $\hat{c}(k)$ does not depend on k.

Proof. Set $Y_i = (X_i - \hat{\mu}_n)\mathbf{1}_{(1\le i\le n)}$, $i \in \mathbb{Z}$.

$$\begin{aligned}\sum_{i,j=1}^{n} a_i a_j \hat{c}(i-j) &= \frac{1}{n}\sum_{i,j=1}^{n} a_i a_j \sum_{l\in\mathbb{Z}} Y_l Y_{l+|i-j|} \\ &= \frac{1}{n}\sum_{l\in\mathbb{Z}}\sum_{i,j=1}^{n} a_i a_j Y_l Y_{l+|i-j|} = \frac{1}{n}\sum_{l'\in\mathbb{Z}}\sum_{i,j=1}^{n} a_i a_j Y_{l'-i} Y_{l'-j} \\ &= \frac{1}{n}\sum_{l'\in\mathbb{Z}}\Big(\sum_{i=1}^{n} a_i Y_{l'-i}\Big)^2 \ge 0.\end{aligned}$$

□

1.13 Example. If X is Gaussian and $\mu = 0$ is known (i.e. $\hat{\mu}_n = \mu = 0$), then $\mathbb{E}[\hat{c}(k)] = \frac{n-k}{n}c(k)$, $n\,\mathrm{Var}(\hat{c}(k)) \to \sum_{l\in\mathbb{Z}}(c(l)^2 + c(l+k)c(l-k))$ if $(c(l))_{l\in\mathbb{Z}} \in \ell^2$ (see class notes $\rightsquigarrow$ products of four Gaussian random variables). $\rightsquigarrow$ $\hat{c}(k)$ has convergence rate $\frac{1}{\sqrt{n}}$ as well (for k fixed).

1.2 Autoregressive and moving average processes

1.14 Definition. A weakly stationary process $(\varepsilon_t,\, t \in \mathbb{Z})$ with mean 0 and autocovariance function $c(t) = \begin{cases}\sigma^2, & t = 0,\\ 0, & t \ne 0.\end{cases}$ is called white noise, $\varepsilon_t \sim \mathrm{WN}(0,\sigma^2)$. If (ε_t) is even i.i.d. and $(\varepsilon_t) \sim \mathrm{WN}(0,\sigma^2)$ we write $(\varepsilon_t) \sim \mathrm{IID}(0,\sigma^2)$.

Consider discrete dynamical systems (with initial values x_0, X_0):

- $x_t = a x_{t-1}$, $t \in \mathbb{N}$ $\rightsquigarrow$ $x_t = a^t x_0$.
  Asymptotics for large t : $\begin{cases} a > 1: & x_t \to \infty,\\ a = 1: & x_t = x_0,\\ 0 < a < 1: & x_t \to 0,\\ a < 0: & \text{similar cases.}\end{cases}$
- $X_t = aX_{t-1} + \varepsilon_t$, $t \in \mathbb{N}$.
  We obtain: $X_t = a^t X_0 + \sum_{i=0}^{t-1} a^i \varepsilon_{t-i}$,
  $\mathbb{E}[X_t] = a^t\,\mathbb{E}[X_0]$ ($\rightsquigarrow$ deterministic dynamics),

$$\mathrm{Cov}(X_t, X_s) \overset{\text{assume } t\ge s}{=} \mathrm{Cov}(a^{t-s}X_s + \sum_{i=0}^{t-s-1} a^i \varepsilon_{t-i}, X_s)$$

$$= a^{t-s}\,\mathrm{Var}(X_s) + \sum_{i=0}^{t-s-1} a^i\,\mathrm{Cov}(\varepsilon_{t-i}, X_s) \overset{\text{supp. } \forall t:\, \mathrm{Cov}(X_0,\varepsilon_t)=0}{=} a^{t-s}\,\mathrm{Var}(X_s).$$

Moreover,

$$\operatorname{Var}(X_s) = a^{2s}\operatorname{Var}(X_0) + \sigma^2 \sum_{i=0}^{s-1} a^{2i} \overset{a\neq\pm 1}{=} a^{2s}\operatorname{Var}(X_0) + \sigma^2 \frac{a^{2s}-1}{a^2-1}.$$

Asymptotics:

I $|a| > 1$: If $\mathbb{E}[X_0] > 0$, then $\mathbb{E}[X_t] \to +\infty$ or $-\infty$ for $a > 1$, $a < -1$ geometrically fast; $\operatorname{Var}(X_t) \to \infty$ holds as well. After normalisation, however, we have that $\mathbb{E}[\frac{X_t}{a^t}]$, $\operatorname{Var}\left(\frac{X_t}{a^t}\right)$ remain bounded (but usually do not tend to zero) $\rightsquigarrow$ unstable behaviour.

II $a = \pm 1$: $a = 1$: random walk, usually $\limsup_{t\to\infty} X_t = +\infty$, $\liminf_{t\to\infty} X_t = -\infty$. $a = -1$: alternating random walk-type process with similar asymptotic properties.

III $|a| < 1$: $\mathbb{E}[X_t] \to 0$, $\operatorname{Var}(X_t) \to \frac{\sigma^2}{1-a^2}$ (independent of X_0).

Correlation for $|a| < 1$:

$$\operatorname{Corr}(X_t, X_s) \overset{t\geq s}{=} \frac{a^{t-s}\operatorname{Var}(X_s)}{\sqrt{\operatorname{Var}(X_s)\operatorname{Var}(X_t)}} \overset{\text{for large } t,s}{\approx} a^{t-s}.$$

More precisely: $\lim_{s\to\infty}\operatorname{Corr}(X_{s+m}, X_s) = a^m$. This means that for large m X_s and X_{s+m} are nearly uncorrelated. The time series 'forgets the initial condition' as $t \to \infty$.

1.15 Definition. For white noise $(\varepsilon_t) \sim \mathrm{WN}(0, \sigma^2)$, $p, q \in \mathbb{N}$; $\varphi_1, \dots, \varphi_p$, $\vartheta_1, \dots, \vartheta_q \in \mathbb{R}$ and random variables $X_0, \dots, X_{-p+1}$ which are uncorrelated to (ε_t)

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t + \vartheta_1\varepsilon_{t-1} + \dots + \vartheta_q\varepsilon_{t-q},\ t \in \mathbb{N}$$

defines an autoregressive-moving average process, ARMA(p, q)-process for short.
With polynomials $\varphi(z) := 1 - \varphi_1 z - \dots - \varphi_p z^p$, $\vartheta(z) := 1 + \vartheta_1 z + \dots + \vartheta_q z^q$ and the backward shift operator $BX_t := X_{t-1}$ ($B^2X_t = X_{t-2}$, $B^0X_t = X_t$ etc.) we obtain more concisely $\varphi(B)X_t \overset{(*)}{=} \vartheta(B)\varepsilon_t$, $t \in \mathbb{N}$.
Any process $(X_t, t \in \mathbb{Z})$ solving $(*)$ is called an ARMA(p, q)-process on $\mathbb{Z}$.
If $\vartheta(z) = 1$, then X is called autoregressive process or AR(p)-process. If $\varphi(z) = 1$, then X is called moving average process or MA(q)-process.

Problem 5: Consider the deterministic dynamics for $x_t \in \mathbb{C}$ with $\varphi(B)x_t = 0$. Show that $x_t = a^t$ is a solution (for suitable initial values) if a^{-1} is a zero of φ. Conclude that in the case where φ has p distinct zeroes, any solution can be written as $x_t = \sum_{j=1}^p c_j a_j^t$ with $c_1, \dots, c_p \in \mathbb{C}$ and $a_1^{-1}, \dots, a_p^{-1}$ zeroes of φ. What happens in the case of multiple zeroes?

Problem 6:

(a) Let $x_t(x_0, \dots, x_{-p+1})$ be the solution of $\varphi(B)x_t = 0$, $t \geq 1$, with initial values $x_0, \dots x_{-p+1}$. Prove that the AR(p)-process X satisfies the <u>variation of constants formula</u>

$$X_t = x_t(X_0, \dots, X_{-p+1}) + \sum_{j=1}^{t} \underbrace{x_{t-j}(1, 0, 0, \dots, 0)}_{\text{'fundamental solution'}} \varepsilon_j.$$

(b) Determine the solution and its expectation as well as its covariance function explicitly for the stochastic Fibonacci dynamics:

$$X_t = X_{t-1} + X_{t-2} + \varepsilon_t,\ X_0 = X_{-1} = 1.$$

(c) Give an example of an AR(2)-process that admits a weakly stationary solution.

1.16 Lemma. *The AR(1)-process on $\mathbb{Z}$ $(X_t,\ t \in \mathbb{Z})$ $X_t = aX_{t-1} + \varepsilon_t$, $t \in \mathbb{Z}$, has a weakly stationary solution if $|a| \neq 1$, For $a \in (-1, 1)$ this solution has the representation $X_t = \sum_{i=0}^{\infty} a^i \varepsilon_{t-i}$, for $|a| > 1$ it has the representation $X_t = -\sum_{i=1}^{\infty} a^{-i} \varepsilon_{t+i}$.*

Proof. The case $|a| < 1$ follows immediately from the formulas above when inserting $X_0 = \sum_{i=0}^{\infty} a^i \varepsilon_{-i}$, cf. also the more general example from the class. The case $|a| > 1$: note that $\sum_{i=1}^{\infty} a^{-i} \varepsilon_{t+i}$ is well-defined as a limit in L^2 since $\sum_{i \geq 1} a^{-2i} < \infty$. We then have $aX_{t-1} = -\sum_{i=1}^{\infty} a^{1-i} \varepsilon_{t-1+i} = -\varepsilon_t + X_t$
$\Rightarrow X$ is AR(1)-process. Weak stationarity is checked by calculating expectation, covariance function as for $|a| < 1$. □

1.17 Definition. A weakly stationary ARMA(p, q)-process is called <u>causal</u> if there is $(\psi_i) \in \ell^1$ such that $X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$, $t \in \mathbb{Z}$. The latter is called an <u>infinite moving average representation</u> (or <u>MA(∞)</u>).

1.18 Remarks.

(a) For the AR(1)-process above X is causal if $|a| < 1$ and not causal for $|a| > 1$.

(b) Compare with the concept of adaptedness for stochastic processes.

Problem 7: Show that there is a weakly stationary solution of an MA(q)-process. Discuss its expectation and autocovariance functions and simulate some examples.

We are now prepared for the main theorem on causal ARMA(p, q)-processes.

First, we need some basic power series calculus for the backward shift operator B.

1.19 Lemma. *If $(X_t,\, t\in\mathbb{Z})$ is a process bounded in L^1 (i.e. $\sup_t \mathbb{E}[|X_t|]<\infty$) and $(a_j)_{j\in\mathbb{Z}}$ in ℓ^1 then the series*

$$a(B)X_t=\sum_{j\in\mathbb{Z}} a_j B^j X_t=\sum_{j\in\mathbb{Z}} a_j X_{t-j}$$

converges absolutely with probability one (=a.s.). If X is bounded in L^2, then the series is bounded in L^2 and converges in L^2 to the same limit.

Proof. By Tonelli theorem:

$$\mathbb{E}[\sum_{j\in\mathbb{Z}}|a_j||X_{t-j}|]=\sum_{j\in\mathbb{Z}}|a_j|\,\mathbb{E}[|X_{t-j}|]\le\|(a_j)\|_{\ell^1}\sup_t\mathbb{E}[|X_t|]<\infty.$$

It follows that $\mathbb{P}(\sum_{j\in\mathbb{Z}}|a_j||X_{t-j}|<\infty)=1$ and the series converges a.s. absolutely.
If X is L^2-bounded, then for $n>m>0$

$$\begin{aligned}\mathbb{E}[\Big(\sum_{m<|j|\le n} a_jX_{t-j}\Big)^2]&=\sum_{m<|j|,|k|\le n} a_ja_k\quad\underbrace{\mathbb{E}[X_{t-j}X_{t-k}]}_{\overset{\text{C.-S.}}{\le}(\mathbb{E}[X_{t-j}^2]\,\mathbb{E}[X_{t-k}^2])^{1/2}}\\ &\le\underbrace{\Big(\sum_{m<|j|\le n}|a_j|\Big)^2}_{(a_j)\in\ell^1}\underbrace{\sup_t\mathbb{E}[X_t^2]}_{<\infty}\overset{m,n\to\infty}{\longrightarrow}0\end{aligned}$$

Hence, the sum forms a Cauchy sequence in L^2 and thus converges in L^2, which must be the same limit. □

1.20 Lemma. *If X is weakly stationary with autocovariance function c_X and if $(a_j)\in\ell^1$, then $Y_t=a(B)X_t=\sum_{j\in\mathbb{Z}}a_jX_{t-j}$, $t\in\mathbb{Z}$, is again weakly stationary with autocovariance function*

$$c_Y(t)=\sum_{j,k\in\mathbb{Z}}a_ja_kc_X(t-j+k).$$

Proof. Y is well-defined by the preceding lemma noting

$$\mathbb{E}[X_t^2]=\mathbb{E}[X_t]^2+\mathrm{Var}(X_t)=\mu_X^2+c_X(0)<\infty.$$

Hence,

$$\begin{aligned}\mathbb{E}[Y_t]&\overset{L^2\text{-conv.}}{=}\lim_{n\to\infty}\mathbb{E}[\sum_{j=-n}^{n}a_jX_{t-j}]=\lim_{n\to\infty}\sum_{j=-n}^{n}a_j\mu_X\\ &=\mu_X\sum_{j=-\infty}^{\infty}a_j=:\mu_Y\text{ (independent of }t),\end{aligned}$$

$$\begin{aligned}\mathbb{E}[Y_tY_s]&\overset{L^2\text{-conv.}}{=}\lim_{n\to\infty}\mathbb{E}[(\sum_{j=-n}^{n}a_jX_{t-j})(\sum_{k=-n}^{n}a_kX_{s-k})]\\ &=\lim_{n\to\infty}\sum_{-n\le j,k\le n}a_ja_k\underbrace{\mathbb{E}[X_{t-j}X_{s-k}]}_{c_X(t-j-s+k)+\mu_X^2}=\Big(\sum_{j,k\in\mathbb{Z}}a_ja_kc_X(t-s-j+k)\Big)+\mu_Y^2.\end{aligned}$$

It is finite:

$$\sum_{j,k\in\mathbb{Z}} \Big| a_j a_k c_X(t-s-j+k) \Big| \le c_X(0)\|a\|_{\ell^1}^2 < \infty$$

and depends on $(t,\, s)$ only via $(t-s)$.
Consequently, Y is weakly stationary and c_Y is as asserted. □

1.21 Remark. The lemma justifies the formal convolution algebra calculations for (a_j), $(b_j) \in \ell^1$:

$$a(B)b(B)X_t = c(B)X_t$$

with $c(z) = \sum_{j=0}^{\infty} c_j z^j$, $c_j = \sum_{k\in\mathbb{Z}} a_k b_{j-k}$ $(c = a * b = b * a)$ for X L^2-bounded.

1.22 Theorem. *Let X be a weakly stationary ARMA$(p,\, q)$-process on $\mathbb{Z}$ with no common zeroes of φ and ϑ on $\{z \in \mathbb{C}\,||z| \le 1\}$. Then X is causal if and only if $\varphi(z) \ne 0$ for $z \in \mathbb{C}$ with $|z| \le 1$. In that case $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ holds where $\psi(z) := \sum_{j=0}^{\infty} \psi_j z^j = \frac{\vartheta(z)}{\varphi(z)}$ for $|z| \le 1$. In particular, such a process X is unique.*

1.23 Remark. Note that $\varphi(z) \ne 0$ for $z \in \mathbb{C}$ with $|z| \le 1$ implies that all solutions of the deterministic equation $\varphi(B)x_t = 0$ are asymptotically stable, i.e. $\lim_{t\to\infty} x_t = 0$ (use Problem 5).

1.24 Corollary. *Suppose $\varphi(z) \ne 0$ for $z \in \mathbb{C}$ with $|z| \le 1$ and define (for white noise $(\varepsilon_j)_{j\in\mathbb{Z}} \sim \mathrm{WN}(0, \sigma^2)$) $X_k := \sum_{j=0}^{\infty} \psi_j \varepsilon_{k-j}$ for $k = 0, \ldots, -p+1$ and with $\psi(z) = \frac{\vartheta(z)}{\varphi(z)}$. Then the ARMA$(p,\, q)$-process $\varphi(B)X_t = \vartheta(B)\varepsilon_t$, $t \ge 1$, with initial values $X_0, \ldots, X_{-p+1}$ is weakly stationary on $\mathbb{N}$ (or $\mathbb{N} \cup \{0, \ldots, -p+1\}$) with $\mu = 0$, $c(t) = \sum_{j=0}^{\infty} \psi_j \psi_{t+j}$.*

1.25 Remark. Often, e.g. in the Gaussian case, $X_0, X_{-1}, \ldots, X_{-p+1}$ can be constructed explicitly without simulating all $(\varepsilon_j)_{j\le 0}$.

Proof of Corollary. Clear from Theorem. □

Proof of Theorem.

'⇐' Suppose $\varphi(z) \ne 0$ for $|z| \le 1$. Since φ has only finitely many zeroes, there is an $\varepsilon > 0$ such that $\frac{1}{\varphi(z)} = \sum_{j=0}^{\infty} \xi_j z^j = \xi(z)$ holds for $|z| \le 1 + \varepsilon$ ($\frac{1}{\varphi}$ is holomorphic there).
This implies $\sum_{j=0}^{\infty} |\xi_j|(1 + \frac{\varepsilon}{2})^j < \infty \Rightarrow (\xi_j) \in \ell^1$.
By the previous lemma,

$$X_t = \underbrace{(\xi\varphi)}_{=1}(B)X_t = \xi(B)(\vartheta(B)\varepsilon_t) = \psi(B)\varepsilon_t$$

with $\psi(z) = \xi(z)\vartheta(z) = \frac{\vartheta(z)}{\varphi(z)}$ for $|z| \le 1$.
$\overset{(\varepsilon_t)\text{ weakly stat.}}{\Longrightarrow}$ X is causal since ψ is holomorphic, $(\psi_j) \in \ell^1$.

'⇒' Suppose X is causal, $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ for some $(\psi_j) \in \ell^1$. Then

$$\vartheta(B)\varepsilon_t = \varphi(B)X_t = \varphi(B)\psi(B)\varepsilon_t.$$

Since $(\varepsilon_t) \sim \mathrm{WN}(0, \sigma^2)$, we have for $s \leq t$

$$\mathbb{E}[\underbrace{(\vartheta(B)\varepsilon_t)}_{=\sum \vartheta_k \varepsilon_{t-k}} \varepsilon_s] = \sigma^2 \vartheta_{t-s}, \ \ \mathbb{E}[(\varphi\psi)(B)\varepsilon_t\varepsilon_s] = \sigma^2 a_{t-s}$$

for $a(z) = (\varphi\psi)(z) = \sum a_j z^j$.
$\overset{\sigma \neq 0}{\Rightarrow} \vartheta_{t-s} = a_{t-s} \Rightarrow \vartheta(z) = a(z) = \varphi(z)\psi(z)$, $|z| \leq 1$.
Since ϑ and φ do not have common zeroes on the unit disk, we cannot have $\varphi(z) = 0$ for some $|z| \leq 1$ (otherwise $\vartheta(z) = 0$ follows by finiteness of ψ on unit disk).

□

Statistical problem: Prediction/Forecasting
Focus on AR(p)-process $X_{t+1} = \varphi_1 X_t + \cdots + \varphi_p X_{t-p+1} + \varepsilon_{t+1}$ ($t \in \mathbb{Z}$) and observations $X_0, \ldots, X_t$ ($t \geq p$).

$$\hat{X}_{t+1} = \varphi_1 X_t + \cdots + \varphi_p X_{t-p+1} + \underbrace{\mathbb{E}[\varepsilon_{t+1}]}_{=0}$$

is the best *linear* predictor of X_{t+1} based on $X_0, \ldots, X_t$:
$\mathbb{E}[(\hat{X}_{t+1} - X_{t+1})^2 | X_0, \ldots, X_t]$ is minimal for this choice (it equals σ^2).
Best nonlinear predictor (in general):

$$\hat{X}_{t+1} = \mathbb{E}[X_{t+1} | X_0, \ldots, X_t].$$

They coincide if $(\varepsilon_t) \sim \mathrm{IID}(0, \sigma^2)$ (and $X_0, \ldots, X_{-p+1}$ independent of $(\varepsilon_t)_{t \geq 0}$).
In practice, we have to estimate $\varphi_1, \ldots, \varphi_p$.

Problem 8: See class notes.

Problem 9:

(a) Prove the optimality of $\hat{X}_{t+1}$ formally.

(b) What is the optimal k-step linear predictor $\hat{X}_{t+k}$?

(c) Show that $\hat{X}_{t+1}$ is also the best linear predictor of X_{t+1} based on $X_t, \ldots, X_{t-p+1}$ for any weakly stationary process (not necessarily AR(p)) when $\varphi_1, \ldots, \varphi_p$ solve $C_p \varphi = c_p$ (see notation below).

1.3 The Yule-Walker estimator and a CLT for martingale differences

Here we focus on causal (weakly stationary) AR(p)-processes on $\mathbb{Z}$ with

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t, \ t \in \mathbb{Z}, \ (\varepsilon_t) \sim \mathrm{WN}(0, \sigma^2).$$

Ansatz: Moment estimation method

1st moments: X has zero mean $\rightsquigarrow$ no information on φ_k.

2nd moments: X has autocovariance function

$$\begin{aligned} c(k) &= \mathrm{Cov}(X_t,\, X_{t-k}) = \mathrm{Cov}(\varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t, X_{t-k}) \\ &= \varphi_1 c(k-1) + \cdots + \varphi_p c(k-p) \text{ for } k \geq 1 \text{ and} \end{aligned}$$

$$\begin{aligned} c(0) &= \mathrm{Cov}(X_t,\, X_t) = \mathrm{Cov}(\varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t, X_t) \\ &= \varphi_1 c(-1) + \cdots + \varphi_p c(-p) + \sigma^2 \end{aligned}$$

Hence, the autocovariance function satisfies a linear recurrence equation and is uniquely determined by its initial values $c(0), \ldots, c(p-1)$, σ^2, given $\varphi_1, \ldots, \varphi_p$.

We can identify $\varphi_1, \ldots, \varphi_p$ from p recurrence equations: (use $c(-k) = c(k)$)

$$\left.\begin{aligned} &c(1) = \varphi_1 c(0) + \cdots + \varphi_p c(p-1) \\ &\vdots \\ &c(p) = \varphi_1 c(p-1) + \cdots + \varphi_p c(0) \end{aligned}\right\} \Rightarrow c_p = C_p \varphi$$

with $c_p = (c(1), \ldots, c(p))^T$, $C_p = (c(i-j))_{1 \leq i,\, j \leq p}$, $\varphi = (\varphi_1, \ldots, \varphi_p)^T$.
If $C_p \in \mathbb{R}^{p \times p}$ is positive definite (i.e. non-singular), then φ can be identified from C_p, c_p: $\varphi = C_p^{-1} c_p$.

Empirical version: Define $\hat{\varphi} = (\hat{\varphi}_1, \ldots, \hat{\varphi}_p)^T$ via $\hat{C}_p \hat{\varphi} = \hat{c}_p$ with empirical autocovariance $\hat{c}(k) = \frac{1}{n} \sum_{i=1}^{n-k} X_i X_{i+k}$ (knowing that $\mathbb{E}[X_t] = 0$).

1.26 Definition. This $\hat{\varphi}$ is called Yule-Walker estimator.

What about σ^2?
The recurrence for $k = 0$ yields $\sigma^2 = c(0) - \langle \varphi,\, c_p \rangle_{\mathbb{R}^p}$
$\rightsquigarrow$ standard estimator: $\hat{\sigma}^2 = \hat{c}(0) - \langle \hat{\varphi},\, \hat{c}_p \rangle_{\mathbb{R}^p}$.

1.27 Example (AR(1))**.**

$$\begin{aligned} \hat{\varphi}_1 = \hat{C}_1^{-1} c_1 = \frac{\sum_{i=1}^{n-1} X_i X_{i+1}}{\sum_{i=1}^{n} X_i^2} &\overset{X \text{ is AR}(1)}{=} \frac{\sum_{i=1}^{n-1} X_i (\varphi_1 X_i + \varepsilon_{i+1})}{\sum_{i=1}^{n} X_i^2} \\ = \varphi_1 \frac{\sum_{i=1}^{n-1} X_i^2}{\sum_{i=1}^{n} X_i^2} + \frac{\sum_{i=1}^{n-1} X_i \varepsilon_{i+1}}{\sum_{i=1}^{n} X_i^2}. \end{aligned}$$

Look at $\varphi_1^* \approx \hat{\varphi}_1$:

$$\varphi_1^* = \frac{\sum_{i=1}^{n-1} X_i X_{i+1}}{\sum_{i=1}^{n-1} X_i^2} = \varphi_1 + \frac{\sum_{i=1}^{n-1} X_i \varepsilon_{i+1}}{\sum_{i=1}^{n-1} X_i^2}.$$

If $(\varepsilon_i) \sim \mathrm{IID}(0, \sigma^2)$ and X causal ($\rightsquigarrow$ ε_{i+1} independent of X_i, $X_{i-1}, \ldots$, ε_i, $\varepsilon_{i-1}, \ldots$),

$$\varphi_1^* = \varphi_1 + \frac{M_n}{\sigma^{-2} \langle M \rangle_n},$$

where $M_n = \sum_{i=2}^n X_{i-1}\varepsilon_i$, $n \geq 2$, is an L^2-martingale w.r.t. $\mathscr{F}_n = \sigma(\varepsilon_k,\, k \leq n)$ (causality: X_k is $\mathscr{F}_k$-measurable) and $\langle M\rangle_n = \sum_{i=2}^n \mathbb{E}[(M_i - M_{i-1})^2|\mathscr{F}_{i-1}]$ where $M_0 = M_1 = 0$.

In Stochastics II: If $\langle M\rangle_n \to \infty$ a.s., then $\frac{M_n}{\langle M\rangle_n^\alpha} \stackrel{\text{a.s.}}{\to} 0$ for L^2-martingales (M_n) with $\mathbb{E}[M_n] = 0$ and $\alpha > \frac{1}{2}$.

We want to prove:

1.28 Theorem. *Let X be a causal (weakly stationary) AR(p)-process with $(\varepsilon_t) \sim \mathrm{IID}(0, \sigma^2)$. Then the Yule-Walker estimator $\hat{\varphi}^{(n)}$ satisfies*

$$\sqrt{n}(\hat{\varphi}^{(n)} - \varphi) \stackrel{d}{\to} \mathrm{N}(0,\, \sigma^2 C_p^{-1}),$$

$C_p = (c(i-j))_{i,\,j=1,\dots,p}$.

1.29 Remark (CLT for Yule-Walker)**.** If the order p is not known and we estimate, assuming an AR(m)-process with $m > p$, then the coefficients $\hat{\varphi}_k^{(n)}$, $k = p+1, \dots, m$, of $\hat{\varphi}^{(n)}$ satisfy each $\sqrt{n}\hat{\varphi}_k^{(n)} \to \mathrm{N}(0,\, \sigma^2)$ and we can provide an asymptotic level-α test for the hypothesis H_0 that $\varphi_k = 0$ (using $\hat{\sigma}^2$ from above and Slutsky's Lemma):

$$\mathbb{P}\left(|\hat{\varphi}_k^{(n)}| \geq \frac{c_\alpha \hat{\sigma}}{\sqrt{n}}\right) \to \alpha$$

if $c_\alpha > 0$ is chosen such that $P(|Z| \geq c_\alpha) = \alpha$ for $Z \sim \mathrm{N}(0,\, 1)$.
The fact that σ^2 is the asymptotic variance of $\sqrt{n}\hat{\varphi}_k^{(n)}$ follows from $(C_m^{-1})_{k,\,k} = \sigma^2$ in the case $m \geq k > p$, for this see Brockwell/Davies.
Other approaches to select the 'right' order of the AR-process are based on model selection criteria like AIC, BIC.

CLT for martingale differences
$\rightsquigarrow$ recall standard CLT: $(\xi_i)_{i\geq 1}$ i.i.d., $\mathbb{E}[\xi_i] = 0$, $\xi_i \in L^2$, $S_n = \sum_{i=1}^n \xi_i \Rightarrow \frac{S_n}{\mathrm{Var}(S_n)^{1/2}} \stackrel{\text{d}}{\to} \mathrm{N}(0,\, 1)$.

Questions

- What if (ξ_i) are not identically distributed?
  $\to$ Lindeberg CLT.

- What if (ξ_i) are uncorrelated?
  $\to$ no CLT: Y, $(\varepsilon_i)_{i\geq 1}$ are independent random variables, $\mathbb{E}[Y] = 0$, $\mathbb{E}[Y^2] = 1$, $\varepsilon_i \sim \mathrm{N}(0,\, 1)$, $\xi_i = Y\varepsilon_i$
  $\rightsquigarrow \frac{S_n}{\mathrm{Var}(S_n)^{1/2}} = Y\varepsilon^{(n)}$, $\varepsilon^{(n)} = \frac{1}{\sqrt{n}}(\varepsilon_1 + \dots + \varepsilon_n) \sim \mathrm{N}(0,\, 1)$.
  For arbitrary Y this is not Gaussian $\mathrm{N}(0,\, 1)$.

  But: CLT holds if ξ_i are martingale differences:
  $\xi_i = M_i - M_{i-1}$, $\mathbb{E}[M_i] = 0 \rightsquigarrow \mathbb{E}[\xi_i\xi_j] \stackrel{i\neq j}{=} 0$.

1.30 Definition. $(\xi_i)_{i\geq 1}$ are called martingale differences w.r.t. $(\mathscr{F}_i)_{i\geq 1}$ if

- $(\mathscr{F}_i)_{i\geq 1}$ is a filtration, $\mathscr{F}_0 = \{\emptyset, \Omega\}$,
- ξ_i is $\mathscr{F}_i$-measurable, $i \geq 1$,
- $\xi_i \in L^2$, $\mathbb{E}[\xi_i | \mathscr{F}_{i-1}] = 0$, $i \geq 1$.

The triangular array

$$\begin{array}{llll} \xi_1^{(1)} & & & \\ \xi_1^{(2)} & \xi_2^{(2)} & & \\ \vdots & & \ddots & \\ \xi_1^{(k)} & \dots & \xi_k^{(k)} & \\ \vdots & & & \ddots \end{array}$$

where $(\xi_i^{(n)})_{i=1,\dots,n}$ are martingale differences w.r.t. $(\mathscr{F}_i^{(n)})_{i=0,\dots,n}$ for each $n \in \mathbb{N}$ is called a martingale difference scheme (MDS). We set

$$(\sigma_i^{(n)})^2 = \mathbb{E}[(\xi_i^{(n)})^2 | \mathscr{F}_{i-1}],$$
$$V_{n,i}^2 = \sum_{j=1}^{i} (\sigma_j^{(n)})^2,\ 1 \leq i \leq n,\ V_n^2 = V_{n,n}^2.$$

We say that $(\xi_i^{(n)})_{i,n}$ satisfies the conditional Lindeberg condition if

$$\sum_{i=1}^{n} \mathbb{E}\left[(\xi_i^{(n)})^2 \mathbf{1}_{(|\xi_i^{(n)}|>\delta)} | \mathscr{F}_{i-1}^{(n)}\right] \xrightarrow{\mathbb{P}} 0 \text{ for all } \delta > 0.$$

Problem 10: The conditional Lindeberg condition implies $\max_{1\leq i\leq n} \sigma_i^{(n)} \xrightarrow{\mathbb{P}} 0$ ('conditional Feller condition').

1.31 Lemma. *$Q(x) = \frac{e^{ix}-1-ix+x^2/2}{x^2/2}$ with $Q(0) = 0$, $M(x) = \frac{x}{3} \wedge 2$, $N(x) = e^{-x} - 1 + x$ satisfy for all $x \in \mathbb{R}$: $|1 - Q(x)| \leq 1$, $|Q(x)| \leq M(|x|)$, $|N(|x|)| \leq \frac{x^2}{2}$.*

Proof. By hand. □

1.32 Lemma. *Let (ξ_n), (η_n) be random variables with $\eta_n \neq 0$ a.s. Suppose φ is a characteristic function and $\lambda_0 \in \mathbb{R}$ with $\varphi(\lambda_0) \neq 0$. If*

(a) $\lim_{n\to\infty} \mathbb{E}[\eta_n^{-1} e^{i\lambda_0 \xi_n} - 1] = 0$,

(b) $\lim_{n\to\infty} \mathbb{E}[|\eta_n^{-1} - \varphi(\lambda_0)^{-1}|] = 0$,

then $\varphi_{\xi_n}(\lambda_0) = \mathbb{E}[e^{i\lambda_0 \xi_n}] \to \varphi(\lambda_0)$ holds.

Proof.

$$\begin{aligned}|\varphi_{\xi_n}(\lambda_0) - \varphi(\lambda_0)| &= |\varphi(\lambda_0)||\,\mathbb{E}[e^{i\lambda_0\xi_n}\varphi(\lambda_0)^{-1} - 1]| \\ &\le \varphi(\lambda_0)\ (\underbrace{|\,\mathbb{E}[e^{i\lambda_0\xi_n}\varphi(\lambda_0)^{-1} - e^{i\lambda_0\xi_n}\eta_n^{-1}]|}_{\le\mathbb{E}[|\varphi(\lambda_0)^{-1}-\eta_n^{-1}|]} + \underbrace{|\,\mathbb{E}[e^{i\lambda_0\xi_n}\eta_n^{-1} - 1]|}_{=|\,\mathbb{E}[\eta_n^{-1}-e^{-i\lambda_0\xi_n}]|}) \to 0.\end{aligned}$$

□

1.33 Theorem. *Let $\xi_i^{(n)}$ be a martingale difference scheme such that $V_n \xrightarrow{\mathbb{P}} 1$ ('norming') and the conditional Lindeberg condition are satisfied. Then*

$$S_n = \sum_{i=1}^{n} \xi_i^{(n)} \xrightarrow{d} \mathrm{N}(0,\,1).$$

Proof.

1. Truncation:
   Put $\eta_j^{(n)} := \xi_j^{(n)}\mathbf{1}_{(V_{n,\,j}^2 \le c)}$ for some $c > 1$, $T_n = \sum_{i=1}^n \eta_i^{(n)}$.
   We shall show:

   (i) $S_n - T_n \xrightarrow{\mathbb{P}} 0$,

   (ii) $(\eta_i^{(n)}, \mathscr{F}_i^{(n)})$ is an MDS satisfying 'norming', 'conditional Lindeberg' and $\mathbb{P}(W_n^2 \le c) = 1$, where

   $$W_n^2 = \sum_{i=1}^{n} \mathbb{E}[(\eta_i^{(n)})^2|\mathscr{F}_{i-1}^{(n)}].$$

   Because of (i) it suffices to prove $T_n \xrightarrow{\mathrm{d}} \mathrm{N}(0,\,1)$ (Slutsky Lemma), i.e. $\varphi_{T_n}(u) \to e^{-u^2/2}$ for all $u \in \mathbb{R}$.

2. Prove (i):
   Write $T_i^{(n)} = \sum_{j=1}^i \eta_j^{(n)}$, $W_{i,\,n}^2 = \sum_{j=1}^i \mathbb{E}[(\eta_j^{(n)})^2|\mathscr{F}_{j-1}^{(n)}]$.

   $$\begin{aligned}\mathbb{P}(\forall j = 1, \dots, n : \xi_j^{(n)} = \eta_j^{(n)}) &\ge \mathbb{P}(\forall j = 1, \dots, n : V_{j,n}^2 \le c) \\ &\ge 1 - \mathbb{P}(|V_n^2 - 1| > c - 1) \overset{\text{'norming'}}{\longrightarrow} 1 - 0 = 1.\end{aligned}$$

   $\Rightarrow$ for $\varepsilon > 0$: $\mathbb{P}(|S_n - T_n| > \varepsilon) \le \mathbb{P}(\exists j = 1, \dots, n : \xi_j^{(n)} \ne \eta_j^{(n)}) \to 0$
   $\Rightarrow S_n - T_n \xrightarrow{\mathbb{P}} 0$.

3. Prove (ii):
   MDS:

   $$\mathbb{E}[\eta_i^{(n)}|\mathscr{F}_{i-1}^{(n)}] \overset{V_{n,\,i}^2 \text{ is } \mathscr{F}_{i-1}^{(n)}\text{-mb.}}{=} \mathbf{1}_{(V_{n,\,i}^2 \le c)}\,\mathbb{E}[\xi_i^{(n)}|\mathscr{F}_{i-1}^{(n)}] = 0. \qquad (*)$$

'Conditional Lindeberg' follows directly from $|\eta_i^{(n)}| \le |\xi_i^{(n)}|$.
'Norming':

$$|W_n^2 - V_n^2| = |\sum_{j=1}^{n} \mathbb{E}[(\eta_j^{(n)})^2 - (\xi_j^{(n)})^2 | \mathscr{F}_{j-1}^{(n)}]| \le \underbrace{V_n^2}_{\overset{\mathbb{P}}{\to} 1} \underbrace{\mathbf{1}_{(\exists j=1,\dots,n: \xi_j^{(n)} \neq \eta_j^{(n)})}}_{\overset{\mathbb{P}}{\to} 0} \overset{\mathbb{P}}{\to} 0.$$

$\Rightarrow W_n^2 \to 1$.

$$W_n^2 = \sum_{j=1}^{n} \mathbb{E}[(\xi_j^{(n)})^2 \mathbf{1}_{(V_{j,n}^2 \le c)} | \mathscr{F}_{j-1}^{(n)}] \overset{\text{a.s.}}{=} \sum_{j=1}^{n} (\sigma_j^{(n)})^2 \mathbf{1}_{(V_{j,n}^2 \le c)} \overset{\text{by def.}}{\le} c \text{ (a.s.)}$$

4. CLT for T_n:
Apply the 2nd lemma above with $\varphi(\lambda) = e^{-\lambda^2/2}$, $\xi_n = T_n$, $\eta_n = e^{-\lambda^2 W_n^2/2}$. To conclude $T_n \overset{\mathrm{d}}{\to} \mathrm{N}(0, 1)$, we have to show
   (a) $\mathbb{E}[e^{i\lambda T_n + \lambda^2 W_n^2/2} - 1] \to 0$ for all $\lambda \in \mathbb{R}$,
   (b) $\mathbb{E}[|e^{\lambda^2 W_n^2/2} - e^{\lambda^2/2}|] \to 0$ for all $\lambda \in \mathbb{R}$.

   Part (b) follows immediately from $W_n \overset{\mathbb{P}}{\to} 1$, the continuity of $x \mapsto e^{\lambda x^2/2}$ (continuous mapping theorem) and the fact that $0 \le W_n^2 \le c$ a.s. (DCT).

5. Prove (a):
Let WLOG $\lambda \neq 0$, $1 \le k \le n$, set

$$\zeta_k^{(n)} = e^{i\lambda T_{k-1}^{(n)} + \frac{1}{2}\lambda^2 W_{n,k}^2} \, (e^{i\lambda \eta_k^{(n)}} - e^{-\frac{1}{2}\lambda^2 (\tau_k^{(n)})^2}),$$

$T_0^{(n)} = \eta_0^{(n)} := 0$, $(\tau_k^{(n)})^2 := \mathbb{E}\left[(\eta_k^{(n)})^2 \middle| \mathscr{F}_{k-1}^{(n)}\right]$. Then

$$\sum_{k=1}^{n} \zeta_k^{(n)} = e^{i\lambda T_n + \frac{1}{2}\lambda^2 W_n^2} - 1 \text{ (telescoping sum).}$$

$$\Rightarrow \left|\mathbb{E}\left[\zeta_k^{(n)} \middle| \mathscr{F}_{k-1}^{(n)}\right]\right| \underset{(*)}{\overset{N,Q \text{ from lemma,}}{=}} \left|e^{i\lambda T_{k-1}^{(n)} + \frac{1}{2}\lambda^2 W_{n,k}^2}\right| \cdot \left|\mathbb{E}\left[\frac{1}{2}\lambda^2 (\eta_k^{(n)})^2 Q(\lambda \eta_k^{(n)}) \middle| \mathscr{F}_{k-1}^{(n)}\right] - N\Big(\frac{1}{2}\lambda^2 (\tau_k^{(n)})^2\Big)\right|$$
$$\le e^{\frac{1}{2}\lambda^2 c} \Big(\mathbb{E}\Big[\frac{1}{2}\lambda^2 (\eta_k^{(n)})^2 M(|\lambda \eta_k^{(n)}|) \Big| \mathscr{F}_{k-1}^{(n)}\Big] + \frac{1}{2}\big(\frac{1}{2}\lambda^2 (\tau_k^{(n)})^2\big)^2 \Big)$$

$$\Rightarrow |\mathbb{E}[e^{i\lambda T_n + \frac{1}{2}\lambda^2 W_n^2} - 1]| \le \sum_{k=1}^{n} \mathbb{E}[|\mathbb{E}[\zeta_k^{(n)} | \mathscr{F}_{k-1}^{(n)}]|]$$
$$\le \frac{1}{2}\lambda^2 e^{\frac{1}{2}\lambda^2 c} \big(\sum_{k=1}^{n} \mathbb{E}[(\eta_k^{(n)})^2 M(|\lambda \eta_k^{(n)}|)] + \frac{1}{4}\lambda^2 c \, \mathbb{E}[\max_{j=1,\dots,n} (\tau_j^{(n)})^2] \big).$$

Problem 10 implies that $\max_{j=1,\dots,n} (\tau_j^{(n)})^2 \xrightarrow{\mathbb{P}} 0$. Moreover, $\tau_j^{(n)} \le c$ such that 2^{nd} term $\to 0$.
By conditional Lindeberg for any $\delta > 0$:

$$\begin{aligned}&\sum_{k=1}^{n} \mathbb{E}[(\eta_k^{(n)})^2 M(|\lambda \eta_k^{(n)}|)] \\ &\le \sum_{k=1}^{n} \Big(\underbrace{2\,\mathbb{E}[\mathbb{E}[(\eta_k^{(n)})^2 \mathbf{1}_{(|\eta_k^{(n)}|>\delta)} | \mathscr{F}_{k-1}^{(n)}]]}_{\sum(\dots) \overset{\text{cond. Lind., DCT}}{\rightarrow} 0} + \underbrace{\frac{\delta|\lambda|}{3}\,\mathbb{E}[(\eta_k^{(n)})^2]}_{\sum(\dots) \overset{\text{'norming'}}{\rightarrow} \frac{\delta|\lambda|}{3}} \Big).\end{aligned}$$

Since this is true for all $\delta > 0$, we conclude (a).

□

Problem 11: Show that the conditional Lyapunov condition

$$\exists \varepsilon > 0 : \sum_{j=1}^{n} \mathbb{E}\left[|\xi_j^{(n)}|^{2+\varepsilon} \middle| \mathscr{F}_{j-1}^{(n)} \right] \xrightarrow{\mathbb{P}} 0$$

implies 'conditional Lindeberg'.

Problem 12:

(a) Let (M_n) be an L^2-martingale, (s_n) be deterministic such that $\frac{\langle M \rangle_n}{s_n^2} \xrightarrow{\mathbb{P}} 1$ and

$$\sum_{i=1}^{n} \mathbb{E}\left[\left|\frac{M_i - M_{i-1}}{s_n}\right|^2 \mathbf{1}_{\left(\left|\frac{M_i - M_{i-1}}{s_n}\right| > \delta\right)} \middle| \mathscr{F}_{i-1} \right] \xrightarrow{\mathbb{P}} 0.$$

Then $\frac{M_n}{s_n} \xrightarrow{\text{d}} \mathrm{N}(0,1)$. (Show that $s_n \to \infty$.)
Do we then also have $\frac{M_n}{\langle M \rangle_n^{1/2}} \xrightarrow{\text{d}} \mathrm{N}(0,1)$?

(b) Formulate and prove by Cramér-Wold device a multivariate MDS-CLT.

(c) Give counterexamples of L^2-martingales where (a) does not hold.

Proof (CLT for Yule-Walker).

1. AR(p)-process: $X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t$, $(\varepsilon_t) \sim \mathrm{IID}(0, \sigma^2)$.
   Rewrite it in 'regression language' as $Y = X\varphi + \varepsilon$ with $Y = (X_1, \dots, X_n)^T$, design matrix

$$X = \begin{pmatrix} X_0 & X_{-1} & \dots & X_{1-p} \\ X_1 & X_0 & & X_{2-p} \\ \vdots & & \ddots & \vdots \\ X_{n-1} & X_{n-2} & \dots & X_{n-p} \end{pmatrix} \in \mathbb{R}^{n \times p},$$

$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$.
Standard Least-Squares estimator:

$$\varphi_n^* = (X^T X)^{-1} X^T Y.$$

$$\frac{1}{n}(X^T X)_{i,j} = \frac{1}{n}\sum_{k=1}^{n} X_{k-i}X_{k-j} \approx \hat{c}(i-j) = (\hat{C}_p)_{i,j},$$
$$\frac{1}{n}(X^T Y)_i = \frac{1}{n}\sum_{k=1}^{n} X_{k-i}X_k \approx \hat{c}(i),\ i,\ j = 1, \ldots p.$$

This means: $\varphi_n^* \approx \hat{\varphi}^{(n)}$, Yule-Walker.
We have $\varphi_n^* = \varphi + (X^T X)^{-1} X^T \varepsilon$.

2. We have $\varphi_n^* - \hat{\varphi}^{(n)} = o_{\mathbb{P}}(n^{-1/2})$ (i.e. $n^{1/2}(\varphi_n^* - \hat{\varphi}^{(n)}) \xrightarrow{\mathbb{P}} 0$)

$$\frac{1}{n}X^T Y - \hat{c}_p = \frac{1}{n}(\sum_{k=1}^{n} X_{k-i}X_k - \sum_{k=1}^{n-i} X_k X_{k+i})_i = \frac{1}{n}(\underbrace{\sum_{k=1}^{i} X_{k-i}X_k}_{\leq p \text{ summands}})_i.$$

Weak stationarity implies that

$$\mathbb{E}[\|\frac{1}{n}X^T Y - \hat{c}_p\|] \leq \frac{c \cdot p}{n} \text{ for some } c > 0$$

$$\Rightarrow \|\frac{1}{n}X^T Y - \hat{c}_p\| = \mathcal{O}_{L^1}(\frac{1}{n})$$
$$\Rightarrow \sqrt{n}\|\frac{1}{n}X^T Y - \hat{c}_p\| \xrightarrow{\mathbb{P}} 0, \text{ i.e. } \|\frac{1}{n}X^T Y - \hat{c}_p\| = o_{\mathbb{P}}(n^{-1/2}).$$

Similarly,

$$\frac{1}{n}X^T X - \hat{C}_p = \frac{1}{n}(\sum_{k=1}^{n} X_{k-i}X_{k-j} - \sum_{k=1}^{n-|i-j|} X_k X_{k+|i-j|})_{i,j}$$
$$= \mathcal{O}_{L_1}(n^{-1}) = o_{\mathbb{P}}(n^{-1/2}).$$

Use continuous mapping theorem to conclude that $\varphi_n^* - \hat{\varphi}^{(n)} = o_{\mathbb{P}}(n^{-1/2})$.
We note for $\varphi_n^* - \varphi = (X^T X)^{-1} X^T \varepsilon$ that

$$M_n^{(i)} := (X^T \varepsilon)_i = X_{1-i}\varepsilon_1 + \cdots + X_{n-i}\varepsilon_n \ (i = 1, \ldots, p)$$

is a martingale in n w.r.t. $\mathscr{F}_n = \sigma(\varepsilon_1, \ldots, \varepsilon_n, X_0, \ldots, X_{-p+1})$:

- $X_k \in L_2$, $(\varepsilon_i) \in L_2 \Rightarrow M_n^{(i)} \in L_1$
  ($M_n^{(i)}$ is even in L_2: $\mathbb{E}[(X_{k-i}\varepsilon_k)^2] \overset{\text{indep.}}{=} \mathbb{E}[X_{k-i}^2]\,\mathbb{E}[\varepsilon_k^2] < \infty$),
- $\mathbb{E}[M_n^{(i)}|\mathscr{F}_{n-1}] = X_{1-i}\varepsilon_1 + \cdots + X_{n-1-i}\varepsilon_{n-1} + \underbrace{\mathbb{E}[\varepsilon_n|\mathscr{F}_{n-1}]}_{=\mathbb{E}[\varepsilon_n]=0} = M_{n-1}^{(i)}$

with quadratic variation

$$\langle M^{(i)}\rangle_n = \sum_{k=1}^{n} \mathbb{E}[(M_k^{(i)} - M_{k-1}^{(i)})^2|\mathscr{F}_{k-1}] = \sigma^2 \sum_{k=1}^{n} X_{k-i}^2 = \sigma^2 (X^T X)_{i,\,i}.$$

Now, $M_n = (M_n^{(1)}, \ldots, M_n^{(p)})^T$ is a vector-valued martingale. Its quadratic covariation matrix $\langle M\rangle_n \in \mathbb{R}^{p\times p}$ satisfies

$$\langle M\rangle_n = \sum_{k=1}^{n} \mathbb{E}[(M_k - M_{k-1})(M_k - M_{k-1})^T|\mathscr{F}_{k-1}] = \sigma^2(X^T X).$$

Hence, $\varphi_n^* - \varphi = \sigma^2 \langle M\rangle_n^{-1} M_n$.

From the chapter on autocovariances we know that $\hat{c}(k) \xrightarrow{\mathbb{P}} c(k)$ (empirical covariances are consistent) if $(c(k))_{k\in\mathbb{Z}}$ decays sufficiently. Here $c(k)$ even decays with geometric rate in k such that this holds (since X is causal). This means $\hat{C}_p \xrightarrow{\mathbb{P}} C_p$ and thus

$$\frac{1}{n} X^T X = \hat{C}_p + \underbrace{(\frac{1}{n} X^T X - \hat{C}_p)}_{\xrightarrow{\mathbb{P}} 0} \xrightarrow{\mathbb{P}} C_p.$$

We define the following martingale difference scheme:

$$\xi_i^{(n)} := (n\cdot \sigma^2 \cdot C_p)^{-1/2}(M_i - M_{i-1}) \in \mathbb{R}^p,\ 1 \le i \le n.$$

It has conditional covariance matrix

$$V_n = V_{n,\,n} = (n\sigma^2 C_p)^{-1} \underbrace{\langle M\rangle_n}_{\sigma^2 X^T X} \xrightarrow{\mathbb{P}} E_p = \mathrm{diag}(1, \ldots, 1) \in \mathbb{R}^{p\times p}$$

such that the norming condition is satisfied.

Check the conditional Lindeberg condition

$$\sum_{i=1}^{n} \mathbb{E}[\|(n\sigma^2 C_p)^{-1/2}(M_i - M_{i-1})\|^2 \mathbf{1}_{(\|(n\sigma^2 C_p)^{-1/2}(M_i - M_{i-1})\| > \delta)} |\mathscr{F}_{i-1}] \xrightarrow{\mathbb{P}} 0.$$

We even have L^1-convergence because of

$$\sum_{i=1}^{n} \mathbb{E}[\|(n\sigma^2 C_p)^{-1/2}(M_i - M_{i-1})\|^2 \mathbf{1}_{(\|(n\sigma^2 C_p)^{-1/2}(M_i - M_{i-1})\| > \delta)}]$$
$$\overset{X \text{ stat.}}{=} \mathbb{E}[\underbrace{\|(\sigma^2 C_p)^{-1/2}(M_1 - M_0)\|^2}_{\mathbb{E}[\ldots]<\infty}\ \underbrace{\mathbf{1}_{(\|(\sigma^2 C_p)^{-1/2}(M_1 - M_0)\| > \delta\sqrt{n})}}_{\to 0 \text{ and } \le 1}] \xrightarrow{\text{DCT}} 0.$$

Hence, we can apply a vector version of the CLT for MDS. It yields

$$(n\sigma^2 C_p)^{-1/2} M_n \xrightarrow{\mathrm{d}} \mathrm{N}(0,\, E_p).$$

We write

$$\sigma^{-2}(\varphi_n^* - \varphi) = \langle M\rangle_n^{-1} M_n = \underbrace{\langle M\rangle_n^{-1}(n\sigma^2 C_p)}_{\xrightarrow{\mathbb{P}} E_p}(n\sigma^2 C_p)^{-1} M_n$$

Then by Slutsky's lemma

$$\Rightarrow \sigma^{-2}(n\sigma^2 C_p)^{1/2}(\varphi_n^* - \varphi) \xrightarrow{\mathrm{d}} \mathrm{N}(0,\, E_p)$$

$$\Rightarrow n^{1/2}(\varphi_n^* - \varphi) \xrightarrow{\mathrm{d}} \mathrm{N}(0,\, \sigma^4(\sigma^2 C_p)^{-1}) = \mathrm{N}(0,\, \sigma^2 C_p^{-1}).$$

3. Fine point: C_p is non-singular, i.e. $C_p > 0$. For $a \in \mathbb{R}^p$:

$$\begin{aligned}\langle C_p a, a\rangle &= \sum_{k,\,l=1}^{p} c(k-l) a_k a_l = \mathrm{Var}(\sum_{k=1}^{p} a_k X_k) \\ &\overset{X \text{ is AR}(p)}{=} \mathrm{Var}(\sum_{k=1}^{p-1} a_k X_k + a_p(\varphi_1 X_{p-1} + \cdots + \varphi_p X_0 + \varepsilon_p)) \\ &\overset{\varepsilon \text{ indep. of } X_k,\, k<p}{=} \mathrm{Var}(\sum_{k=1}^{p-1} a_k X_k + a_p(\varphi_1 X_{p-1} + \cdots + \varphi_p X_0)) + a_p^2\sigma^2.\end{aligned}$$

Hence, $\langle C_p a, a\rangle = 0 \Rightarrow a_p = 0$ and continuing in the same way we obtain $a_p = a_{p-1} = \cdots = a_1 = 0 \Leftrightarrow a = 0$ and thus $C_p > 0$ and C_p non-singular.

□

Problem 13: Consider the Yule-Walker estimator of an AR(1)-process $X_t = \varphi_1 X_{t-1} + \varepsilon_t$, $(\varepsilon_t) \sim \mathrm{IID}(0, \sigma^2)$ and show that in the 'exploding case' $|\varphi_1| > 1$ the estimator converges to φ_1 (in probability) with geometric speed in n, i.e. $\hat{\varphi}_1^{(n)} - \varphi = o_{\mathbb{P}}(r^n)$ for some $r \in (0,1)$.

Problem 14: Consider the causal (weakly stationary) AR(1)-process with $(\varepsilon_t) \sim \mathrm{N}(0, \sigma^2)$. Determine the Maximum-Likelihood estimator (MLE) of φ_1. Discuss its difference to the Yule-Walker estimator.

Question: Is there another sequence of estimators $\tilde{\varphi}^{(n)}$ of φ based on $X_1, \ldots, X_n$ which is better in the sense that $\tilde{\varphi}^{(n)}$ converges with faster rate than $n^{-1/2}$ to φ (in probability) or

$$\sqrt{n}(\tilde{\varphi}^{(n)} - \varphi) \xrightarrow{\mathrm{d}} \mathrm{N}(0,\, V)$$

with $V < \sigma^2 C_p^{-1}$ (i.e. $\sigma^2 C_p^{-1} - V$ is positive semi-definite and $\sigma^2 C_p^{-1} - V \neq 0$)?

Tool: Fisher information.

Excursion: Suppose $\hat{g} : \Omega \to \mathbb{R}$ is an unbiased estimator of $g(\vartheta)$ $(g : \Theta \to \mathbb{R})$, i.e. $\hat{g}$ is measurable on $(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$, Θ non-empty index set, $\mathbb{E}_\vartheta[\hat{g}] = g(\vartheta)$ for all $\vartheta \in \Theta$, and that $\hat{g} \in L^2(\mathbb{P}_\vartheta)$, $\vartheta \in \Theta$. Moreover, suppose that $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ is dominated by a σ-finite measure μ on $(\Omega, \mathscr{F})$, i.e. $\mathbb{P}_\vartheta \ll \mu$ for all $\vartheta \in \Theta$, and let $p_\vartheta = \frac{d\mathbb{P}_\vartheta}{d\mu}$ be the densities (Radon-Nikodym derivatives). We want to derive a lower bound on

$$\mathbb{E}_\vartheta[(\hat{g} - \underbrace{g(\vartheta)}_{\mathbb{E}_\vartheta[\hat{g}]})^2] = \mathrm{Var}_\vartheta(\hat{g}).$$

For each $H \in L^2(\mathbb{P}_\vartheta)$ Cauchy-Schwarz inequality yields

$$\mathbb{E}_\vartheta[(\hat{g} - g(\vartheta))H]^2 \le \mathbb{E}_\vartheta[(\hat{g} - g(\vartheta))^2]\,\mathbb{E}_\vartheta[H^2]$$

$$\Rightarrow \mathbb{E}_\vartheta[(\hat{g} - g(\vartheta))^2] \ge \frac{\mathbb{E}_\vartheta[(\hat{g} - g(\vartheta))H]^2}{\mathbb{E}_\vartheta[H^2]} \text{ for all } H \in L^2(\mathbb{P}_\vartheta).$$

Goal: find H such that the numerator is independent of $\hat{g}$.
Fisher's idea: $H_\vartheta = \frac{\mathrm{d}}{\mathrm{d}\vartheta}(\log p_\vartheta)\mathbf{1}_{(p_\vartheta>0)} = \frac{\frac{\mathrm{d}}{\mathrm{d}\vartheta}p_\vartheta}{p_\vartheta}\mathbf{1}_{(p_\vartheta>0)}$, $\vartheta \in \Theta \subseteq \mathbb{R}^d$.
Then formally:

$$\begin{aligned}\mathbb{E}_{\vartheta_0}[H_{\vartheta_0}] &= \int_\Omega H_{\vartheta_0} \underbrace{p_{\vartheta_0}\mathrm{d}\mu}_{\mathrm{d}\mathbb{P}_{\vartheta_0}} = \int_{\{p_{\vartheta_0}>0\}} \frac{\mathrm{d}}{\mathrm{d}\vartheta}p_\vartheta\Big|_{\vartheta=\vartheta_0}\mathrm{d}\mu \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}\vartheta}\int_{\{p_{\vartheta_0}>0\}} p_\vartheta \mathrm{d}\mu\right)\Big|_{\vartheta=\vartheta_0} = \left(\frac{\mathrm{d}}{\mathrm{d}\vartheta}\underbrace{\int_{\{p_\vartheta>0\}} p_\vartheta \mathrm{d}\mu}_{=1}\right)\Big|_{\vartheta=\vartheta_0} = 0.\end{aligned}$$

For the change of the integration boundary above note:

$$G(\vartheta) := \int_\Omega \mathbf{1}_{(p_{\vartheta_0}=0)}p_\vartheta \mathrm{d}\mu. \text{ If } G \in C^1, \text{ then } G'(\vartheta_0) = 0.$$

Hence,

$$\begin{aligned}\mathbb{E}_{\vartheta_0}[(\hat{g} - g(\vartheta_0))H_{\vartheta_0}] &= \mathrm{Cov}_{\vartheta_0}(\hat{g}, H_{\vartheta_0}) = \mathbb{E}_{\vartheta_0}[\hat{g}(H_{\vartheta_0} - \mathbb{E}_{\vartheta_0}[H_{\vartheta_0}])] \\ &= \int \hat{g}\frac{\frac{\mathrm{d}}{\mathrm{d}\vartheta}p_\vartheta\big|_{\vartheta=\vartheta_0}}{p_{\vartheta_0}}\mathbf{1}_{(p_{\vartheta_0}>0)}p_{\vartheta_0}\mathrm{d}\mu = \frac{\mathrm{d}}{\mathrm{d}\vartheta}\Big(\int_{\{p_{\vartheta_0}>0\}} \hat{g}p_\vartheta \mathrm{d}\mu\Big)\Big|_{\vartheta=\vartheta_0}.\end{aligned}$$

Since $\hat{g}$ is unbiased, we have

$$\int \hat{g}p_\vartheta \mathrm{d}\mu = \mathbb{E}_\vartheta[\hat{g}] = g(\vartheta)$$

$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}\vartheta}\Big(\int \hat{g}p_\vartheta \mathrm{d}\mu\Big)\Big|_{\vartheta=\vartheta_0} = \frac{\mathrm{d}}{\mathrm{d}\vartheta}g(\vartheta)\big|_{\vartheta=\vartheta_0} = g'(\vartheta_0)$$

$\rightsquigarrow$ numerator $= g'(\vartheta_0)^2$.
<u>Cramér-Rao inequality</u>:

$$\mathbb{E}_{\vartheta_0}[(\hat{g} - g(\vartheta_0))^2] \ge \frac{g'(\vartheta_0)^2}{\mathbb{E}_{\vartheta_0}[(\frac{\mathrm{d}}{\mathrm{d}\vartheta}(\log p_\vartheta)\big|_{\vartheta=\vartheta_0})^2]} =: \frac{g'(\vartheta_0)^2}{I(\vartheta_0)}$$

where $I(\vartheta_0) = \mathbb{E}_{\vartheta_0}[(\frac{\mathrm{d}}{\mathrm{d}\vartheta}(\log p_\vartheta)\big|_{\vartheta=\vartheta_0})^2]$ is the Fisher information at $\vartheta = \vartheta_0$. (This holds for unbiased estimators $\hat{g}$ of $g(\vartheta)$ under regularity conditions on (p_ϑ) and $\hat{g}$).
$\rightsquigarrow$ Formal versions and proofs:

- Lehmann/Casella: Theory of Point Estimation ([5]),
- van der Vaart: Asymptotic Statistics ([8]).

1.34 Remark. If $\hat{g}$ is biased, i.e. $\mathbb{E}_\vartheta[\hat{g}] = g(\vartheta) + b(\vartheta)$ for some b, we obtain from above in terms of $\tilde{g}(\vartheta) = g(\vartheta) + b(\vartheta)$:

$$\operatorname{Var}_\vartheta(\hat{g}) \geq \frac{\tilde{g}'(\vartheta)^2}{I(\vartheta)}.$$

The bias-variance decomposition thus yields

$$\mathbb{E}_\vartheta[(\hat{g} - g(\vartheta))^2] \geq b(\vartheta)^2 + \frac{(g'(\vartheta) + b'(\vartheta))^2}{I(\vartheta)}.$$

Problem 15: Formulate and prove the Cramér-Rao inequality for $\vartheta \in \Theta \subseteq \mathbb{R}^d$, i.e. for $d \geq 2$ (with $g : \Theta \to \mathbb{R}$).

Asymptotic efficiency lower bound:
Hajek-Le Cam convolution theorem: If the statistical model is (asymptotically) regular (e.g. LAN), then any 'reasonable' estimator $\hat{g}^{(n)}$ of $g(\vartheta)$ satisfies

$$\sqrt{I^{(n)}(\vartheta_0)}(\hat{g}^{(n)} - g(\vartheta_0)) \xrightarrow{\mathrm{d}} Q_{\vartheta_0}$$

for some limit distribution Q_{ϑ_0} and we have

$$Q_{\vartheta_0} = \mathrm{N}(0,\, g'(\vartheta_0)^2) * R_{\vartheta_0}$$

for some law R_{ϑ_0} ($*$ denotes the convolution).

Interpretation: Since convolution of measures spreads the probability distribution (e.g. increases variance if it exists), the most concentrated limit law we can obtain is $\mathrm{N}(0,\, g'(\vartheta_0)^2)$ (meaning $R_{\vartheta_0} = \delta_0$). Therefore, estimators $(\hat{g}^{(n)})$ with

$$\sqrt{I^{(n)}(\vartheta_0)}(\hat{g}^{(n)} - g(\vartheta_0)) \xrightarrow{\mathrm{d}} \mathrm{N}(0,\, g'(\vartheta_0)^2)$$

are called asymptotically efficient.

Superficial similarity to Cramér-Rao bound:

$$\hat{g}^{(n)} - g(\vartheta_0) \overset{\mathrm{d}}{\approx} \mathrm{N}\Big(0,\, \frac{g'(\vartheta_0)^2}{I^{(n)}(\vartheta_0)}\Big).$$

Note that $\hat{g}^{(n)}$ was not supposed to be unbiased.

Let us now look at the Yule-Walker estimator for a causal AR(1)-process

$$X_t = \vartheta X_{t-1} + \varepsilon_t,\ \vartheta \in (-1,\,1),\ (\varepsilon_t) \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0,\,\sigma^2).$$

Here $\Theta = (-1,\,1)$, $g(\vartheta) = \vartheta$, $g'(\vartheta) = 1$. Write μ_ϑ for the Lebesgue density of X_0 under $\mathbb{P}_\vartheta$. One can prove that this AR(1)-model is indeed 'regular'.
The random vector $(X_0, \dots, X_n)$ has Lebesgue density ($\mu = \lambda_{\mathbb{R}^{n+1}}$):

$$p_\vartheta^{(n)}(x_0, \dots, x_n) = \mu_\vartheta(x_0)\varphi_{0,\,\sigma^2}(x_1 - \vartheta x_0) \cdot \ldots \cdot \varphi_{0,\,\sigma^2}(x_n - \vartheta x_{n-1})$$

with $\varphi_{\mu,\,\sigma^2}$ density of $\mathrm{N}(\mu,\,\sigma^2)$, i.e. ε_i has density $\varphi_{0,\,\sigma^2}$.

Log-Likelihood:

$$\log p_\vartheta^{(n)}(x_0, \dots, x_n) = \log(\mu_\vartheta(x_0)) + \sum_{k=1}^n \log(\varphi_{0,\,\sigma^2}(x_k - \vartheta x_{k-1})).$$

Score function:

$$\frac{\mathrm{d}}{\mathrm{d}\vartheta} \log p_\vartheta^{(n)}(x_0, \dots, x_n) \ = \frac{\mathrm{d}}{\mathrm{d}\vartheta} \log(\mu_\vartheta(x_0)) + \sum_{k=1}^n \big(- \frac{1}{\sigma^2}\big) x_{k-1}(x_k - \vartheta x_{k-1}).$$

$$\begin{aligned}
&\mathbb{E}_{\vartheta_0}[(\frac{\mathrm{d}}{\mathrm{d}\vartheta} \log p_\vartheta^{(n)}(X_0, \dots, X_n)\big|_{\vartheta=\vartheta_0})^2] \\
&\quad = \mathbb{E}_{\vartheta_0}[(\frac{\mathrm{d}}{\mathrm{d}\vartheta} \log(\mu_\vartheta(X_0))\big|_{\vartheta=\vartheta_0} + \sum_{k=1}^n \big(- \frac{1}{\sigma^2}\big) X_{k-1}\varepsilon_k)^2] \\
&\quad \overset{(*)}{=} \mathrm{Var}_{\vartheta_0}(\frac{\mathrm{d}}{\mathrm{d}\vartheta} \log(\mu_\vartheta(X_0))\big|_{\vartheta=\vartheta_0}) + \ \sum_{k=1}^n \frac{1}{\sigma^4}\, \mathbb{E}_{\vartheta_0}[X_{k-1}^2]\sigma^2 \\
&\quad \overset{X \text{ stat.}}{=} \mathrm{Var}_{\vartheta_0}(\frac{\mathrm{d}}{\mathrm{d}\vartheta} \log(\mu_\vartheta(X_0))\big|_{\vartheta=\vartheta_0}) \ + \frac{n}{\sigma^2} \underbrace{\mathbb{E}_{\vartheta_0}[X_0^2]}_{=c_{\vartheta_0}(0)}.
\end{aligned}$$

(For $(*)$ regularity conditions are required $\rightsquigarrow$ regular model.)

$$\begin{aligned}
&\Rightarrow I^{(n)}(\vartheta_0) = \frac{2\frac{\vartheta_0^2}{(1-\vartheta_0^2)^2} + \sigma^2 n c_{\vartheta_0}(0)}{\sigma^4} \\
&\Rightarrow \frac{I^{(n)}(\vartheta_0)}{n} \to \frac{c_{\vartheta_0}(0)}{\sigma^2}.
\end{aligned}$$

This means that an estimator $(\tilde{\vartheta}^{(n)})$ with

$$\sqrt{n}(\tilde{\vartheta}^{(n)} - \vartheta) \overset{\mathrm{d}}{\to} \mathrm{N}\left(0,\, \frac{\sigma^2}{c_{\vartheta_0}(0)}\right)$$

is asymptotically efficient. This is the case for the Yule-Walker estimator.

Problem 16: Investigate whether the Yule-Walker estimator for causal AR(p)-processes, $p \geq 2$, is also asymptotically efficient (in a natural generalisation).

Final remark: In the 'explosive' case (e.g. AR(1) with $|\vartheta| > 1$) the Fisher information grows geometrically in n and the Yule-Walker estimator also converges with geometric rate in n.

2 Statistics for continuous-time processes

2.1 Diffusion processes

2.1 Definition. A (time-inhomogeneous) diffusion process in $\mathbb{R}^d$ is a process $(X_t,\, t \geq 0)$ solving the stochastic differential equation (SDE)

$$\mathrm{d}X_t = b(X_t,\, t)\mathrm{d}t + \sigma(X_t,\, t)\mathrm{d}W_t,\, t \geq 0, \qquad (*)$$

with initial condition $X_0 = X^{(0)}$. Here $b : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^{d\times m}$ and W is m-dimensional Brownian motion.

The intuition is that (after 'division by dt')

$$\dot{X}_t = \frac{\mathrm{d}X_t}{\mathrm{d}t} = b(X_t,\, t) + \sigma(X_t,\, t)\dot{W}_t,$$

where $\dot{W}_t$ is Gaussian white noise ('equivalent of i.i.d. N(0, 1)-random variables in continuous time'). Since white noise can only be defined in a distributional sense, the Itô interpretation in terms of integrated quantities is nowadays preferred.

Rigorous definition: X is a strong solution of the SDE $(*)$, where W is defined on some $(\Omega, \mathscr{F}, \mathbb{P})$and $X^{(0)}$ is independent of W on $(\Omega, \mathscr{F}, \mathbb{P})$, if

(a) $(X_t,\, t \geq 0)$ is adapted to the completion by null sets of

$$\mathscr{F}_t^0 = \sigma(W_s,\, 0 \leq s \leq t;\, X^{(0)});$$

(b) X is a continuous process;

(c) $\mathbb{P}(X_0 = X^{(0)}) = 1$;

(d) $\mathbb{P}(\int_0^t (\|b(X_s,\, s)\| + \|\sigma(X_s,\, s)\|^2)\mathrm{d}s < \infty) = 1$ for all $t > 0$ (with $\|\cdot\|$ any norm);

(e) With probability one:

$$\forall t \geq 0 : X_t = X_0 + \int_0^t b(X_s,\, s)\mathrm{d}s + \int_0^t \sigma(X_s,\, s)\mathrm{d}W_s.$$

The stochastic integral is taken in Itô's sense and obtained as the limit of sums

$$0 = t_0 < t_1 < \cdots < t_m = t : \sum_{i=1}^{m} \sigma(X_{t_{i-1}}, t_{i-1})(W_{t_i} - W_{t_{i-1}})$$

where $\Delta := \max_i |t_i - t_{i-1}| \to 0$.

2.2 Theorem (Standard existence and uniqueness result for SDEs)**.** *Suppose the drift coefficient b and the diffusion coefficient σ satisfy the global Lipschitz and linear growth conditions*

(i) $\|b(x, t) - b(y, t)\| + \|\sigma(x, t) - \sigma(y, t)\| \leq K\|x - y\|$,

(ii) $\|b(x, t)\| + \|\sigma(x, t)\| \leq K(1 + \|x\|)$

for all $x, y \in \mathbb{R}^d$, $t \geq 0$ and some constant K. Then the SDE $(*)$ has a strong solution which is also unique, provided $X^{(0)} \in L^2$.

If $(X_t, t \in [0, T])$ is observed (continuous-time observations), then by taking refined partitions, we can calculate the quadratic (co-)variation

$$\int_0^t \sigma(X_s, s)\sigma(X_s, s)^T \mathrm{d}s$$

for all $t \in [0, T]$:

$$\sum_{i=1}^{m} (X_{t_i} - X_{t_{i-1}})(X_{t_i} - X_{t_{i-1}})^T \overset{\Delta \to 0}{\underset{\text{a.s.}}{\to}} \int_0^t \sigma(X_s, s)\sigma(X_s, s)^T \mathrm{d}s.$$

By taking the derivative in t, we thus identify $(\sigma\sigma^T)(X_t, t) \in \mathbb{R}^{d \times d}$ for all $t \in [0, T]$. Note that we cannot hope for more: if x is not visited by $(X_t, t \in [0, T])$ there is no chance to learn about $(\sigma\sigma^T)(x, t)$ for some t.
Moreover, we cannot find out more about $\sigma \in \mathbb{R}^{d \times m}$ itself, because X also solves an SDE of the form:

$$\mathrm{d}X_t = b(X_t, t) + (\sigma\sigma^T)^{1/2}(X_t, t)\mathrm{d}\tilde{W}_t$$

with $\tilde{W}$ a d-dimensional Brownian motion.

Résumé: Continuous-time observations identify the diffusion part as far as possible and the main interest is the drift part.

Main tool for drift statistics: Girsanov theorem to obtain the likelihood. [Liptser/Shiryaev: Statistics of Random Processes ([6])]

2.3 Theorem (Theorem 7.19 in [6])**.** *Let $(X_t, t \in [0, T])$, $(Y_t, t \in [0, T])$ be two real diffusion processes with*

$$\mathrm{d}X_t = b_X(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t,$$
$$\mathrm{d}Y_t = b_Y(Y_t, t)\mathrm{d}t + \sigma(Y_t, t)\mathrm{d}W_t$$

and $X_0 = Y_0$ a.s.
Suppose for Y there is a unique strong solution and $(b_X - b_Y)(x, t) = 0$ if $\sigma(x, t) = 0$. If

$$\mathbb{P}(\int_0^T \mathbf{1}_{(\sigma(X_s, s)>0)} \frac{(b_X^2 + b_Y^2)(X_s, s)}{\sigma^2(X_s, s)} \mathrm{d}s < \infty)$$
$$= \mathbb{P}(\int_0^T \mathbf{1}_{(\sigma(Y_s, s)>0)} \frac{(b_X^2 + b_Y^2)(Y_s, s)}{\sigma^2(Y_s, s)} \mathrm{d}s < \infty) = 1,$$

then the laws $\mathbb{P}_T^X$, $\mathbb{P}_T^Y$ of X and Y on $C([0, T])$ (with Borel-σ-algebra) are equivalent with Radon-Nikodym derivative/density/likelihood:

$$\frac{\mathrm{d}\,\mathbb{P}_T^Y}{\mathrm{d}\,\mathbb{P}_T^X}((X_t)_{t\in[0, T]})$$
$$= \exp\left\{ \int_0^T \mathbf{1}_{(\sigma(X_s, s)>0)} \left(\frac{b_Y - b_X}{\sigma^2}\right)(X_s, s)\mathrm{d}X_s \; - \frac{1}{2}\int_0^T \mathbf{1}_{(\sigma(X_s, s)>0)} \left(\frac{b_Y^2 - b_X^2}{\sigma^2}\right)(X_s, s)\mathrm{d}s\right\}.$$

2.4 Examples.

1. Brownian motion with drift:
   $b_X(X_t, t) = b_X(t)$, $b_Y(X_t, t) = b_Y(t)$, $\sigma(X_t, t) = \sigma > 0$, $X^{(0)} = 0$, i.e.

   $$X_t = \int_0^t b_X(s)\mathrm{d}s + \sigma\mathrm{d}W_t,$$
   $$Y_t = \int_0^t b_Y(s)\mathrm{d}s + \sigma\mathrm{d}W_t$$

   $\rightsquigarrow$ all conditions above are satisfied and

   $$\frac{\mathrm{d}\,\mathbb{P}_T^Y}{\mathrm{d}\,\mathbb{P}_T^X}(X) = \exp\left\{ \int_0^T \frac{(b_Y - b_X)(s)}{\sigma^2}\mathrm{d}X_s \; - \frac{1}{2}\int_0^T \frac{(b_Y^2 - b_X^2)(s)}{\sigma^2}\mathrm{d}s\right\}.$$

   $\rightsquigarrow$ if b_Y, b_X are constant in t, then X_T is a sufficient statistics, i.e. for all statistical puropses it suffices to use X_T, not the trajectory $(X_T, t \in [0, T])$,
   $\rightsquigarrow$ enormous data reduction without loss of information on b_X, b_Y.
   Example: MLE for $\mathrm{d}X_t = \vartheta\mathrm{d}t + \sigma\mathrm{d}W_t$, $\vartheta \in \mathbb{R}$ unknown, is $\hat{\vartheta}_{\mathrm{MLE}} = \frac{X_T}{T}$.

2. Ornstein-Uhlenbeck process:
   It is the solution of the SDE

   $$\mathrm{d}X_t = aX_t\mathrm{d}t + \sigma\mathrm{d}W_t$$

for some initial value $X^{(0)}$.
Variation of constants formula gives

$$X_t = e^{at}X^{(0)} + \int_0^t e^{a(t-s)}\sigma \mathrm{d}W_s.$$

If $X^{(0)}$ is Gaussian or deterministic, then (X_t) is a Gaussian process.
It is easy to see that all conditions in Girsanov's theorem are satisfied for $b_Y(x, t) = ax$, $b_X(x, t) = 0$ (for $a = 0$) and thus

$$\frac{\mathrm{d}\,\mathbb{P}_T^Y}{\mathrm{d}\,\mathbb{P}_T^X} = \exp\left\{\int_0^T \frac{aX_s}{\sigma^2}\mathrm{d}X_s - \frac{1}{2}\int_0^T \frac{a^2X_s^2}{\sigma^2}\mathrm{d}s\right\}.$$

Writing $\mathbb{P}_T^a$ instead of $\mathbb{P}_T^Y$, we have

$$\frac{\mathrm{d}\,\mathbb{P}_T^Y}{\mathrm{d}\,\mathbb{P}_T^X} = \frac{\mathrm{d}\,\mathbb{P}_T^a}{\mathrm{d}\,\mathbb{P}_T^0}\left(= \frac{\mathrm{d}\,\mathbb{P}_T^a}{\mathrm{d}\,\mathbb{P}_T^{\sigma W}}\right) =: \mathscr{L}(a).$$

The MLE is then

$$\hat{a}_T = \frac{\int_0^T X_s \mathrm{d}X_s}{\int_0^T X_s^2 \mathrm{d}s} \overset{\text{plug in } X}{=} \frac{\int_0^T X_s(aX_s\mathrm{d}s + \sigma\mathrm{d}W_s)}{\int_0^T X_s^2\mathrm{d}s}$$

$$= a + \frac{\int_0^T X_s\sigma\mathrm{d}W_s}{\int_0^T X_s^2\mathrm{d}s} = a + \frac{M_T}{\sigma^{-2}\langle M\rangle_T}$$

with $M_t = \int_0^t X_s\sigma\mathrm{d}W_s$.

Problem 17:

(a) Show that a strictly stationary solution of $\mathrm{d}X_t = aX_t\mathrm{d}t + \sigma\mathrm{d}W_t$ exists if $a < 0$. It has the representation (cf. MA(∞)-representation of AR(1))

$$X_t = \sigma\int_{-\infty}^t e^{a(t-s)}\mathrm{d}\tilde{W}_s$$

where $(\tilde{W}_s,\, s \in \mathbb{R})$ is two-sided Brownian motion, i.e. $(\tilde{W}_t,\, t \geq 0)$ and $(\tilde{W}_{-t},\, t \geq 0)$ are independent Brownian motions.
If $a \geq 0$, then no weakly stationary solution exists.

(b) Consider the observations $(X_0, X_\Delta, \ldots, X_{n\Delta})$ with $\Delta > 0$ and $T = n\Delta$ (discrete observations). Estimate a by discretising the continuous-time MLE $\hat{a}_T$ and secondly by identifying $(X_{k\Delta},\, k \geq 0)$ as an AR(1)-process and using the Yule-Walker estimator.

3. Cox-Ingersoll-Ross (Bessel) process:
   It solves
$$\mathrm{d}X_t = (\vartheta_1 - \vartheta_2 X_t)\mathrm{d}t + \sigma\sqrt{X_t}\mathrm{d}W_t,$$
   $X^{(0)} > 0$; $\vartheta_1,\, \vartheta_2,\, \sigma > 0$.
   One can show that there is a unique strong solution (although diffusion coefficient is not Lipschitz at $X_t = 0$) with $X_t \geq 0$ for all t a.s. If $2\vartheta_1 > \sigma^2$, then even $X_t > 0$ for all t a.s.
   Assuming $2\vartheta_1 > \sigma^2$ and $2\vartheta_1^{(0)} > \sigma^2$ and considering $\mathbb{P}_T^{\vartheta}$ ($\vartheta = (\vartheta_1,\, \vartheta_2)$) as the law of (X_t) on $C([0,\, T])$ we have
$$\frac{\mathrm{d}\,\mathbb{P}_T^{\vartheta}}{\mathrm{d}\,\mathbb{P}_T^{\vartheta^{(0)}}} = \exp\left\{ \int_0^T \frac{(\vartheta_1 - \vartheta_1^{(0)}) - (\vartheta_2 - \vartheta_2^{(0)})X_s}{\sigma^2 X_s}\mathrm{d}X_s - \frac{1}{2}\int_0^T \frac{(\vartheta_1 - \vartheta_2 X_s)^2 - (\vartheta_1^{(0)} - \vartheta_2^{(0)}X_s)^2}{\sigma^2 X_s}\mathrm{d}s \right\}$$
   by Girsanov's theorem ($\sigma(X_s,\, s) > 0$).
   The MLE $\hat{\vartheta} = (\hat{\vartheta}_1,\, \hat{\vartheta}_2)$ is obtained from $\nabla_{\vartheta} \log\left(\frac{\mathrm{d}\,\mathbb{P}_T^{\vartheta}}{\mathrm{d}\,\mathbb{P}_T^{\vartheta^{(0)}}}\right) = 0$:
$$\hat{\vartheta}_1 = \frac{\int_0^T \frac{1}{X_s}\mathrm{d}X_s \int_0^T X_s\mathrm{d}s - \int_0^T 1\mathrm{d}s \int_0^T 1\mathrm{d}X_s}{\int_0^T \frac{1}{X_s}\mathrm{d}s \int_0^T X_s\mathrm{d}s - \left(\int_0^T 1\mathrm{d}s\right)^2},$$
$$\hat{\vartheta}_2 = \frac{\int_0^T 1\mathrm{d}s \int_0^T \frac{1}{X_s}\mathrm{d}X_s - \int_0^T 1\mathrm{d}X_s \int_0^T \frac{1}{X_s}\mathrm{d}s}{\int_0^T \frac{1}{X_s}\mathrm{d}s \int_0^T X_s\mathrm{d}s - \left(\int_0^T 1\mathrm{d}s\right)^2}.$$

4. General linear parametrisation:
   Consider
$$\mathrm{d}X_t = \langle \vartheta,\, b(X_t,\, t)\rangle\mathrm{d}t + \sigma(X_t,\, t)\mathrm{d}W_t,$$
   $X_0 = X^{(0)}$ with $\vartheta = (\vartheta_1, \ldots, \vartheta_k)^T \in \Theta \subseteq \mathbb{R}^k$ (unknown parameter), $b : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^k$ such that all conditions for Girsanov's theorem are satisfied; suppose $\mathbf{0} \in \Theta$ and $\sigma(x,\, t) > 0$. Then
$$\frac{\mathrm{d}\,\mathbb{P}_T^{\vartheta}}{\mathrm{d}\,\mathbb{P}_T^{0}} = \exp\left\{ \int_0^T \frac{\langle \vartheta,\, b(X_t,\, t)\rangle}{\sigma^2(X_t,\, t)}\mathrm{d}X_t - \frac{1}{2}\int_0^T \frac{\langle \vartheta,\, b(X_t,\, t)\rangle^2}{\sigma^2(X_t,\, t)}\mathrm{d}t \right\}.$$
   MLE is obtained from $\nabla_{\vartheta} \log\left(\frac{\mathrm{d}\,\mathbb{P}_T^{\vartheta}}{\mathrm{d}\,\mathbb{P}_T^{0}}\right)$:
$$\hat{\vartheta}_T^{\mathrm{MLE}} = \underbrace{\left(\int_0^T \left(\frac{b \cdot b^T}{\sigma^2}\right)(X_t,\, t)\mathrm{d}t\right)^{-1}}_{=:I_T \in \mathbb{R}^{k \times k}} \underbrace{\int_0^T \left(\frac{b}{\sigma^2}\right)(X_t,\, t)\mathrm{d}X_t}_{\in \mathbb{R}^k} \in \mathbb{R}^k,$$

provided the matrix is non-singular.
Under the law $\mathbb{P}_T^{\vartheta_0}$ we then obtain:

$$\hat{\vartheta}_T^{\mathrm{MLE}} = I_T^{-1}\Bigg(\int_0^T \frac{b(X_t,\, t)b(X_t,\, t)^T\vartheta_0 \mathrm{d}t + b(X_t,\, t)\sigma(X_t,\, t)\mathrm{d}W_t}{\sigma^2(X_t,\, t)}\Bigg)$$

$$= \vartheta_0 + I_T^{-1}\underbrace{\Bigg(\int_0^T \Big(\frac{b}{\sigma}\Big)(X_t,\, t)\mathrm{d}W_t\Bigg)}_{=:M_T} = \vartheta_0 + \underbrace{\langle M\rangle_T^{-1}}_{=I_T^{-1}} M_T.$$

If there is a deterministic sequence $A_T \in \mathbb{R}^{k\times k}$, A_T strictly positive definite, with $A_T^{-1}\langle M\rangle_T \overset{\mathbb{P}}{\to} E_k$ and the conditional Lindeberg condition is satisfied, then

$$A_T^{1/2}(\hat{\vartheta}_T - \vartheta_0) \overset{\text{under } \mathbb{P}_T^{\vartheta_0}}{\longrightarrow} \mathrm{N}(0,\, E_k).$$

If (X_t) is strictly stationary and ergodic, then we can take $A_T = T\cdot I_1$ where I_1 is the Fisher information matrix for observations $(X_t,\, t\in[0,\, 1])$. In particular, then $\hat{\vartheta}_T - \vartheta_0$ is of order $\mathcal{O}_{\mathbb{P}}(T^{-1/2})$.

Problem 18: Consider the stationary Ornstein-Uhlenbeck process

$$\mathrm{d}X_t = aX_t \mathrm{d}t + \sigma \mathrm{d}W_t,$$

$a < 0$, and the estimator

$$\hat{a}_T = \frac{\int_0^T X_t \mathrm{d}X_t}{\int_0^T X_t^2 \mathrm{d}t}.$$

Prove that $\sqrt{T}(\hat{a}_T - a)$ is asymptotically normal. By calculating the Fisher information prove that it is even efficient.

2.2 Nonparametric drift estimation

Suppose we observe a time-homogeneous diffusion process

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t,$$
$$X_0 = X^{(0)},$$

on $[0,\, T]$, we know the diffusion coefficient σ, but we do not know b and do not want to impose a particular parametric form on b. We merely assume that $x \mapsto b(x)$ has a certain Hölder smoothness:

$$|b(x) - b(y)| \le R|x-y|^\alpha$$

for all $x,\, y \in \mathbb{R}$, $\alpha \in (0,\, 1]$.
Idea: The drift $b(x)$ is the mean of the infinitesimal increment of X_t given $X_t = x$:

$$b(x) = \lim_{h\downarrow 0} \mathbb{E}\left[\frac{X_{t+h} - X_t}{h}\middle| X_t = x\right].$$

Hence we should use $\mathrm{d}X_t$ for estimating b.
$\rightsquigarrow$ Nadaraja-Watson-type estimator:

$$\hat{b}_{T,\,h}(x) = \frac{\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}X_t}{\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}t}.$$

Note:

$$\hat{b}_{T,\,h}(x) = \underbrace{\frac{\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)b(X_t)\mathrm{d}t}{\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}t}}_{=\int_0^T \tilde{\mathbf{1}}_{[x-h,\,x+h]}(X_t)b(X_t)\mathrm{d}t} + \frac{\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\sigma(X_t)\mathrm{d}W_t}{\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}t}$$

with $\tilde{\mathbf{1}}_{[x-h,\,x+h]}(X_t) \propto \mathbf{1}_{[x-h,\,x+h]}(X_t)$, $\int_0^T \tilde{\mathbf{1}}_{[x-h,\,x+h]}(X_t)\mathrm{d}t = 1$.
$\int_0^T \tilde{\mathbf{1}}_{[x-h,\,x+h]}(X_t)b(X_t)\mathrm{d}t$ is a convex combination of values $b(y)$ for $y \in [x - h,\, x+h]$, hence it lies in $[\min_{|y-x|\le h} b(y),\ \max_{|y-x|\le h} b(y)]$. Since $b \in C^\alpha$,

$$\Big|\int_0^T \tilde{\mathbf{1}}_{[x-h,\,x+h]}(X_t)b(X_t)\mathrm{d}t - b(x)\Big| \le Rh^\alpha,$$

which is a deterministic bound. It tends to zero when $h \downarrow 0$.

We look at the stochastic error term

$$\frac{\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\sigma(X_t)\mathrm{d}W_t}{\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}t}.$$

Suppose that (X_t) is stationary, then the numerator satisfies

$$\mathbb{E}\Big[\Big(\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\sigma(X_t)\mathrm{d}W_t\Big)^2\Big]$$

$$\overset{\text{Itô isometry}}{=} \int_0^T \mathbb{E}[\mathbf{1}_{[x-h,\,x+h]}(X_t)^2\sigma(X_t)^2]\mathrm{d}t$$

$$\overset{X\text{ stat.}}{=} T\,\mathbb{E}[\mathbf{1}_{[x-h,\,x+h]}(X_0)\sigma(X_0)^2]$$

$$\underset{\text{dens. of }X_0}{\overset{\mu\text{ inv. Lebesgue}}{=}} T\int_{x-h}^{x+h}\sigma^2(y)\mu(y)\mathrm{d}y \le 2Th||\sigma^2\mu||_\infty \sim Th.$$

Stationarity of X, existence of the invariant Lebesgue density μ and finiteness of σ^2 are necessary assumptions.
For the denominator:

$$\mathbb{E}[\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}t] \underset{\text{Fubini}}{\overset{X\text{ stat.,}}{=}} T\,\mathbb{E}[\mathbf{1}_{[x-h,\,x+h]}(X_0)]$$

$$\underset{\text{density}}{\overset{\mu\text{ invar.}}{=}} 2Th\bigg(\frac{1}{2h}\int_{x-h}^{x+h}\mu(y)\mathrm{d}y\bigg).$$

<u>Hope</u>: The denominator 'concentrates' around $2Th\mu(x)$ as $T\to\infty$, $h\to 0$ such that the stochastic error is of order (in probability) $\mathcal{O}_{\mathbb{P}}\Big(\frac{\sqrt{Th}}{Th}\Big) = \mathcal{O}_{\mathbb{P}}\Big(\frac{1}{\sqrt{Th}}\Big)$.

2.5 Proposition (Durrett: Stochastic Calculus ([2]))**.** *If*

$$G := \int_{-\infty}^{\infty}\frac{1}{\sigma^2(x)}\exp\bigg(\int_0^x \frac{2b}{\sigma^2}(z)\mathrm{d}z\bigg)\mathrm{d}x < \infty$$

and the SDE has a strong solution for any initial condition, then there is a stationary solution X of the SDE with invariant Lebesgue density

$$\mu(x) = \frac{1}{G\sigma^2(x)}\exp\bigg(\int_0^x\frac{2b}{\sigma^2}(z)\mathrm{d}z\bigg),\; x\in\mathbb{R}\,.$$

2.6 Proposition. *Suppose there are $A,\,\gamma>0$ such that $\operatorname{sgn}(x)\frac{2b}{\sigma^2}(x)\le-\gamma$ for all x with $|x|>A$, that b is bounded on $[-A,\,A]$ and $\underline{\sigma}^2 := \inf_{x\in\mathbb{R}}\sigma^2(x)>0$, then there is a stationary solution X of the SDE and for any function $f:\mathbb{R}\to\mathbb{R}$ with $\mathbb{E}[f(X_0)]=0$ and $f\in L^1(\mathbb{R})$ we have*

$$\mathbb{E}\Big[\Big(\int_0^T f(X_t)\mathrm{d}t\Big)^2\Big] \le ||f||_{L^1}^2(C+C'T)$$

with constants $C,\,C'>0$ depending only on $A,\,\gamma,\,\underline{\sigma}^2,\,\sup_{|x|\le A} b(x)$.

2.7 Remark. The condition $\operatorname{sgn}(x)\frac{2b}{\sigma^2}(x) \le -\gamma$ $(*)$ means for $x > 0$ that the drift is negative for $x > A$ and strong enough to push the diffusion process back to the direction of the origin such that an equilibrium can be obtained. For $x < 0$ the situation is symmetric. An easy example is the Ornstein-Uhlenbeck process with $b(x) = ax$ and $a < 0$.

Proof.

1. Condition $(*)$ implies $G < \infty$, using that $\frac{2b}{\sigma^2}$ is bounded in $[-A, A]$ and $\frac{1}{\sigma^2}$ is bounded on $\mathbb{R}$.

2. Find F such that $LF = f$ with the Markov generator

$$LF(x) = \frac{\sigma^2(x)}{2}F''(x) + b(x)F'(x).$$

Then by Itô's formula

$$\begin{aligned}\mathrm{d}F(X_t) &= F'(X_t)\mathrm{d}X_t + \frac{1}{2}F''(X_t)\mathrm{d}\langle X\rangle_t \\ &= \underbrace{(F'(X_t)b(X_t) + \frac{1}{2}F''(X_t)\sigma^2(X_t))}_{=LF(X_t)=f(t)}\mathrm{d}t + F'(X_t)\sigma(X_t)\mathrm{d}W_t.\end{aligned}$$

$$\Rightarrow \int_0^T f(X_t)\mathrm{d}t = F(X_T) - F(X_0) - \int_0^T F'(X_t)\sigma(X_t)\mathrm{d}W_t$$

$$\begin{aligned}\Rightarrow \mathbb{E}[(\int_0^T f(X_t)\mathrm{d}t)^2] &\le 3\Big(\mathbb{E}[F(X_T)^2] + \mathbb{E}[F(X_0)^2] + \mathbb{E}\Big[\Big(\int_0^T F'(X_t)\sigma(X_t)\mathrm{d}W_t\Big)^2\Big]\Big) \\ &\overset{X\text{ stat.}}{\underset{\text{Itô-iso.}}{=}} 6\,\mathbb{E}[F(X_0)^2] + 3T\,\mathbb{E}[F'(X_0)^2\sigma(X_0)^2].\end{aligned}$$

3. Check that

$$F(x) = \int_0^x \frac{2}{\sigma^2(y)\mu(y)}\Big(\int_{-\infty}^y f(z)\mu(z)\mathrm{d}z\Big)\mathrm{d}y$$

satisfies $LF = f$.

$$\begin{aligned}F'(x) &= \frac{2}{\sigma^2(x)\mu(x)}\int_{-\infty}^x f(z)\mu(z)\mathrm{d}z \\ &\overset{\text{prop. 2.5}}{=} 2\int_{-\infty}^x f(z)\frac{1}{\sigma^2(z)}\exp\Big(\int_x^z \frac{2b}{\sigma^2}(y)\mathrm{d}y\Big)\mathrm{d}z \\ &\overset{\int_{\mathbb{R}} f(z)\mu(z)\mathrm{d}z=0}{=} -2\int_x^\infty f(z)\frac{1}{\sigma2(z)}\exp\Big(\int_x^z \frac{2b}{\sigma^2}(y)\mathrm{d}y\Big)\mathrm{d}z.\end{aligned}$$

$$F''(x) = \frac{2f(x)}{\sigma^2(x)} + 2\int_{-\infty}^x f(z)\frac{1}{\sigma^2(z)}\Big(-\frac{2b}{\sigma^2}(x)\Big)\exp\Big(\int_x^z \frac{2b}{\sigma^2}(y)\mathrm{d}y\Big)\mathrm{d}z.$$

Hence

$$LF(x) = \left(\frac{\sigma^2}{2}F'' + bF'\right)(x) = (f(x) - b(x)F'(x)) + b(x)F'(x) = f(x).$$

4. Bound $F'(x)$, $F(x)$.
   For $x > 0$:

$$|F'(x)| \leq \frac{2}{\underline{\sigma}^2}\int_x^\infty |f(z)| \underbrace{\exp\left(\int_x^z \frac{2b}{\sigma^2}(y)\mathrm{d}y\right)}_{\sup_{x,\,z>0}(\dots)\leq C_1} \mathrm{d}z \leq C_2||f||_{L^1}.$$

   For $x < 0$ the same bound applies. We obtain $|F'(x)| \leq C_3||f||_{L^1}$ and thus

$$\mathbb{E}[F'(X_0)^2\sigma^2(X_0)] \leq C_3^2||f||_{L^1}^2 \int_{-\infty}^{\infty} \sigma^2(x)\mu(x)\mathrm{d}x \leq C_4||f||_{L^1}^2.$$

   The bound for $|F(x)|$ and then $\mathbb{E}[F(X_0)^2]$ follows in the same way.

□

Problem 19: Generalise this proposition by relaxing the conditions $\mathrm{sgn}(x)\frac{2b}{\sigma^2}(x) \leq -\gamma$, $\underline{\sigma}^2 > 0$. Follow the constants more explicitly.

Applying this proposition to the denominator, we obtain for diffusions satisfying its conditions:

$$\mathbb{E}[\left(\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t) - \mathbb{E}[\mathbf{1}_{[x-h,\,x+h]}(X_t)]\mathrm{d}t\right)^2] \\ \leq (C + C'T)||\mathbf{1}_{[x-h,\,x+h]}(X_t) - \underbrace{\mathbb{E}[\mathbf{1}_{[x-h,\,x+h]}(X_t)]}_{=\int_{x-h}^{x+h}\mu(x)\mathrm{d}x \leq 2h||\mu||_\infty}||_{L^1}^2 \leq (C + C'T)C_1h^2.$$

We have as $T \to \infty$, $h \downarrow 0$:

$$\left.\begin{aligned} &\mathbb{E}[\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}t] \geq C_2Th, \\ &\mathrm{Var}(\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}t) \leq C_3Th^2. \end{aligned}\right\} \Rightarrow \quad \begin{aligned} &\mathbb{E}[\tfrac{1}{Th}\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}t] \geq C_2 > 0, \\ &\mathrm{Var}(\tfrac{1}{Th}\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}t) \leq C_3T^{-1} \to 0. \end{aligned}$$

We thus have

$$\mathbb{P}(\frac{1}{Th}\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}t \geq \frac{C_2}{2}) \to 1.$$

Hence the stochastic error term is $\mathcal{O}_{\mathbb{P}}\left(\frac{\sqrt{Th}}{Th}\right) = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{Th}}\right)$ in the sense that

$$\sqrt{Th}\frac{\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\sigma(X_t)\mathrm{d}W_t}{\int_0^T \mathbf{1}_{[x-h,\,x+h]}(X_t)\mathrm{d}t}$$

is tight (i.e. bounded in probability). This implies the following theorem.

2.8 Theorem. *Suppose the SDE satisfies the conditions of the previous proposition. Then for the stationary solution (X_t) and a drift b with*

$$|b(x) - b(y)| \le R|x-y|^{\alpha}$$

we find

$$|\widehat{b}_{T,\,h}(x_0) - b(x_0)| \le Rh^{\alpha} + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{Th}}\right).$$

Hence, if $h = h_T \downarrow 0$, but $Th_T \to \infty$, then $\widehat{b}_{T,\,h}(x_0)$ is a consistent estimator of $b(x_0)$.

2.9 Corollary. *If we choose $h_T \sim T^{-\frac{1}{2\alpha+1}}$ (optimally in order), then we obtain*

$$|\widehat{b}_{T,\,h}(x_0) - b(x_0)| = \mathcal{O}_{\mathbb{P}}\left(T^{-\frac{\alpha}{2\alpha+1}}\right).$$

2.10 Remark. One can show that this rate $T^{-\frac{\alpha}{2\alpha+1}}$ is optimal in a minimax sense over α-Hölder continuous drifts b. For the most interesting Lipschitz case ($\alpha = 1$) the rate is $T^{-1/3}$ (compared to $T^{-1/2}$ for parametric problems).

2.3 Nonparametric volatility estimation with high frequency data

Consider the diffusion process

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t.$$

We observe $X_0,\, X_\Delta, \ldots, X_{N\Delta}$ ($\Delta << 1$).
Intuition: We look at X_0, X_Δ and at the increment:

$$\frac{X_\Delta - X_0}{\Delta} = \underbrace{\frac{1}{\Delta}\int_0^\Delta b(X_s)\mathrm{d}s}_{\sim b(X_0) \text{ if } b \text{ cts.}} + \underbrace{\frac{1}{\Delta}\int_0^\Delta \sigma(X_s)\mathrm{d}W_s}_{\mathbb{E}[\ldots]=0}.$$

To access σ, we look at the square:

$$\frac{(X_\Delta - X_0)^2}{\Delta} = \underbrace{\frac{1}{\Delta}\Big(\int_0^\Delta b(X_s)\mathrm{d}s\Big)^2}_{\sim\Delta}$$

$$+2\underbrace{\frac{1}{\Delta}\int_0^\Delta b(X_s)\mathrm{d}s}_{\sim 1}\underbrace{\int_0^\Delta \sigma(X_s)\mathrm{d}W_s}_{\sim\sqrt{\Delta}} + \underbrace{\frac{1}{\Delta}\Big(\int_0^\Delta \sigma(X_s)\mathrm{d}W_s\Big)^2}_{\mathbb{E}[\dots]\overset{\text{Itô}}{=}\frac{1}{\Delta}\mathbb{E}[\int_0^\Delta \sigma^2(X_s)\mathrm{d}s]\sim\sigma^2(X_0)}$$

Consider the process $\mathrm{d}B_t = \sigma \mathrm{d}W_t$, $\sigma > 0$ and the observations $B_0, B_\Delta, \dots, B_{N\Delta}$, $N\Delta = T$.

$$\hat{\sigma}^2 := \frac{1}{N}\sum_{n=0}^{N-1}\frac{(B_{(n+1)\Delta} - B_{n\Delta})^2}{\Delta} = \frac{1}{N}\sum_{n=0}^{N-1}\sigma^2 Y_n^2,$$

where (Y_n) are i.i.d. N(0, 1).
Then $\mathbb{E}[\hat{\sigma}] = \sigma^2$ and

$$\mathbb{E}[(\hat{\sigma} - \sigma^2)^2] = \mathbb{E}[(\frac{1}{N}\sum_{n=0}^{N-1}\sigma^2(Y_n^2 - 1))^2]$$

$$= \sigma^4\,\mathbb{E}[(\frac{1}{N}\sum_{n=0}^{N-1}(Y_n^2 - 1))^2] = \sigma^4\frac{1}{N}\underbrace{\mathrm{Var}(Y_0^2 - 1)}_{=2}.$$

$\Rightarrow \mathbb{E}[(\hat{\sigma} - \sigma^2)^2]^{1/2} = \frac{\sqrt{2}\sigma^2}{\sqrt{N}}$.
What has made the computation easy?

1. σ is constant,

2. increments are independent.

L^2 error bounds for the Florens-Zmirou estimator

2.11 Definition. Set $0 < m < M$ and define $\Theta(m, M) = \{\sigma \in C^1(\mathbb{R}) : m \le \inf_{x\in\mathbb{R}}\sigma(x) \le \sup_{x\in\mathbb{R}}\sigma(x) \le M, \sup_{x\in\mathbb{R}}|\sigma'(x)| \le M\}$. Note that each $\sigma \in \Theta$ satisfies the global Lipschitz and linear growth conditions, hence the corresponding equation

$$\begin{aligned} dX_t &= \sigma(X_t)dW_t, \\ X_0 &= X^{(0)} \in L^2, \end{aligned}$$

has a unique strong solution. For $\Delta > 0$ we observe a path $t \to X_t$ at equidistant times $0, \Delta, 2\Delta, ..., N\Delta = 1$. When $x \in \mathbb{R}$ is visited by the observed path (i.e.

$X_t = x$ for some $t \in (0,1)$) we define the Florens-Zmirou ([4]) estimator of the diffusion coefficient σ^2 by

$$\hat{\sigma}^2_{FZ}(x, h_\Delta) \;=\; \frac{\sum_{n=0}^{N-1} \mathbf{1}_{(|X_{n\Delta}-x|<h_\Delta)} \frac{1}{\Delta}(X_{(n+1)\Delta} - X_{n\Delta})^2}{\sum_{n=0}^{N-1} \mathbf{1}_{(|X_{n\Delta}-x|<h_\Delta)}}.$$

2.12 Definition. For any Borel set A define its occupation measure as $\mu(A) = \int_0^1 \mathbf{1}_A(X_s)ds$, i.e. the amount of time the path $(X_t)_{0\le t\le 1}$ stayed in A. Then the measure μ has a Lebesgue density L ([7], [1]) called the local time (chronological local time) of X at time one. For every positive Borel measurable function f the occupation formula $\int_0^1 f(X_s)ds = \int_{\mathbb{R}} f(x)L(x)dx$ holds.

2.13 Lemma. *For every $p > 2$ we have $\sup_{(\sigma,b)\in\Theta} \mathbb{E}[L^p(x)] < C_p$.*

Proof. By the Tanaka formula

$$L(x) = |X_1 - x| - |X_0 - x| - \int_0^1 \operatorname{sgn}(X_s - x)dX_s \le |X_1 - X_0| + \Big|\int_0^1 \operatorname{sgn}(X_s - x)dX_s\Big|.$$

Using the Burkholder-Davis-Gundy inequality (see stochastic analysis notes) we obtain

- $\mathbb{E}[|X_1 - X_0|^p] = \mathbb{E}\left[|\int_0^1 \sigma(X_s)dW_s|^p\right] \le \tilde{C}_p \,\mathbb{E}\left[|\int_0^1 \sigma^2(X_s)ds|^{\frac{p}{2}}\right] \le \tilde{C}_p M^p.$
- $\mathbb{E}\left[|\int_0^1 \operatorname{sgn}(X_s - x)dX_s|^p\right] \le \tilde{C}_p \,\mathbb{E}\left[|\int_0^1 \operatorname{sgn}^2(X_s - x)\sigma^2(X_s)ds|^{\frac{p}{2}}\right] \le \tilde{C}_p M^p.$

□

2.14 Theorem. *Consider an interval K, some positive $\nu > 0$ and let $\mathcal{L} = \{\inf_{x\in K} L_T(x) \ge \nu\}$, $h_\Delta \sim \Delta^{\frac{1}{3}}$. Then for every $x \in int(K)$ we have*

$$\sup_{\sigma\in\Theta} \mathbb{E}\left[\mathbf{1}_{\mathcal{L}} \cdot |\hat{\sigma}^2_{FZ}(x, h_\Delta) \wedge M^2 - \sigma^2(x)|^2\right] \le C\Delta^{\frac{2}{3}},$$

where the constant C depends only on the set K and level ν.

Notation: We will write $f_\sigma \lesssim g_\sigma$ (resp. $g_\sigma \gtrsim f_\sigma$) if we have $f_\sigma \le C\cdot g_\sigma$ for every $\sigma \in \Theta$ with some constant $C > 0$ depending only on K and ν.

Proof. (a) (Bias and martingale part) For $n = 0, ..., N-1$ define

$$\eta_n = \frac{1}{\Delta}\Big(\int_{n\Delta}^{(n+1)\Delta} \sigma(X_s)dW_s\Big)^2 - \frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta} \sigma^2(X_s)ds.$$

- $\mathbb{E}[\eta_n|\mathscr{F}_n] = 0$ and in particular $\mathbb{E}[\eta_n\eta_m] = 0$ for $n \ne m$.
- $\mathbb{E}[\eta_n^2|\mathscr{F}_n] \lesssim 1$. Indeed, by the Burkholder-Davies-Gundy inequality:

$$\begin{aligned}\Delta^2\,\mathbb{E}[\eta_n^2|\mathscr{F}_n] &\lesssim \mathbb{E}\Big[\Big(\int_{n\Delta}^{(n+1)\Delta} \sigma(X_s)dW_s\Big)^4|\mathscr{F}_n\Big] + \mathbb{E}\Big[\Big(\int_{n\Delta}^{(n+1)\Delta} \sigma^2(X_s)ds\Big)^2|\mathscr{F}_n\Big] \\ &\lesssim \mathbb{E}\Big[\Big(\int_{n\Delta}^{(n+1)\Delta} \sigma^2(X_s)ds\Big)^2|\mathscr{F}_n\Big] + \Delta^2 \lesssim \Delta^2.\end{aligned}$$

We decompose the estimation error into martingale and bias parts:

$$|\hat{\sigma}^2_{FZ}(x,h_\Delta)-\sigma^2(x)| =$$

$$= \Big| \frac{\sum_{n=0}^{N-1} \mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\big(\frac{1}{\Delta}\big(\int_{n\Delta}^{(n+1)\Delta}\sigma(X_s)dW_s\big)^2-\sigma^2(x)\big)}{\sum_{n=0}^{N-1}\mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}} \Big|$$

$$\lesssim \underbrace{\Big|\frac{\sum_{n=0}^{N-1}\mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\eta_n}{\sum_{n=0}^{N-1}\mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}}\Big|}_{M_{x,\Delta}} + \underbrace{\Big|\frac{\sum_{n=0}^{N-1}\mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\big(\frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta}\sigma^2(X_s)ds-\sigma^2(x)\big)}{\sum_{n=0}^{N-1}\mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}}\Big|}_{B_{x,\Delta}}.$$

(b) (The ”good” high-probability set) Denote by $\omega(\Delta)$ the modulus of continuity of the path $(X_t)_{t\in(0,1)}$, i.e.

$$\omega(\Delta) = \sup_{\substack{0\le s,t\le 1\\ |t-s|<\Delta}} |X_t - X_s|.$$

Set $0<\epsilon<1/6$ and let $\alpha = 3/2-3\epsilon \in (1, 3/2)$. Define the event $\mathcal{R} = \{\omega(\Delta) < h_\Delta^\alpha\}$. Then for every $p>1$ holds

$$\mathbb{P}(\mathcal{R}^c) \lesssim h_\Delta^{-p\alpha}\Big(\Delta\log\big(2\Delta^{-1}\big)\Big)^{\frac{p}{2}} \lesssim \Delta^{\epsilon p}\log\big(2\Delta^{-1}\big)^{\frac{p}{2}}. \tag{$*1$}$$

In particular $\mathbb{P}(\mathcal{R}^c)\lesssim \Delta^{2/3}$ for p big enough.

Proof. (Proof of ($*1$))

Set $p>0$. By Markov's inequality we just have to show that there exists a constant C_p depending only on p and the upper bound of σ, such that

$$\mathbb{E}[\omega(\Delta)^p] \le C_p\Big(\Delta\log\big(\frac{2T}{\Delta}\big)\Big)^{\frac{p}{2}}. \tag{$*2$}$$

- ($*2$) holds for Brownian motion - [3].
- Let $dX_t = \sigma(X_t)dW_t$. By the Dambis-Dubin-Schwarz theorem $X_t = B_{\int_0^t\sigma^2(X_s)ds}$ for some Brownian motion B. Consequently

$$|X_t - X_s| = \big|B_{\int_0^t\sigma^2(X_s)ds} - B_{\int_0^s\sigma^2(X_s)ds}\big| \le \omega^B(|t-s|M^2)$$

□

(c) (Bias part error) When $|X_{n\Delta}-x|<h_\Delta$ we have

$$\begin{aligned}\frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta}|\sigma^2(X_s)-\sigma^2(x)|ds &\lesssim \frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta}|X_s-x|ds\\ &\le \frac{1}{\Delta}\int_{n\Delta}^{(n+1)\Delta}|X_s-X_{n\Delta}|ds + |X_{n\Delta}-x|\\ &\lesssim \omega(\Delta)+h_\Delta.\end{aligned}$$

Consequently $\mathbf{1}_{\mathcal{R}}\cdot B_{x,\Delta}\lesssim h_\Delta$.

(d) (Martingale part error) Denote $\sum_{n=0}^{N-1} \mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}} = N(x,h_\Delta)$. Then, on the event $\mathcal{R}$ we have

$$\Big|\frac{N(x,h_\Delta)}{Nh_\Delta} - \frac{1}{h_\Delta}\int_{x-h_\Delta}^{x+h_\Delta} L(z)dz\Big| \lesssim \frac{1}{h_\Delta}\int_{\{h_\Delta-h_\Delta^\alpha \le |z-x|<h_\Delta+h_\Delta^\alpha\}} L(z)dz. \tag{$*3$}$$

Indeed by the triangle inequality

$$\begin{aligned}
\Big|\frac{1}{N}\sum_{n=0}^{N-1}\mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}} - \int_0^1 \mathbf{1}_{\{|X_s-x|<h_\Delta\}}ds\Big| &\le \\
&\le \sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta}\big|\mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}} - \mathbf{1}_{\{|X_s-x|<h_\Delta\}}\big|ds \\
&= \sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta}\mathbf{1}_{\{h_\Delta\le|X_s-x|<h_\Delta+\omega(\Delta)\}}ds \\
&\quad + \sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta}\mathbf{1}_{\{h_\Delta-\omega(\Delta)\le|X_s-x|<h_\Delta\}}ds \\
&= \int_0^1 \mathbf{1}_{\{h_\Delta-h_\Delta^\alpha\le|X_s-x|<h_\Delta+h_\Delta^\alpha\}}ds \\
&= \int_{\{h_\Delta-h_\Delta^\alpha\le|z-x|<h_\Delta+h_\Delta^\alpha\}} L(z)dz.
\end{aligned}$$

Denote for simplicity $\{z:\ h_\Delta - h_\Delta^\alpha \le |z-x| < h_\Delta + h_\Delta^\alpha\} = A$ and observe that the Lebesgue measure of A is $4h_\Delta^\alpha$. Using first Markov's and next Hölder's inequalities we obtain

$$\begin{aligned}
\mathbb{P}\Big(\frac{1}{h_\Delta}\int_A L(z)dz \ge c\Big) &\lesssim \mathbb{E}\Big[\frac{1}{h_\Delta^p}\big(\int_A L(z)dz\big)^p\Big] \\
&\lesssim \frac{h_\Delta^{\alpha(p-1)}}{h_\Delta^p}\int_A \mathbb{E}[L^p(z)]dz \lesssim h_\Delta^{(\alpha-1)p} \lesssim \Delta^{\frac{2}{3}}
\end{aligned}$$

for p big enough. Consequently there exists a high probability event $Q \subseteq \mathcal{R}$, $\mathbb{P}(Q^c) \lesssim \Delta^{2/3}$, such that $\frac{N(x,h_\Delta)}{Nh_\Delta}$ is bounded from below on $Q \cap \mathcal{L}$. Now using martingale properties of η_n we obtain:

$$\begin{aligned}
\mathbb{E}\Big[\mathbf{1}_{Q\cap\mathcal{L}}\cdot M^2_{x,\Delta}\Big] &= \mathbb{E}\Big[\Big(\frac{1}{N(x,h_\Delta)}\sum_{n=0}^{N-1}\mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\eta_n\Big)^2\cdot\mathbf{1}_{Q\cap\mathcal{L}}\Big] \\
&\lesssim \frac{1}{N^2h_\Delta^2}\,\mathbb{E}\Big[\Big(\sum_{n=0}^{N-1}\mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\eta_n\Big)^2\mathbf{1}_{Q\cap\mathcal{L}}\Big] \\
&\lesssim \frac{1}{N^2h_\Delta^2}\,\mathbb{E}\Big[\sum_{n,m=0}^{N-1}\mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\mathbf{1}_{\{|X_{m\Delta}-x|<h_\Delta\}}\eta_n\eta_m\Big] \\
&= \frac{1}{N^2h_\Delta^2}\,\mathbb{E}\Big[\sum_{n=0}^{N-1}\mathbf{1}_{\{|X_{n\Delta}-x|<h_\Delta\}}\,\mathbb{E}[\eta_n^2|\mathscr{F}_n]\Big] \\
&\lesssim \frac{1}{N^2h_\Delta^2}\,\mathbb{E}\Big[N(x,h_\Delta)\Big].
\end{aligned}$$

Finally

$$\begin{aligned}
\frac{1}{Nh_\Delta}\,\mathbb{E}\Big[N(x,h_\Delta)\Big] &\lesssim \frac{1}{Nh_\Delta}\,\mathbb{E}\Big[N(x,h_\Delta)\mathbf{1}_{\mathcal{R}}\Big]+\frac{1}{Nh_\Delta}\,\mathbb{E}\Big[N(x,h_\Delta)\mathbf{1}_{\mathcal{R}^c}\Big] \\
&\lesssim \mathbb{E}\Big[\frac{1}{h_\Delta}\int_{x-h_\Delta}^{x+h_\Delta}L(z)dz+\frac{1}{h_\Delta}\int_A L(z)dz\Big]+h_\Delta^{-1}\,\mathbb{P}(\mathcal{R}^c) \\
&\lesssim \frac{1}{h_\Delta}\int_{(x-h_\Delta,x+h_\Delta)\cup A}\mathbb{E}[L(z)]dz+h_\Delta^{-1}\Delta^{\frac{2}{3}} \\
&\lesssim 1.
\end{aligned}$$

(e) (Conclusion) We have shown

$$\mathbb{E}[\mathbf{1}_{\mathcal{L}\cap Q}\cdot|\sigma^2_{FZ}(x,h_\Delta)-\sigma^2(x)|^2]\lesssim\mathbb{E}[\mathbf{1}_{\mathcal{L}\cap Q}\cdot M^2_{x,\Delta}+\mathbf{1}_{\mathcal{R}}\cdot B^2_{x,\Delta})]\lesssim\frac{1}{Nh_\Delta}+h_\Delta^2\sim\Delta^{\frac{2}{3}}.$$

Furthermore

$$\mathbb{E}[\mathbf{1}_{\mathcal{L}\cap Q^c}\cdot|\sigma^2_{FZ}(x,h_\Delta)\wedge M^2-\sigma^2(x)|^2]\lesssim\mathbb{P}(Q^c)\lesssim\Delta^{\frac{2}{3}}.$$

□

2.15 Corollary. *Let*

$$\Theta^*=\Theta(m,M)\times\{b\in C(\mathbb{R}):b\text{ is Lipschitz and }\sup_{x\in\mathbb{R}}b(x)\le M\}.$$

For $(\sigma,b)\in\Theta^*$ *consider a diffusion* Y *defined by the SDE* $dY_t=b(Y_t)dt+\sigma(Y_t)dW_t$, $Y_0=x_0$. *Then for the event* $\mathcal{L}$ *and* x *defined as before, given that* $h_\Delta\sim\Delta^{\frac{1}{3}}$, *we have*

$$\sup_{(\sigma,b)\in\Theta^*}\mathbb{E}_{\sigma,b}\big[\mathbf{1}_{\mathcal{L}}\cdot|\sigma^2_{FZ}(x,h_\Delta)\wedge M^2-\sigma^2(x)|\big]\le C(\mathcal{L})\Delta^{\frac{1}{3}}.$$

Proof. Using boundedness of the coefficients b and σ one can easily verify the assumptions of the Girsanov's theorem. The laws of the diffusions X and Y on $C([0,1])$ are equivalent and

$$\frac{d\,\mathbb{P}_Y}{d\,\mathbb{P}_X}(X) = \exp\Big(\int_0^1 \frac{b(X_s)}{\sigma^2(X_s)}dX_s - \frac{1}{2}\int_0^1 \frac{b^2(X_s)}{\sigma^2(X_s)}ds\Big)$$
$$= \exp\Big(\int_0^1 \frac{b(X_s)}{\sigma(X_s)}dW_s - \frac{1}{2}\int_0^1 \frac{b^2(X_s)}{\sigma^2(X_s)}ds\Big).$$

Denote $\mathbf{1}_{\mathcal{L}}\cdot|\sigma^2_{FZ}(x,h_\Delta)\wedge M^2 - \sigma^2(x)| = \mathcal{E}_{x,\Delta}$. By Cauchy-Schwarz we obtain

$$\begin{aligned}
\mathbb{E}_{\sigma,b}\big[\mathcal{E}_{x,\Delta}\big] &= \mathbb{E}\Big[\mathcal{E}_{x,\Delta}\frac{dP_Y}{dP_X}(X)\Big] \\
&= \mathbb{E}\Big[\mathcal{E}_{x,\Delta}\exp\Big(\int_0^1 \frac{b(X_s)}{\sigma(X_s)}dW_s - \frac{1}{2}\int_0^1 \frac{b^2(X_s)}{\sigma^2(X_s)}ds\Big)\Big] \\
&\leq \mathbb{E}\Big[\mathcal{E}_{x,\Delta}\exp\Big(\int_0^1 \frac{b(X_s)}{\sigma(X_s)}dW_s\Big)\Big] \\
&\leq \mathbb{E}[\mathcal{E}^2_{x,\Delta}]^{\frac{1}{2}}\,\mathbb{E}\Big[\exp\Big(2\int_0^1 \frac{b(X_s)}{\sigma(X_s)}dW_s\Big)\Big]^{\frac{1}{2}}.
\end{aligned}$$

We just have to argue that $\mathbb{E}\Big[\exp\Big(\int_0^1 \frac{2b(X_s)}{\sigma(X_s)}dW_s\Big)\Big]$ is uniformly bounded. Since

$$\mathbb{E}\Big[\exp\Big(\int_0^1 2(b\sigma^{-1})^2(X_s)ds\Big)\Big] < \infty$$

by the Novikov's condition the process $M_t = \exp\Big(\int_0^t 2(b\sigma^{-1})(X_s)dW_s - \int_0^t 2(b\sigma^{-1})^2(X_s)ds\Big)$ is a martingale and consequently

$$\mathbb{E}\Big[\exp\Big(\int_0^1 2(b\sigma^{-1})(X_s)dW_s\Big)\Big] = \mathbb{E}\Big[\exp\Big(\int_0^1 2(b\sigma^{-1})^2(X_s)ds\Big)\Big].$$

□

2.16 Theorem. *(Florens-Zmirou, 1993)*
Let X satisfy

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad t\in[0,1],$$

where b is a bounded function with two bounded derivatives, σ has three continuous and bounded derivatives and furthermore $m < \sigma < M$ for some positive $0 < m < M$. If Nh^3_Δ tends to zero, then

$$\sqrt{Nh_\Delta}\Big(\frac{\sigma_{FZ}(x,h_\Delta)}{\sigma^2(x)} - 1\Big) \xrightarrow{D} L(x)^{-1/2}Z,$$

where Z is a standard normal variable independent of $L(x)$.

2.4 Introduction to high-frequency statistics

Setting: Fix $T > 0$; $X = (X_t)_{0 \le t \le T}$.

$X_t = x_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s$, $0 \le t \le T$,

$x_0 \in \mathbb{R}$, $W = (W_t)_{0 \le t \le T}$ standard Brownian motion,

(A0) $b : [0, T] \to \mathbb{R}$, $\sigma : [0, T] \to \mathbb{R}$ are deterministic functions; b and σ are bounded.

Data: $n \ge 1$, $\mathscr{G}_n = (0 = t_{0,n} < t_{1,n} < \cdots < t_{n,n} = T)$

(particular case: $t_{i,n} = \frac{iT}{n}$).

$|\mathscr{G}_n| = \max_{1 \le i \le n} |t_{i,n} - t_{i-1,n}|$.

We observe $X_0 = X_{t_{0,n}}, \dots, X_{t_{n,n}} = X_T$, which is equivalent to the observations X_0, $\Delta X_{t_{i,n}} = X_{t_{i,n}} - X_{t_{i-1,n}}$; $i = 1, \dots, n$.

$\Delta t_{i,n} = t_{i,n} - t_{i-1,n}$.

Objective: Pick $g : [0, T] \to \mathbb{R}$. Estimate $\Lambda(g) = \int_0^T g(s)\sigma_s^2 \mathrm{d}s$.

2.17 Examples.

(1) $g(t) = 1$. $\Lambda(1)$ is called integrated volatility.

(2) $g_h(t) = \frac{1}{h}\mathbf{1}_{[t_0 - h, t_0]}(t)$, $h > 0$.

$\Lambda(g_h) = \frac{1}{h}\int_{t_0-h}^{t_0} \sigma_s^2 \mathrm{d}s \approx \sigma_{t_0}^2$ for $h \downarrow 0$ if σ^2 is smooth.

Note: $\mathscr{L}(X_t) = \mathrm{N}(x_0 + \int_0^t b_s \mathrm{d}s, \int_0^t \sigma_s^2 \mathrm{d}s)$,

$\mathscr{L}(\Delta X_{t_{i,n}}) = \mathrm{N}(\int_{\Delta t_{i,n}} b_s \mathrm{d}s, \int_{\Delta t_{i,n}} \sigma_s^2 \mathrm{d}s)$ and the $\Delta X_{t_{i,n}}$ are independent.

Problem 20: $b_s = b$, $\sigma_s = \sigma > 0$ (constant), $\vartheta = (b, \sigma^2)$.

(i) Compute the MLE in that setting and find conditions on $\mathscr{G}_n$ in order to have consistency.

(ii) Assume that b is known. Compute the Fisher information for the parameter σ^2.

$$\Delta X_{t_{i,n}} \stackrel{\mathrm{d}}{=} \int_{\Delta t_{i,n}} b_s \mathrm{d}s + \left(\int_{\Delta t_{i,n}} \sigma_s^2 \mathrm{d}s \right)^{1/2} \xi_{i,n} \text{ where } \xi_{i,n} \stackrel{\mathrm{d}}{=} \mathrm{N}(0, 1).$$

(A1) $b = 0$.

$(\Delta X_{t_{i,n}})^2 = \int_{\Delta t_{i,n}} \sigma_s^2 \mathrm{d}s \xi_{i,n}^2 \approx \sigma_{t_{i-1}}^2 \Delta t_{i,n}$.

$\rightsquigarrow \widehat{\Lambda}_n(g) = \sum_{i=1}^n g(t_{i-1,n})(\Delta X_{t_{i,n}})^2$.

Error decomposition:

$$\widehat{\Lambda}_n(g)-\Lambda(g) = \underbrace{\sum_{i=1}^{n} g(t_{i-1,n})(\overbrace{(\Delta X_{t_{i,n}})^2 - \int_{\Delta t_{i,n}} \sigma_s^2 \mathrm{d}s}^{=:\eta_{i,n}})}_{=:M_n} + \underbrace{\sum_{i=1}^{n} \int_{\Delta t_{i,n}} \sigma_s^2 (g(t_{i-1,n}) - g(s))\mathrm{d}s}_{=:R_n}.$$

Look at R_n. Define

$$P_{\mathscr{G}_n} g(t) = \sum_{i=1}^{n} g(t_{i-1,n}) \mathbf{1}_{(t\in\Delta t_{i,n})}.$$

Then we have

$$R_n = \sum_{i=1}^{n} \int_{\Delta t_{i,n}} \sigma_s^2 (g(t_{i-1,n}) - g(s))\mathrm{d}s = \int_0^T \sigma_s^2 (P_{\mathscr{G}_n} g(s) - g(s))\mathrm{d}s.$$

We give a very rough bound:

$$|R_n| \leq ||\sigma^2||_{L^\infty} \underbrace{||P_{\mathscr{G}_n} g - g||_{L^1}}_{\mathscr{M}(g,\mathscr{G}_n)}.$$

For M_n:

$$\mathbb{E}[(\Delta X_{t_{i,n}})^2] = \int_{\Delta t_{i,n}} \sigma_s^2 \mathrm{d}s,$$

$$\mathbb{E}[M_n^2] = \sum_{i=1}^{n} g(t_{i-1,n})^2\, \mathbb{E}[\eta_{i,n}^2],$$

$$\mathbb{E}[\eta_{i,n}^2] = \mathbb{E}[((\Delta X_{t_{i,n}})^2 - \int_{\Delta t_{i,n}} \sigma_s^2 \mathrm{d}s)^2] = \Big(\int_{\Delta t_{i,n}} \sigma_s^2 \mathrm{d}s\Big)^2 \underbrace{\mathbb{E}[(\xi_{i,n}^2 - 1)^2]}_{=2}.$$

Hence,

$$\mathbb{E}[M_n^2] = 2\sum_{i=1}^{n} g(t_{i-1,n})^2 \Big(\int_{\Delta t_{i,n}} \sigma_s^2 \mathrm{d}s\Big)^2 \leq 2||\sigma^4||_{L^\infty} \underbrace{\sum_{i=1}^{n} g(t_{i-1,n})^2 (\Delta t_{i,n})^2}_{\tilde{\mathscr{M}}(g,\mathscr{G}_n)^2}.$$

2.18 Proposition. *Work under (A0) and (A1). Then*

$$\mathbb{E}[(\widehat{\Lambda}_n(g) - \Lambda(g))^2] \leq C||\sigma^4||_{L^\infty}(\mathscr{M}(g,\mathscr{G}_n)^2 + \tilde{\mathscr{M}}(g,\mathscr{G}_n)^2)$$

(with C constant).

Consider
(A2(α)) $|g(t)-g(s)| \leq R|t-s|^\alpha$ (for $0 < \alpha \leq 1$) and $|g(t)| \leq R$ for all $t \in [0, T]$.
Then

$$\mathscr{M}(g,\mathscr{G}_n) = \sum_{i=1}^{n} \int_{\Delta t_{i,n}} |g(t_{i,n}) - g(s)|\mathrm{d}s \leq R\sum_{i=1}^{n} (\Delta t_{i,n})^{\alpha+1} \leq TR|\mathscr{G}_n|^\alpha,$$

$$\tilde{\mathscr{M}}(g,\mathscr{G}_n)^2 \leq R^2 T|\mathscr{G}_n|.$$

2.19 Corollary. *Assume moreover A2(α). Then*

$$\mathbb{E}[(\widehat{\Lambda}_n(g) - \Lambda(g))^2] \le C_T ||\sigma^4||_{L^\infty} |\mathscr{G}_n|^{1\wedge 2\alpha}.$$

2.20 Remark. $|\mathscr{G}_n| \le \frac{c}{n} \rightsquigarrow$ rate $n^{-(1\wedge 2\alpha)}$.

Towards a CLT: We want

$$\sqrt{n}(\widehat{\Lambda}_n(g) - \Lambda(g)) = \sqrt{n}M_n + \underbrace{\sqrt{n}R_n}_{\overset{!}{\to}0}.$$

Take (A3) $|\mathscr{G}_n|^\alpha = o\Big(\frac{1}{\sqrt{n}}\Big)$.

$$\sqrt{n}M_n = \sum_{i=1}^n g(t_{i-1,\,n})\sqrt{n} \int\limits_{\Delta t_{i,\,n}} \sigma_s^2 \mathrm{d}s(\xi_{i,\,n}^2 - 1).$$

Recall the CLT for independent random variables with Lindeberg condition: Let $\tilde{\eta}_{1,\,n}, \tilde{\eta}_{2,\,n}, \dots, \tilde{\eta}_{n,\,n}$ be independent random variables such that

(i) $\mathbb{E}[\tilde{\eta}_{i,\,n}] = 0$,

(ii) $v_n = \sum_{i=1}^n \mathbb{E}[\tilde{\eta}_{i,\,n}^2]$,

(iii) $\exists c > 0$ such that $\frac{1}{v_n}\sum_{i=1}^n \mathbb{E}[\tilde{\eta}_{i,\,n}^2 \mathbf{1}_{(\tilde{\eta}_{i,\,n}) > c\sqrt{v_n}}] \to 0$.

Then

$$\frac{1}{\sqrt{v_n}} \sum_{i=1}^n \tilde{\eta}_{i,\,n} \xrightarrow{\mathrm{d}} \mathrm{N}(0,\, 1).$$

Choose $\tilde{\eta}_{i,\,n}$ such that $\sqrt{n}M_n = \sum\limits_{i=1}^n \tilde{\eta}_{i,\,n}$. If v_n converge to some v^2, then

$$\sqrt{n}M_n \xrightarrow{\mathrm{d}} \mathrm{N}(0,\, v^2).$$

Identify v_n:

$$v_n = \sum_{i=1}^n \mathbb{E}[\tilde{\eta}_{i,\,n}^2] = 2n \sum_{i=1}^n g(t_{i-1,\,n})^2 \underbrace{\Big(\int\limits_{\Delta t_{i,\,n}} \sigma_s^2 \mathrm{d}s\Big)^2}_{\approx \sigma^4_{t_{i-1,\,n}}(\Delta t_{i,\,n})^2} \to 2\cdot T \int_0^T g(s)^2 \sigma_s^4 \mathrm{d}s$$

if σ^2 is continuous and provided

(A4) $\sum\limits_{i=1}^n |n\Delta t_{i,\,n} - T|\Delta t_{i,\,n} \to 0$ and

(A5) $\sigma_s^2 > 0$ for all s; $\{t : g(t)^2 > 0\}$ contains an open set.

2.21 Theorem. *Work under (A0)-(A5). Then*

$$\sqrt{n}(\widehat{\Lambda}_n(g) - \Lambda(g)) \xrightarrow{d} \mathrm{N}\Big(0,\, 2\cdot T\int\limits_0^T g^2(s)\sigma_s^4 \mathrm{d}s\Big).$$

Problem 21: What can you say if $g = g_h(t) = \frac{1}{h}\mathbf{1}_{[t_0 - h,\, t_0]}(t)$?

2.5 Volatility estimation from high frequency data in a nutshell

2.5.1 Direct observation model

Consider the semi-martingale (continuous semi-martingale if there are no jumps)

$$X_t = X_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s + \text{Jumps}. \qquad \text{(SM/CSM)}$$

Main objective in (CSM): $\langle X, X\rangle_1 = \int_0^1 \sigma_s^2 \mathrm{d}s$.

Functional stable CLT for realised volatility in (CSM) (see Jacod):

$$\sqrt{n}\Big(\sum_{i=1}^{\llcorner n-t \lrcorner} (X_{\frac{i}{n}} - X_{\frac{i-1}{n}})^2 - \int_0^t \sigma_s^2 \mathrm{d}s\Big) \overset{\text{st.}}{\to} \int_0^t \sqrt{2}\sigma_s^2 \mathrm{d}B_s$$

with B_s Brownian motion and $B \perp W$. 'st.' denotes stable convergence in law.

$$\Rightarrow \sqrt{n}\Big(\sum_{i=1}^{n} (\Delta_i^n X)^2 - \int_0^1 \sigma_s^2 \mathrm{d}s\Big) \overset{\text{st.}}{\to} \mathrm{N}(0,\, 2\int_0^1 \sigma_s^4 \mathrm{d}s).$$

Consider the case

$$X_t = X_0 + \int_0^t \sigma \mathrm{d}W_s. \qquad \text{(M)}$$

In (M) for $t_i = \frac{i}{n}$: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (\sqrt{n}\Delta_i^n X)^2$.

In (M) for general t_i: $\hat{\sigma}^2 = \sum_{i=1}^n \alpha_i (\sqrt{n}\Delta_i^n X)^2$.

We would like to have $\sum_{i=1}^n \alpha_i \overset{(*)}{=} 1$ such that $\hat{\sigma}^2$ is unbiased.

The variance is $\sum_{i=1}^n \alpha_i^2 2\sigma^4 n^2 (\Delta t_i)^2$. We try to minimise it:

$$\frac{\mathrm{d}}{\mathrm{d}\alpha_j}\Big(\sum_{i=1}^n \alpha_i^2 2\sigma^4 n^2 (\Delta t_i)^2 + \lambda(\sum_{i=1}^n \alpha_i - 1)\Big) = 0$$

$$\Rightarrow \alpha_j = \frac{-\lambda}{4\sigma^4 n^2 (\Delta t_j)^2} = \frac{1}{n^2(\Delta t_j)^2 G}$$

with $G = \sum_{i=1}^n \frac{1}{n^2(\Delta t_i)^2}$ (calculate using $(*)$).

If we now set $I_{n,i} = \frac{1}{2\sigma^4(\Delta t_i)^2 n^2}$; $I_n = \sum_{i=1}^n I_{n,i}$, we obtain

$$\mathrm{Var}(\hat{\sigma}^2) = \sum_{i=1}^n \frac{1}{n^4(\Delta t_i)^4 G^2} 2\sigma^4 n^2 (\Delta t_i)^2 = 2\sigma^4 G^{-1} = I_n^{-1}.$$

Estimating spot volatility in (CSM)
Set K_n to be the size of the window for relevant observations around $s \in (0, 1)$. Then

$$\hat{\sigma}_s^2 = \frac{n}{2K_n + 1} \sum_{i=\llcorner sn \lrcorner - K_n}^{\llcorner sn \lrcorner + K_n} (\Delta_i^n X)^2.$$

For the bias we compute

$$\mathbb{E}[\hat{\sigma}_s^2 - \sigma_s^2] \approx \frac{n}{2K_n + 1} \sum_{i=\llcorner sn \lrcorner - K_n}^{\llcorner sn \lrcorner + K_n} (\sigma_{\frac{i}{n}}^2 n^{-1} - \sigma_s^2 n^{-1}) \approx K_n^{-1} \sum_{i=\llcorner sn \lrcorner - K_n}^{\llcorner sn \lrcorner + K_n} (\sigma_{\frac{i}{n}}^2 - \sigma_s^2).$$

We look at the modulus of continuity to characterise the smoothness of σ and assume

$$\sup_{\tau \in [s,\, t]} |\sigma_\tau^2 - \sigma_s^2| \leq |t - s|^\alpha.$$

Then

$$\mathbb{E}[\hat{\sigma}_s^2 - \sigma_s^2] \approx K_n^{-1} \sum_{j=1}^{K_n} \left(\frac{j}{n}\right)^\alpha \approx \frac{K_n^\alpha}{n^\alpha}.$$

$$\mathrm{Var}(\hat{\sigma}_s^2) \approx \frac{n^2}{4K_n^2} \sum_i 2\sigma_{\frac{i}{n}}^4 n^{-2} \approx K_n^{-1} 2\sigma_s^4.$$

Bias and variance are balanced if $K_n \propto n^{\frac{2\alpha}{2\alpha+1}}$; then

$$(\hat{\sigma}_s^2 - \sigma_s^2) = \mathcal{O}_{\mathbb{P}}\Big(n^{\frac{-\alpha}{2\alpha+1}}\Big).$$

2.5.2 Noisy observation model

The model is

$$Y_{t_i} = X_{t_i} + \varepsilon_i, \qquad i = 0, \ldots, n.$$

We assume $\varepsilon \perp X$, ε_i i.i.d., $\mathbb{E}[\varepsilon_i] = 0$, $\mathrm{Var}(\varepsilon_i) = \eta^2$ and $\mathbb{E}[\varepsilon_i^8] < \infty$. We observe

$$\Delta_i^n Y = \underbrace{\Delta_i^n X}_{\mathcal{O}_{\mathbb{P}}(n^{-1/2})} + \underbrace{\varepsilon_i - \varepsilon_{i-1}}_{\mathcal{O}_{\mathbb{P}}(1)}$$

and get

$$\mathbb{E}[\sum_{i=1}^n (\Delta_i^n Y)^2] = 2n\eta^2 + o(n),$$
$$\mathbb{E}[\Delta_i^n Y \Delta_{i-1}^n Y] = -\eta^2.$$

Spectral volatility estimation
Idea: split $[0, 1]$ in bins $[kh,\, (k+1)h)$, $k = 0, \ldots, h^{-1} - 1$. Approximate σ_t:

$$\sigma_t = \sigma_{kh} \mathbf{1}_{[kh,\,(k+1)h)}(t).$$

Take the family of functions

$$\Phi_{jk}(t) = \sqrt{\frac{2}{h}} \sin(j\pi h^{-1}(t-(k-1)h))\mathbf{1}_{[kh,\,(k+1)h)}(t), \qquad j \geq 1.$$

Φ_{jk} are orthonormal: $\langle \Phi_{jk},\, \Phi_{mk} \rangle = \delta_{jm}$.
Define the spectral statistics

$$S_{jk} = \sum_{i=1}^{n} Y_{t_i} \Phi_{jk}(t_i), \qquad j \geq 1.$$

Summation by parts decomposition yields

$$S_{jk} \approx \sum_{i=1}^{n} X_{t_i} \Phi_{jk}(t_i) - \sum_{i=1}^{n-1} \varepsilon_i {\Phi'}_{jk}(t_i)\Delta t_i.$$

Assume additionally $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0,\, \eta^2)$. Then

$$S_{jk} \sim \mathrm{N}(0,\, \sigma_{kh}^2 + \pi^2 j^2 h^{-1} \eta^2) \qquad j \geq 1$$

and S_{jk} are independent. We find optimal weights w_{jk} for the integrated volatility estimator

$$\widehat{IV}_n = \sum_{k=0}^{h^{-1}-1} \sum_{j=1}^{\infty} w_{jk} (S_{jk}^2 - \pi^2 j^2 h^{-2} \hat{\eta}^2) h :$$

$w_{jk} = I_k^{-1} I_{jk}$ with $I_k = \sum_{j=1}^{\infty} I_{jk}$, $I_{jk} = \frac{1}{2}(\sigma_{kh}^2 + \pi^2 j^2 h^{-2} \eta^2)^{-1}$.
Problem: σ_{kh} are unknown. The solution is to use two-stage methods ($\rightsquigarrow$ estimate weights first). The final result is

$$n^{1/4}(\widehat{IV}_n - \int_0^1 \sigma_s^2 \mathrm{d}s) \xrightarrow{\text{st.}} \mathrm{N}(0,\, 8 \int_0^1 \sigma_s^3 \eta \mathrm{d}s).$$

References

[1] T. Björk. The Pedestrian's Guide to Local Time. Lecture notes, 2013.

[2] Richard Durrett. Stochastic Calculus - A Practical Introduction. CRC Press, Boca Raton, Fla, 0002. aufl. edition, 1996.

[3] M. Fischer and D. Nappo. On the Moments of the Modulus of Continuity of Itô Processes. Stochastic Analysis and Applications, 2010.

[4] Danielle Florens-Zmirou. On Estimating the Diffusion Coefficient from Discrete Observations. J. Appl. Prob., 1993.

[5] E.L. Lehmann and George Casella. Theory of Point Estimation. Springer Science & Business Media, Berlin Heidelberg, 2nd ed. 1998. corr. 4th printing 2003 edition, 1998.

[6] Robert Liptser, Albert N. Shiryaev, and B. Aries. Statistics of Random Processes - I. General Theory. Springer Science & Business Media, Berlin Heidelberg, 2nd rev. and exp. ed. 2001 edition, 2001.

[7] Daniel Revuz and Marc Yor. Continuous Martingales and Brownian Motion. Springer Science & Business Media, Berlin Heidelberg, 3. aufl. edition, 1999.

[8] A. W. van der Vaart. Asymptotic Statistics. Cambridge University Press, Cambridge, 2000.