An Introduction to Inertial Sensors Stochastic Calibration

Stéphane Guerrier, Roberto Molinari, Yuming Zhang, Haotian Xu, Gaetan Bakalli, Ahmed Radi and Mucyo Karemera

2018-06-05

Contents

Chapter 1

Introduction

TO DO (Introduction to the text)

Chapter 2

Introduction to Time Series Analysis

In this chapter, we will provide an introduction to time series analysis. This chapter is organized with the following outline:

- Definition and descriptive analysis of time series;
- Latent time series processes and composite stochastic process;
- Dependence within time series;
- Concept of stationarity;
- Concept of linear processes;
- Fundamental representations of time series.

2.1 Time Series

Definition 2.1 (Time Series)**.** A time series is a stochastic process (i.e. a sequence of random variables (r.v.)), defined on a common probability space denoted as $(X_t)_{t=1,\dots,T}$ (i.e. $\{X_1,\ X_2,\ ....,\ X_T\}$). Note that the time t is not continuous and belongs to discrete index sets. Therefore, we implicitly assume that

- t is not random, i.e. the time at which each observation is measured is known, and
- the time between two consecutive observations is constant.

When recording values of a time series over an extended period of time, it is usually difficult to discern any trend or pattern of the time series by simply looking at the values. However, when these data points are displayed on a plot with time on x-axis and X_t on y-axis, some features of the time series jump out. So it is often useful to understand a time series process by performing a **descriptive analysis**, especially when we have data of small or moderate size.

When we perform descriptive analysis, we usually check the following in the time series data/graph:

- Trends
  - Seasonal (e.g. business cycles)
  - Non-seasonal (e.g. impact of economic indicators on stock returns)
  - Local fluctuation (e.g. vibrations observed before, during and after an earthquake)
- Changes in the statistical properties
  - Mean (e.g. economic crisis)
  - Variance (e.g. earnings)
  - States (e.g. bear/bull in finance)
- Model deviations (e.g. outliers)

Figure 2.1: Johnson and Johnson Quarterly Earnings

Figure 2.2: Monthly Precipitation Data

Example 2.1 (Johnson and Johnson Quarterly Earnings)**.** A traditional example of a time series is the quarterly earnings of the company Johnson and Johnson. In the graph below, we present these earnings between 1960 and 1980.

```
# Load simts package
library(simts)
```

```
##
## Attaching package: 'simts'
## The following object is masked _by_ '.GlobalEnv':
##
##     hydro
```

```
# Load data
data(jj, package = "astsa")
# Construct gts object
jj = gts(jj, start = 1960, freq = 4)
# Plot time series
plot(jj, main = "Johnson and Johnson Quarterly Earnings",
     ylab = "Quarterly Earnings per Share ($)")
```

As we can see from the graph, the data contains a non-linear increasing trend as well as a yearly seasonal component. In addition, we can notice that the variability of the data seems to increase with time. By plotting the time series graph for this data, we can discern some important features of the data, which is helpful for us to conduct further analysis such as selecting suitable models for this data.

Example 2.2 (Monthly Precipitation Data)**.** Now we consider another data set coming from the domain of hydrology. The data records monthly precipitation (in mm) over a certain period of time (1907 to 1972) and is interesting for scientists in order to study water cycles. The data are presented in the graph below:

```
# Load data
data(hydro, package = "simts")
# Construct gts object
hydro = gts(hydro, start = 1907, freq = 12)
# Plot time series
plot(hydro, main = "Monthly Precipitation Data",
     ylab = "Mean Monthly Precipitation (mm)")
```

From the above time series graph, we can observe some extreme observations (i.e. outliers) in this data more easily than if we simply look at all the values of this data.

Example 2.3 (Inertial Sensor Data)**.** Now we consider the data coming from the calibration procedure of an Inertial Measurement Unit (IMU). The signals coming from an IMU are usually measured at high frequencies over a long time and are often characterized by linear trends and numerous underlying stochastic processes. We present the time series graph of some data from an IMU below.

```
# Load data
data(imu6, package = "imudata")
# Construct gts object
imu = gts(imu6[,1], freq = 100*60*60)
```

Figure 2.3: Inertial Sensor Data

```
# Plot time series
plot(imu, main = "Inertial Sensor Data",
     ylab = expression(paste("Error ", (rad/s^2))))
```

As we can see from the graph, although a linear trend and other processes are present in this data, it is practically impossible to discern any feature of the data based on the time series graph. In general, the descriptive analysis in classical time series analysis is not appropriate for the analysis of inertial sensors as these data are usually very large in order to perform parameter estimation.

2.2 Latent Time Series Processes and Composite Stochastic Process

We first introduce some latent time series processes that are commonly used, especially in the calibration procedure of inertial sensors.

Definition 2.2 (Gaussian White Noise)**.** The Gaussian White Noise (WN) process with parameter $\sigma^2 \in \mathbb{R}^+$ is defined as

$$X_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

where "iid" stands for "independent and identically distributed".

Definition 2.3 (Quantization Noise)**.** The Quantization Noise (QN) process with parameter $Q^2 \in \mathbb{R}^+$ is a process with Power Spectral Density (PSD) of the form

$$S_X(f) = 4Q^2 \sin^2\left(\frac{\pi f}{\Delta t}\right)\Delta t, \quad f < \frac{\Delta t}{2}.$$

Definition 2.4 (Drift)**.** The Drift (DR) process with parameter $\omega \in \Omega$, where Ω is either $\mathbb{R}^+$ or $\mathbb{R}^-$, is defined as

$$X_t = \omega t.$$

Definition 2.5 (Random Walk)**.** The Random Walk (RW) process with parameter $\gamma^2 \in \mathbb{R}^+$ is defined as

$$X_t = X_{t-1} + \epsilon_t \;\text{ where }\; \epsilon_t \overset{iid}{\sim} \mathcal{N}\left(0, \gamma^2\right) \;\text{ and }\; X_0 = 0.$$

Definition 2.6 (Auto-Regressive)**.** The Auto-Regressive process of Order 1 (AR1) with parameter $\phi \in (-1, +1)$ and $\upsilon^2 \in \mathbb{R}^+$ is defined as

$$X_t = \phi X_{t-1} + Z_t, \quad Z_t \overset{iid}{\sim} \mathcal{N}(0, \upsilon^2).$$

Definition 2.7 (Gauss Markov)**.** The Gauss Markov process of Order 1 (GM) with parameter $\beta \in \mathbb{R}$ and $\sigma_G^2 \in \mathbb{R}^+$ is defined as

$$X_t = \exp(-\beta\Delta_t)X_{t-1} + Z_t, \quad Z_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_G^2(1 - \exp(-2\beta\Delta t)))$$

where Δt denotes the time between X_t and X_{t-1}.

Remark 2.1 (GM and AR1)**.** A GM process is a one-to-one reparametrization of an AR1 process. In the following, we will only discuss AR1 processes but all results remain valid for GM processes.

With the above defined latent time series processes, we introduce the composite stochastic process, which is widely used in the estimation procedure of inertial sensor stochastic calibration.

Definition 2.8 (Composite Stochastic Process)**.** A composite stochastic process is a sum of latent processes. We implicitly assume that these latent processes are independent.

Example 2.4 (2*AR1 + WN)**.** The composite stochastic process of "2*AR1 + WN" is given as

$$Y_t = \phi_1 Y_{t-1} + Z_t, \quad Z_t \overset{iid}{\sim} \mathcal{N}(0, \upsilon_1^2), \tag{2.1}$$

$$W_t = \phi_2 W_{t-1} + U_t, \quad U_t \overset{iid}{\sim} \mathcal{N}(0, \upsilon_2^2), \tag{2.2}$$

$$Q_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \tag{2.3}$$

$$X_t = Y_t + W_t + Q_t, \tag{2.4}$$

where Y_t, W_t and Q_t are independent and only X_t is observed.

2.3 Dependence within Time Series

One of the main purpose of time series analysis is to make predictions. That is, if $(X_t)_{t=1,\dots,T}$ is an identically distributed but not independent sequence, what is the best predictor for X_{T+h} for $h > 0$ (i.e. an estimator of $\mathbb{E}[X_{T+h}|X_T, ...]$)? In order to answer this question, we need to understand the dependence between $X_1, \dots, X_T$. Before we start to consider the dependence within time series, let us first take a review on independence.

Definition 2.9 (Independence of Events)**.** Two events A and B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

In general, $A_1, \dots, A_n$ are independent if

$$\mathbb{P}(B_1 \dots B_n) = \mathbb{P}(B_1) \dots \mathbb{P}(B_n) \text{ for all } B_i = A_i \text{ or } S, \quad i = 1, \dots, n$$

where S is the sample space.

Definition 2.10 (Independence of Random Variables)**.** Two random variables X and Y with Cumulative Distribution Functions (CDF) $F_X(x)$ and $F_Y(y)$ respectively are independent if and only if their joint CDF $F_{X,Y}(x, y)$ is such that

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$

In general, random variables $X_1, \dots, X_n$ with CDF $F_{X_1}(x_1), \dots, F_{X_n}(x_n)$ respectively are independent if and only if their joint CDF $F_{X_1,\dots,X_n}(x_1, \dots, x_n)$ is such that

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n).$$

Definition 2.11 (iid sequence)**.** The sequence $X_1, X_2, \dots, X_T$ is said to be independent and identically distributed (i.e. iid) if and only if

$$\mathbb{P}(X_i < x) = \mathbb{P}(X_j < x) \ \ \forall x \in \mathbb{R}, \forall i, j \in \{1, \dots, T\}$$

and

$$\mathbb{P}(X_1 < x_1, X_2 < x_2, \dots, X_T < x_T) = \mathbb{P}(X_1 < x_1) \dots \mathbb{P}(X_T < x_T) \ \ \forall T \geq 2, x_1, \dots, x_T \in \mathbb{R}.$$

Now we start to consider the dependence, specifically the **linear dependence**, within time series. Notice that it is difficult to consider the dependence between T random variables at a time. So we need to consider only two random variables at a time.

Definition 2.12 (AutoCovariance)**.** AutoCovariance (ACV) denoted as $\gamma_X(t, t+h)$ is defined as

$$\gamma_X(t, t+h) = \mathrm{Cov}(X_t, X_{t+h}) = \mathbb{E}(X_t X_{t+h}) - \mathbb{E}(X_t)\mathbb{E}(X_{t+h})$$

where

$$\mathbb{E}(X_t) = \int_{-\infty}^{\infty} x f(x) dx \text{ and } \mathbb{E}(X_t, X_{t+h}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2) dx_1 dx_2$$

where $f(x)$ denotes the density of X_t and $f(x_1, x_2)$ denotes the joint density of X_t and X_{t+h}.

Remark 2.2 (Properties of ACV)**.**

1. ACV is symmetric, i.e. $\gamma_X(t, t+h) = \gamma_X(t+h, t)$ as $\mathrm{Cov}(X_t, X_{t+h}) = \mathrm{Cov}(X_{t+h}, X_t)$. Under stationarity (will be discussed very soon), $\gamma_X(h) = \gamma_X(-h)$, i.e. ACV is an even function.

2. Variance of the process $\mathrm{Var}(X_t) = \gamma_X(t, t) \geq 0$. Under stationarity, $\mathrm{Var}(X_t) = \gamma_X(0)$ and $\mid \gamma_X(h) \mid \leq \gamma_X(0)$ by Cauchy-Schwarz inequality.

3. Scale dependent: ACV $\gamma_X(t, t+h)$ is scale dependent like any covariance. So $\gamma_X(t, t+h) \in \mathbb{R}$.
   - If $\mid \gamma_X(t, t+h) \mid$ is "close" to 0, then X_t and X_{t+h} are "less (linearly) dependent".
   - If $\mid \gamma_X(t, t+h) \mid$ is "far" from 0, then X_t and X_{t+h} are "more (linearly) dependent".

However in general, it is difficult to assess what "close" and "far" from zero mean.

4. In general, $\gamma_X(t, t+h) = 0$ does not imply X_t and X_{t+h} are independent. However, if X_t and X_{t+h} are joint normally distributed, then $\gamma_X(t, t+h) = 0$ implies that X_t and X_{t+h} are independent.

Another measure of linear dependence which is related to the ACV is the AutoCorrelation. This is arguably the most commonly used metric in time series analysis.

Definition 2.13 (AutoCorrelation)**.** AutoCorrelation (ACF) denoted as $\rho_X(t, t+h)$ is defined as

$$\rho_X(t, t+h) = \mathrm{Corr}(X_t, X_{t+h}) = \frac{\mathrm{Cov}(X_t, X_{t+h})}{\sqrt{\mathrm{Var}(X_t)}\sqrt{\mathrm{Var}(X_{t+h})}}.$$

Remark 2.3 (Properties of ACF)**.**

1. $\mid \rho_X(t, t+h) \mid \leq 1$ and $\mid \rho_X(t, t) \mid = 1$.

2. ACF is symmetric, i.e. $\rho_X(t, t+h) = \rho_X(t+h, t)$ as $\mathrm{Corr}(X_t, X_{t+h}) = \mathrm{Corr}(X_{t+h}, X_t)$. Under stationarity, $\gamma_X(h) = \gamma_X(-h)$, i.e. ACF is an even function.

3. Scale invariant: ACF $\rho_X(t, t+h)$ is scale free like any correlation. Moreover, if $\rho_X(t, t+h)$ is "close" to ± 1, then this implies that there is "strong" (linear) dependence between X_t and X_{t+h}.

We can simplify the notations $\gamma_X(t, t+h)$ and $\rho_X(t, t+h)$ to be $\gamma(t, t+h)$ and $\rho(t, t+h)$ when there is no ambiguity (i.e. only one time series is considered).

Notice that both ACV and ACF are appropriate to measure linear dependence only. Besides linear dependence, other forms of dependence such as monotonic or nonlinear dependence also exist. However, both ACV and ACF are less helpful to measure these dependence as they might have ACV and ACF to be zero. Here is an example:

It is worth noting that correlation does NOT imply causation. For example, if $\rho(t, t+h) \neq 0$, it does not imply that $X_t \to X_{t+h}$ is causal. Actually, real causation doesn't exist in statistics but there exists approximated metric to measure this concept such as Granger causality (see **?**). This idea is clearly illustrated in the image below:

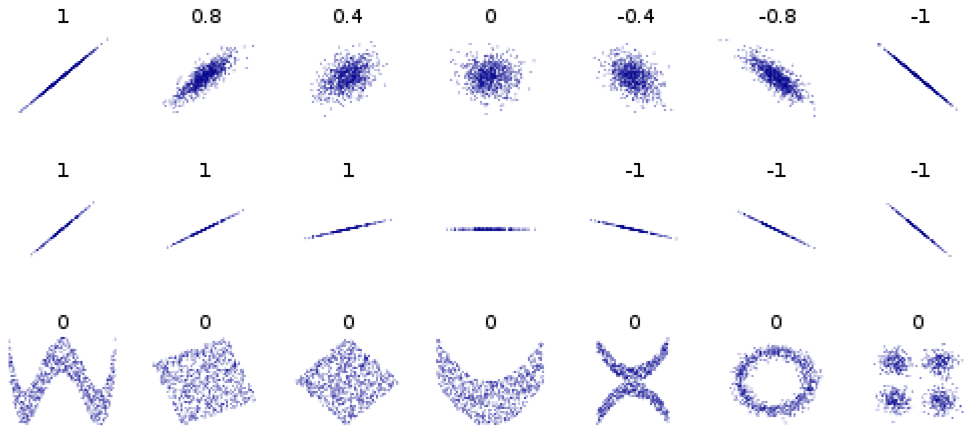

Figure 2.4: Different forms of dependence and their ACF values

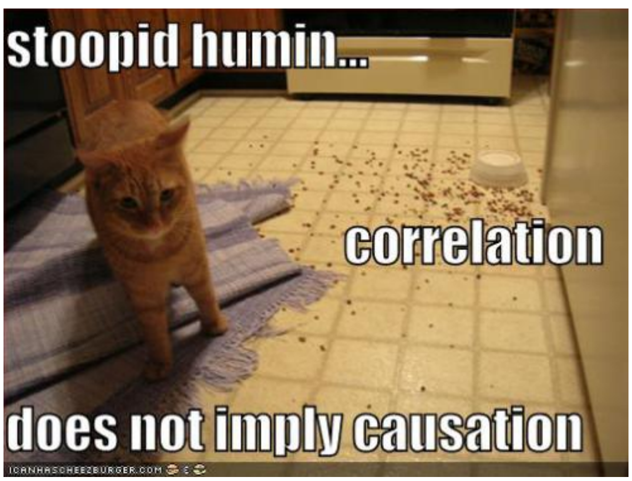

Figure 2.5: Correlation does NOT imply causation.

2.4 Stationarity

In this section we are going to introduce the concept of stationarity, one of the most important characteristics of time series data. First let us consider an example of **non-stationary processes**.

Example 2.5 (Non-Stationary Process).

$$X_t \sim \mathcal{N}\left(0, Y_t^2\right) \quad \text{where } Y_t \text{ is unobserved and such that } \; Y_t \overset{iid}{\sim} \mathcal{N}\left(0, 1\right).$$

In this case, it is clear that the estimation of $\text{Var}(X_t)$ is difficult since only X_t is useful for the estimation. So in fact, X_t^2 is our best guess for $\text{Var}(X_t)$.

On the other hand, let us consider an example of **stationary processes** where averaging becomes meaningful for such process.

Example 2.6 (Stationary Process).

$$X_t = \theta W_{t-1} + W_t \quad \text{where} \quad W_t \overset{iid}{\sim} \mathcal{N}\left(0, 1\right).$$

In this case, we can guess that a natural estimator of $\text{Var}(X_t)$ can be $\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{T} X_i^2$. That is, now averages are meaningful for such process.

We formalize the above idea by introducing the concept of stationarity. There exist two forms of stationarity, which are defined below:

Definition 2.14 (Strong Stationarity)**.** The time series X_t is strongly stationary if the joint probability distribution is invariant under a shift in time, i.e.

$$\mathbb{P}(X_t \leq x_0, \ldots, X_{t+k} \leq x_k) = \mathbb{P}(X_{t+h} \leq x_0, \ldots, X_{t+h+k} \leq x_k)$$

for any time shift h and any $x_0, x_1, x_2, \cdots, x_k$ belong to the domain of $X_t, \cdots, X_{t+k}$ and $X_{t+h}, \cdots, X_{t+h+k}$.

Definition 2.15 (Weak Stationarity)**.** The time series $(X_t)_{t\in\mathbb{N}}$ is weakly stationary if the mean and auto-covariance are finite and invariant under a shift in time, i.e.

$$\begin{aligned} \mathbb{E}\left[X_t\right] &= \mu < \infty, \\ \mathbb{E}\left[X_t^2\right] &= \mu_2 < \infty, \\ \text{Cov}(X_t, X_{t+h}) &= \text{Cov}(X_{t+k}, X_{t+h+k}) = \gamma(h). \end{aligned}$$

for any time shift h. For convenience, we use the abbreviation "stationary" to indicate "weakly stationary" by default.

The stationarity of X_t is important because it provides a framework in which averaging makes sense. The concept of averaging is essentially meaningless unless properties like mean and covariance are either fixed or evolve in a known manner.

Remark 2.4 (Implication on the ACV and ACF)**.** If a process is weakly stationary or strongly stationary and $\text{Cov}(X_t, X_{t+h})$ exists for all $h \in \mathbb{Z}$, then we have both ACV and ACF only depend on the lag between observations, i.e.

$$\begin{aligned} \gamma(t, t+h) &= \text{Cov}(X_t, X_{t+h}) = \text{Cov}(X_{t+k}, X_{t+h+k}) = \gamma(t+k, t+h+k) = \gamma(h), \\ \rho(t, t+h) &= \text{Corr}(X_t, X_{t+h}) = \text{Corr}(X_{t+k}, X_{t+h+k}) = \rho(t+k, t+h+k) = \rho(h). \end{aligned}$$

Remark 2.5 (Relation between Strong and Weak Stationarity)**.** In general, neither type of stationarity implies the other one. However,

- If X_t is Normal (Gaussian) with $\sigma^2 = \text{Var}(X_t) < \infty$, then weak stationarity implies strong stationarity.
- If X_t is strongly stationary, $\mathbb{E}(X_t) < \infty$ and $\mathbb{E}(X_t^2) < \infty$, then X_t is weakly stationary.

Example 2.7 (Strong Stationarity does NOT imply Weak Stationarity)**.** An iid Cauchy process is strongly but not weakly stationary as the mean of the process does not exist.

Example 2.8 (Weak Stationarity does NOT imply Strong Stationarity)**.** Let $X_t \overset{iid}{\sim} \exp(1)$ (i.e. exponential distribution with $\lambda = 1$) and $Y_t \overset{iid}{\sim} \mathcal{N}(1, 1)$. Then, let

$$Z_t = \begin{cases} X_t & \text{if } t \in \{2k \vert k \in \mathbb{N}\} \\ Y_t & \text{if } t \in \{2k+1 \vert k \in \mathbb{N}\} \end{cases},$$

we have Z_t is weakly stationary but not strongly stationary.

Remark 2.6 (Stationarity of Latent Time Series Processes)**.**

- (Weakly) Stationary: WN, QN, AR1
- (Weakly) Non-Stationary: DR, RW

AR1 is weakly stationary. Consider an AR1 process defined as:

$$X_t = \phi X_{t-1} + Z_t, \quad Z_t \overset{iid}{\sim} \mathcal{N}(0, \nu^2),$$

where $\mid \phi \mid < 1$ and $\nu^2 < \infty$. Then we have

$$\begin{aligned} X_t &= \phi X_{t-1} + Z_t = \phi \left[\phi X_{t-2} + Z_{t-1}\right] + Z_t = \phi^2 X_{t-2} + \phi Z_{t-1} + Z_t \\ &\vdots \\ &= \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j Z_{t-j}. \end{aligned}$$

By taking the limit in k (which is perfectly valid as we assume $t \in \mathbb{Z}$), we obtain

$$X_t = \lim_{k \to \infty} X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}.$$

So we have

$$\begin{aligned} \mathbb{E}\left[X_t\right] &= \sum_{j=0}^{\infty} \phi^j \mathbb{E}[Z_{t-j}] = 0, \\ \text{Var}\left(X_t\right) &= \text{Var}\left(\sum_{j=0}^{\infty} \phi^j Z_{t-j}\right) = \sum_{j=0}^{\infty} \phi^{2j} \text{Var}\left(Z_{t-j}\right) = \nu^2 \sum_{j=0}^{\infty} \phi^{2j} = \frac{\nu^2}{1-\phi^2} < \infty. \end{aligned}$$

Moreover, assuming for notational simplicity that $h > 1$, we obtain

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

In general, when $h \in \mathbb{Z}$ we obtain

$$\text{Cov}\left(X_t, X_{t+h}\right) = \phi^{|h|} \text{Cov}(X_t, X_t) = \phi^{|h|} \frac{\nu^2}{1-\phi^2},$$

which is a function of the lag h only. Therefore, this AR1 process is weakly stationary. □

2.5 Linear Processes

In this section we introduce the concept of linear processes. As a matter of fact, the stationary models considered so far can all be represented as linear processes.

Definition 2.16 (Linear Processes)**.** A stochastic process (X_t) is said to be a linear process if it can be expressed as a linear combination of an iid Gaussian sequence (i.e. white noise process), i.e.:

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$$

where $W_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ and $\sum\limits_{j=-\infty}^{\infty} |\psi_j| < \infty$.

Notice that the condition $\sum\limits_{j=-\infty}^{\infty} |\psi_j| < \infty$ is required in the definition of linear processes in order to ensure that the series has a limit and is related to the absolutely summable covariance structure, which is defined below.

Definition 2.17 (Absolutely Summable Covariance Structure)**.** A process (X_t) is said to have an absolutely summable covariance structure if $\sum\limits_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$.

Remark 2.7 (Properties of Linear Processes)**.** 1. All linear processes are stationary since

$$\begin{aligned}\mathbb{E}[X_t] &= \mu,\\ \gamma(h) &= \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}.\end{aligned}$$

2. All linear processes have absolutely summable covariance structures.

ACV of Linear Processes.

$$\begin{aligned}\gamma(h) = \text{Cov}(X_t, X_{t+h}) &= \text{Cov}(\mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}, \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t+h-j})\\ &= \text{Cov}(\sum_{j=-\infty}^{\infty} \psi_j W_{t-j}, \sum_{j=-\infty}^{\infty} \psi_j W_{t-(j-h)})\\ &= \text{Cov}(\sum_{j=-\infty}^{\infty} \psi_j W_{t-j}, \sum_{j=-\infty}^{\infty} \psi_{j+h} W_{t-j})\\ &= \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \text{Cov}(W_{t-j}, W_{t-j})\\ &= \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}.\end{aligned}$$

□

All linear processes have absolutely summable covariance structures.

$$\begin{aligned}
\sum_{h=-\infty}^{\infty} \mid \gamma(h) \mid &= \sum_{h=-\infty}^{\infty} \sigma^2 \mid \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \mid \\
&\leq \sigma^2 \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \mid \psi_j \psi_{j+h} \mid \\
&= \sigma^2 \sum_{j=-\infty}^{\infty} \sum_{h=-\infty}^{\infty} \mid \psi_j \mid \cdot \mid \psi_{j+h} \mid \\
&= \sigma^2 \sum_{j=-\infty}^{\infty} \mid \psi_j \mid \sum_{h=-\infty}^{\infty} \mid \psi_{j+h} \mid \\
&= \sigma^2 \Big(\sum_{j=-\infty}^{\infty} \mid \psi_j \mid \Big)^2 < \infty
\end{aligned}$$

So with the assumption that $\sum_{j=-\infty}^{\infty} \mid \psi_j \mid < \infty$, we obtain that all linear processes have absolutely summable covariance structures. Notice that here we have shown that the condition $\sum\limits_{j=-\infty}^{\infty} |\psi_j| < \infty$ is actually stronger than $\sum\limits_{h=-\infty}^{\infty} |\gamma(h)| < \infty$. □

Example 2.9 (AR1 is a linear process)**.** When we prove above that AR1 is weakly stationary, we have shown that for an AR1 process $X_t = \phi X_{t-1} + Z_t, \quad Z_t \overset{iid}{\sim} \mathcal{N}(0, \nu^2)$, it can be represented as

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}.$$

Therefore, AR1 is a linear process.

2.6 Fundamental Representations of Time Series

We conclude this chapter by summarizing the fundamental representations of time series. If two processes have the same fundamental representations, then these two processes are the same. There are two most commonly used fundamental representations of time series, i.e.

- ACV and ACF;
- the Power Spectral Density (PSD).

Remark 2.8 (ACV and ACF as fundamental representation)**.** If we consider a zero mean normally distributed process, it is clear that its joint distribution is fully characterized by the autocovariances $\mathbb{E}[X_t X_{t+h}]$ since the joint probability density only depends on these covariances. Once we know the autocovariances we know everything there is to know about the process and therefore: if two processes have the same autocovariance function, then they are the same process.

Remark 2.9 (PSD as fundamental representation)**.** The Power Spectral Density (PSD) is defined as

$$S_X(f) = \int_{-\infty}^{\infty} \gamma_X(h)\, e^{-ifh} dh,$$

where f is a frequency. Hence, the PSD is a Fourier transform of the autocovariance function which describes the variance of a time series over frequencies (with respect to lags h).

Given that the definition of the PSD, as for the autocovariance function, once we know the PSD we know everything there is to know about the process and therefore: if two processes have the same PSD, then they are the same process.

Chapter 3

A Review of the Properties of Statistical Estimators

TO DO

Chapter 4

Allan Variance Calibration Techniques

- The Allan Variance (AV) is a statistical technique originally developed in the mid-1960s to study the stability of precision oscillators (see e.g. **?**).
- It can provide information on the types and magnitude of various superimposed noise terms (i.e. composite stochastic processes).
- This method has been adapted to characterize the properties of a variety of devices including inertial sensors (see **?**).
- The AV is a measure of variability developed for long term memory processes and can in fact be interpreted as a Haar wavelet coefficient variance (see **?**). We will discuss this connection further on.

Definition 4.1 (Allan Variance)**.** We consider the AV at dyadic scales (τ_j) starting from local averages of the process which can be denoted as

$$\bar{X}_t^{(j)} \equiv \frac{1}{\tau_j} \sum_{i=1}^{\tau_j} X_{t-\tau_j+i} \, ,$$

where $\tau_j \equiv 2^j, \; j \in \{x \in \mathbb{N} \, : \, 1 \leq x < \log_2(T) - 1\}$ therefore determines the number of consecutive observations considered for the average. Then, the AV is defined as

$$\text{AV}_j \left(X_t \right) \equiv \frac{1}{2} \, \mathbb{E} \left[\left(\bar{X}_t^{(j)} - \bar{X}_{t-\tau_j}^{(j)} \right)^2 \right] .$$

Remark 4.1 (Alternative scale definition)**.** The definition of the AV is actually valid for $\tau_j = \lfloor 2^j \rfloor$ with $j \in \{x \in \mathbb{R} \, : \, 1 \leq x < \log_2(T) - 1\}$. In some case, it could use to consider this alternative definition (see e.g. **?**) but we shall restrict ourself here to the case where $j \in \{x \in \mathbb{N} \, : \, 1 \leq x < \log_2(T) - 1\}$.

Remark 4.2 (Notation of the Allan Variance)**.** For notational simplicity, we may sometimes replace $\text{AV}_j \left(X_t \right)$ by simply ϕ_j^2 when the dependence of the AV to the process (X_t) is evident.

As highlighted earlier, the AV is, among others, a widely and commonly used approach in engineering for sensor calibration as it is linked to the properties of the process (X_t) as shown in the following lemma (see e.g. **?**, for a proof).

Lemma 4.1 (AV connection to PSD)**.** *For a stationary process (X_t) with PSD $S_X(f)$ we have*

$$\phi_j^2 \equiv AV_j \left(X_t \right) = 4 \int_0^{\infty} \frac{\sin^4(\pi f \tau_j)}{(\pi f \tau_j)^2} S_X(f) df.$$

Therefore, this result establishes a direct connection between the AV and PSD. A natural question is therefore whether the mapping PSD $\mapsto$ AV is one-to-one. **?** (see Theorem 1) showed that this is actually not the case. This is illustrated in the following section ????? (ADD REF).

4.1 Spectral Ambiguity of the AV

Consider two (independent) stochastic processes (X_t) and (Y_t) with respective PSD $S_X(f)$ and $S_Y(f)$. Suppose that $S_X(f) \neq S_Y(f)$, then the two processes will have the same AV if

$$\Delta \equiv \int_0^\infty \frac{\sin^4(\pi f \tau_j)}{(\pi f \tau_j)^2} \Phi(f) df = 0,$$

where $\Phi(f) \equiv S_X(f) - S_Y(f)$. To show that it is possible that $\Delta = 0$ when $\Phi(f) \neq 0$, we will use the following critical identity:

$$\sin^4(x) = \sin^2(x) - \frac{1}{4}\sin^2(2x). \tag{4.1}$$

First, we note that Δ may be expressed using (**??**) as follows:

$$\begin{aligned}
\Delta &= \int_0^\infty \frac{\sin^4(\tau\pi f)}{(\tau\pi f)^2} \Phi(f) df \\
&= \lim_{n \to -\infty} \int_{2^n}^\infty \frac{\sin^2(\tau\pi f) - \frac{1}{4}\sin^2(2\tau\pi f)}{(\tau\pi f)^2} \Phi(f) df.
\end{aligned}$$

Second, by the change of variable $u = 2f$ in the second term we obtain

$$\Delta = \lim_{n \to -\infty} \left[\int_{2^n}^\infty \frac{\sin^2(\tau\pi f)}{(\tau\pi f)^2} \Phi(f) df - \frac{1}{2} \int_{2^{n+1}}^\infty \frac{\sin^2(\tau\pi u)}{(\tau\pi u)^2} \Phi(f) du \right].$$

Now suppose that $\Phi(f) = 2\Phi(2f)$. In this case, we have $\Phi(f) = 2\Phi(u)$ and therefore we obtain

$$\Delta = \lim_{n \to -\infty} \int_{2^n}^{2^{n+1}} \frac{\sin^2(\tau\pi f)}{(\tau\pi f)^2} \Phi(f) df = 0.$$

Remark 4.3. This result demonstrates that the mapping from PSD to Allan variance is not necessarily one-to-one. **?** showed that in the continuous case (i.e. $\tau_j \in \mathbb{R}$) $\Delta = 0$ if and only if $\Phi(f) = 2\Phi(2f)$. However, the "only if" part of this results (while conjectured) is unknown in the discrete case.

4.2 Properties of the Allan Variance

One reason of explaining the widespread use of the Allan variance for sensor calibration is due to the following additivity property, which is particularly convenient to identify composite stochastic processes (see Definition REF MISSING!).

Corollary 4.1 (Additivity of the AV)**.** *Consider two (independent) stochastic processes* (X_t) *and* (Y_t) *with respective PSD* $S_X(f)$ *and* $S_Y(f)$*. Suppose that we observe the process* $Z_t = X_t + Y_t$*. Then, we have*

$$AV_j(Z_t) = AV_j(X_t) + AV_j(Y_t).$$

Proof. The proof of this result is direct from Lemma REF MISSING HERE. Indeed, since $S_Z(f) = S_X(f) + S_Y(f)$, we have

$$
\begin{aligned}
\mathrm{AV}_j\left(Z_t\right) &= 4\int_0^{\infty} \frac{\sin^4(\pi f \tau_j)}{(\pi f \tau_j)^2} S_Z(f) df \\
&= 4\int_0^{\infty} \frac{\sin^4(\pi f \tau_j)}{(\pi f \tau_j)^2} S_X(f) df + 4\int_0^{\infty} \frac{\sin^4(\pi f \tau_j)}{(\pi f \tau_j)^2} S_Y(f) df \\
&= \mathrm{AV}_j\left(X_t\right) + \mathrm{AV}_j\left(Y_t\right).
\end{aligned}
$$

□

While Lemma **??** is an important results which is very convenient to determine the theoretical AV of a certain stochastic process. However, the applicability of this results is often limited since the integral defined in (REF MISSING HERE) can be intractable. An alternative to Lemma **??** has been proposed by **?** and is far advantageous from a computational standpoint.

Lemma 4.2 (AV connection to ACF)**.** *For a stationary process* (X_t) *with variance* σ_X^2 *and ACF* $\rho(h)$ *we have*

$$
AV_j\left(X_t\right) = \frac{\sigma_X^2}{\tau_j^2}\left(\tau_j\left[1-\rho(\tau_j)\right] + \sum_{i=1}^{\tau_j - 1} i\left[2\rho(\tau_j - i) - \rho(i) - \rho(2\tau_j - i)\right]\right).
$$

The proof of this result is instructive and is presented in **?**.

Remark 4.4. Using Lemma **??**, the exact form of the AV for different stationary processes, such as the general class of ARMA models, can easily be derived. Moreover, **?** provided the theoretical AV for non-stationary processes such as the random walk and ARFIMA models for which the AV, as mentioned earlier, represents a better measure of uncertainty compared to other methods.

Remark 4.5. Lemma **??** was extended to non-stationary processes in **?**.

Example 4.1 (Theoretical AV of an MA(1) process)**.** From the autocovariance we obtain

$$
\rho(h) = \operatorname{corr}\left(X_t, X_{t-h}\right) = \begin{cases} 1 & \text{if } h = 0 \\ \frac{\theta}{1+\theta^2} & \text{if } |h| = 1 \\ 0 & \text{if } |h| > 1. \end{cases}
$$

We can now apply the formula given in Lemma **??**, which leads to

$$
\begin{aligned}
\mathrm{AV}_j\left(X_t\right) &= \frac{\left(1+\theta^2\right)\sigma^2}{\tau_j^2}\left(\tau_j + \sum_{i=1}^{\tau_j-1} i\left[2\rho(\tau_j - i) - \rho(i) - \rho(2\tau_j - i)\right]\right) \\
&= \frac{\left(1+\theta^2\right)\sigma^2}{\tau_j^2}\left(\tau_j + 2\sum_{i=1}^{\tau_j-1} i\rho(\tau_j - i) - \sum_{i=1}^{\tau_j-1} i\rho(i) - \sum_{i=1}^{\tau_j-1} i\rho(2\tau_j - i)\right) \\
&= \frac{\left(1+\theta^2\right)\sigma^2}{\tau_j^2}\left(\tau_j + 2(\tau_j - 1)\rho(1) - \rho(1)\right) \\
&= \frac{\left(1+\theta^2\right)\sigma^2}{\tau_j^2}\left(\tau_j + (2\tau_j - 3)\frac{\theta}{1+\theta^2}\right).
\end{aligned}
$$

•

4.3 Estimation

Several estimators of the AV have been introduced in the literature. The most commonly is (probably) the Maximum-Overlapping AV (MOAV) estimator proposed by **?**, which is defined as follows:

Definition 4.2 (Maximum-Overlapping AV Estimator)**.** The MOAV is defined as:

$$\hat{\phi}_j^2 \equiv \widehat{\mathrm{AV}}_j\left(X_t\right) = \frac{1}{2\left(T - 2\tau_j + 1\right)} \sum_{k=2\tau_j}^{T} \left(\bar{X}_k^{(j)} - \bar{X}_{k-\tau_j}^{(j)}\right)^2. \tag{4.2}$$

We will now study the properties of this estimator through the following lemmas.

4.3.1 Consistency

Lemma 4.3 (Consistency)**.** *Let (X_t) be such that:*

- $(X_t - X_{t-1})$ *is a (strongly) stationary process,*
- $(X_t - X_{t-1})^2$ *has absolutely summable covariance structure,*
- $\mathbb{E}\left[(X_t - X_{t-1})^4\right] < \infty$,

Then, we have

$$\widehat{AV}_j\left(X_t\right) \xrightarrow{\mathcal{P}} AV_j\left(X_t\right).$$

Proof. TO BE ADDED ☐

Remark 4.6 (Connection to Wavelet Variance)**.** This result is closely related by the results of **?** on the wavelet variance. We shall explore the connection between the AV and wavelet variance in the next section.

4.3.2 Asymptotic Normality

Compare to consistency, the asymptotic normality requires stronger conditions given in the following lemma.

Lemma 4.4 (Asymptotic normality)**.** *Let (X_t) be such that:*

- $(X_t - X_{t-1})$ *is a (strongly) stationary process.*
- $(X_t - X_{t-1})$ *is strong mixing process with mixing coefficient $\alpha(n)$ such that $\sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta}{2+\delta}} < \infty$ for some $\delta > 0$.*
- $\mathbb{E}\left[(X_t - X_{t-1})^{4+\delta}\right] < \infty$ *for some $\delta > 0$.*

Then, under these conditions we have that

$$\sqrt{T}\left(\widehat{AV}_j\left(X_t\right) - AV_j\left(X_t\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_T^2/T),$$

where $\sigma_T^2 \equiv \sum_{h=-\infty}^{\infty} cov\left(\left(\bar{X}_t^{(j)} - \bar{X}_{t-\tau_j}^{(j)}\right)^2, \left(\bar{X}_{t+h}^{(j)} - \bar{X}_{t+h-\tau_j}^{(j)}\right)^2\right)$.

Proof. TO BE ADDED ☐

4.3.3 Confidence Interval of the MOAV Estimator

Based on the asymptotic normality results (Lemma **??**), we can construct the $1-\alpha$ confidence intervals for $\widehat{\text{AV}}_j\left(X_t\right)$ as %

$$\text{CI}\left(\text{AV}_j\left(X_t\right)\right) = \left[\widehat{\text{AV}}_j\left(X_t\right) \pm z_{1-\frac{\alpha}{2}} \frac{\sigma_T}{T}\right],$$

% where $z_{1-\frac{\alpha}{2}} \equiv \mathbf{\Phi}^{-1}\left(1-\frac{\alpha}{2}\right)$ is the $\left(1-\frac{\alpha}{2}\right)$ quantile of a standard normal distribution.\ However, the so called "Long-Run Variance" σ_T^2 is usually unknown. Many methods have been proposed to consistently estimate it under mild conditions (see e.g. **?**).

Remark 4.7. Gaussian-based confidence intervals are often problematic with the AV as the lower limit of CI can very well be negative. Next, we will discuss an alternative method to construct the CI for such statistic.

4.4 Allan variance based estimation

4.4.1 Allan Variance log-log Representation

As illustrated in Lemmas **??** and **??** the AV depends on the properties of the stochastic process (X_t). We will see that "log-log" representation of the AV is often useful for the identify various processes that may compose (X_t).

For example, let's suppose that X_t is a white noise process. We showed in REF MISSING HERE that the theoretical AV of such process is given

$$\phi_j^2 \equiv \text{AV}_j(X_t) = \frac{\sigma^2}{\tau_j}.$$

Therefore, we have that the Allan Deviation or AD (i.e. $\sqrt{\text{AV}_j(X_t)}$ or ϕ_j) is such that

$$\log\left(\phi_j\right) = \log\left(\sqrt{\frac{\sigma^2}{\tau_j}}\right) = \log\left(\sigma\right) - \frac{1}{2}\log(\tau_j). \tag{4.3}$$

Thus, the log of the AD is linear in τ_j with a slope of $-1/2$ and with intercept $\log(\sigma)$. Let us start by considering a simple simulated example.

4.4.2 Allan Deviation of a WN process

Simulation based on a white noise process with $\sigma^2 = 10^2$ and $T = 10^5$.

```r
# Load packages
library(av)       # Package for Allan Variance functions
library(simts)    # Package for time series simulations

# Simulate white noise
Xt = gen_gts(WN(sigma2 = 1), n = 10^5)

# Compute allan variance
av = avar(Xt)

# Allan Variance log-log Representation
plot(av)
```

Figure 4.1: ADD A NICE CAPTION

Chapter 5

The Generalized Method of Wavelet Moments

Chapter 6

Extensions