Temporal Time Series Analysis

Joaquín Rapela

Gatsby Computational Neuroscience Unit
University College London

January 13, 2025

Contents

Contents

Background

- On Spring 2023 I helped in the discussion sessions of the Neuroinformatics course by Prof. Ken Harris, for UCL undergraduate students in Neuroscience.
- Suggested to Klara Olofsdotter (SWC PhD program coordinator) and Sonja Hofer (SWC PhD program faculty coordinator) to ask SWC PhD students to take this course. They liked the idea.
- With Gatsby Unit PhD students and postdoctoral scholars, as well as researchers from elsewhere, we offered Neuroinformatics 2024, Ken taught the first five lectures and we taught the remaining ones.
- On 2025 we renamed the course *Statistical Neuroscience* and we invited lecturers and students from the Francis Crick Institute.

A few of our motivations to run this course

1. Learn by teaching.
2. Gain more teaching experience.
3. Provide SWC PhD students with relevant neural data-analysis tools.
4. Contribute to better interactions between the SWC and the Gatsby Unit. Build a common language.

Course structure

Refer to the course repo.

Course logistics

Lectures : Monday 1-3pm, SWC lecture theatre.

Practicals : Friday 2-3:30pm, SWC lecture theatre.

Office Hours : Joaquin, Wednesday 4-5pm, or arranged by appointment.

Worksheets : assigned on Mondays, due on the following Monday before 1pm. Worksheets by SWC PhD students will be graded. Solutions to all worksheets will be provided.

Participation : in-class participation, and off-class participation (e.g., by email), is greatly encouraged.

Contents

Contents

What is a time series?

Definition 1

A **time series** is a set of observations $\{x_t\}$, each observation recorded at a specific time t.

Key concepts in probability theory

sample space The set, $\mathcal{S}$, of all possible outcomes of a particula experiment is called the **sample space** for this experiment.

event An **event** is any collection of possible outcomes of an experiment, that is, any subset of $\mathcal{S}$ (including $\mathcal{S}$ itself).

probability function A **probability function** is a function that assigns a real number (i.e., the probability) to events.

random variable A **random variable** is a function from the sample space $\mathcal{S}$ into the real numbers. For example in an experiment tossing two dice a random variable X could be the sum of the obtained numbers, or in an experiment tossing a coin 25 times a random variable Y could be the number of heads in the 25 tosses.

The probability that a random variable takes a given value is the probability of the event associated with that value.

$$P_X(X = x_i) = P(\{s_j \in \mathcal{S} : X(s_j) = x_i\})$$

Time series model

The purpose of this course is to study techniques to draw inferences from neural time series. Before doing so, it is necessary to set up a hypothetical proability model to represent the time series. Then it is possible to estimate model parameters, check for goodness of fit to the data, and possibly use the estimated model to enhance our understanding of the mechanisms generating the time series.

Definition 2 (time series model)

A **time series model** for the observed data $\{x_t\}$ is a specification of the joint distribution (or possibly the means and covariances) of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization.

What is time series analysis?

- Time series analysis characterises **data that is correlated in time**.
- These correlations severely **restrict the applicability of conventional techniques** assuming data samples that are independent and identically distributed.
- These correlations allow to **forecast** future values of a time series based on present and past values.

Relevance of time series analysis

economics
: daily stock market quotations, monthly unemployment figures.

social scientists
: birthrates, school enrolment.

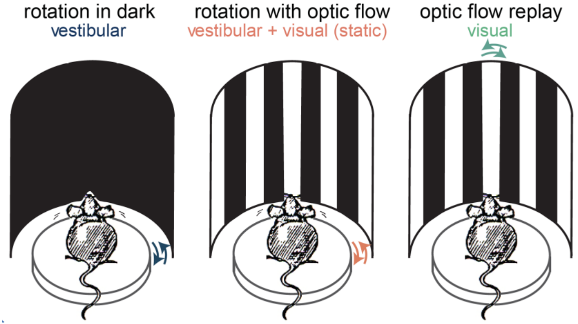
epidemiology
: number of influenza cases observed over some time period.

medicine
: blood pressure measurements traced over time.

Examples of SWC time series analysis

1. aeon project: kinematic inference.

2. integration of visual/vestibular information, with Prof. Sepi Keshavarsi.



Temporal vs spectral time series analysis

temporal time series analysis focuses on the analysis of lagged relationship (e.g., how does what happened today affect what will happen tomorrow?).

spectral time series analysis centres on the analysis of rhythms (e.g., can we observe rhythmic activity in local field potentials recorded from human brains?)

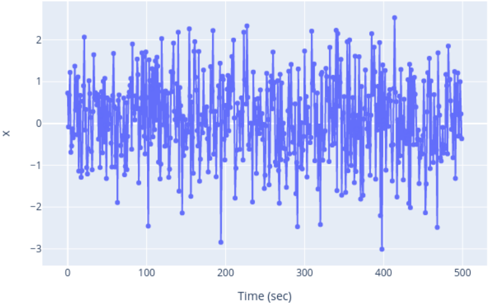
Contents

Generation of time series: white noise process

The first step to generate time series is to generate **white noise process**, $\{w_t\}$ (i.e., independent Gaussian random variables with zero mean and fixed variance, example). For $t = 0, \pm 1, \pm 2, \ldots$

$$E\{w_t\} = 0$$

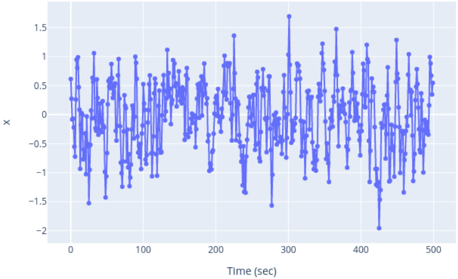
$$\gamma(t,s) = \mathrm{Cov}(w_t, w_s) = \begin{cases} \sigma_w^2 & s = t \\ 0 & s \neq t \end{cases}$$



Generation of time series: moving average model

In a white noise process w_t, for any pairs of time points, t_1 and t_2, the random variables w_{t_1} and w_{t_2} are uncorrelated. The **moving average model** adds serial correlation to white noise (example).

$$\nu_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1}) \tag{1}$$



Moving average model of order q

Definition 3

An **moving average model** of order q, abbreviated as **MA**(q), is of the form

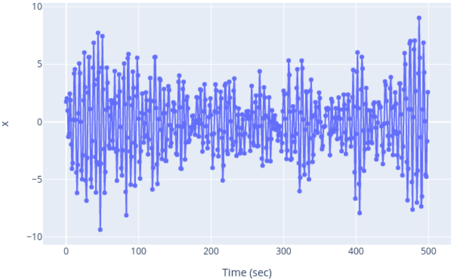
$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_q w_{t-q}$$

for $t = 0, \pm 1, \pm 2, \ldots$, where $\{w_t\}$ is a white noise process and $\theta_1, \ldots, \theta_p$ are constants ($\theta_p \neq 0$).

Generation of time series: autoregressive model

Many neural time series, like local field potential recordings, exhibit oscillations of the type of sine waves. The **autoregressive model** generates oscillations (example).

$$x_t = x_{t-1} - 0.9x_{t-2} + w_t$$



Autoregressive model of order p

Definition 4

An **autoregressive model** of order p, abbreviated as **AR**(p), is of the form

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + w_t$$

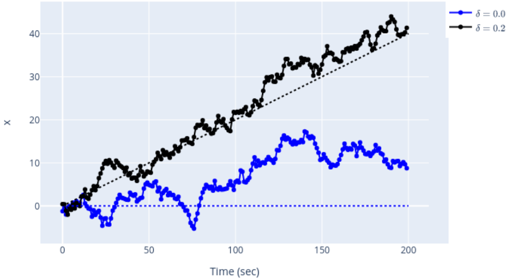
for $t = 0, \pm 1, \pm 2, \ldots$, where w_t is a white noise process, $\phi_1, \ldots, \phi_p$ are constants ($\phi_p \neq 0$).

Generation of time series: random walk with drift

The **random walk with noise** model is used to characterise trends in time series (example).

$$x_t = \delta + x_{t-1} + w_t \qquad (2)$$

for $t = 1, 2, \ldots$, with initial condition $x_0 = 0$, and where $\{w_t\}$ is a white noise random process.



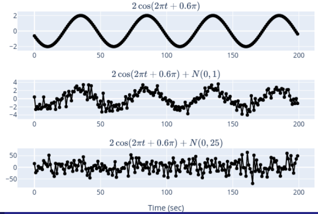
Generation of time series: signal plus noise

Many realistic models of time series assume an underlying signal with a periodic variation contaminated by adding a random noise (example).

$$x_t = 2\cos(2\pi\frac{t}{50} + 2\pi\frac{15}{50}) + w_t$$

$$A\cos(2\pi\omega t + \phi)$$

for $t = 0, \pm1, \pm2, \ldots$, where $A = 2, \omega = 1/50, \phi = 2\pi 15/50$. Lecture on spectral analysis of time series.



Contents

Mean function

Definition 5 (Mean function)

The mean function, μ_t, is defined as $\mu_t = E\{x_t\}$.

Example (Mean function of a moving average model)

Calculate the mean function of the moving average model in Eq. 8.

$$E\{\nu_t\} = \frac{1}{3}(E\{w_{t-1}\} + E\{w_t\} + E\{w_{t+1}\}) = \frac{1}{3}(0 + 0 + 0) = 0$$

Mean function

Example (Mean function of the autoregressive model of order 1)

Calculate the mean function of the autoregressive model of order 1, AR(1), in Eq. 3.

$$x_t = \phi x_{t-1} + w_t \tag{3}$$

For $|\phi| < 1$, the AR(1) model in Eq. 3 can be represented as a moving average of infinite order MA(∞). See MA(∞) representation of AR(1) random process in Appendix. Then

$$x_t = \sum_{i=0}^{\infty} \phi^i w_{t-i}$$

$$E\{x_t\} = \sum_{i=0}^{\infty} \phi^i E\{w_{t-i}\} = \sum_{i=0}^{\infty} \phi^i 0 = 0$$

Mean function

Example (Mean function of the random walk with drift model)

Calculate the mean function of the random noise with drift model, in Eq. 2.

The random noise with drift model in Eq. 2 can be represented as

$$x_t = t\delta + \sum_{i=1}^{t} w_i$$

For a proof see Alternative representation of the random walk with drift in the Appendix. Then

$$E\{x_t\} = \delta t + \sum_{i=1}^{t} E\{w_i\} = \delta t + \sum_{i=1}^{t} 0 = \delta t$$

In the figure of samples of the random noise with drift random process the dotted lines plot the mean of the process.

Autocovariance function

Definition 6 (Autocovariance function)

The autocovariance function is defined as
$\gamma(s, t) = cov(x_s, x_t) = E\{(x_s - \mu_s)(x_t - \mu_t)\}$.

Note

For $s = t$ the autocovariance function reduces to the variance function, because $\gamma(t, t) = E\{(x_t - \mu_t)^2\} = var(t)$.

Definition 7 (Autocorrelation function)

The autocorrelation function is defined as $\rho(s, t) = \frac{\gamma(t,s)}{\sqrt{\gamma(t,t)\gamma(s,s)}}$.

Autocovariance function

Example (Autocovariance function of moving average)

Calculate the autocovariance function of the moving average model in Eq. 8.

$$\gamma_\nu(s,t) = cov(\nu_s, \nu_t) = cov(\frac{1}{3}(w_{s-1} + w_s + w_{s+1}), \frac{1}{3}(w_{t-1} + w_t + w_{t+1}))$$

If s=t:

$$\begin{aligned}
\gamma_\nu(t,t) &= cov(\nu_t, \nu_t) = cov(\frac{1}{3}(w_{t-1} + w_t + w_{t+1}), \frac{1}{3}(w_{t-1} + w_t + w_{t+1})) \\
&= \frac{1}{9}\left(cov(w_{t-1}, w_{t-1}) + cov(w_t, w_t) + cov(w_{t+1}, w_{t+1})\right) \\
&= \frac{1}{9}\left(\sigma_w^2 + \sigma_w^2 + \sigma_2^2\right) = \frac{3}{9}\sigma_w^2
\end{aligned}$$

Autocovariance function

Example (Autocovariance function of moving average)

If s=t+1:

$$
\begin{aligned}
\gamma_\nu(t+1,t) &= cov(\nu_{t+1},\nu_t) \\
&= cov(\frac{1}{3}(w_t + w_{t+1} + w_{t+2}), \frac{1}{3}(w_{t-1} + w_t + w_{t+1})) \\
&= \frac{1}{9}\left(cov(w_t, w_t) + cov(w_{t+1}, w_{t+1})\right) \\
&= \frac{1}{9}\left(\sigma_w^2 + \sigma_w^2\right) = \frac{2}{9}\sigma_w^2
\end{aligned}
$$

Autocovariance function

Example (Autocovariance function of moving average)

If s=t+2:

$$
\begin{aligned}
\gamma_\nu(t+2,t) &= cov(\nu_{t+2},\nu_t) \\
&= cov(\frac{1}{3}(w_{t+1} + +w_{t+2} + w_{t+3}), \frac{1}{3}(w_{t-1} + w_t + w_{t+1})) \\
&= \frac{1}{9}\left(cov(w_{t+1}, w_{t+1})\right) \\
&= \frac{1}{9}\sigma_w^2
\end{aligned}
$$

Autocovariance function

Example (Autocovariance function of moving average)

$$\gamma_\nu(s,t) = \begin{cases} \frac{3}{9}\sigma_w^2 & \text{if} \quad s = t, \\ \frac{2}{9}\sigma_w^2 & \text{if} \quad |s-t| = 1, \\ \frac{1}{9}\sigma_w^2 & \text{if} \quad |s-t| = 2, \\ 0 & \text{if} \quad |s-t| > 2. \end{cases}$$

Autocovariance function

Example (Autocovariance function of AR(1))

Calculate the autocovariance function of the autoregressive model of order 1 in Eq. 7.

For $|\phi| < 1$, an AR(1) model (Eq. 3) can be represented as a moving average of infinite order MA(∞). See MA(∞) representation of AR(1) random process in Appendix.

$$x_t = \sum_{i=0}^{\infty} \phi^i w_{t-i}$$

Autocovariance function

Example (Autocovariance function of AR(1))

$$\gamma(t-h,t) = E\{(x_{t-h}-\mu_{t-h})(x_t-\mu_t)\} = E\{x_{t-h}x_t\} = E\left\{\left(\sum_{i=0}^{\infty}\phi^i w_{t-h-i}\right)\left(\sum_{j=0}^{\infty}\phi^j w_{t-j}\right)\right\} \tag{4}$$

$$= E\left\{\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\phi^i\phi^j w_{t-h-i}w_{t-j}\right\} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\phi^i\phi^j E\{w_{t-h-i}w_{t-j}\} \tag{5}$$

$$= \sum_{i=0}^{\infty}\phi^i\phi^{i+h}E\{w_{t-h-i}^2\} = \phi^h\sigma_w^2\sum_{i=0}^{\infty}\phi^{2i} = \phi^h\sigma_w^2\frac{1}{1-\phi^2} \quad \square \tag{6}$$

Autocovariance function

Example (Autocovariance function of AR(1))

Note

1. the last equality in Eq. 4 requires $|\phi| < 1$ to use the MA(∞) representation of AR(1) random process, which guarantees that $\sum_{i=0}^{\infty} \phi^{2i} = \frac{1}{1-\phi^2}$ in Eq. 6.
2. because the noise w_t is uncorrelated, the expectation in the right hand side of Eq. 5 will be different from zero only when $t - h - i = t - j$, or when $j = h + i$. Thus, the double summation in this equation will reduce to a single summation replacing j by $h + i$ in the lefmost summation in Eq. 6.

Contents

Strictly stationary time series

Definition 8 (Strict stationarity)

A **strictly stationary time series** is one for which the probabilistic behaviour of every collection of values

$$\{x_{t_1}, \dots, x_{t_n}\}$$

is identical to that of any shifted set

$$\{x_{t_1+h}, \dots, x_{t_n+h}\}$$

That is

$$P(x_{t_1} < c_1, \dots, x_{t_k} < c_k) = P(x_{t_1+h} < c_1, \dots, x_{t_k+h} < c_k)$$

for all $k = 1, 2, \dots$, all time points $t_1, t_2, \dots, t_k$, all numbers $c_1, c_2, \dots, c_k$, and all time shifts $h = 0, \pm 1, \pm 2, \dots$.

Weakly stationary time series

Definition 9 (Weak or wide-sense stationarity)

A **weakly** or **wide-sense stationary time series** is a finite-variance process such that:

1. the mean function, μ_t, is constant and does not depend on time t, and
2. the autocovariance function, $\gamma(s, t)$, depends on s and t only through their difference $|s - t|$.

Contents

Sample mean, autocovariance and autocorrelation

Definition 10 (Sample mean)

Let $x_1, \ldots, x_n$ be observations from a time series. The **sample mean** of $x_1, \ldots, x_n$ is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Definition 11 (Sample autocovariance)

The **sample autocovariance function** is

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{i=1}^{n-|h|} (x_{i+|h|} - \bar{x})(x_i - \bar{x}), \quad -n < h < n$$

Sample mean, autocovariance and autocorrelation

Definition 10 (Sample autocorrelation)

The **sample autocorrelation function** is

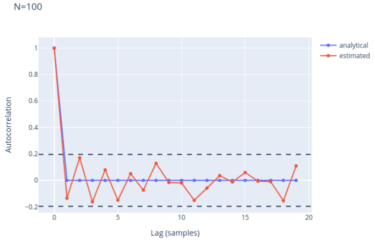
$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

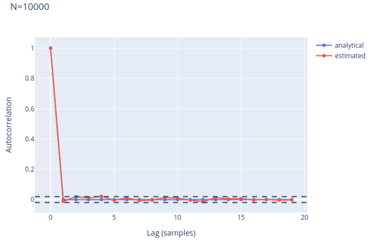
Theorem 11 (Distribution of sample autocorrelation for white noise)

For white noise, and a sample of size n, the sample autocorrelations, $\hat{\gamma}(h), h > 0$, are approximately independent and identically distributed (with a Normal distribution of mean 0 and standard deviaiton $1/\sqrt{n}$), for large n (Brockwell and Davis, 1991). Hence 95% of the sample autocorrelations should fall between the bound $\pm 1.96/\sqrt{n}$

Analytical and estimated autocorrelation function for white noise

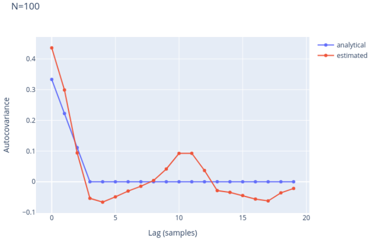
Simulate a white noise time series with N=100 and N=100,000 samples. For each N, plot the analytical and estimated autocorrelation function. Include the 95% confidence interval of the autocorrelation function. Solution.

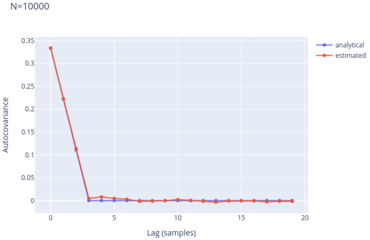




Analytical and estimated autocovariance function for MA

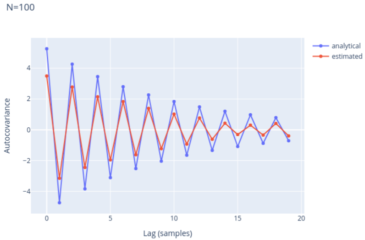
Simulate the previous moving average time series with N=100 and N=100,000 samples. For each N, plot the analytical and estimated autocovariance function. Solution.

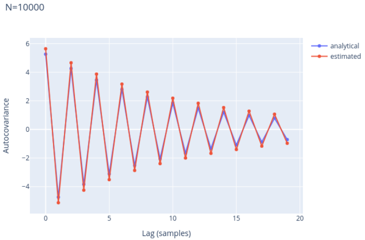




Analytical and estimated autocovariance function for AR(1)

Simulate an AR(1) time series with N=100 and N=100,000 samples, $\phi = -0.9$ and $\sigma_w = 1.0$. For each N, plot the analytical and estimated autocovariance function.





Contents

Forecasting

Forecasting is the problem of predicting the value of $x_{n+h}, h > 0$, of a stationary time series, in term of the previous m values$\{x_n, \ldots, x_{n-(m-1)}\}$. The mean of such predictor is

$$\text{mean}(\text{pred}(x_{m+h}|x_n, \ldots, x_{n-(m-1)})) = \mu + \mathbf{a_m}^\intercal \begin{bmatrix} x_n - \mu \\ \ldots \\ x_{n-(m-1)} - \mu \end{bmatrix}$$

and its variance is

$$\text{var}(\text{pred}(x_{n+h}|x_n, \ldots, x_{n-(m-1)})) = \gamma(0) - \mathbf{a_m}^\intercal \gamma_m(h)$$

Forecasting

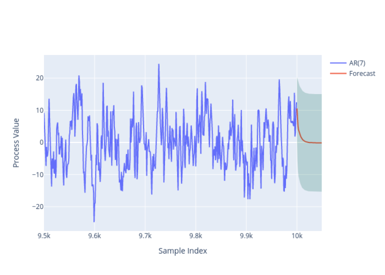
with

$$
\begin{aligned}
\Gamma_m \mathbf{a_m} &= \gamma_m(h) \\
\Gamma_m = [\gamma(i-j)]_{i,j=1}^m &= \begin{bmatrix}
\gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) & \dots & \gamma(m-1) \\
\gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) & \dots & \gamma(m-2) \\
\gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) & \dots & \gamma(m-3) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\gamma(m-1) & \gamma(m-2) & \gamma(m-3) & \gamma(m-4) & \dots & \gamma(0)
\end{bmatrix} \\
\mathbf{a_m} &= [a_1, \dots, a_m]^\intercal \\
\gamma_m(h) &= [\gamma(h), \gamma(h+1), \dots, \gamma(h+m-1)]^\intercal
\end{aligned}
$$

AR(p) forecasting example

Example (Forecasting with an AR(p) model)

Simulate N=10,000 samples from an AR(7) stochastic process with $\phi = [5.0/6, -1.0/6, 0.5/6, -0.25/6, 0.5/6, -0.1/6, 0.05/6]$ and $\sigma_w = 5.0$. Use the last 500 samples to forecast 50 samples (i.e., $n = 10,000, m = 500, h = 1, \ldots, 50$).



Summary

References

- Brockwell and Davis (2002)
- Shumway and Stoffer (2016)
- Priestley (1981)

Contents

MA(∞) representation of AR(1) random process

Claim 1

Let $|\phi| < 1$*, then*

$$x_t = \phi x_{t-1} + w_t \quad \textit{if and only if} \tag{7}$$

$$x_t = \sum_{i=0}^{\infty} \phi^i w_{t-i} \tag{8}$$

MA(∞) representation of AR(1) random process

Proof.

We first show that x_t, as defined in Eq. 8, satisfies Eq. 7.

$$\phi x_{t-1} = \phi \sum_{i=0}^{\infty} \phi^i w_{t-1-i} = \phi \sum_{j=1}^{\infty} \phi^{j-1} w_{t-j} = \sum_{j=1}^{\infty} \phi^j w_{t-j}$$

$$\phi x_{t-1} + w_t = \sum_{j=0}^{\infty} \phi^j w_{t-j} = x_j$$

MA(∞) representation of AR(1) random process

Proof.

We now show that Eq. 8 is the unique solution to Eq. 3. Suppose y_t is stationary and satisfies Eq. 3, then

$$\begin{aligned}
y_t &= \phi y_{t-1} + w_t \\
&= \phi(\phi y_{t-2} + w_{t-1}) + w_t = \phi^2 y_{t-2} + \phi w_{t-1} + w_t \\
&= \phi^{t+1} y_{t-(t+1)} + \phi^t w_t + \ldots + \phi w_{t-1} + w_t \\
&= \phi^{k+1} y_{t-(k+1)} + \sum_{i=0}^{k} \phi^k w_{t-i}
\end{aligned}$$

$$E\left\{\left(y_t - \sum_{i=0}^{k} \phi^i w_{t-i}\right)^2\right\} = \phi^{2k+2} E\{y^2_{t-(k+1)}\} = \phi^{2k+2}\sigma^2$$

$$E\left\{\left(y_t - \sum_{i=0}^{\infty} \phi^i w_{t-i}\right)^2\right\} = \lim_{k\to\infty} \phi^{2k+2}\sigma^2 = 0$$

MA(∞) representation of AR(1) random process

Proof.

Thus y_t equals $\sum_{i=0}^{\infty} \phi^i w_{t-i}$ in the mean-squared sense. $\square$

Alternative representation of the random walk with drift

Claim 2

For $t = 1, 2, \ldots$, the random noise with drift model in Eq. 2 can be represented as

$$x_t = \delta t + \sum_{i=1}^{t} w_i$$

Alternative representation of the random walk with drift

Proof.

We prove this claim by induction. We define a property P_t, we demonstrate the P_1 holds, and then we demonstrate that if P_t holds then P_{t+1} also holds. This strategy then demonstrates that P_t holds for $t = 1, 2, \ldots$. Define

$$P_t: \quad x_t = \delta t + \sum_{i=1}^{t} w_i$$

We first prove P_1

$$P_1: \quad x_1 = \delta + x_0 + w_1 = \delta + w_1 \tag{9}$$

Next we assume that P_t holds and prove that P_{t+1} also holds

$$\begin{aligned} P_t \rightarrow P_{t+1}: \quad x_{t+1} &= \delta + x_t + w_{t+1} = \delta + \left(\delta t + \sum_{i=1}^{t} w_i\right) + w_{t+1} \qquad (10) \\ &= \delta(t+1) + \sum_{i=1}^{t+1} w_i \quad \square \end{aligned}$$

Alternative representation of the random walk with drift

Note

1. the first equality in Eq. 9 follows from the definition of the random walk with drift in Eq. 2.
2. the last equality in Eq. 10 uses the hypothesized P_t.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: Theory and methods*. Springer-Verlag, 2nd edition.

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer.

Priestley, M. (1981). Spectral analysis and time series.

Shumway, R. H. and Stoffer, D. S. (2016). *Time series analysis and its applications*. Springer, 4 edition.