Time series

Kristian Wichmann

March 26, 2017

1 Stochastic processes and time series

Definition 1.1. *Let T be a set, called the index set, and $(\Omega, \mathcal{F})$ be a measurable space. Then a stochastic process is a set of probability distributions and associated random variables $\{X_t | t \in T\}$, i.e. $\mathcal{F}$-measurable functions $X_t : \Omega \to \mathbf{R}$.*

Definition 1.2. *For each $\omega \in \Omega^T$ we can define a function $x : T \to \mathbf{R}$ by:*

$$x(t) = x_t = X_t(\omega(t)) \tag{1.1}$$

These are known as realizations or sample-paths of the stochastic process.

Definition 1.3. *A stochastic process for which the index set $T \subseteq \mathbf{Z}$ is called a time series.*

As long as there's no chance of confusion, we will use the term 'time series' interchangeably for the stochastic process itself, and any relevant realizations of it.

2 White noise

White noise is a time series W_t with the following characteristics:

- The W_t's are pairwise uncorrelated.
- $E(W_t) = 0$
- The variances for all W_t are equal and finite.

We then write $W_t \sim \text{wn}(0, \sigma_w^2)$, where σ_w^2 is the variance.

A slightly stronger criterium, is for the random variables to be i.i.d. In this case we may write $W_t \sim \text{iid}(0, \sigma_w^2)$.

Even stronger, the noise may be normally distributed. We then write $W_t \sim \text{iid } N(0, \sigma_w^2)$.

2.1 Moving averages of white noise

In order to smooth out a white noise time series W_t, one could average over the values in the immediate vicinity of each t. Such a series might be realized as:

$$V_t = \frac{1}{3}(W_{t-1} + W_t + W_{t+1}) \tag{2.1}$$

V_t is not a white noise, as the first criterium is no longer satisfied. Such a linear combination of time series data/variables is known as a *filter*.

2.2 Autoregression

Imagine two initial values x_1, x_2 being given. Then we may form a time series from a white noise w_t as follows:

$$x_t = \alpha x_{t-2} + \beta x_{t-1} + w_t \tag{2.2}$$

Here α and β are some constants. This is an example of an *autoregression* - using the last few data points to make predictions about the next one. More about these later.

3 Random walks

3.1 Simple random walk

Let the random variables $Y_1, Y_2, Y_3, \ldots$ be i.i.d. with the distribution:

$$P(Y_t = 1) = 1/2, \quad P(Y_t = -1) = 1/2 \tag{3.1}$$

Now let a time series be defined as:

$$X_0 = 0, \quad X_t = \sum_{i=1}^{t} Y_i \tag{3.2}$$

This is known as the *simple random walk*.

3.1.1 Asymptotic behaviour

For each of the Y's we have:

$$E[Y_t] = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (-1) = 0, \quad \text{var}[Y_t] = \frac{1}{2}1^2 + \frac{1}{2}(-1)^2 = 1 \tag{3.3}$$

So, according to the central limit theorem, for large t, X_t will be approximately normally distributed:

$$X_t \sim N(0, t), \quad t \gg 1 \tag{3.4}$$

This means that the standard deviation for large t is $\sqrt{t}$.

3.2 Random walk

Instead of the discrete variables Y_t given above, we might use white noise $W_t \sim \text{wn}(0, \sigma_w^2)$. The result is the *random walk* with variance σ_w^2:

$$X_0 = 0, \quad X_t = \sum_{i=1}^{t} W_i \tag{3.5}$$

The second equation could also have been expressed $X_t = X_{t-1} + W_t$.

3.2.1 Asymptotic behaviour

Again, we may apply the central limit theorem. For large t we approximately have:

$$X_t \sim N(0, t\sigma_w^2), \quad t \gg 1 \tag{3.6}$$

So the standard deviation for large t is $\sqrt{t}\sigma_w$.

3.3 Random walk with drift

We now modify the random walk by adding a linear term:

$$X_0 = 0, \quad X_t = \delta t + \sum_{i=1}^{t} W_i \tag{3.7}$$

Here δ is some constant. The second equation could also have been expressed $X_t = \delta + X_{t-1} + W_t$.

3.3.1 Asymptotic behaviour

In this case, the central limit theorem gives us

$$X_t \sim N(\delta t, t\sigma_w^2), \quad t \gg 1 \tag{3.8}$$

The standard deviation for large t is still $\sqrt{t}\sigma_w$.

4 Time series descriptors

Formally, a time series can be described by the family of finite point distribution functions:

$$F(c_1, c_2, \ldots, c_n) = P(x_{t_1} \leq c_1, x_{t_2} \leq c_2, \ldots x_{t_n} \leq c_n) \tag{4.1}$$

Here $t_1, t_2, \ldots, t_n \in T$. While this family does contain all the information about the time series, there's a number of other descriptors that are usually a lot easier easier to compute and/or visualize.

4.1 The mean function

For a time series X_t, this is simply the series of expectation values:

$$\mu_X(t) = \mu_t = E(X_t) \tag{4.2}$$

4.1.1 Examples

The different types of white noise all have a mean function of zero by definition. A moving average is a linear combination of such, so the mean function is still zero. The simple random walk and the random walk are sums of variables with expectation zero, so again we get zero. However, the random walk with drift has a non-zero mean function:

$$\mu(t) = \delta t \tag{4.3}$$

4.2 The autocovariance function

For a time series X_t, the *autocovariance function* is:

$$\gamma_X(s,t) = \text{cov}(X_s, X_t) \tag{4.4}$$

When $s = t$ we get the *variance function* for the series.

Using the definition of covariance, this can be written:

$$\text{cov}(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)] = E[X_s X_t - X_s\mu_t - \mu_s X_t + \mu_s\mu_t)] \tag{4.5}$$

Here $\mu_s = E(X_s), \mu_t = E(X_t)$. Since these are numbers, this is:

$$E(X_s X_t) - E(X_s)\mu_t - \mu_s E(X_t) + \mu_s\mu_t = E(X_s X_t) - \mu_s\mu_t - \mu_s\mu_t + \mu_s\mu_t \tag{4.6}$$

So the autocovariance function can also be calculated as follows:

$$\gamma_X(s,t) = E(X_s X_t) - \mu_s\mu_t \tag{4.7}$$

4.2.1 Examples

For the various types of white noise, these are uncorrelated pr. definition when $s \neq t$. When $s = t$ we simply get the variance σ_w^2. This may be summed up as:

$$\gamma_W(s,t) = \delta_{s,t}\sigma_w^2 \tag{4.8}$$

For the moving average, things get a little more interesting. First, when $s = t$:

$$\gamma_V(t,t) = \text{cov}\left(\frac{1}{3}(W_{t-1} + W_t + W_{t+1}), \frac{1}{3}(W_{t-1} + W_t + W_{t+1})\right) \tag{4.9}$$

Since the W's are uncorrelated unless the times are equal, this is:

$$\frac{1}{9}(\text{cov}(W_{t-1}, W_{t-1}) + \text{cov}(W_t, W_t) + \text{cov}(W_{t+1}, W_{t+1})) = \frac{1}{3}\sigma_w^2 \tag{4.10}$$

When $s = t \pm 1$ we get a similar calculation, but now only two of the terms contribute, so we get $\gamma_V(t \pm 1, t) = \frac{2}{9}\sigma_w^2$. Finally when $s = t \pm 2$ only one term contribute so $\gamma_V(t \pm 2, t) = \frac{1}{9}\sigma_w^2$. Otherwise, the autocorrelation function is zero. Here γ only depends on $|s - t|$.

For a simple random walk we have:

$$\gamma_X(s, t) = \text{cov}\left(\sum_{i=1}^{s} Y_i, \sum_{i=1}^{t} Y_i\right) \tag{4.11}$$

Since the Y's are i.i.d. we only get a contribution at equal times. Hence we get:

$$\gamma_X(s, t) = \min\{s, t\} \tag{4.12}$$

For a random walk, the situation is the same as above, but now each contribution is σ_w^2, so:

$$\gamma_X(s, t) = \min\{s, t\} \cdot \sigma_w^2 \tag{4.13}$$

For a random walk with drift, the added constants do not correlate with the Y's so the result is the same as the case without drift.

4.3 The autocorrelation function

It is often convenient to normalize the autocovariance function to get the *autocorrelation function* (or ACF for short):

$$\rho_X(s, t) = \frac{\gamma_X(s, t)}{\sqrt{\gamma_X(s, s)\gamma_X(t, t)}} \tag{4.14}$$

Like ordinary correlations, the autocorrelation function takes on values between -1 and 1.

4.3.1 Examples

For white noise, the autocorrelation is zero unless $s = t$. Then we get:

$$\rho_W(t, t) = \frac{\sigma_w^2}{\sqrt{\sigma_w^2 \sigma_w^2}} = 1 \tag{4.15}$$

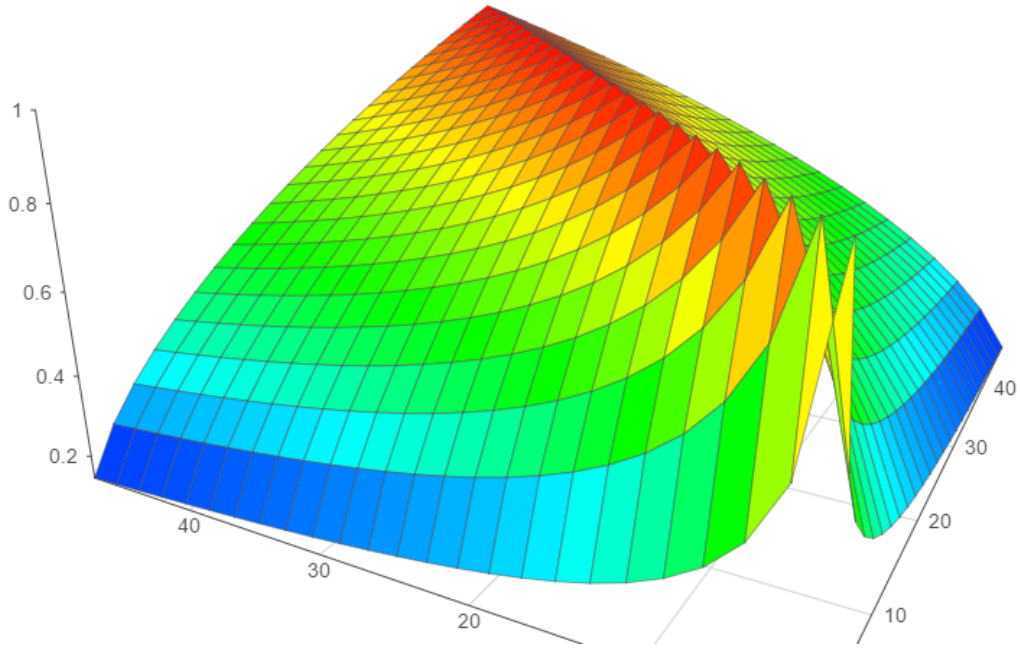

Figure 1: The autocorrelation function for a random walk.

To sum up $\rho_W(s,t) = \delta_{s,t}$.

For a simple random walk we get:

$$\rho_X(s,t) = \frac{\min\{s,t\}}{\sqrt{st}} \tag{4.16}$$

For a random walk (with or without drift) we get the same result:

$$\rho_X(s,t) = \frac{\min\{s,t\} \cdot \sigma_w^2}{\sqrt{s\sigma_w^2 t\sigma_w^2}} = \frac{\min\{s,t\}}{\sqrt{st}} \tag{4.17}$$

Figure 1 graphs ρ_X as a function of s and t. Note how is is 1 along the $s = t$ line.

One may wonder what happens to adjacent times:

$$\rho_X(t,t+1) = \frac{t}{\sqrt{t}\sqrt{t+1}} = \sqrt{\frac{t}{t+1}} \tag{4.18}$$

As $t \to \infty$, this fraction tends to 1. In other words, the adjacent values of a random walk becomes progressively more correlated for large times. Figure 2 graphs this.

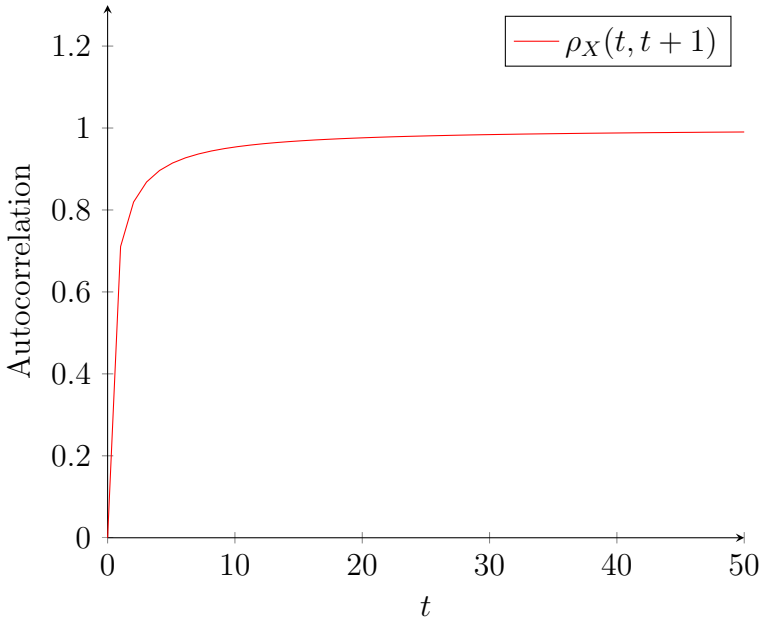

Figure 2: Autocorrelation of adjacent times for random walk.

5 Correlation between time series

If we have two time series X_t and Y_t, we may wish to evaluate if they are correlated in time. Analogously to the last section, we define the *cross-covariance function* as:

$$\gamma_{XY}(s,t) = \text{cor}(X_s, Y_t) \tag{5.1}$$

Similarly, the *cross-correlation function* is:

$$\rho_{XY}(s,t) = \frac{\gamma_{XY}(s,t)}{\gamma_X(s,s)\gamma_Y(t,t)} \tag{5.2}$$

6 Stationarity

6.1 Strict stationarity

A time series X_t with finite variance for all times, is said to be *strictly stationary* if its properties are invariant under time translation in the following sense:

$$P(x_{t_1} \leq c_1, x_{t_2} \leq c_2, \ldots x_{t_n} \leq c_n) = P(x_{t_1+h} \leq c_1, x_{t_2+h} \leq c_2, \ldots x_{t_n+h} \leq c_n) \tag{6.1}$$

This should be true for all times and translations, and any number of points in time n.

This is a very strong condition. Specifically, for $n = 1$ we get:

$$P(x_s \leq c) = P(x_t \leq c) \tag{6.2}$$

This means, that the distribution must be the same for all times. Specifically, the mean function μ_X must be a constant.

For $n = 2$ it similarly implies, that the autocovariance - and hence the autocorrelation - can only depend on the different $t - s$.

6.2 Weak stationarity

As noted, strict stationarity is very strong. And it is both rare and hard to demonstrate in a model. Therefore, a weaker condition known as *weak stationarity* is useful. A time series X_t is called weakly stationary (or simply stationary) if:

- It has finite variance at all times.
- The mean function μ_X is a constant.
- The autocovariance function is time translation invariant: $\gamma_X(t, s) = \gamma_X(s + h, t + h)$. In other words a function $\gamma_X(h) = \gamma_X(t, t + h)$ exists.

From the last section it is clear, that a strictly stationary time series is also (weakly) stationary. The converse is not generally true.

Because of this, for a stationary time series we may write $\mu_X(t) = \mu$ and $\gamma_X(s, t) = \gamma_X(t - s)$. Therefore, the autocorrelation function can be written:

$$\rho_X(h) = \frac{\gamma_X(t, t + h)}{\sqrt{\gamma_X(t, t)\gamma_x(t + h, t + h)}} = \frac{\gamma_X(h)}{\gamma_X(0)} \tag{6.3}$$

In addition, because of the symmetry of the covariance, we also have:

$$\gamma_X(h) = \gamma_X(t, t + h) = \text{cov}(X_t, X_{t+h}) = \tag{6.4}$$

$$\text{cov}(X_{t+h}, X_t) = \gamma_X(t + h, t) = \gamma_X(-h) \tag{6.5}$$

This implies, that the autocorrelation function is a function of the absolute value of h only: $\rho_X(h) = \rho_X(|h|)$.

6.3 Weak dependence

A stationary time series is said to be *weakly dependent*, if the autocovariance function $\gamma_X(h)$ tends to zero as $h \to \infty$.

6.3.1 Examples

All the types of white noise are clearly stationary. If the white noise is i.i.d., they are also strictly stationary, since equation 6.1 reduces to a product of distribution functions.

Moving averages are stationary as is evident from the properties derived in previous sections.

Random walks turn out not to be stationary, as correlation is dependent on both times chosen, not just the difference.

6.4 Joint stationarity

Two time series X_t and Y_t are said to be *jointly stationary* if they are both stationary and if the cross-covariance function is a function of lag only:

$$\gamma_{XY}(t+h,t) = \gamma_{XY}(h) \tag{6.6}$$

This function satisfies:

$$\gamma_{XY}(h) = \gamma_{XY}(t+h,t) = \text{cov}(X_{t+h}, Y_t) = \tag{6.7}$$

$$\text{cov}(Y_t, X_{t+h}) = \gamma_{YX}(t, t+h) = \gamma_{YX}(-h) \tag{6.8}$$

We can now define the *cross-correlation function* (or CCF for short):

$$\rho_{XY}(h) = \frac{\gamma_{XY}(h)}{\sqrt{\gamma_X(0)\gamma_Y(0)}} \tag{6.9}$$

6.4.1 Example

Let W_t be a white noise time series. Define two new time series as:

$$X_t = W_t + W_{t-1}, \qquad Y_t = W_t - W_{t-1} \tag{6.10}$$

First, we need to check that each series is stationary in itself. It is easily seen, that $\gamma_X(t,t) = \gamma_Y(t,t) = 2\sigma^2$, $\gamma_X(t+1,t) = \gamma_Y(t+1,t) = \sigma^2$, and $\sigma_Y(t+1,t) = \gamma_Y(t-1,t) = -\sigma^2$ are the only non-zero autocovariance function values. So both are stationary.

What about the cross-covariance function?

$$\gamma_{XY}(t,t) = \text{cov}(W_t + W_{t-1}, W_t - W_{t_1})\sigma^2 - \sigma^2 = 0 \tag{6.11}$$

$$\gamma_{XY}(t+1,t) = \text{cov}(W_{t+1} + W_t, W_t - W_{t-1}) = \sigma^2 \tag{6.12}$$

$$\gamma_{XY}(t-1,t) = \text{cov}(W_{t-1} + W_{t-2}, W_t - W_{t-1}) = -\sigma^2 \tag{6.13}$$

All other values are zero. Hence, X_i and Y_i are jointly stationary with cross-correlation function:

$$\rho_{XY}(0) = 0, \quad \rho_{XY}(\pm 1) = \frac{\pm\sigma^2}{\sqrt{2\sigma^2 \cdot 2\sigma^2}} = \pm\frac{1}{2} \tag{6.14}$$

7 MA(1) time series

We've seen moving averages before above. The general MA(q)-model is:

$$X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q} \tag{7.1}$$

Here ε_t is white noise with variance σ^2. We will focus on the case $q = 1$:

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1} \tag{7.2}$$

7.1 Autocovariance and autocorrelation

We've seen before, that moving averages are stationary, so autocovariance depends only on the time difference h. It is clear, that it is zero for $h > q$. For $q = 1$, this means there's only two non-zero cases:

$$\gamma_X(0) = \mathrm{var}(\varepsilon_t + \theta\varepsilon_{t-1}) = \mathrm{var}(\varepsilon_t) + \theta^2\mathrm{var}(\varepsilon_{t-1}) = (1+\theta^2)\sigma^2 \tag{7.3}$$

And:

$$\gamma_X(\pm 1) = \mathrm{cov}(\varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_{t+1} + \theta\varepsilon_t) = \theta\sigma^2 \tag{7.4}$$

The corresponding autocorrelation function is then:

$$\rho_X(h) = \begin{cases} 1 & \text{for } h = 0 \\ \frac{\theta}{1+\theta^2} & \text{for } h = \pm 1 \end{cases} \tag{7.5}$$

8 AR(1) time series

As noted above, an autoregressive model is one, where the values depend linearly on the ones immediately before it, plus some noise. A series depending on the last p values is denoted as AR(p) for short. Here we will look closer at the case where $p = 1$. In other words, the model is:

$$X_t = \rho X_{t-1} + \varepsilon_t \tag{8.1}$$

Again, ε_t is white noise with variance σ^2.

8.1 Autocovariance and autocorrelation for stationary AR(1)

Let's start by considering the variance of X_t:

$$\mathrm{var}(X_t) = \mathrm{var}(\rho X_{t-1} + \varepsilon_t) = \rho^2\mathrm{var}(\rho X_{t-1}) + \sigma^2 \tag{8.2}$$

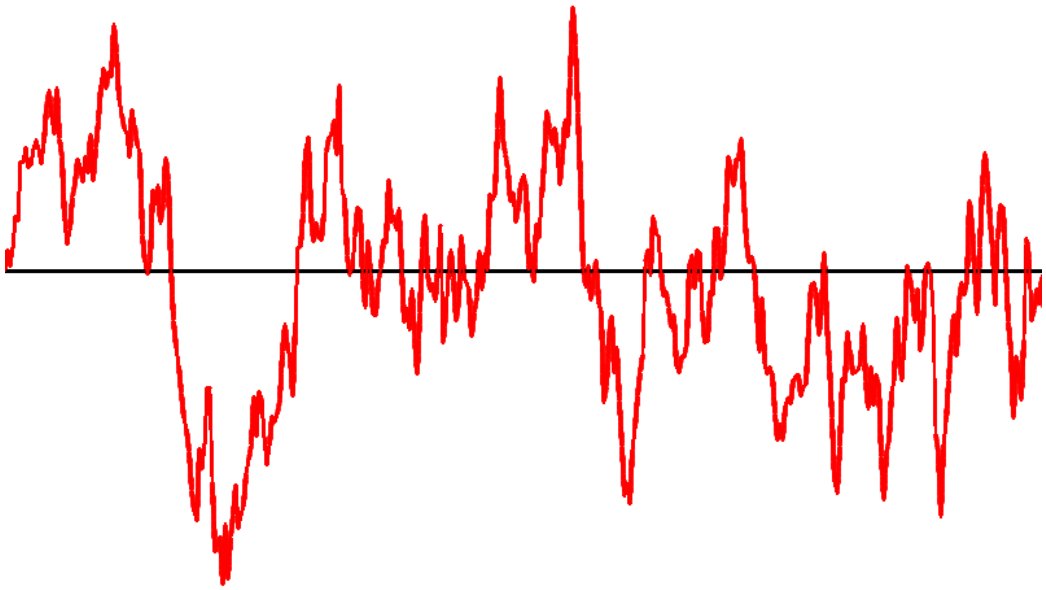

Figure 3: An example of an AR(1) process with $\rho = 0.95$.

Now, let us assume the series to be stationary. The specifically, the variance has to be constant, and therefore $\text{var}(X_t) = \text{var}(X_{t-1})$. This means:

$$\text{var}(X_t) = \rho^2 \text{var}(\rho X_t) + \sigma^2 \Leftrightarrow \text{var}(X_t) = \frac{\sigma^2}{1-\rho^2} \tag{8.3}$$

This shows some restrictions on ρ if stationarity is required. First of all $\rho = \pm 1$ gives an infinite variance. And for $|\rho| > 1$, the variance is negative according to the equation, which is also problematic. So we will assume $|\rho| < 1$ in the following. Figures 3, 4 and 5 show examples of time series with different values of ρ and the same σ^2. Note that the closer ρ is to 1, the longer it generally takes for the series to cross the time axis.

Now, we turn to the autocovariance function for $h \neq 0$. Let's start with $h = 1$. Here we see that:

$$X_t = \rho X_{t-1} + \varepsilon_t = \rho(\rho X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \rho^2 X_{t-2} + \rho \varepsilon_{t-1} + \varepsilon_t \tag{8.4}$$

In fact, for any $h > 0$, we could iterate this substitution process to get:

$$X_t = \rho^h X_{t-h} + \rho^{h-1} \varepsilon_{t-h+1} + \rho^{h-2} \varepsilon_{t-h+2} + \cdots + \varepsilon_t \tag{8.5}$$

Now, it is easy to compute the autocovariance, since only the X-term con-

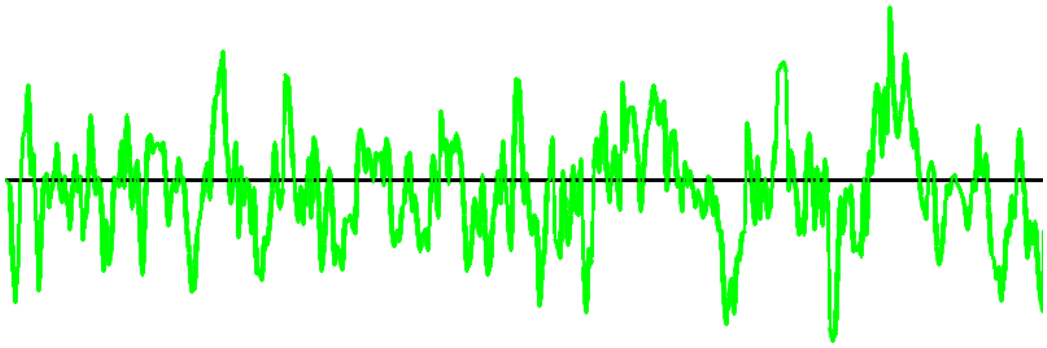

Figure 4: An example of an AR(1) process with $\rho = 0.7$.

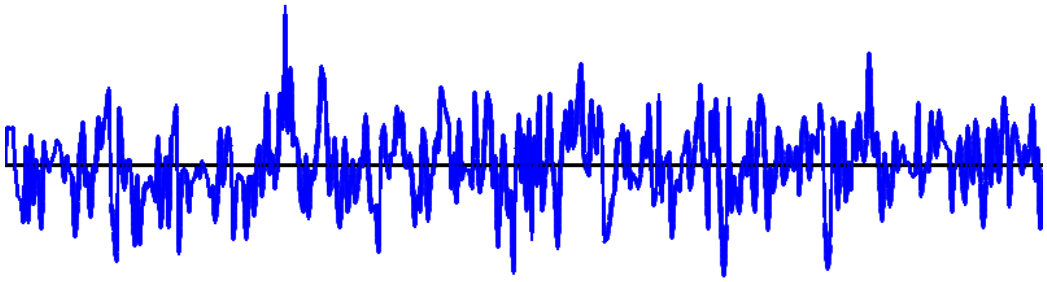

Figure 5: An example of an AR(1) process with $\rho = 0.2$.

tribute:

$$\gamma_X(h) = \text{cov}(X_{t-h}, X_t) = \text{cov}(X_{t-h}, \rho^h X_{t-h}) = \rho^h \text{var}(X_{t-h}) = \rho^h \frac{\sigma^2}{1-\rho^2} \tag{8.6}$$

Here equation 8.3 has been used. This means that the autocorrelation takes on the simple, exponential form:

$$\rho_X(h) = \rho^h \tag{8.7}$$

9 Lag, backshift and differencing

9.1 The lag and backshift operators

The *lag operator* L for a time series is defined as:

$$LX_t = X_{t-1} \tag{9.1}$$

The lag operator is invertible (at least when the index set is $\mathbb{Z}$):

$$L^{-1}X_t = X_{t+1} \tag{9.2}$$

In general L raised to an arbitrary integer power n is:

$$L^n X_t = X_{t-n} \tag{9.3}$$

The *backshift operator* B is identical to L, except for when conditional expectation values are calculated. Let E_t denote the expectation operator based on a realization of the time series up to time $t-1$. The lag operator shifts this realization time, while the backshift operator does not. In other words:

$$L^n E_t[X_{t+j}] = E_{t-n}[X_{t+j-n}], \quad B^n E_t[X_{t+j}] = E_t[X_{t+j-n}] \tag{9.4}$$

9.2 MA(q)-model with lag operator

The MA(q) model is:

$$X_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{9.5}$$

We might rewrite this using the lag operator:

$$X_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i L^i \varepsilon_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \tag{9.6}$$

9.3 AR(p)-model with lag operator

The general autoregressive model of order n is:

$$X_t = \rho_1 X_{t-1} + \rho_2 X_{t-2} + \ldots + \rho_p X_{t-p} + \varepsilon_t \tag{9.7}$$

This may be rewritten using the lag operator:

$$X_t = \rho_1 L X_t + \rho_2 L^2 X_t + \ldots + \rho_p L^p X_t + \varepsilon_t \tag{9.8}$$

The X-dependent part may be written as a *lag polynomial*:

$$X_t = (\rho_1 L + \rho_2 L^2 + \ldots + \rho_p L^p) X_t + \varepsilon_t \tag{9.9}$$

Or rearranging:

$$\underbrace{\left(1 - \sum_{i=1}^{p} \rho_i L^i\right)}_{\phi(L)} X_t = \varepsilon_t \tag{9.10}$$

Here, we have introduced $\phi(z)$, the *characteristic polynomial* of the process.

9.4 Rewriting the characteristic polynomial

According to equation 9.10, the characteristic polynomial can be written:

$$\phi(z) = -\rho_p z^p - \rho_{p-1} z^{p-1} - \ldots - \rho_1 z + 1 \tag{9.11}$$

According to the fundamental theorem of algebra, ϕ has p roots and can be factorized as follows:

$$\phi(z) = -\rho_p (z - z_1)(z - z_2) \cdots (z - z_p) \tag{9.12}$$

According to Vieta's equations, we have:

$$z_1 z_2 \cdots z_p = (-1)^p \frac{1}{-\rho_p} \Leftrightarrow -\rho_p = \frac{(-1)^p}{z_1 z_2 \cdots z_p} = (-\phi_1)(-\phi_2) \cdots (-\phi_p) \tag{9.13}$$

Here, the *reciprocal roots* $\phi_i = 1/z_i$ have been introduced. Equation 9.12 can now be written as:

$$\phi(z) = \prod_{i=1}^{p} (-\phi_i)(z - z_1) = \prod_{i=1}^{p} (1 - \phi_i z) \tag{9.14}$$

9.5 Differencing

When examining a time series, it is often useful to *difference* it. This means calculating (and usually plotting) the series $\Delta X_t = X_t - X_{t-1}$. It is essentially a discrete differentiation. Using the lag operator, this can be written as:

$$\Delta X_t = (1 - L)X_t \tag{9.15}$$

Similarly, higher order differings can be defined through:

$$\Delta^d X_t = (1 - L)^d X_t \tag{9.16}$$

Again, these are examples of lag polynomial operators, since they're expressed as a polynomial form of L.

9.6 I(d) time series

A time series of *order of integration* d, I(d), is a series which reduces to white noise after being differenced d times (usually, the lowest of such n's is used). So:

$$\Delta^d X_t = (1 - L)^d X_t = \varepsilon_t \tag{9.17}$$

Note the similarity to equations 9.10 and 9.14 - the left side corresponds to a characterstic polynomial with d-fold degenrate unit root.

As an example, an I(1) process is simply a random walk, since:

$$(1 - L)X_t = X_t - X_{t-1} = \varepsilon_t \Leftrightarrow X_t = X_{t-1} + \varepsilon_t \tag{9.18}$$

In general, such series are non-stationary.

10 MA(q) and AR(p) models

With the theory from the last section, we can now take a closer look at the general moving average and autoregressive models.

10.1 AR(p)

Using the decomposition of the characteristisc polynomial for equation 9.14, the model is:

$$(1 - \phi_i L)(1 - \phi_2 L) \cdots (1 - \phi_p)X_t = \varepsilon_t \tag{10.1}$$

10.2 The Koyck transformation

Reconsider an AR(1) model:

$$(1 - \phi L)X_t = \varepsilon_t \tag{10.2}$$

If we want to solve for X_t, we can think of this as inverting the operator $1 - \phi L$ and applying the inverse to each side of the operator. Now if L was simply a number, the inverse would be:

$$(1 - \phi L)^{-1} = \frac{1}{1 - \phi L} \tag{10.3}$$

If we further knew that $|\phi L| < 1$ we could write this as a geometric series:

$$(1 - \phi L)^{-1} = \sum_{i=0}^{\infty} \phi^i L^i \tag{10.4}$$

This is a good candidate for an inverse, so let's check that it actually is the inverse:

$$\left[\sum_{i=0}^{\infty} \phi^i L^i\right](1 - \phi L)\varepsilon_t = \left[\sum_{i=0}^{\infty} \phi^i L^i\right](\varepsilon_t - \phi\varepsilon_{t-1}) \tag{10.5}$$

Let the sum work on each term in the parenthesis:

$$\sum_{i=0}^{\infty} \phi^i L^i \varepsilon_t - \sum_{i=0}^{\infty} \phi^i L^i \phi\varepsilon_{t-1} = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} - \sum_{i=0}^{\infty} \phi^{i+1}\varepsilon_{t-i-1} = \varepsilon_t \tag{10.6}$$

In the last step, ee see that all terms in the left sum is exactly the ones in the right one, except for ε_t. Of course we have assumed convergence of all series, which means that we must have $|\phi| < 1$. Now, use this inverse on both sides of equation 10.2:

$$X_t = \left[\sum_{i=0}^{\infty} \phi^i L^i\right]\varepsilon_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-1} \tag{10.7}$$

This is known as the *Koyck transformation* - we have turned an AR(1) model into a MA(∞) model! Of course, we could have gotten the same result by repeatedly applying equation 8.1

10.3 Invertibility - the "opposite Koyck transformation"

It turns out that for an MA(1) process we can essentially use the same trick we did for the Koyck transformation. Write the model as:

$$X_t = \varepsilon_t - \theta\varepsilon_{t-1} = (1 - \theta L)\varepsilon_t \tag{10.8}$$

Here, the sign of θ is opposite of the original formulation. Now, we can try to isolate the noise ε_t by applying the inverse $(1-\theta L)^{-1}$ to both sides of the equation. Now, this only makes sense if $|\theta| < 1$. This criterion determines whether the model in invertible or not. Assuming this to be true, we have:

$$\varepsilon_t = \left[\sum_{i=0}^{\infty} \theta^i L^i\right] X_t = \sum_{i=0}^{\infty} \theta^i X_{t-i} = X_t + \theta X_{t-1} + \theta^2 X_{t-2} + \dots \tag{10.9}$$

This can be rewritten as:

$$X_t = \varepsilon_t - \theta X_{t-1} - \theta^2 X_{t_2} - \dots \tag{10.10}$$

In other words we have turned the MA(1) model into an AR(∞) model.

11 ARMA models

Seeing the similarities in the lag operator formulation of the autoregressive and moving average models, we wish to unify them in the *autoregressive moving average* or simply ARMA model.

12 ARIMA models

We now wish to combine the AR, I, and MA models into one. The result is the *autoregressive integrated moving average model* or *ARIMA* for short.

13 Markov chains

A *Markov chain* is a time series, in which the conditional distribution of X_{n+1} given the realizations of $X_0, X_1, \dots, X_n$ only depends on the realization of X_n. Formally:

$$P(X_{n+1} = s | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = P(X_{n+1} = s | X_n = x_n) \tag{13.1}$$

13.1 Markov chains with a finite number of states

In the case where each X only has a finite number of realizations n, the Markov chain can be conveniently specified in matrix form. Assume the realization of X_n is state i, then we might ask what to probability of X_{n+1}

being realized as state j. This probability is called p_{ij}. These probabilities can be neatly organized in matrix form:

$$A = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{n1} \\ p_{12} & p_{22} & \cdots & p_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1n} & p_{2n} & \cdots & p_{nn} \end{pmatrix} \tag{13.2}$$

Since the n states exhaust the possibilities, each column must sum to 1:

$$\sum_{j=1}^{n} p_{ij} = 1, \quad i \in 1, 2, \ldots n \tag{13.3}$$

14 Martingales