STAT 5170: Applied Time Series

Course notes for part A of learning unit 2

Section 2A.1: Framework and examples.

Definition: A *time series* is a set of measurements collected in sequence,

$$x_1, x_2, \ldots x_n$$

In this notation, the value x_1 is measured first, x_2 second, *etc.* In general, denote by x_t the measurement at time t. We will study a time series as a sequence of *random variables.*

From previous classes you are likely to be familiar the following properties that would be exhibited in a series of random variables:

- If the x_t exhibit no dependencies and share a common marginal distribution, then $x_1, x_2, \ldots x_n$ is an *independent and identically distributed* (i.i.d.) random sample. Random-variable series of this type exhibit a property that is sometimes called *invariance to permutation*, which means that the order in which the measurements were taken does not matter. Independent and identically distributed random samples are the focus of most introductory-level statistics classes.
- The x_i in a series of random variables may be independent, but not identically distributed because the mean $\mu_t = E[x_t]$, or some other characteristic of their marginal distributions, exhibit *trends* over time or with respect to a set of *covariate measurements.* In a random-variable series of this type, the order in which the measurements were taken matters because it determines the trend of the series. Independent but non-identically distributed random samples that exhibit trends in the mean are the focus of introductory-level statistics classes on regression analysis.

Many of the time series we will discussed in the course exhibit the following properties:

- The time-points at which measurements are equally spaced, $t = 1, 2, \ldots$.
- The measurements, x_t, are themselves statistically *dependent*, but dependency is weaker between measurements that are taken at time points further apart. So, *e.g.*,
  - x_t correlated with x_{t+1};
  - x_t correlated with x_{t+2}, but not as strongly as the correlation between x_t and x_{t+1};
  - x_t correlated even less strongly with x_{t+3}.

  Correlation of measurements within the same random-variable—*i.e.,* correlation of a series with itself—is called *autocorrelation.*

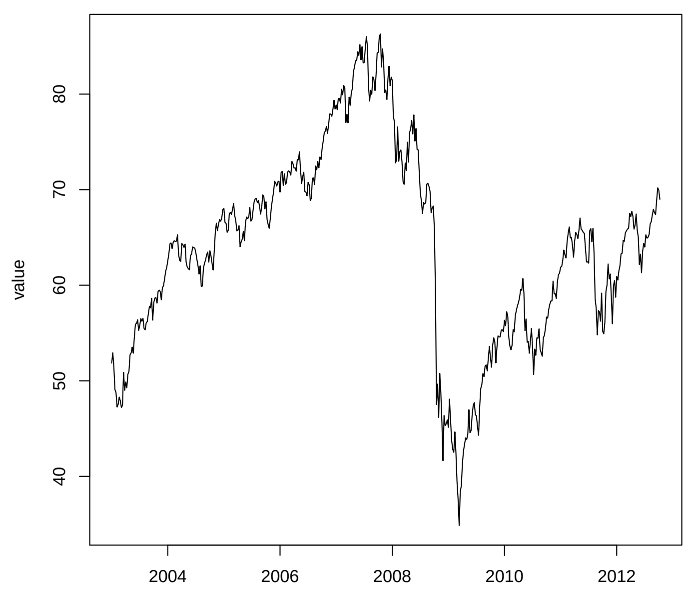
- Dependency across the measurements of a time series implies that we are not in either of the two situations, above, of i.i.d. random samples or regression models for independent response values. In time series models, the order in which measurements are taken matters, in part because order determines the dependency structure of among the measured values. On the other hand, many of the time series we will study are such that the marginal distributions of the x_t are identical. This is a type of *stationarity* property, about which we will learn much more later. A stationarity time series might be said to be identically distributed, but not necessarily independent.

- Some of the time series we will discuss exhibit trends over time. Such trends take on a variety forms, which include linearly increasing or decreasing trends, nonlinear increasing or decreasing trends, other nonlinear trends, seasonally periodic trends, or other periodic trends. Familiar regression techniques will sometimes help to understand and work with such series, after modification to accommodate dependencies across measurements.

Some examples of time series and particular interesting characteristics that we will closely examine later are as follows.

Example: S&P 500 Index

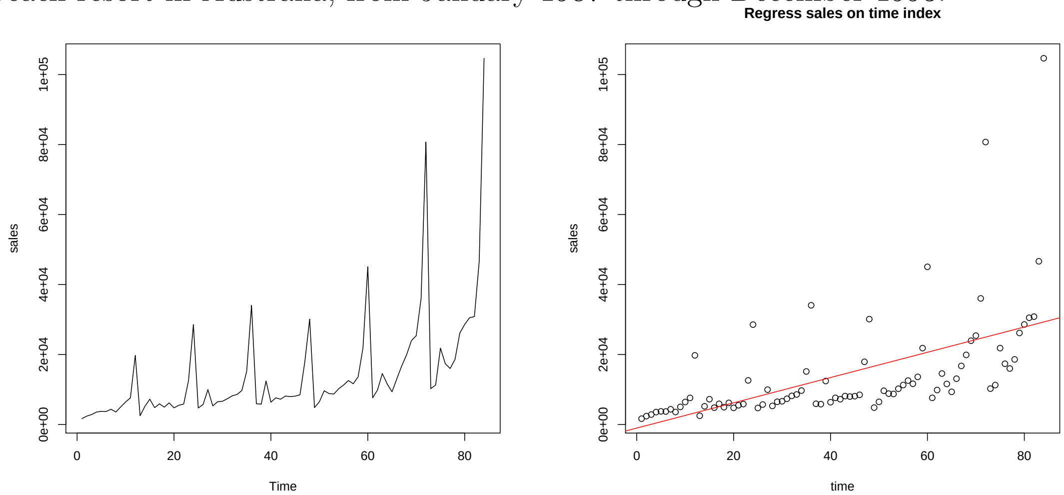
The following is an example of a time series that exhibits periods of increasing trends over time, with a substantial interruption.



Observe the pattern of dependencies between nearby measurements, which is exhibited as sweeping arcs or broad, meandering paths within sections of the series. □

Example: Monthly beach-shop souvenir sales in Australia

The right panel of the figure, below, displays the series of monthly sales of a souvenir shop at a beach resort in Australia, from January 1987 through December 1993.

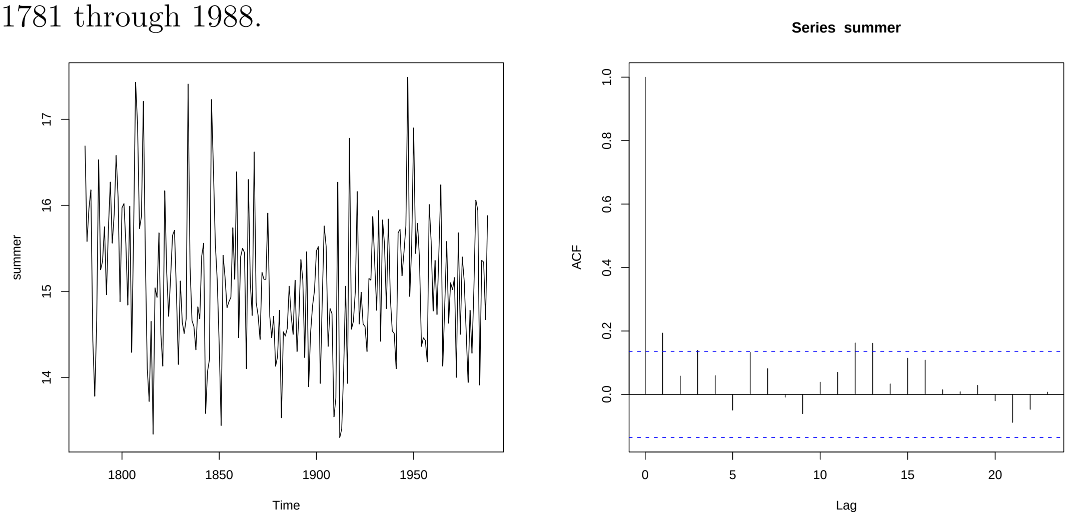


The left panel is a plot of fitted values that results from a linear regression analysis of this time series, using time as a the covariate. Observe the increasing trend in the series, and strongly correlated residual values.

The regularly occurring peaks in these series suggest a *seasonal* periodic pattern. □

Example: Average summer temperatures in Munich

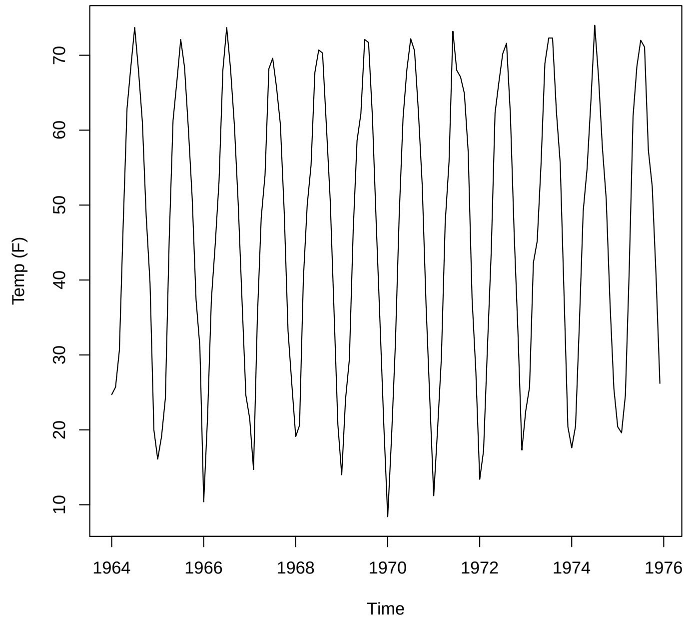
The right panel of the following figure displays the average summer temperatures in Munich, from 1781 through 1988.



The left panel displays a type of statistical diagnostic called a sample autocorrelation function. This statistic summarizes the direction and strength of sample correlations between measurements x_t and x_{t+k} across a range of *lag* values, $k = 0, 1, 2, \ldots$. Observe, for example, at $k = 1$ the sample correlation between values x_t and x_{t+1} is right around 0.2. □

Example: Average monthly temperatures in Dubuque, IA

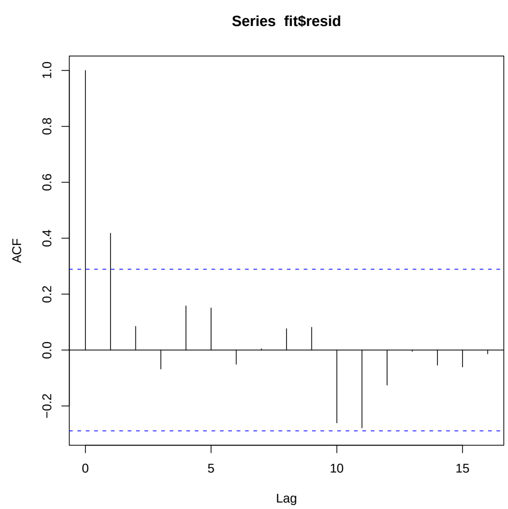
The following is an example of a time series that is *highly* periodic. These data are of average monthly temperatures in Dubuque, IA from January 1964 through December 1975.



☐

Example: Marriages in the Church of England

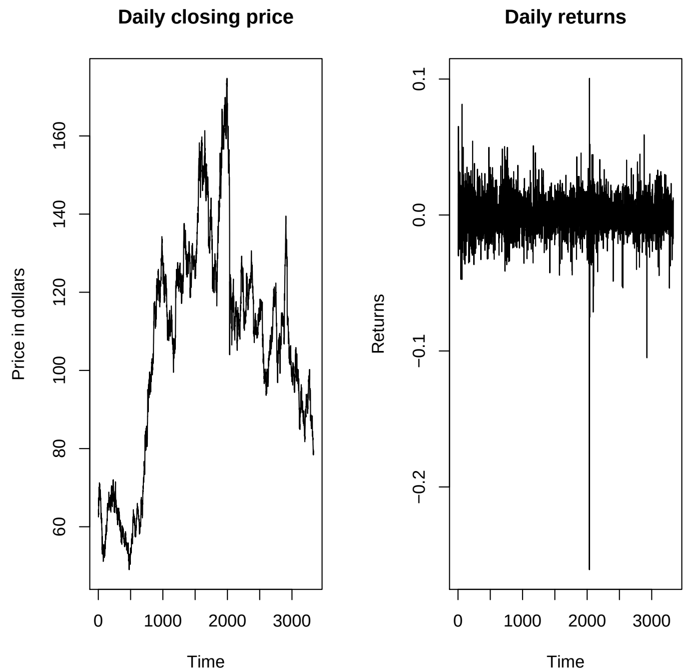
The right panel, below, displays a time series made famous by the statistician Udny Yule. They record the marriages within the Church of England as a percent of all marriages from 1866 through 1911. Observe the *negative* trend in these data.



The left panel shows the sample autocorrelation function calculated on the fitted residuals from a regression analysis of the marriage data against time. ☐

Example: IBM stock

The values x_t displayed in the left panel of the following figure are the daily closing IBM stock prices from January 1980 through October 1992.



The values displayed in the right panel are the stock returns, $r_t = (x_t - x_{t-1})/x_{t-1}$. The latter series exhibits a stationary pattern, overall; a strong feature is a huge shock to the returns series. □

As the study of time series has developed, two distinct ways of approaching them has emerged:

- The *time domain* approach to time series analysis focuses on understanding dependencies between lagged measurements. In this approach, autocorrelation is a key organizing concept. Most of the discussion in this course follows the time domain approach.

- The *frequency domain* approach to time series analysis focuses on understanding periodicity at small and large scales. The spectral density, a concept we will define later, is central to this approach.

Time series is an especially interesting topic because it involves sophisticated ideas from *probability*, which are needed to develop useful time-series models, and equally compelling approaches to *statistical inference*, which are needed for analysis. In this course, we focus on *Bayesian methods* of statistical inference, which is often implemented by numerical calculation that involves simulating random-variable distributions. We will also discuss a handful classical, non-Bayesian inference techniques, such as the use of sample autocorrelation functions. This will provide tools for addressing the variety of objectives of time-series analysis, which can include description, interpretation, and prediction (*i.e.*, forecasting).

Section 2A.2: Elementary concepts.

Let us now begin our dive into the material.

Definition: A *time series model* is a mathematical expression that describes relationships between random variables that define a time series.

A time series model may be used for a variety of purposes, among which is to understand patterns of correlation within time series data. A strategy for doing this is to start by studying the mathematical characteristics of a time series model, in the process identifying important *parameters* of the model; then look for patterns that are implied by those characteristics in the data to gain an understanding of the phenomenon that generated the data. A precise way to frame this is in terms on an inference problem, wherein the focus is on the model's parameters, which are treated a quantities in the model formulation whose values are left unspecified. It is *uncertain* what values of those parameters led to the data that were generated and measured, but our goal is to *reduce our uncertainty* about those values by analyzing the data. This is one part of *time series analysis*, which can give rise to inferences in the mode of *estimation* and *hypothesis testing*. Subsequent inferential analysis might proceed in the form of *forecasting* (a.k.a., prediction) or *model checking*. We will discuss all of these modes of inference later.

In this portion of the unit, we discuss several of the most basic time series models. Before starting that discussion it is helpful to define the most elementary notation:

Notation: Often we just write (x_t) to refer to a time series. In some instances is helpful to be explicit about the span of time the time series model is intended to cover. In particular, (x_t) may be defined for, say,

a finite sequence	$x_1, \ldots, x_n$
an infinite sequence	$x_0, x_1, \ldots$
a doubly-infinite sequence	$\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2 \ldots$

When left unspecified, in most cases it would make sense to assume the span of time is a doubly-infinite sequence. An important, advanced topic closely related to time series analysis is the study of continuous-time stochastic processes, where the span of time is a continuous range of values. We will not discuss such processes here. Nevertheless, do note that a time series model is sometimes referred to as a (discrete-time) stochastic process.

Our first basic time series model is as follows:

Definition: The t.s. (w_t) is *white noise* if

(*i.*) $E[w_t] = 0$ and $Var[w_t] = \sigma_w^2$ for every t, and

(*ii.*) $Cov[w_s, w_t] = 0$ for every $s \neq t$.

If, in addition, the w_t are i.i.d. $N(0, \sigma_w^2)$, then the time series is *Gaussian white noise.*

White noise very much like the i.i.d. random-variable series you worked with in previous classes. Do notice, however, that the two main conditions in the definition of white noise refer only to the means and covariances of the series. As we will see, some approaches to statistical analysis attempt to describe uncertainty, or make predictions, without assuming anything more about the white noise processes involved the inference context. The additional condition that defines the Gaussian case is a distributional assumption, which offers a much more complete formulation of the model, and makes possible more sophisticated forms of inference.

The next time series model is more interesting, and will motivate a number of helpful concepts and techniques in time series analysis:

Definition: The time series (x_t) is a *random walk* with initial value x_0 if

$$x_t = \delta + x_{t-1} + w_t \text{ for } t = 1, 2, \ldots$$

where δ is a constant *drift* parameter and (w_t) is white noise. The case of "no drift" is when $\delta = 0$. If, in addition, (w_t) is Gaussian white noise, then (x_t) is said to be a *Gaussian random walk.*

A couple of very basic properties of random-walk time series are as follows:

Property: Suppose (x_t) is a random walk.

1. The first property gives a representation of a random walk entirely in terms of the white-noise time series that defines it. One may write

$$x_t = \delta t + x_0 + \sum_{i=1}^{t} w_i$$

   It follows that if x_0 is constant (rather than itself a random variable), then $E[x_t] = \delta t + x_0$ and $Var[x_t] = \sigma_w^2 t$.

   Proof. A short proof of this property is by mathematical induction.

   (i.) $x_1 = \delta t + x_0 + w_1$
   (ii.) $x_t = \delta t + x_{t-1} + w_t = \delta t + \left\{\delta(t-1) + x_0 + \sum_{i=1}^{t-1} w_i\right\} + w_t = \delta t + x_0 + \sum_{i=1}^{t} w_i$

   □

2. The second property shows how to *decorrelate* the random-walk time series. Define the *difference* time series (y_t) according to

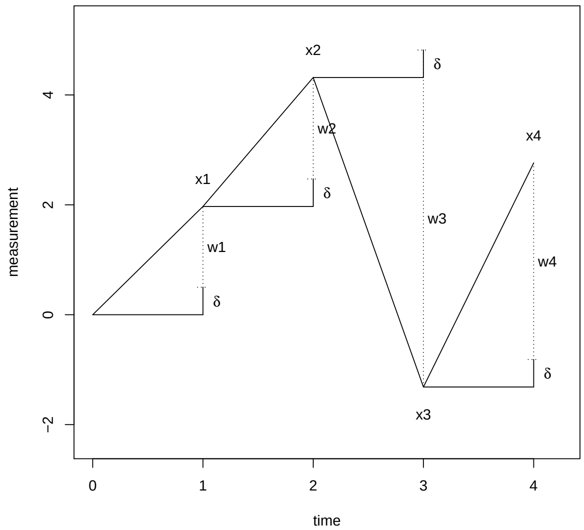
$$y_t = x_{t+1} - x_t$$

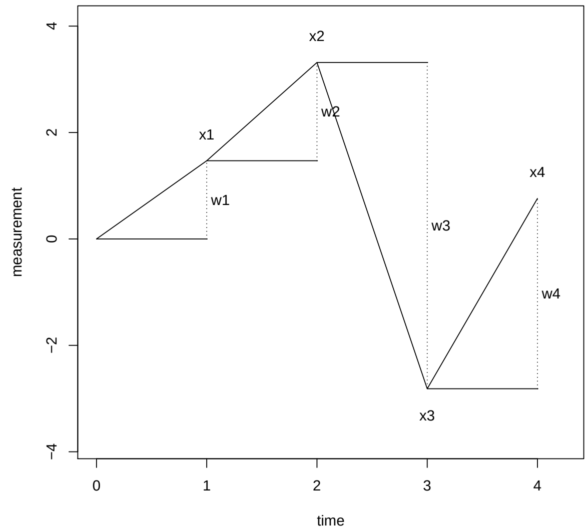
   It follows that (y_t) is uncorrelated.

   Proof. The one-line proof of this property is by direct substitution, which shows that (y_t) is white noise: $y_t = \delta + x_t + w_{t+1} - x_t = \delta + w_{t+1}$. □

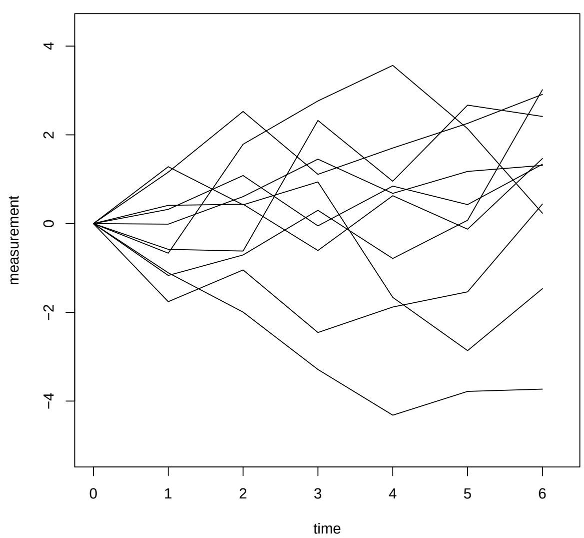
Example: Sample path of a random walk

A random walk evolves by moving from the last position to the current position by a random step. The graphics below illustrate this process for the case of a "positive drift" ($\delta > 0$, left panel) and "no drift" ($\delta = 0$, right panel):





In the latter illustration, one gets a sense that where there is no drift, the random walk remains centered around its initial value, which reflects the property $E[x_t] = x_0$. However, the property $Var[x_t] = t\sigma_w^2$ indicates that the time series also becomes more variable over time. This property is illustrated in the following plot of ten independently simulated random walks (without drift).



Observe that the terminal points of the series vary much more widely that points near the initial value.

The next basic time series model might be regarded as an extension of the random walk model:

Definition: The time series (x_t) is an $AR(p)$ *autoregressive* time series if

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t$$

where $\phi_1, \ldots, \phi_p$ are constant parameters and (w_t) is Gaussian white noise.

Notice here that the definition makes an explicit distributional assumption by requiring that (w_t) is *Gaussian* white noise. This assumption is not strictly necessary for some purposes, but as the concepts involved in our discussion of these models becomes more complicated, it will greatly simplify matters by always assuming that the definition of an autoregressive time series involves a distributional assumption.

Incidentally, can you see how a Gaussian random walk is a special case of an autoregressive time series?

As we will eventually see, the following basic time series model plays a complementary role to autoregressive time series, though it is perhaps conceptually simpler to grasp.

Definition: The time series (x_t) is an $MA(q)$ *moving-average* time series if

$$x_t = w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

where $\theta_1, \ldots, \theta_q$ are constant parameters and (w_t) is Gaussian white noise.

Now that we have laid out a few interesting time series models, we may begin discussing some of the basic conceptual tools that would be used to understand their properties.

Definition:

(*i.*) The *mean function* of a time series (x_t) is $\mu_t = E[x_t]$

(*ii.*) The *autocovariance function* of a time series (x_t) is

$$\gamma(s,t) = Cov[x_s, x_t] = E[(x_s - \mu_s)(x_t - \mu_t)]$$

(*iii.*) The *autocorrelation function*, or ACF, of a time series (x_t) is

$$\rho(s,t) = Corr[x_s, x_t] = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}.$$

Among these concepts, the mean function is simplest to work with. For example, it is quite easy to deduce that a moving-average time series model has mean zero: If (x_t) is $MA(q)$, then

$$\begin{aligned} \mu_t = E[x_t] &= E[w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}] \\ &= E[w_t] + E[\theta_1 w_{t-1}] + \cdots + E[\theta_q w_{t-q}] = 0 \end{aligned}$$

The following examples offer basic illustrations of the use of autocovariance and autocorrelation functions:

Example: Autocovariance function of a random walk

Suppose (x_t) is a random walk with fixed initial value x_0. If $0 < s < t$, the representations

$$x_s = \delta s + x_0 + \sum_{i=1}^{s} w_i \quad \text{and} \quad x_t = \delta t + x_0 + \sum_{i=1}^{t} w_i$$

allow us to write

$$\begin{aligned} x_t &= \delta t + x_0 + \sum_{i=1}^{t} w_i = \delta s + x_0 + \sum_{i=1}^{s} w_i + \delta(t-s) + \sum_{i=s+1}^{t} w_i \\ &= \delta(t-s) + x_s + \sum_{i=s+1}^{t} w_i. \end{aligned}$$

Subsequently, using the formula $Var[x_s] = s\sigma_w^2$ and the property that x_s is uncorrelated with any w_i such that $i > s$, the autocovariance function is

$$\begin{aligned} \gamma(s,t) &= Cov\left[x_s, \delta(t-s) + x_s + \sum_{i=s+1}^{t} w_i\right] \\ &= Cov[x_s, x_s] + Cov\left[x_s, \sum_{i=s+1}^{t} w_i\right] = Var[x_s] = s\sigma_w^2. \end{aligned}$$

A similar deduction will show that if $0 < t < s$ then $\gamma(s,t) = t\sigma_w^2$. A succinct representation that covers all cases is therefore

$$\gamma(s,t) = \min\{s,t\}\sigma_w^2.$$

<u>Example:</u> Mean and autocovariance function of a $MA(2)$ time series

Suppose (x_t) is an MA(2) moving average time series, defined by the relationship

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2},$$

where θ_1 and θ_2 are constant parameters, and (w_t) is Gaussian white noise.

Because each measurement, x_t is linear combination of white noise, the mean function is quickly deduced as

$$\mu_t = E[x_t] = E[w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}] = E[w_t] + \theta_1 E[w_{t-1}] + \theta_2 E[w_{t-2}] = 0.$$

The autocovariance function is deduced by considering individual distances between s and t:

If $s = t$, then

$$\begin{aligned} \gamma(s,t) &= Var\left[w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}\right] \\ &= Var[w_t] + \theta_1^2 Var[w_{t-1}] + \theta_2^2 Var[w_{t-2}] \\ &= (1 + \theta_1^2 + \theta_2^2)\sigma_w^2. \end{aligned}$$

If $s = t-1$, then

$$\begin{aligned} \gamma(s,t) &= Cov\left[w_s + \theta_1 w_{s-1} + \theta_2 w_{s-2}, w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}\right] \\ &= Cov\left[w_{t-1} + \theta_1 w_{t-2} + \theta_2 w_{t-3}, w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}\right] \\ &= \theta_1 Var[w_{t-1}] + \theta_1\theta_2 Var[w_{t-2}] \\ &= \theta_1(1 + \theta_2)\sigma_w^2. \end{aligned}$$

If $s = t - 2$, then

$$\begin{aligned}
\gamma(s,t) &= Cov\left[w_s + \theta_1 w_{s-1} + \theta_2 w_{s-2}, w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}\right] \\
&= Cov\left[w_{t-2} + \theta_1 w_{t-3} + \theta_2 w_{t-4}, w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}\right] \\
&= \theta_2 Var[w_{t-2}] \\
&= \theta_2 \sigma_w^2.
\end{aligned}$$

If $s < t - 2$, then $\gamma(s,t) = 0$.

A parallel pattern emerges in cases where $s > t$. A summary of the autocovariance function for all cases is

$$\gamma(s,t) = \begin{cases} (1+\theta_1^2+\theta_2^2)\sigma_w^2 & \text{if } s = t \\ \theta_1(1+\theta_2)\sigma_w^2 & \text{if } |s-t| = 1 \\ \theta_2\sigma_w^2 & \text{if } |s-t| = 2 \\ 0 & \text{otherwise} \end{cases}$$

The following defines a very broad class of time series models, which, as we will eventually see, intersects with the autoregressive and moving-average models.

Definition: The time series (x_t) is said to be a *linear process* if it has the representation

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j} \text{ where } \sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

and (w_t) is mean-zero white noise. It is a *Gaussian linear process* if (w_t) is Gaussian white noise. Notice that index covers the entire range of all integers, $j = 0, \pm 1, \pm 2, \ldots$.

The mean function of a linear process is

$$\mu_t = E[x_t] = \mu + \sum_{j=-\infty}^{\infty} \psi_j E[w_{t-j}] = \mu,$$

and the corresponding autocovariance function is

$$\begin{aligned}
\gamma(s,t) &= Cov[x_s, x_t] = Cov\left[\mu + \sum_{j=-\infty}^{\infty} \psi_j w_{s-j}, \mu + \sum_{k=-\infty}^{\infty} \psi_k w_{t-k}\right] \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k Cov[w_{s-j}, w_{t-k}] \\
&= \sigma_w^2 \sum_{k=-\infty}^{\infty} \psi_{k+s-t} \psi_k,
\end{aligned}$$

where $\sigma_w^2 = Var[w_t]$.

An important specially case of a linear process is a *one-sided* linear process, which defined when the index variable starts at zero. That is,

$$x_t = \mu + \sum_{j=0}^{\infty} \psi_j w_{t-j},$$

This is sometimes called the $MA(\infty)$ representation of a time series. Its autocovariance function is

$$\gamma(t+h,t) \quad = \quad \sigma_w^2 \sum_{k=0}^{\infty} \psi_{k+h}\psi_k,$$

which is quickly derived from the autocovariance formula for a linear process by equating $h = s - t$. $\square$