variance σ^2, and you randomly select a sample of data $\{y_1, y_2, \ldots, y_n\}$, then an unbiased (point) estimator of μ is

$$\overline{y} = \frac{y_1 + \cdots + y_n}{n},$$

and a common unbiased (point) estimator of σ^2 is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

In Stat 252, you studied *many* of the properties of these (point) estimators.

Exercise 2.1.1. Being sure to carefully list all of the necessary assumptions, prove that

- $\overline{y}$ is an unbiased estimator of μ, and
- s^2 is an unbiased estimator of σ^2.

This is one reason for dividing by $n - 1$ in the definition of the sample variance s^2, instead of dividing by (the more natural) n.

You will also recall that using only a single number (such as $\overline{y}$ or s^2) to estimate a parameter is not that beneficial because it is unlikely that point estimator will equal the parameter exactly. Instead, you learned that much more information is provided by a *confidence interval.*

Most confidence intervals that you have encountered are normal-distribution-based. This means that if $\hat{\theta}$ is a point estimator of a parameter θ, then

$$\hat{\theta} \pm z_\alpha \sqrt{\hat{V}(\hat{\theta})},$$

where $\hat{V}(\theta)$ is the (estimated) variance of $\hat{\theta}$, is a $(1-\alpha)\%$ confidence interval for θ.

You will also recall that the interpretation of a confidence interval is that *before* the sample is drawn,

$$P\left(|\theta - \hat{\theta}| \geq z_\alpha \sqrt{\hat{V}(\hat{\theta})}\right) = \alpha.$$

After the sample is drawn, no such probability statement is true. All you can do is simply *hope* that your confidence interval is not one of the unlucky $\alpha\%$.

For the analysis of time series, the approach will be similar. However, as we have indicated previously, we no longer have a single random variable to worry about, but an entire stochastic process.

2.2 Objectives of Time Series Analysis

As with most of Statistics, the objective of time series analysis is to draw inference from a time series. This is done by providing a concise description of the historic series. It may include some summary statistics, but more likely, it will require a *function* rather than a single number to summarize the time series' essential features.

Another, more challenging, objective is to *forecast* future values of the series. (It should be noted that many of the methods of analysis of time series were developed with this specific goal in mind.)

A problem that occurs in the biological and medical sciences is *monitoring* a time series in order to detect changes in behaviour as they occur. You might immediately think of an electrocardiogram (ECG) trace. It is important for a doctor observing a patient's ECG trace to recognize an abnormality when it occurs.

Finally, there is a rôle to be played in time series analysis for *accommodation* of the serial dependence when making inferences about the basic stochastic structural parameters.

Definition 2.2.1. Consider an observed time series $\{x_t\}$. A *time series model* for the observed data $\{x_t\}$ is a specification of the stochastic process $\{X_t\}$ of which $\{x_t\}$ is conjectured to be a realization. By the specification of the stochastic process, we mean the specification of its joint distribution, or possibly only its means and covariances.

Remark. It is important to realize that in many practical problems, the time series that we observe is the *only* realization of that series which will *ever* occur. Our analysis is only furthered, however, if we imagine it to be one of the many processes that *might* have occurred.

2.3 The First Step in Time Series Modelling

As you were probably told in every class you have ever had in which you encountered data, you should ***ALWAYS*** begin your analysis by producing a graph (or several) of the data. For time series data, the most obvious graph is a time series plot. (See Section 1.5.)

Specifically for time series data, you should examine the plot to see whether there appears to be

- a trend (whether increasing or decreasing),
- a seasonal component (as evidenced by a cyclic pattern),
- any apparent sharp changes in behaviour,
- any outlying behaviour.

It should be noted that if you have discrete data, there are arguments for and against joining successive points with straight lines.

- By joining succesive points, it might be easier to digest, especially if multiple time series are superimposed on one graph.
- The zig-zag appearance of the trace made by joining points gives the false impression of continuous observation. Doing so may give the wrong impression, especially for time series with missing data.

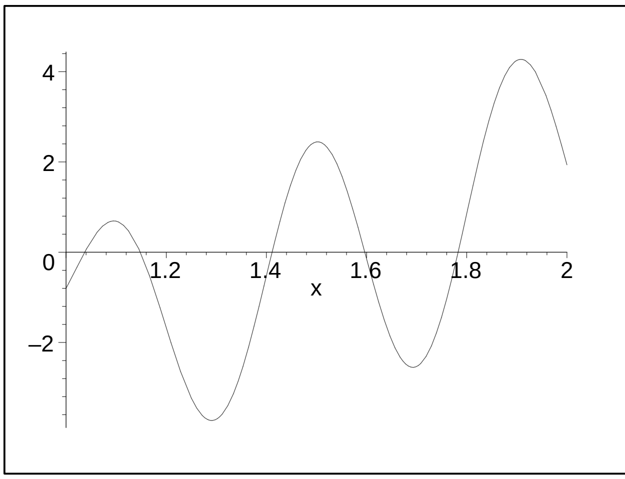
One final consideration is that you should carefully choose the *aspect ratio* (or *shape parameter*) for the actual display of the time series.

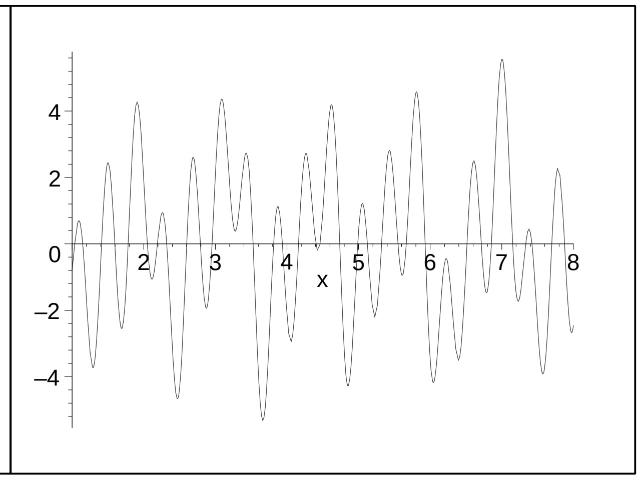
Remark. For those who are familiar with the difference between the *fullscreen* and *widescreen* versions of VHS/DVD movies, then you will already understand the concept of *aspect ratio* (fullscreen is 4:3 and widescreen is 16:9).

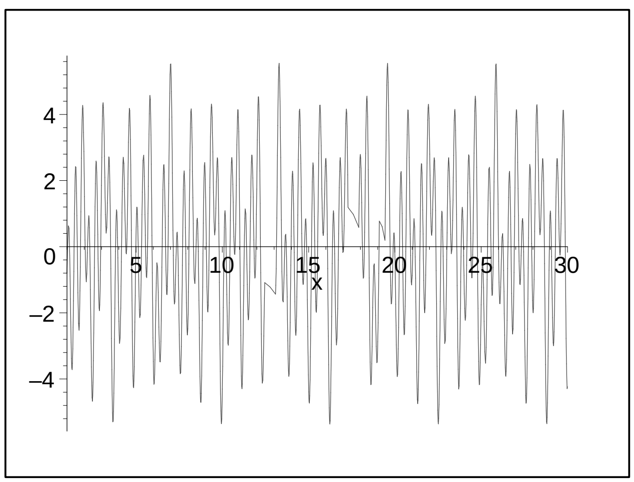
Example 2.3.1. The following three plots are *all* graphs of the function

$$f(t) = \sin(10t) - 2\cos(5t) + 0.5\sin(3t-1) + 2.5\cos(16t+1.1) - 0.001\sin(8t) + 0.3t^{-3/2}.$$

In the pictures below, the viewing window is *fixed.* However, the apparent loss of smoothness occurs because the domain is restricted in the first picture to $[1,2]$, in the second to $[1,8]$, and in the third to $[1,30]$.







2.4 Sample Functions

Again, to stress the matter, a *statistic* is a number computed from data which is most probably used as an *estimate* of a population *parameter*. That is, the parameter is unknown, and the investigator desires knowledge of the parameter. A sample of data is drawn (e.g., in Stat 257 a survey is conducted, in Stat 471 a time series is observed) from which statistics are computed and used to make inferences about the parameter in question.

For the analysis of time series, the following list of *sample functions* will be important. Note that we are no longer only talking about single numbers (such as the sample mean and sample variance), but entire *functions* (such as the sample ACF).

In all of the definitions which follow, suppose that $\{x_1, x_2, \ldots, x_n\}$ is an observed time series.

Sample Mean

$$\overline{x} := \frac{1}{n}\sum_{t=1}^{n} x_t$$

Sample Autocovariance Function (ACVF)

$$\hat{\gamma}(h) := \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \overline{x})(x_t - \overline{x}), \quad -n < h < n$$

Notice that $\hat{\gamma}(-h) = \hat{\gamma}(h)$.

Sample Covariance Matrix

The sample covariance matrix is simply the matrix of sample autocovariance functions given by

$$\hat{\Gamma}_n := [\hat{\gamma}(i-j)]_{1\le i,j\le n}.$$

Notice that $\hat{\Gamma}_n$ is nonnegative definite.

Sample Autocorrelation Function (ACF)

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -n < h < n$$

Sample Correlation Matrix

The sample correlation matrix is simply the matrix of sample autocorrelation functions given by

$$\hat{R}_n := [\hat{\rho}(i-j)]_{1\le i,j\le n}$$

Notice that $\hat{R}_n$ is nonnegative definite, and that each diagonal entry of $\hat{R}_n$ is 1 since $\hat{\rho}(0) = 1$.

Example 2.4.1. It is important to note that all these sample functions can be computed for *ANY* data set. Here is an elementary example. Suppose that the observed data set is $\{0, 4, 8, 4, 0, -4, 0, -4\}$. Viewing this as a "time series" means that $\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\} = \{0, 4, 8, 4, 0, -4, 0, -4\}$. The sample mean is therefore

$$\overline{x} = \frac{1}{8}\sum_{t=1}^{8} x_t = \frac{0+4+8+4+0-4+0-4}{8} = \frac{8}{8} = 1,$$

and the sample autocovariance function is

$$\hat{\gamma}(h) := \frac{1}{8}\sum_{t=1}^{8-|h|}(x_{t+|h|} - 1)(x_t - 1), \quad -8 < h < 8.$$

Thus, we can easily compute that

$$\begin{aligned}
\hat{\gamma}(0) &= \frac{1}{8}\sum_{t=1}^{8}(x_t - 1)(x_t - 1)\\
&= \frac{1}{8}\left[(x_1-1)^2 + (x_2-1)^2 + (x_3-1)^2 + (x_4-1)^2 + (x_5-1)^2 + (x_6-1)^2 + (x_7-1)^2 + (x_8-1)^2\right]\\
&= \frac{1}{8}\left[(0-1)^2 + (4-1)^2 + (8-1)^2 + (4-1)^2 + (0-1)^2 + (-4-1)^2 + (0-1)^2 + (-4-1)^2\right]\\
&= \frac{120}{8}
\end{aligned}$$

$$
\begin{aligned}
\hat{\gamma}(1) = \hat{\gamma}(-1) &= \frac{1}{8}\sum_{t=1}^{7}(x_{t+1}-1)(x_t-1) \\
&= \frac{1}{8}\left[(x_2-1)(x_1-1)+(x_3-1)(x_2-1)+(x_4-1)(x_3-1)+(x_5-1)(x_4-1)+(x_6-1)(x_5-1)\right. \\
&\qquad \left. +\ (x_7-1)(x_6-1)+(x_8-1)(x_7-1)\right] \\
&= \frac{1}{8}\left[(4-1)(0-1)+(8-1)(4-1)+(4-1)(8-1)+(0-1)(4-1)+(-4-1)(0-1)\right. \\
&\qquad \left. +\ (0-1)(-4-1)+(-4-1)(0-1)\right] \\
&= \frac{51}{8}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\gamma}(2) = \hat{\gamma}(-2) &= \frac{1}{8}\sum_{t=1}^{6}(x_{t+2}-1)(x_t-1) \\
&= \frac{1}{8}\left[(x_3-1)(x_1-1)+(x_4-1)(x_2-1)+(x_5-1)(x_3-1)+(x_6-1)(x_4-1)+(x_7-1)(x_5-1)\right. \\
&\qquad \left. +\ (x_8-1)(x_6-1)\right] \\
&= \frac{1}{8}\left[(8-1)(0-1)+(4-1)(4-1)+(0-1)(8-1)+(-4-1)(4-1)+(0-1)(0-1)\right. \\
&\qquad \left. +\ (-4-1)(-4-1)\right] \\
&= \frac{5}{8}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\gamma}(3) = \hat{\gamma}(-3) &= \frac{1}{8}\sum_{t=1}^{5}(x_{t+3}-1)(x_t-1) \\
&= \frac{1}{8}\left[(x_4-1)(x_1-1)+(x_5-1)(x_2-1)+(x_6-1)(x_3-1)+(x_7-1)(x_4-1)+(x_8-1)(x_5-1)\right] \\
&= \frac{1}{8}\left[(4-1)(0-1)+(0-1)(4-1)+(-4-1)(8-1)+(0-1)(4-1)+(-4-1)(0-1)\right] \\
&= \frac{-39}{8}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\gamma}(4) = \hat{\gamma}(-4) &= \frac{1}{8}\sum_{t=1}^{4}(x_{t+4}-1)(x_t-1) \\
&= \frac{1}{8}\left[(x_5-1)(x_1-1)+(x_6-1)(x_2-1)+(x_7-1)(x_3-1)+(x_8-1)(x_4-1)\right] \\
&= \frac{1}{8}\left[(0-1)(0-1)+(-4-1)(4-1)+(0-1)(8-1)+(-4-1)(4-1)\right] \\
&= \frac{-36}{8}
\end{aligned}
$$

$$\begin{aligned}
\hat{\gamma}(5) = \hat{\gamma}(-5) &= \frac{1}{8}\sum_{t=1}^{3}(x_{t+5}-1)(x_t-1) \\
&= \frac{1}{8}\left[(x_6-1)(x_1-1)+(x_7-1)(x_2-1)+(x_8-1)(x_3-1)\right] \\
&= \frac{1}{8}\left[(-4-1)(0-1)+(0-1)(4-1)+(-4-1)(8-1)\right] \\
&= \frac{-33}{8}
\end{aligned}$$

$$\begin{aligned}
\hat{\gamma}(6) = \hat{\gamma}(-6) &= \frac{1}{8}\sum_{t=1}^{2}(x_{t+6}-1)(x_t-1) \\
&= \frac{1}{8}\left[(x_7-1)(x_1-1)+(x_8-1)(x_2-1)\right] \\
&= \frac{1}{8}\left[(0-1)(0-1)+(-4-1)(4-1)\right] \\
&= \frac{-14}{8}
\end{aligned}$$

$$\begin{aligned}
\hat{\gamma}(7) = \hat{\gamma}(-7) &= \frac{1}{8}\sum_{t=1}^{1}(x_{t+7}-1)(x_t-1) \\
&= \frac{1}{8}\left[(x_8-1)(x_1-1)\right] \\
&= \frac{1}{8}\left[(-4-1)(0-1)\right] \\
&= \frac{5}{8}
\end{aligned}$$

The sample autocorrelation function is

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\hat{\gamma}(h)}{120/8}, \quad -8 < h < 8,$$

so that

$$\hat{\rho}(0)=1, \quad \hat{\rho}(1)=\hat{\rho}(-1)=51/120, \quad \hat{\rho}(2)=\hat{\rho}(-2)=5/120, \quad \hat{\rho}(3)=\hat{\rho}(-3)=-39/120,$$

$$\hat{\rho}(4)=\hat{\rho}(-4)=-36/120, \quad \hat{\rho}(5)=\hat{\rho}(-5)=-33/120, \quad \hat{\rho}(6)=\hat{\rho}(-6)=-14/120,$$

$$\hat{\rho}(7)=\hat{\rho}(-7)=5/120.$$

As for the sample covariance and correlation matrices, we have

$$\hat{\Gamma}_8 = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \hat{\gamma}(2) & \hat{\gamma}(3) & \hat{\gamma}(4) & \hat{\gamma}(5) & \hat{\gamma}(6) & \hat{\gamma}(7) \\ \hat{\gamma}(-1) & \hat{\gamma}(0) & \hat{\gamma}(1) & \hat{\gamma}(2) & \hat{\gamma}(3) & \hat{\gamma}(4) & \hat{\gamma}(5) & \hat{\gamma}(6) \\ \hat{\gamma}(-2) & \hat{\gamma}(-1) & \hat{\gamma}(0) & \hat{\gamma}(1) & \hat{\gamma}(2) & \hat{\gamma}(3) & \hat{\gamma}(4) & \hat{\gamma}(5) \\ \hat{\gamma}(-3) & \hat{\gamma}(-2) & \hat{\gamma}(-1) & \hat{\gamma}(0) & \hat{\gamma}(1) & \hat{\gamma}(2) & \hat{\gamma}(3) & \hat{\gamma}(4) \\ \hat{\gamma}(-4) & \hat{\gamma}(-3) & \hat{\gamma}(-2) & \hat{\gamma}(-1) & \hat{\gamma}(0) & \hat{\gamma}(1) & \hat{\gamma}(2) & \hat{\gamma}(3) \\ \hat{\gamma}(-5) & \hat{\gamma}(-4) & \hat{\gamma}(-3) & \hat{\gamma}(-2) & \hat{\gamma}(-1) & \hat{\gamma}(0) & \hat{\gamma}(1) & \hat{\gamma}(2) \\ \hat{\gamma}(-6) & \hat{\gamma}(-5) & \hat{\gamma}(-4) & \hat{\gamma}(-3) & \hat{\gamma}(-2) & \hat{\gamma}(-1) & \hat{\gamma}(0) & \hat{\gamma}(1) \\ \hat{\gamma}(-7) & \hat{\gamma}(-6) & \hat{\gamma}(-5) & \hat{\gamma}(-4) & \hat{\gamma}(-3) & \hat{\gamma}(-2) & \hat{\gamma}(-1) & \hat{\gamma}(0) \end{pmatrix}$$

$$= \begin{pmatrix} 120/8 & 51/8 & 5/8 & -39/8 & -36/8 & -33/8 & -14/8 & 5/8 \\ 51/8 & 120/8 & 51/8 & 5/8 & -39/8 & -36/8 & -33/8 & -14/8 \\ 5/8 & 51/8 & 120/8 & 51/8 & 5/8 & -39/8 & -36/8 & -33/8 \\ -39/8 & 5/8 & 51/8 & 120/8 & 51/8 & 5/8 & -39/8 & -36/8 \\ -36/8 & -39/8 & 5/8 & 51/8 & 120/8 & 51/8 & 5/8 & -39/8 \\ -33/8 & -36/8 & -39/8 & 5/8 & 51/8 & 120/8 & 51/8 & 5/8 \\ -14/8 & -33/8 & -36/8 & -39/8 & 5/8 & 51/8 & 120/8 & 51/8 \\ 5/8 & -14/8 & -33/8 & -36/8 & -39/8 & 5/8 & 51/8 & 120/8 \end{pmatrix},$$

and

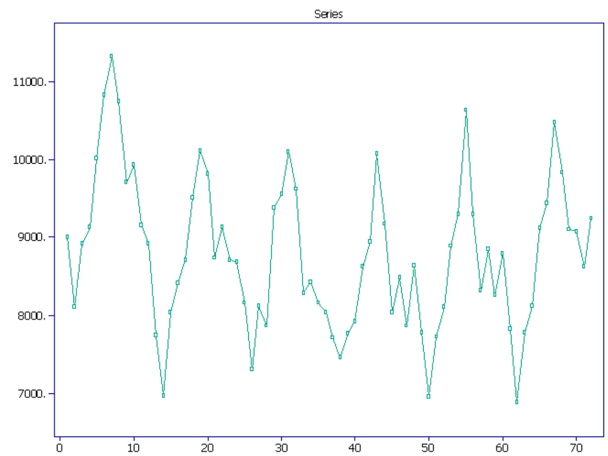
$$\hat{R}_8 = \begin{pmatrix} \hat{\rho}(0) & \hat{\rho}(1) & \hat{\rho}(2) & \hat{\rho}(3) & \hat{\rho}(4) & \hat{\rho}(5) & \hat{\rho}(6) & \hat{\rho}(7) \\ \hat{\rho}(-1) & \hat{\rho}(0) & \hat{\rho}(1) & \hat{\rho}(2) & \hat{\rho}(3) & \hat{\rho}(4) & \hat{\rho}(5) & \hat{\rho}(6) \\ \hat{\rho}(-2) & \hat{\rho}(-1) & \hat{\rho}(0) & \hat{\rho}(1) & \hat{\rho}(2) & \hat{\rho}(3) & \hat{\rho}(4) & \hat{\rho}(5) \\ \hat{\rho}(-3) & \hat{\rho}(-2) & \hat{\rho}(-1) & \hat{\rho}(0) & \hat{\rho}(1) & \hat{\rho}(2) & \hat{\rho}(3) & \hat{\rho}(4) \\ \hat{\rho}(-4) & \hat{\rho}(-3) & \hat{\rho}(-2) & \hat{\rho}(-1) & \hat{\rho}(0) & \hat{\rho}(1) & \hat{\rho}(2) & \hat{\rho}(3) \\ \hat{\rho}(-5) & \hat{\rho}(-4) & \hat{\rho}(-3) & \hat{\rho}(-2) & \hat{\rho}(-1) & \hat{\rho}(0) & \hat{\rho}(1) & \hat{\rho}(2) \\ \hat{\rho}(-6) & \hat{\rho}(-5) & \hat{\rho}(-4) & \hat{\rho}(-3) & \hat{\rho}(-2) & \hat{\rho}(-1) & \hat{\rho}(0) & \hat{\rho}(1) \\ \hat{\rho}(-7) & \hat{\rho}(-6) & \hat{\rho}(-5) & \hat{\rho}(-4) & \hat{\rho}(-3) & \hat{\rho}(-2) & \hat{\rho}(-1) & \hat{\rho}(0) \end{pmatrix}$$

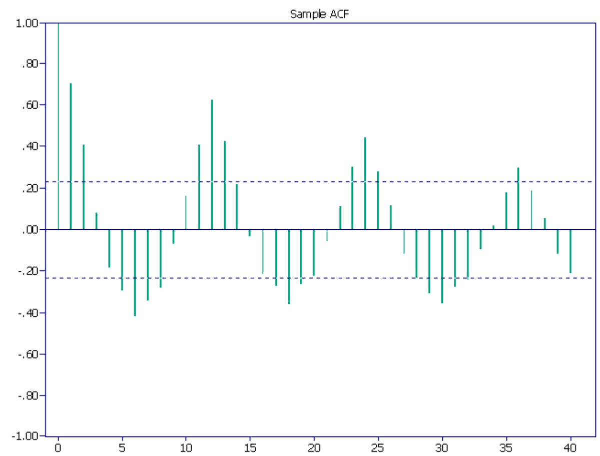
$$= \begin{pmatrix} 1 & 51/120 & 5/120 & -39/120 & -36/120 & -33/120 & -14/120 & 5/120 \\ 51/120 & 1 & 51/120 & 5/120 & -39/120 & -36/120 & -33/120 & -14/120 \\ 5/120 & 51/120 & 1 & 51/120 & 5/120 & -39/120 & -36/120 & -33/120 \\ -39/120 & 5/120 & 51/120 & 1 & 51/120 & 5/120 & -39/120 & -36/120 \\ -36/120 & -39/120 & 5/120 & 51/120 & 1 & 51/120 & 5/120 & -39/120 \\ -33/120 & -36/120 & -39/120 & 5/120 & 51/120 & 1 & 51/120 & 5/120 \\ -14/120 & -33/120 & -36/120 & -39/120 & 5/120 & 51/120 & 1 & 51/120 \\ 5/120 & -14/120 & -33/120 & -36/120 & -39/120 & 5/120 & 51/120 & 1 \end{pmatrix}.$$

Exercise 2.4.2. It is now your turn! Consider the data set (or observed time series, if you prefer) $\{x_1, x_2, x_3, x_4, x_5, x_6\} = \{0, 1, 0, -1, -2, -3\}$. Compute (by hand) the sample mean, sample ACVF, sample ACF, sample covariance matrix, and sample correlation matrix.

As you might have gathered, it is often tedious (and painful!) to compute the sample ACF by hand. In fact, in practice a graphical representation is usually much more informative.

Exercise 2.4.3. Do Problem 1.17 on page 43 of [2]. Here is a partial solution, namely the time series plot of the DEATHS.TSM data along with the sample ACF.





3 The *Probability* of Time Series Analysis

As you are no doubt aware from your previous statistics courses, the language of Statistics is probability. That is to say, although it is **TRIVIAL** and **BORING** to compute the summary statistics of a collection of numbers (as in Example 2.4.1 and Exercise 2.4.2), it is **FASCINATING** to know that a random sample can be modelled as a collection of iid random variables from which quite **INTERESTING** ideas such as confidence intervals develop.

3.1 Summarizing Random Variables

Recall that for a random variable X we can compute its moments $\mathbb{E}(X^k)$. (See Section 1.3.) One important assumption that is often made is that X has a finite second moment, as in, for example, the statements of the Central Limit Theorem, or the Strong Law of Large Numbers.

Definition 3.1.1. If X is a random variable, then the *mean* of X is the number $\mu := \mathbb{E}(X)$. Note that $-\infty \leq \mu \leq \infty$. If $-\infty < \mu < \infty$, then we say that X has a *finite mean*, or that X is an *integrable random variable*, and we write $X \in L^1$.

Example 3.1.2. Suppose that X is a Cauchy-distributed random variable. That is, X is a continuous random variable with density function

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{x^2+1}.$$

Carefully show that $X \notin L^1$.

Definition 3.1.3. If X is a random variable with $\mathbb{E}(X^2) < \infty$, then we say that X has a *finite second moment* and write $X \in L^2$. If $X \in L^2$, then we define the *variance* of X to be the number $\sigma^2 := \mathbb{E}\left((X-\mu)^2\right)$. The *standard deviation* of X is the number $\sigma := \sqrt{\sigma^2}$. (As usual, this is the positive square root.)

Remark. It is an important fact that if $X \in L^2$, then it must be the case that $X \in L^1$. This follows from the so-called Cauchy-Schwarz Inequality. (See Exercises 3.1.18 and 3.1.19 below.)