Statistical process control for serially correlated data
Wieringa, Jakob

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1999

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 2

Statistical Process Control

In this chapter, we will discuss the basics of Statistical Process Control (SPC). Concepts that are used in subsequent chapters will be defined. In Section 2.1, the meaning of the word “quality” is explained. In Section 2.2, we will focus on the relation between quality and variation. Two important concepts, namely, “common causes of variation” and “special causes of variation”, will be discussed. Furthermore, we will argue that the philosophy behind SPC is not only applicable to manufacturing environments, it is useful in all areas of an organization. In Section 2.3, the Shewhart control chart for independent observations is introduced. In subsequent chapters, we want to study the effect of serial correlation on the behavior of the control chart. In Section 2.4, we review some of the time series models that are used throughout this thesis.

2.1 Quality defined

The word “quality” is used by a lot of people in different contexts. In most companies it is recognized that quality of a service or a product is very important. However, it is very hard to give a perfect definition that discriminates products or services of bad quality from products or services of high quality. Various authors have attempted to formulate such a definition.

Shewhart (1931) argues that there are two aspects of quality. Firstly, there is an objective concept of quality, resulting in quantitatively measurable physical characteristics, that are independent of a second, subjective, aspect of quality. The latter has to do with what we think, feel or sense. Shewhart recognizes that the subjective side of quality is commercially interesting,
but that it is necessary to establish standards of quality in a quantitative (objective) manner. Deming (1982) also stresses the subjective side of quality:

Quality can be defined only in terms of the agent (p. 168).

Furthermore, he emphasizes that impressions of quality are not static. They change over time. This creates problems in defining quality, since it is difficult to translate future needs of the user into measurable characteristics.

In a vivid case history starting on page 41, Crosby (1979) defines quality as “conformance to requirements”. In this view, a product is of good quality if it meets its specifications. This definition encloses an important part of what customers perceive as “quality”. A manufacturer will most certainly receive a lot of ‘quality complaints’ if a large part of his production exceeds tolerances that were agreed upon with the customers. On the other hand, the definition is too narrow. For example, the specifications themselves are part of what is perceived as “quality”. A product that is produced in conformance with specifications that are not popular with customers will not be ranked as a high quality product.

In Juran and Gryna (1988), a broader definition is given: “quality is fitness for use”. In this view, “quality” is defined as a relative notion. Different usages of the product will result in different requirements with regard to the product. Let us consider shoes as an example. The requirements of a person looking for high quality jogging shoes will differ from the requirements of the same person looking for high quality elegant shoes. Montgomery (1996) considers ‘conformance to requirements’ as one of two aspects of ‘fitness for use’. The other aspect is ‘quality of design’, which emphasizes the intentional design differences between types of a product. The conformance aspect is how well the product conforms to the specifications required by the design.

Nowadays, these definitions of “quality” are labeled as traditional. It is recognized that the quality of a product or a service is not a single, identifiable characteristic. Garvin (1987) distinguishes the following eight dimensions of quality, which, when taken together, incorporate more aspects than the ‘traditional’ definitions of quality.

1. **Performance** is one of the traditional measures of quality. It refers to the basic functioning of a product. For example, for an automobile, performance would include acceleration, handling, cruising speed, and comfort;
2. **Features** that are added to the basic functioning of a product attribute to a higher quality;

3. **Reliability** is another traditional measure of quality. A reliable product rarely fails. This aspect, which is sometimes reformulated as ‘being free of deficiencies’ is a very important dimension of quality;

4. **Conformance** is related to reliability. It refers to the degree to which a product meets pre-established requirements. This dimension of quality is very important in situations where products are used as the components in a more complex assembly. Specifications on the individual components are usually expressed as a target and a tolerance. If each of the components is just slightly too big or too small, a tight fit is unlikely, and the final product may not perform as intended by the designer, or may wear out early;

5. **Durability** is a measure of product life, either economically (expected cost of repair exceeds current product value) or physically (repair is impossible). A product that lasts longer is usually viewed as being of higher quality;

6. **Serviceability** relates to the time and effort that is needed to repair a product. The breaking down of a product is usually viewed as an annoyance, but a prompt repair may relieve part of the irritation;

7. **Aesthetics** is a subjective dimension of quality. It refers to the look, feel, sound, taste or smell of a product. It is greatly influenced by the preferences of the individual customer. On this dimension of quality it is usually not possible to meet the needs of every customer;

8. **Perceived quality** is also a very subjective dimension. When customers do not have full information about a product they may base their quality image on past experiences, the reputation of the manufacturer, the quality of other products from the same manufacturer, or the name of the product.

Realizing that quality is not a one-dimensional characteristic of a product and that quality is determined at various levels in the production process,
CHAPTER 2. STATISTICAL PROCESS CONTROL

Garvin stresses that manufacturers should not strive to be first on all eight dimensions of quality. Rather, he should select a number of dimensions on which to compete.

Traditionally, the quality control departments in factories compete with regard to the conformance dimension of quality. It is their responsibility to ensure that requirements set on a quality characteristic are met. Such requirements are usually stated in the form of specification limits. All parts within limits are classified as conforming. The objective then is to produce “zero defects”. Sullivan (1984), argued that this conformance-to-specification-limits approach effectively prevents ongoing quality improvement. As long as all outcomes of a production process are within specification limits, a process engineer would have great difficulty in convincing his plant manager to make any investment for the improvement of quality. Sullivan advocates defining quality as “uniformity around the target”. In this more modern point of view, which was put forward by, among others, Deming and Taguchi, any deviation from target reduces reliability and increases costs, in the form of plant and customer loss. Operational objectives directed towards achieving ongoing quality improvement should not be stated in terms of specification limits such as “zero defects”. The attention for quality improvement will diminish as soon as the manufacturing process is able to produce amply within specs. A more continuous drive for ongoing quality improvement will be obtained if the aim is to reduce variation around the target.

This relation between “quality” and “variation” is summarized in the following phrase which can be found in Montgomery (1996):

“Quality is inversely proportional to variability”.

This statement demonstrates the contribution of statistical methods to quality improvement projects. By acknowledging that variation is present in process outcomes, and that they are to some extent uncertain, it becomes necessary to employ methods which take uncertainty explicitly into account.

In this thesis, techniques are described that are aimed at improving quality through reduction of variability. We realize that this only covers one of the many facets of quality, but nevertheless, it can make an important contribution.
2.2 Reducing variability

In the previous section, an attempt was made to define the use of the word “quality”. As a result of the close relation between quality and variation, it was natural to use terminology that expresses uncertainty in process outcomes such as “rarely”, “degree to which”, “probability”, and “variation”. It was also argued that quality improvement can be obtained by reduction of variability. The approach towards reduction of variability is based on the idea that the phenomena causing variation in process outcomes can be classified in two groups: common causes of variation and special causes of variation. In Subsection 2.2.1, the difference between these classes is discussed. Moreover, we will concern ourselves with the question how this distinction can be utilized to improve the quality of process outcomes. In many practical cases, the implementation of this approach is limited to manufacturing processes. However, the basic ideas are applicable in all areas of the organization, from production to sales. It is important that the whole organization is aware of the opportunities of quality improvement by reduction of variation. This concept, which is sometimes called “Statistical Thinking”, is discussed in Subsection 2.2.2.

The information concerning the presence of one or both types of variation is in practice based on samples from a process. These samples should be taken with care. This is discussed in Subsection 2.3.1.

2.2.1 Common causes and special causes of variation

Dr. Walter A. Shewhart is unmistakably the founding father of Statistical Process Control. The concepts he developed in the twenties have been written down in his pioneering work, “Economic Control of Quality of Manufactured Product” (1931). His ideas are still very relevant and the most important tool, the control chart (which will be discussed in the next section and, in fact, throughout this thesis), is nowadays used worldwide in a virtually unaltered form.

Shewhart compares a manufacturer that tries to make all products conform with a certain target to a marksman, aiming at a bull’s-eye. Just as it is impossible for the marksman to hit the bull’s-eye with every shot, it is not to be expected that every product will exactly comply with the target. This is not a problem in itself, since ‘small’ variations around the target are acceptable for most customers. The problem is in Shewhart’s words:
“how much may the quality of a product vary and yet be controlled? In other words, how much variation should we leave to chance?”

It is important to realize that this formulation of the problem explicitly states that, even if the quality of a certain product is ‘controlled’, some variability must be allowed for. The marksman may do his utmost to hit the bull’s-eye with each shot, but he cannot hope to hit the target every time. A complex of causes of variation exerts its influence on the outcome of each shot. Both the manufacturer and the marksman are not always able to do what they want to do and cannot hope to precisely understand why.

The marksman’s example is just one of many that can be given to illustrate that in all aspects of our lives, there is some variation that is considered to be ‘normal’ or ‘acceptable’, and does not call for action. The underlying sources that are responsible for this type of variation were originally called chance causes of variation by Shewhart. Nowadays, the term common causes is used, a reformulation that is due to Deming. An important aspect of the variation due to common causes is that it is to some extent predictable: it is possible, based on earlier experiences, to determine limits that bound the effect of the common causes. In the marksman’s example, numerous common causes such as wind, trembling of the marksman’s hands, and the precision of the rifle, affect the result of each shot. As long as the hits are within some range of the bull’s-eye, the marksman leaves the variation to chance.

If the observed variability is such that it exceeds the boundaries of ‘normal’ variation we are inclined to undertake action. The presence of something out of the ordinary, not within the class of common causes, is suspected. Such causes of variation belong to the class of special causes (in Shewhart’s terminology: assignable causes). If a single shot of the marksman deviates more than normally from the bull’s-eye because he was startled by some noise of the bystanders, he might want to ask them to keep quiet or he might want to use earplugs.

In a manufacturing environment, it is, in most cases, not difficult to visualize a large number of common causes, the joint effect of which causes variation in the outcomes. This variation is inherently part of the process, and is always present, from day to day, from hour to hour. Usually, it is not within the power of an operator to remove common causes of variation; such variation is ‘left to chance’. If it is necessary to remove common causes of variation, this requires in most cases a profound revision of the process,
which is the responsibility of the owner of the process, i.e. management. Special causes of variation are not part of the process, and occur only accidentally. However, when a special cause of variation is present, it will have a large effect on the outcomes of the manufacturing process. If removal is possible, a special cause can usually be eliminated without revising the process. In many cases, an operator can be instructed to recognize and remove special causes of variation, thereby improving the quality of the outcomes of the process.

It is important to realize that the responsibility for reducing the effect of special causes of variation lies on a different management level than the reduction of common causes of variation. Counteracting the effect of special causes of variation can be delegated to operators, whereas reducing the effect of common causes of variation is the responsibility of the owner of the process. It is important for an operator to know whether or not special causes are present, so that he can undertake action to remove this cause of variation. But it is even more important to know when to leave the process alone when only common causes of variation are affecting the outcomes.

What happens in practice, is that operators try to counteract the effect of common causes of variation as if it was a special cause of variation. In many cases, this will result in larger variation in the outcomes. This phenomenon of intervening in a stable process when it would have been better to do nothing, is called 'tampering'. It results in frustration, because of unsuccessful searches for special causes of variation, and in waste of time and money.

It is therefore of critical importance to be able to distinguish situations where only common causes of variation affect the outcomes of a process, from situations where also special causes are present. If only common causes of variation are present, the manufacturing process is said to be “statistically in control”. This does not mean that there is no variation, or that there is small variation. It does mean that the outcomes are predictable, within statistical limits. In Shewhart’s words:

“... a phenomenon will be said to be controlled when, through the use of past experience, we can predict, at least within limits, how the phenomenon may be expected to vary in the future. Here it is understood that prediction within limits means that we can state, at least approximately, the probability that the observed phenomenon will fall within the given limits.”
The predictability of a process that is statistically in control is the basis for the control chart, a tool that can be utilized to distinguish between situations where only common causes of variation affect the outcomes of a process, and situations where also special causes are present. The control chart will be discussed in Section 2.3 for the case of independent observations.

In the case of independent observations, the term ‘an in-control process’ is associated with a sequence of independently and identically distributed observations. In most applications and in many textbooks, these requirements are adopted as necessary and sufficient conditions for an in-control process. However, it must be noted that Shewhart’s original definition does not require independence of successive observations from an in-control process. It only demands that we can predict how the process may be expected to vary in the future. In Subsection 2.5, we will extend the definition of an in-control process to include cases where local deviations of the mean are allowed, due to stationary serial correlation. Such a definition is entirely in line with Shewhart’s definition of an in-control process.

In the next subsection, which is based on an article by Snee (1990), it is shown that the philosophy behind SPC is not restricted to manufacturing processes alone.

### 2.2.2 Statistical thinking

Application of the idea of improving quality by reduction of variability is in most cases limited to manufacturing environments. However, the concept can be applied at many other levels of an organization, as well. In Snee (1990), the place of Statistical Process Control in the much broader context of Total Quality Management is discussed. Successful implementation of such a strategy requires a new way of thinking, which he calls statistical thinking. These ideas can provide a useful contribution to efforts aimed at quality improvement at all levels in an organization, from production to sales. Snee describes the essence of statistical thinking as follows:

> “... all work is a series of interconnected processes and identifying, characterizing, quantifying, controlling, and reducing variation provide opportunities for improvement.”

Even more insightful is the schematic presentation that Snee gave of statistical thinking in quality improvement, see Figure 2.1.
2.2. REDUCING VARIABILITY

In the first two boxes of Figure 2.1, it is indicated why statistical thinking is a logical approach to follow in all activities that are aimed at improving quality. All work that is done can be viewed as a process. A process can be defined as (see Nolan and Provost (1990)): "a set of causes and conditions that repeatedly come together to transform inputs into outcomes". The inputs might include people, materials, or information. The outcomes include products, services, behavior, or people.

In all such processes, variation is encountered. Careful analysis of this variation, combined with knowledge of the process may lead to reduction of variation. It is therefore important for a manager to realize that close cooperation with those who work with the process (e.g. operators), is an absolute necessity for successful quality improvement. The people that work with the process posses much of the knowledge that is needed to reduce variation.

Reduction of variation may be accomplished by one of two paths. Reduction of variation may be brought about by removing special causes of variation, which is the responsibility of the people working with the process. The other path is to reduce the effect of common causes of variation, which requires the management to undertake action. Removal of common causes requires a different approach than removal of special causes of variation.

Figure 2.1: Statistical Thinking in Quality Improvement
However,

“Deming and his colleagues point out that managers typically treat all problem as due to special-cause variation, when, in fact, more than 85% of problems are due to defects in a system (common-cause variation), which only management can change. The result is that management spends too much time ‘fire-fighting,’ solving the same problem again and again because the system was not changed”

Snee (1990) page 120.

2.3 The Shewhart control chart

In the foregoing section, we argued that it is very important to detect the presence of special causes of variation. A tool is needed for this, since the effect of a possible special cause is hidden in the variation due to common causes. Shewhart developed the control chart for this purpose. It will be discussed in this section.

The control chart is based on the idea that if the process is in a state of statistical control, the outcomes are predictable. Based on previous observations, it is possible for a given set of limits to determine the probability that future observations fall within these limits.

Observations that are utilized for monitoring the process are usually grouped according to time, amount of production, or some other descriptive statistic. Sampling and subgrouping should be carried out with care. In Subsection 2.3.1, it will be discussed how to take and group the observations. From each subgroup sample, descriptive statistics such as the mean, the range or the sample standard deviation are computed.

In its original form, the control chart is a simple time plot of a sequence of subgroup statistics. The points in the plot are compared to limits, which indicate the bandwidth of the variation due to common causes. These limits are called control limits. The width of these limits is such that, as long as all points are within the control limits, it is reasonable to assume that the underlying process is statistically in control. A point outside the control limits is called an out-of-control signal, as it indicates that more variation is present than can be attributed to the effect of common causes of variation. However, due to the random nature of the observations, there is a small probability that an out-of-control signal is encountered while the process is
2.3. THE SHEWHART CONTROL CHART

Statistically in control. Such a signal is called a *false out-of-control signal*. In Figure 2.2, an example of a control chart is depicted.

![Control Chart Diagram](image)

**Figure 2.2:** Illustration of a control chart.

Control limits are not to be confused with specifications or other targets for the process. They are simply a prediction of the variation that will occur due to common causes. If the variation due to common causes is relatively large, all points on the control chart may be within the control limits, but the process outcomes might fail to meet the specification limits. The presence of special causes does not necessarily mean that there is large variation, or that the specifications are not met. It does mean that there is some source of variation that causes the measurements to be more variable than can be attributed to common causes.

Specification limits are agreements between manufacturer and its customers concerning the tolerated deviation from target. They are in no way related to the actual performance of the process. Control limits on the other hand indicate the magnitude of the variability of the process when only common causes are present. They reveal what Nelson (1988) calls ‘the heartbeat of the process’.

It is not advisable to plot specification limits in a control chart. Not only to avoid confusion: understanding the difference between the two concepts is hard enough. A more important reason is that specification limits relate to single products, whereas control limits are usually computed for a grouped variable such as a sample mean. A sample mean that is both within control limits and specification limits may give the impression that the process is performing as required, also in situations where some of the
individual observations exceed the specification limits.

2.3.1 Rational subgroups

Briefly, a control chart is a tool designed to judge whether a process is statistically in control or not. The control chart utilizes samples of observations, mostly drawn in subgroups, to obtain information about the behavior of the underlying process. Statistics that summarize the information in the subsamples are compared to limits which represent the variation due to common causes. If an out-of-control signal is observed, action is required to track down the special cause that is responsible. Sampling and subgrouping of the observations must be carried out in such a way that the search for a causal relation between the out-of-control signal and some underlying disturbing phenomenon is facilitated.

The way the observations are sampled and grouped may have a large effect on the behavior of a control chart. This phenomenon is illustrated in Section 6.1, where serially correlated measurements result in control limits that are too tight, leading to too many out-of-control alarms. Subgrouping of the data should be carried out with care, especially in situations where subsequent observations are serially correlated.

In addition to providing us with the control chart, Shewhart also gave guidelines for sampling subgroups of observations from a process. To this end, he introduced the concept of “rational subgroups”. A rational subgroup is a sample in which all of the items are produced under conditions in which only common causes are responsible for the observed variation. Special causes do not occur within a rational subgroup, but only between subgroups. If the observations can be grouped according to these requirements, then appropriate control limits can be determined that discriminate between in-control situations and out-of-control situations.

Linking an out-of-control signal to a specific special cause of variation is facilitated by preserving the time order in the data points on the control chart. The time at which an out-of-control signal is given may give a hint as to when the special cause has occurred.

A control chart must be sensitive enough to detect the effect of special causes of variation, but must not generate too many false out-of-control signals. In practice, a balance between these two must be struck by determining the width of the control limits. How this is done for the classical Shewhart chart is the subject of the following two subsections. Following Does and Schriever (1992), we distinguish two situations.

In Subsection 2.3.2, the situation is discussed where we have $k$ rational
subgroups of size $n$ of past observations available. These observations are used to set up the control chart. This is sometimes called “Phase I”, Does and Schriever (1992) call it “analysis of past data”. In Phase I, the magnitude of the variation due to common causes is determined, so that the width of the control limits can be determined for what is sometimes called “Phase II”. In Phase II, rational subsamples become available one by one, as they are drawn online from the process. Does and Schriever (1992) call this phase “performance of current control. Determining control limits for the Shewhart chart in Phase II will be discussed in Subsection 2.3.3.

2.3.2 Phase I: setting up of the control chart

In order to be able to compute the control limits for Phase I, some formalization of the foregoing is needed. Suppose that, from a certain manufacturing process, $k$ rational subgroups of size $n$ are sampled. Let us denote the $j^{th}$ observation of the $i^{th}$ subsample by $X_{ij}$, where $i = 1, \cdots, k$, and $j = 1, \cdots, n$. For the time being, we will assume that for all $i$, the individual observations $X_{i1}, \cdots, X_{in}$ are identically and independently distributed within sample $i$, with distribution function $F_i$. Furthermore, observations in different samples are also assumed to be independent. The independence assumption will be loosened in subsequent chapters.

Note that the foregoing includes a formalization of the concept of rational subgroups. Distribution functions $F_1, \cdots, F_k$ are used to model the joint effect of common and special causes of variation. Within each sample, the distribution function does not change. However, due to the presence of special causes, the distribution function may change over time.

If there are no special causes present, we assume that the distribution functions do not change over time. Monitoring the process for special causes of variation can then be formalized as testing the hypothesis

$$H_0 : F_1 = F_2 = \cdots = F_k = F_0$$

against the alternative

$$H_1 : \text{there are } s, t \in \{1, 2, \cdots, k\} \text{ such that } F_s \neq F_t.$$
these parameters. An out-of-control signal on one of these charts is an
dication of the presence of a special cause of variation.

The control chart for a one-dimensional parameter $\eta_i$ of $F_i$ is set up in
the following way. Let $T_i = T(X_{i1}, \ldots, X_{in})$ be an estimate of $\eta_i$ based
on the sample at time $i$. Furthermore, let $M_k$ and $V_k$ be statistics based
on the $k$ samples such that $M_k = M(X_{11}, \ldots, X_{1n}, \ldots, X_{k1}, \ldots, X_{kn})$ is
a consistent estimator of the location of the distribution of $T_i$ under $H_0$, and
$V_k = V(X_{11}, \ldots, X_{1n}, \ldots, X_{k1}, \ldots, X_{kn})$ is a consistent estimator of
the spread of the distribution of $T_i$ under $H_0$.

In the control chart, realizations of $T_i$ are plotted against $i$. These
values are compared to control limits of the form

$$L = M_k + c(n, k, p) V_k,$$

where $c(n, k, p)$ is the $p^{th}$ percentile of the null distribution of $(T_i - M_k)/V_k$.
For the LCL and the UCL of a control chart, different constants $c(n, k, p_{LCL})$
and $c(n, k, p_{UCL})$ must be determined.

In situations where a location parameter or a spread parameter of the
distribution function of $T_i$ is known, these values are used in (2.1) instead
of their estimates.

Note that the elements of the sequence $\{T_1, \ldots, T_k\}$ are mutually inde-
dendent, but this does not, in general, hold for $T_i, M_k$, and $V_k$, since they
are (partly) based on the same set of observations.

In most literature on SPC it is assumed that $F_i$ is a normal distribution
function with expectation $\mu_i$ and variance $\sigma_i^2$. The mean and/or variance
of the observations may change over time due to the presence of special
causes of variation. With these assumptions, the process is in control if
and only if $\mu_i = \mu$ for some $\mu$ and if $\sigma_i = \sigma$ for some $\sigma$ for all $i = 1, \ldots, k$.
It is for this reason that in a lot of cases, a production process is monitored
using two control charts, one for the standard deviation, and one for the
mean of the process.

Shewhart did not consider statistical arguments to determine the con-
stants $c(n, k, p_{LCL})$ and $c(n, k, p_{UCL})$. He decided to choose, “based on eco-
nomic considerations”

$$c(n, k, p_{LCL}) = -3$$

and

$$c(n, k, p_{UCL}) = 3.$$
These values turn out to work well in a lot of practical cases. The underlying statistical arguments for a control chart for the mean of normal observations are the following. Suppose that we are testing the hypothesis $H_0 : \mu_1 = \cdots = \mu_k \equiv \mu$ where $\mu$ is known, assuming a constant known variance $\sigma^2$. Furthermore, assume that the sample means $T_i = 1/n(X_{1i} + \cdots + X_{ni})$ are plotted in a control chart with limits $LCL = \mu - 3\sigma/\sqrt{n}$, and $UCL = \mu + 3\sigma/\sqrt{n}$. Then we have $p_{LCL} = (1 - p_{UCL}) = 0.00135$, so that under these assumptions a false out-of-control signal is quite unlikely.

If an out-of-control signal is generated in Phase I, a search is initiated for a responsible special cause of variation. If this can be found, action should be taken to prevent it from re-occurring. In cases where a special cause is found and removed, the corresponding sample does not provide information about the in-control state of the process. Therefore, in such cases, it should be removed from the data set, and the remaining $k - 1$ subsamples should be compared to re-estimated control limits.

This procedure should be repeated until no out-of-control signals are generated, or when underlying special causes either cannot be found or cannot be removed. At the end of Phase I, we have a data set at our disposal of, say, $m \leq k$ subsamples that provides information concerning the variability that can be attributed to common causes of variation. This information is needed for Phase II, when samples are drawn online.

### 2.3.3 Phase II: current control

In Phase II, we have an estimate of the in-control distribution available based on $m$ samples from Phase I. This distribution function will be denoted by $F_0$. Each time a sample $X_{f1}, \cdots, X_{fn}$ becomes available at time $f > k$, we want to test the hypothesis

$$H_0 : F_f = F_0$$

against the alternative

$$H_1 : F_f \neq F_0.$$
The control charts discussed in this section are Shewhart-type control charts. Their performance under various levels of first-order autocorrelation (including the special case of independence) will be studied in Chapter 3. More efficient control charts are also developed, such as the EWMA control chart and the CUSUM control chart. Their performance for various levels of first-order serial correlation will be discussed in Chapter 4 and Chapter 5, respectively.

2.4 Serial correlation

In the previous sections, we made the assumption that the measurements that are used to monitor the process are independently distributed. This is standard practice in most literature on SPC. In this thesis, it is investigated how the performance of control charts is affected by serial correlation in the observations. This is motivated by many practical situations, where the assumption that the observations are (approximately) uncorrelated cannot be justified, see also Chapter 1. In this section, we will shortly review some of the types of serial correlation that will be considered in subsequent chapters. We will restrict ourselves to serial correlation that can be successfully modelled using ARIMA($p,d,q$) models. This class of models is discussed in a large number of texts on time series, such as Box and Jenkins (1976), Pandit and Wu (1983), or Montgomery, Johnson and Gardiner (1990). We refer to such texts for a deeper discussion of time series analysis.

Throughout this thesis, a serially correlated sequence of variables will be denoted by \( \{Y_1, \ldots, Y_T\} \), whereas a sequence of independent random variables will be denoted by \( \{X_1, \ldots, X_T\} \).

Before we turn to the discussion of ARIMA models, we introduce some of the terminology that is used in time series analysis. This will facilitate the discussion in subsequent chapters. Much of it is taken from Anderson (1976).

A time series model that has proved to be useful in practice is the AR(1) model. This model deserves and will receive most of the attention in subsequent chapters. It is one of the simplest special cases of ARIMA($p,d,q$) models. It will be discussed in Subsection 2.4.2. More general ARIMA($p,d,q$) models will be discussed in Subsection 2.4.3.

2.4.1 Some time series analysis terminology

We define a *time series* as a set of observations that are ordered in time. We will only consider discrete series, with observations drawn at equal
time intervals. A sequence of observations \( \{y_1, y_2, \cdots, y_T\} \) can be viewed as a single realization of some underlying stochastic process. Each \( y_i \) is the realization of some random variable \( Y_i \), which has an associated probability density function \( f_{Y_i}(\cdot) \). We assume that for any set of \( Y_i \)'s, say \( Y_{j_1}, \cdots, Y_{j_r} \), a joint probability density function \( f_{Y_{j_1}, \cdots, Y_{j_r}}(\cdot) \) exists. If a statistical process is such that \( f_{Y_{i+n_1}, \cdots, Y_{i+n_m}}(\cdot) \) is independent of \( i \) for any positive integer \( m \) and for any choice of \( n_1, \cdots, n_m \) then the probabilistic structure does not change over time and the process is said to be strictly stationary. Otherwise, it is nonstationary. If the definition of strict stationarity only holds for \( m \leq p \) for some positive integer \( p \), then the process is said to be stationary of order \( p \).

A Gaussian process is defined by the property that the probability density function associated with any finite subset of \( \{ \cdots, Y_{-2}, Y_{-1}, Y_0, Y_1, Y_2, \cdots \} \) is multivariate normal. For a Gaussian process, a sufficient condition for strict stationarity is stationarity of order 2, since all moments of higher order are then precisely fixed. Stationarity of order 2 is sometimes also called weak stationarity.

It is important to make the distinction between (weakly) stationary models and nonstationary models, since observations from the former wander around a fixed mean. Observations from a nonstationary model may wander away from a specified target value if no action is taken. That is, nonstationary processes require some form of Automated Process Control (APC) if the quality characteristic should fall between certain specification limits. Stationary processes are mean reverting and need not be controlled by APC techniques. However, the performance of such processes can be improved by applying APC, see Box and Luceño (1997b).

Weak stationarity implies that for all \( i \)

\[
E(Y_i) = \mu
\]

and

\[
\text{Cov}(Y_i, Y_{i-k}) = \gamma_k,
\]

where \( \mu \) is the constant mean of the process, and \( \gamma_k \), the autocovariance at lag \( k \) (integer), is also constant. In particular, \( Y_i \) has constant variance \( \sigma^2 = \gamma_0 \).

Furthermore, we have for all integers \( k \),

\[
\gamma_{-k} = \gamma_k
\]
since
\[ \text{Cov}(Y_t, Y_{t+k}) = \text{Cov}(Y_{t+k}, Y_t) = \text{Cov}(Y_t, Y_{t-k}). \]

Hence it is only necessary to consider \( \gamma_k \) for \( k > 0 \). The set \( \{\gamma_0, \gamma_1, \cdots\} \) is sometimes called the \textit{autocovariance function}. We define \( \rho_k \), the \textit{autocorrelation coefficient} at lag \( k \) by
\[ \rho_k = \frac{\gamma_k}{\gamma_0}. \]

The set \( \{\rho_0, \rho_1, \cdots\} \) is called the \textit{AutoCorrelation Function} (ACF) of the process. An estimate of the ACF can be used to identify which models within the class of ARIMA(\( p,d,q \)) can be used to model the random behavior in a given set of observations.

### 2.4.2 The AR(1) process

A process \( \{Y_t\} \) is said to be a first-order autoregressive process (AR(1) process) if it is generated by
\[ Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t \quad \text{for } t \in \mathbb{Z}, \tag{2.2} \]
where \( \phi \) is some constant satisfying \( \phi \in (-1, 1) \), and \( \{\varepsilon_t\} \) is a sequence of i.i.d. disturbances, \( \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2) \) for \( t \in \mathbb{Z} \). Subsequent observations of model (2.2) are serially correlated, since
\[ \text{Cov}(Y_t, Y_{t-k}) = \phi^k \sigma_Y^2, \]
where
\[ \sigma_Y^2 = \text{Var}(Y_t) = \frac{\sigma_\varepsilon^2}{1 - \phi^2}. \]

Model (2.2) is strictly stationary since \( |\phi| < 1 \).

Obviously, \( \text{E}(Y_t) = \mu \) for all \( t \) in model (2.2). A model that includes the possibility of a shift in \( \text{E}(Y_t) \equiv \mu_t \) is
\[ Y_t - \mu_t = \phi(Y_{t-1} - \mu_{t-1}) + \varepsilon_t \quad \text{for } t \in \mathbb{Z}. \tag{2.3} \]

First-order autoregressive models are useful when disturbances affect not only the current outcome of the process, but also have an (exponentially declining effect) on future outcomes. Such situations occur for example when a tank containing raw material is refilled from time to time with raw material of varying quality, see also the example considered in Subsection 1.1.3.
2.4. SERIAL CORRELATION

2.4.3 ARIMA processes

More general AR processes are of the form

\[ Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \varepsilon_t \quad \text{for } t \in \mathbb{Z}, \quad (2.4) \]

If we introduce the backward shift operator \( B \), where \( BY_t = Y_{t-1} \), then (2.4) can be rewritten as

\[ \phi(B)(Y_t - \mu) = \varepsilon_t, \]

where \( \phi(B) \) is a polynomial in \( B \) of degree \( p \). Autoregressive models of order \( p \) are stationary if all roots of the polynomial \( \phi(\cdot) \) are outside the unit circle (see Box and Jenkins (1976), Section 3.2).

A Moving Average (MA) process of order \( q \) is generated by

\[ Y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}. \]

Moving average models of any order are always stationary.

A useful class of models for time series is formed from a combination of MA and AR processes. A mixed autoregressive moving average process containing \( p \) AR terms and \( q \) MA terms is abbreviated to an ARMA(\( p,q \)) process, and is given by

\[ Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q} \quad \text{for } t \in \mathbb{Z}, \]

Or, rewritten using the backward shift operator \( B \):

\[ \phi(B)(Y_t - \mu) = \theta(B)\varepsilon_t, \quad (2.5) \]

where \( \phi(B) \) and \( \theta(B) \) are polynomials of degrees \( p \) and \( q \), respectively. Such processes are stationary if the roots of the polynomial \( \phi(\cdot) \) lie outside the unit circle. The class of ARMA(\( p,q \)) models can be used to model a wide range of stationary time series, with only a few parameters to estimate. In practice, values of \( p \) and \( q \) larger than 2 are rarely encountered.

However, many time series encountered in practice are nonstationary. In order to fit a stationary model it is necessary to remove the nonstationarity first. In many cases, this can be obtained by taking successive differences of the observations one or more times. That is, in case of first differences,
we consider \( \{y_2 - y_1, y_3 - y_2, \ldots, y_T - y_{T-1}\} \) and try to fit a stationary ARMA\((p,q)\) model to these new observations. If taking first differences is not enough to obtain stationarity, the observations are differenced once more, and so on, until stationarity is obtained. Taking differences can be expressed in terms of the backward shift operator as \((1 - B)y_t\). The notation \((1 - B)^d y_t\) is used to indicate that successive differences were taken \(d\) times.

If we replace \(Y_t - \mu\) in Equation (2.5) by random variables that are differenced \(d\) times, we have

\[
\phi(B)(1 - B)^d(Y_t - \mu) = \theta(B)\varepsilon_t.
\]

Such models are called AutoRegressive Integrated Moving Average (ARIMA) models of order \((p,d,q)\). The term ‘integrated’ is used because the stationary model which is fitted to the differenced data has to be summed or ‘integrated’ to provide a model for the nonstationary data. ARIMA\((p,d,q)\) models are capable of describing certain types of nonstationary time series. An important special case is the IMA\((1,1)\) model, which is often encountered in practice. In Chapter 8, an example of an IMA\((1,1)\) process is discussed.

The ARIMA\((p,d,q)\) models were popularized by Box and Jenkins (1976). Box and Jenkins (1963) themselves developed one of the first control charts to account for serial correlation. By assuming that the quality characteristic was drifting away from its target value according to an ARIMA\((0,1,1)\) model, they derived a chart with action limits that are determined such that the total costs of running the process are minimized. The cost minimization procedure required trading off the cost of being off-target with the cost of resetting the machine.

### 2.5 Control charts for serially correlated data

In Section 2.3, it was discussed that the control chart is the tool to detect special causes of variation. The control limits that are drawn on the control chart bound the variation due to common causes of variation. For the proper placement of control limits in the case of serial correlated data, it is of crucial importance to decide which process behaviors are part of the process, and which are attributed to special causes. Gilbert, Kirby and Hild (1997) state the following.
“For example, if autoregressive behavior is a normal, unchangeable part of the process, then a chart that gives out-of-control signals because of the presence of autocorrelation is not very useful. On the other hand, if autocorrelation in a process is a symptom of a problem that should be addressed, then the control chart should detect the presence of the autocorrelation.”

Crowder, Hawkins, Reynolds and Yashchin (1997) share this view. They argue that the cause of autocorrelation should be assessed before the data is analyzed and interpreted. As argued in Chapter 1, we will consider cases where autocorrelation in process data is unremovable and part of the process. Control charts should not signal because of autocorrelation, but give out-of-control signals because of the presence of special causes of variation.

Consequently, the definition of an in-control process that is most commonly used in practice and Shewhart’s original definition do not necessarily agree. In practice, the term ‘an in-control process’ is more often than not associated with a sequence of independently and identically distributed observations. As was discussed in Subsection 2.2.1, Shewhart’s original definition of an in-control process only requires that we can predict (within statistically determined limits) how the process may be expected to vary in the future. A process that exhibits serial correlation is predictable. For this reason we extend the definition of an in-control process to include observations which may be serially correlated. Alwan (1988) refers to such processes as being ‘in control in a broader sense’.