Quality Engineering - Class Notes (experimental)

Jonathan Rosenblatt

November 8, 2015

Preface

This text accompanies my Quality Engineering course at the Dept. of Industrial Engineering at the Ben-Gurion University of the Negev. It has several purposes:

- Help me organize and document the course material.

- Help students during class so that they may focus on listening and not writing.

- Help students after class, so that they may self-study.

At its current state it is experimental. It can thus be expected to change from time to time, and include mistakes. I will be enormously grateful to whoever decides to share with me any mistakes found.

I also ask for the readers' forgiveness for my Wikipedia quoting style. It is highly unorthodox to cite Wikipedia as one would cite a peer reviewed publication. I do so, in this text, merely for technical convenience.

I hope the reader will find this text interesting and useful.

Contents

List of Figures

List of Definitions

Chapter 1

Introduction

Quality Engineering is the study and design of practices aimed improving the "quality" of production. Production is understood in a wide sense, and includes services as well. Quality is understood in many senses. Here are several definitions compiled verbatim from Montgomery (2007) and Wikipedia (2015d):

1. Montgomery: "The reciprocal of variability".

2. American Society for Quality: A combination of quantitative and qualitative perspectives for which each person has his or her own definition; examples of which include, "Meeting the requirements and expectations in service or product that were committed to" and "Pursuit of optimal solutions contributing to confirmed successes, fulfilling accountabilities. In technical usage, quality can have two meanings: (a) The characteristics of a product or service that bear on its ability to satisfy stated or implied needs. (b) A product or service free of deficiencies."

3. Subir Chowdhury: "Quality combines people power and process power".

4. Philip B. Crosby: "Conformance to requirements."

5. W. Edwards Deming: "The efficient production of the quality that the market expects".

6. W. Edwards Deming: "Costs go down and productivity goes up as improvement of quality is accomplished by better management of design, engineering, testing and by improvement of processes."

7. Peter Drucker: "Quality in a product or service is not what the supplier puts in. It is what the customer gets out and is willing to pay for."

8. Victor A. Elias: "Quality is the ability of performance, in each Theme of Performance, to enact a strategy."

9. ISO 9000: "Degree to which a set of inherent characteristics fulfills requirements."

10. Joseph M. Juran: "Fitness for use.".

11. Noriaki Kano and others, present a two-dimensional model of quality: "must-be quality" and "attractive quality." The former is near to "fitness for use" and the latter is what the customer would love, but has not yet thought about. Supporters characterize this model more succinctly as: "Products and services that meet or exceed customers' expectations."

12. Robert Pirsig: "The result of care."

13. Six Sigma: "Number of defects per million opportunities."

14. Genichi Taguchi: "Uniformity around a target value."

15. Genichi Taguchi: "The loss a product imposes on society after it is shipped."

16. Gerald M. Weinberg: "Value to some person".

17. Jonathan D. Rosenblatt: "The efficient fulfilment of a promise".

Collecting ideas

1. Quality is not only about production.

2. Quality is the means, not the end.

3. Quality may deal with the **design** or with **conformance** to a given design.

Almost all of the above definitions, may apply to different characteristics, we call *dimensions of quality*. Following Wikipedia (2015b) :

Dimensions of Quality

Performance Performance refers to a product's primary operating characteristics. This dimension of quality involves measurable attributes; brands can usually be ranked objectively on individual aspects of performance.

Features Features are additional characteristics that enhance the appeal of the product or service to the user.

Reliability Reliability is the likelihood that a product will not fail within a specific time period. This is a key element for users who need the product to work without fail.

Conformance Conformance is the precision with which the product or service meets the specified standards.

Durability Durability measures the length of a product's life. When the product can be repaired, estimating durability is more complicated. The item will be used until it is no longer economical to operate it. This happens when the repair rate and the associated costs increase significantly.

Serviceability Serviceability is the speed with which the product can be put into service when it breaks down, as well as the competence and the behavior of the service person.

Aesthetics Aesthetics is the subjective dimension indicating the kind of response a user has to a product. It represents the individual's personal preference.

Perceived Quality Perceived Quality is the quality attributed to a good or service based on indirect measures.

1.1 Terminology and Concepts

1.1.1 Basic Terminology

Quality Characteristics A.k.a. *Critical to Quality Characteristics* (CTQs). May be physical, sensory, or temporal properties of a process/product. Obviously related to the dimensions of quality.

Quality Engineering "The set of operational, managerial, and engineering activities that a company uses to ensure that the quality characteristics of a product are at the nominal or required levels and that the variability around these desired levels is minimum." (Montgomery, 2007)

Variables Continuous measurements of some CTQ.

Attributes Discrete measurements of some CTQ.

Target Value The desired level of a particular CTQ. A.k.a. *nominal* value.

USL & LSL Largest and smallest allowable values of a CTQ.

Specifications The set of permissible values for all CTQs. Either a set of target values, or USL-LSL intervals.

Non-conformity A non conforming product is one that fails to meet the specification.

Fallout The same as non-conformity.

Defect A non-conformity that is serious enough to affect the use of the product.

DPMO Defect per million opportunities.

PPM Parts per million. Interchangeable with DPMO.

1.1.2 Statistical Terminology

Exploratory Data Analysis (EDA) An assumption free quantitative inspection of data; "Story telling"; no inference.

Inference Data analysis with the intention of generalizing from a sample to a population.

Causal Inference Inference, with the intention of claiming causal relations between quantities under study.

Predictive Analytics Data analysis with the intention of making predictions with future data. Can be seen as inference, without aiming at causality.

Design of experiments (DOE) By far the best and most established way for causal inference. The *random assignment* of units to groups allows to interpret statistical correlations as causal.

Statistical Process Control (SPC) Data analysis with the intention of identifying anomalous behaviour with respect to history[1].

Computer Simulation Well, just what the name implies.

---

[1]Akin to *anomaly detection*, or *novelty detection*, in the machine learning literature.

Control Chart A graphic visualization of the historical evolution of one (or several) CTQs.

(Un)Controllable Inputs Each process has inputs that affect the behaviour of the CTQ. Some are controllable, and some are not.

Factorial Design In the language of DOE, controllable inputs are *factors*. A factorial design, is an experiment where factors are varied in order to study their effect on the CTQ.

Off/On-line process control SPC can be performed on or off line. On-line, a.k.a. *in-process control*, meaning control happens as the process evolves, and off-line meaning before it starts or after it has finished.

Engineering control A.k.a. *automatic control*, or *feedback control*. SPC that triggers an intervention that keeps the process in control

Outgoing/Ingoing Inspection Refers to the stage at which SPC is performed. As inputs come in (ingoing), or as outputs come out (outgoing).

1.2 Some History

1875	Frederick W. Taylor introduces "Scientific Management"
	Henry Ford refines the assembly line to refine productivity and quality.
	AT&T begins systematic inspections.
	W.S. Gosset publishes the t-test.
1900	AT&T Bell labs establish a quality department.
1907	B. P. Dudding at General Electric in England uses statistical methods to control the quality of electric lamps
	R.A. Fisher inaugurates *design of experiments*.
1920	W. A. Shewhart introduces the *control chart* concept in a Bell Laboratories technical memorandum.
1924	
1928	Acceptance sampling refined by H. F. Dodge and H. G. Romig at Bell Labs.
1933	British textile and woolen industry and German chemical industry begin use of designed experiments for product/process development.
	Deming is invited to Japan to help occupation forces in rebuilding Japanese industry.
1946	
1948	G. Taguchi begins study and application of experimental design.

Table 1.1: Adapted from (Montgomery, 2007, Table 1.1).

1951	A. V. Feigenbaum publishes the first edition of his book, Total Quality Control.
1954	G. E. P. Box and K. B. Wilson publish fundamental work on designed experiments; focus is on chemical industry. Applications of designed experiments in the chemical industry grow steadily after this.
1960	Joseph M. Juran is invited by the Japanese to lecture on quality management and improvement.
	Courses in statistical quality control become widespread in industrial engineering academic programs.
	ISO publishes the first quality systems standard.
	Motorola's six-sigma initiative begins.
	Motorola's six-sigma approach spreads to other industries.
1987	ISO 9000:2000 standard is issued. Emphasis on supply-chain management and supplier quality. Expansion beyond the traditional industrial setting into financial services, health care, insurance.
1997	
2000	

Table 1.2: Adapted from (Montgomery, 2007, Table 1.1).

1.3 Management Aspects of Improving Quality

The founding fathers of QC have many do's-and-don'ts for managers. See Montgomery (2007, Sec 1.4) for details. As usual, we collect recurring ideas:

> 1. The responsibility for quality rests with management.
>
> 2. QC is not a one-time project, but an on-going process. It may advance continuously, or incrementally.
>
> 3. QC is (or should be) manifested in organizational structure, training, recruitment, incentives, knowledge management, to name a few.

1.4 Programs and Initiatives

1.4.1 Zero Defects Program (ZD)

Quoting Wikipedia (2015g):

> ...a management-led program to eliminate defects in industrial production that enjoyed brief popularity in American industry from 1964 to the early 1970's. Quality expert Philip Crosby later incorporated it into his "Absolutes of Quality Management" and it enjoyed a renaissance in the American automobile industry, as a performance goal more than as a program, in the 1990s. Although applicable to any type of enterprise, it has been primarily adopted within supply chains wherever large volumes of components are being purchased (common items such as nuts and bolts are good examples).

1.4.2 Quality is Free Initiative

Quoting Montgomery (2007):

> ...in which management worked on identifying the cost of quality (or the cost of *nonquality*, as the Quality is Free devotees so cleverly put it). Indeed, identification of quality costs can be very useful, but the Quality is Free practitioners often had no idea about what to do to actually improve many types of complex industrial processes.

1.4.3 Value Engineering Program (VE)

Quoting Wikipedia (2015f):

Value engineering (VE) is systematic method to improve the "value" of goods or products and services by using an examination of function. Value, as defined, is the ratio of function to cost. Value can therefore be increased by either improving the function or reducing the cost. It is a primary tenet of value engineering that basic functions be preserved and not be reduced as a consequence of pursuing value improvements.

1.4.4 Total Quality Management (TQM)

TQM originates in the 1980's with the ideas of Deming and Juran. It is a very wide framework that attempts at capturing the company-wide efforts required for QC. According to Montgomery (2007, p.23):

TQM has only had **moderate success** for a variety of reasons, but frequently because there is insufficient effort devoted to widespread utilization of the technical tools of variability reduction. Many organizations saw the mission of TQM as one of training. Consequently, many TQM efforts engaged in widespread training of the workforce in the philosophy of quality improvement and a few basic methods. This training was usually placed in the hands of human resources departments, and much of it was ineffective. The **trainers often had no real idea about what methods should be taught**, and success was usually measured by the percentage of the workforce that had been "trained," not by whether any measurable impact on business results had been achieved.

. . . Another reason for the erratic success of TQM is that many managers and executives have regarded it as **just another "program" to improve quality**. During the 1950's and 1960's, programs such as Zero Defects and Value Engineering abounded, but they had little real impact on quality and productivity improvement.

1.4.5 Six-Sigma

Quoting Montgomery (2007):

Products with many components typically have many opportunities for failure or defects to occur. Motorola developed the Six-Sigma program in the late 1980s as a response to the demand

for their products. The focus of six-sigma is reducing variability in key product quality characteristics to the level at which failure or defects are extremely unlikely.

Assume a device has m components. The failure probability of component $j \in 1, \ldots, m$ is α_j. What is the probability of the device failing, when assuming independent failures?

$$
\begin{align}
P(\text{failure}) &= P(\text{at least one failure}) \tag{1.1} \\
&= 1 - P(\text{no failure}) \\
&= 1 - \prod_{j=1}^{m}(1 - \alpha_j)
\end{align}
$$

Assuming all components have the same fallout rate, we omit the index j in α_j.

The failure probability α is implied by the CTQs, and its specification limits (USL, LSL). Denoting the target value of the CTQ by T, then $USL = T + \delta$ and $LSL = T - \delta$. Three-sigma means that the production variability, σ, is small enough so that

$$3\sigma = \delta.$$

Assuming

$$CTQ \sim \mathcal{N}\left(T, \sigma\right),$$

we can compute:

$$
\begin{align}
\alpha &= 1 - P(LSL < CTQ < USL) \tag{1.2} \\
&= 1 - P(|CTQ| < \delta) \tag{1.3} \\
&= 1 - P(|CTQ| < 3\sigma) = 0.0027. \tag{1.4}
\end{align}
$$

The 3-sigma quality guarantee is also known as $2,700$ defective parts per million (ppm) for now obvious reasons. Plugging the 3-sigma performance in Eq.(1.1) returns

PPM

$$P(\text{3-sigma failure}) < 1 - (1 - 0.0027)^m$$

Figure 1.1 illustrates the probability of failure against the number of components. For simple devices, the 3-sigma criterion may suffice. But now imagine the number of components in a car, a cellular phone, ... The 3-sigma rule is just not good enough. This is where 6-sigma requirement comes along. It implies that the production is process is so accurate that

$$6\sigma = \delta.$$

Figure 1.1: The probability of failure as a function of components under the 3-sigma standard.

Figure 1.2: The probability of failure as a function of components under the 6-sigma standard.

Updating Eq.(1.2) we get that the defective *ppm* of 6-sigma is 0.002. This is obviously excellent news, except for the typically tremendous effort involved in achieving this level of quality.

According to Montgomery (2007), the 6-sigma methodology has gained more success than it predecessors:

> The reason for the success of six-sigma in organizations outside the traditional manufacturing sphere is that variability is every-where, and where there is variability, there is an opportunity to improve business results.

1.4.6 Lean Systems

Quoting Wikipedia (2015c) (my own emphasis in bold):

> Essentially, lean is centered on making obvious what **adds value** by **reducing everything else**. Lean manufacturing is a man-agement philosophy derived mostly from the Toyota Production System (TPS) (hence the term Toyotism is also prevalent) and identified as "lean" only in the 1990s.

1.4.7 Design for Six-Sigma (DFSS)

Quoting Wikipedia (2015a) (my own emphasis in bold):

It is based on the use of **statistical tools** like linear regression and enables empirical research similar to that performed in other fields, such as social science. While the tools and order used in Six Sigma require a process to be in place and functioning, DFSS has the objective of **determining the needs of customers** and the business, and driving those needs into the product solution so created. DFSS is relevant for relatively simple items / systems. It is used for product or process design in contrast with process improvement.

1.4.8 Quality Systems and Standards

The first quality standard was issued by the International Standards Organization (ISO) in 1987. Current quality standards are known as the *ISO9000* series. These include: _{ISO9000}

ISO9000:2000 Quality Management System-Fundamentals and Vocabulary.

ISO9001:2000 Quality Management System-Requirements.

ISO9004:2000 Quality Management System-Guidelines for Performance Improvement.

In Israel, it is the Standards Institute of Israel[2] that may give ISO9000 (like any ISO) certifications upon inspecting the candidate organization. As emphasized by Montgomery (2007, p.24), ISO9000 is a set of rules and best practices, mostly oriented at *knowledge management*. It may help to *preserve* quality, but it does not, nor does it aim to, *improve* quality. As such, it will not be the focus of our course, which will focus on *statistical tools*.

Extra Info 1. [TODO: Just-in-Time, Poka-Yoke]

1.5 DMAIC

There are many names for the process of quantitative re-evaluations of performance against a given target: *data driven decision making* (DDD), *Shewart cycle*, etc. We will focus on one such framework, illustrated in Figure 1.3 known as DMAIC: Define, Measure, Analyze, Improve, Control.

Here are some general observations on DMAIC:

---

[2]https://portal.sii.org.il/heb/qualityauth/certificationtypes/ qualitylinks/iso9001/

Figure 1.3: The DMAIC cycle.
`http://www.sapartners.com/sigma-academy/`

1. It is aimed at promoting improvement and creative thinking.

2. It is not part of the six-sigma methodology, but will typically take part in its implementation.

What do the stages of DMAIC mean [3]?

Define the problem, improvement activity, opportunity for improvement, the project goals, and customer (internal and external) requirements.

Measure process performance.

Analyze the process to determine root causes of variation, poor performance (defects).

Improve process performance by addressing and eliminating the root causes.

Control the improved process and future process performance.

In the following chapter we give a set of statistical tools required for *measuring*, *analyzing* and *controlling* a process.

---

[3]`http://asq.org/learn-about-quality/six-sigma/overview/dmaic.html`

Chapter 2

Exploratory Data Analysis-EDA

In this chapter, we give a short review of methods for *exploratory data analysis* (EDA), a.k.a. *descriptive statistics*. Recall that our goal is an assumptions-free description of our data. EDA thus consist of computing interpretable summaries of the data, called *summary statistics*, and visualizations.

Descriptive Statistics

2.1 Summary Statistics

We now distinguish between summary statistics that apply to attributes, categorical by definition, and variables, continuous by definition.

2.1.1 Summarizing Categorical Data

Univariate

Summarizing a vector of categorical data can naturally be done by tabulating it, i.e., computing the frequency and relative frequency of each category. Clearly averages, medians, and the likes are incomputable, since categorical data has no ordering, nor does it admit simple operations such as summation.

Extra Info 2. Variability of categorical data can clearly not be measured by its variance, since it does not admit a summation operation. It is, however, possible to define different measures of variability that do apply. The *entropy* is such an example.

Entropy

Bivariate

Generalizing the univariate case to bivariate, or multivariate, one can keep tabulating. I.e., compute the frequency, and relative frequency, of combinations of categories.

2.1.2 Summarizing Continuous Data

Continuous variables admit many more mathematical manipulations than categorical attributes.

Univariate

We start by presenting the most natural summaries of the data. Without going into the formal definition, we refer to them as *summary of location.* These include:

Location Summaries

Definition 1 (The Mean)**.** The *mean*, or *average*, is defined as

$$\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_i \tag{2.1}$$

Definition 2 (The Median)**.** The median is the observation that is smaller than half of the sample and larger than half of the sample.

Definition 3 (α-Trimmed Mean)**.** The α-trimmed mean is the average of the observations left after ignoring the largest and the smallest $(100\alpha)\%$ of them.

The naïve average is the 0-trimmed mean, and the median is the 0.5-trimmed mean.

From summaries of location, we move to summaries of *scale.*

Summary of Scale

Definition 4 (The Standard Deviation)**.**

$$s(x) := \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{2.2}$$

For the following, we require the definition of the sample quantiles, themselves **not** a scale summary.

Definition 5 (α Quantile)**.** The α-quantile of the sample is the observation that is larger than $(100\alpha)\%$, and smaller then $(100(1 - \alpha))\%$ of the sample.

The empirical maximum and minimum are then $x_{1.0}$ and $x_{0.0}$, respectively.

Definition 6 (The Range).

$$Range(x) := \max_i\{x_i\} - \min_i\{x_i\} = x_{1.0} - x_{0.0} \tag{2.3}$$

Definition 7 (The Inter Quantile Range- IQR).

$$IQR(x) := x_{0.75} - x_{0.25} \tag{2.4}$$

Definition 8 (The Median Absolute Deviation- MAD).

$$MAD(x) := \{|x_i - x_{0.5}|\}_{0.5} \tag{2.5}$$

Note that the MAD may be sometimes scaled by 1.4826, so that it estimates σ. Such is the behaviour of the **R** function `mad()`.

After summaries of scale, we move to summaries of *skewness*, or *asymmetry*.

Definition 9 (Yule Skewness Measure).

$$YULE(x) := \frac{\frac{1}{2}(x_{0.75} + x_{0.25}) - x_{0.5}}{IQR(x)} \tag{2.6}$$

Bivariate

From univariate data x, we move to bivariate x, y. Clearly we can apply univariate summaries component-wise. We want, however, to summarize the *joint* behaviour of the data. For this purpose, we assume that data comes in pairs, implying that x and y are of same length.

Definition 10 (Covariance). The sample covariance, or *empirical* covariance is defined as

$$Cov(x, y) := \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n - 1} \tag{2.7}$$

Definition 11 (Pearson's Correlation Coefficient). *Pearson's Correlation Coefficient*, or *Pearson's Moment Product Correlation Coefficient*, is defined as

$$r(x, y) := \frac{(n-1)Cov(x, y)}{S(x)S(y)} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{S(x)S(y)} \tag{2.8}$$

We can dwell into the meaning and intuition underlying Pearson's correlation coefficient, but we will not. The curious reader is reffered to Rodgers and Nicewander (1988).

The next measure of association captures a more general association.

Definition 12 (Spearman's Correlation Coefficient)**.** Spearman's correlation coefficient is merely Pearson's correlation coefficient computed on the *ranks* of x and y.

We conclude by noting that *regression coefficients* are also a measure of association.

Multivariate Data

Multivariate data, both continuous (variables), and discrete (attributes), admits a vast realm of method for summary and visualization. Clearly, associations between several variables can be very complicated so that the more we try to summarize, the more information we give up. On the other hand, and unlike the univariate and bivariate case, our minds will need some type of simplification since they cannot grasp the raw data (did you ever try to imagine how $\mathbb{R}^4$ looks like?). As usual, we emphasize that our purpose is to summarize the joint association in the data. For component-wise summaries, we can always apply the univariate summaries one variable at a time.

By far the most popular measures of joint association are the covariance matrix and correlation matrix.

Definition 13 (Covariance Matrix)**.** For multivariate data consisting of $x_1, \ldots, x_j, \ldots, x_p$ vectors, each with n entries: $x_{j,1}, \ldots, x_{j,n}$, we define the (sample) covariance matrix to be a $p \times p$ matrix whose elements are the (sample) covariances between corresponding vectors:

$$\hat{\Sigma}_{k,l} := Cov(x_k, x_l). \tag{2.9}$$

Extra Info 3 (Sample Covariance Matrix)**.** The matrix $\hat{\Sigma}$ has many useful properties. The curious reader is referred to Petersen and Pedersen (2006), and references therein, for more details.

Definition 14 (Correlation Matrix)**.** For multivariate data consisting of $x_1, \ldots, x_j, \ldots, x_p$ vectors, each with n entries: $x_{j,1}, \ldots, x_{j,n}$, we define the (sample) correlation matrix to be a $p \times p$ matrix whose elements are the (Pearson) correlations between corresponding vectors:

$$\hat{R}_{k,l} := r(x_k, x_l) \tag{2.10}$$

Extra Info 4 (Multivariate Data Analysis)**.** Multivariate analysis is an important, and very actively studied field in statistics and machine learning. A non-comprehensive list of methods that belong to this realm include Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Factor Analysis (FA), Independent Component Analysis (ICA), Dimensionality Reduction, Manifold Learning, Self Organizing Maps, etc. Ask me for reference books or courses if this topic interests you.

PCA, SVD,ICA

2.2 Visualization

2.2.1 Visualizing Categorical Data

Univariate

Much like computing summaries, there is not much to be said about visualizing univariate categorical variables. The most natural, and perhaps only visualization, is the *bar plot*, illustrated in Figure 2.1.

Remark 1 (Pie Chart)**.** About those pie charts. There is really no reason to use them. Ever[1].

Figure 2.1: The Bar-Plot.
`http://www.r-tutor.com/elementary-statistics/qualitative-data/`
`bar-graph`

---

[1]`http://www.businessinsider.com/pie-charts-are-the-worst-2013-6`.
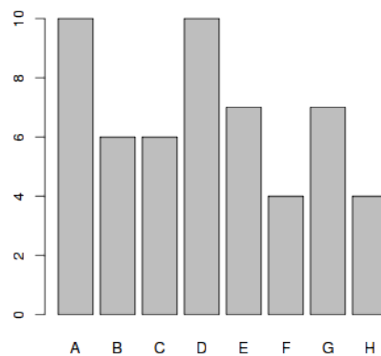
Bivariate

Visualizing a two-way cross-table can be done using an extension of the bar-plot. Several extensions exist. By far, the most informative and recommended figure, in this author's view, is the *mosaic plot*, illustrated in Figure 2.2.

Mosaic
Plot

Figure 2.2: Mosaic Plot.
http://www.statmethods.net/advgraphs/mosaic.html

2.2.2   Visualizing Continuous Data

Univariate

Visualization of univariate continuous vectors can present the raw data, or it distribution (i.e.- discarding the indexes). The most basic visualizations are the *dotchart*, *histogram*, *boxplot*, *stem-and-leaf plot*. These are illustrated in figures 2.3, 2.4, 2.5, 2.6 respectively.

Bivariate

The simultaneous visualization of two continuous variables, can naturally be done with a *scatter plot*. More sophisticated visualization, which generalizes the histogram into two dimensions, is the *hexbin plot*. These are illustrated in figures 2.7, and 2.8, respectively.

Figure 2.3: Dot Plot.
http://stackoverflow.com/questions/15109822/
r-creating-scatter-plot-from-data-frame

Figure 2.4: Histogram. Notice the ticks on the x axis. These are the raw data points. Make sure you always add them, with the **rug()** **R** command.
http://compbio.pbworks.com/w/page/16252882/Basic

Figure 2.5: Boxplot.
http://support.sas.com/documentation/cdl/en/statug/63033/HTML/
default/viewer.htm

Figure 2.6: Stem-and-leaf plot.
https://www.mathsisfun.com/data/stem-leaf-plots.html

Figure 2.7: Scatter Plot.
http://texample.net/tikz/examples/scatterplot/

Figure 2.8: HexBin plot (a 2D histogram).
http://www.r-bloggers.com/5-ways-to-do-2d-histograms-in-r/

Multivariate Data

Since we cannot possibly visualize data in more than 3-dimensions, and we clearly prefer data in 1 or 2 dimensions, the visualization of multivariate data will typically consist of summarizing the data into $1D$ or $2D$, and then applying the above mentioned visualization techniques.

An important exception is due to the observation that a computer image, is essentially a matrix. We can thus visualize matrices, with a simple image, and in particular, covariance and correlation matrices, as illustrated in Figure 2.9.

A second exception is when the data has both continous variables and discrete attributes. Endlessly many combinations are then possible. The author strongly recommends to visit Hans Rosling's *Gap Minder* at `http://www.gapminder.org/world` for an excellent interactive visualization.

Gap
Minder

Figure 2.9: Image of covariance matrix.
`http://cs.brown.edu/courses/csci1950-g/results/final/sghosh/`

2.2.3   On-Line Visualization

For the purpose of quality control, we may often want an *on-line* visualization, and not *off-line*, as the ones previously discussed. This is the purpose of *dashboards*, illustrated in Figure 2.10.

Dash-
board

29

Figure 2.10: Dashboard.
`http://www.iconics.com/Home/Products/AnalytiX/Quality-AnalytiX.aspx`

Chapter 3

Statistical Inference

The idea of extrapolating knowledge from a *sample* to a population is known as *statistical inference*. It encompasses the ideas of *parameter estimation*, *confidence intervals*, and *hypothesis testing*. We will assume the reader is familiar with these, but recall some required terminology. The QC and SPC terminology are not always consistent with statisticians' terminology. When new names are given to old ideas, we will emphasize this in the text.

Null/Alternative Hypothesis Some statement about the world we wish to test with data. The frequentist argument follows a Popperian philosophy: to show the alternative hypothesis is true, we will show that the null hypothesis is not true. In the context of quality control, the null hypothesis will be the process is *in statistical control*, while the alternative will be that it is *out of control*.

In Statistical Control

Statistical Test The procedure of inferring from data on the truthfulness of the alternative hypothesis.

Assumptions As the name suggests, these are assumptions. We stress that unlike hypothesis, assumptions are not being tested in a statistical test.

Test Statistic The function of the data to be computed for the purpose of inference. As such, it is a random variable.

Null/Alternative Distribution The distribution of the test statistic under the null/alternative hypothesis.

Type I/II error See Figure 3.1.

False/True Positive/Negative See Figure 3.1.

	Null hypothesis (H_0) is true	Null hypothesis (H_0) is false
Reject null hypothesis	Type I error **False positive**	Correct outcome True positive
Fail to reject null hypothesis	Correct outcome True negative	Type II error **False negative**

Figure 3.1: Type I/II error confusion table.
`https://infocus.emc.com/william_schmarzo/`
`beware-of-false-positives/`

Rejection Region The collection of event that will lead us to reject the null hypothesis, and believe in the alternative hypothesis.

p-value A.k.a. *observed significance.* The null probability of the observed (or "more extreme") event.

Significance Level A.k.a. α. The probability of a false positive.

Power The probability of a true positive.

i.i.d. "Independent and identically distributed" (i.i.d.) is an assumption made on the sampling distribution, meaning that samples are statistically independent, and all originating from the same distribution.

The following sections of this chapter present particular statistical inference methods we will be using in the following chapters.

3.1 Goodness of Fit- GOF

Goodness of fit (GOF) deals with the inference on the sampling distribution, a.k.a., the generative process. It can be approached via rigorous hypothesis testing, or by visualizations.

3.1.1 QQplot and QQnorm

The fundamental idea of the *quantile-quantile plot* (QQplot) is to compare the empirical quantiles in the sample, to the theoretical quantiles implied by the assumed distribution. If the theory and observations agree, we conclude our assumptions are plausible. For the particular case of testing the normality of the data, the corresponding QQplot is known as a *QQnorm plot*.

Figure 3.2 illustrates a QQnorm plot of normal distributed data, while Figure 3.3 is the same for non-normal data.

Figure 3.2: A QQplot of Gaussian distributed data.

Figure 3.3: A QQnorm plot of non Gaussian distributed data.

3.1.2 Chi-Square GOF Test

The Chi-Square GOF test (not to be confused with the Chi-Square independence test), tests a hypothesis on the sampling distribution of discrete (attributes) data. Note that it is very general, since all continuous variables may be discretized, simply by binning.

Definition 15 (Chi-Square GOF Test). Assume an i.i.d. sample $x_1, \ldots, x_n$. The Chi-Square GOF is a tests $H_0 : \mathbf{x}_i \sim P$ versus $H_1 : \mathbf{x}_i \nsim P$. P is assumed to be discrete with K categories and $p_k := P(\mathbf{x}_i \in k)$. The test statistic, X^2, is defined as

$$X^2 := \sum_{k=1}^{K} \frac{(obs_k - exp_k)^2}{exp_k}, \tag{3.1}$$

where $obs_k := \#\{x_i \in k\}$, and $exp_k := p_k n$. The approximate null distribution of X^2 is χ^2_{K-1}.

3.1.3 Kolmogorov–Smirnov GOF Test

The Kolmogorov-Smirnof GOF test, tests a hypothesis on the sampling distribution of continuous (variable) data.

Definition 16 (Kolmogorov–Smirnov GOF Test). Assume an i.i.d. sample $x_1, \ldots, x_n$. The Chi-Square GOF is a tests $H_0 : \mathbf{x}_i \sim P$ versus $H_1 : \mathbf{x}_i \nsim P$. P is assumed to be continuous. The test statistic, D, is depicted in Figure 3.4.
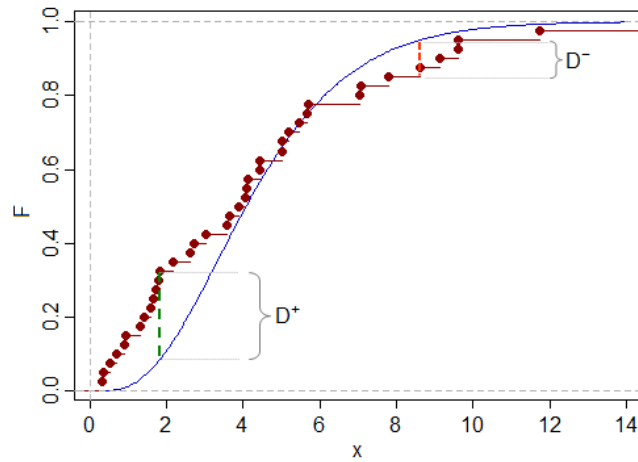
Figure 3.4: Kolmogorov-Smirnov Test. The D statistic is D_+ in the figure.

The null distribution of D is the *Kolmogorov distribution* obtained from tables.

Kol-
mogorov
Distribu-
tion

Extra Info 5 (GOF tests)**.** There are endlessly many more GOF tests, such as Anderson-Darling, Jarque–Bera, Shapiro–Wilk, Kuiper's test, etc. Wikipedia is a good place for further reading.

Chapter 4

System Capability Analysis

In a *system capability analysis*, we essentially use statistical tools to measure the variability in a production process. This analysis can answer questions that are raised at the measuring, analyzing, and improving stages of the DMAIC cycle. The methods we will discus compare between the processes variability and its specifications. Note, however, that a simple statistical analysis of the production process, without relating to specification, may also qualify as a system capability analysis.

We naturally want production processes that adhere to specifications, and want to quantify the level of adherence. The quantification is performed by comparing the variability in a CTQ to the specification. In this chapter we assume the process's capability is fixed over time. In Chapter 9 we revisit the same problems, when allowing the process's capability to vary over time.

The particular setup discussed in this chapter is also known as *product specification*. When we use the actual time, and ordering of data samples, as in a control chart, we will no longer regard it as product specification but rather as a bona-fide capability analysis. This is the subject of Chapter 9.

Product Specification

Because capability analysis, or product specification, is essentially the study of the CTQ's distribution, it can be approached with the aforementioned statistical tools such as univariate summary statistics and visualizations presented in Sections 2.1 and 2.2. To test a particular hypothesis want to be tested on the distribution of the CTQ, we may call upon the inference tools from Chapter 3.

4.1 Process Capacity Indexes

Classical statistical devices do not incorporate the designed process's capabilities. *Process capability ratios* (PCR), or *process capability indexes*, are

merely population parameters that also depend on specifications. The first, Process Capability Index and most basic PCR is the C_p of a particular CTQ.

Definition 17 (C_p).

$$C_p := \frac{USL - LSL}{6\sigma}, \tag{4.1}$$

where σ is the standard deviation of the CTQ. Eq.(4.1) readily offers an interpretation of the C_p: it measures how accurate is our process compared to a 3-sigma process: $C_p = 1$ for 3-sigma, $C_p = 2$ for 6-sigma, etc.

Clearly C_p is a process parameter, that needs to be estimated.

Definition 18 ($\hat{C}_p$).

$$\hat{C}_p := \frac{USL - LSL}{6\hat{\sigma}}, \tag{4.2}$$

where $\hat{\sigma}$ is some estimate of the standard deviation of the CTQ.

The most natural $\hat{\sigma}$ is the sample standard deviation s, but we will explore other options in the following.

There is a relation between C_p and the probability of non-conformance. To explore this relation we introduce the following notation:

> **Collecting Notation**
> $T := (USL + LSL)/2$, the target value.
> $\delta := (USL - LSL)/2$, the specification tolerance. $USL = T + \delta, LSL = T - \delta$.
> $\mu := \mathbf{E}[CTQ]$, the expected CTQ.
> $p_{NC} := 1 - P(CTQ \in [LSL, USL])$, the probability of non compliance.

With our new notation Eq.(4.1) is now $C_p = \frac{\delta}{3\sigma}$. Assuming $CTQ \sim \mathcal{N}(\mu, \sigma^2)$, and that the process is centred so that $\mu = T$, then C_p is related to p_{NC} via

$$p_{NC} = 2\,\Phi(-3C_p) \tag{4.3}$$

As a sanity check, we check this relation for a 3-sigma process. A 3-sigma process implies that $C_p = 1$, and Eq.(4.3) returns $p_{NC} = 0.0027$, as we have already seen in the introduction (Section 1.4.5).

Montgomery (2007) recommends the following C_p values:

	C_p Value	Implied ppm
Existing processes	1.33	66
New processes	1.50	6.8
Safety, strength, or critical parameter, existing process	1.50	6.8
Safety, strength, or critical parameter, new process	1.67	0.5
Six Sigma quality process	2.00	0.002

To derive Eq.(4.3), and thus the ppm column in the table, we had to call upon several assumptions. Namely:

1. The process is in statistical control, i.e. μ is fixed over time

2. The CTQ has a normal distribution.

3. The process is centred, i.e. $\mu = T$.

4. T is mid-way between LSL and USL.

4.1.1 Non-Conformance for a Non-Gaussian CTQ

The first assumption we will now relax is the Gaussianity of CTQ. We start by checking what is the fallout rate, if we were completely wrong about the distribution of the CTQ. Chebyshev's inequality provides a universal bound on p_{NC} for a $C_p = 1$ process.

Theorem 4.1.1 (Chebyshev's inequality). *For any random variable* $\mathbf{x}$*, with* $\mu := \mathbf{E}[\mathbf{x}]$*, and* $\sigma^2 := \mathbf{E}[(\mathbf{x} - \mu)^2]$*, then*

$$P(|\mathbf{x} - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{4.4}$$

If $C_p = 1$ then $\delta = 3\sigma$. Plugging $k = 3$ in the inequality returns $p_{NC} < 0.11$. This means that a 3-sigma process, assumingly with $2,700ppm$, may actually have $111,111ppm$, if we were wrong about the Gaussianity assumption.

The moral of the story is that by assuming the correct distribution of the CTQ, we may save a lot of resources. We can assume normality, as we typically do, but there are other alternatives:

1. Transformations: it is quite possible that the CTQ is not Gaussian in its original scale, but it is Gaussian in a different scale. You should always inspect a QQnorm (Sec. 3.1.1) plot after a *log* or *sqrt* transformation.

2. Assume a different distribution: We derived $p_{NC}(C_p)$ (eq. 4.3) under a normality assumption, but it may certainly be derived for different distributions.

3. The denominator of C_p is a range that leaves 0.00135 probability of the Gaussian tail outside the range. When relaxing the normality assumption, σ is no longer related to the tail probability as it was before. We may still, however, directly plug $CTQ_{0.00135}$ and $CTQ_{1-0.00135}$ quantiles to get a CPR in the same spirit of the C_p. This is known as the $C_p(q)$ index we now define.

Figure 4.1: C_{pk} and C_p.
`https://www.spcforexcel.com/knowledge/process-capability/`
`interactive-look-process-capability`.

Definition 19 ($C_p(q)$)**.**

$$C_p(q) := \frac{USL - LSL}{CTQ_{1-0.00135} - CTQ_{0.00135}} \tag{4.5}$$

4.1.2 Process Capability of a Non-Centred Process

We will now relax the assumption of $\mu = T$, while still in statistical control.

Definition 20 (C_{pk})**.**

$$C_{pk} := \min\{C_{pu}, C_{pl}\} \tag{4.6}$$

where $C_{pu} := \frac{USL - \mu}{3\sigma}$ and $C_{pl} := \frac{\mu - LSL}{3\sigma}$.

For a non-centred process, this definition is more informative on the probability of non-compliance. Indeed, Eq.(4.3) will not hold, but we can derive an updated version:

$$p_{NC} \approx \Phi(-3C_{pk}). \tag{4.7}$$

Generally, $C_{pk} \leq C_p$, with equality holding for centred processes ($\mu = T$). An illustration is given in Figure 4.1.

The C_{pk} index is motivated by conserving the relation between the index and the fallout rate p_{NC}, which is not captured by C_p when the process is not

centred. Indeed if C_p can be interpreted as "accuracy compared to a (centred) 3-sigma process", then C_{pk} can be interpreted as "accuracy compared to a (equally non-centred) 3-sigma process". The 3-sigma process is used as a benchmark for historical reasons. In practice, it is actually very common to report the sigma-level as a capability index. This implies the 1-sigma process as a benchmark: $\min\{\frac{USL-\mu}{\sigma}, \frac{\mu-LSL}{\sigma}\}$. Three-sigma would thus be reported as 3, 6-sigma as 6, etc.

Another index, known as C_{pm}, or the *Taguchi capability index*, also deals with the non centring but slightly differently. It is motivated by the observation that for $\mu = T$, then $\sigma = \sqrt{\mathbf{E}\left[(CTQ - T)^2\right]}$. This relation does not hold if $\mu \neq T$, which leads us to the following definition:

Definition 21 (C_{pm}).

$$C_{pm} := \frac{USL - LSL}{6\sqrt{\mathbf{E}\left[(CTQ - T)^2\right]}} \tag{4.8}$$

$$= \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}} \tag{4.9}$$

$$= \frac{C_p}{\sqrt{1 + (\frac{\mu-T}{\sigma})^2}}. \tag{4.10}$$

Eq.(4.10) readily shows that just like the C_{pk}, then $C_{pm} \leq C_p$.

4.1.3 Interval Estimation for Capability Indexes

Since the various process capability indexes are merely population parameters, we can also construct confidence intervals (CIs) for them, which may be very informative if sample sizes are small. The simplest case is that of C_p. Being a monotone transformation of σ, we can call upon confidence intervals for the variance of a normal population, so that with probability $1 - \alpha$:

$$C_p \in \left[\hat{C}_p\sqrt{\frac{\chi^2_{\alpha/2,n-1}}{n-1}}, \hat{C}_p\sqrt{\frac{\chi^2_{1-\alpha/2,n-1}}{n-1}}\right]. \tag{4.11}$$

This equation is simply derived from Eq.(4.12). Intervals for the other capability indexes, are available in Montgomery (2007) and references therein.

4.1.4 Testing Hypotheses on Capability

Consider a supply contract, which requires production to have $C_p > 1.5$. It may easily be the case, that the $C_p > 1.5$, even if $\hat{C}_p < 1.5$, especially if the

sample size is small. It thus makes a lot of sense, to design hypothesis tests on process capabilities. We observe that for an i.i.d. sample from a Gaussian population, where $\hat{C}_p$ is estimated with s, then

$$(n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2 \sim \chi^2_{n-1} \tag{4.12}$$

so that $(n-1)\left(\frac{C_p}{\hat{C}_p}\right)^2$ may serve as test statistic.

Example 1 (C_p test for 6-sigma compliance)**.**

$$H_0 : C_p \leq 2$$
$$H_1 : C_p > 2$$
$$(n-1)\left(\frac{2}{\hat{C}_p}\right)^2 \overset{H_0}{\sim} \chi^2_{n-1}$$

so that the $1-\alpha$ rejection region in $\hat{C}_p$ scale is

$$\hat{C}_p > \sqrt{\frac{4(n-1)}{\chi^2_{n-1,\alpha}}}.$$

Note that we should not be testing this hypothesis with the confidence interval in Eq.(4.11) because this particular hypothesis is directional. Now for a more general case:

$$H_0 : C_p \leq a; H_1 : C_p > a \Rightarrow \text{reject if } \hat{C}_p > \sqrt{\frac{a^2(n-1)}{\chi^2_{n-1,\alpha}}},$$

$$H_0 : C_p \geq a; H_1 : C_p < a \Rightarrow \text{reject if } \hat{C}_p < \sqrt{\frac{a^2(n-1)}{\chi^2_{n-1,1-\alpha}}},$$

$$H_0 : C_p = a; H_1 : C_p \neq a \Rightarrow \text{reject if } \hat{C}_p < \sqrt{\frac{a^2(n-1)}{\chi^2_{n-1,1-\alpha/2}}} \text{ or } \hat{C}_p > \sqrt{\frac{a^2(n-1)}{\chi^2_{n-1,\alpha/2}}}$$

4.1.5 Process Performance Indices

Process *performance* indices measure compliance to specification of a process out of statistical control. These include the P_p and P_{pk} indices. Besides mentioning their existence, we will not give them further attention, since we adopt Montgomery (2007)'s view that their use is strongly discouraged.

Remark 2. At this point, I hope you are wondering why isn't the fallout rate, p_{NC}, not used as a capability index. Well, it is this author's view, that is it probably the best index of them all.

Chapter 5

Statistical Process Control-SPC

Statistical process control (SPC), a.k.a. *change detection*, or *novelty detection*, deals with the quantitative analysis of a "process", which may be a production line, a service, or any other repeated operation. As such, SPC may be found in the Analyze, Improve, and Control stages of the DMAIC cycle. The purpose of the SPC, in the terms coined by Shewart, is to seperate the variability in the process into *assignable* causes of variation and *chance* causes of variation. These are also known as *special* and *common* causes of variation, respectively. A process is said to be in *statistical control* if all its variation is attributable to chance causes. If this is not the case, we call it *out of control* and we will seek the assignable causes, so that we may reduce variability by removing them.

<div style="float:right; font-size:small;">Change Detection</div>

<div style="float:right; font-size:small;">Causes of variation</div>

All the previously mentioned statistical tools may be called upon for this analysis. In the context of process control, a subset of tools has gained the nick-name "The Magnificent Seven". These include:

Histogram and stem-and-leaf plot. As described in Chapter 2.

Check Sheet. [TODO: add figure]

Pareto chart. An ordered bar plot of the events due to the various assignable variability causes. [TODO: add figure]

Cause-and-effect diagram. A visualization of candidate assignable variability causes. [TODO: add figure]

Defect concentration diagram. A visual inspection of the location of defects on the product. [TODO: add figure]

Scatter plot. As described in Chapter 2.

Control chart. A powerful analysis tool to which we devote the rest of this chapter.

Extra Info 6 (A mathematically rigorous treatment of SPC). The contents of this chapter is mostly derived from Montgomery (2007). For a more mathematically rigorous treatment of the topic see Basseville et al. (1993). For an **R** oriented exposition of the topic, see Qiu (2013).

5.1   A soft start. The $\bar{x}$-chart

We demonstrate the concepts and utility of Control Charts with the simplest, yet most popular of them all, the $\bar{x}$-chart . The chart borrows its name from the fact that it is essentially a visualization of the time evolution of the average ($\bar{x}$) of the CTQ of a sample of products. The chart is also augmented with visual aids that help in determining if the process is *in control*, i.e., if it is consistent with its own history. Process capability analysis may benefit from the ideas of control charts, as we explain in Chapter 9. We emphasize, again, that control charts have no information on the specifications of the process, merely on its own history.

An illustration of a $\bar{x}$-chart is given in Figure 5.1. The ingredients of this chart are the centerline, lower and upper control limits (LCL, UCL), and $\bar{x}$ evolving in time. If at each period $t = 1, \ldots, \tau$ we compute the average of n samples, we denote $\bar{x}_t := 1/n \sum_{i=1}^{n} x_{it}$.

Figure 5.1 makes it evident $\bar{x}$-chart requires several design decisions. A standard design decision is setting the center line as the grand average of the process:

$$\hat{\mu}_0 = 1/\tau \sum_{t=1}^{\tau} \bar{x}_t, \tag{5.1}$$

where μ_0 denotes the in-control mean of the process. Notation follows from treating the in-control process as a null hypothesis.

If it is unclear to you, how may we compute the grand average of a process that is still evolving and has not finished, you are right! We thus introduce the ideas of *Phase I* and *Phase II*. Initially we assume the process it out of control, we identify and remove assignable causes of variation, until we are left with a "well-behaved" subset of data points. We call this Phase I, and we use it to initialize required quantities such as the centre line. Eq.(5.1) thus implies that in Phase I we were left with τ samples assumingly in statistical

Phase I/II

Figure 5.1: $\bar{x}$-chart .
https://mvpprograms.com/help/P-mvpstats/spc/
WhatAreControlCharts

control. After the chart has been calibrated, and major assignable sources of variability removed, we can finally start monitoring the process, known as Phase II.

Other design decisions to be made are:

1. UCL and LCL [1].

2. Sample size in each sample (n).

3. Rational groupings (within-period sampling scheme).

4. Frequency of samples (between-period sampling scheme).

5. Other stopping rules.

These design decisions ultimately govern the error rates of the chart, which in turn, incur financial costs. For now we will restrict attention to type I/II error rates, until Section 5.4 where we consider these choices as economical optimization problems.

For ease of exposition, control chart design is demonstrated for the $\bar{x}$-chart , but equally applies to other control charts, presented in Section 5.5. We start by a type I error rate analysis. Denote α_t the false alarm probability

---

[1]Do not confuse with USL and LSL!

at period t. How do our design choices affect α_t?

$$\alpha_t := 1 - P_{H_0}(\bar{x}_t \in [UCL, LCL]) \tag{5.2}$$

$$= 2P_{H_0}(\bar{x}_t < UCL) \tag{5.3}$$

$$= 2P_{H_0}(Z < \frac{UCL - \mu}{\sigma_{\bar{x}}}) \tag{5.4}$$

$$= 2P_{H_0}(Z < -L) \tag{5.5}$$

$$= 2\Phi(-k) \tag{5.6}$$

The above follows from assuming that $UCL := \mu + L\sigma_{\bar{x}}, LCL := \mu - L\sigma_{\bar{x}}$, $\mathbf{x}_{it} \sim \mathcal{N}(\mu, \sigma^2)$, and denoting $\sigma_{\bar{x}} := \frac{\sigma}{n}$. A typical design choice is $L = 3$, known as *3-sigma control limits*, implying a false alarm rate of $\alpha_t = 0.0027$. Since we assumed the process is fixed over time, then so is α_t and we can simply write $\alpha_t = \alpha$. 3-Sigma Control Limits

Remark 3 (3-Sigma Control Limits vs. 3-Sigma Capability)**.** Do not confuse these two similar ideas. 3-Sigma Control Limits is a statement on the false alarm rate of a process with respect to its own **history**. 3-Sigma Capability is a statement on the non-compliance rate of a process with respect to its **specification**.

A power analysis for our design choices follows the same lines. Denote β_t, and $\pi_t = 1 - \beta_t$ the type II error rate, and power, at period t. We then have

$$\pi_t := 1 - P_{H_1}(\bar{x}_t \in [UCL, LCL]) \tag{5.7}$$

and the rest follow from the distribution of $\bar{x}$ when the process is out of control. Since the out-of-control shift is stable, we can again omit the time index and write $\pi = \pi_t$. Assuming the out-of-control process is a shift of magnitude $k\sigma$, i.e.: $\mathbf{x} \sim \mathcal{N}(\mu_1, \sigma^2); \mu_1 = \mu_0 + k\sigma$, we plot in Figure 5.2, the detection power of a 3-sigma $\bar{x}$-chart , as a function of k. This is known in the statistical literature as a *power function*, and in the engineering literature as the *true positive rate* operator characteristic (TPR-OR). Operator Characteristic

Extra Info 7 (Operator Characteristics)**.** Many operator characteristics have been proposed to study the performance of control charts, statistical tests, or binary classifiers in general. You may be already familiar with some, such as the ROC-curve. The curious reader is referred to Wikipedia (2015e) for more information.

Another important related quantity is the *average run length* (ARL), which is the expected number of periods between two crossings of control

Figure 5.2: Power function of the 3-sigma $\bar{x}$-chart with $n = 5, 10, 20$ and $\mu_1 = \mu_0 + k\sigma$.

limits, i.e., the expected periods between alarms. We denote by ARL_0 the ARL when the process is under statistical control, and ARL_1 otherwise[2]. If $\bar{x}_t$ are statistically independent, then clearly the number of periods until a crossing is geometrically distributed. Using the expectation of a geometric random variable we can conclude that

$$ARL_0 = 1/\alpha, \tag{5.8}$$

$$ARL_1 = 1/\pi. \tag{5.9}$$

Clearly we can convert to time units by multiplying the ARL by the duration of sampling interval. This is known as the *average time to signal* (ATS). It is quite common to design a control chart so that it achieves a particular ATS_0.

Now assume that we are unhappy with some control chart. It simply makes too many false alarms, or takes too long to detect loss of statistical control. What can we do about it? Well, this is exactly the same question as when increasing the power or lowering the type I error of a statistical hypothesis test. This is obviously no coincidence, since control charts are nothing but a statistical test! Here are some action courses:

1. Increase L. This is the same as shrinking the rejection region: it will decrease the false alarm rate, at the cost of power.

---

[2]Note that it is implied that the process has a *stable* distribution, even though it is out if control.

46

2. Increase n. Brilliant! Statistically, there is nothing to lose. It may, however, cost time and money.

3. Increase the sampling frequency. Brilliant again! Nothing to lose, except time and money...

4. Change the sampling scheme within period. We elaborate on this in Section 5.1.2.

5. Add other stopping rules: this acts just like growing the rejection region. It will increase power, at the cost of type I error. We elaborate in Section 5.1.3.

6. Pool together more time points or more CTQs. We elaborate on this in sections 5.2 and 5.3, respectively.

5.1.1 Control Limits and the Alarm Rate

As previously discussed, L governs the tradeoff between type I and type II errors, or sensitivity versus specificity. It is very common to set $L = 3$. For a Normally distributed CTQ, this implies $2,700$ false alarms per million periods. This also implies an ARL_0 of $1/\alpha \approx 370$ periods. We may obviously, discard this $L = 3$ convention, and directly set UCL and LCL so they guarantee some desirable false alarm rate.

If normality of $\bar{x}_t$ can be assumed, then one may estimate σ from phase I, and set LCL and UCL by finding the L that solves $2\Phi(-L) = \alpha$. If normality cannot be assumed, there are many ways to go about. Here are some options:

1. Increase n: even if $\mathbf{x}_i t$ is non normal, for large enough n, then $\bar{x}_t$ will be via the central limit theorem (CLT).

2. Use empirical quantiles: If phase I was returned enough data, then we may estimate $\mathbf{x}_{\alpha/2}$ and $\mathbf{x}_{1-\alpha/2}$ using the empirical quantiles of phase I. The false alarm rate will be α since $P(\mathbf{x} \notin [\mathbf{x}_{\alpha/2}, \mathbf{x}_{1-\alpha/2}]) = \alpha$.

3. If some other distribution can be assumed then we may compute the false alarm rate of particular limits either analytically, or computationally (by simulation).

5.1.2 Rational Groupings

Recall that at each period we compute the average of n samples. How should we draw these samples? All at the same time from the same machine? At

different times from the same machine? Many configurations are possible, and the correct approach depends on the type of out-of-control behaviour one seeks. *Rational groupings* merely reminds us to sample "rationally" in each period. Quoting Montgomery (2007)'s words of caution:

> ...we can often make any process appear to be in statistical control just by stretching out the interval between observations in the sample.

5.1.3 Other Stopping Rules

The assumption that we may only create alarms if $\bar{x}$ exceeds some control limits is needlessly restrictive. A first relaxation is by allowing multiple regions. It is quite common to define *warning limits* which only call for inspection, and *action limits*. Each may have its own alarm rate. We may even change the sapling scheme if limits are breached. Increasing the sampling rate once the warning limits have been breached is known as *adaptive sampling*, or *variable sampling*.

Adaptive Sampling

Another approach is to define multiple sets of stopping rules. Here is an example:

1. One or more points outside of the 3-sigma control limits.

2. Two of three consecutive points outside the 2-sigma warning limits but still inside the 3-sigma control limits.

3. Four of five consecutive points beyond the 1-sigma limits.

4. A run of 8 consecutive points on one side of the center line.

The above set of rules is known as the Western Electric Rules, a.k.a. the *WECO* rules. Augmenting the set of rules is the same as increasing a rejection region. It adds more sensitivity, at the cost of false alarms. If the rules are properly selected, the gain in sensitivity is worth the increase in false alarms.

WECO

As a quick exercise, we may compute α for m independent rules, each with α^* type I error itself:

$$\alpha = 1 - (1 - \alpha^*)^m. \tag{5.10}$$

Having 4 rules, like WECO, each at $\alpha^*0.0027$ implies that we actually have $\alpha = 0.01$.

Extra Info 8 (Stopping Rules)**.** There are many sets of stopping rules. These include WECO, Nelson, AIAG, Juran, Hughes, Duncan, Gitlow, Westgard, and more. See `http://www.quinn-curtis.com/spcnamedrulesets.htm` for a quick introduction.

5.2 Pooling Information Over Periods

Assume an out-of-control process is simply a mild shift of the controlled-process. This shift may be hard to detect in Shewart chart, especially if n is not too large (as seen in Figure 5.2). If the shift persists, we may gain power, i.e., sensitivity, by pooling several periods together. We now present several ways to pool information from history. These are typically applied in Phase II, where out-of-control processes are expected to have only mild shifts, and not major ones as in Phase I.

Remark 4 (No longer Shewart)**.** The name *Shewart control chart* is reserved to charts plotting one period at a time. When several periods are pooled together, we will no longer call this "Shewart".

Remark 5 (One observation at a time)**.** The following charts have a continuous flavour. As such, it is both favourable, and common, to compute them using one observation at a time, and not by grouping, as is typically for Shewart charts.

5.2.1 Run Test Chart

[TODO]

5.2.2 Moving Average Chart- MA

One way to pool information from different periods is by a *moving average.*

Definition 22 (Moving Average- MA)**.** The moving average at period t is defined as

$$M_t := \frac{x_t + \cdots + x_{t-w+1}}{w} \tag{5.11}$$

Assuming $x_t \sim \mathcal{N}\left(\mu_0, \sigma_x^2\right)$ then clearly

$$M_t \sim \mathcal{N}\left(\mu_0, \frac{\sigma_x^2}{w}\right). \tag{5.12}$$

The control limits on M_t are typically

$$UCL := \mu_0 + 3\sigma_{M_t} = \mu_0 + 3\,\frac{\sigma_x}{\sqrt{w}} \tag{5.13}$$

$$UCL := \mu_0 - 3\sigma_{M_t} = \mu_0 - 3\,\frac{\sigma_x}{\sqrt{w}}. \tag{5.14}$$

The fallout rate of this criterion is trivially $\alpha = 0.0027$. The ARL_0 is no longer simple to compute. This is because the pooling of periods has compromised independence between periods, and Eqs.(5.8,5.9) are no longer valid. Do not despair as the ARL may still be computed. You can always use simulation to compute it, or try using the spc **R** package.

We are free to choose the magnitude of w. If w is too small, there is no real pooling from history. At the limit, where $w = 1$, we are back to the classical Shewart chart. If w is too large, then each new observation has very small importance, and it may take a long time to detect a shift. Which is the right intermediate value of w, is left for you to decide.

5.2.3   Exponentially Weighted Moving Average Chart- EWMA

The moving average gives all observations the same importance. We want to change this, giving more importance to new observations so that we may capture drifts quickly when they occur. The *Exponentially Weighted Moving Average* (EWMA), a.k.a. the *geometric moving average* (GMA), does just that.

GMA

Definition 23 (Exponentially Weighted Moving Average- EWMA)**.**

$$z_t := \lambda x_t + (1 - \lambda)z_{t-1} \tag{5.15}$$

By recursive substitution, we have

$$z_t := \lambda \sum_{j=0}^{t-1} (1 - \lambda)^j x_{t-j} + (1 - \lambda)^t z_0, \tag{5.16}$$

and

$$z_t \sim \mathcal{N}\left(\mu_0, \sigma_{z_t}^2\right), \tag{5.17}$$

$$\sigma_{z_t}^2 = \sigma_x^2 \left(\frac{\lambda}{2 - \lambda}\right)(1 - (1 - \lambda)^{2t}). \tag{5.18}$$

Eq.(5.17) may be used to construct control limits for EWMA: It is however, more economic to observe that for large λ and t: $(1 - (1-\lambda)^{2t}) \approx 1$ so that we may use

$$
\begin{aligned}
UCL &:= \mu_0 + 3\sigma_{z_t} \approx \mu_0 + L \sqrt{\sigma_x^2 \left(\frac{\lambda}{2-\lambda} \right)}, \\
UCL &:= \mu_0 - 3\sigma_{z_t} \approx \mu_0 - L \sqrt{\sigma_x^2 \left(\frac{\lambda}{2-\lambda} \right)},
\end{aligned}
\tag{5.19}
$$

with $L = 3$ being the typical choice. By now, you should immediately know what is the fallout rate of these limits. By now, you should also know that because of the dependence between z_t's, computing the ARL is not as simple as for Shewart charts. The `xewma.arl()` **R** function, in package `spc`, permits doing so easily. Its output for various λ and L is illustrated in Figure 5.3.

Figure 5.3: ARL_0 for EWMA.
Code   from   `http://users.phhp.ufl.edu/pqiu/research/book/spc/`
`r-codes/fig53.r`

Note that if we do not opt for the approximation in Eq.(5.19), and use the exact Eq.(5.17) then the alarm limits change over time.

In the MA chart, we used the choice of w to balance between quick response (small w) and sensitivity (large w). EWMA has no window-width parameter, since it looks into all of history. On the other hand, we can control it by choosing λ. Large λ gives more importance to the present. At the limit, $\lambda = 1$, EWMA collapses to a standard Shewart chart.

5.2.4   Filtered Derivative Chart

[TODO]

5.2.5   CUSUM Chart

The *cumulative sum* chart is similar to the EWMA in that it pools information from the history. The CUSUM simply sums all past deviations from the centre line. If the process is in control, deviation will cancel each other, and their sum will vary around 0. If the process is ot of control, a drift will appear. The statistic to be plotted is

$$C_t := \sum_{j=0}^{t} (x_j - \mu_0) = C_{t-1} + (x_t - \mu_0) \tag{5.20}$$

Observing that when under control then $C_t \sim \mathcal{N}\left(\mu_0, t\sigma_x^2\right)$, we could set

$$\begin{aligned} UCL &:= \mu_0 + 3\sigma_{C_t} = \mu_0 + L\sqrt{t\sigma_x^2}, \\ UCL &:= \mu_0 - 3\sigma_{C_t} = \mu_0 - L\sqrt{t\sigma_x^2}. \end{aligned} \tag{5.21}$$

You may encounter these limits in your favourite software (`qcc` package in **R** ?), but it less often discussed in the literature. This is because CUSUMs were introduced by Page (1954), which offered different control limits. Montgomery (2007) adopts Page's view and presents limits in two forms: the *decision interval* (DI) form, a.k.a. the *tabular* form, and the graphical form known as a *V-mask*. These two control limits are equivalent. Before we present them, we try to offer some intuition for the difference between the limits in Eq.(5.21) and those of Page (1954).

The fundamental difference between the control limits of Page (1954), and the ones presented until now, is that Page designed limits for the particular history of each process, while the limits until now, including Eq.(5.21) do not adapt to the particular history of the process. As such, Page's control limits are said to be *adaptive*.

Adaptive Contol Limits

[TODO: explain V-mask and tabular]

5.2.6   Shiryaev-Roberts Procedure

[TODO]

5.2.7   Combined Shewart and Running Window Charts

Extra Info 9 (Local Methods)**.**

5.3 Multivariate Control Charts

Example 2 (Intensive Care Unit)**.** Consider an intensive care unit. The CTQs are the patient's blood pressure, temperature, etc. We want to sound an alarm if the patient's condition deteriorates. Clearly, we can apply previous the univariate methodology above. It is possible, that the deterioration is mild, so that it is not picked up by any CTQ individually (low power), but could have been noticed were we to aggregate signal from all CTQs.

This is the concern of the current section.

5.3.1 Mass Univariate Control

A first natural approach is to raise an alarm when **any** of the processes exceeds its respective control limits. For m independent processes, with false alarm rate α each, then the joint false alarm rate is

$$\alpha^* = 1 - (1 - \alpha)^m. \tag{5.22}$$

Clearly we could set $\alpha = 1 - \sqrt[m]{1 - \alpha^*}$, so that the joint false alarm rate is under control, but we would not be enjoying the added sensitivity of pooling many CTQs together.

5.3.2 Hotteling's T^2

Hotteling's T^2 statistic is a generalization of the t-test. To see this we write the t-statistic in the following weird form:

$$t(x) = (\bar{x}_t - \mu_0)(s_t^2(x))^{-1}(\bar{x}_t - \mu_0). \tag{5.23}$$

This notation readily extends to the multivariate case. For p CTQs, then $\bar{x}_t$ and μ_0 are p-length vectors, and $s_t^2(x)$ is replaced with the $p \times p$ matrix $\hat{\Sigma}_t(x)$.

Definition 24 (Hotteling's T^2)**.**

$$T_t^2 := n(\bar{x}_t - \hat{\mu}_0)'\hat{\Sigma}_t^{-1}(\bar{x}_t - \hat{\mu}_0). \tag{5.24}$$

To derive the control limits, we will be assuming that $\mathbf{x}_{it}$ is p-variate Gaussian distributed, $\mathcal{N}(\mu_0, \Sigma_{p \times p})$. We need to be very clear, however, on how μ_0 and Σ_t^{-1} are estimated. We thus consider the following scenario:

1. T^2 is computed at Phase II with a single observation at a time ($n = 1$)

2. $\mu_0, \hat{\Sigma}_t^{-1}$ were estimated from m observations in Phase I.

Eq.(5.24) thus yields

$$T_t^2 := (x_t - \hat{\mu}_0)' \hat{\Sigma}_t^{-1} (x_t - \hat{\mu}_0). \tag{5.25}$$

and

$$T_t^2 \sim F_{p,m-p} \frac{p(m-1)(m+1)}{m(m-p)} \overset{m\to\infty}{\rightsquigarrow} \chi_p^2. \tag{5.26}$$

We can thus construct the control limit for this scenario:

$$UCL := F_{1-\alpha,p,m-p} \frac{p(m-1)(m+1)}{m(m-p)} \approx \chi_{1-\alpha,p}^2. \tag{5.27}$$

Note that the T^2 statistic is *non directional*: it will increase in the presence of both positive and negative drift.

We may consider the distribution of T^2 under many configurations: where $\mu_0, \hat{\Sigma}_t^{-1}$ are estimated at Phase II, where we take n observation in each period, etc. For our current purposes, we will use the fact that if $n \gg p$, or $m \gg p$, then

$$T_t^2 \rightsquigarrow \chi_p^2, \tag{5.28}$$

and thus

$$UCL \approx \chi_{1-\alpha,p}^2. \tag{5.29}$$

Extra Info 10. For exact results and references on the various scenarios described, we refer the reader to (Qiu, 2013, Ch.7).

Since the above limits have an (approximate) type-I error rate of α, and the periods are typically independent, then we can readily apply Eq.(5.8) to compute ARL_0.

5.3.3 Extra Info- Other Pooling Statistics

Hotteling's T^2 implicitly targets a weak signal on many coordinates. In the signal detection literature, this is known as a *dense* signal. In our intensive care unit example (Example 2), it may be the case the a patient's deterioration is mildly manifested in many CTQs. It may however, be the case, that deterioration is manifested in a small subset of CTQs. This is further emphasized in Example 3.

Example 3 (Cyber Monitoring System)**.** Consider a server farm. All servers dump their status into logs. These include CPU loads, temperature, network I/O, etc. The administrator is worried about an imminent attack, and it thus parsing the logs for CTQs, and inspecting the on-line status on his dashboard. He knows that the cyber-attacker is no amateur, so that if any fingerprint is left in the logs, it will be very subtle.

The intensive care example, and the cyber security example, motivate our search for a statistic, that unlike Hotteling's T^2 , is sensitive to *sparse* signals. I.e., a signal that is manifested in very few of the p CTQs. For this purpose, we merely offer several candidate multivariate statistics, with references where appropriate.

Sparse
Signal

Max pooling Where we control the process using the max over CTQs. Useful when we expect signal in a handful of CTQs.

Higher Criticism Where pooling is performed using the Higher Criticism statistic. Appropriate for an intermediate *rare-weak* sparsity pattern. See details in (Jin et al., 2005).

Skewness Statistic Where pooling is performed using the empirical third moment. Appropriate for an intermediate sparsity pattern. See details in (Jin et al., 2005).

5.4 Economical Design of Control Charts

Up until now, our design of control charts was driven by type-I error rates, and ARLs. Economical consideration were merely implied. In this section, economical consideration take the driver's seat. We present a toy model, to demonstrate the economical optimization of design parameters in a $\bar{x}$-chart . Before beginning, a few remarks are in order.

Remark 6 (Economical Design of Control Charts)**.**

1. According to Montgomery (2007)

> Saniga and Shirland (1977) and Chiu and Wetherill (1975) report that very few practitioners have implemented economic models for the design of control charts.

Hmmmm.. Have things changed since 1977?

2. A comprehensive theoretical analysis of the optimization of a quality control system may be found in Girshick and Rubin (1952). Again, Montgomery (2007) is skeptic:

> The optimal control rules are difficult to derive, as they depend on the solution to complex integral equations. Consequently, the model's use in practice has been very limited.

In light of the above skepticism, and following the lines of Duncan (1956), we aim at the modest goal of an economical optimal $\bar{x}$-chart . Our target function is optimizing the expected income per hour, with respect to the design parameters:

$$\max_{n,L,h}\{\mathbf{E}\,[C/T]\} \tag{5.30}$$

where C is the income between two productions halts, i.e., a *cycle*; T is the cycle duration; n is the number of samples per period; L governs the control limits via $UCL := \mu_0 + L\sigma_{\bar{x}} = \mu_0 + L\sigma_x/\sqrt{n}$; h is the time interval between sample periods.

We now need to establish how $\mathbf{E}\,[C/T]$ is related to n, L, h. Here is our set of assumptions and notation:

1. When in control (IC), production is centred on μ_0, assumed known.

2. When out of control (OC), $\mu_1 = \mu_0 \pm \delta\sigma_x$.

3. When OC, production may proceed (!).

4. Search and repair costs are not part of C.

5. OCs occur as a Poisson process, with rate λ events per hour. The expected time from a sampling to an OC events is thus

$$\tau := \frac{1 - (1 + \lambda h)e^{-\lambda h}}{\lambda(1 - e^{-\lambda h})} \tag{5.31}$$

6. The power to detect an OC is

$$\pi := \Phi(-L - \delta/\sqrt{n}) + (1 - \Phi(L - \delta/\sqrt{n})) \tag{5.32}$$

7. The false alarm rate

$$\alpha := 2\Phi(-L) \tag{5.33}$$

8. Because of the Poisson process assumption, $\mathbf{E}\left[C/T\right] = \mathbf{E}\left[C\right]/\mathbf{E}\left[T\right]$.

9. The expected cycle length:

$$\mathbf{E}\left[T\right] = \frac{1}{\lambda} + \frac{h}{\pi} - \tau + D \qquad (5.34)$$

where $\frac{1}{\lambda}$ is time IC; $\frac{h}{\pi} - \tau$ is the time the process is OC until detection; D is a fixed time to identify the assignable cause.

10. The expected income per cycle

$$\mathbf{E}\left[C\right] = V_0\frac{1}{\lambda} + V_1\left(\frac{h}{\pi} - \tau + D\right) - a_3 - \frac{a_3' e^{-\lambda h}}{1 - e^{-\lambda h}} - (a_1 + a_2 n)\frac{\mathbf{E}\left[T\right]}{h}$$

where V_0 is the net income per cycle when IC; V_1 is the net income when OC; $(a_1 + a_2 n)$ is the fixed and variable cost of taking a sample; a_3 is the cost of finding an assignable cause; a_3' is the cost of investigating a false alarm.

Given all the above, we may now plug Eq.(5.30) into our favourite numerical solver to find the optimal h, L, n.

5.5 Shewart Charts With Other Test Statistics

We have been focusing on the $\bar{x}$-chart for ease of exposition. There are, however, many cases where the mean is not an appropriate test statistic. Examples include:

1. A discrete CTQ, where only the number of non-compliances can be counted.

2. Where the departure from statistical control is not only a shift in μ.

The following charts are designed for those cases. Practically all of the ideas presented for the $\bar{x}$-chart may be adapted to these other test statistics after appropriate adaptations. The reader is referred to Montgomery (2007) for the details.

5.5.1 R Chart

Where $\bar{x}$ is replaced by the range. Sensitive to variability changes. Popularized by its ease of implementation, in the pre-PC age.

5.5.2 s Chart

Where $\bar{x}$ is replaced by s. Sensitive to variability changes. Usually more sensitive than the range.

5.5.3 s^2 Chart

Like the s chart, only in variance-scale.

5.5.4 p and np Chart

Where $\bar{x}$ is replaced by the proportion (p), or number (np), of non-conforming units. Appropriate for attributes, i.e., discrete QTCs.

5.5.5 c Chart

Like a np chart, but where the number of nonconforming units is replaced with the total number of nonconformances, allowing multiple defects per unit.

5.5.6 u Chart

Like the c chart, but allowing a variable number of units per sample (varying n).

5.5.7   Regression Control Chart

In a *regression control chart*, the test statistic can be a regression coefficient. When compounded with multivariate charts, a regression control chart may accommodate several regression coefficient, or the residuals.

5.6   Extensions of Control Charts

We have been discussing the simplest of $\bar{x}$-chart chart, for ease of exposition. Several immediate extensions are possible. The reader is referred to Montgomery (2007) for details.

One sided charts Just like hypothesis testing, where we may consider non-directional or directional hypotheses, we may consider directional control chart. Obviously the control limits and the test statistic may need appropriate updates.

Autocorrelation and time series models The assumption of independence between sampling periods may be relaxed, by adopting a quantitative or qualitative model for temporal dependence.

Running Window Versions The ideas of pooling information over periods with a moving averages, EWMA, and CUSUM may be extended to all type of Shewart charts.

Multiple Charts When inspecting a process, rarely does one inspect a single chart at a time. A typical dashboard would include several charts running in parallel, as depicted in Figure 2.10.

5.6.1 Cuescore Charts

[TODO]

Chapter 6

Design of Experiments - DOE

6.1 Terminology

Alias

Balanced Design

Blocking

Factor Encodings

Confounding

Control Group

Design

Design Matrix

Effect

Experimental Unit

Factors

Fixed Effect

Random Effect

Interaction

Lack of fit error

Pure Error

Orthogonality

Replication

Resolution

Response

Screening Design

Test Plan

Treatment

Treatment Group

Variance Components

6.2 Randomization

6.3 Full Factorial Designs

6.4 Fractional Factorial Designs

6.5 Robust Parameter Design (RPD)

6.6 Latin-Square Design

6.7 Plackett–Burman Design

6.8 Computer Experiments

Chapter 7

Acceptance Sampling

Chapter 8

Reliability

Chapter 9

Revisiting System Capability Analysis

[TODO]

9.1 System Capability with Control Charts

9.2 System Capability with Designed Experiments

Appendix A

Notation

In this text we use the following notation conventions:

x A column vector, or scalar, as implied by the text.

$:=$ An assignment, or definition. $A := a$ means that A is defined to be a.

$\prod_{i=1}^{n}$ The product operator: $\prod_{i=1}^{n} x_i := x_1 \times \cdots \times x_n$

$\#\{A\}$ The count operator. Returns the number of elements of the set A. Also known as the *cardinality*.

$\Phi(t)$ The standard Gaussian CDF at t: $\Phi(t) := P(Z < t)$.

$\phi(t)$ The standard Gaussian density at t: $\phi(t) := \frac{\partial}{\partial t}\Phi(t)$.

x' We use $'$ for the transpose operation. For a $1 \times p$ vector x, then x' is $p \times 1$.

$\mathbf{x}_n \rightsquigarrow P$ Convergence in distribution: for large enough n, then $\mathbf{x}_n$ is distributed like P.

Appendix B

SPC in R

Here is a list of packages worth knowing for SPC in R. The list is taken from Qiu (2013).

cpm [TODO]

mnspc [TODO]

msqc [TODO]

qcc [TODO]

spc [TODO]

Bibliography

M. Basseville, I. V. Nikiforov, and others. *Detection of abrupt changes: theory and application*, volume 104. Prentice Hall Englewood Cliffs, 1993.

A. J. Duncan. The Economic Design of X Charts Used to Maintain Current Control of a Process. *Journal of the American Statistical Association*, 51 (274):228–242, 1956.

M. A. Girshick and H. Rubin. A Bayes Approach to a Quality Control Model. *The Annals of Mathematical Statistics*, 23(1):114–125, 1952.

J. Jin, J.-L. Starck, D. L. Donoho, N. Aghanim, and O. Forni. Cosmological non-Gaussian Signature Detection: Comparing Performance of Different Statistical Tests. *EURASIP J. Appl. Signal Process.*, 2005:2470–2485, 2005.

D. C. Montgomery. *Introduction to statistical quality control*. John Wiley & Sons, 2007.

E. S. Page. Continuous Inspection Schemes. *Biometrika*, 41(1-2):100–115, 1954.

K. B. Petersen and M. S. Pedersen. *The matrix cookbook*. Citeseer, 2006.

P. Qiu. *Introduction to Statistical Process Control*. Chapman and Hall/CRC, Boca Raton, 2013.

J. L. Rodgers and W. A. Nicewander. Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, 42(1):59–66, 1988.

Wikipedia. *Design for Six Sigma — Wikipedia, The Free Encyclopedia*. 2015a. [Online; accessed 29-October-2015].

Wikipedia. *Eight dimensions of quality — Wikipedia, The Free Encyclopedia*. 2015b. [Online; accessed 28-October-2015].

Wikipedia. *Lean manufacturing — Wikipedia, The Free Encyclopedia.* 2015c. [Online; accessed 29-October-2015].

Wikipedia. *Quality (business) — Wikipedia, The Free Encyclopedia.* 2015d. [Online; accessed 28-October-2015].

Wikipedia. *Receiver operating characteristic — Wikipedia, The Free Encyclopedia.* 2015e. [Online; accessed 6-November-2015].

Wikipedia. *Value engineering — Wikipedia, The Free Encyclopedia.* 2015f. [Online; accessed 29-October-2015].

Wikipedia. *Zero Defects — Wikipedia, The Free Encyclopedia.* 2015g. [Online; accessed 29-October-2015].