

# Six sigma and its application to software development

C. C. Shelley

Oxford Software Engineering  
9 Spinners Court, 53 West End, Witney, Oxfordshire, England, OX28 1NH  
shelley@osel.netkonect.co.uk

## ABSTRACT

*The term 'six sigma' has its origins in statistical process control. More recently it has become better known as the name of a wide ranging set of data based process improvement techniques. This paper begins by describing the statistical term and examines its applicability for measuring software and software development processes. Related, more widely applicable but less sensitive statistical process control methods are then described. Comparison of these two approaches is used to provide the basis for a notional 'quality maturity model' which can be used to provide criteria for the selection of appropriate techniques in particular software development environments. The origins of the new six sigma process management methods are then described and its themes, characteristics and techniques are reviewed with comparisons to the quality maturity model. Finally the applicability of both the statistical six sigma and related techniques and the new improvement methodology are assessed.*

## 1. Introduction

The term 'six sigma' tends to polarize opinion when mentioned in software engineering circles. Technically inclined engineers will identify the technical limitations of the statistical analysis and in the process eliminate other related and potentially valuable statistical analyses. Those familiar with the new six sigma management methods can find themselves promoting methods and tools for improvement that need to be treated with caution in a software development environment. This paper describes both six sigmas and related areas and

assesses their value to software development organizations attempting to improve their software processes.

## 2. Six Sigma as a statistical term

Manufactured items generally have one or more design characteristics that have values assigned to them: the weight of a tablet, the light output of a lamp, the diameter and length of a shaft. The processes that produce these items are imperfect so the products will vary in their design characteristics. If an item exceeds the tolerances of the design characteristics then it is defective and is either scrapped or reworked.

To ensure that the number of defectives is kept to an acceptable level random sampling and analysis of the variability of the design characteristics is undertaken. It is often found that the variation in design characteristics conforms, more or less, to the normal distribution or 'bell shaped' curve, although there is a tendency for distributions to have a slight amount of skew, or a tail on the higher value side of the curve. This curve is described by the average value and the spread. The average is usually the mean, but, with marked deviation from the normal distribution, the median or mode may be used with an empirical correction,  $\text{mean} - \text{mode} = 3(\text{mean} - \text{median})$  being applied. The spread can be characterized by the minimum and maximum values but it is more usual to use the standard deviation, or sigma (?).

By comparing the average value and the spread with the design characteristic tolerances (known as the upper and lower specification limits or USL and LSL) a manufacturing process can be managed. It is usual for the random variation to be corrected first,

reducing the standard deviation (the spread), then the systematic variation is corrected, to place the average value mid way between the specification limits. The random and systematic variations are then managed to control the process and maintain an acceptably low level of defectives outside the USL and LSL.

For a sample with a standard deviation ( $\sigma$ ) if the USL and LSL are at  $\pm 1\sigma$  then 69% of the sample will be within limits; if at  $\pm 2\sigma$  then 95%, and at  $\pm 3\sigma$ , 99.999999%. This  $3\sigma$  level of control is equivalent to a defective rate of 2 parts per billion ( $10^9$ ). In practice a degree of process 'drift' of about  $\pm 1.5\sigma$  is usually assumed giving a defective rates of 3 to 4 parts per million. This very low rate of defectives has led to the term  $6\sigma$  being equated with 'zero defects'.

The use of standard deviations and specification limits is used to describe 'process capability' or  $C_p$  of industrial processes:

$$C_p = (USL - LSL) / 6\sigma \text{ or } 2T / 6\sigma$$

If  $C_p > 1$  the process is capable; if  $< 1$ , incapable.

This assumes no process drift. To account for drift another index,  $C_{pk}$ , is used:

$$C_{pk} = \text{lesser of } (USL - \text{process mean}) / 3\sigma \text{ or } (LSL - \text{process mean}) / 3\sigma$$

If  $C_{pk} < 1$  the process is incapable.

This form of statistical analysis is widely used in manufacturing and often manifests itself in the form of control charts for variables. This form of measurement and analysis is being promoted for use in software development as part of the maturation of software development into an engineering discipline.

The adoption of this form of measurement should be considered carefully. There are a number of constraints that limit its use that, while evident in production engineering, become very significant in software development. Specifically: the process is a replication process – tablets, lamps etc. (is software development a replication process?); the data are of at least interval scale type, the distribution is continuous and the distribution is, approximately, normal. Such conditions are rare in software development.

<sup>1</sup> NB process mean, not sample mean, i.e. approximately the mean of sample means.

However there are related statistical process control techniques that are far more suitable for software development.

### 3. Related statistical process controls

The normal distribution is one of three distributions. The normal distribution is continuous, i.e. values can take any value on a continuous scale. The other two, the binomial and Poisson, are discrete. That is, the data can only take values corresponding to positive integers. The binomial distribution is applicable when random samples are being taken from batches of products in which individual items have the same probability of being defective. If  $n$  items are selected and their probability that each item will be defective is  $p$  then the probability of obtaining  $r$  defectives is:

$$\frac{n!}{r!(n-r)!} \cdot p^r (1-p)^{n-r}$$

for all integral values from 0 to  $n$ .

When  $n$  is very large and  $p$  is very small this becomes:

$$\frac{m^r e^{-m}}{r!}$$

where  $m$  is the mean number of defects per sample. This limiting form is the Poisson distribution. This gives the probability of obtaining  $r$  defects, for all positive integral values of  $r$ , in a large sample in which the probability of obtaining a defect is constant and small for every item in the sample.

*"The type of case in which the Poisson distribution must be used in quality control is in counting the number of defects in manufactured items such as a roll of cloth or a radio set."*<sup>3</sup>

These discrete distributions are more appropriate to much of software development where *counting* of defects (and other entities) is very much easier and more common than continuous *measurement*. They are the basis for the control of attributes used where quantitative evaluations of quality are problematic. The extreme case is where no characteristic can be measured quantitatively. A similar problem occurs where a complex manufactured item may contain defects but not itself be completely defective (like software?). The control of attributes is a less sensitive technique than the control of variables but

<sup>2</sup> For large values of  $n$ , the binomial distribution does tend to form the normal curve – and even for comparatively low values of  $n$  where  $p$  and  $q (= 1-p)$  are nearly equal.

<sup>3</sup> 'Essentials of Quality Control', Huitson and Keen, 1965.

has the advantages that it is more robust, more widely applicable and, in general attributes are quick to assess, to the extent that variables are often converted to attributes for assessment.

The process for assessing attributes is by ongoing sampling from the process under examination. The distribution of defectives will follow a binomial distribution with average  $M = nP$  where  $n$  is the sample size and  $P$  is the average proportion of defective items produced by the process. From this the probability of a sample with  $m$  defectives in a sample can be calculated and compared with warning or action limits.

A control chart, known as an 'np-chart' is used to represent process data. This is set up by collecting samples and dividing the total number of defectives by the number of samples to establish an initial estimate of  $M$ . Where samples contain more than one item and each item can contain more than one defect a 'standardized c chart' or 'u chart' can be used.

Where small samples are unavoidable cumulative sum (cumsum) charts are particularly useful and should be used. Their value is in their ability to reveal small changes in the process difficult to detect in np or u charts. In practice simple plots of cumulative counts (of defects or changes for example) serve much the same purpose and are occasionally found in use by software developers and support groups.

#### 4. A 'Quality Maturity Model'

The use of defect, or other data, to control a process can be considered as the second of a set of four quality management types.

*Type 1* is simply quality control. QC activity, reviews or tests, examines the product and defects identified are removed, or the item is scrapped, or subsequent processes that will be applied to the item are modified.

*Type 2* is process control, discussed above. The focus shifts from the items being produced to the process used to produce them. Defects identified in items examined are used to analyse and control the manufacturing process. In process control defect data are disassociated from the items containing them.

*Type 3* quality management replaces the tolerance limits of type 2 with Taguchi's 'loss function' where deviation from the target specification is a function (usually the square) of the deviation. This leads to the idea of minimization of the loss and

consequential continuous process improvement, guided by experiments to identify the sensitive parameters.

*Type 4* quality management is directed at reducing the influence of the design parameters themselves, effectively designing a product that is robust or insensitive to the many variations in manufacture. Instead of working to reduce variability of the variables or attributes within tolerance or control limits the intent is to expand the limits with no effect on the performance of the product.

For each of these types analytical techniques and methods have been developed by production engineers. However the ability of an organization to apply these methods should be assessed carefully based on perceived quality maturity, in much the same manner as the Capability Maturity Model can be used to provide guidance on appropriate software development practices. This is particularly important in organizations adopting the new six sigma, described below, where a number of the type 3 and 4 analytical tools are proposed.

### 5. The new Six Sigma

The new six sigma is very different to the statistical analysis. It is a collection of management methods and tools developed to aid process improvement. Developed in Motorola during the 1980's to reduce defects in manufacturing it was later extended to business processes and service operations.

It has been adopted by General Electric and publicized by Jack Welch, who was the GE CEO. Six sigma can be viewed as a development or enhancement of TQM type practices.

Like TQM it depends on the use of facts and data, much of it statistical in nature. The objectives of six sigma activities are increased customer satisfaction, reduced cycle time and reduced defects – all appropriate to software development. The distinguishing characteristics of six sigma are claimed to be the involvement of middle management, accountability for results, the need for controls to ensure the persistence of improvements, and financial accountability.

The central technique from which six sigma derives its name is a simplified (simplistic to statistical purists) version of the statistical analysis described above: for any process 'defect opportunities' are identified and defects per defect opportunity calculated, where a defect is a failure to meet parameters 'critical to quality' (CTQ). These CTQ limits are equivalent to the production engineer's upper and lower specification limits. What was

originally conceived as the tails of a normal distribution extending beyond the specification limits has been simplified to be a simple count of defects per defect opportunity with little concern over the nature of the distribution. Yields in terms of 'sigmas' are calculated but with process drift included. So a six sigma process in the new six sigma would have a defect yield of 99.9997% (= 3.4 ppm).

Six sigma has six elements or themes supported by its numerous tools and methods: 1 - customer focus – six sigma improvements are defined in terms of customer satisfaction and value, 2 - data and fact driven management – six sigma provides managers with methods to determine what information they need and equips them to acquire and analyse it, 3 - six sigma focuses on process as the primary entity of interest, 4 - six sigma encourages management to be proactive by requiring the identification of improvement goals, , setting priorities and questioning ways of working, 5 – boundaryless collaboration.- in working to deliver customer satisfaction six sigma works to break down barriers and eliminate process disconnects and internal competition, 6 - strive for perfection and tolerate failure – continuous improvement.

Six sigma can be implemented at any of a number of levels in an organization. At the lowest level it can be introduced at a tactic for problem solving. Use it for addressing persistent problems. This approach is recommended for organizations not wishing to trigger major organization change. The next level is for strategic improvement. Aim improvements at one or two business critical issues with the intention of addressing organizational weaknesses or opportunities. The most ambitious level is business transformation where six sigma is the vehicle to drive the organization to re-invent itself – as General Electric did.

Which ever level of implementation is adopted the same roles and responsibilities are adopted. These are the 'belts'. A black belt is a full time role dedicated to tackling business critical issues familiar with the tools and techniques of six sigma. A master black belt will coach black belts in their roles, and a 'green belt' performs the same role as a black belt but in a part time capacity. Typically, black belts are expected to have completed several weeks training. (This may be a major disincentive for many organizations, especially smaller ones , that cannot afford this level of investment.) The belts will be working under the sponsorship of a champion or sponsor who is accountable for the results of the process improvement activity. Finally there is the implementation leader who is responsible for the whole six sigma effort. This is usually a senior manager who reports to, or is, the head of the

organization. There is no formal certification for the belt roles, some organizations have certification for black belt and master black belts, others take a more relaxed view and judge belts by their results.

Process improvement activity is generally managed as projects within which the six sigma tools and techniques are applied. One of the most valuable models is the 'DMAIC' problem solving framework. This provides a structure to the improvement activity which otherwise can lack focus and fail to deliver worthwhile results. DMAIC is an acronym: *Define* the problem – who is the customer, what are the benefits? *Measure* – gather data to quantify the problem. *Analyse* the data and determine 'sigmas' or performance measures for the processes that are the focus of the improvement project. *Analyse* the data, consider the causes and explore the 5Ms and 1P<sup>4</sup>. *Implement. Control* – develop monitoring processes, prepare a response plan and provide key management metrics – including financial – and hand off.

Within this framework the six sigma toolset is deployed as determined by the black belt(s). The tools are grouped into five groups:

*Ideas and Organization* - brainstorming, affinity diagramming SIPOC (Supplier, Input, Process, Output, Customer), multivoting, structure tree, flowcharting, Ishikawa diagramming.

*Data Gathering* – sampling, operation definitions, VOC (voice of the customer), check sheets, MSA (measurement systems analysis, something that the software measurement community could benefit from!)

*Process and Data Analysis* – process flow analysis, value analysis, graphics.

*Statistical Analysis* – significance, correlation and regression, DoE (Design of Experiments).

*Implementation and Process Management* – PM methods, FMEA (Failure Mode and Effects Analysis), Process documentation.

These, and other, six sigma tools vary widely in power and sophistication. The six sigma 'belts' are trained to deploy these tools, however even with this training care should be taken to ensure that tools are carefully selected to ensure that they are not used beyond their intended scope.

---

<sup>4</sup> Methods, Machines, Materials, Measures, Mother Nature (environment) and People. It is easy to be dismissive of these simple *aide-memoires* but they can be very useful in providing groups with a common way of working.

Careful chosen and used these tools are very effective for understanding and improving manufacturing, service delivery and other processes. However the software development community can claim that software development is not a manufacturing (i.e. replication) process but a less well defined *design* process. While there is some truth to this there are significant amounts of software development that use, or depend on, potentially well defined, repetitive processes, configuration and release management being examples, that can benefit from the application of six sigma improvement efforts. Consequently if an organization is embarking on a six sigma improvement programme special pleading by the software developers that their ways of working places them out of scope will be unhelpful. Even those less well defined processes in software development will benefit from careful scrutiny and improvement efforts.

Where (software) design processes are particularly the focus of improvement a variant of the standard six sigma project approach may be appropriate. DFSS (Design For Six Sigma) is based around a selection of the six sigma tools that includes the 'design of experiments' collection of techniques. These techniques enable the critical design parameters for products (software) or process improvement (software development) examined and tuned to meet product or organizational needs. In outline, instead of attempting to control all the variables that affect the performance of a product or process and exploring one variable at a time an orthogonal array of variables can be developed. Trials based on combinations indicated by the array produce data that is analysed using techniques such as ANOVA (Analysis of Variance) to indicate the most critical parameters, or combinations of parameters, to watch. These techniques while an order of magnitude more powerful than the traditional TQM 'seven tools' require a degree of process and statistical sophistication that will usually place them in the upper levels of 'quality maturity' and are currently out of the reach of most software development organizations.

## 6. Applicability to software development

The original statistical six sigma has no value for software development. The types of data and data distributions required for the control of variables are rare and the degree of sensitivity to process provided by six sigma is beyond the practical limits of software development. Defect densities for mission critical software are around 0.1 to 0.5 defects/KLOC with military systems having densities in the range 5 to 50 defects/KLOC. Assuming the inclusion of

process drift and consequent six sigma equivalent of 3 to 4 defects per million lines of code this is 0.004 defects/KLOC.

While control of variables to six sigma precision is of limited use the related control of attributes is remarkably applicable. The use of defect counts or qualitative assessment of product, especially complex products is exactly what is required in software development. Furthermore the use of sophisticated production engineering and quality management techniques may well find applications in the more engineering oriented software development organizations.

The targets of the new six sigma, customer satisfaction, reduced cycle time, and reduced defects, would also seem to be entirely appropriate to software development. The criticism that the new six sigma is only applicable to production or replication process is not convincing. The statistical approaches have been so simplified that they could find useful, tactical application in most software environments capable of counting process or product defects, even if they lack the rigour of formal statistical process control. The DMAIC problem solving framework is also applicable to software development, and the attitude the new six sigma fosters is also highly desirable. Where specific design and development process issues are to be dealt with then DFSS is available to those organizations capable of exploiting it.

## 7. Closing remarks

The original six sigma analytical method is not applicable to software development, however there are less well known aspects of statistical process control that are, and these await adoption and development by software development organizations.

The new six sigma, while expensive, and with some inappropriate statistical tools, and perhaps with over long project cycle times can never-the-less, with care, be adapted and exploited for software process improvement too.