

Shewhart and the Probability Approach

The difference is much greater than how we compute the limits

In theory, there is no difference between theory and practice. In practice, there is. During the past three months we have looked at three families of probability models and found that they share some remarkable properties. These properties provide a theoretical explanation of how and why process behavior charts work as they do. However, the ultimate justification for Shewhart's approach lies in the fact that it works in practice.

Two types of variation, two actions

When a process displays routine variation, we may think of the differences between items as being created by a constant system of a large number of "common causes" in which, according to Walter A. Shewhart, "no cause produces a predominating effect." When this happens, according to W. Edwards Deming, "It will not be profitable to try to determine the cause of individual variations." On the other hand, when a process displays exceptional variation, "it will be profitable to try to determine and remove the cause of the exceptional variation."

Given this distinction, the process behavior chart is a technique for detecting which of these two types of variation is displayed by a given process. The objective is to give the user a guide for taking appropriate action—to look for assignable causes when the data display exceptional variation, and to avoid looking for assignable causes when the data display routine variation.

It is important to note what is deliberately absent from the previous paragraph. There is no mention of estimating parameters for a specific model; there is no mention of fitting a probability model; and there is no mention of statistical inference regarding the process location or dispersion. Knowing when to look for assignable causes of exceptional variation, and when to avoid doing so, requires the *characterization* of the process behavior rather than the *estimation* of parameters for some probability model. Those who advocate using the probability approach miss this distinction.

Two types of mistakes

In the characterization problem, there are two types of mistakes that can be made. The first is to attribute an outcome to an assignable cause when it is simply the result of common causes. The second is to attribute an outcome to common causes when it is, in truth, the result of an assignable cause. It is impossible to avoid both of these types of mistakes. So this cannot be the objective. Instead, a realistic objective is to minimize the overall economic loss from making these mistakes. To this end, Shewhart created the process behavior chart with three-sigma limits.

Shewhart's use of three-sigma limits, as opposed to any other multiple of sigma, did not stem from any specific mathematical computation. Rather, Shewhart found that the use of three-sigma limits "seems to be an acceptable economic value," and that the choice of three sigma was justified by "empirical evidence that it works." This pragmatic approach is markedly different from the strictly mathematical approach commonly taught by those who have not understood what Shewhart was doing. In fact, to have a practical, sensible, and robust approach to the construction of process behavior charts, Shewhart deliberately avoided any dependence on unnecessary mathematical detail.

Process improvement

Notice what process behavior charts do—they seek to identify if the process is behaving one way or another. This, in effect, is the same as asking if the process exists as a well-defined entity, where the past can be used to predict the future, or if the process is so ill-defined and unpredictable that the past gives little clue to the future. As Shewhart said, "*We are not concerned with the functional form of the universe [i.e. probability model], but merely with the assumption that a universe exists.*"

We are not trying to find some *exact probability model* to describe the process, but we are trying to determine if the process fits (at least approximately) a *very broad model* of behavior that Shewhart described as "predictable within limits." This reversal of the roles of "process" and "model," and the broad rather than the precise nature of the "model," are not addressed by those who seek to use the probability approach. Many of those who propose "alternatives" to process behavior charts begin with ways of finding detailed models to describe the current process.

Of course, the problem with such approaches is that a production process is always changing. Left to itself, it will deteriorate. With appropriate help, it can also be improved. The question is how to go about the job of improving it. By comparing the process to the broad model of "predictable behavior" we can make the appropriate choice for process improvement. The emphasis is not upon the use of

the model to describe the process, but upon the characterization of the process behavior as a starting point for process improvement.

The probability approach

In the August 1967 issue of *Industrial Quality Control*, E. S. Pearson wrote, “While the broad lines of Shewhart’s attack could not be questioned, it seemed that there was scope for sharpening the statistical procedures which he was using...” Here we have Pearson explaining what he did in 1935 when he filled in what he perceived to be mathematical gaps in Shewhart’s argument. In doing this he created the “probability approach” to process behavior charts that is contained in British Standard 600. In the subsequent 80 years, many of our fellow statisticians have followed Pearson down this same path of trying to add things to Shewhart’s approach. Unfortunately, when they do this they fall into the very trap that Shewhart had been careful to avoid.

The problem is that, to enable the development of a sophisticated mathematical argument, it is usually the case that assumptions need to be made that are unduly restrictive or indeed impossible in terms of the real world. There is no exception for process behavior charts. Indeed, in this case, the assumptions required for the probability approach beg the fundamental question that Shewhart set out to answer!

Specifically, Pearson argued that the data would need to be normally distributed before the process behavior chart limits would work as advertised. This approach makes a probability model a prerequisite for using a process behavior chart. However, before the outcomes for a production process can be said to follow any particular probability model, the process has to be operated predictably over an extended period of time. Of course, the only way to determine if this is indeed the case is to use a process behavior chart with that production process. This leaves us with a logical paradox: Which comes first, the chart or the probability model?

Even more unfortunately, this weakened version of SPC (though it is often perceived as stronger due to its mathematical rigor) has spread via the technical journals and is essentially the only version known to many in academic circles. Exact mathematical methods are both easier to teach and more impressive to the unwary, but they have seriously reduced the potential of what may be accomplished through the use of process behavior charts. And just what is the probability approach? The common version uses limits which are calculated so that, *supposedly*, while the process is operated predictably, there are precisely 27 chances in 10,000 that a point will fall outside the limits.

Shewhart's rebuttal

In order to understand the fallacy of the probability approach, it is instructive to consider Shewhart's argument on pages 275–277 of his 1931 book, *Economic Control of Quality of Manufactured Product* (Martino Fine Books, 2015 reprint). On these pages Shewhart seems to toy briefly with the idea of using the probability approach. Paraphrasing his argument on these pages, he points out that *if* a process were *exactly* stable, i.e., if it did unwaveringly fit some precise mathematical model, and *if* we knew the details of its underlying (fixed) statistical distribution, we *could* then work in terms of probability limits. However, he notes that, in practice, we never know the appropriate type of statistical distribution.

Although statisticians usually plump, almost as if it is a foregone conclusion, for their favorite distribution, the Gaussian or normal distribution, Shewhart disposes of the use of the normal distribution on page 12 of his 1939 book, *Statistical Method from the Viewpoint of Quality Control* (Dover Publications, 1986 reprint). And in *Economic Control* he notes that, even if the process were exactly stable, and if a normal distribution were appropriate (neither of which we would ever know), we would still never know the value of the mean or standard deviation parameter for that normal distribution. We could only *estimate* these parameters from the data. Because the probability calculations depend upon knowing *all* of these things, it will always be impossible in practice to compute the required probabilities.

Furthermore, in *Out of the Crisis* (MIT Press, 2000), W. Edwards Deming points out that, in practice as opposed to mathematical theory, *exactly stable processes never exist*. Real processes are never entirely free of perturbations. This identifies a vast gulf between the common mathematical assumptions and the real world. What does this imply? Surely not that we are to spend all of our time looking for these perturbations! No, of course not. What we need is guidance as to when the perturbations are troublesome enough to warrant attention.

Shewhart's process behavior chart, with its three-sigma limits, provides this guidance. *No calculations from the normal distribution, or any other distribution, were involved in the choice of the multiplier of 3.0*. Certainly, Shewhart did then check that this multiplier turned out to be reasonable under the artificial conditions of a normal distribution—and plenty of other circumstances as well. (As we have just seen in the past three months, fixed-width three-sigma limits will filter out more than 98 percent of the probable noise for just about any realistic probability model we might choose.) But that is a far cry from deducing that the choice of three-sigma was *based* on the assumption of a normal distribution. On pages 334–335 of *Out of the Crisis*, Deming says:

“The calculations that show where to place the control limits on a chart have their basis in the theory of probability. It would nevertheless be wrong to attach any particular figure to the probability that a statistical signal for detection of a special cause could be wrong, or that the chart could fail to send a signal when a special cause exists. The reason is that no process, except in artificial demonstrations by use of random numbers, is steady, unwavering.

“It is true that some books on the statistical control of quality and many training manuals for teaching control charts show a graph of the normal curve and proportions of area thereunder. Such tables and charts are misleading and derail effective study and use of control charts.

“Rules for detection of special causes and for action on them are not tests of hypotheses that the system is in a stable state.”

Thus, a major problem with the probability approach to process behavior charts is that it is totally out of contact with reality. The assumptions used for the mathematical treatment are turned into prohibitions, which are mistakenly imposed upon practice.

Restrictions such as the following are commonly encountered: “The data have to be normally distributed.” “The process behavior chart works because of the Central Limit Theorem—therefore you have to have subgroups of at least five observations.” “The chart will not work with serially correlated (autocorrelated) data—the observations must be independent of each other before you can use a process behavior chart.” “You have to transform the data to make them look normal before putting them on a chart.”

These, and others like them, are the unintended byproducts of the probability approach, and they are all completely wrong.

Today, 80 years after Pearson’s mistake in confusing Shewhart’s fixed-width limits with the traditional fixed-coverage limits used in statistical inference, the probability approach has morphed into the following steps:

1. Assume your dataset is homogeneous (and your process is predictable).
2. Decide which family of probability models to use.
3. Fit a probability model to your dataset.
4. Select some arbitrarily small value as the fixed coverage for the tail areas.
5. Use the probability model to find corresponding critical values.
6. Compute your fixed-coverage limits.
7. Use these limits to evaluate your data for evidence of unpredictable behavior.

The complexity of this approach makes it appear to be rigorous and exact. However, the rigor is undermined by the fact that the assumption in item one begs the question asked in item seven.

In contrast to the above, Shewhart's approach consists of the following steps:

1. Compute generic, fixed-width, three-sigma limits.
2. Use these limits to evaluate your data for evidence of unpredictable behavior.

Summary

The process behavior chart is not found at the end of a mathematical syllogism. It is an empirical procedure that is consistent with theory and that works in practice. This empirical justification is what takes it outside the restrictions of the probability approach and makes it a tool for the real world. Although other tools may be compared with the process behavior chart by means of the probability approach, and although other tools may even be simple enough and robust enough to work in the real world, such theoretical comparisons do not justify the use of any technique in practice.

Many statisticians consider a process behavior chart to be a sequential test of hypotheses procedure with a variable Type I error rate. Although the process behavior chart may be described, on a purely mechanical level, in such terms if the necessary (but unrealistic) assumptions are made, the use of the chart is completely different. Process behavior charts are much broader than this narrow view of their nature. They examine the data for behavior compatible with the *existence* of a universe (without necessarily implying that one can or does exist), rather than merely looking for a change in one or more parameters of a well-defined universe. No probabilities can be attached to the judgement regarding the existence of a universe simply because there is no probability structure without a well-defined universe. This is why the use of the probability approach severely limits the actual usefulness of the process behavior chart.

Some modern authors have understood the logical inconsistency of the original probability approach, and have sought to remedy this problem by defining a two-stage approach. In stage 1 the charts are used as an empirical tool. Then when the process is thought to be operating predictably, they move to stage 2 and use some probability model to "refine" the limits to achieve some fixed risk of getting a point outside the limits. Of course, the problem with this approach is that, as Deming said, no process ever completely gets to stage 2.

But more troubling than the complexity of the probability approach is the limited view of what process behavior charts are for. This limited view leads to a narrow perception of how to use the

charts. For example, some think that process behavior charts are only useful for process *monitoring*. The probability approach does not allow for much else because there most certainly is no distribution, and no probability, when a process displays exceptional variation. Although a process behavior chart *can* be used for process monitoring, the point is that monitoring is only a *minor* part of what charts can do, rather than being *all* that they can do. The probability approach generally relegates the process behavior chart to the role of a mere monitoring procedure after the process is presumed, by some undefined means or another, to find itself in a satisfactory state. The chart's function is perceived to be an early-warning device of when the process moves away from this supposedly satisfactory state.

The crucial difference between Shewhart's work and the probability approach is that his work was developed in the context, and with the purpose, of process *improvement* as opposed to process *monitoring*. From his perspective, a major purpose for creating the chart was to provide help to get the process *into* a "satisfactory state," which you might then be content to monitor (if not persuaded by arguments for the need of continual improvement).

This difference is far more important than you might at first appreciate. It gets right to the heart of the divide between the main approaches to the whole quality issue. On the one hand, we have approaches that regard quality merely in terms of conformance to requirements, meeting specifications, and zero defects. On the other hand, we have Deming's demand for continual improvement—a never-ending fight to reduce variation. The probability approach can only cope with the former. Shewhart's own work was inspired by the need for the latter.

ABOUT THE AUTHORS

Donald J. Wheeler

Dr. Donald J. Wheeler is a Fellow of both the American Statistical Association and the American Society for Quality, and is the recipient of the 2010 Deming Medal. As the author of 25 books and hundreds of articles, he is one of the leading authorities on statistical process control and applied data analysis.

Dr. Wheeler welcomes your questions. You can contact him at djwheeler@spcpress.com.

Henry R. Neave

*Dr. Henry R. Neave is the recipient of the 2001 Deming Medal and an Honorary Fellow of the Chartered Quality Institute (U.K.). He is author of the bestselling *The Deming Dimension* (SPC Press, 1990) and was a founder of the British Deming Association, where he served as director of research. He retired as the W. Edwards Deming Professor of Leadership and Management from the Nottingham Trent University Business School (England) in 2005.*