A probability chart for statistical process analysis, with applications

Nina F. Thornhill*, Roddy J. Hutchison† and Barry T. Haughie†

**Department of Electronic and Electrical Engineering, University College London, Torrington Place, London WC1E 7JE, UK*

†BP Chemicals Ltd, Bo'Ness Road, Grangemouth, FK3 9XH, UK

This paper reports an application of the Kolmogorov–Smirnov test for the purpose of detecting changes in the distribution of a sequence of measurements. The ‘probability chart’ of the title gave a value reflecting whether the distribution was constant in the neighbourhood of each observation. The chart had the advantage that the threshold for detection of a change was a dimensionless probability value. The non-parametric nature of the test made it suitable for measurements sampled from a non-Gaussian distribution.

Keywords: change detection; non-parametric statistics

This paper gives an account of a technique for statistical process analysis that detects changes in the local median value or in the local scatter of a sequence of measurements while taking into consideration the limitations of real production measurements. The test was applied to measurements of the melt flow rate of high density polyethylene (HDPE) from BP Chemicals’ Rigidex® processes. It was used in the automated retrospective analysis of large volumes of plant data for assessment of long term process performance and for highlighting significant changes. Other applications illustrate the simplicity of tuning of the method; these applications were to North Sea well-head pressure data and to detection of disturbances in an averaging level control loop.

The analyses were based upon testing of the hypothesis that the samples in two overlapping windows were drawn from the same population. In this context a ‘sample’ means a collection of observed values or measurements that have been grouped together for statistical purposes while the sample in a window is a set of several consecutive measurements. The results of the test, which was applied at every observation point, were displayed on a chart. The chart tended to show changes in the distribution as rapid transitions from a high probability value to a low probability and changes were easily detected by setting a threshold at a probability of 0.5. A key point was that the threshold for detection did not depend upon the values of the measurements so the

same detection threshold could be applied without re-tuning for different polyethylene grades. Testing the distribution in the neighbourhood of any particular observation point needed a forward look as well as a backward look. The procedure was therefore acausal and was best suited to retrospective analysis. In any on-line use, there would be a delay of several observation intervals before the procedure could signal a change.

There are several well known methods¹ for examining the distributions of numerical values within a sample. Tests such as Student’s *t*-test and the *F*-test examine whether the means or the standard deviations indicate that the samples are statistically the same. Those procedures assume the samples were drawn from a Gaussian probability distribution and on the strength of the Central Limit Theorem are frequently used with no prior testing of that assumption. They can, however, give errors when the Gaussian assumption is invalid. Other methods involve the testing of a sample against a hypothesised probability distribution, the parameters of which may be derived from the sample itself. A more general type of test makes no assumptions about the form or the parameters of the distributions and asks merely for a decision about whether two samples have been drawn from the same distribution. This work has applied such a test (the Kolmogorov–Smirnov test) because the probability distribution was unknown and was certainly non-Gaussian, and has compared the results with those from a test that did assume a Gaussian distribution.

Koninckx² has reported an application of a non-parametric test (the Wilcoxon test) to the detection of steady states in a boiler case study. The Wilcoxon method is a

*Author to whom correspondence should be addressed

rank order test that also makes no Gaussian assumptions. It is sensitive to changes in the median value of the distribution. Another type of technique reported in the literature detects changes in sequences of measurements by abstracting a qualitative representation. The qualitative techniques have often been directed towards non-steady batch process data³. Cheung and Stephanopoulos⁴ introduced a triangular representation for process trends in which selected points in the trend become the vertices of a fitted triangle. They showed how the triangles could be classified according to several types of features in the data. Love and Simaan⁵ detected constant episodes, ramps and spikes in measurements from a continuous process. They fitted least-squares straight lines to short measurement sequences and distinguished between a ramp and a constant episode when the slope of the line exceeded a threshold. The slope threshold was, however, a number that had to be tuned to the characteristics of the measurements. The threshold problem often seems to arise at the point where a qualitative decision must be abstracted from numerical data and this is true also of the industry-standard statistical process control (SPC) individuals and cu-sum chart methods. It is desirable to have a decision threshold that does not depend upon the expected range of the measured values otherwise the procedure has to be re-tuned for every application. An advantage of the proposed statistical test is that the decision threshold is independent of the measured values.

The next section of the paper outlines the statistical and computational methods and gives details of the polyethylene product measurement that provided the main application. Two other applications are introduced as well. The "Results and Discussion" section will present the outcomes of the applications and will contrast the results with those from a method that inappropriately assumes Gaussian statistics and with two types of SPC charts. The paper ends with a brief conclusion.

Methods

This section will describe the nature of the HDPE product measurements and outlines the development and application of the statistical procedure that formed the basis of the probability chart. The section will also outline some secondary statistical procedures and some additional tests. The purpose of the secondary statistical procedures was to display the features which the probability chart responded to so that its performance could be evaluated. The additional tests were introduced to compare the performance of the probability chart with other techniques.

This section outlines two further applications which illustrate the simplicity of the tuning of the procedure.

HDPE application

The procedures were applied to records stored in a data base that contained on-line process measurements and off-line polymer property measurements for HDPE

grades from several reactors. The data base contained more data than could be inspected manually. The present work concentrated on measurements of the polymer melt flow rate which is a measure of the melt viscosity. Because replicate tests tend to be scattered, BS 3412 (ISO 1133, 1988, condition 4) indicates that melt flow rate results should be reported only to two significant figures, so shifts or trends in the melt flow rate records may be obscured by scatter or by the limited resolution. *Figure 1(a)* shows melt flow rate measurements which contained changes in the distribution.

For the secondary statistical tests the samples were the sub-sequences of measurements within a window in the vicinity of an observation point. An observation point is the location on the x -axis at which a measurement, y , is plotted. The window $W(x_i)$ is the set of 20 observation points $\{x_j; j = i - 10, \dots, i, \dots, i + 9\}$. For the probability tests the two samples being compared were in the windows $V_1(x_i) = W(x_{i-4})$ and $V_2(x_i) = W(x_{i+4})$ separated by 8 points. These definitions ensured the reference observation point, x_i , was the right hand one of the two central points of the total of 28 points.

Several considerations led to the choice of the sizes of the windows and their separations. Firstly, both the size and separation were larger than any correlation effects due to process dynamics. The separation reflected the time scale of a typical process change; in the vicinity of a change one of the windows should sample mostly the new distribution while the other still samples the previous distribution. The size of an individual window was 20 points so that an analytical approximation could be used for the sampling distribution of the D statistic (the D statistic is outlined in a later section). Windows larger than that in the HDPE application could fail to detect short term changes in the distribution; conversely it would be possible to reduce the window size and to use a look-up table for the D statistic had the data records suggested that this was required. The section entitled "Statistical and computational methods" will give a guideline for the choice of window size relative to the separation.

Further applications

Two other applications were also studied. Each illustrated a feature of the tuning of the probability chart.

The procedure was applied to North Sea oil and gas well-head pressure data. The algorithm was used to determine the timing of offshore well-head choke changes which moved production from one steady value to another.

The method has also been applied to retrospective detection and timing of a disturbance in a continuous refinery process. The reason for presenting that application is to illustrate the decisions to be made in the choice of window size.

Statistical and computational methods

The aim of the statistical analysis was to evaluate whether the distribution of samples in the two windows

Figure 1 (a) Melt flow rate measurements for one HDPE grade. Scatter and stratification tend to obscure changes in the distribution. (b) The Kolmogorov–Smirnov probability plot for the measurements of (a). (c) The percentile statistics. Shaded regions indicate the periods of change detected by the probability chart. (d) A t-test probability plot (adjusted for unequal variances). The t-test indicates changes where none exist. (e) A statistical process control I-chart. The marked points indicate where the SPC rules detected changes. (f) A cu-sum plot. The changes in slope at observations numbers 40 and 90 correspond with changes in the mean value.

V_1 and V_2 in the vicinity of each observation point in a process trend could plausibly be considered the same.

The method used the D statistic in the Kolmogorov–Smirnov two sample test⁶ and implemented the test as suggested by Press *et al.*⁷. The test is suited to samples drawn from arbitrary distributions and is appropriate for continuous measurements. It is noteworthy that although the true melt flow rate values are continuous the specification of two significant figures has the effect of binning the melt flow rate measurements. The Kolmogorov–Smirnov test is known to behave conservatively with binned values so that if the test rejects a hypothesis when the measurements are binned one can have confidence in that decision. The value of D is the magnitude of the maximum difference between the cumulative sample frequency distributions of the two samples being compared. Given measurements y_j in a sample ($j = 1$ to N), the cumulative sample frequency function $S_N(y)$ is equal to $n(y)/N$, where $n(y)$ is the number of measurements less than or equal to y . The Kolmogorov–Smirnov two-sample statistic for samples from V_1 and V_2 with cumulative sample frequency functions $(S_N(y))_1$ and $(S_N(y))_2$ is:

$$D = \max |(S_N(y))_1 - (S_N(y))_2|$$

Figure 2 shows a graphical representation of D .

The sampling distribution of D can be found from published tables in the case of small samples or from an analytical approximation in the case of larger samples. The sampling distribution of D indicates the chances that the observed value of D could plausibly have arisen

if two random samples were taken from the same distribution; that is to say, it tests the probability of the null hypothesis of equal distributions. Large values of D correspond to small probabilities. Returning to the question of window sizes; large windows with small separations would give a limited range of D values and reduce the sensitivity of the method because the samples in the two windows would have most of their points in common. In practice, it is reasonable to choose the window size in the range 1.5 to 4 times the separation.

The “probability chart” of the title was the plot of the probability of D against observation number. Siegel and Castellan⁶ have discussed choices of a significance threshold. At issue is the level of risk of mistakenly rejecting a correct null hypothesis (a Type I error, or false positive result) compared with the danger of failing to reject an incorrect null hypothesis. Since this application needed sensitivity, the threshold was set at 50%, albeit at a risk of generating Type I errors. It

Figure 2 Graphical representation of the “ D ” statistic. The value of “ D ” is the largest distance between the cumulative frequency functions for the samples within two windows.

seems plausible that the occurrence of several consecutive values below the 50% threshold would improve the statistical confidence of each decision but we have not proved that assertion.

Percentile statistics were derived for the sample in the window $W(x_i)$ in order to give an insight into the performance of the probability chart. The k th percentile point is the value of y such that $n(y)/N$ is less than or equal to $k/100$. The statistics were the interquartile distance (75th percentile to 25th percentile) and the median value (50th percentile).

Alternative tests for the purposes of comparison comprised the t -test and the SPC individuals and cu-sum charts. The t -test assumed the samples were drawn from Gaussian distributions and returned a probability value for the null hypothesis that the means were equal (the test was adjusted for unequal variances). The SPC individuals chart was analysed with industry-standard rules. The action and warning limits for an individuals chart are usually taken as $\pm 3.09\sigma$ and $\pm 1.96\sigma$ (BS 5700, 1984) where σ is the standard deviation of all the in-control samples since the record began. For a Gaussian distribution in-control regions for the action limits and warning limits respectively would accommodate 99.8% and 95.0% of the measurements.

Tuning the algorithm

There are choices to be made in implementing the probability chart. The choices concern the number and placement of bins to use in the cumulative distribution and the number of samples to include in the moving window.

This section gives guidelines for the above tuning decisions rather than an algorithmic procedure. It is, however, important to realise that the decisions are easily made and, once made, apply to all processes with similar characteristics. The polyethylene application, for example, used 8 bins with approximately equal contents to span any polyethylene data set. The window sizes were chosen to contain 20 samples, separated by 8 samples. But once those parameters were fixed, they could be applied to all grades on all reactors. There was no need to set up individual thresholds by an inspection of every grade/reactor combination.

Three to eight bins sufficed to characterise the cumulative frequency for the applications presented in this paper. For applications with a lot of scatter in the measurements it was most effective to allocate the bin boundaries such that the bins contained approximately equal numbers of measurements in the record being examined. Of course, if the measurements are themselves stratified (measured to 2 s.f. for example) then exactly equal numbers may not be possible.

In cases where there is little scatter, and where large changes in steady value are to be detected, it has proved more effective to use smaller numbers of equally spaced bins containing unequal numbers of measurements. If a data record stays at a value for an extended period of time with little scatter, then that value should lie safely in the centre of a bin. The boundaries of bins contain-

ing approximately equal numbers of measurements, by contrast, tend to cluster close to the steady values. The choice of equally-filled or equally-spaced bins is therefore guided by the nature of the process.

Window sizes must, as already indicated, be wide enough to average over any known correlations due to process dynamics. An evaluation of the autocorrelation of data from the process can help. For example, in the polyethylene application, nearest neighbours had a correlation coefficient of about 0.4 so a window that averaged over 20 samples would certainly be insensitive to process correlations. Correlations also cover the effects of non-white noise. In the level control example, for instance, the scatter on the measurements had a periodic structure with a wavelength of about 4 to 5 sample intervals; the window in that application was 40 samples wide, averaging over several wavelengths.

Results and discussion

This results section first presents the outcome of the proposed probability chart for a sample of data from the polyethylene process together with plots of the percentile statistics which indicated that the probability chart was sensitive to changes in the distribution. The second sub-section presents results of the t -test and standard SPC methods on the same data sets and compares them with the results from the probability chart. The t -test gave unsatisfactory results because it was overly sensitive at times where there were no changes, while the SPC individuals chart was not sensitive enough. The SPC cu-sum chart indicated changes, but detection would rely upon an experienced operator applying a measurement dependent threshold.

This section also shows results from well-head pressure records and from an averaging level control loop. The well-head pressure example shows that the probability chart was insensitive to individual gross errors in the data, errors that were caused by data transmission problems. It also shows the effect of different placement of bin boundaries.

The averaging level control example illustrates the effect of a window size which is too small, as well as showing a determination of the onset of two process disturbances using a correctly-sized window.

The probability charts

Figure 1(b) shows the probability plot for the melt flow rate measurements of Figure 1(a). The shaded regions in Figure 1(c), which shows the percentile statistics, indicate points where the probability chart detected changes in the distribution showing that the chart was sensitive to shifts in the median and interquartile distance. Although some of the changes detected by the probability chart were visible by eye the probability chart overcomes any danger of overlooking features amongst the large volume of data. The chart also detected other more subtle changes.

An attractive feature of the probability chart was that it appeared to switch rapidly from high to low probability when the distribution changed. The changes were easy to detect because the plot gave unambiguous answers to the question of whether the distribution was changing.

The comparative tests

Figure 1(d) presents the probability results from the *t*-test for the same representative episodes and Figures 1(e) and 1(f) represent the SPC individuals and cu-sum charts. The *t*-test was unsatisfactory because it showed numerous false positives, detecting changes where none had occurred. The reason for its failure was that the distribution of the melt flow rate measurements had more outliers in the tails of the distribution than a Gaussian distribution. A composite Gaussian test for both mean and standard deviation (not shown) was even more unsatisfactory.

The SPC rules detected some but not all of the significant changes in the individuals chart (Figure 1(e)). One might increase the sensitivity of the individuals chart by an ad-hoc compression of the limits, but such an arbitrary method has no theoretical basis and could not be applied in a systematic manner across the range of HDPE grades from the different reactors. In Figure 1(f), an experienced person would detect changes in the cu-sum slope that indicate shifts in the mean value at observation numbers 40 and 90 and a downward trend through to observation number 100 showing that the

values were consistently below the mean during that time. The probability chart did not detect that feature. Any decisions about the cu-sum slopes, however, use measurement dependent threshold values and moreover their statistical significance would not be assessed.

A comparison of SPC charts with the proposed probability chart is not really a fair one. SPC charts are designed for on-line use and are thus causal, using data only up to the present time. They do not have the advantage the retrospective probability chart has of being able to take a forward look at future data so we should not expect the performance to be the same. Nevertheless, the comments about choices of thresholds remain valid.

The other examples

Figure 3 shows the probability chart associated with a simple record of well-head pressure data. The data samples are spaced by 1 minute in this example, as compared with the 4 hourly sampling interval of the polyethylene application. The chart in Figure 3(b) is a chart with three equally spaced bins, with the steady values well away from any bin boundary. The data record showed occasional gross errors due to data transmission problems from the off-shore location (the negative-going spikes). The probability chart was insensitive to these individual errors because one or two such samples make no great difference to the value of the *D* statistic. A measure based upon local variances, by contrast, would be sensitive to such errors.

Figure 3 A daily record of well-head pressure from the North Sea. (a) shows the whole record, including outliers due to data transmission errors. (b) is a close-up view on an expanded scale. (c) is a probability chart derived using equally-spaced bins in the cumulative frequency function. It detects the changes in steady state while being insensitive to individual outliers. (d) is a probability chart derived using bins with equal numbers of values. It detects the more detailed changes occurring in a steady state.

FIGURE 4

Figure 4 The level signal from a vessel under averaging level control. (a) shows the level measurement. (b) is a probability chart derived using a wide window that averages over the periodic noise signal. (c) is a probability chart derived using a window that was not wide enough.

Table 1 Tuning parameters used in the probability charts presented in this paper. The number of bins indicates the number of steps used to define the cumulative distribution. The bin type indicates whether the bins were spaced equally along the ‘value’ axis in *Figure 2* or whether the bin boundaries were chosen so that contents were similar. The table also shows how many measurements were included in each window and the separation of the windows

Figure number	Number of bins	Type of bins	Window size	Window separation
1(b)	8	Equal contents	20	8
3(c)	3	Equal spacing	10	6
3(d)	3	Equal contents	10	6
4(b)	3	Equal spacing	40	10
4(c)	3	Equal spacing	10	6

Figure 3(d) shows an analysis that can be achieved with the well-head data by using equally-filled bins. In that case the bin boundaries cluster close to the steady values. The probability chart then reveals more details of the episodes during which the pressure is running close to a constant value. In this example, we can see a periodic ramp-and-reset behaviour in the pressure measurement. The physical origin of the effect has not, however, been studied further.

Figure 4 shows the detection of disturbances in the averaging level control example. A probability chart using a wide window gave a clear signature of the onset of the disturbance and the return to the steady-state. On the other hand, *Figure 4(c)* shows the results from a chart with a narrower window (a window size that violates the guidelines given in the “Tuning the algorithm” section). The narrower window sampled differences due to the periodic noise and gave spurious indications of changes.

Record of tuning parameters

Table 1 lists the numbers of bins and the window sizes for the probability charts presented in Figures 1, 3 and 4.

Conclusions

This paper has shown that the probability chart could detect changes in the scatter and in the median value of melt flow rate measurements from a production grade of high density polyethylene. The procedure submitted the distributions of samples in two neighbourhoods of each observation point to a comparison by the Kolmogorov–Smirnov two-sample test. The result was plotted on a chart in which changes in the distributions were indicated in an almost binary manner. The detection threshold for the chart, in contrast with thresholds from many other methods, was a probability value that was independent of the expected range of the measured values. This attractive feature made the procedure readily adaptable to results from other HDPE grades with no need for re-tuning. The detected changes appeared to coincide with local variations in the median value or in the scatter as measured by the interquartile distance.

This paper also compared the results of the probability chart with other methods for detecting changes. Submitting the data samples to a comparison by the t -test rather than the Kolmogorov–Smirnov test gave numerous false positive results because the Gaussian assumptions of that test were not met. Simple SPC individuals and cu-sum charts detected some features, but were less satisfactory for another reason, which was that the detection threshold depended on the measured values.

Further examples showed the effects of tuning the chart. The effects of a choice of allocating the values in the data record into bins with similar numbers of measurements in each or into equally spaced bins was illustrated, and so was the effects of the choice of window size. The additional examples emphasise the universal applicability of the chart. Transferring the algorithm from one process or another requires adjustments of just a few easily understood parameters.

The probability chart gives a means for rapidly screening data from a large data base and has pinpointed times where there has been significant change. In the HDPE application it directed attention to specific points in the data base or to dates in the operators' log books where we might expect to find explanations for disturbances. It proved useful in the on-shore determination of offshore process changes in the well-head monitoring application. The level control example was presented only as a demonstration for the purposes of the paper. Nevertheless, a refinery is an integrated site and the means to track disturbances from one unit to another has potential value in the design of multivariable and supervisory control systems.

Acknowledgements

This work was supported by a Royal Society/Science and Engineering Research Council Industrial Fellowship and by BP Chemicals Ltd. The authors are grateful for the support of these organisations.

We wish also to thank BP Oil Ltd (Grangemouth, Scotland), Total Oil Marine plc (Aberdeen, Scotland) and Scomag Ltd (Motherwell, Scotland) for providing additional test data and production trials of the work presented in this paper.

References

1. Kreyszig, E. *Advanced Engineering Mathematics*, 6th ed, Wiley, New York, 1988
2. Koninckx, J. PhD thesis, Department of Chemical and Nuclear Engineering, University of Maryland at College Park, 1988
3. Marshall, C.T., Thornhill, N.F., Lilly, M.D., Buckland, B.C. and Fish, N.M. *Modeling and Control of Biotechnical Processes, IFAC Symposia Series Number 10*, Pergamon Press, Oxford, 1992, 283
4. Cheung, J.T-Y. and Stephanopoulos, G. *Comput. Chem. Engng.* 1990, 14, 495
5. Love, P.L. and Simaan, M. *Pattern Rec.* 21, 333
6. Siegel, S. and Castellan, N.J. Jr. *Nonparametric Statistics*, McGraw-Hill, New York, 1988
7. Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. *Numerical Recipes*, CUP, Cambridge, 1986