Chapter 1

CONTROL CHARTS FOR DATA HAVING A SYMMETRICAL DISTRIBUTION WITH A POSITIVE KURTOSIS

CASTAGLIOLA PHILIPPE

Dept. d'Automatique et de Productique
Ecole des Mines de Nantes, La Chantrerie 4, Nantes, France

1. Introduction

The designing of a "classical" control chart (Shewhart, EWMA, CUSUM) supposes that the probability density function of the quality characteristic X has to be normal or approximately normal. However, in some situations, it has been shown that this condition does not hold [Jacobs (1990)]. In order to design control charts when the underlying population is non-normal (without increasing the sample sizes), different approaches may be used:

- Use classical Shewhart control charts anyway. Many authors studied the effect of non-normality on Shewhart control charts: [Burr (1967), Schilling and Nelson (1976), Balakrishnan and Kocherlakota (1986), Chan, Hapuarachchi and Macpherson (1988)]. One important conclusion of these studies is that classical Shewhart control charts give good results unless the population is highly skewed.
- Assume that the distribution of the underlying population is known and then derive specific control limits which verify the type I error α. Such an approach was chosen by Ferrell (1958), Nelson (1979). Ferrell assumed a log-normal distribution for the underlying population and proposed control limits for the geometric midrange and the geometric range, whereas Nelson assumed a Weibull distribution and derived control limits for the median, range, scale, and location.
- Use distribution free control charts that provide a type I error close enough to the theoretical one. This approach was first considered by Cowden (1957) who proposed to split the skewed distribution into two parts at its mode, and to consider the two new distributions as two half-normal distributions having the same mean, but different standard deviations. Another very similar approach, the Weighted Variance control chart (WV control chart), was proposed by Choobineh and

RECENT ADVANCES IN RELIABILITY AND QUALITY ENGINEERING

Ballard (1987) who suggested to split the skewed distribution into two parts at its mean, instead of its mode, and then compute the standard deviations of the two new distributions using the semivariance approximation of Choobineh and Branting (1986). Finally, we can cite the recent works of Seppala (1995) who suggests to use the "Bootstrap" in the computation of control limits, Willemain and Runger (1996) who proposes to use the notion of "Statistically Equivalent Blocks" to design nonparametric control charts, and Castagliola (1997) who proposes an extension of the Weighted Variance method called the "Scaled Weighted Variance" method.

- Transform the data in order to make them quasi-normal. This approach was chosen by Pyzdek (1992), Farnum (1997) who used the Johnson system of distributions as a general tool for transforming the data to normality.

The method proposed in this paper, which follows the last approach, is devoted only to the designing of "classical" control charts (mean, median, standard deviation, range, EWMA, CUSUM, etc.) for data having a symmetrical distribution with a positive kurtosis (leptokurtic distribution). This method is based on the properties of the symmetrical Johnson S_U distributions which will be examined in the following section.

2. The Symmetrical Johnson S_U Distributions

Let us focus on transformations of form $Z = a + bg(Y)$ of the random variable Y, where a and $b > 0$ are two parameters, where g is a monotone increasing function, and where Z is a $(0, 1)$ normal random variable. It is very easy to show that the random variable Y has the following characteristics:

- cumulative distribution:
$$F_Y(y) = \Phi[a + bg(y)]$$
- inverse cumulative distribution:
$$F_Y^{-1}(\alpha) = g^{-1}\left[\frac{\Phi^{-1}(\alpha) - a}{b}\right]$$
- density function:
$$f_Y(y) = bg'(y)\phi[a + bg(y)]$$
- noncentral moments of order s:
$$m_s(Y) = \int_{-\infty}^{+\infty}\left[g^{-1}\left(\frac{z-a}{b}\right)\right]^s \phi(z)dz \tag{1}$$

If c and $d > 0$ are two additional parameters such that $Y = (X - c)/d$, then we can straightforwardly deduce the characteristics of the random variable X, i.e., $F_X(x) = F_Y[(x - c)/d]$ and $F_X^{-1}(\alpha) = c + dF_Y^{-1}(\alpha)$. There are a large number of possibilities for choosing an adequate function g. Johnson (1949) has proposed a very popular system of distributions based on a set of three different functions: