A Control Chart to Monitor a Multivariate Binomial Process

Una carta di controllo per monitorare un processo binomiale multivariato

Paolo Cozzucoli\textsuperscript{1}, Salvatore Ingrassia\textsuperscript{2}

\textsuperscript{1} Dipartimento di Economia e Statistica, Università della Calabria
\textsuperscript{2} Dipartimento di Economia e Metodi Quantitativi, Università di Catania

e-mail: cozzucoli@unical.it

Keywords: multivariate binomial process, two-sided control chart, ARL

1. Introduction

The most applied statistical methods for monitoring multivariate attribute processes have been developed assuming that they have a multinomial distribution, see e.g. Marcucci (1985) and Cassady and Nachlas (2006). However this assumption is not always reasonable; indeed, it is more general and correct to suppose that in each item it is possible to identify one or more of $k$ ordered and not mutually exclusive quality defects. In this case, the appropriate probabilistic model to monitor the process is the multivariate binomial distribution. Specifically, the sampling data may be modeled as coming from multivariate binomial distributions; let $X = (X_1,\ldots,X_k)$ be a $k$-component multivariate binomial random vector with binomial marginal distribution $X_i \sim B(n, p)$ and dependence structure specified by $D$ and we write $X \sim MVBin(p, n, D)$ with $p=(p_1,\ldots,p_k)$. The multivariate binomial distribution arises as follows: let $Y_1,Y_2,\ldots,Y_n$ be iid multivariate Bernoulli vectors $Y \sim MVBin(p,1,D)$, where $D$ denotes a particular probability distribution on the $2^k$ binary $k$-tuples subject to the constraint that $E(Y_i)=p$; then $X = \sum_{i=1}^{n} Y_i \sim MVBin(p, n, D)$, see e.g. Westfall and Young (1989).

2. An Index and a Control Chart for the Overall Defectiveness

Let $C = (C_1,\ldots,C_i,\ldots,C_k)$ be the vector of $k$ ordered quality defects. $C_i$ is the minimum unserious defect and $C_k$ is the maximum serious defect. Since different defects bring to the process different losses of quality then we may define a vector of weights that are numerical evaluations of the defectiveness degree of the product. In this case the sampling procedure, based on a sample of $n$ items, determines a vector $X = (X_1,\ldots,X_i,\ldots,X_k)$ of random variables associated with the $p = (p_1,\ldots,p_i,\ldots,p_k)$ probability vector. Specifically, $X_i$ is the number of items in the sample with the $C_i$ defect and $p_i$ is the probability of observing in the selected item the $C_i$ defect. Consequently, the multivariate random variable $X$ has a multivariate binomial distribution with parameters $(n, p)$. Let $c = (c_1,\ldots,c_i,\ldots,c_k)$ be a vector of increasing weights associated to the $C$ vector, where, $c_i < c_{i+1}$. In general, $c_i$ indicates the degree of quality loss that the $i-th$ defect introduces into the system. These weights may be
selected considering the dysfunction, dissatisfaction, economical loss, increasing costs, or demerit caused by the specific defect. So, for fixed $c$ and $p$ vectors, a normalized index of the overall defectiveness degree is $\xi = \sum_{i=1}^{k} c_i^t p_i$, where $\frac{c_i}{\sum c_i}$ measures the weighted degree of overall defectiveness; specifically, $\min(\xi) = 0$, if $p_i = 0, i=1,...,k$ and $\max(\xi) = 1$, if $p_i = 1, i=1,...,k$. Therefore, to monitor the overall defectiveness parameter $\xi$, we can use the following sampling statistic $\hat{\xi} = \sum_{i=1}^{k} c_i^t \hat{p}_i$, where $\hat{p}_i = \frac{X_i}{n}$. Following the Shewhart procedure, we can define a control chart for the sampling statistics $\hat{\xi}$. So, let $p_0 = (p_{0,1},..., p_{0,k})$ be the specified parameter vector under the in control process hypothesis. Usually, the vector $p_0$ is unknown and it is estimated using $m$ preliminary samples of size $n$ taken from the process in control. Let $X_t = (X_{1t},..., X_{kt},..., X_{nt})$, $t=1,2,...,m$, be a set of $m$ preliminary samples of size $n$ taken from the multivariate binomial process $X$ with parameters $(n,p)$. Specifically, $X_{it}$ is the number of items in the $t-th$ sample with $C_i$ defect. An unbiased estimator of $p_i$ parameter is $\bar{p}_i = \frac{1}{m} \sum_{t=1}^{m} X_{it}$. Therefore, the estimated control chart limits are $[UCL, LCL] = \left[ \sum_{i=0}^{k} c_i^t \bar{p}_i \pm 3\sqrt{\frac{1}{n}\left[ \left( \sum_{i=1}^{k} c_i^t \bar{p}_i \right)^2 - \left( \sum_{i=1}^{k} c_i \bar{p}_i \right)^2 \right]} \right]$. The proposed control chart signals the process deterioration when $\hat{\xi}$ is plotted out the UCL or the process improvement when $\hat{\xi}$ is plotted out the LCL. Finally, some considerations on the choice of the weights in vector $c$ are necessary. In fact, it is necessary to consider also the influence of the weights $c_i$ on the performance, in terms of ARL, of the corresponding control chart. Assuming that the loss of quality due to the $i-th$ defect is proportional to the loss of quality due to the others, we suggest to use weights that are in terms of a geometric progression; that is, $c_i = kc_{i-1}$. So, it is possible, almost from numerical evaluation, to evaluate the influence of the weights system on the ARL of the control chart.

References

