Controlling Non-Normal Multivariate Processes using Information Theoretic Control Charts

A. Sachlas, Takis Papaioannou, and S. Bersimis

Abstract

Nowadays one of the most rapidly developing areas of process control is *Multivariate Statistical Process Control* (*MSPC*). *MSPC* refers to process monitoring problems in which several related variables are of interest. The most widely used tool of *MSPC* is control charting, which may be used efficiently in order to monitor and control processes in a multivariate environment. Till today, the majority of the charts appeared in the literature require the assumption of multivariate normality. In cases, that the multivariate normal is not an appropriate model, the literature on alternative multivariate charting techniques is poor. The aim of this article is to present thoroughly the use of *Information Theory* to the area of *MSPC* as well as to introduce some valuable tools for controlling processes which are related to non-normal variables.

Index Terms

Quality Control; Process Control; Multivariate Statistical Process Control; Information Theory; Maximum Entropy; Shannon's Entropy.

I. Introduction

Multivariate Statistical Process Control (*MSPC*) is one of the most rapidly developing section of statistical process control, because today industrial processes are far too complex and sensitive than ever. Specifically, in most of the cases, the stability of the process is directly or indirectly related to many characteristics. Monitoring these process characteristics independently can be very misleading due to the fact that in most of the cases, these characteristics are strongly related to each other. Process monitoring problems in which several related variables are of interest are collectively known as *Multivariate Statistical Process Control*. The most valuable tool for controlling a multivariate process is the same as in the univariate case, the control chart.

The first who introduced *MSPC* techniques was Hotelling (1947). Since then a great number of authors have dealt with *MSPC* problems. Many authors gave multivariate extensions for almost all the kinds of univariate control charts, such as multivariate Shewhart type control charts, multivariate CUSUM control charts, and multivariate EWMA control charts (see e.g., Alt and Smith (1988), Crosier (1988), Pignatiello and Runger (1990), Lowry and Montgomery (1995)). For an extended overview of the literature in the area of *MSPC* charting techniques see Bersimis et al. (2006).

In most control procedures appearing in the literature, a multivariate normal distribution is assumed. This is a major drawback of any control procedure, because of the fact that in most practical problems hundreds of variables are of interest and the multivariate normal is not the best fitted model for them. A solution to this drawback may be the use of *Information Theory*, which is a flexible theory that provides solutions to a lot of problems in statistics (Brockett (1991), Ebrahimi (2000), Golan et al.(1996), Kapur (1989), Kullback (1959), Thomas and Cover (1991)).

The most celebrated tool of *Information Theory* which can be used in order to produce a data generating density approximation is *Maximum Entropy (ME)*. The use of this method demands only partial information about the data generating distribution in the form of moment constraints. The selected distribution is the one that maximizes *Shannon's Entropy* under these moment constraints. Alwan et al. (1998) gave a detailed framework for the use of *ME* in process control problems. Specifically, they described an information theoretic process control algorithm for monitoring process moments using the *ME* and *Minimum Discrimination Information (MDI)* principles of inference, aiming firstly at the univariate case and secondly at the multivariate case. Specifically, they just briefly described the use of their method in a multivariate environment.

In this paper we extend Alwan et al.'s (1998) procedure in the multivariate case in order to construct multivariate control charts with optimum properties when the underlying distribution of the process is not normal. This paper is organized as follows. Section 2 reviews briefly the most usually used techniques of MSPC. In Section 3 we formulate Alwan's procedure in the multivariate case, when the underlying distribution is not normal. In Section 4, we present the use of the information theoretic procedure and give a short numerical study of the properties of the chart introduced in Section 3, in case of multivariate log-normal data, pointing out the advantage of using one chart for controlling all the parameters of the process. Finally, some concluding remarks are given in Section 5.

II. MSPC Basics

This paper focuses on the simultaneous monitoring, through charting, of p distinct quality characteristics (variables) $x_1, x_2, ..., x_p$. The simpler *MSPC* charting technique is the *chi-square control chart*. We have already stated that the major drawback of most of the procedures presented till today is that a multivariate normal distribution is assumed. This assumption must be made also for the *chi-square control chart*. Thus, in order to establish *chi-square control chart*, we have to assume that the in-control joint probability distribution of the vector $\mathbf{x} = (x_1, x_2, ..., x_m)^T$ is the p-variate normal distribution with mean vector μ_0 and variance-covariance matrix $\mathbf{\Sigma}_0$, that is $\mathbf{x} \sim N_p(\mu_0, \mathbf{\Sigma}_0)$.

Thus, in that case rational subgroups of size $n > 1$ are inspected and the mean sample vector $\overline{\mathbf{x}}_i$ of the i-th subgroup is calculated. In a *chi-square control chart* for monitoring the process mean, the test statistic

$$D_i^2 = n(\overline{\mathbf{x}}_i - \mu_0)^T \mathbf{\Sigma}_0^{-1} (\overline{\mathbf{x}}_i - \mu_0), \quad i \geq 1$$

is plotted in the chart; this statistic represents the Mahalanobis distance between $\overline{\mathbf{x}}_i$ and μ_0 in the Euclidean space R^p and follows a chi-square distribution with p degrees of freedom ($D_i^2 \sim \chi_p^2$). The upper control limit of the *chi-square control chart* is given by $UCL = \chi_{p,a}^2$ where $\chi_{p,a}^2$ denotes the upper a percentage point of the chi-square distribution. If $D_i^2 > \chi_{p,a}^2$ there is evidence that the process is out-of-control due to assignable causes, otherwise the process is assumed to be in-control and no action is deemed necessary.

It is worth noting that, in a *chi-square control chart* there is no need for a lower control limit since the discrimination between the in-control and out-of-control states is determined by the magnitude of $D_i^2 \geq 0$; extreme values of the statistic D_i^2 indicate that the point $\overline{\mathbf{x}}_i$ is far away from μ_0 and therefore the process can be declared out-of-control, while small values of D_i^2 indicate that the point $\overline{\mathbf{x}}_i$ is close to μ_0 and therefore it is reasonable to assume that the process is in-control. Apparently, the performance of the *chi-square control chart* depends solely on the distance of the out-of-control mean from the in-control mean and not on the particular direction where the former is placed as compared to the latter (directional invariance).

In addition, it is worth here to note that this control chart is more sensitive to changes in the mean vector of the process. Thus, another control chart must be used in order to control process dispersion. In that case, keeping the assumption of multivariate normality and that there are samples of size $n > 1$ available, a chart for the process dispersion, presented by Alt (1985), may be based on the sequence of the following statistic

$$W_i = -pn + pn \ln n - n \ln \left[|\mathbf{A}_i| \, |\mathbf{\Sigma}_0|^{-1} \right] + trace \left(\mathbf{\Sigma}_0^{-1} \mathbf{A}_i \right)$$

for the i-th, $i = 1, 2, ...$ sample, where $\mathbf{A}_i = (n-1)\mathbf{S}_i$, and $\mathbf{S}_i$ is the sample variance-covariance matrix of the i-th rational subgroup. This statistic follows an asymptotic χ^2-distribution with $p \times (p+1)/2$ degrees of freedom (Anderson, 1958). Thus, the control chart for the process dispersion, with known mean vector μ_0 and known variance-covariance matrix $\mathbf{\Sigma}_0$ has upper control limit $\chi^2_{p(p+1)/2,\alpha}$. Hence, if the value of the test statistic W_i exceeds $\chi^2_{p(p+1)/2,\alpha}$, the chart signals a potential out-of-control process.

These two control charts are the most frequently used in the industry, worldwide. As we will see in the next section, by appropriate using information theoretic methodology we can combine these two charts in one single chart having the ability to detect changes in both the mean vector and the covariance matrix.

III. Integrating Information Theoretic Criteria in Multivariate SPC

Suppose that $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$, is a p-variate random vector with probability density function $f(\mathbf{x})$ or probability function $p(\mathbf{x})$ in the continuous and discrete case respectively. Thus, the entropy is defined by

$$H[f(\mathbf{x})] = -\int \ldots \int f(\mathbf{x}) \ln f(\mathbf{x})\, d\mathbf{x}.$$

What we have as prior information in the multivariate case are $p \times 1$ vectors and/or $p \times p$ matrices of moments and since we do not know the underlying distribution we shall seek the one that maximizes the entropy. This is the well known *ME* principle. Several multivariate distributions can be obtained as *ME* solutions with appropriate moments as we can see in Table 1. For example, for the multivariate normal distribution we need the $p \times 1$ mean vector μ and the $p \times p$ variance-covariance matrix $\mathbf{\Sigma}$ or for the multivariate gamma distribution we only need to know the p values of the arithmetic means and the p values of the geometric means of the variates. In some cases, in order to take a maximum entropy distribution, apart from the moments we have to ascribe a prior distribution to the variates as well (Kapur, 1989). Kapur (1989) provides details for the derivation of a lot of univariate and multivariate distributions while Zografos (1998) gave the maximum entropy characterization of Pearson's type II and VII multivariate distributions, from which multivariate t-distribution and its special case - multivariate Cauchy distribution - can be derived.

The class of the multivariate distributions with moment constraints is

$$\Omega_0 = \left\{ f(\mathbf{x}|\omega_0) : \int \ldots \int f(\mathbf{x}|\omega_0) g_k(\mathbf{x})\, d\mathbf{x} = \omega_{k0},\ k = 0, 1, 2, ..., K \right\},$$

which consists of many distributions that satisfy the constraints. In order to find the *ME* distribution i.e. the model for the in-control state, we maximize $H[f(\mathbf{x}|\omega_0)]$ with respect to f over Ω_0. The solution, whenever it exists, has the form

$$f_0^*(\mathbf{x}|\omega_0) = C(\omega_0) \exp\{\eta(\omega_0) g_k(\mathbf{x})\},$$

where $C(\omega_0)$ is a normalizing constant and $\eta(\omega_0)$ is the vector of Lagrangian multipliers.

In the *monitoring state*, the class of multivariate distributions is

$$\Omega_t = \left\{ f(\mathbf{x}|\mathbf{m}_t) : \int \ldots \int f_t(\mathbf{x}|\mathbf{m}_t) g_k(\mathbf{x})\, d\mathbf{x} = \mathbf{m}_{kt}, k = 1, 2, ..., K \right\},$$

which has the same form as that in the in-control state. In this stage, the *ME* procedure selects the multivariate distribution belonging to Ω_t that is closest to the $f_0^*(\mathbf{x}/\omega_0)$.

In order to measure the discrepancy between any distribution F_t and the in-control distribution F_0^* we use the Kullback-Leibler discrimination information function

$$K(f_t : f_0^*) = E_{f_t}\left[\ln \frac{f_t(\mathbf{x})}{f_0^*(\mathbf{x})}\right] = \int \ldots \int f_t(\mathbf{x}) \ln \frac{f_t(\mathbf{x})}{f_0^*(\mathbf{x})}\, d\mathbf{x}.$$

This function is continuous and convex for nonnegative $f_t(\mathbf{x})$ and equals $\mathbf{0}$ if and only if the equation $f_t(\mathbf{x}) = f_0^*(\mathbf{x})$ holds almost everywhere.

The relationship (Equation 4) that Alwan et al. (1998) showed about the Kullback-Leibler discrimination information function and the entropy, holds in the multivariate case as well. So we have

$$K(f_t : f_0^*) \quad = \quad -H[f_t(\mathbf{x}|\mathbf{m}_t)] - \ln C(\omega_0) - \eta(\omega_0)\mathbf{m}_t, \tag{1}$$

which means that instead of minimizing $K(f_t : f_0^*)$ over Ω_t we can maximize the entropy $H[f_t(\mathbf{x}|\mathbf{m}_t)]$. The form of the MDI solution is the same as that of the ME distribution in the in-control state, i.e.

$$f_t^*(\mathbf{x}|\mathbf{m}_t) = C(\mathbf{m}_t)\exp\{\eta(\mathbf{m}_t)g_k(\mathbf{x})\},\ t = 1, 2, ...,$$

where $C(\mathbf{m}_t)$ is a normalizing constant and $\eta(\mathbf{m}_t)$ is the vector of Lagrangian multipliers. This model is the closest information theoretic distribution to the model $f(\mathbf{x}|\omega_0)$ which satisfies constraints (Kullback, 1959). We observe that the two estimates of the unknown distribution belong to the same parametric families and they only differ on the value of their parameters. This provides the grounds to use information theory for constructing control charts. The parameters of the ME model in the in-control state are given by the initial moment's vector ω_0 while in the monitoring state the parameters of the MDI distribution are given by the data generating moment's vectors $\mathbf{m}_t$, $t = 1, 2, ...$. The Kullback-Leibler discrimination information function in each stage measures the discrepancy between monitoring state's model $f_t^*(\mathbf{x}|\mathbf{m}_t)$ and the in-control state model $f_0^*(\mathbf{x}|\omega_0)$. So when this function takes a "sufficiently" small value, then the process is judged in control.

Evaluating the maximum entropy value

$$H[f_t^*(\mathbf{x}|\mathbf{m}_t)] = -\ln C(\mathbf{m}_t) - \eta(\mathbf{m}_t)\mathbf{m}_t$$

and substituting the above equation in equation (1), we take the MDI control function for monitoring the multivariate parameter shifts as

$$I_{m_t} \quad = \quad 2nK(f_t^* : f_0^*) = 2n\left\{\ln\frac{C(\mathbf{m}_t)}{C(\omega_0)} + [\eta(\mathbf{m}_t) - \eta(\omega_0)]\mathbf{m}_t\right\}. \tag{2}$$

The function I_{m_t} is the minimum information discrepancy between the two estimates of the unknown distribution of the process variable $f(\mathbf{x}|\omega_0)$ and is a function of vectors ω_0 and $\mathbf{m}_t$. Because $K(f_t^* : f_0^*) = 0$ if and only if $f_t^*(\mathbf{x}|\mathbf{m}_t) = f_0^*(\mathbf{x}|\omega_0)$, $I_{m_t} = 0$ means that the process distribution is unchanged and the process is in control. However, when the function takes a value which is beyond the control limit, this means that the values of the moments are significantly different and the process is judged out of control.

Under some regularity conditions (Kullback, 1959, p. 98), the information control function I_{m_t} asymptotically follows a χ^2 distribution with K degrees of freedom, where K depends on the number of the distributions' parameters.

As we have already mentioned, in order to test whether the process is in-control we use the MDI principle, i.e. I_{m_t} given by equation (2). The MDI function $K(f_t^* : f_0^*)$ as well as the moment equations for the multivariate distributions given in Table I are given in Table II.

IV. Numerical Experimentation

In this section, we illustrate the use of the proposed method. Specifically, we present the appropriate procedure for controlling processes that may be modeled as multivariate log-normal distribution. Moreover, we discuss in detail, an extensive numerical study that was performed in order to examine the overall performance of the proposed methodology. Additionally, in order to assess the performance of the proposed control procedure we compare its performance with the performance of the chi-square control chart. This control chart is one of the most frequently used for the specific problem. These comparisons were conducted by the aid of a simulation study. Specifically, we applied the competing methods to 10000 simulated data.

Let us consider a process involving three continuous quality characteristics $(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3)$, where $Y_i = Exp(X_i)$, $i = 1, 2, 3$ and that the stability of this process is evaluated in terms of the values of these three characteristics. Assume that the vector $\mathbf{x} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ follows a $N_3(\mu_0, \mathbf{\Sigma}_0)$, with in control mean $\mu_0 = (0, 0, 0)$ and variance-covariance matrix

$$\mathbf{\Sigma}_0 = \begin{bmatrix} 1 & & \\ -.7 & 1 & \\ -.4 & .3 & 1 \end{bmatrix}.$$

TABLE I

MULTIVARIATE MAXIMUM ENTROPY DISTRIBUTIONS

Parameters of interest	Distribution
Means $\omega_{1i} = E(r_i),\ i = 1, 2, ..., p,\ \sum_{i=1}^{p} r_i = n$	Multinomial
Means $\omega_{1i} = E(r_i),\ i = 1, 2, ..., p,\ \sum_{i=1}^{p} q_i < 1$	M. Geometric
Means $\omega_{1i} = E(r_i),\ i = 1, 2, ..., p$	M. Poisson
Mean vector μ and variance-covariance matrix $\mathbf{\Sigma}$	M. Normal
Variance-covariance matrix $\mathbf{\Sigma}$	M. Normal
Mean vector μ and second order moments about the origin	M. Normal
$E(\ln Y_i),\ E[(\ln Y_i)^2],\ E(\ln Y_i \ln Y_j),\ i = 1, 2, ..., p,\ j = 1, 2, ..., p$	M. Log-normal
$\omega_1 = E[\ln\{\frac{X_1}{\theta_1} + \frac{X_2}{\theta_2} + ... + \frac{X_p}{\theta_p} + 1\}],\ x_i \geq \theta_i$	M. Pareto
$\omega_{1i} = E(X_i),\ \omega_{2i} = E(\ln X_i),\ i = 1, 2, ..., p$	M. Gamma
$\omega_1 = E[\ln[1 + v^{-1}(\mathbf{X} - \mu)^T \mathbf{\Sigma}^{-1}(\mathbf{X} - \mu)]]$	M. t
$\omega_1 = E[\ln[1 + \mathbf{X}^T \mathbf{X}]]$	M. Cauchy
$\omega_1 = E[\{(\mathbf{X} - \mu)^T \mathbf{\Sigma}^{-1}(\mathbf{X} - \mu)\}^s]$ and $\omega_2 = E[\ln[(\mathbf{X} - \mu)^T \mathbf{\Sigma}^{-1}(\mathbf{X} - \mu)]]$	M. Kotz type
$\omega_1 = E[\ln(1 + \sum_{i=1}^{p} d_i \mathbf{X}_i^{c_i})]$ and $\omega_2 = E[\ln(1 + d_i^{\frac{1}{\gamma_i}} \mathbf{X}_i)]$	M. Burr with $\alpha = 1$
$\omega_1 = E[\ln(1 + \sum_{i=1}^{p} (\frac{\mathbf{X}_i - \lambda_i}{\theta_i})^{\frac{1}{\gamma_i}})]$ and $\omega_2 = E[\ln(\frac{\mathbf{X}_i - \lambda_i}{\theta_i})]$	M. Pareto III

TABLE II

MDI FUNCTION FOR SEVERAL MULTIVARIATE DISTRIBUTIONS

ME Distribution	Moment Equations	MDI Function $K(f_t^* : f_0^*)$
M. Gamma	$\omega_{2i} = \psi(m_{i+1})$	$\sum_{i=1}^{p} \ln \frac{\Gamma(\omega_i)}{\Gamma(m_i)} + (\omega_{2,i+1} - m_{2,i+1})\psi(m_{2,i+1})$
M. Poisson	$\omega_{1i} = q_i \left(1 + \sum_{i=1}^{p} q_i\right) / \left(\sum_{i=1}^{p} q_i\right)$	$\sum_{i=1}^{p} \ln \frac{q_{ti}}{q_{0i}} \mathbf{m}_{1t}$
M. Pareto	$\omega_{1i} = \sum_{i=1}^{p} \frac{1}{a+i-1}$	$\sum_{i=1}^{p} \ln \frac{\omega_{i1}}{m_{i1}} - \sum_{i=1}^{p} \ln \frac{\omega_2+i-1}{m_2+i-1} - (m_2+n)\mathbf{m}_{1t} + (\omega_2+n)E_{f_t^*}\omega_{10}$
M. Geometric	$\omega_{1i} = \frac{q_i}{1-\sum_{i=1}^{p} q_i}$	$\ln \frac{(1-q_{t1}-q_{t2}-...-q_{tp})}{(1-q_{01}-q_{02}-...-q_{0p})} + \sum_{i=1}^{p} \ln \frac{q_{ti}}{q_{0i}} \mathbf{m}_{1t}$
M. Normal	$\omega_1 = \mu,\ \omega_2 = \mathbf{\Sigma}$	$\frac{1}{2}(\mathbf{m}_{1t} - \omega_{10})^T \omega_{20}^{-1}(\mathbf{m}_{1t} - \omega_{10}) + \frac{1}{2}\left(tr(\omega_{20}^{-1}\mathbf{m}_{2t}) - \ln\left\|\omega_{20}^{-1}\mathbf{m}_{2t}\right\| - p\right)$
M. Log-normal	$\omega_1 = E[\ln \mathbf{Y}] = \mu,\ \omega_2 = \mathbf{\Sigma}$	$\frac{1}{2}(\mathbf{m}_{1t} - \omega_{10})^T \omega_{20}^{-1}(\mathbf{m}_{1t} - \omega_{10}) + \frac{1}{2}\left(tr(\omega_{20}^{-1}\mathbf{m}_{2t}) - \ln\left\|\omega_{20}^{-1}\mathbf{m}_{2t}\right\| - p\right)$
Multinomial	$\omega_{1i} = Np_i$	$\sum_{i=1}^{p} \ln \frac{p_{ti}}{p_{0i}} \mathbf{m}_{1t}$
M. Pearson II	$\omega_{1i} = \psi(m+1) - \psi(m+1+\frac{p}{2})$	$\ln \frac{\Gamma(m+1)\Gamma(\frac{p}{2}+\hat{m}+1)}{\Gamma(\hat{m}+1)\Gamma(\frac{p}{2}+m+1)} + \frac{1}{2}\ln\left\|\mathbf{\Sigma}_0\mathbf{S}_t^{-1}\right\| + (\hat{m}-m)\mathbf{m}_{1t}$
M. Pearson VII	$\omega_{1i} = \psi(m) - \psi(m-\frac{p}{2})$	$\ln \frac{\Gamma(m-\frac{p}{2})\Gamma(\hat{m})}{\Gamma(m)\Gamma(\hat{m}-\frac{p}{2})} - \frac{1}{2}\ln\left\|\mathbf{\Sigma}_0\mathbf{S}_t^{-1}\right\| - (\hat{m}-m)\mathbf{m}_{1t}$
M. t	$\omega_{1i} = \psi(\frac{v+p}{2}) - \psi(\frac{v}{2})$	$\ln\left\|\mathbf{\Sigma}_0\mathbf{S}_t^{-1}\right\|$
M. Cauchy	$\omega_{1i} = \psi(\frac{p+1}{2}) - \psi(\frac{1}{2})$	$\ln\left\|\mathbf{S}_t^{-1}\right\|$

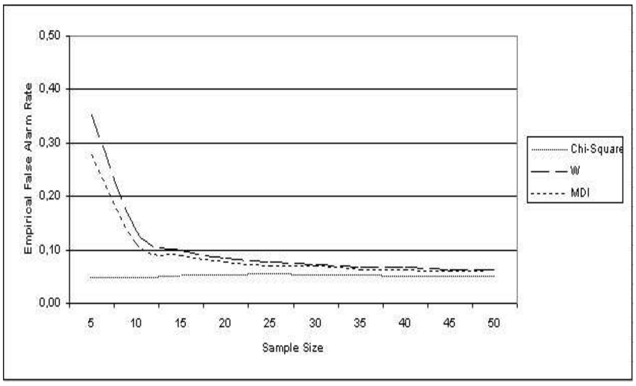
In this way vector $\mathbf{y}$ follows a $LN_3(\mathbf{v}_0, \mathbf{D}_0)$ (Kotz et al., 2000).

From Table I we have that the parameters of interest in the log-normal setting are $E[\ln Y_i]$, $E[(\ln Y_i)^2]$ and $E[\ln Y_i \ln Y_j]$, $i, j = 1, 2, 3$. However, we have that $E[\ln Y_i] = E[X_i]$, $E[(\ln Y_i)^2] = E[X_i^2]$ and $E[\ln Y_i \ln Y_j] = E[X_i X_j]$, $i, j = 1, 2, 3$. So, controlling the process for the parameters $E[\ln Y_i]$, $E[(\ln Y_i)^2]$ and $E[\ln Y_i \ln Y_j]$ is equivallent to controlling the process for the parameters of the multivariate normal distribution.

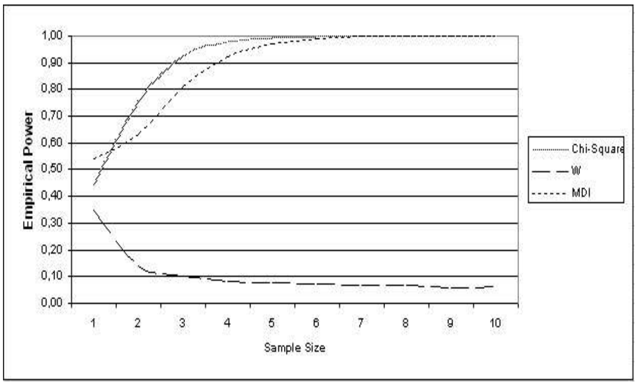
In order to monitor the three dimensional process and identify a probable shift in the process mean (i.e. the parameter $E[\ln Y_i]$ of the log-normal distribution) using the chi square control chart we plot the following statistic

$$D_i^2 = n(\overline{\mathbf{x}}_i - \mu_0)^T \mathbf{\Sigma}_0^{-1}(\overline{\mathbf{x}}_i - \mu_0), \quad i \geq 1$$

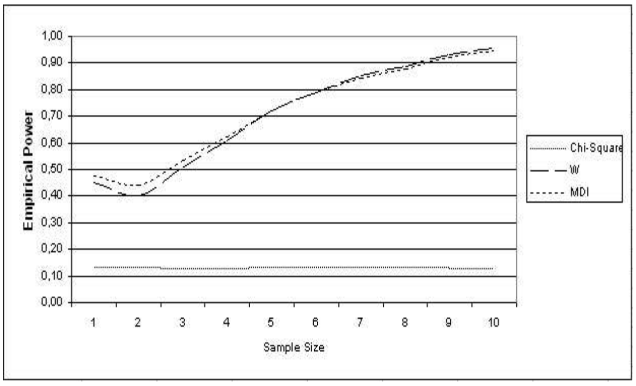
against an appropriate control limit which is given by $UCL = \chi^2_{p,a}$, where as already presented $\chi^2_{p,a}$



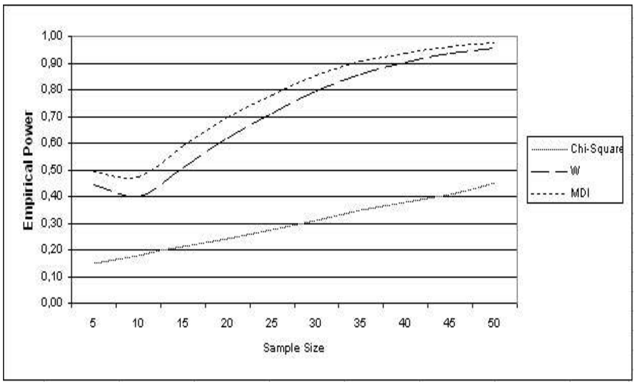
(a) In-Control state



(b) Mean Changes



(c) Covariance Matrix Changes



(d) Changes at Both Parameters

Fig. 1. Empirical ARL

denotes the upper a percentage point of the chi-square distribution.

Additionally, if someone wants to detect changes in the process dispersion (i.e. the parameters $E[(\ln Y_i)^2]$ and $E[\ln Y_i \ln Y_j]$ of the log-normal distribution) may chart the following statistic

$$W_i = -pn + pn \ln n - n \ln \left[|\mathbf{A}_i| \, |\boldsymbol{\Sigma}_0|^{-1} \right] + trace \left(\boldsymbol{\Sigma}_0^{-1} \mathbf{A}_i \right).$$

Alternatively, we may monitor and control the three dimensional process, both for changes in the mean vector and the covariance structure, by plotting the MDI function, under the multivariate normal setup, which is of the form

$$MDI_i = \frac{1}{2} (\overline{\mathbf{x}}_i - \mu_0)^T \boldsymbol{\Sigma}_0^{-1} (\overline{\mathbf{x}}_i - \mu_0) + \frac{1}{2} \left(tr(\boldsymbol{\Sigma}_0^{-1} \mathbf{A}) - \ln \left| \boldsymbol{\Sigma}_0^{-1} \mathbf{A} \right| - p \right)$$

against an appropriate control limit which is given by $UCL = \chi^2_{p+p(p+1)/2,a}$, where $\chi^2_{p+p(p+1)/2,a}$ denotes the upper a percentage point of the chi-square distribution with $p + p(p+1)/2$ degrees of freedom. This control chart is a Phase II control chart for controlling the mean vector and has the advantage that may capture both changes in the mean level and the dispersion of the process.

In Table III we give the power of the new control chart in detecting either shifts in the mean vector or changes in the covariance matrix. Specifically, we compare our methodology to the χ^2 control chart as well as to the control chart for the process dispersion, for various choices of the sample size n and the non-centrality parameter $\lambda^2(\mu_1)$, which is given by $\lambda(\mu_1) = \sqrt{n(\mu_1 - \mu_0)^T \boldsymbol{\Sigma}_0^{-1} (\mu_1 - \mu_0)} = \sqrt{n \delta^T \boldsymbol{\Sigma}_0^{-1} \delta}$, where $\mu_1 = \mu_0 + \delta$ is a specific out-of-control mean vector, keeping in mind that the $\boldsymbol{\Sigma}_0$ is still in-control.

As we may observe, from Table III and the equivalent plots in Figure 1 in the in-control state the empirical false rate increases as n increases and tends to 5% given that MDI follows asymptoticaly the χ^2 distribution. Of course the false rate of the chi-square chart is better than that of MDI and W as it

TABLE III

EMPIRICAL FALSE RATES AND EMPIRICAL POWER OF THE PROPOSED CHART

	In control					*Shift in the Mean*			
Sample Size	MDI	χ^2	W	λ	Sample Size	MDI	χ^2	W	λ
5	0,278	0,050	0,353	0	5	0,543	0,443	0,350	2,250
10	0,111	0,048	0,135	0	10	0,634	0,753	0,139	3,182
15	0,089	0,053	0,100	0	15	0,813	0,930	0,103	3,897
20	0,077	0,054	0,086	0	20	0,923	0,983	0,084	4,500
25	0,071	0,055	0,076	0	25	0,973	0,997	0,078	5,031
30	0,071	0,053	0,072	0	30	0,991	1,000	0,072	5,511
35	0,064	0,054	0,069	0	35	0,998	1,000	0,068	5,952
40	0,064	0,051	0,067	0	40	1,000	1,000	0,066	6,363
45	0,061	0,051	0,063	0	45	1,000	1,000	0,060	6,749
50	0,063	0,050	0,064	0	50	1,000	1,000	0,062	7,115
	Shift in the Covariance Matrix					*Shift at both parameters*			
Sample Size	MDI	χ^2	W	λ	Sample Size	MDI	χ^2	W	λ
5	0,477	0,134	0,453	0	5	0,493	0,153	0,446	0,770
10	0,441	0,132	0,401	0	10	0,474	0,182	0,399	1,089
15	0,534	0,129	0,508	0	15	0,590	0,215	0,507	1,334
20	0,621	0,129	0,607	0	20	0,697	0,243	0,619	1,541
25	0,719	0,131	0,720	0	25	0,781	0,276	0,710	1,723
30	0,787	0,133	0,791	0	30	0,854	0,314	0,796	1,887
35	0,846	0,132	0,854	0	35	0,906	0,349	0,859	2,038
40	0,880	0,134	0,889	0	40	0,939	0,379	0,900	2,179
45	0,922	0,134	0,931	0	45	0,961	0,412	0,935	2,311
50	0,945	0,126	0,954	0	50	0,974	0,452	0,954	2,436

follows the exact χ^2 distribution. Comparing the two asymptotic charts the MDI gives equivalent results to that of W.

When we have shifts in the mean vector, the MDI chart gives equivalent results to those of the chi-square chart. The W chart does not give good results as it targets to the process dispersion. As far as shifts in the covariance matrix are concerned the MDI chart gives equivalent results to those of the W chart while the chi-square chart fails as it targets to the mean of the process. Finally, in the case of shifts at both the parameters the proposed MDI chart is at least equivalent to the W chart. Logically, the chi-square chart fails again as it targets only to the mean of the process.

V. CONCLUSIONS

In this article, we gave a thorough presentation of the use of information theory in the field of multivariate SPC. As we showed, in the case where the controlling process is modeled as multivariate log-normal distribution, the performance of the control charts that are based on information theoretic criteria have at least equivalent empirical power to the corresponding control charts that are used today. Their major advantage is their good performance when we have changes at both the mean vector and the covariance matrix in the normal case or the parameters $E[\ln Y_i]$, $E[(\ln Y_i)^2]$ and $E[\ln Y_i \ln Y_j]$ in the log-normal case. This means that we can monitor a process for all the parameters with a single chart. Thus the information theoretic methodology integrate two charts in one single chart having the ability to detect changes at both the mean vector and the covariance matrix.

Information theoretic control charts are not a new idea as Alwan et al. (1998) where the first to introduce them in the case of normal multivariate processes. In this paper we tried to show their good performance in non-normal multivariate processes suchs as the multivariate log-normal case. This case is under further investigation by the authors and further results will be reported soon.

REFERENCES

[1] F.B. Alt, Multivariate Quality Control. The Encyclopedia of Statistical Sciences, Kotz S., Johnson, NL, Read CR (eds), New York: John Wiley, pp. 110-122, 1985.

[2] F.B. Alt and N.D. Smith, Multivariate Process Control. Handbook of Statistics, P.R. Krishnaiah and C.R. Rao (eds), Elsevier Science Publishers: North- Holland V.7, 333-351, 1988.
[3] L.C. Alwan, N. Ebrahimi and E.S. Soofi, Information theoretic framework for process control, European Journal of Operational Research 111, 526-542, 1998.
[4] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, 2nd ed, New York: John Wiley, 1958.
[5] S. Bersimis, S. Psarakis and J. Panaretos, Multivariate Statistical Process Control Charts: An Overview, Quality and Reliability Engineering International, in press, 2006.
[6] P.L. Brockett, Information-theoretic approach to actuarial science: A unification and extention of relevant theory and applications. Transactions of the Society of Actuaries 43, 73-135, 1991.
[7] R.B. Crosier, Multivariate generalizations of cumulative sum quality-control schemes, Technometrics, 30, 291-303, 1988.
[8] N. Ebrahimi, The maximum entropy method for lifetime distributions. Sankya: The Indian Journal of Statistics 62 A, 236-243, 2000.
[9] A. Golan, G. Judge and D. Miller, Maximum entropy econometrics: robust estimation with limited data. Wiley, New York, 1996.
[10] JL. Guerrero, Multivariate mutual information. Communications in Statistics - Theory and Methods, 23:1319-1339, 1994.
[11] JL. Guerrero, Testing variability in multivariate quality control: A conditional entropy measure approach. Information Sciences, 86: 179-202, 1995.
[12] H. Hotelling, Multivariate quality control, illustrated by the air testing of sample bombsights, Techniques of Statistical Analysis (eds., C. Eisenhart, M.W. Hastay, and W.A. Wallis), 111-184, McGraw-Hill, New York, 1947.
[13] J.N. Kapur, Maximum-Entropy Models in Science and Engineering. Wiley, New York, 1989.
[14] S. Kotz, N. Balakrishnan and N.L. Johnson, Continuous multivariate distributions. Wiley, New York, 2000.
[15] S. Kullback, Information Theory and Statistics. Wiley, New York, 1959.
[16] S. Kullback and R.A. Leibler, On Information and Sufficiency, Annals of Mathematical Statistics 22, 79-86, 1951.
[17] C.A. Lowry and D.C. Montgomery, A review of multivariate control charts, IIE Transactions, 27, 800-810, 1995.
[18] J.J.Jr. Pignatiello and G.C. Runger, Comparisons of multivariate CUSUM charts, Journal of Quality Technology, 22, 173-186, 1990.
[19] J. Thomas and T. Cover, Elements of Information Theory, John Wiley & Sons, New York, 1991.
[20] K. Zografos, On maximum entropy characterization of Pearson's type II and VII multivariate distributions. J. Multivariate Analysis 71, 67-75, 1998.