An Improved $|S|$ Control Chart for Multivariate Process Variability Monitoring based on Cornish-Fisher Correction

Emanuel Pimentel Barbosa, Mario Antonio Gneri and Ariane Meneguetti
Depto Estatística, Imecc/UNICAMP, Cidade Universitária Zeferino Vaz
Distrito de Barão Geraldo, CEP 13083-859, Campinas, SP, Brazil
e-mail: emanuel@ime.unicamp.br (correspondent author)

Abstract

This paper presents an improved version of the generalized variance $|S|$ control chart for multivariate process dispersion monitoring, based on the Cornish-Fisher formula for non-normality correction of the usual normal based 3-sigma limits chart. The exact sample distribution of $|S|$ doesn't have a simple known form for dimension $p > 2$, and we show here that the information from its 3^{rd} and 4^{th} order moments or cumulants are sufficient for a satisfactory approximation.

The performance of this corrected control chart is compared (in terms of false alarm risk) with the original normal based chart and the exact distribution based chart (for $p = 2$ and $p = 3$) where in the last case ($p = 3$), the exact distribution is obtained by simulation methods. This study shows that the control limits corrections do remove the drawback of excess of false alarm associated with the traditional normal based $|S|$ control chart. Finally, the proposed new chart is illustrated with a numerical example of application with real data.

KEY WORDS: Cornish-Fisher, False Alarm, Generalized Variance, Multivariate Process, Variability Monitoring.

1 INTRODUCTION

The modern statistical process control took place when Walter A. Shewhart (1926) developed the concept of a control chart based on the monitoring of the process mean level ($\bar{X}$ chart) and process dispersion (R or S charts). In the multivariate setting (Fuchs and Kenett, 1998), the basic monitoring tools for process level and process variability, are respectively the Hotelling T^2 statistic (Hotelling, H., 1931, 1947) and the statistics based on the sample variance-covariance S matrix.

The two more important and used statistics based on S, for process variability monitoring and testing, are the likelihood-ratio LR and the generalized variance $|S|$ (Wilks, S., 1932; Anderson, T.W., 1958; Alt, F.A., 1984; Aparisi, F. et al, 1999).

The study of control charts for process variability monitoring based on the $|S|$ statistic has received attention in the literature (for instance, Alt, F.A., 1984; Aparisi, F. et al, 1999; Djauhari, M.A., 2005 and others), and is the object of the present paper.

Although this statistic is considered simpler than the LR for applications in control charts, there are two drawbacks in its use for process monitoring. One is about its theoretical properties (see Johnson and Wichern, 1982 or Aparisi et al, 1999) and the other is practical, about its computational implementation. The theoretical limitations can be alleviated with the joint use of $|S|$ and univariate charts (see Alt, F.A., 1984). And how to overcome the practical difficulty is shown here in detail.

The question is that the exact distribution of the sample $|S|$ does not have a simple form for dimension $p > 2$ (Mathai, A.M., 1972; Pham-Gia and Turkkan, 2008) and it is common to approximate it by normal distribution, as in Djauhari, M.A.(2005) and others, considering a sort of 3-sigma limits (Shewhart) control charts. However, this sort of approximation, although used in practice, is very bad, since it has a very serious drawback: it produces a severe increase in the false alarm risk (what is shown in this paper at section 4).

The solution we present here is to work with a good approximation to the exact distribution of $|S|$, correcting its non-normality through the Cornish-Fisher formula (Cornish & Fisher, 1960; Lee & Lee, 1992). The information provided by the 3^{rd} and 4^{th} order moments or cumulants of the $|S|$ sample distribution will be sufficient for a satisfactory approximation, avoiding the drawback of false alarm increase.

The organization of the paper is the following. The standard normal based $|S|$ control chart and corresponding moments formulae are reviewed at section 2. The new $|S|$ control chart based on Cornish-Fisher correction is presented at section 3 with special emphasis to the cases of dimensions $p = 2$ and $p = 3$. A false alarm comparative study to show the advantages of the new chart in relation to the traditional chart, is presented at section 4 where the exact reference distribution, in the case of $p = 3$, is obtained by simulation. The proposed new chart is illustrated with a numerical example of application with real data at section 5. Final comments and conclusions are presented at section 6, followed by the references.

2 The Normal-based $|S|$ Control Chart

2.1 The Sample $|S|$ and Its Basic Properties

(i) **notation and definition:** Let $(X_1, X_2, ..., X_n)$ be a random sample from a p-variate normal with parameters μ and Σ, for $n = 2, 3, ...$ and $p = 2, 3, ...$ Then, the statistics *sample mean* $\bar{X}$ and *sample variance-covariance matrix* S, are:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad ; \qquad S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T$$

The sample generalized variance, is $|S| = det(S)$, where det is the determinant function.

(ii) **distribution:** From Anderson, T.W. (1958, 1984), it is known that,

$$|S| \sim \frac{|\Sigma|}{(n-1)^p} Z_1 Z_2 ... Z_{\text{p}} \qquad \text{where} \quad Z_k \sim \chi^2_{n-k} \quad \text{independent,} \quad k = 1, 2, ..., p$$

As a consequence, when $p = 2$, we have (Alt, F.A., 1984; Aparisi, F. et al, 1999), $|S| \sim \frac{|\Sigma|}{4(n-1)^2}(\chi^2_{2n-4})^2$. For $p > 2$, the $|S|$ distribution can be obtained numerically by simulation of S, using a Wishart generator algorithm (Smith & Hocking, 1972) through the software Matlab or R for instance.

Also, since each of the p chi-square variables in the distribution of $|S|$ can be expressed in terms of the meijer G function (Springer, M., 1979), with density given by

$$h_{Z_i}(z) = \frac{1}{2\,\Gamma(\frac{n-i-1}{2})} G^{\,1\;\;0}_{\,0\;\;1}\left(\tfrac{z}{2}\,|\,\tfrac{(n-i-2)}{2}\right) \quad i = 1, 2, ..., p \quad ,$$

then, the product of these independent G functions is given by

$$h(y) = \frac{1}{2^p}\left(\prod_{i=1}^{p} \frac{1}{\Gamma(\frac{n-i-1}{2})}\right) G^{\,p\;\;0}_{\,0\;\;p}\left(\tfrac{y}{2^p}\,|\,\tfrac{(n-3)}{2}, \tfrac{(n-4)}{2}, ..., \tfrac{(n-p-2)}{2}\right) \quad , \; y > 0$$

where $Y = Z_1 Z_2 ... Z_p$. See Springer, M. (1979) for the definition and properties of the G function.

It can be implemented, using the recent version of the Symbolic Math Toolbox of Matlab. The exact quantile of interest for the one-sided $|S|$ control chart is the value $y_0 = \frac{|\Sigma| x_0}{(n-1)^p}$ such that

$$\int_0^{x_o} h(y)dy = 1 - \alpha$$

(iii) **moments (ordinary and central):** From the general expression above for $|S|$, and using moment formulas for χ^2 r.v., it is obtained the r$^{\text{th}}$ (ordinary) moment formulae:

$$\alpha_r = \mathbb{E}\left(|S|^r\right) = \left(\frac{2}{n-1}\right)^{pr} |\Sigma|^r \prod_{k=1}^{p} \frac{\Gamma(r + \frac{n-k}{2})}{\Gamma(\frac{n-k}{2})}$$

In particular, the first four ordinary moments are given by:

$$\alpha_1 = \mathbb{E}\left(|S|\right) = \frac{(n-2)}{(n-1)}|\Sigma| \quad ; \quad \alpha_2 = \mathbb{E}\left(|S|^2\right) = \frac{(n+1)n(n-2)}{(n-1)^3}|\Sigma|^2$$

$$\alpha_3 = \mathbb{E}\left(|S|^3\right) = \frac{(n+3)(n+2)(n+1)n(n-2)}{(n-1)^5}|\Sigma|^3$$

$$\alpha_4 = \mathbb{E}\left(|S|^4\right) = \frac{(n+5)(n+4)(n+3)(n+2)(n+1)n(n-2)}{(n-1)^7}|\Sigma|^4$$

The central moments are given by:

$$\mu_k = \mathbb{E}\left(|S| - \mu\right)^k = \sum_{r=0}^{k} \alpha_r \binom{k}{r}(-\mu)^{k-r} \text{ ; for } k = 1, 2, 3, 4, \text{ we have,}$$

$$\mu = \alpha_1 = \mathbb{E}\left(|S|\right) \quad , \quad \mu_2 = \alpha_2 - \alpha_1^2 = Var\left(|S|\right)$$

$$\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 \quad , \quad \mu_4 = \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4$$

(iv) **cumulants:** From the Cumulants Generating Function (logarithm of the moments generating function), with the central moments previouslly standardized, we have:

$$K_1 = \mu^* = \frac{\alpha_1}{\sigma} \quad , \quad K_2 = \mu_2^* = \frac{\mu_2}{\sigma^2} \quad , \quad K_3 = \mu_3^* = \frac{\mu_3}{\sigma^3} \quad , \quad K_4 = \mu_4^* - 3\mu_2^{*2} = \frac{\mu_4}{\sigma^4} - 3\left(\frac{\mu_2}{\sigma^2}\right)^2$$

where $\sigma = \sqrt{\mu_2}$.

2.2 Normal $|S|$ Control Chart: Limits and False Alarm Risk

(i) **limits:** From the moments formulae of section 2.1, the 3-sigma control limits for the $|S|$ chart, in the case of $p = 2$, is given by,

$$\mathbb{E}(|S|) \pm 3\sqrt{Var(|S|)} = |\Sigma|\left(b_1 \pm 3\sqrt{b_2}\right) \text{ , where } b_1 = \frac{(n-2)}{(n-1)} \text{ and } b_2 = \frac{(n-2)(4n-2)}{(n-1)^3}$$

In the general p-dimensional case, we have the same expression for the control limits,

$$LCL = |\Sigma|\left(b_1 - 3\sqrt{b_2}\right) \quad ; \quad UCL = |\Sigma|\left(b_1 + 3\sqrt{b_2}\right)$$

but now, b_1 and b_2 are given by,

$$b_1 = \frac{1}{(n-1)^p}\prod_{i=1}^{p}(n-i) \quad ; \quad b_2 = b_1\left[\prod_{i=1}^{p}\frac{(n-i+2)}{(n-1)^p} - b_1\right]$$

obtained from the general moment formula of section 2.1, after some algebra (see also, Djauhari, M., 2005).

The corresponding empirical limits are given by

$$LCL = |\widehat{\Sigma}|\left(b_1 - 3\sqrt{b_2}\right) \quad ; \quad UCL = |\widehat{\Sigma}|\left(b_1 + 3\sqrt{b_2}\right)$$

where $|\widehat{\Sigma}| = |\bar{S}|/b_1$ and $\bar{S}$ is the matrix of average variances and average covariances based on the m calibration samples (phase I). These formulae will be applied in the two numerical examples presented at section 5.

(ii) **false alarm α risk:** In the case of $p = 2$, we have the exact distribution for $|S|$ in simple form (given at 2.1 *(ii)*), where $|\Sigma| = |\Sigma_0|$ under H_0: $\Sigma = \Sigma_0$.

The reference value for α is the usual 0.0027, which is pre-fixed. However, the actual α risk, is given by

$$\begin{aligned}
\alpha\, risk &= \mathbb{P}(Reject\, H_0 \,|\, H_0\, is\, true) = 1 - \mathbb{P}(LCL \le |S| \le UCL \,|\, H_0\, is\, true) \\
&= 1 - \mathbb{P}\left(|\Sigma_0|\left(b_1 - 3\sqrt{b_2}\right) \le \frac{|\Sigma_0|}{4(n-1)^2}\left(\chi^2_{2n-4}\right)^2 \le |\Sigma_0|\left(b_1 + 3\sqrt{b_2}\right)\right) \\
&= 1 - \left[F_{\chi^2_{2n-4}}\left(2(n-1)\sqrt{b_1 + 3b_2^{1/2}}\right) - F_{\chi^2_{2n-4}}\left(2(n-1)\sqrt{b_1 - 3b_2^{1/2}}\right)\right]
\end{aligned}$$

For $p > 2$, the α risk is obtained numerically, by simulation of $|S|$, as shown at section 4.

3 The Cornish-Fisher Corrected $|S|$ Control Chart

3.1 The Control Limits

In order to better approximate the exact distribution of the sample $|S|$ statistic, we correct its non-normality (skewness and kurtosis) using the information from the 3rd and

4^{th} order cumulants of $|S|$ (K_3 and K_4 from section 2.1 *(iv)*), through the Cornish-Fisher expansion formula (Cornish & Fisher, 1960; Lee & Lee, 1992).

In this formula, the p-quantile of $|S|^* = \frac{|S|-\mu}{\sigma}$, denoted $q_{|S|^*}(p)$, is corrected, starting from the p-quantile of $Z \sim N(0,1)$, denoted $q_Z(p)$, as follows:

$$q_{|S|^*}(p) \approx q_Z(p) + K_3 \frac{q_Z^2(p) - 1}{6} + K_4 \frac{q_Z^3(p) - 3q_Z(p)}{24} - K_3^2 \frac{2q_Z^3(p) - 5q_Z(p)}{36}$$

Since $|S| = \mu + \sigma|S|^*$, then $q_{|S|}(p) = \mu_{|S|} + q_{|S|^*}(p)\, \sigma_{|S|} = b_1\, |\Sigma| + q_{|S|^*}(p)\, \sqrt{b_2\, |\Sigma|^2}$, and the quantile formula is: $q_{|S|}(p) = |\Sigma| \left[b_1 + q_{|S|^*}(p) \sqrt{b_2} \right]$, where $p = \frac{\alpha_0}{2} = 0.00135$ and $p = 1 - \frac{\alpha_0}{2} = 0.99865$ will give the control limits of the chart.

3.2 The Chart Performance: False Alarm α Risk

For dimension 2, since $|S| \sim \frac{|\Sigma|}{4(n-1)^2} (\chi^2_{2n-4})^2$, we have

$$\alpha\, risk = \mathbb{P}(Reject\ H_0 \,|\, H_0\ is\ true) = 1 - \mathbb{P}(LCL \leq |S| \leq UCL \,|\, H_0\ is\ true)$$

$$= 1 - \mathbb{P}(\underbrace{\left[b_1 + q_{|S|^*}(\alpha/2) \sqrt{b_2} \right]}_{CF_1} |\Sigma_0| \leq \frac{|\Sigma_0|}{4(n-1)^2} \left(\chi^2_{2n-4}\right)^2 \leq \underbrace{\left[b_1 + q_{|S|^*}(1-\alpha/2) \sqrt{b_2} \right]}_{CF_2} |\Sigma_0|)$$

$$= 1 - \mathbb{P}\left(2(n-1)\sqrt{CF_1} \leq \chi^2_{2n-4} \leq 2(n-1)\sqrt{CF_2}\right)$$

$$\alpha\, risk = 1 - \left[F_{\chi^2_{2n-4}} \left(2(n-1)\sqrt{CF_2}\right) - F_{\chi^2_{2n-4}} \left(2(n-1)\sqrt{CF_1}\right) \right]$$

where CF_1 is the lower CF quantile for $|S|$ and CF_2 is the upper CF quantile for $|S|$.

For dimension greatter than 2, the false alarm risk should be obtained numerically by simulation.

4 False Alarm Comparative Study

4.1 The case of dimension 2

From previous sections (2.2 and 3.2), we have obtained the expressions for the exact false alarm risk for the traditional normal-based $|S|$ chart and the Cornish-Fisher corrected $|S|$ chart, which are

$$\alpha\, risk\, (normal) = 1 - \left[F_{\chi^2_{2n-4}} \left(2(n-1)\sqrt{b_1 + 3b_2^{1/2}}\right) - F_{\chi^2_{2n-4}} \left(2(n-1)\sqrt{b_1 - 3b_2^{1/2}}\right) \right]$$

$$\alpha\, risk\, (CF\ corrected) = 1 - \left[F_{\chi^2_{2n-4}} \left(2(n-1)\sqrt{CF_2}\right) - F_{\chi^2_{2n-4}} \left(2(n-1)\sqrt{CF_1}\right) \right]$$

The ideal exact-distributed $|S|$ chart, taken as a reference, has obviously $\alpha\, risk = \alpha_0$, the pre-fixed risk of type-I (reject H_0 when H_0 is true), that is considered as the usual 0.0027. The comparison results, are presented in form of table and graphics as well, as a function of the sample size n, for two-sided and one-sided charts.

The table below (Table 1) shows the false alarm risks for two-sided charts for n from 3 to 60, and after that, the same results are presented graphically (Fig. 1).

Table 1: α **risk (two-sided)**

n	$Normal\ Approx.$	$Cornish-Fisher$	n	$Normal.Approx.$	$Cornish-Fisher$
3	0.01971	0.00061	9	0.01737	0.00225
4	0.02081	0.00096	10	0.01670	0.00392
5	0.02042	0.00117	15	0.01409	0.00467
6	0.01968	0.00130	20	0.01234	0.00395
7	0.01888	0.00139	30	0.01014	0.00336
8	0.01810	0.00144	60	0.00719	0.00297
ref.	**0.00270**	**0.00270**	**ref.**	**0.00270**	**0.00270**

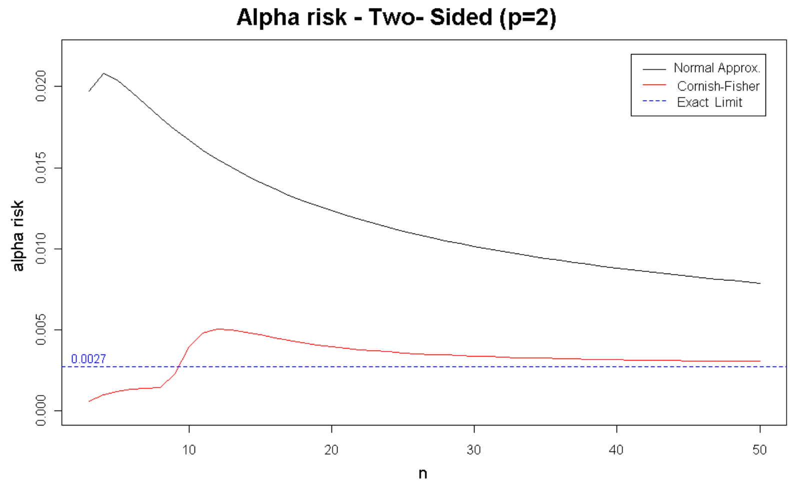


Figure 1: α risk of $|S|$ exact versus approximated normal versus Cornish-Fisher

From the table and figure above, it is clear that our corrected chart presents false alarm risks much closer to the reference risk ($\alpha_0 = 0.0027$) than the traditional normal-based chart.

If we consider only the upper risk (probab. of crossing the upper limit when H_0 is true), since it is the more important, the comparative results are even stronger, in favor of the corrected chart, as shown in Figure 2 below.

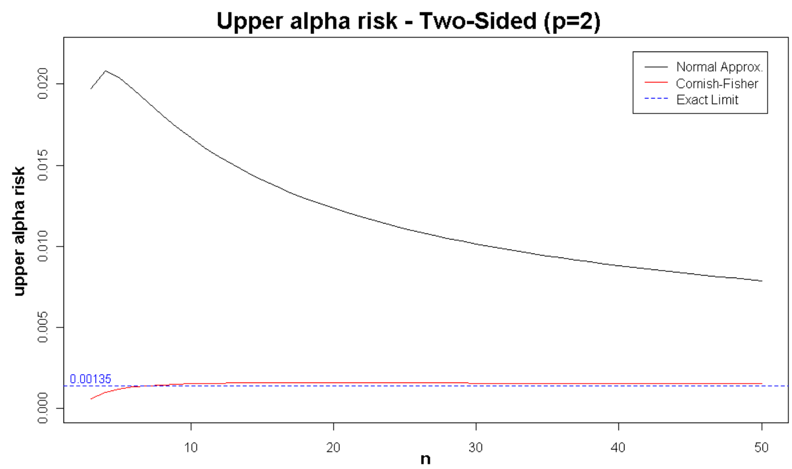


Figure 2: Upper α risk of $|S|$ exact versus approximated normal versus Cornish-Fisher

And in this case, only one term of correction in the Cornish-Fisher formula is sufficient.

In this way, in the one-sided chart (only upper limit), we use only one correction term, and the results are presented below, at Table 2 and Figure 3.

Table 2: α **risk (one-sided)**

n	$Normal\ Approx.$	$Cornish - Fisher$	n	$Normal\ Approx.$	$Cornish - Fisher$
3	0.01971	0.00100	9	0.01737	0.00259
4	0.02081	0.00161	10	0.01670	0.00265
5	0.02042	0.00198	15	0.01409	0.00281
6	0.01968	0.00222	20	0.01234	0.00285
7	0.01888	0.00239	30	0.01014	0.00287
8	0.01810	0.00250	60	0.00719	0.00284
ref.	**0.00270**	**0.00270**	**ref.**	**0.00270**	**0.00270**

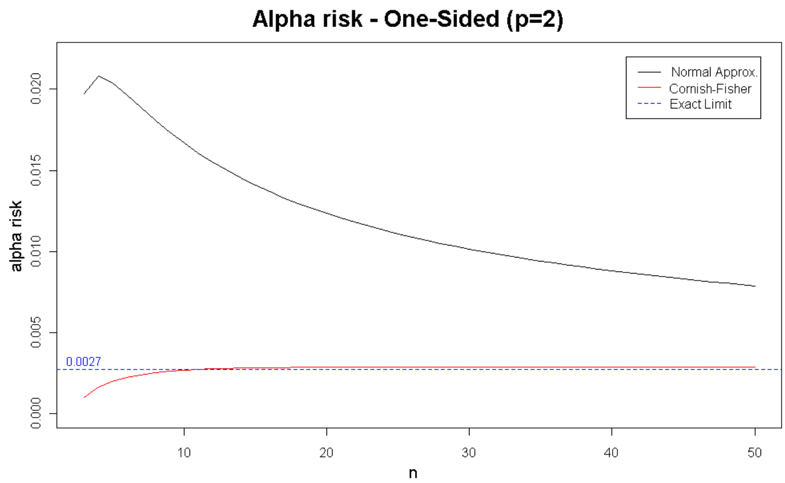


Figure 3: α risk (one-sided) of $|S|$ exact versus approximated normal versus Cornish-Fisher

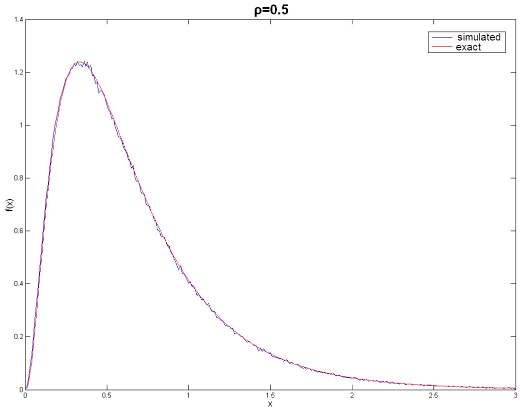
From the table and figure above we can see that the correction (just one term) produce excellent results, where the excess of false alarm is substantially reduced, almost eliminated for $n \geq 8$ say.

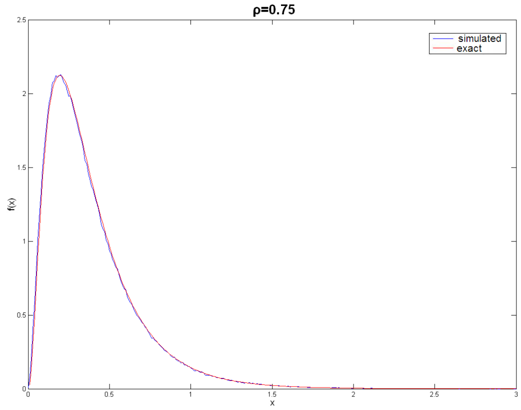
4.2 The case of dimension 3

In this case, the α risks were obtained by simulation of the $|S|$ sample distribution, through a Wishart random generator (algorithm AS53, from Smith & Hocking, 1972) available in the Matlab software, considering around $N = 1$ million samples for each sample size n. Before the study with dimension 3, we have tested the simulator in the case of dimension 2, where we know the exact distribution. This distribution depends on the Σ parameter, and we have considered, without loss of generality, that $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where the particular ρ value ($\rho = 0.5$ and 0.75 in the figures) does not have impact on the degree of approximation.

Just to illustrate the excellence of the method, with $n = 10$ and $N = 10^6$, we can not distinguish between the exact PDF of $|S|$ and the simulated PDF, as show at Figure 4 below.

Figure 4: **PDF of $|S|$ (p=2) exact versus simulated - $n = 10$, $N = 1$ million and $p = 2$**





Based on this test, we have adopted the value $N = 10^6$. For dimension 3, we have considered just the one-sided (upper limit) chart, and CF correction with only one term. The results are shown bellow, at Table 3, where the quantiles from the 3 methods are presented, and Figure 5 (false alarm risk).

Table 3: **Quantiles (one-sided)**

n	$Exact(sim.)$	$Normal\ Approx.$	CF	n	$Exact(sim.)$	$Normal Approx.$	CF
4	2.68949	1.03849	5.50433	9	2.42403	1.30405	2.59090
5	2.89698	1.23081	4.23237	10	2.29517	1.29008	2.41570
6	2.82795	1.29575	3.54973	15	1.89068	1.21041	1.90241
7	2.68866	1.31462	3.11760	20	1.64740	1.14353	1.64311
8	2.53757	1.31384	2.81575	30	1.38473	1.04938	1.37043

From the table above, we can see that the Cornish-Fisher quantiles are closer to the exact quantiles than the normal ones; in particular, for $n = 15$ or more, the Cornish-Fisher approximation is very good.

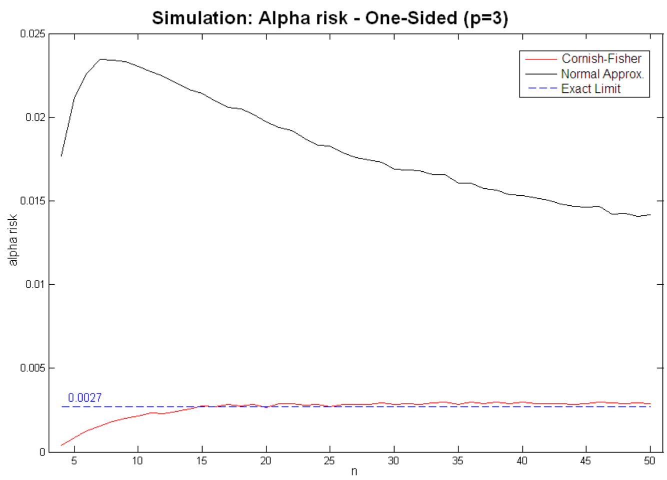


Figure 5: α risk (one-sided) of $|S|$ exact versus approximated normal versus Cornish-Fisher through simulations ($N = 1$ million samples for each sample size n)

From the figure above, we can see that this one term CF correction reduces almost all the excess of false alarm present in the normal-based chart.

5 Numerical Illustrations

5.1 An Example with $p = 2$

In order to illustrate the application of our corrected $|S|$ chart with real data, we consider an adaptation of the example 11.1 from Montgomery, D.C. (2008).

The tensile strength and diameter of a textile fiber are two important quality characteristics that are to be jointly controlled. The quality engineer has decided to use $n = 10$ fiber specimens in each sample. He has taken 20 preliminary samples shown at Table 4 below.

Table 4: **Variability data for the textile fiber example - p=2**

k	S^2_{1k}	S^2_{2k}	S_{12k}	$\|S_k\|$	k	S^2_{1k}	S^2_{2k}	S_{12k}	$\|S_k\|$
1	1.25	0.87	0.80	0.4475	11	1.45	0.79	0.78	0.5371
2	1.26	0.85	0.81	0.4149	12	1.24	0.82	0.81	0.3607
3	1.30	0.90	0.82	0.4976	13	1.26	0.55	0.72	0.1746
4	1.02	0.85	0.81	0.2109	14	1.17	0.76	0.75	0.3267
5	1.16	0.73	0.80	0.2068	15	1.48	1.07	0.82	0.9112
6	1.01	0.80	0.76	0.2304	16	1.74	1.27	0.83	1.5209
7	1.25	0.78	0.75	0.4125	17	1.80	1.42	0.70	2.0660
8	1.40	0.83	0.80	0.5220	18	1.42	1.00	0.79	0.7959
9	1.19	0.87	0.83	0.3464	19	1.31	0.89	0.76	0.5883
10	1.17	0.86	0.95	0.1037	20	1.29	0.85	0.68	0.6341

Based on the table above (preliminary samples), taking the average of the variances and covariances, we obtain an estimate of Σ, given by

$$\bar{S} = \begin{pmatrix} \bar{S}^2_1 & \bar{S}_{12} \\ & \bar{S}^2_2 \end{pmatrix} = \begin{pmatrix} 1.3085 & 0.7885 \\ & 0.8880 \end{pmatrix}$$

which result in $|\bar{S}| = 0.5402$. The constants b_1 and b_2 are given by

$$b_1 = \frac{(n-2)}{(n-1)} = 0.8889 \qquad \text{and} \qquad b_2 = \frac{(n-2)(4n-2)}{(n-1)^3} = 0.4170$$

since $n = 10$. As $|\bar{S}|$ is a biased estimator of $|\Sigma|$ we consider its bias-corrected version (see Djauhari, M.A., 2005), given by $\widehat{|\Sigma|} = |\bar{S}|/b_1 = 0.5402/0.8889 = 0.6077$, as the corresponding numerical value for $|\Sigma|$ in the control limits expression.

In this way, the control limits for the 3 charts (normal-based, CF corrected and exact) in the two-sided case, are given by:

(i) normal-based: $LCL = max\left\{0,\, \widehat{|\Sigma|}\left(b_1 - 3\sqrt{b_2}\right)\right\} = 0$, $UCL = \widehat{|\Sigma|}\left(b_1 + 3\sqrt{b_2}\right) = 1.7176$

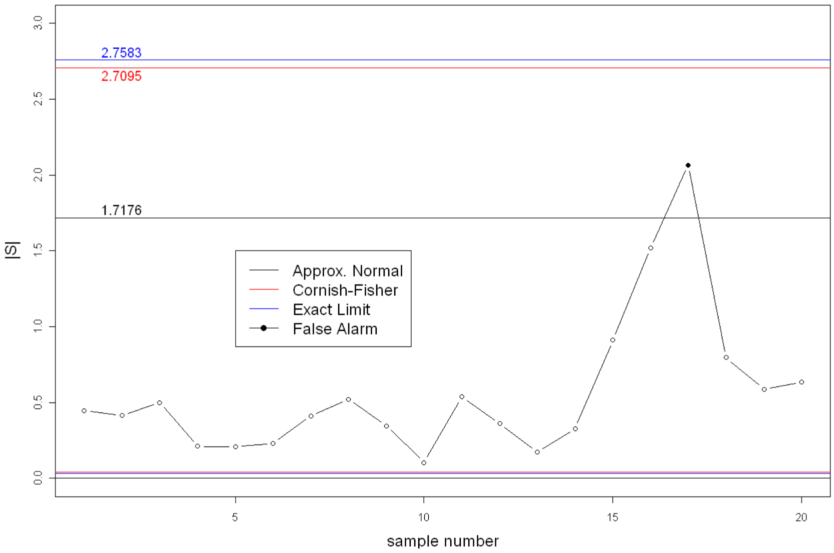
(ii) CF corrected: $LCL = q_{|S|}(0.00135) = max\left\{0,\, \widehat{|\Sigma|}\left[b_1 + q_{|S|^*}(0.00135)\sqrt{b_2}\right]\right\} = 0.0388$

$$UCL = q_{|S|}(0.99865) = \widehat{|\Sigma|}\left[b_1 + q_{|S|^*}(0.99865)\sqrt{b_2}\right] = 2.7095$$

(iii) exact limits: $LCL = |\widehat{\Sigma}|(q_{\chi^2_{2n-4}}(0.00135))^2/4(n-1)^2 = 0.0321$
$UCL = |\widehat{\Sigma}|(q_{\chi^2_{2n-4}}(0.99865))^2/4(n-1)^2 = 2.7583$

A chart with the 3 control limits and the data is presented at Figure 6 below.

Figure 6: $|S|$ **Control Chart - Two-sided,** $p = 2$



From the figure it is clear that the correct limit (from the exact chart) is not reached, and therefore the alarm from the normal-chart should be considered false at the pre-fixed α_0 level. This drawback does not happen with the Cornish-Fisher corrected chart since it practically represent the exact limit chart.

Now, if we consider the one-sided case, the upper limit for the 3 charts, are given by:
(i) normal-based: $UCL = |\widehat{\Sigma}| \left(b_1 + q_Z(0.9973)\sqrt{b_2}\right) = 1.6321$
(ii) CF corrected: $UCL = q_{|S|}(0.9973) = |\widehat{\Sigma}| \left[b_1 + q_{|S|^*}(0.9973)\sqrt{b_2}\right] = 2.4679$
(iii) exact limit: $UCL = |\widehat{\Sigma}|(q_{\chi^2_{2n-4}}(0.9973))^2/4(n-1)^2 = 2.4602$

The figure 7b below illustrate this situation, where the CF corrected chart practically coincide with the exact limit chart, avoiding the false alarm drawback that happens with the traditional normal-approximated $|S|$ control chart. In this case, the figure shows that there is no sign of problems with the process variability (see also figures 8a and 8b) nor with the process level (figure 7a) expressed through the Hotelling statistic.

Figure 7a: **Hotelling T^2 Control Chart - One-Sided, $p = 2$**

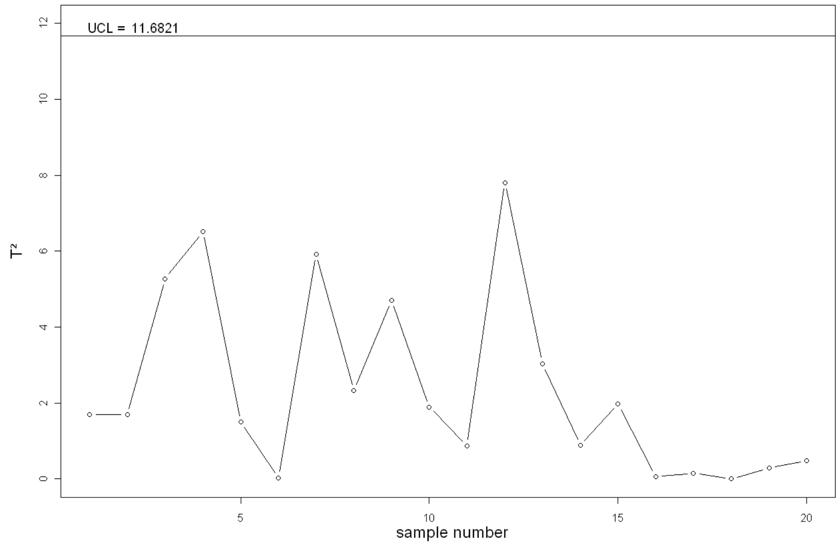


Figure 7b: **$|S|$ Control Chart - One-sided, $p = 2$**

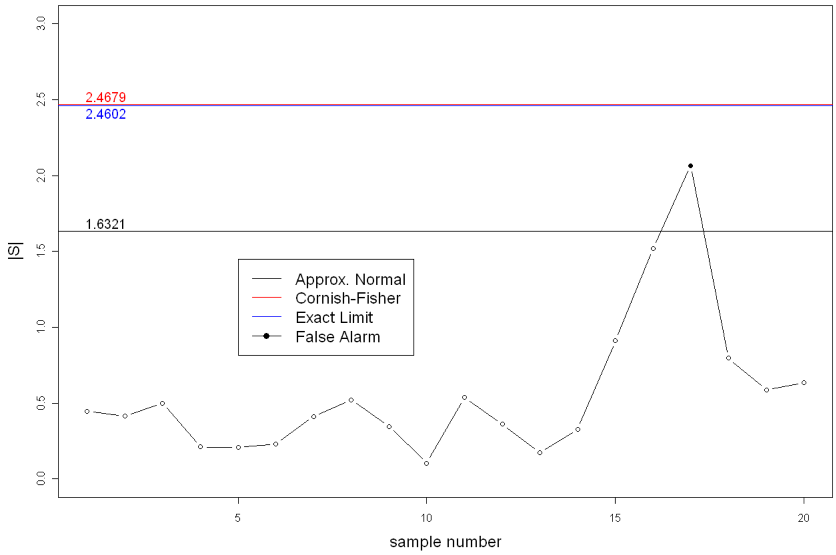


Figure 8a: **S_1^2 Control Chart - One-Sided**

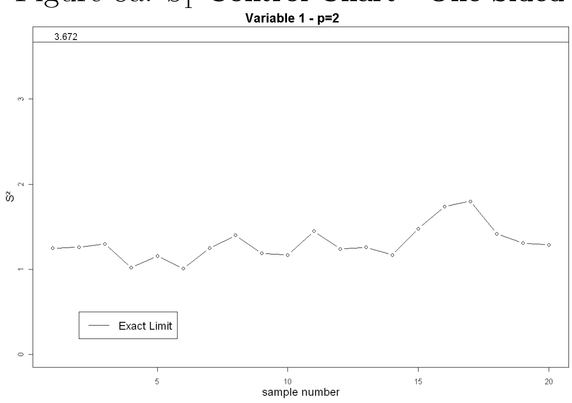
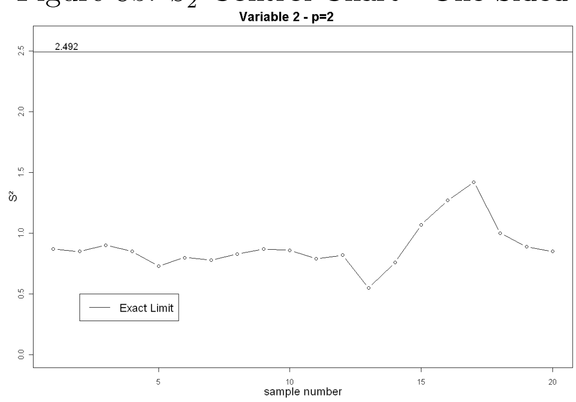


Figure 8b: **S_2^2 Control Chart - One-Sided**



5.2 An Example with $p = 3$

In order to illustrate the application of the proposed $|S|$ corrected chart when $p = 3$, we consider one adaptation of case 1 of Fuchs & Kenett (1998) where some simulations were made from the original multivariate data in order to get samples of size $n = 15$.

These original data are from a capability study of a process to produce aluminium bolts, which were used as base to generate 70 simulated samples of dimension $p = 3$. A sketch of aluminium bolt and the variables measured (X_1, X_2, X_3) is presented in the picture below.

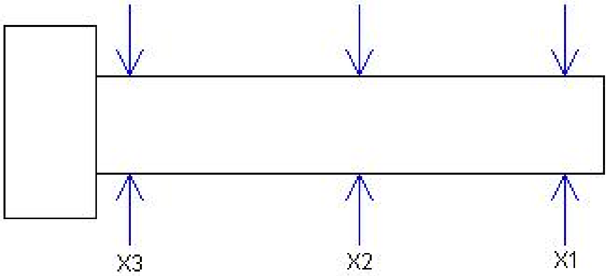


Figure 9: Aluminium bolt sketch and the measured variables

The first 30 samples (from the total of 70) were generated by simulation from the first 5 original samples, in order to be used as calibration samples for the phase I (control chart implementation). From these original samples, it was obtained the parameters (means, variances and covariances) used in the simulation, shown at table 5 below.

Table 5: **Parameters used in the simulation -** $p = 3$

k	μ_{1k}	μ_{2k}	μ_{3k}	$S^2_{1k}\,(10^{-4})$	$S^2_{2k}\,(10^{-4})$	$S^2_{3k}\,(10^{-4})$	$S_{12k}\,(10^{-4})$	$S_{13k}\,(10^{-4})$	$S_{23k}\,(10^{-4})$
1	9.997	9.986	9.989	3.81	6.54	4.21	2.78	1.55	2.26
2	9.997	9.993	9.989	3.64	8.35	4.27	1.86	0.46	1.39
3	10.001	9.991	9.985	4.12	5.69	4.41	1.12	1.57	1.49
4	9.998	9.988	9.992	4.03	3.74	3.46	1.46	1.83	2.76
5	10.003	9.990	9.989	3.67	6.57	4.49	1.50	1.02	2.50

The multivariate normal samples simulated based on the parameters above (table 5) were multiplied by 100, and then computed the $|S|$ statistics for each sample; as shown at table 6 below.

Table 6: **Generated values of the sample** $|S|$ **statistic -** $p = 3$

Sample k	S^2_{1k}	S^2_{2k}	S^2_{3k}	S_{12k}	S_{13k}	S_{23k}	$\|S_k\|$
1	6.7046	6.7999	5.5515	5.7386	2.8837	3.3595	49.2486
2	4.6107	12.4917	2.5024	5.8136	1.8154	3.2702	38.1021
3	4.2813	4.7268	3.2870	3.1535	1.9047	1.4954	25.0735
.	.					.	.
.	.					.	.
30	5.5042	7.4888	3.9771	3.2764	1.3976	2.2112	99.9524
Sum	127.1023	185.485	121.2132	59.5676	43.9286	59.0504	
Mean	4.2367	6.1828	4.0404	1.9856	1.4643	1.9683	**71.6842**

From the table margins (mean values), it is obtained the “average generalized variance matrix" given by

$$\bar{S} = \begin{pmatrix} \bar{S}_1^2 & \bar{S}_{12} & \bar{S}_{13} \\ & \bar{S}_2^2 & \bar{S}_{23} \\ & & \bar{S}_3^2 \end{pmatrix} = \begin{pmatrix} 4.2367 & 1.9856 & 1.4643 \\ & 6.1828 & 1.9683 \\ & & 4.0404 \end{pmatrix}$$

resulting that $|\bar{S}| = 71.6842$. Since $|\bar{S}|$ is a biased estimator of $|\Sigma|$, the control limits are corrected by the constants (section 2.2),

$$b_1 = \frac{(n-2)(n-3)}{(n-1)^2} = 0.7959 \qquad \text{and} \qquad b_2 = \frac{(n-2)(n-3)6}{(n-1)^3} = 0.3411$$

Then, the control limits (normal, CF corrected and exact), considering $\widehat{|\Sigma|} = |\bar{S}|/b_1$, are given by

(i) normal-approximated: $UCL = \widehat{|\Sigma|}\left(b_1 + q_Z(0.9973)\sqrt{b_2}\right) = 218.030$

(ii) Cornish-Fisher corrected: $UCL = \widehat{|\Sigma|}\left[b_1 + q_{|S|^*}(0.9973)\sqrt{b_2}\right] = 342.680$

(iii) exact limit: $UCL = \dfrac{\widehat{|\Sigma|}\, x_0}{(n-1)^p} = 339.876$

where $x_0 = 10355$ was obtained integrating numerically the $h(y)$ density (section 2.1 *(ii)*) using the *"meijerG"* function from the Matlab Symbolic Math Toolbox version 5.

From figure 10b below it is clear that the alarm from the normal chart should be considered a false alarm at the established α risk of 0.0027, since the exact limit is not reached by any point. The CF corrected chart does not present this drawback since its UCL practically coincide with the exact limit.

Therefore, the figure shows that there is no evidence of process change in variability (see also figure 10c for the univariate charts) as well as no change in the process level (figure 10a) as expressed by the Hotelling statistics.

Figure 10a: **Hotelling T^2 Control Chart - One-Sided, $p = 3$**

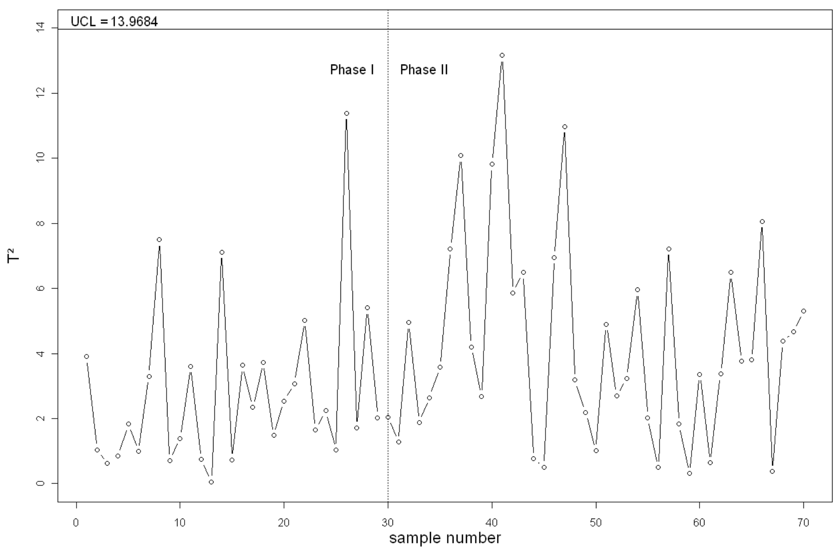


Figure 10b: $|S|$ **Control Chart - One-Sided,** $p = 3$

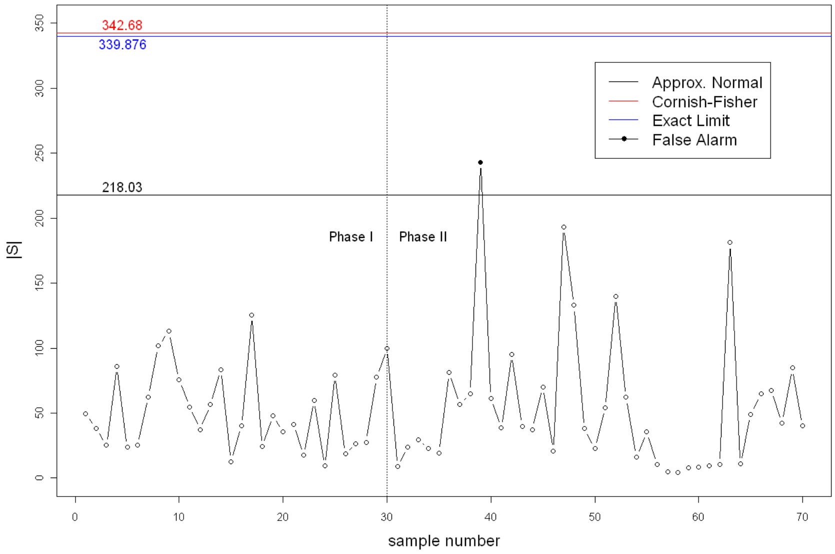
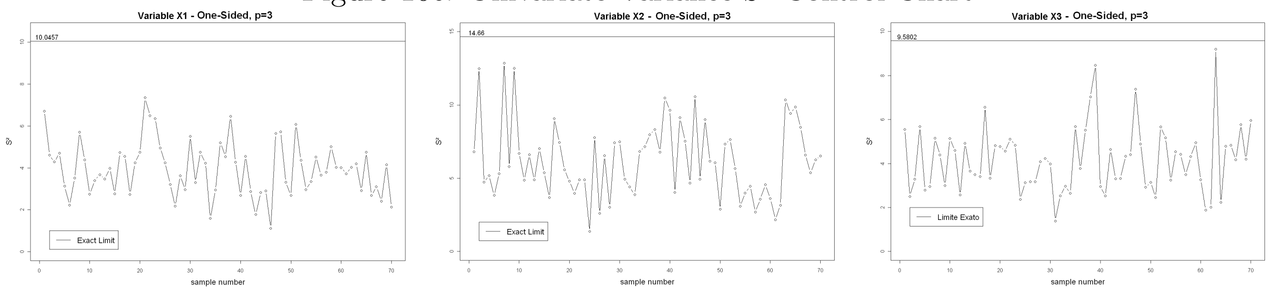


Figure 10c: **Univariate Variance** S^2 **Control Chart**



6 Final Comments and Conclusions

This paper has presented a simple correction in the traditional $|S|$ control chart for multivariate process dispersion monitoring based on the Cornish-Fisher expansion formula. For the very important case of a one-sided chart, the correction is based on only one term involving the 3^{rd} order moment or cumulant, with a very simple implementation.

At the same time, the gain in terms of practically eliminating the false alarm drawback of the traditional $|S|$ chart is enormous. With this new chart we can now control the false alarm risk at the pre-fixed level, avoiding the serious consequences of an incontrolled increase of this risk. We have presented the full details specifically for the cases of dimensions 2 or 3, but the method is general, and valid for higher dimensions.

References

[1] Alt, F.A. (1984).*Multivariate Quality Control.* In S. Kotz; N.L. Johnson and C.R. Reid (Eds). The Encyclopedia of Statistical Sciences, pp 110-122, Wiley, NY.

[2] Anderson, T.W. (1958, 1984).*An Introduction to Multivariate Statistical Analysis.* Wiley.

[3] Aparisi, F.; Jabaloyes, J. and Carrión, A. (1999).*Statistical Properties of the $|S|$ Multivariate Control Chart.* Comm. in Stat. Theory Meth., 28 (11), pp 2671-2686.

[4] Cornish, E.A. and Fisher, R.A. (1960).*The Percentage Points of Distributions Having Known Cumulants.* Technometrics 2, pp 209-225.

[5] Djauhari, M.A. (2005).*Improved Monitoring of Multivariate Process Variability.* J. of Quality Tech., 37 (1), pp 32-39.

[6] Fuchs, C. and Kenett, R. (1998).*Multivariate Quality Control.* Marcel Dekker.

[7] Hotelling, H. (1931).*The Generalization of Student's Ratio.* Ann.Math.Stat., vol. 2 (3), pp 360-378.

[8] Hotelling, H. (1947).*Multivariate Quality Control.* In Techniques of Statistical Analysis. Eds.: C. Eisenhart; M. Hastay and W.A. Wallis. McGraw-Hill, NY, pp 111-184.

[9] Johnson, R.A. and Wichern, D.W. (1982).*Applied Multivariate Statistical Analysis.* Prentice Hall.

[10] Lee, Y. and Lee, M.C. (1992).*On the Derivation and Computation of the Cornish-Fisher Expansion.* Austr. J. of Stat., 34 (3), pp 443-450.

[11] Mathai, A.M. (1972).*The Exact Non-Central Distribution of the Generalized Variance.* Ann. Inst. Stat. Math., 24, pp 53-65.

[12] Matlab 2009. The MathWorks Inc, and Symbolic Math Toolbox (version 5).

[13] Montgomery, D.C. (2008).*Introduction to Statistical Quality Control.* $6^{\underline{th}}$ ed., Wiley.

[14] Pham-Gia, T. and Turkkan, N. (2008).*Exact Expression of the Density of the Sample Generalized Variance and Applications.* Statist. Papers, DOI 10.1007, Springer.

[15] R Development Core Team (2010).*R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, URL http://www.R-project.org

[16] Shewhart, W. A. (1926).*Quality Control Charts.* Bell System Technical Journal $n^{\underline{o}}$ 5, pp 593-603.

[17] Smith, W.B. and Hocking, R.R. (1972).*Algorithm AS53: Wishart Variate Generator.* JRSS, Series C (Applied Statistics), vol 21, $n^{\underline{o}}$ 3, pp 341-345.

[18] Springer, M. (1979).*The Algebra of Random Variables.* Wiley, New York.

[19] Wilks, S.S. (1932).*Certain Generalizations in the Analysis of Variance.* Biometrika, vol 24, pp 471-494.