PHASE I AND PHASE II - CONTROL CHARTS FOR THE VARIANCE AND GENERALIZED VARIANCE

R. van Zyl [1]
Quintiles International
e-mail: *ruaan.vanzyl@quintiles.com*

A.J. van der Merwe
University of the Free State

Key words: Generalized Variance, Phase 1, Phase 2, Predictive Density, Shewart-type Control Charts, Variance.

Abstract: By extending the results of Human, Chakraborti and Smit (2010), Phase I control charts are derived for the generalized variance when the mean vector and covariance matrix of multivariate normally distributed data are unknown and estimated from m independent samples, each of size n. In Phase II predictive distributions based on a Bayesian approach are used to construct Shewart-type control limits for the variance and generalized variance. The posterior distribution is obtained by combining the likelihood (the observed data in Phase I) and the uncertainty of the unknown parameters via the prior distribution. By using the posterior distribution the unconditional predictive density functions are derived.

1. Introduction

Quality control is a process which is used to maintain the standards of products produced or services delivered. As mentioned by Human et al. (2010) monitoring spread is important in practice since it is an indicator of process quality and also the spread must be first monitored before the mean control chart should be constructed and examined. If any of the sample variances plot on or outside the control limits, they should be examined and if they are discarded, revised values should be calculated for the estimators as well as for the control limits. Although the main aim of this paper is to monitor sample variances and generalized variances that are too large, i.e., upper one-sided control limits, it is also important to look at samples with very small variances, i.e., two-sided control-charts. Observations with very small variances or no variance at all seem suspicious and could have been tampered with (i.e., artificially changed) to suit the quality control procedure.

It is nowadays commonly accepted by most statisticians that statistical process control should be implemented in two phases:

[1] Corresponding author.

AMS: 62F15, 62P30

1. Phase I where the primary interest is to assess process stability; and
2. Phase II where online monitoring of the process is done.

As in Human et al. (2010) the Phase I control charts will be constructed using the false alarm probability (FAP) which is the overall probability of at least one false alarm. The nominal FAP value that will be used is $FAP_0 = 0.05$. The main difference between our method and that of Human et al. (2010) is in Phase II where we are using a Bayesian procedure to study the in-control run length.

Bayarri and Garcia-Donato (2005) gave the following reasons for recommending Bayesian analysis for the determining of control chart limits:

- Control charts are based on future observations and Bayesian methods are very natural for prediction.
- Uncertainty in the estimation of the unknown parameters is adequately handled.
- Implementation with complicated models and in a sequential scenario poses no methodological difficulty, the numerical difficulties are easy handled via Monte Carlo methods.
- Objective Bayesian analysis is possible without introduction of external information other than the model, but any kind of prior information can be incorporated into the analysis if desired.

In this article, control chart limits will be determined for the sample variance, S^2 , and the generalized variance $|S|$. Average run-lengths and false alarm rates will also be calculated in the Phase II setting, using a Bayesian predictive distribution.

2. Example

The data presented in Table 1 represent measurements of inside diameters and represent the number of 0.0001 inches above 0.7500 inches (Duncan, 1965). The measurements are taken in samples of $j = 1,2,\ldots,n$ each ($n = 5$) over time. Also shown in Table 1 are the sample variances, S_i^2 for $i = 1,2,\ldots m$ samples ($m = 10$). These data will be used to construct a Shewart type Phase I upper control chart for the variance, and also to calculate the run-length for future samples of size $n = 5$ taken repeatedly for the process.

From the data in Table 1 the sample variances are calculated by

$$S_i^2 = \frac{1}{n-1}\sum_{j=1}^{m}(y_{ij}-\bar{y}_i)^2 .$$

The pooled sample variance is then determined as

$$S_p^2 = \frac{1}{m}\sum_{i=1}^{m} S_i^2 = 10.72.$$

Table 1: Data for Constructing a Shewart-Type Phase I Upper Control Chart for the Variance.

Sample Number/ Time (i)	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}	s_i^2
1	15	11	8	15	6	16.5
2	14	16	11	14	7	12.3
3	13	6	9	5	10	10.3
4	15	15	9	15	7	15.2
5	11	14	11	12	5	11.3
6	13	12	9	6	10	7.5
7	10	15	12	4	6	19.8
8	9	12	9	8	8	2.7
9	8	12	14	9	10	5.8
10	10	10	9	14	14	5.8

3. Statistical Calculation of the Control Limits in Phase I

The upper control limit, using the data by Duncan (1965) will be obtained as described by Human et al. (2010).

It is well known that

$$\frac{(n-1)\,S_i^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Also, if the underlying distribution is Normal,

$$\frac{m\,(n-1)\,S_p^2}{\sigma^2} \sim \chi^2_{m(n-1)} = \sum_{i=1}^{m} \chi^2_{n-1}.$$

Therefore

$$Y_i = \frac{(n-1)\,S_i^2/\sigma^2}{m\,(n-1)\,S_p^2/\sigma^2} = \frac{X_i}{\sum_{i=1}^{m} X_i}$$

where $X_i \sim \chi^2_{n-1}$ $(i = 1, 2, \ldots, m)$.

The distribution of $Y_{max} = max\,(Y_1, Y_2, \ldots, Y_m)$ obtained from $100,000$ simulations is illustrated in Figure 1 . The value b is then calculated such that the False Alarm Probability (FAP) is at a level of 0.05 (also shown in the figure).

The upper control limit is then determined as:

$$UCL = mbS_p^2 = 10\,(0.3314)\,(10.72) = 35.526.$$

The data from Duncan (1965) are presented visually in Figure 2. The figure includes the upper control limit as determined above.

In the case of a two-sided control chart the joint distribution of $Y_{min} = min\,(Y_1, Y_2, \ldots, Y_m)$ and $Y_{max} = max\,(Y_1, Y_2, \ldots, Y_m)$ must first be simulated. The equal tail values given in Table 2, Page 868 of Human et al. (2010) are calculated in such a way that the FAP does not exceed 0.05. For $m = 10$

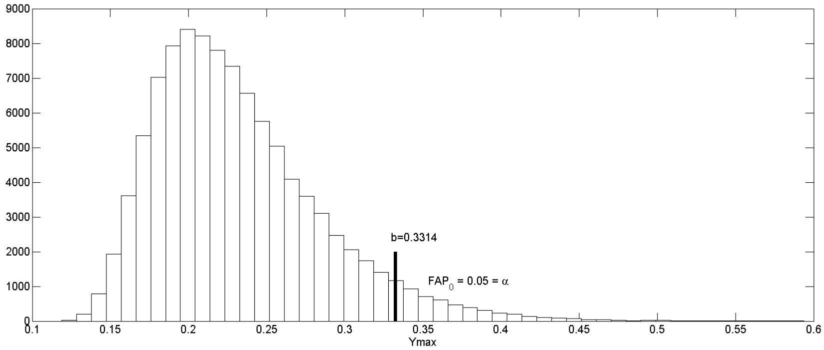


Figure 1: Distribution of $Y_{max} = max\left(Y_1, Y_2, \ldots, Y_m\right)$ (100,000) simulations.

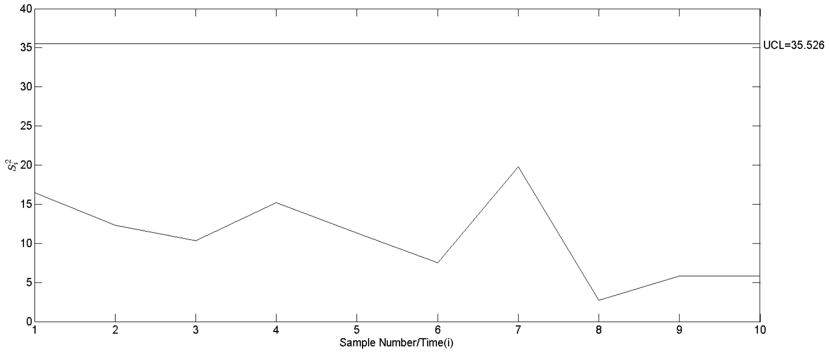


Figure 2: Shewart-type Phase I Upper Control Chart for the Variance - $FAP_0 = 0.05$.

and $n = 5$ it follows that the lower control limit, $LCL = ma_1 S_p^2 = (10)(0.0039)(10.72) = 0.4181$ and the upper control limit, $UCL = mb_1 S_p^2 = (10)(0.3599)(10.72) = 38.581$. Since the correlation coefficient between Y_{min} and Y_{max} for $m = 10$ and $n = 5$ is -0.2731, the shortest two-sided control limits such that the FAP does not exceed 0.05 is given by $L\tilde{C}L = (10)(0.0002)(10.72) = 0.0214$ and $U\tilde{C}L = (10)(0.3314)(10.72) = 35.5261$. This interval is 7% shorter than the equal tail interval.

4. Upper Control Limit for the Variance in Phase II

In the first part of this section, the upper control limit in a Phase II setting will be derived using the Bayesian predictive distribution.

Theorem 1 Assume $Y_{ij} \sim^{iid} N\left(\mu_i, \sigma^2\right)$ where Y_{ij} denotes the j^{th} observation from the i^{th} sample where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. The mean μ_i and variance σ^2 are unknown.

Using the Jeffrey's prior $p\left(\mu_1, \mu_2, \ldots, \mu_m, \sigma^2\right) \propto \sigma^{-2}, \sigma^2 > 0, -\infty < \mu_i < \infty, i = 1, 2, \ldots, m$ it can be proved that the posterior distribution of σ^2 is given by

$$p\left(\sigma^2|data\right) = \left(\frac{\tilde{S}}{2}\right)^{\frac{1}{2}k} \frac{1}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}(k+2)} \exp\left(-\frac{\tilde{S}}{2\sigma^2}\right), \sigma^2 > 0 \tag{1}$$

an Inverse Gamma distribution with $k = m(n-1)$ and $\tilde{S} = m(n-1)S_p^2$.

Proof. The proof is given in the Appendix. ■

The posterior distribution given in (1) is presented in Figure 3.

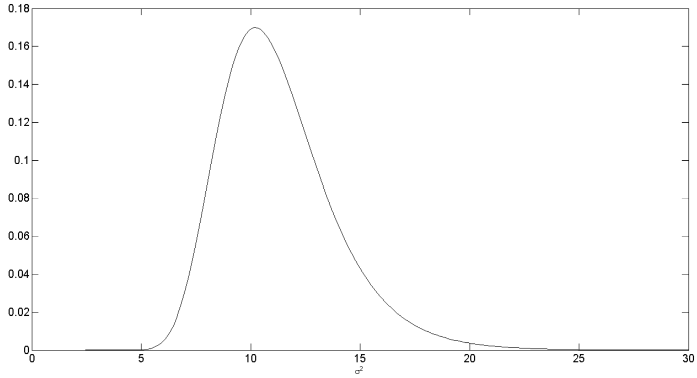


Figure 3: Distribution of $p\left(\sigma^2|data\right)$-Simulated Values.

A predictive distribution derived using a Bayesian approach will be used to obtain the control limits in a Phase II setting. Let S_f^2 be the sample variance of a future sample of n observations from the Normal distribution. Then for a given σ^2 it follows that

$$\frac{(n-1)S_f^2}{\sigma^2} = \frac{vS_f^2}{\sigma^2} \sim \chi_v^2$$

which means that

$$f\left(S_f^2|\sigma^2\right) = \left(\frac{v}{2\sigma^2}\right)^{\frac{1}{2}v} \frac{1}{\Gamma\left(\frac{v}{2}\right)} \left(S_f^2\right)^{\frac{1}{2}v-1} \exp\left(-\frac{vS_f^2}{2\sigma^2}\right). \tag{2}$$

Theorem 2 If S_f^2 is the sample variance of a future sample of n observations from the Normal distribution then the unconditional predictive density of S_f^2 is given by

$$f\left(S_f^2|data\right) = S_p^2 F_{n-1,m(n-1)}$$

where S_p^2 is the pooled sample variance and $F_{n-1,m(n-1)}$ the F-distribution with $n-1$ and $m(n-1)$ degrees of freedom.

Proof. The proof is given in the Appendix. ■

The upper control limit in the Phase II setting is then derived as

$$S_p^2 F_{n-1,m(n-1)}(1-\beta).$$

At $\beta = 0.0027$ we therefore obtain the upper control limit as

$$S_p^2 F_{n-1,m(n-1)}(0.9973) = (10.72)(4.8707) = 52.214.$$

The distribution of the predictive density of S_f^2 including the derived upper control limit is presented in Figure 4.

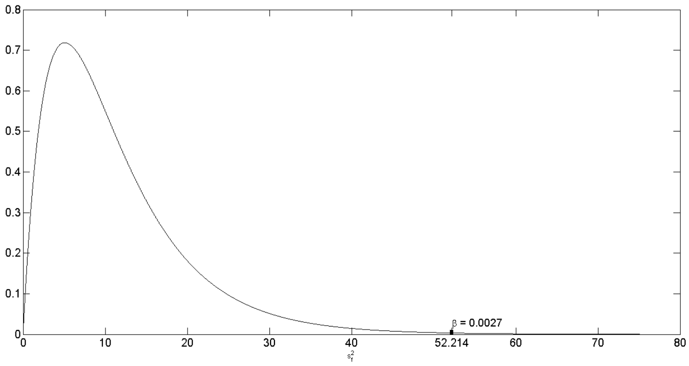


Figure 4: Distribution of $f\left(S_f^2|data\right)$.

Assuming that the process remains stable, the predictive distribution for S_f^2 can also be used to derive the distribution of the run-length, that is the number of samples until the control chart signals for the first time.

The resulting rejection region of size β using the predictive distribution for the determination of the run-length is defined as

$$\beta = \int_{R(\beta)} f\left(S_f^2|data\right) dS_f^2$$

where

$$R(\beta) = (52.214, \infty),$$

is the upper one-sided control limits.

Given σ^2 and a stable process, the distribution of the run-length r is Geometric with parameter

$$\psi\left(\sigma^2\right) = \int_{R(\beta)} f\left(S_f^2|\sigma^2\right) dS_f^2$$

where $f\left(S_f^2|\sigma^2\right)$ given in (2) is the predictive distribution of a future sample variance given σ^2.

The value of the parameter σ^2 is however unknown and its uncertainty is described by the posterior distribution $p\left(\sigma^2|data\right)$ defined in (1).

Theorem 3 For a given σ^2 the parameter of the Geometric distribution is

$$\psi\left(\sigma^2\right) = \psi\left(\chi^2_{m(n-1)}\right) \text{ for given } \chi^2_{m(n-1)}$$

which means that it is only dependent on $\chi^2_{m(n-1)}$ and not on σ^2.

Proof. The proof is given in the Appendix. ∎

In Figure 5 the distributions of $f\left(S_f^2|data\right)$ and $f\left(S_f^2|\sigma^2\right)$ for $\sigma^2 = 9$ and $\sigma^2 = 20$ are presented to show the different shapes of the applicable distributions.

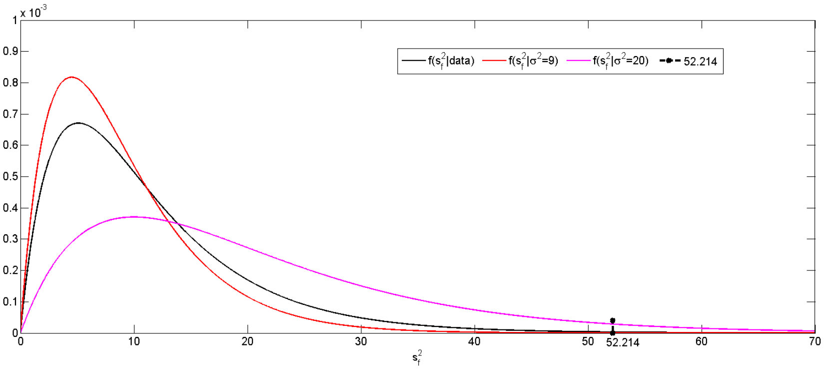


Figure 5: Distributions of $f\left(S_f^2|\sigma^2\right)$ and $f\left(S_f^2|data\right)$ showing $\psi\left(\sigma_1^2\right)$.

As mentioned, by simulating σ^2 from $p\left(\sigma^2|data\right)$ the probability density function $f\left(S_f^2|\sigma^2\right)$ as well as the parameter $\psi\left(\sigma^2\right)$ can be obtained. This must be done for each future sample. Therefore by simulating a large number of σ^2 values from the posterior distribution a large number of $\psi\left(\sigma^2\right)$ values can be obtained. A large number of geometric and run length distributions with different parameter values ($\psi\left(\sigma_1^2\right)$, $\psi\left(\sigma_2^2\right)$,...,$\psi\left(\sigma_l^2\right)$) will therefore be available. The unconditional run length distribution is obtained by using the Rao-Blackwell method, i.e., the average of the conditional run length distributions.

In Table 2(a) results for the run-length at $\beta = 0.0027$ for $n = 5$ and different values for m are presented for the upper control limit for the variance. The table presents the mean, median, 95% equal tail interval and calculated β value to obtain a run-length of 370 (the expected run length at $\beta = 0.0027$ is $\frac{1}{0.0027} \approx 370$ if σ^2 is known).

In the case of the diameter example the mean run-length is 29754 and the median run-length 1354. The reason for these large values is the uncertainty in the parameter estimate because of the small sample size and number of samples ($n = 5$ and $m = 10$). To obtain a mean run-length 370 β must be 0.0173 instead of 0.0027.

For a two-sided control chart, the upper control limit is $UCL = S_p^2 F_{n-1,m(n-1)}\left(1-\frac{\beta}{2}\right)$ and the lower control limit is $LCL = S_p^2 F_{n-1,m(n-1)}\left(\frac{\beta}{2}\right)$. $R(\beta)$ are all those values larger than UCL and smaller than LCL. For the diameter example $UCL = (10.72)(5.4445) = 58.365$ and $LCL = (10.72)(0.02583) = 0.2769$. In this case the mean run-length is 500.

In Table 2(b) results for the run-length at $\beta = 0.0027$ for $n = 5$ and different values of m are presented for a two-sided control chart.

Table 2: Mean and Median Run-length at $\beta = 0.0027$ for $n = 5$ and Different Values of m.

m	n	Mean	Median	95% Equal Tail Interval	Calculated β for Mean Run Length of 370
10	5	29 754	1 354	(54;117 180)	0.0173
50	5	654	470	(121;2 314)	0.0044
100	5	482	411	(156;1 204)	0.0035
200	5	422	391	(197;829)	0.0031
500	5	389	379	(244;596)	0.0028
1 000	5	379	374	(274;517)	0.0028
5 000	5	371	370	(322;428)	0.0027
10 000	5	370	370	(335;410)	0.0027

(a) Upper Control Limit.

m	n	Mean	Median	95% Equal Tail Interval
10	5	500	552	(92;661)
50	5	399	427	(184;492)
100	5	385	398	(227;474)
200	5	377	383	(263;459)
500	5	373	375	(300;432)
1 000	5	371	372	(320;416)
5 000	5	370	370	(347;391)
10 000	5	370	370	(354;385)

(b) Upper and Lower Control Limit.

From Table 2 it can be noted that as the number of samples increase (larger m) the mean and median run-lengths converge to the expected run-length of 370.

Further, define $\bar{\psi}\left(\sigma^2\right) = \frac{1}{l}\sum_{i=1}^{l}\psi\left(\sigma_i^2\right)$. From Menzefricke (2002) it is known that if $l \to \infty$, then $\bar{\psi}\left(\sigma^2\right) \to \beta = 0.0027$ and the harmonic mean of the unconditional run length will be $\left(\frac{1}{\beta}\right) = \frac{1}{0.0027} = 370$. Therefore it does not matter how small m and n are, the harmonic mean of the run length will be $\frac{1}{\beta}$ if $l \to \infty$.

5. Phase I Control Charts for the Generalized Variance

Assume $\underline{Y}_{ij} \sim^{idd} N\left(\underline{\mu}_i, \Sigma\right)$ where $\underline{Y}_{ij}\,(p \times 1)$ denotes the jth observation vector from the ith sample, $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. The mean vector $\underline{\mu}_i\,(p \times 1)$ and covariance matrix, $\Sigma\,(p \times p)$ are unknown.

Define $\bar{\underline{Y}}_i = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$ and $A_i = \sum_{j=1}^{n}\left(\underline{Y}_{ij} - \bar{\underline{Y}}_i\right)\left(\underline{Y}_{ij} - \bar{\underline{Y}}_i\right)'$ $(i = 1, 2, \ldots, m)$.

From this it follows that

$$\underline{\bar{Y}}_i \sim N\left(\underline{\mu}_i, \frac{1}{n}\Sigma\right), (i = 1,2,\ldots,m),$$

$$A_i = (n-1) S_i \sim W_p (n-1, \Sigma),$$

$$A = \sum_{i=1}^{m} A_i \sim W_p (m(n-1), \Sigma)$$

and

$$S_p = \frac{1}{m(n-1)} A.$$

The generalized variance of the ith sample is defined as the determinant of the sample covariance matrix, i.e., $|S_i|$.

Define

$$T_i = \frac{|A_i|}{|\sum_{i=1}^{m} A_i|} = \frac{|A_i^*|}{\left|\sum_{i=1}^{m} A_i^*\right|}$$

where $A_i^* \sim W_p (n-1, I_p)$.

Also

$$T = max(T_1, T_2, \ldots, T_m) = max(T_i), i = 1,2,\ldots,m.$$

Now

$$T_i = \frac{|A_i|}{|\sum_{i=1}^{m} A_i|} = \frac{|S_i|}{m^p \left|S_p\right|}.$$

Therefore a $(1-\beta)\,100\%$ upper control limit for $|S_i|$ $(i = 1,2,\ldots,m)$ is $m^p \left|S_p\right| T_{1-\beta}$.

Figure 6 presents a histogram of 100,000 simulated values of $max(T_i)$ for the two dimensional case (p=2, m=10 and n=6). The upper control limit given in Table 3 is presented on the figure. Table 3 also presents the upper control limit for the one dimensional (p=1) and the three dimensional (p=3) situations as well as the lower control limit for all three dimensions.

Table 3: Upper 95% Control Limit, $T_{0.95}$ for $T = max(T_i)$ for the Generalized Variance in Phase I for $m = 10$, $n = 6$ and $p = 1,2$ and 3.

p	m	n	$T_{0.95}$	$T_{0.025}$	$T_{0.975}$
1	10	6	0.30259	0.00665	0.32655
2	10	6	0.04429	0.00014	0.05122
3	10	6	0.00445	$1.822967e^{-6}$	0.00544

By using a Bayesian procedure a predictive distribution will be derived to obtain control chart limits in Phase II.

Using the Jeffrey's prior

$$p(\underline{\mu}, \Sigma) \propto |\Sigma|^{-\frac{1}{2}(p+1)} \quad -\infty < \underline{\mu} < \infty, \Sigma > 0$$

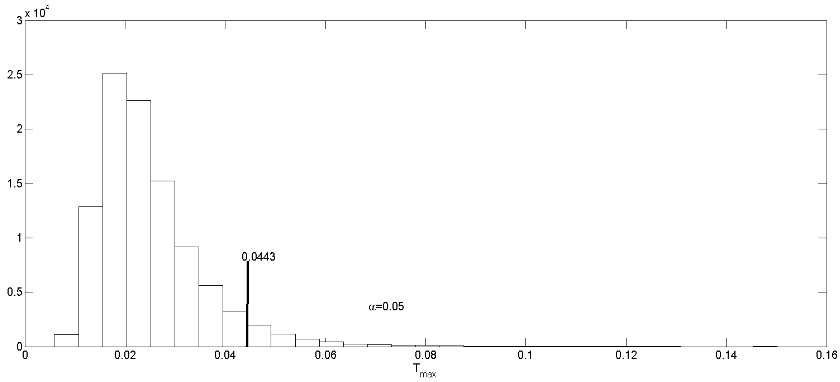


Figure 6: Histogram of $max(T_i)$-100,000 Simulations.

the posterior distribution of Σ is derived as

$$|\Sigma|\,|data \sim |A| \prod_{i=1}^{p} \left(\frac{1}{\chi^2_{m(n-1)+1-i}} \right) \tag{3}$$

and the predictive distribution of a future sample generalized variance $|S_f|$ given Σ as

$$|S_f|\,|\Sigma \sim \left| \frac{1}{n-1} \Sigma \right| \prod_{i=1}^{p} \chi^2_{n-i}. \tag{4}$$

By combining (3) and (4) the unconditional predictive distribution is given by

$$|S_f^2|\,|data \sim \left(\frac{1}{n-1} \right)^p |A| \left(\prod_{i=1}^{p} \frac{n-i}{m(n+1)+1-i} \right) F^* \tag{5}$$

where

$$F^* = \prod_{i=1}^{p} F_{n-i, m(n-1)+i-i}.$$

Equation (5) can be used to obtain the control chart limits.

Similarly for the variance, the rejection region of size β is defined as

$$\beta = \int_{R(\beta)} f\left(|S_f|\,|data \right) d\,|S_f|.$$

Given Σ and a stable process, the distribution of the run-length r is Geometric with parameter

$$\psi(|\Sigma|) = \int_{R(\beta)} f\left(|S_f|\,|\Sigma \right)$$

where $f\left(\left|S_f\right| |\Sigma \right)$ is given in (4).

Theorem 4 For a given value of $|\Sigma|$, and simulated from the posterior distribution described in (3), the parameter of the Geometric distribution is

$$\psi(|\Sigma|) = P\left\{\prod_{i=1}^{p}\chi^2_{n-i} \geq \left(\prod_{i=1}^{p}\chi^2_{m(n-1)+1-i}\right)\left(\prod_{i=1}^{p}\frac{n-i}{m(n-1)+1-i}\right)F^*_{1-\beta}\right\}$$

for a given $\prod_{i=1}^{p}\chi^2_{m(n-1)+1-i}$.

Proof. The proof is given in the Appendix. ■

For further details see Menzefricke (2002, 2007, 2010a, 2010b) .

Mean and median run-length results at $\beta = 0.0027$ for $n = 50$, $m = 50$ and 100 for the one, two and three dimensional cases are presented in Table 4.

Table 4: Mean and Median Run-length at $\beta = 0.0027$ for $n = 50$, $m = 50$ and 100 and $p = 1,2$ and 3.

p	m	n	Mean	Median	95% Equal Tail Interval
1	50	50	482	414	(185;1 198)
1	100	50	431	402	(197;841)
2	50	50	466	404	(162;1 128)
2	100	50	423	396	(205;819)
3	50	50	461	407	(165;1 063)
3	100	50	424	399	(209;786)

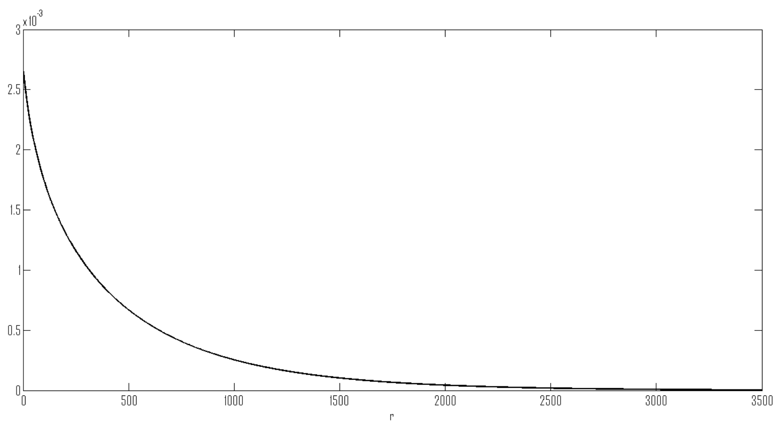
6. Guidelines for Practitioners for the Implementation of the Proposed Control Chart

It was mentioned in the introductory section and also by Chakraborti, Human and Graham (2008) that it is now generally accepted that statistical process control should be implemented in two phases. Phase I is the so-called retrospective phase and Phase II the prospective or monitoring phase.

The construction of Phase I control charts should be considered as a multiple testing problem. The distribution of a set of dependent variables (ratios of chi-square random variables) can therefore be used to calculate the control limits so that the false alarm probability is not larger than FAP_0. Tables are provided by Human et al. (2010) for the charting constants for S^2 for each Phase I chart, for a FAP_0 of 0.01 and 0.05 respectively. These tables can easily be implemented by practitioners. Further the charting constants can also be adjusted so that the difference between the upper and lower control limit is a minimum. Tables of the “adjusted” charting constants for a FAP_0 and 0.05 can therefore also be made available to practitioners. Similar tables can be drawn up for the generalized variance.

Since the Phase I control charting problem is considered to be a multiple hypothesis testing problem, Bartletts test can be used instead of two-sided control charts. Bartletts test is a likelihood ratio test and it is uniformly the most powerful test for testing the homogeneity of variances. Under the null hypothesis it is the ratio of the geometric and arithmetic mean of a set of independent chi-square random variables. Tables of critical values can therefore be easily calculated. A similar test is available for the generalized variance.

Once the Phase I control chart has been calculated, statistical process control should move to Phase II. According to Chakraborti et al. (2008) Phase II control chart performance should be measured in terms of some attribute of the run length. This is exactly what this paper has tried to achieve. Predictive distributions based on a Bayesian approach are used to construct Shewart-type control limits for the variance and generalized variance. By using Monte Carlo simulation methods the distributions of the run length and average run length can easily be obtained. It has also been shown that an increase in the number of samples m and sample size n leads to a convergence of the run length towards the expected value of 370 at $\beta = 0.0027$. In the case of small sample sizes and number of samples the average run length can be very large because of the uncertainty in the estimation of the parameter σ^2. This is especially the case for the upper control limit. If practitioners are interested in a certain average run length, they are advised to adjust the nominal value of β to obtain that specific run length. For $m = 10$ and $n = 5$, $\beta = 0.0173$ an average run length will be 370. For the "diameter" example, the upper control limit will then be $UCL = (3.406)(10.72) = 36.512$. In Figure 7 the predictive distribution of the run length is displayed for the $(1 - 0.0173)\,100\% = 98.27\%$ two-sided control limit. As mentioned for given σ^2 the run length r is geometric with parameter $\psi\left(\sigma^2\right)$. The unconditional run length as given in Figure 7 is therefore obtained using the Rao-Blackwell method, i.e., the average of a large number of unconditional run lengths.

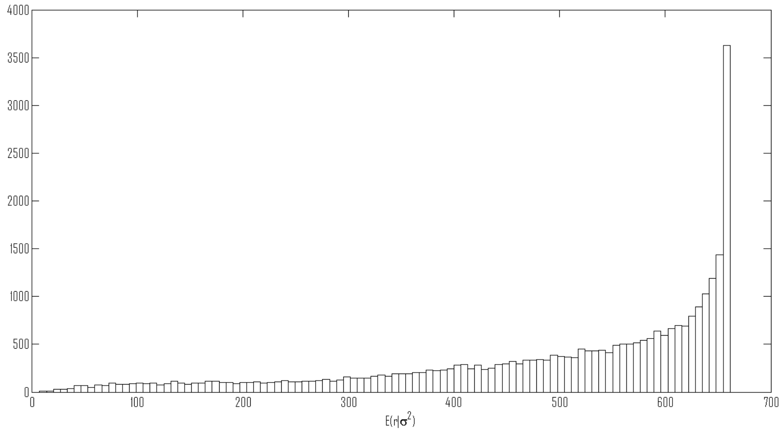


$$E\,(r|data) = 498.6473;\ Median\,(r|data) = 319;\ Var\,(r|data) = 274473.1449$$

$$95\%\ Equal - tail = [9; 1961]$$

Figure 7: Predictive Distribution of the "Run-length" $f\,(r|data)$for $m = 10$ and $n = 5$ - Two-sided Control Chart.

In Figure 8 the distribution of the average run length is given. Also the harmonic mean of the run length is $\frac{1}{\beta}$. Therefore if $\beta = 0.0027$, the harmonic mean is $\frac{1}{0.0027} = 370.37$ and if $\beta = 0.0173$, the harmonic mean is $\frac{1}{0.0173} = 57.8$ and the arithmetic mean is 370.



$$Mean = 500;\ Median = 552;\ Variance = 25572.95$$

$$95\%\ Equal - tail = [92; 661]$$

Figure 8: Distribution of the Average "Run-length" - Two-sided Control Chart.

7. Conclusion

Phase I and Phase II control chart limits have been constructed using Bayesian methodology. In this article we have seen that due to Monte Carlo simulation the construction of control chart limits using the Bayesian paradigm are handled with ease. Bayesian methods allow the use of any prior to construct control limits without any difficulty. It has been shown that the uncertainty in unknown parameters are handled with ease in using the predictive distribution in the determination of control chart limits. It has also been shown that an increase in number of samples m and the sample size n leads to a convergence in the run-length towards the expected value of 370 at $\beta = 0.0027$.

Acknowledgements

We would like to thank the anonymous referee and the editor whose contributions improved and made this publication possible.

References

BAYARRI, M. AND GARCIA-DONATO, G. (2005). A Bayesian sequential look at u-control charts. *Technometrics*, **47** (2), 142–151. doi:10.1198/004017005000000085.

CHAKRABORTI, S., HUMAN, S., AND GRAHAM, M. (2008). Phase I statistical process control charts: An overview and some results. *Quality Engineering*, **21** (1), 52–62. doi: 10.1080/08982110802445561.

DUNCAN, A. (1965). *Quality Control and Industrial Statistics*, volume 3. Richard D. Irwin, Inc.: Homewood, Illinois.

HUMAN, S., CHAKRABORTI, S., AND SMIT, C. (2010). Shewart-type control charts for variation in Phase I data analysis. *Computational Statistics and Data Analysis*, **54**, 863–874. doi:10.1016/j.csda.2009.09.026.

MENZEFRICKE, U. (2002). On the evaluation of control chart limits based on predictive distributions. *Communications in Statistics - Theory and Methods*, **31** (8), 1423–1440. doi: 10.1081/STA-120006077.

MENZEFRICKE, U. (2007). Control chart for the generalized variance based on its predictive distribution. *Communications in Statistics - Theory and Methods*, **36** (5), 1031–1038. doi: 10.1080/03610920601036176.

MENZEFRICKE, U. (2010a). Control chart for the variance and the coefficient of variation based on their predictive distribution. *Communications in Statistics - Theory and Methods*, **39** (16), 2930–2941. doi:10.1080/03610920903168610.

MENZEFRICKE, U. (2010b). Multivariate exponentially weighted moving average chart for a mean based on its predictive distribution. *Communications in Statistics - Theory and Methods*, **39** (16), 2942–2960. doi:10.1080/03610920903168628.

Appendix

Theorem 1

Proof. The likelihood function, i.e., the distribution of the data is

$$L\left(\mu_1,\mu_2,\ldots,\mu_m,\sigma^2|data\right)=\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}mn}\prod_{i=1}^{m}\prod_{j=1}^{n}\exp\left\{-\frac{1}{2}\left(y_{ij}-\mu_i\right)^2/\sigma^2\right\}.$$

Deriving the posterior distribution as Poster $\propto$ Likelihood $\times$ Prior, and using the Jeffrey's prior it follows that

$$\mu_i|\sigma^2,data\sim N\left(\bar{y}_i,\frac{\sigma^2}{n}\right),i=1,2,\ldots,m$$

and

$$p\left(\sigma^2|data\right)=\left(\frac{\tilde{S}}{2}\right)^{\frac{1}{2}k}\frac{1}{\Gamma\left(\frac{k}{2}\right)}\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}(k+2)}\exp\left(-\frac{\tilde{S}}{2\sigma^2}\right),\sigma^2>0$$

an Inverse Gamma distribution with $k=m(n-1)$ and $\tilde{S}=m(n-1)S_p^2$. ■

Theorem 2

Proof. For a given σ^2 it follows that

$$\frac{(n-1)S_f^2}{\sigma^2}=\frac{vS_f^2}{\sigma^2}\sim\chi_v^2,$$

which means that

$$f\left(S_f^2|\sigma^2\right)=\left(\frac{v}{2\sigma^2}\right)^{\frac{1}{2}v}\frac{1}{\Gamma\left(\frac{v}{2}\right)}\left(S_f^2\right)^{\frac{1}{2}v-1}\exp\left(-\frac{vS_f^2}{2\sigma^2}\right)$$

where $v=n-1$ and $S_f^2>0$.
The unconditional predictive density of S_f^2 is given by

$$\begin{aligned} f\left(S_f^2|data\right) &= \int_0^\infty f\left(S_f^2|\sigma^2\right)p\left(\sigma^2|data\right)d\sigma^2 \\ &= \frac{(v)^{\frac{1}{2}v}(\tilde{S})^{\frac{1}{2}k}\left(S_f^2\right)^{\frac{1}{2}v-1}\Gamma\left(\frac{v+k}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{v}{2}\right)\left(\tilde{S}+vS_f^2\right)^{\frac{1}{2}(v+k)}} \qquad S_f^2>0\end{aligned}$$

where $v=n-1$, $k=m(n-1)$ and $\tilde{S}=kS_p^2=m(n-1)S_p^2$.
Therefore

$$f\left(S_f^2|data\right)=S_p^2F_{n-1,m(n-1)}.$$

■

Theorem 3

Proof. For a given σ^2

$$\begin{aligned}
\psi\left(\sigma^2\right) &= P\left(S_f^2 > S_p^2 F_{n-1,m(n-1)}\left(1-\beta\right)\right) \\
&= P\left(\frac{\sigma^2\chi^2_{n-1}}{n-1} > S_p^2 F_{n-1,m(n-1)}\left(1-\beta\right)\right) && \text{for given } \sigma^2 \\
&= P\left(\frac{m(n-1)S_p^2}{\chi^2_{m(n-1)}}\frac{\chi^2_{n-1}}{n-1} > S_p^2 F_{n-1,m(n-1)}\left(1-\beta\right)\right) && \text{for given } \chi^2_{m(n-1)} \\
&= P\left(\chi^2_{n-1} > \frac{1}{m}\chi^2_{m(n-1)} F_{n-1,m(n-1)}\left(1-\beta\right)\right) && \text{for given } \chi^2_{m(n-1)} \\
&= \psi\left(\chi^2_{m(n-1)}\right) && \text{for given } \chi^2_{m(n-1)}.
\end{aligned}$$

■

Theorem 4

Proof. For a given $|\Sigma|$

$$\begin{aligned}
\psi\left(|\Sigma|\right) &= P\left\{\left|S_f\right| > \left(\tfrac{1}{n-1}\right)^p |A| \left(\prod_{i=1}^p \frac{n-i}{m(n-1)+1-i} F^*_{1-\beta}\right)\right\} \\
&= P\left\{\left|\tfrac{1}{n-1}\Sigma\right| \prod_{i=1}^p \chi^2_{n-i} \geq \left(\tfrac{n}{n-1}\right)^p |A| \left(\prod_{i=1}^p \frac{n-i}{m(n-1)+1-i}\right) F^*_{1-\beta}\right\} \\
&= P\left\{|A| \prod_{i=1}^p \left(\frac{1}{\chi^2_{m(n-1)+1-i}}\right) \prod_{i=1}^p \chi^2_{n-i} \geq |A| \left(\prod_{i=1}^p \frac{n-i}{m(n-1)+1-i}\right) F^*_{1-\beta}\right\} \\
&= P\left\{\prod_{i=1}^p \chi^2_{n-i} \geq \left(\prod_{i=1}^p \chi^2_{m(n-1)+1-i}\right) \left(\prod_{i=1}^p \frac{n-i}{m(n-1)+1-i}\right) F^*_{1-\beta}\right\}
\end{aligned}$$

for a given $\prod_{i=1}^p \chi^2_{m(n-1)+1-i}$. ■

Manuscript received, 2015-03-03, revised, 2015-08-24, accepted, 2015-08-28.