SMALL SAMPLE UNI- AND MULTIVARIATE CONTROL CHARTS FOR MEANS

F.W. Scholz and T.J. Tosch
Boeing Computer Services
F.W. Scholz, P.O. Box 24346, MS: 7L-22, Seattle, WA 98124-0346

Keywords: Moving range, moving squared range, Satterthwaite method, Wishart distribution.

1 Introduction

Several characteristics, say d of them, are measured on the output of a process. Such characteristics may or may not be correlated. A case has been made that one should not monitor such processes by keeping separate control charts on each of the d characteristics. What is often suggested is a single control chart based on Hotelling's T^2. See Alt (1985) and Murphy (1987) on this point and for further entries into the literature. It is usually assumed that the observations come in subgroups of size k each and each such subgroup is used to compute an estimated covariance matrix in order to capture the local variation pattern. These covariance matrices are then averaged in order to obtain a more stable estimate to be used in the formation of the Hotelling T^2 criterion.

Often subgroups are of size $k = 1$ or the variation within a subgroup is not representative of the between subgroup variation, i.e., we may be dealing with batch effects. In this latter case it would make more sense to average the observations in each subgroup and let the variation of these averages speak for themselves, i.e., we will again deal with $k = 1$. It is this latter situation (subgroups of size $k = 1$) that we address here and we follow the univariate strategy of an X-chart. Such charts, also called charts for individual measurement, are discussed in Montgomery (1991), chapter 6-4. Since an X-chart is typically based on substantially fewer data points, say $n = 20$, the variance estimate obtained from the moving range formula is not yet very stable. This should be viewed in contrast to the fairly stable estimate based on $n = 20$ subgroups of size $k = 5$. In that case we would have $20 \cdot 4 = 80$ degrees of freedom in estimating σ as opposed to 19 degrees of freedom when the group size is $k = 1$. Here the 19 degrees of freedom are based on using the sample variance of all 20 observations. This however has the drawback of completely ignoring the time order of the observations and any trend in the observations could be mistaken for natural variability. In the univariate situation such trends would easily be visible on the control chart but not so in a multivariate situation. For this reason one prefers σ estimates that are based on local variation, such as the moving range or the moving squared range. The latter is more easily adapted to the multivariate situation and was proposed by Holmes and Mergen (1993), however without allowing for the degree of freedom loss due to the overlap of the ranges.

We will show that with $n = 20$ the moving squared range estimate of σ has only roughly 13 degrees of freedom, since the local variabilty estimates use overlapping data values. For the same reason the moving range estimate would presumably have similarly reduced degrees of freedom, but here the effect is more difficult to assess analytically. This reduction in degrees of freedom is still better than the 10 degrees of freedom one gets when using the 10 nonoverlapping, consecutive data pairs to estimate σ. To account for this small sample instability in the σ estimates we propose to use control limits based on the Student-t distribution (F distribution in the multivariate case) with appropriately adjusted degrees of freedom.

2 The Univariate Case

Before discussing the general multivariate setup we will illustrate this issue with the univariate situation. Here we assume that we observe $X_1, \ldots, X_n$ i.i.d. $\sim N(\mu, \sigma^2)$ when the process is under control. Compute

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \widetilde{\sigma}_n^2 = \frac{1}{2(n-1)}\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2 .$$

For obvious reasons $\widetilde{\sigma}_n^2$ is called the *moving squared range* estimate of σ^2. It is an unbiased estimate for

σ^2 and in the Appendix it is shown (as special case of the multivariate case) that $f \cdot \widetilde{\sigma}_n^2 \approx \sigma^2 \chi_f^2$ where $f = 2(n-1)^2/(3n-4)$ are the effective degrees of freedom which account for the overlap. We form the following T statistic which compares a future independent observation X with the past history of the process

$$T = \sqrt{\frac{n}{n+1}}\,\frac{X - \overline{X}_n}{\widetilde{\sigma}_n}\ .$$

Using the fact that $\overline{X}_n$ is independent of $\widetilde{\sigma}_n^2$ we can approximate the distribution of T by a Student t distribution with f degrees of freedom when the process is in control. The quality of this approximation is illustrated in Figure 1. Here $N = 1000$ samples of size $n = 20$ and $N = 1000$ future observations were generated from a standard normal population. $\overline{X}_{20}$ and $\widetilde{\sigma}_{20}$ were computed for each such sample and the corresponding T ratio was computed for each sample and its corresponding future observation. The sorted T ratios are plotted against the corresponding quantiles of a t distribution with $f = 12.9$ degrees of freedom. The point pattern follows the main diagonal exceptionally well and thus appears to confirm the validity of the approximation.

A future observation X is said to be out of control when $|T| > t_{f,.99865}$, where the .99865 point is chosen to parallel the pointwise false alarm rate in the conventional control chart based on $\pm 3\sigma$ limits. Using limits based on the t distribution results in wider control limits than would be used in ordinary X-charts. In these latter charts one usually treats the μ and σ estimates as though they agree with the underlying process parameters. This amounts to setting $f = \infty$ above. The difference between this and our small sample treatment is illustrated in Figure 2a, where the solid control limits are based on the t distribution and the dashed control limits are the conventional ones ($f = \infty$), i.e., $\pm 3\sqrt{n/(n+1)}$. In either case the limits are based on a training sample of size $n = 20$ (the training sample points are not shown here) and the long run behavior for $N = 1000$ future observations is exhibited. Note that the dashed lines are violated more frequently than would be desired by the nominal exception rate of 2.7/1000. In order to point out another feature we have replicated Figure 2a in Figure 2b. A comparison of the two figures illustrates the "random effect" of the training sample. Namely, sometimes the training sample shows unusually high dispersion. In that case (Figure 2b) the T ratios are scaled down too much and will be far away from either set of control limits. On the other hand, if the training sample shows unusually low dispersion, then the T ratios will be inflated and violate either set of control limits more than they should. This is not illustrated in the interest of saving space. In any case, the nominal exception rate of 2.7/1000 should be interpreted as averaged over all these training sample random effects. If there are too many exceptions that turn out to be false alarms one could draw the conclusion that the initial training sample of size $n = 20$ is not very representative of the process and one should then recalibrate the limits taking the later data into account. In fact, such a recalibaration can take place after each new observation is obtained. This way the price of wider t-based control limits is only temporary. Such an updated control chart is illustrated in Figure 3, using an initial training sample of size $n = 20$ and 80 update observations. Strictly speaking, the first 20 points should not have been plotted, since they do not represent future observation. The narrowing control limits show the updating effect. The dashed lines also show a very mild updating effect due to the factor $\sqrt{n/(n+1)}$. After some updating (to downplay the random effect of the training sample) one may want to stick with the established control limits in order to be more sensitive to slow mean drifts.

3 The Multivariate Case

Here we assume that our observations are vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ i.i.d. with d-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We will form $n-1$ unbiased estimates of $\boldsymbol{\Sigma}$ based on local difference vectors $\mathbf{y}_\nu = \mathbf{x}_{\nu+1} - \mathbf{x}_\nu$, $\nu = 1, \ldots, n-1$, namely

$$\begin{aligned}
\mathbf{S}_\nu &= \frac{1}{2} \cdot \mathbf{y}_\nu \cdot \mathbf{y}_\nu' \\
&= \frac{1}{2} \begin{pmatrix} y_{\nu,1} \cdot y_{\nu,1} & y_{\nu,1} \cdot y_{\nu,2} & \cdots & y_{\nu,1} \cdot y_{\nu,d} \\ y_{\nu,2} \cdot y_{\nu,1} & y_{\nu,2} \cdot y_{\nu,2} & \cdots & y_{\nu,2} \cdot y_{\nu,d} \\ \cdots & & & \cdots \\ y_{\nu,d} \cdot y_{\nu,1} & y_{\nu,d} \cdot y_{\nu,2} & \cdots & y_{\nu,d} \cdot y_{\nu,d} \end{pmatrix}
\end{aligned}$$

and as pooled estimate for $\boldsymbol{\Sigma}$ use

$$\widetilde{\mathbf{S}}_n = \frac{1}{(n-1)} \sum_{\nu=1}^{n-1} \mathbf{S}_\nu = \frac{1}{2(n-1)} \sum_{\nu=1}^{n-1} \mathbf{y}_\nu \cdot \mathbf{y}_\nu'\ .$$

In the Appendix it is shown that

$$\begin{aligned}
\widetilde{F} &= \frac{f-d+1}{f \cdot d}\,\frac{n}{n+1} (\mathbf{x} - \overline{\mathbf{x}}_n)' \widetilde{\mathbf{S}}_n^{-1} (\mathbf{x} - \overline{\mathbf{x}}_n) \\
&\approx F_{d,f-d+1}\ ,
\end{aligned}$$

where

$$\overline{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \ , \quad f = \frac{2(n-1)^2}{3n-4}$$

and $\mathbf{x}$ is a future observation. The quality of this F-approximation is illustrated with QQ-plots in Figures 4a-4c. Here the dimension is $d = 5$ and all components have common correlation $\rho = 0, .2, .5$. The sample size is $n = 20$, which results in effective degrees of freedom of 5 and 8.89. As in the univariate case independent F-ratios were computed, sorted, and compared against corresponding quantiles from the approximating F-distribution. The approximation appears to be reasonable for various values of ρ.

This can be used to judge whether a future $\mathbf{x}$ is out of line with the past history of the process by comparing $\widetilde{F}$ with the .9973 point of the $F_{d,f-d+1}$ distribution. These F-based limits are much wider than those that would result if one took the estimated parameters as "known" true parameters and thus applied the appropriate χ^2_d limits, adjusted by the factor $(f-d+1)n/[fd(n+1)]$ which appears in the definition of $\widetilde{F}$. The difference is illustrated in Figures 5a-5b for common correlation $\rho = .2$. As in Figures 2a-2b we again observe the random effect of the start-up sample of size $n = 20$. The dashed line is based on the χ^2_5 distribution (modified by the above factor) and leads to many false alarms. This is much more pronounced here than in the univariate case, since another 4 degrees of freedom are lost due to $d = 5$. Figure 6 is the counterpart to Figure 3 using multivariate data with common correlation $\rho = .2$. It should be apparent that there is a much stronger case for updating the limits in the multivariate case.

4 Appendix

Let $\mathbf{X}' = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be the matrix of n data vectors in R^d, where $\mathbf{x}_i \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\mathbf{y}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$ for $i = 1, \ldots, n-1$. In matrix form this can be written as

$$\begin{pmatrix} -1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & -1 & 1 & \ldots & 0 & 0 \\ \ldots & & & & \ldots & & \ldots \\ 0 & 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \mathbf{x}_3' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{x}_2' - \mathbf{x}_1' \\ \mathbf{x}_3' - \mathbf{x}_2' \\ \mathbf{x}_4' - \mathbf{x}_3' \\ \vdots \\ \mathbf{x}_n' - \mathbf{x}_{n-1}' \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1' \\ \mathbf{y}_2' \\ \mathbf{y}_3' \\ \vdots \\ \mathbf{y}_{n-1}' \end{pmatrix}$$

or

$$\mathbf{DX} = \mathbf{Y}$$

where D is the above differencing matrix. Consider next the following unbiased estimate $\widetilde{\mathbf{S}}_n$ of $\boldsymbol{\Sigma}$:

$$\begin{aligned}\widetilde{\mathbf{S}}_n &= \frac{1}{2(n-1)}\sum_{i=1}^{n-1}\mathbf{y}_i\mathbf{y}_i' = \frac{1}{2(n-1)}\mathbf{Y}'\mathbf{Y} \\ &= \frac{1}{n-1}\mathbf{X}'\mathbf{A}\mathbf{X} \ ,\end{aligned}$$

where $\mathbf{A} = \mathbf{D}'\mathbf{D}/2$. We will argue that a certain multiple of $\widetilde{\mathbf{S}}_n$ has a distribution which can be approximated by a Wishart distribution $W_d(f, \boldsymbol{\Sigma})$ for some f. The idea behind this is basically the same as the Satterthwaite method of approximating the distribution of quadratic forms by an appropriate multiple (α) of a chi-square random variable with f degrees of freedom. The multiplier α and degree of freedom f are obtained by matching the first two moments of the quadratic form and the approximating distribution. That the same can be done in the multivariate case hinges on the following theorem which may be found in Seber (1984, p.24).

Theorem: Let $\mathbf{X}' = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, where $\mathbf{x}_i \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$, and let $\mathbf{v} = \boldsymbol{X\ell}$, where $\boldsymbol{\ell}$ is a d-vector of constants. Let $\mathbf{A}$ be an $n \times n$ symmetric matrix of rank r. Then $\mathbf{X}'\mathbf{A}\mathbf{X} \sim W_d(r, \boldsymbol{\Sigma})$ if and only if $\mathbf{v}'\mathbf{A}\mathbf{v} \sim \sigma^2_\ell \chi^2_r$ for any $\boldsymbol{\ell} \in R^d$, where $\sigma^2_\ell = \boldsymbol{\ell}'\boldsymbol{\Sigma}\boldsymbol{\ell}$.

Without loss of generality we assume $\boldsymbol{\mu} = \mathbf{0}$ when considering the distribution of $\widetilde{\mathbf{S}}_n$. Denoting $v_i = \mathbf{x}_i'\boldsymbol{\ell} \sim N(0, \sigma^2_\ell)$ and $\mathbf{v}' = (v_1, \ldots, v_n)$ we will invoke the Satterthwaite approximation paradigm, namely, for some α the quadratic form

$$\frac{1}{\alpha}\mathbf{v}'\mathbf{A}\mathbf{v} \approx \sigma^2_\ell \chi^2_f$$

for appropriate α and f. First note that

$$\begin{aligned} E\left(\frac{1}{\alpha}\mathbf{v}'\mathbf{A}\mathbf{v}\right) &= E\left(\frac{1}{2\alpha}\sum_{i=1}^{n-1}(v_{i+1}-v_i)^2\right) \\ &= \frac{n-1}{\alpha}\sigma^2_\ell \end{aligned}$$

and

$$\mathrm{var}\left(\frac{1}{\alpha}\mathbf{v}'\mathbf{A}\mathbf{v}\right) = \frac{1}{4\alpha^2}\sum_{i=1}^{n-1}\mathrm{var}\left((v_{i+1}-v_i)^2\right)$$

$$+\frac{2}{4\alpha^2}\sum_{1\le i<j\le n-1}\operatorname{cov}\left((v_{i+1}-v_i)^2(v_{j+1}-v_j)^2\right)$$

$$=\frac{\sigma_\ell^4}{\alpha^2}(3n-4)\ .$$

Equating these to the corresponding mean and variance of $\sigma_\ell^2\chi_f^2$ we get the following two equations

$$\frac{n-1}{\alpha}\sigma_\ell^2 = f\sigma_\ell^2 \quad\text{and}\quad \frac{\sigma_\ell^4}{\alpha^2}(3n-4) = 2f\sigma_\ell^4$$

which yield

$$\alpha = \frac{3n-4}{2(n-1)} \quad\text{and}\quad f = \frac{2(n-1)^2}{3n-4} \approx \frac{2n}{3}\ .$$

Since α and f do not depend on $\boldsymbol{\ell}$, it appears reasonable to claim that

$$(n-1)\widetilde{\mathbf{S}}_n/\alpha = f\widetilde{\mathbf{S}}_n \approx W_d(f,\boldsymbol{\Sigma})\ .$$

Now note that

$$\overline{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i \sim N_d(\boldsymbol{\mu},\boldsymbol{\Sigma}/\mathrm{n})$$

is independent of $\widetilde{\mathbf{S}}_n$. Further, if $\mathbf{x}\sim N_d(\boldsymbol{\mu},\boldsymbol{\Sigma})$ is a future independent observation, then

$$\mathbf{x}-\overline{\mathbf{x}}_n \sim N_d(\mathbf{0},\boldsymbol{\Sigma}(1{+}1/\mathrm{n}))$$

is independent of $\widetilde{\mathbf{S}}_n$. From this it follows (see Seber 1984, p.30) that

$$\widetilde{F} \;=\; \frac{f-d+1}{f\cdot d}\frac{n}{n+1}(\mathbf{x}-\overline{\mathbf{x}}_n)'\widetilde{\mathbf{S}}_n^{-1}(\mathbf{x}-\overline{\mathbf{x}}_n)$$

$$\approx F_{d,f-d+1}\ .$$

References

Alt, F.B. (1985). "Multivariate quality control," *Encyclopedia of Statistical Sciences*, Vol. 6, eds. S. Kotz, N.L. Johnson, and C. Read, John Wiley & Sons, New York, 110-122.

Holmes, D.S. and Mergen, A.E. (1993). "Improving the performance of the T^2 control chart," *Quality Engineering* **5**, 619-625.

Montgomery, D.C. (1991). *Introduction to Statistical Quality Control* 2nd ed., John Wiley & Sons, New York.

Murphy, B.J. (1987). "Selecting out of control variables with the T^2 multivariate quality control procedure." *The Statistician* **36**, 571-583.

Seber, G.A.F. (1984), *Multivariate Observations*, John Wiley & Sons, New York.

Figure 1: QQ-Plot for the Student t Approximation

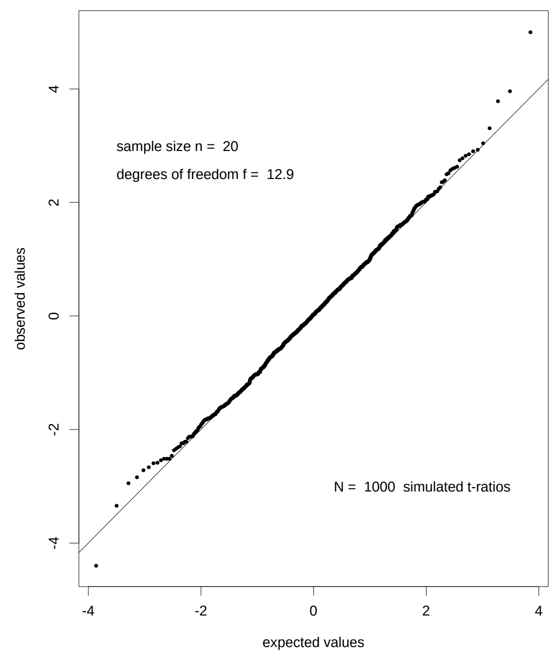

Figure 2a: Univariate X-Chart Limits Based on t Approximation

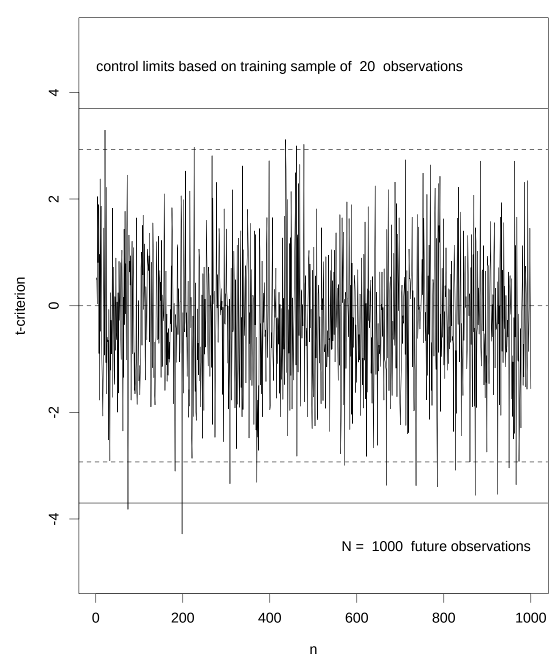

Figure 2b: Univariate X-Chart Limits Based on t Approximation

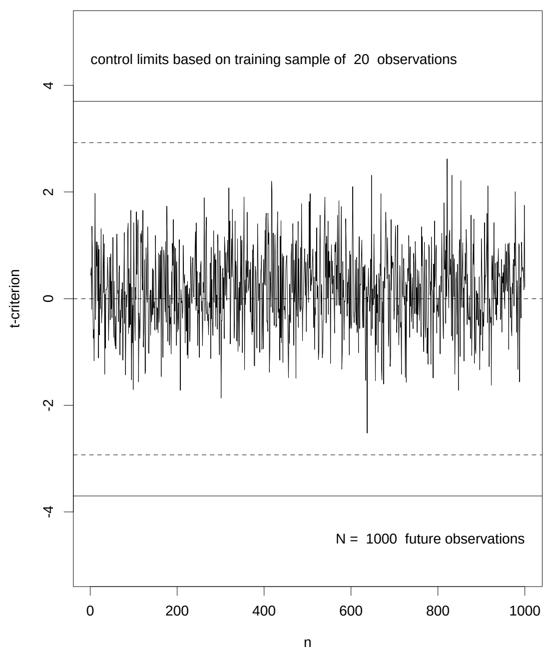

Figure 3: Univariate X-Chart Limits Based on Updated t Approximation

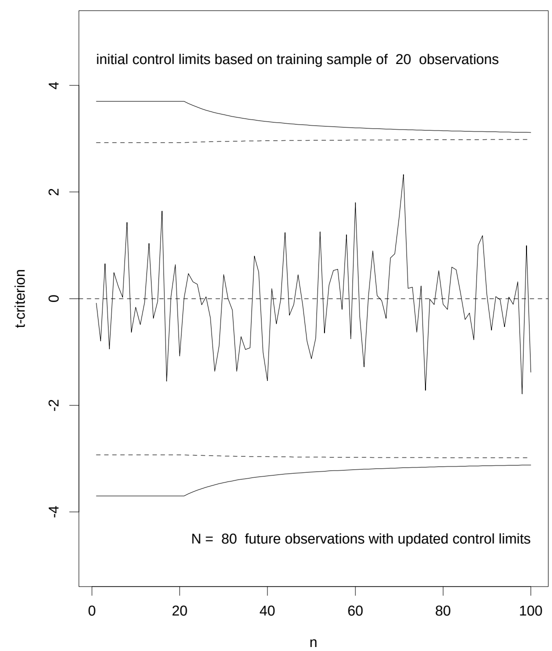

Figure 4a: QQ-Plot for the F-Approximation

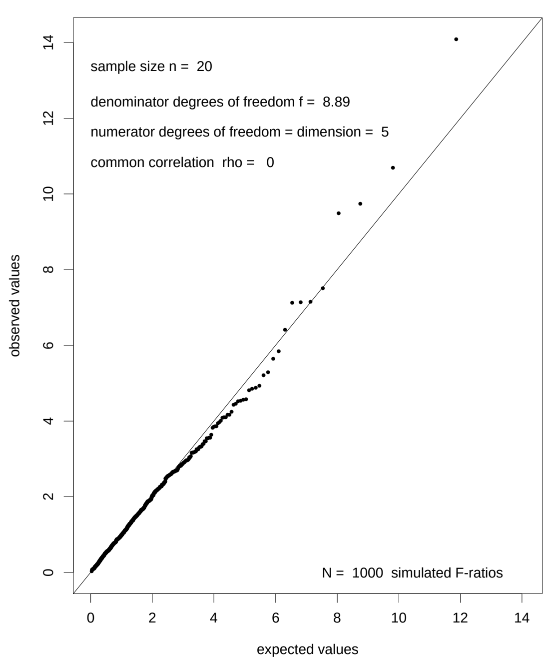

Figure 4b: QQ-Plot for the F-Approximation

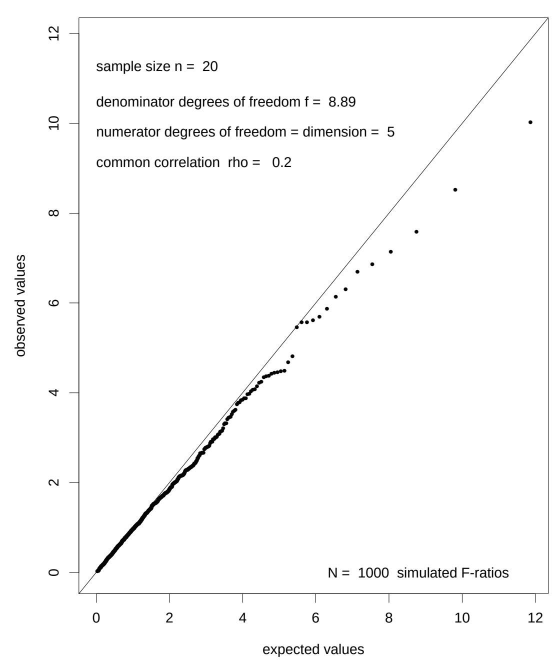

Figure 4c: QQ-Plot for the F-Approximation

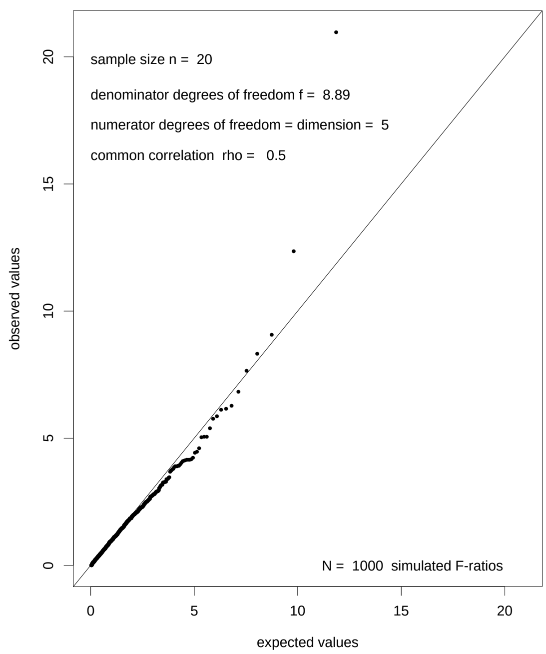

Figure 5a: Multivariate X-Chart Limits Based on F Approximation

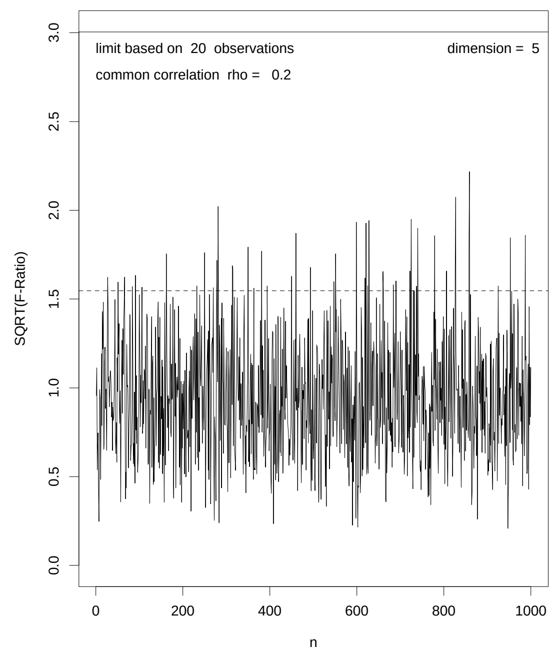

Figure 5b: Multivariate X-Chart Limits Based on F Approximation

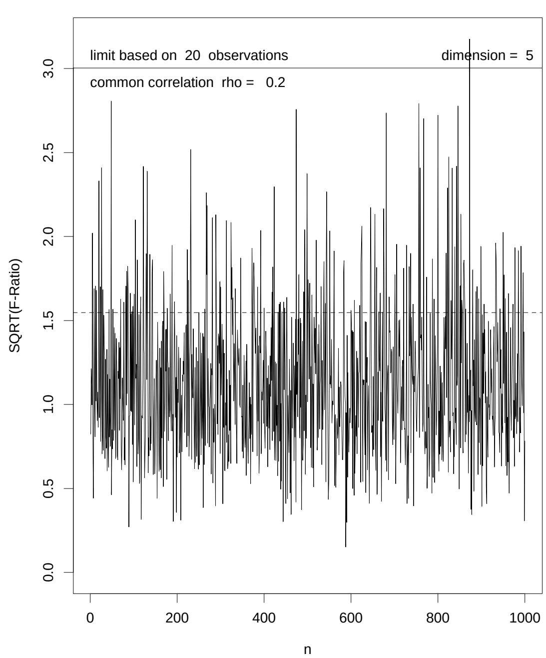

Figure 6: Multivariate X-Chart Limits Based on Updated F Approximation