บทความวิจัย (Research Article)

Performance Comparison of the Hotelling's T2 and DMEWMA Control Charts Using Bivariate Copulas for Small Shifts

Sasigarn Kuvattana*

Applied Statistics, Basic Science, Maejo University, Phrae Campus, Maejsai, Rong Kwang, Phrae 54140

Received: September 18, 2019; Accepted: December 28, 2019

Abstract

Multivariate control charts are an important tool in statistical process control for identifying an out-of-control process. Most multivariate control charts were designed to assume that the observations are an independence and normal distribution, but it is not valid in practice. This paper proposes the copulas modeling for dependence and non-normal multivariate cases and compare bivariate copulas on Hotelling's T^2 and double multivariate exponentially weighted moving average (DMEWMA) control charts. Observations are from an exponential distribution with Monte Carlo simulation when the parameter shifts are 1.02, 1.04, 1.06, 1.08, and 1.1. The level of dependence of observations is measured by Kendall's tau as 0.8 and -0.8 for normal, Frank and Clayton copulas. The performance of control charts is based on the average run length (ARL) in each copula. The results show that in the case of one and two-parameter shifts, the performance of the Hotelling's T^2 is better than DMEWMA control chart for all modifications.

Keywords: ARL; copula; Hotelling's T^2; DMEWMA; Monte Carlo simulation; exponential distribution

1. Introduction

Statistical process control (SPC) is an effective tool in simple manufacturing processes with only one process output variable or quality characteristic. In practice, most process monitoring and control scenarios involve more than one variable. A univariate control chart is the most type of SPC procedure for a single process characteristic. Multivariate methods that consider the variables jointly are required. Multivariate statistical process control (MSPC) charts have been regarded as the multivariate extensions of the univariate charts (Montgomery, 2013). Several multivariate quality control charts have been proposed to monitor the quality characteristics. Most multivariate detection procedures are based on a multi- normality assumption and independence, but many processes assume non- normality and correlation.

*Corresponding author: sasigarn@mju.ac.th
DOI: 10.14456/tjst.2020.78

Copulas approach is a representation by Sklar (1959 and 1973), which has become a popular tool for modeling non- linearity, asymmetrically, and tail dependence in several fields; it can be used in the study of dependence or association between random variables. The copulas can estimate the joint distribution of nonlinear outcomes and describe the dependence structure among variables through the joint distribution by eliminating the effect of univariate marginals. Bivariate copulas are the simplest case for the description of dependent random variables, and they can be used with control charts. Recent papers have applied copulas on control charts such as, Sukparungsee *et al*. (2018) proposed five types of copulas on the Hotelling's T^2 control chart; Fatahi *et al*. (2011) studied the joint distribution of two correlated zero-inflated Poisson (ZIP) distributions using the copula function approach; Fatahi *et al*. (2012) develop a copula-based bivariate ZIP control chart which can be used for monitoring correlated rare events; Dokouhaki and Noorossana (2013) applied the Markov approach for modeling the auto-correlated data, and the copula approach is used making the joint distribution of two auto-correlated binary data series; Hryniewicz (2012) presented the concept of copulas to model dependencies of other types on Shewhart control charts for auto-correlated and normal data. Verdier (2013) proposed a new approach for the non-normal multivariate case and constructed a tolerance region obtained from a density level set estimation on the Hotelling's T^2 control chart.

This paper presents a comparative performance of the Hotelling's T^2 and double multivariate exponentially weighted moving average (DMEWMA) control charts when observations are from an exponential distribution with the mean shifts and use bivariate copulas for specifying dependence between random variables.

2. Hotelling's T^2 control chart

Suppose that $\overline{\mathbf{X}}$ and $\mathbf{S}$ are the sample mean vector and covariance of the matrix, respectively. Let *m* and *p* be the number of samples and the number of quality characteristics observed in each sample, respectively. The Hotelling's T^2 statistic is $T^2 = \left(\mathbf{X}-\overline{\mathbf{X}}\right)' \mathbf{S}^{-1} \left(\mathbf{X}-\overline{\mathbf{X}}\right)$ (1), where $\overline{\mathbf{X}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_i$ and $\mathbf{S} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{X}-\overline{\mathbf{X}}\right)\left(\mathbf{X}-\overline{\mathbf{X}}\right)'$.

Statistical process control is usually split into two phases (Montgomery, 2013). Phase I constitutes a retrospective analysis, constructing trial control limits to determine if the process has been in control. Once this is achieved, the controlled data are used in Phase II to monitor the process. The control chart has a centerline representing the average value of the quality characteristic corresponding to the in-control process. Two other horizontal lines, called the upper control limit (UCL) and the lower control limit (LCL), are shown on the chart.

For phase I, the control limits are $\text{UCL} = \frac{(m-1)^2}{m}\beta_{\alpha,\, p/2,\, (m-p-1)/2}$ and LCL = 0 where $\beta_{\alpha,\, p/2,\, (m-p-1)/2}$ is an upper α percentage point of beta distribution with parameters *p*/2 and (*m*-*p*-1)/2 (Bersimis *et al.*, 2007). For the phase II control limits for this statistic are $\mathrm{UCL} = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha,\, p,\, (m-p)}$ and LCL=0 where $F_{\alpha,\, p,\, (m-p)}$ is *F* distribution with parameters *p* and *(m-p)*. Note that this article will mostly focus on Phase II control charts and their performance.

3. Double multivariate exponentially weighted moving average control chart (DMEWMA)

Suppose that $\mathbf{x}_1, \mathbf{x}_2, \ldots$ are $p \times 1$ random vectors each representing the *p*-variate normal distribution $\mathrm{N}(\mu_0, \Sigma_0)$, with mean vector μ_0 and variance-covariance matrix Σ_0 for in-control. The MEWMA statistic $\mathbf{y}_i$ and $\mathbf{z}_i$ for *i* = 1, 2,…; which is written by $\mathbf{y}_i = \Lambda\, \mathbf{x}_i + (\mathbf{I} - \Lambda)\mathbf{y}_{i-1}$ (2) and $\mathbf{z}_i = \Lambda\, \mathbf{y}_i + (\mathbf{I} - \Lambda)\mathbf{z}_{i-1}$ (3) where $\mathbf{z_0} = \mathbf{0}$ and Λ is a diagonal matrix with entries $\lambda_1, \ldots, \lambda_p$. The first equation in (2) is just the MEWMA statistic $\mathbf{y}_i$ calculated from $\mathbf{x}_i$ and the second equation in (3) is the MEWMA statistic $\mathbf{z}_i$ calculated from $\mathbf{y}_i$. The original data $\mathbf{x}_i$ is double smoothed, such that $\mathbf{z}_i$ is called the DMEWMA statistics, with $\Lambda = \lambda \mathbf{I}, 0 \le \lambda \le 1$ and $\mathbf{y}_0 = \mathbf{z}_0 = \boldsymbol{\mu}_0$. The DMEWMA control chart statistics T_{di}^2 is $T_{di}^2 = \mathbf{z}_i' \sum_{\mathbf{Z}_i}^{-1} \mathbf{z}_i$ (4) where $\sum_{\mathbf{Z}_i}^{-1}$ is the inverse of the exact variance-covariance matrix of $\mathbf{z}_i$, and $\sum_{\mathbf{Z}_i} = \frac{\lambda^4\left[1+(1-\lambda)^2-(i+1)^2(1-\lambda)^{2i}+(2i^2+2i-1)(1-\lambda)^{2i+2}-i^2(1-\lambda)^{2i+4}\right]}{\left[1-(1-\lambda)^2\right]^3} \sum_0$ (5) let $i \to \infty$, the asymptotic variance-covariance matrix is: $\sum_{\mathbf{Z}_i} = \frac{\lambda\,(2-2\lambda+\lambda^2)}{(2-\lambda)^3} \sum_0$ (6) if $T_{di}^2 > h$ then the signal gives an out-of-control, where *h* is the control limit (Alkahtani and Schaeffer, 2012; Abdella *et al.*, 2018).

4. Copulas modeling

Copulas introduced by Sklar (see Sklar, 1959). According to Sklar's theorem for a bivariate case, let *X* and *Y* be continuous random variables with joint distribution function *H* and marginal cumulative distribution *F*(*x*) and *F*(*y*) , respectively. Then $H(x, y) = C(F(x), F(y); \theta)$ with a copula $C : [0,1]^2 \to [0,1]$, where θ is a parameter of the copula called the dependence parameter, which measures dependence between the marginals. For the purposes of the statistical method, it is desirable to parameterize the copula function. Let θ denote the association parameter of the bivariate distribution, and there exists a copula *C*. Then *F*(*x*) = *u* and *F*(*y*) = *v*, where *u* and *v* are uniformly distributed variates (Trivedi and Zimmer, 2005). The copulas function has two families as follows:

4.1 Elliptical copulas

Elliptical copulas are simply the copulas of elliptical distributions. Simulation from elliptical distributions is easy, and as a consequence of Sklar's Theorem is a simulation from elliptical copulas. Two common elliptical copulas are the Gaussian (normal) and Student's t distributions. Each of these copulas can be extended to *d*-dimensional space, but this paper only focuses on the Normal copula. The Normal copula is an elliptical copula is defined as: $C(u,v;\theta) = \Phi_N(\Phi^{-1}(u), \Phi^{-1}(v);\theta)$; $-1 \le \theta \le 1$ (7) where $\Phi_N(u,v)$ is the cumulative probability distribution function of the bivariate normal distribution, $\Phi^{-1}(u)$ and $\Phi^{-1}(v)$ are the inverse of the cumulative probability function of the univariate normal distribution.

4.2 Archimedean copulas

Let Φ be a class of continuous, strictly decreasing functions $\phi:[0,1] \to [0,\infty]$ such that $\phi(1)=0, \phi'(t)<0$ and $\phi''(t)>0$ for all $0<t<1$ (Nelsen, 2006; Genest and McKay, 1986; Genest and Rivest, 1993). These are two types of Archimedean copulas generated as follows:

4.2.1 Frank Copula

$$C(u,v;\theta) = -\frac{1}{\theta} ln(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}), \quad (8)$$

where $\phi(t) = -ln(\frac{e^{-\theta t}-1}{e^{-\theta}-1})$; $\theta \in (-\infty,\infty) \setminus 0$.

4.2.2 Clayton copula

$$C(u,v;\theta) = \left[max(u^{-\theta}+v^{-\theta}-1,0)\right]^{-1/\theta}, \quad (9)$$

where $\phi(t) = (t^{-\theta}-1)/\theta$; $\theta \in (-1,\infty) \setminus 0$.

5. Dependence measures for data

Generally, a parametric measure of the linear dependence between random variables is the correlation coefficient, and nonparametric measures of dependence are Spearman's rho and Kendall's tau. According to the earlier literature, copulas can be used to study dependence or association between random variables. The values of Kendall's tau are easy to calculate, so this measure is used for observation dependencies.

Let *X* and *Y* be continuous random variables whose copula is *C* then Kendall's tau for *X* and *Y* is given by $\tau_c = 4\iint_{\mathbf{I}^2} C(u,v)\, dC(u,v) - 1$ where τ_c is Kendall's tau of copula *C* and the unit square $\mathbf{I}^2$ is the product $\mathbf{I} \times \mathbf{I}$ where $\mathbf{I} = [0,1]$ and the expected value of the function $C(u,v)$ of uniform (0,1) random variables U and V whose joint distribution function is *C*, i.e., $\tau_c = 4E[C(U,V)] - 1.$

Genest and McKay (1986) considered Archimedean copula *C* generated by ϕ, then $\tau_{Arch} = 4\int_0^1 \frac{\phi(t)}{\phi'(t)} dt + 1$ where τ_{Arch} is Kendall's tau of Archimedean copula *C*.

5.1 Normal copula

$$\tau = \frac{arcsin(\theta)}{\pi/2} \; ; \; \theta \in [-1,1] \quad (10)$$

5.1.1 Frank copula

$$\tau = 1 + \frac{4\left(\frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1} dt - 1\right)}{\theta} \; ; \; \theta \in (-\infty,\infty) \setminus \{0\} \quad (11)$$

5.1.2 Clayton copula

$$\tau = \frac{\theta}{\theta+2} \; ; \; \theta \in [-1,\infty) \setminus \{0\} \quad (12)$$

6. Performance of control charts

The basic characteristic that describes control charts' performance is the Average Run Length (ARL), which is the average number of points that must be plotted before a point indicates an out-of-control. ARL is classified into ARL_0 and ARL_1, where ARL_0 is the Average Run Length when the process is in-control and ARL_1 is the Average Run Length when the process is out-of-control.

In this paper, we use a Monte Carlo simulation in R statistical software (Mächler and Zurich, 2011-2012; Yan, 2007) with the number of simulation runs 50,000 and sample sizes 1,000. Observations were from an exponential distribution with in-control parameter $\alpha = 1.$ The shift size is reported in terms of the quantity $\delta = \mu - \mu_0$ and large values of δ correspond to bigger shift in the mean. The value $\delta = 0$ and the process mean $\mu = 1$ are in-control. The process means are 1.02, 1.04, 1.06, 1.08, and 1.1 for out-of-control.

The simulation results were carried out to evaluate the performance of the Hotelling's T^2 control chart, and the DMEWMA control chart with $\lambda = 0.05.$ Copulas estimations are restricted to the cases of positive and negative dependence. For all copula models, the setting θ corresponds with Kendall's tau. The level of dependence is measured by Kendall's tau values $(-1 \le \tau \le 1)$ which are defined to 0.8 and -0.8, respectively.

7. Numerical results

The results are presented in Tables 1-4, and the different values of exponential parameters denote μ_1 for the variables *X* and μ_2 for the variables *Y*. For in-control, control charts were chosen by setting the desired ARL_0 = 370 for each copula. Tables 1-2 show strong positive dependence $(\tau = 0.8)$ and Tables 3-4 show strong negative dependence $(\tau = -0.8)$.

Table 1 ARL_0 and ARL_1 values of control charts with Kendall's tau equal to 0.8 in the case of one parameter shifts at μ_1 or μ_2.

Parameters shifts		Types of control charts					
		Hotelling's T^2			DMEWMA		
μ_1	μ_2	Normal	Frank	Clayton	Normal	Frank	Clayton
1	1	369.873	370.159	370.009	369.739	370.036	369.904
1	1.02	355.470	***352.910***	356.129	358.012	356.476	359.604
1	1.04	***336.555***	340.471	343.186	344.642	340.732	344.840
1	1.06	***320.981***	328.888	329.263	330.497	328.521	332.529
1	1.08	***304.142***	315.771	314.523	316.389	314.808	316.656
1	1.1	***286.338***	301.977	299.246	304.480	301.490	302.728
1	1	369.873	370.159	370.009	369.739	370.036	369.904
1.02	1	***353.405***	354.389	355.886	359.709	358.398	358.315
1.04	1	***338.413***	344.341	344.282	344.070	343.506	342.606
1.06	1	***319.972***	329.466	329.822	328.720	326.768	331.008
1.08	1	***303.122***	316.329	314.385	315.598	314.321	315.439
1.1	1	***287.430***	300.675	299.802	303.648	300.635	300.328

The result in Table 1 shows the mean shifts of when is fixed at 1, for the parameter shift = 1.02, the ARL1 value of the Frank copula on the Hotelling's T2 control chart is less than the other copulas; and the other shifts, the ARL1 values of the Normal copula on the Hotelling's T2 control chart are less than the other copulas. When is fixed at 1, the

Table 2 ARL_0 and ARL_1 values of control charts with Kendall's tau equal to 0.8 in the case of two parameter shifts at μ_1 and μ_2 .

Parameters shifts		Type of control charts					
		Hotelling's T^2			DMEWMA		
μ_1	μ_2	Normal	Frank	Clayton	Normal	Frank	Clayton
1	1	369.873	370.159	370.009	369.739	370.036	369.904
1.02	1.02	***339.930***	344.763	343.462	346.905	346.749	345.967
1.04	1.04	***313.508***	322.770	317.983	325.373	320.750	321.439
1.06	1.06	***288.406***	298.431	296.771	301.847	298.082	302.819
1.08	1.08	***266.017***	280.742	274.405	283.412	279.967	279.428
1.1	1.1	***245.511***	259.140	254.223	266.924	261.138	260.614

Table 3 ARL_0 and ARL_1 values of control charts with Kendall's tau equal to -0.8 in the case of one parameter shifts at μ_1 or μ_2 .

Parameters shifts		Type of control charts					
		Hotelling's T^2			DMEWMA		
μ_1	μ_2	Normal	Frank	Clayton	Normal	Frank	Clayton
1	1	370.012	369.959	370.109	370.030	369.968	369.841
1	1.02	356.011	356.207	***354.420***	358.121	358.255	356.991
1	1.04	341.964	341.236	***338.637***	343.603	344.349	343.688
1	1.06	325.497	325.807	***323.094***	330.954	328.921	329.797
1	1.08	312.953	***311.910***	312.068	318.684	315.088	319.627
1	1.1	300.119	296.065	***296.004***	305.383	305.584	305.930
1	1	370.012	369.959	370.109	370.030	369.968	369.841
1.02	1	***356.399***	356.833	356.942	357.208	356.907	357.052
1.04	1	342.439	***341.682***	342.127	342.793	342.144	343.917
1.06	1	325.547	329.568	***324.481***	332.379	331.893	329.941
1.08	1	***310.445***	311.245	312.284	317.710	316.308	319.074
1.1	1	297.094	***297.005***	299.235	305.751	303.375	305.368

Table 4 ARL_0 and ARL_1 values of control charts with Kendall's tau equal to -0.8 in the case of two parameter shifts at μ_1 and μ_2 .

Parameters shifts		Type of control charts					
		Hotelling's T^2			DMEWMA		
μ_1	μ_2	Normal	Frank	Clayton	Normal	Frank	Clayton
1	1	370.012	369.959	370.109	370.030	369.968	369.841
1.02	1.02	***343.269***	344.380	344.000	345.630	343.936	344.504
1.04	1.04	318.851	***317.002***	317.303	321.720	321.080	321.481
1.06	1.06	293.561	293.140	***290.101***	300.171	298.786	301.591
1.08	1.08	272.739	***268.715***	270.753	280.218	278.643	279.727
1.1	1.1	252.297	***249.462***	250.027	262.172	260.552	262.442

ARL1 values of the Normal copula on the Hotelling's T2 control chart are less than the other copulas for all shifts. Table 2 shows two-parameter shifts of the ARL1 values of the Normal copula on the Hotelling's T2 control chart are less than the other copulas for all shifts.

Table 3 shows the mean shifts of $\tau = -0.8$ when μ_1 is fixed at 1, for the parameter shifts μ_2 =1.02, 1.04, 1.06, and 1.1; the ARL_1 values of the Clayton copula on the Hotelling's T^2 control chart are less than the other copulas; except for the shift μ_2 =1.08, the ARL_1 value of the Frank copula on the Hotelling's T^2 control chart is less than the other copulas. When μ_2 is fixed at 1, the ARL_1 values of the Normal copula on the Hotelling's T^2 control chart are less than the other copulas at μ_1 =1.02 and 1.08; the ARL_1 values of the Frank copula on the Hotelling's T^2 control chart are less than the other copulas at μ_1 = 1.04 and 1.1; and the ARL_1 value of the Clayton copula on the Hotelling's T^2 control chart is less than the other copulas at μ_1 =1.06.

Table 4 shows two-parameter shifts of $\tau = -0.8,$ the ARL_1 value of the Normal copula on the Hotelling's T^2 control chart is less than the other copulas at the shift is 1.02; the ARL_1 values of the Frank copula on the Hotelling's T^2 control chart are less than the other copulas at the shifts are 1.04, 1.08 and 1.1; the ARL_1 value of the Clayton copula on the Hotelling's T^2 control chart is less than the other copulas at the shift is 1.06.

8. Conclusions

The results presented two control charts for dependence measures of two variables by copulas modeling based on ARL property. The ARL comparisons indicate that the Hotelling's T^2 control chart performs better than the DMEWMA control chart for all shifts when the parameter shifts are 1.02, 1.04, 1.06, 1.08, and 1.1. For strong positive dependence $(\tau = 0.8)$, in the case of one and two parameter shifts; the ARL_1 values of the Normal copula on the Hotelling's T^2 control chart are less than the other copulas for almost all shifts. For strong negative dependence $(\tau = -0.8)$, in the case of one and two parameter shifts; the ARL_1 values of three copulas on the Hotelling's T^2 control chart are less than the other copulas, but three types of copulas would be more sensitive.

9. Acknowledgements

This paper was partially supported by the grants project from Maejo University, Phrae Campus, in 2018.

10. References

Abdella, G.M., Al-Khalifa, N.K. and Hamouda, A.M., 2018, The performance assessment for the dMEWMA in capturing changes in simple linear profiles, pp. 1591-1606, International Conference on Industrial Engineering and Operations Management, Washington, DC.

Alkahtani, S. and Schaffer, J., 2012, A double Multivariate Exponentially Weighted moving average (dMEWMA) control chart for a process location monitoring, Commun. Stat. Simulat. 41: 1-15.

Bersimis, S., Psarakis, S. and Panaretos, J., 2007, Multivariate statistical process control charts: An overview, Qual. Reliab. Eng. Int. 23: 517-543.

Dokouhaki, P. and Noorossana, R., 2013, A copula Markov CUSUM chart for monitoring the bivariate auro-correlated binary observation, Qual. Reliab. Eng. Int. 29: 911-919.

Fatahi, A.A., Dokouhaki, P. and Moghaddam, B.F., 2011, A bivariate control chart based on copula function, pp. 292-296, IEEE International Conference on Quality and Reliability (ICQR), Bangkok.

Fatahi, A.A., Noorossana, R., Dokouhaki, P. and Moghaddam, B.F., 2012, Copula-based bivariate ZIP control chart for monitoring rare events, Commun. Stat. Theor. M. 41: 2699-2716.

Genest, C. and MacKay, J., 1986, The joy of copulas: bivariate distributions with uniform marginal, Am. Sat. 40: 280-283.

Genest, C. and Rivest, L.P., 1993, Statistical inference procedures for bivariate Archimedean copulas, J. Am. Stat. Assoc. 88: 1034-1043.

Hryniewicz, O., 2012, On the robustness of the Shewhart control chart to different types of dependencies in data, Front. Stat. Qual. Contr. 10: 20-33.

Khoo, B.C., Atta, M.A. and Phua, H.N., 2009, A study on the performances of MEWMA and MCUSUM charts for skewed distributions, pp. 817-822, 10th Islamic Countries Conference on Statistical Science, Cairo.

Kuvattana, S., Sukparungsee, S., Busababodhin, P. and Areepong, Y., 2018, Bivariate copulas on the Hotelling's T^2 control chart, Commun. Stat. Simulat. 47: 413-419.

Mächler, M. and Zurich, E., 2011, Numerically stable Frank copula functions *via* multiprecision: R package Rmpfr, Available Source: http://cran.r-project.org/web/packages/copula/vignettes/Frank-Rmpfr.pdf, March 1, 2016.

Montgomery, D. C. , 2013, Statistical Quality Control: A Modern Introduction, 7th Ed. , John Wiley & Sons, Singapore.

Nelsen, R. B. , 2006, An introduction to copulas, 2nd Ed, Springer, New York.

Sklar, A. ,1959, Fonctions de repartition a n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8: 229-231.

Sklar, A. , 1973, Random variables, joint distributions, and copulas, Kybernetica 9: 449-460.

Trivedi, P. K. , and Zimmer, D. M. , 2005, Copula Modeling: An Introduction for Practitioners, Foundations and Trends in Econometrics.

Verdier, G. , 2013, Application of copulas to multivariate control charts, J. Stat. Plan. Infer. 143: 2151-2159.

Yan, J. , 2007, Enjoy the joy of copulas: With a package copula, J. Stat. Softw. 21: 1-21.