Efficiency of Bivariate Copula on the CUSUM Chart

Sasigarn Kuvattana, Saowanit Sukparungsee, Piyapatr Busababodhin, and Yupaporn Areepong

***Abstract*—The objective of this paper will propose four types of copulas on the CUSUM control chart when observations are exponential distribution. We use the Monte Carlo simulation to investigate the value of Average Run Length (ARL). The dependence of random variables are used and measured by Kendall's tau in each copula. The numerical results show that negative dependence Normal copula is better than the others. For positive dependence, in the case of two parameter shifts, Normal copula is better than others and Gumbel copula is better than others in the case of one parameters shift.**

***Index Terms*—Copula, Average run length, CUSUM chart, Monte Carlo simulation**

I. Introduction

STATISTICAL Process Control (SPC) is a method for monitoring, controlling and improving quality of production in many areas of applications. These areas are in industry, finance and economics, health care, environment sciences and other fields. Control charts are statistical and visual tools designed to detect shifts in a process and they are designed and evaluated under the assumption that the observations are from processes are independent and identically distributed (i.i.d.). A Univariate control chart is devised to monitor the quality of a single process characteristic [1]. However, modern process often monitor more than one quality characteristic and they are referred to as multivariate statistical process control charts.

Multivariate statistical process control (MSPC) is one of the most rapidly developing sections of statistical process control [2] and lead to an interest in the simultaneous inspection of several related quality characteristics [3, 4]. There are multivariate extensions for all kinds of univariate control charts, such as multivariate Shewhart control chart, multivariate exponentially weighted moving average control chart (MEWMA) and multivariate cumulative sum control chart (MCUSUM) [5]. Multivariate Shewhart control chart is used to detect large shifts in the mean vectors. The MEWMA and MCUSUM are commonly used to detect small or moderate shifts in the mean vectors [6].

Most of multivariate detection procedures are based on a multi-normality assumption and independence but many processes are often non-normality and correlated. Many multivariate control charts are the lack of the related joint distribution and copula can specify this property. Copulas introduced by Sklar [7], are useful devices which give a representation of a multivariate distribution function in terms of its univariate marginal distribution [8]. The copula approach has become a popular tool for modeling nonlinearity, asymmetricality and tail dependence in several fields [9] and it can be used in the study of dependence or association between random variables. Copulas modeling can estimate joint distribution of nonlinear outcomes and explain the dependence structure among variables through the joint distribution by eliminating the effect of univariate marginals. A bivariate copula is the simplest case for the description of dependent random variables and it can apply to control chart.

Recently, several papers use copula in control chart such as, copula based on bivariate ZIP control chart [10, 11], copula Markov CUSUM chart [12], Shewhart control charts for autocorrelated and normal data [13], new control chart based on a nonparametric Kendall's tau statistics [14], non-normal multivariate cases for the Hotelling T^2 control chart [15] and bivariate copula on the Shewhart control chart [16].

This paper presents the work on the CUSUM control chart when observations are exponential distribution with the means shifts and use a bivariate copula function for specifying dependence between random variables.

II. The Multivariate Cumulative Sum Control Chart

The multivariate cumulative sum (MCUSUM) control chart is the multivariate extension of the univariate cumulative sum (CUSUM) chart. The MCUSUM chart was initially proposed by Crosier [17]. The MCUSUM chart may be expressed as follows:

$$C_t = [(\mathbf{S}_{t-1} + \mathbf{X}_t - \mathbf{a})' \textstyle\sum^{-1} (\mathbf{S}_{t-1} + \mathbf{X}_t - \mathbf{a})]^{1/2},\ t = 1,\ 2,\ 3, \ldots \quad (1)$$

where covariance $\left(\sum\right)$ and $\mathbf{S}_t$ are the cumulative sums expressed as:

$$\mathbf{S}_t = \begin{cases} \mathbf{0}, & \text{if } C_t \le k \\ (\mathbf{S}_{t-1} + \mathbf{X}_t - \mathbf{a})\left(1 - \dfrac{k}{C_t}\right), & \text{if } C_t > k \end{cases} \quad (2)$$

Manuscript received January 10, 2015; revised February 3, 2015. This work was supported in part by the Graduate Colleges, King Mongkut's University of Technology North Bangkok.

S. Kuvattana is with the Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand (e-mail: sasigarn2010@gmail.com).

S. Sukparungsee is with the Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand (e-mail: swns@kmutnb.ac.th).

P. Busababodhin is with the Department of Mathematics, Faculty of Science, Mahasarakham University, Mahasarakham, 41150, Thailand (e-mail: piyapatr99@gmail.com).

Y. Areepong is with the Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand (e-mail:yupaporna@kmutnb.ac.th).

with $\mathbf{S}_t = 0,$ the reference value $k > 0$ and $\mathbf{a}$ is the aim point or target value for the mean vector [18]. The control chart statistics for MCUSUM chart is

$$Y_t = [\mathbf{S}_t' \sum^{-1} \mathbf{S}_t]^{1/2},\ t = 1,\ 2,\ 3,\ldots \qquad (3)$$

The signal gives an out-of-control if $Y_t > h$ where h is the control limit [19].

III. Copula Function

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 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 $F(x) = u,\ F(y) = v$ where u and v are uniformly distributed variates [20]. This paper focuses on Normal copula and three types of Archimedean copulas which are Clayton, Frank and Gumbel [21].

A. Normal copula

$$C(u,v;\ \theta) = \Phi_N(\Phi^{-1}(u), \Phi^{-1}(v);\ \theta);\ \ -1 \le \theta \le 1, \qquad (4)$$

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.

B. Archimedean copulas

Let a class Φ of functions $\phi:[0,1] \to [0,\infty]$ with continuous, strictly decreasing, such that $\phi(1) = 0,\ \phi'(t) < 0$ and $\phi''(t) > 0$ for all $0 < t < 1$ [21-23]. Archimedean copulas of three types and these types are generated as follow:

Clayton copula

$$C(u,v;\ \theta) = \left[max(u^{-\theta} + v^{-\theta} - 1,\ 0)\right]^{-1/\theta}, \qquad (5)$$

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

Frank copula

$$C(u,v;\theta) = -\frac{1}{\theta} ln(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}), \qquad (6)$$

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

Gumbel copula

$$C(u,v;\theta) = \exp(-[(-ln\ \mathrm{u})^{\theta} + (-ln\ \mathrm{v})^{\theta}]^{1/\theta}), \qquad (7)$$

where $\phi(t) = [-ln(t)]^{\theta}$; $\theta \in [1,\infty)$.

IV. Dependence Measures For Data

Generally, a parametric measure of the linear dependence between random variables is correlation coefficient and nonparametric measures of dependence are Spearman's rho and Kendall's tau. According to the earlier literature, the copulas can be used in the study of dependence or association between random variables and 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 $\tau_c = 4\iint_{\mathrm{I}^2} C(u,v)\ dC(u,v) - 1$ where τ_c is Kendall's tau of copula C and the unit square I^2 is the product $\mathrm{I}\times\mathrm{I}$ where $\mathrm{I} = [0,1]$ 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., $\tau_c = 4E[C(U,V)] - 1$ [22].

Genest and McKay [21] 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.

TABLE I
KENDALL'S TAU OF COPULA FUNCTION

Copula	Kendall's tau	Parameter space of θ
Normal	$arcsin(\theta)/(\pi/2)$	$[-1,1]$
Clayton	$\theta/(\theta+2)$	$[-1,\infty)\setminus\{0\}$
Frank	$1+4\left(\frac{1}{\theta}\int_0^{\theta}\frac{\mathrm{t}}{\mathrm{e}^{\mathrm{t}}-1}\ \mathrm{dt}\ \text{-1}\right)/\theta$	$(-\infty,\infty)\setminus\{0\}$
Gumbel	$(\theta-1)/\theta$	$[1,\infty)$

V. Average Run Length

The basic characteristics that describe the performance of control charts is the Average Run Length (ARL). 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 [24]. Theoretically, an acceptable ARL_0 should be enough large when the process is in control and ARL_1 should be small when the process is out-of-control. The copula approach focus on four types: the Normal, Clayton, Frank and Gumbel. These copulas are implemented in the R statistical software [25-27] with the number of simulation runs 50,000. Observations were from exponential distribution with parameter (α) equal to 1 for in control process ($\mu_0 = 1$) and the shifts of the process level (δ) by $\mu = \mu_0 + \delta$. The shifts in process mean are equal to 1, 2 and 3, and sample size is 1,000.

The simulation experiments carried out to assess the performance of CUSUM control chart. Copula estimations are restricted to the cases of dependence (positive and negative dependence). For all copula models, setting θ correspondes with Kendall's tau. The level of dependence is measured by Kendall's tau values ($-1 \le \tau \le 1$). For moderate and strong dependence, Kendall's tau values are defined to 0.5 and 0.8, respectively.

VI. Simulation Results

The results from simulation experiments are presented in Table II – V for the different values of Kendall's tau and denote μ_1 for the variables X and μ_2 for the variables Y. The CSUM control chart was chosen by setting the desired $ARL_0 =$ 370 for each copulas. Table II and III show positive dependence $(\tau > 0)$ and Table IV and V show negative dependence $(\tau < 0)$. For example, Table II shows moderate and strong positive dependence when the shifts in one of exponential parameters. In the case of moderate dependence $(\tau = 0.5)$, for small shifts $(\mu_1 = 2, \mu_2 = 2)$, the ARL_1 values of Normal copula are less than the other copula. For moderate and large shifts $(3 \le \mu_1 \le 4,\ 3 \le \mu_2 \le 4)$, the Gumbel values for ARL_1 are less than the other copula. In the case of strong positive dependence $(\tau = 0.8)$, for all shifts $(2 \le \mu_1 \le 4,\ 2 \le \mu_2 \le 4)$, the ARL_1 values of Gumbel copula are less than the other copula. For example, Table III shows moderate and strong positive dependence and the same shifts in two of exponential parameters, the Normal copula give smaller values of ARL_1 than the other copula for almost all shifts.

TABLE II
ARL OF CUSUM CONTROL CHART WITH KENDALL'S TAU VALUES EQUAL TO 0.5 AND 0.8 FOR THE CASE OF ONE OF EXPONENTIAL PARAMETERS SHIFT.

τ	Parameters		ARL_0 and ARL_1			
	μ_1	μ_2	Normal	Clayton	Frank	Gumbel
0.5	1	1	370.017	370.047	370.016	370.197
	1	2	*41.357*	42.626	42.433	42.246
	1	3	12.203	13.216	12.807	*12.076*
	1	4	5.724	6.325	6.058	*5.613*
	2	1	*41.243*	42.613	42.851	42.166
	3	1	12.191	13.100	12.834	*12.036*
	4	1	5.779	6.265	6.109	*5.602*
0.8	1	1	369.860	370.041	370.129	370.059
	1	2	25.383	38.588	36.230	*24.583*
	1	3	6.387	10.622	9.336	*6.116*
	1	4	2.779	4.720	4.074	*2.654*
	2	1	25.372	38.869	36.091	*24.921*
	3	1	6.360	10.643	9.335	*6.129*
	4	1	2.752	4.691	4.045	*2.668*

TABLE III
ARL OF CUSUM CONTROL CHART WITH KENDALL'S TAU VALUES EQUAL TO 0.5 AND 0.8 FOR THE CASE OF TWO OF EXPONENTIAL PARAMETERS SHIFT.

τ	Parameters		ARL_0 and ARL_1			
	μ_1	μ_2	Normal	Clayton	Frank	Gumbel
0.5	1	1	370.017	370.047	370.016	370.197
	2	2	22.482	*21.785*	22.500	24.425
	3	3	*6.232*	6.309	6.441	6.697
	4	4	*2.442*	2.507	2.560	2.654
0.8	1	1	369.860	370.041	370.129	370.059
	2	2	*27.655*	30.540	31.957	28.758
	3	3	*7.911*	9.762	10.005	8.263
	4	4	*3.304*	4.419	4.513	3.547

TABLE IV
ARL OF CUSUM CONTROL CHART WITH KENDALL'S TAU VALUES EQUAL TO -0.5 AND -0.8 FOR THE CASE OF ONE OF EXPONENTIAL PARAMETERS SHIFT.

τ	Parameters		ARL_0 and ARL_1		
	μ_1	μ_2	Normal	Clayton	Frank
-0.5	1	1	370.098	369.858	369.961
	1	2	*40.574*	45.504	42.004
	1	3	*12.449*	13.960	12.654
	1	4	*5.914*	6.621	6.083
	2	1	*40.302*	45.344	41.924
	3	1	*12.410*	13.947	12.870
	4	1	*5.947*	6.651	6.117
-0.8	1	1	369.840	370.073	369.898
	1	2	*44.554*	47.062	45.447
	1	3	*12.724*	14.110	13.109
	1	4	*5.758*	6.455	5.956
	2	1	*44.329*	47.452	44.831
	3	1	*12.670*	13.940	13.162
	4	1	*5.770*	6.377	6.004

TABLE V
ARL OF CUSUM CONTROL CHART WITH KENDALL'S TAU VALUES EQUAL TO -0.5 AND -0.8 FOR THE CASE OF TWO OF EXPONENTIAL PARAMETERS SHIFT.

τ	Parameters		ARL_0 and ARL_1		
	μ_1	μ_2	Normal	Clayton	Frank
-0.5	1	1	370.098	369.858	369.961
	2	2	*20.317*	22.194	21.139
	3	3	*5.513*	5.913	5.670
	4	4	*2.084*	2.129	2.101
-0.8	1	1	369.840	370.073	369.898
	2	2	*28.865*	28.945	29.362
	3	3	8.110	*8.041*	8.190
	4	4	3.053	*2.968*	3.079

VII. Conclusion

Dependence measures of two or more variables can be investigated in term of various copulas. In this paper show CUSUM control chart for four types of copulas and level of dependence are measured by Kendall's tau values. Table VI summarize from Table II – V which show that for negative dependence Normal copula is better than the others and for positive dependence, in the case of the shift in two of parameters, Normal copula is better than the others and in the case of the shift in one of parameters, Gumbel copula is better than the others.

TABLE VI
EFFICIENCY OF COPULA DEPENDS ON POSITIVE AND NEGATIVE DEPENDENCES

Parameters shifts	Positive Dependence		Negative Dependence	
	moderate	strong	moderate	strong
one parameter	Normal	Normal	Normal	Clayton
two parameters	Gumbel	Gumbel	Normal	Normal

Acknowledgement

The authors would like to thank the King Mongkut's University of Technology North Bangkok and Thailand Ministry of Science and Technology for the financial support during this research.

References

[1] S.K. Kim, W. Jitpitaklert, S.-K. Park, S.-J. Hwang, "Data mining model-based control charts for multivariate and autocorrelated processes," Expert Systems with Applications, vol. 39, pp. 2073-2081. 2012.

[2] C. Fuchs and R.S. Kenett, *Multivariate quality control: theory and applications*. New York: Marcel Dekker,INC. 1985.

[3] C.A. Lowry and D.C. Montgomery, "A review of multivariate control charts," *IIE Transactions*, vol. 27, pp. 800-810. 1995.

[4] W.H. Woodall, D.C. Montgomery, "Research issues and ideas in statistical process control," *Journal of Quality Technology*, vol. 31, pp. 376-386. 1999.

[5] S. Bersimis, J. Panaretos and S. Psarakis, "Multivariate Statistical Process Control Charts and the Problem of Interpretation: A Short Overview and Some Applications in Industry," *Published in: Proceedings of the 7th Hellenic European Conference on Computer Mathematics and its Applications*, Athens Greece, 2005.

[6] H.M. Midi and A. Shabbak, "Robust Multivariate Control Charts to Detect Small Shifts in Mean," *Mathematical Problems in Engineering*, vol. 2011, Special section, pp.1-19. 2011.

[7] A. Sklar, "Fonctions de repartition a n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris* , vol. 8, pp. 229-231. 1959.

[8] J.-M. Kim, E.A. Sungur, T.C. and T.Y. Heo, "Generalized bivariate copulas and their properties," *Model Assisted Statistics and Applications*, vol. 6, pp. 21-136. 2011.

[9] X. Chen, W.B. Wu and Y. Yi, "Efficient Estimation of Copula-Based Semiparametric Markov Models," *The Annals of Statistics*, vol. 37, no. 6B, pp. 4214–4253. 2009.

[10] A.A. Fatahi, , P. Dokouhaki and B.F. Moghaddam, "A bivariate control chart based on copula function," *IEEE International Conference on Quality and Reliability (ICQR)*, Bangkok, Thailand, 14-17 Sept, 2011, pp. 292 – 296.

[11] A.A. Fatahi, R. Noorossana, P. Dokouhaki and B.F. Moghaddam, "Copula-based bivariate ZIP control chart for Monitoring rare events," *Communications in statistics theory and methods*, vol. 41, pp. 2699 – 2716, 2012.

[12] P. Dokouhaki and R. Noorossana, "A copula Markov CUSUM chart for Monitoring the bivariate auto-correlated binary observation," *Quality and Reliability Engineering International*, vol. 29, pp. 911-919, 2013.

[13] O. Hryniewicz, "On the robustness of the Shewhart control chart to different types of dependencies in data," *Frontiers in Statistical Quality Control*, vol. 10, pp. 20-33, 2012.

[14] O. Hryniewicz and A. Szediw, "Sequential signals on a control chart based on nonparametric statistical tests," *Frontiers in Statistical Quality Control*, vol. 9, pp. 99-108. 2010.

[15] G. Verdier, "Application of copulas to multivariate control charts," *Journal of Statistical Planning and Inference*, vol. 143, pp. 2151–2159, 2013.

[16] S. Kuvattana, S. Sukparungsee, P. Busababodhin and Y. Areepong, "Efficiency of bivariate copula on the Shewhart control chart," 21st International Conference on Computational Statistics(COMPSTAT2014), Geneva, 19-22 August, 2014.

[17] R.B. Crosier, "Multivariate generalizations of cumulative sum quality-control schemes," *Technometrics*, vol. 30, pp. 291–303, 1988.

[18] B. C. Khoo, M. A. Atta and H. N. Phua, "A Study on the Performances of MEWMA and MCUSUM charts for Skewed Distributions," in *Proc.10th Islamic Countries Conference On Statistical Science*, Cairo, Egypt, 2009. pp 817 – 822.

[19] C. C. Alves, R. W. Samohyl and E. Henning, "Application of multivariate cumulative sum control charts(MCUSUM) for monitoring a machining process," *16th Int. Conf. on Industrial Engineering and Operations Management,* Sao Carlos, Brazil, 12-15 Oct, 2010, pp. 1-7.

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

[21] C. Genest and R. J. McKay, "The joy of copulas: bivariate distributions with uniform marginals," *American Statistician*, vol. 40, pp. 280–283, 1986.

[22] R. B. Nelsen, *An introduction to copulas*. 2nd ed. New York. Springer. 2006.

[23] C. Genest and L.-P. Rivest, "Statistical inference procedures for bivariate Archimedean copulas," *Journal of the American Statistical Association*, vol. 88, pp. 1034–1043, 1993.

[24] J. Busaba, S. Sukparungsee and Y. Areepong, "Numerical Approximations of Average Run Length for AR(1) on Exponential CUSUM," *Proceedings of IMECS 2012*, Hong Kong, 14-16 March, 2012.

[25] J. Yan, "Enjoy the joy of copulas: With a package copula," *Journal of Statistical Software*, vol. 21, pp. 1-21, 2007.

[26] M. Hofert, M. M̈achler and A.J. McNeil, "Likelihood inference for Archimedean copulas in high dimensions under known margins," *Journal of Multivariate Analysis*, vol.110, pp. 133–150, 2012.

[27] M. M̈achler and E. Zurich. (2013, March). Numerically stable Frank copula functions via multiprecision: R Package Rmpfr. [Online]. Available: http://cran.r-project.org/web/packages/copula/vignettes/Frank-Rmpfr.pdf