

# Directional Dependence of Genes using Survival Truncated FGM Type Modification Copulas

**Jong-Min Kim**<sup>1</sup>

*Statistics Discipline, Division of Science and Mathematics, University of Minnesota-Morris, Morris, MN, 56267, USA*

**Yoon-Sung Jung**

*Department of Statistics, Kansas State University, Manhattan, KS, 66506, USA*

**Tim Soderberg**

*Chemistry Discipline, Division of Science and Mathematics, University of Minnesota-Morris, Morris, MN, 56267, USA*

**Summary.** A multivariate distribution can be represented in terms of its underlying margins by binding these margins together using a copula function (Sklar, 1959). Here, we propose a new class of survival FGM type modification truncated copulas which quantify dependency and incorporate directional dependence. In addition, we apply our proposed methods to the analysis of directional dependence relationships between genes. Finally, we employ the Akaike Information Criterion (AIC) to check the goodness of fit for our proposed copula models.

*Keywords:* Survival copula; Farlie-Gumbel-Morgenstern copula; Directional dependence

## 1 Introduction

With the availability of increasingly large sets of gene expression data, there is a need for new methods to analyze these data for gene-gene dependence relationships. The ability to reconstruct gene networks from large sets of microarray data will facilitate analysis of cellular function at the molecular level, and will have a profound impact on many areas of biomedical research.

In statistics, there are two approaches to describing dependence structure: (i) setting up a functional relationship between the variables, and (ii) specifying the joint distribution of the variables. The second approach eliminates the effect of univariate marginals, which has nothing to do with

---

<sup>1</sup>*Address for correspondence:* Jong-Min Kim, Statistic Discipline, Division of Science and Mathematics, University of Minnesota-Morris, Morris, MN, 56267, USA Email: [jongmink@morris.umn.edu](mailto:jongmink@morris.umn.edu)

the dependence structure, and is much more general than approach (i). Copulas, first developed by Sklar (1959), are devices which give a representation of a multivariate distribution function in terms of its univariate marginal distributions. Recently, Rodríguez-Lallena and Úbeda-Flores (2004) have proposed a new class of bivariate copulas depending on two univariate functions. This class is a generalization of Farlie-Gumbel-Morgenstern (FGM) families of copulas. In this paper, we describe survival truncated FGM type modification copulas and their use in the analysis of directional dependence among genes. We investigate several properties of the new class of survival truncated FGM type modification copulas, including dependence properties and measures of association between two random variables given one truncated random variable.

This paper is organized as follows: Section 2 contains a description of the copulas and survival copulas. The new class of survival truncated copulas with directional dependence is explored in Section 3, while in section 4 the gene-gene directional dependence application is introduced. In Section 5 we summarize our results and suggest future research directions.

## 2 Definitions and Preliminaries

In this section, we review the basic concepts of copulas, focusing on some preliminary properties of copulas. In addition, we present some standard bivariate copula families. A copula is a multivariate distribution function defined on the unit  $[0, 1]^n$ , with uniformly distributed marginals. In this paper, we focus on a bivariate (two-dimensional) copula, where  $n = 2$ .

Sklar (1973) shows that any bivariate distribution function ( $F_{XY}$ ), can be represented as a function of its marginal distribution of  $X$  and  $Y$  ( $F_X$  and  $F_Y$ ) by using a two-dimensional copula  $C(\cdot, \cdot)$ . More specifically, the copula may be written as

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)) = C(u, v).$$

Therefore, the copula function represents how the multivariate function  $F_{XY}(x, y)$  is copuled with its marginal distribution functions,  $F_X(x)$  and  $F_Y(y)$ . Also, it describes the dependence mechanism between two random variables by eliminating the influence of the marginals or any monotone transformation on the marginals.

Two additional properties of copulas are the continuity property and differentiability property.

For the continuity property, let  $C$  be a copula. Then,

$$|C(u_2, v_2) - C(u_1, v_1)| \geq |u_2 - u_1| + |v_2 - v_1|,$$

$\forall u_1, u_2, v_1, v_2 \in [0, 1]$ ,  $u_1 < u_2$  and  $v_1 < v_2$ , hence, every copula  $C$  is uniformly continuous on its domains [Nelsen (1999)].

Let  $X, Y$  be random variables with continuous distribution functions  $F_X$  and  $F_Y$ , respectively. Then the Spearman's  $\rho$  and Kendall's  $\tau$  are given, respectively, by

$$\rho_C = 12 \int_0^1 \int_0^1 [C(u, v) - uv] dudv, \quad (1)$$

and

$$\tau_C = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1, \quad (2)$$

[Nelsen (1999)].

Cherubini (2004) describes the survival copula as follows:

**Definition 1** (Cherubini, 2004) *The survival copula associated with the copula  $C$ , is*

$$\bar{C}(v, z) = v + z - 1 + C(1 - v, 1 - z) \quad (3)$$

It is easy to verify that  $\bar{C}$  has copula properties. Once computed in  $(1 - v, 1 - z)$ , it represents the probability that two standard uniform variates with copula  $C$  are greater than  $v, z$  respectively, since

$$\begin{aligned} \bar{C}(1 - v, 1 - z) &= 1 - v + 1 - z - 1 + C(v, z) \\ &= 1 - P(U_1 \leq v) + 1 - P(U_2 \leq z) - 1 + P(U_1 \leq v, U_2 \leq z) \\ &= P(U_1 > v) + P(U_2 > z) - 1 + P(U_1 \leq v, U_2 \leq z) \\ &= P(U_1 > v, U_2 > z) \end{aligned} \quad (4)$$

It is also possible to express, via survival copula, the conditional probability

$$P(U_1 > v | U_2 > z) = \frac{1 - v - z + C(v, z)}{1 - z} = \frac{\bar{C}(1 - v, 1 - z)}{1 - z}.$$

### 3 Survival Truncated Copulas and Directional Dependence

Rodríguez-Lallena and Úbeda-Flores (2004) developed a wide class of bivariate copulas which depend on two univariate functions, and describe the dependency of the copulas in different ways. In this section, we generalize the class of bivariate copulas suggested by Rodríguez-Lallena and Úbeda-Flores (2004).

**Lemma 1** *Let  $(X, Y)$  be a continuous random pair whose associated copula  $C_\theta$  is given by  $C_\theta(u, v) = uv + \theta u^a v^b (1-u)^c (1-v)^d$  for every  $(u, v)$  in  $[0, 1]^2$  with  $a, b, c, d \geq 1$ . Rodríguez-Lallena and Úbeda-Flores (2004) prove that  $C_\theta$  is a copula if and only if  $-\frac{1}{\max\{\nu\gamma, \omega\delta\}} \leq \theta \leq -\frac{1}{\min\{\nu\delta, \omega\gamma\}}$ , where  $\omega = -\nu = 1$  if  $a = c = 1$ ,  $\delta = -\gamma = 1$  if  $b = d = 1$  and*

$$\begin{aligned}\nu &= -\left(\frac{a}{a+c}\right)^{a-1} \left(1 + \sqrt{\frac{c}{a(a+c-1)}}\right)^{a-1} \left(\frac{c}{a+c}\right)^{c-1} \left(1 - \sqrt{\frac{a}{c(a+c-1)}}\right)^{c-1} \sqrt{\frac{ac}{a+c-1}}, \\ \omega &= \left(\frac{a}{a+c}\right)^{a-1} \left(1 - \sqrt{\frac{c}{a(a+c-1)}}\right)^{a-1} \left(\frac{c}{a+c}\right)^{c-1} \left(1 + \sqrt{\frac{a}{c(a+c-1)}}\right)^{c-1} \sqrt{\frac{ac}{a+c-1}}, \\ \gamma &= -\left(\frac{b}{b+d}\right)^{b-1} \left(1 + \sqrt{\frac{d}{b(b+d-1)}}\right)^{b-1} \left(\frac{d}{b+d}\right)^{d-1} \left(1 - \sqrt{\frac{b}{d(b+d-1)}}\right)^{d-1} \sqrt{\frac{bd}{b+d-1}}, \\ \delta &= \left(\frac{b}{b+d}\right)^{b-1} \left(1 - \sqrt{\frac{d}{b(b+d-1)}}\right)^{b-1} \left(\frac{d}{b+d}\right)^{d-1} \left(1 + \sqrt{\frac{b}{d(b+d-1)}}\right)^{d-1} \sqrt{\frac{bd}{b+d-1}},\end{aligned}$$

otherwise. Moreover, the range for  $\theta$  contains the interval  $[-1, 1]$  for all  $a, b, c, d \geq 1$ . The case  $a = b = c = d = 1$  produces the Farlie-Gumbel-Morgenstern family of copulas (and the smallest range for  $\theta$ , i.e., the interval  $[-1, 1]$ ). In general, the bigger the parameters  $a, b, c, d$  are, the bigger the range for  $\theta$  is (for instance: if  $a = b = c = d = 2$ , then  $\theta \in [-27, 27]$ ; if  $a = 2, b = 3, c = 4, d = 5$ , then  $\theta \in [-840.445, 939.403]$ ; etc.).

Based on Rodríguez-Lallena and Úbeda-Flores (2004), Jung et al. (2007) introduced three different types of FGM distributions as follows:

- Type I:  $C_I(u, v) = uv + \theta uv(1-u)(1-v)$  where  $0 \leq u, v \leq 1$
- Type II:  $C_{II}(u, v) = uv + \theta u^\alpha v^\beta (1-u)(1-v)$  where  $\alpha \geq 1; \beta \geq 1; 0 \leq u, v \leq 1$
- Type III:  $C_{III}(u, v) = uv + \theta uv(1-u)^\alpha (1-v)^\beta$  where  $\alpha \geq 1; \beta \geq 1; 0 \leq u, v \leq 1$

Table 1 shows some special forms of  $f(u)$  and  $g(v)$  for each type of FGM function considered in this paper.

Table 1: Forms of  $f(u)$  and  $g(v)$  for each type

Type	$f(u)$	$g(v)$
I	$\sqrt{\theta}u(1-u)$	$\sqrt{\theta}v(1-v)$
II	$\sqrt{\theta}u^\alpha(1-u)$	$\sqrt{\theta}v^\beta(1-v)$
III	$\sqrt{\theta}u(1-u)^\alpha$	$\sqrt{\theta}v(1-v)^\beta$

The ranges of  $\theta$  at each Type in Table 1 are

$$\begin{aligned}
 \text{Type I} &: -1 \leq \theta \leq 1, \quad a = b = c = d = 1. \\
 \text{Type II} &: -\frac{1}{\max\{\nu\gamma, \omega\delta\}} \leq \theta \leq -\frac{1}{\min\{\nu\gamma, \omega\delta\}}, \quad a = \alpha, \quad b = \beta, \quad c = d = 1. \\
 \text{Type III} &: -\frac{1}{\max\{\nu\gamma, \omega\delta\}} \leq \theta \leq -\frac{1}{\min\{\nu\gamma, \omega\delta\}}, \quad a = b = 1 \quad c = \alpha, \quad d = \beta.
 \end{aligned} \tag{5}$$

By using Definition 1 and Lemma 1, the survival FGM type modification copulas are as follows:

- Survival Type I:  $\bar{C}_I(u, v) = uv + \theta uv(1-u)(1-v) = C_I(u, v)$  where  $0 \leq u, v \leq 1$
- Survival Type II:  $\bar{C}_{II}(u, v) = uv + \theta u^\alpha v^\beta (1-u)(1-v) = C_{II}(u, v)$   
where  $\alpha \geq 1; \beta \geq 1; 0 \leq u, v \leq 1$
- Survival Type III:  $\bar{C}_{III}(u, v) = uv + \theta uv(1-u)^\alpha (1-v)^\beta = C_{III}(u, v)$   
where  $\alpha \geq 1; \beta \geq 1; 0 \leq u, v \leq 1$

The truncation dependence on invariant copula is defined as follows: (Sungur, 1999)

**Definition 2** *If a 3-dimensional copula can be represented as*

$$\begin{aligned}
 C(u_1, u_2, u_3) &= C_{12} \left( \frac{C_{13}(u_1, u_3)}{u_3}, \frac{C_{23}(u_2, u_3)}{u_3} \right) u_3 \\
 &= C_{13} \left( \frac{C_{12}(u_1, u_2)}{u_2}, \frac{C_{23}(u_2, u_3)}{u_2} \right) u_2 \\
 &= C_{23} \left( \frac{C_{12}(u_1, u_2)}{u_1}, \frac{C_{13}(u_1, u_3)}{u_1} \right) u_1,
 \end{aligned}$$

*then it will be called a truncation invariant copula.*

We consider several different types of the Farlie-Gumbel-Morgenstern (FGM) distribution which have the specific form of the Rodríguez-Lallena and Úbeda-Flores (2004) copula family,  $C(u, v) =$

$uv + f(u)g(v)$ . Let  $C_{12}$ ,  $C_{13}$  and  $C_{23}$  be members of the Farlie-Gumbel-Morgenstern class of copulas:

$$\{C_{ij}; C_{ij}(u_i, u_j) = u_i u_j [1 + \theta_{ij}(1 - u_i)(1 - u_j)]\}.$$

Also, let  $\theta_{12}$ ,  $\theta_{13}$ , and  $\theta_{23}$  be the dependence parameters of  $C_{12}$ ,  $C_{13}$  and  $C_{23}$ , respectively. Provided that  $\theta_{12}$ ,  $\theta_{13}$ , and  $\theta_{23}$  lead to compatible 2-dimensional copulas, the partially truncated invariant 3-dimensional copula with respect to  $U_3$  is

$$\begin{aligned} C(u_1, u_2, u_3) &= u_1 u_2 u_3 [1 + \theta_{13}(1 - u_1)(1 - u_3)] [1 + \theta_{23}(1 - u_2)(1 - u_3)] \\ &\times \{1 + \theta_{12}(1 - u_1)(1 - u_2) [1 + \theta_{13}u_1(1 - u_3)] [1 + \theta_{23}u_2(1 - u_3)]\} \end{aligned}$$

If  $\theta_{12} = \theta_{23} = \theta_{13} = \theta$ , then

$$\begin{aligned} C(u_1, u_2, u_3) &= u_1 u_2 u_3 [1 + \theta(1 - u_1)(1 - u_3)] [1 + \theta(1 - u_2)(1 - u_3)] \\ &\times \{1 + \theta(1 - u_1)(1 - u_2) [1 + \theta u_1(1 - u_3)] [1 + \theta u_2(1 - u_3)]\} \end{aligned} \quad (6)$$

which will be referred to as the equi-dependence structure. It can be easily shown that such generated copulas are partially truncated invariants with respect to all possible truncation regions. The equi-dependence structure employed by Cook and Johnson (1981) to describe data which is not elliptically symmetric has been extensively discussed in Sungur (1999).

Under the assumption that each trivariate copula has the same parameter, we let  $\theta_{12}$ ,  $\theta_{13}$ ,  $\theta_{23}$  and  $\theta_{123}$  be the dependence parameters of  $C_{12}$ ,  $C_{13}$ ,  $C_{23}$  and  $C_{123}$  respectively then  $\theta_{12} = \theta_{23} = \theta_{13} = \theta_{123} = \theta$ . We can show the survival FGM type modification copulas as follows:

- Survival FGM Type Modification Copula Type I:

$$\begin{aligned} &\bar{C}_I(u_1, u_2, u_3) \\ &= \bar{C}_{I,12} \left( \frac{\bar{C}_{I,13}(u_1, u_3)}{1 - u_3}, \frac{\bar{C}_{I,23}(u_2, u_3)}{1 - u_3} \right) \cdot (1 - u_3) \\ &= C_{I,12} \left( \frac{C_{I,13}(u_1, u_3)}{1 - u_3}, \frac{C_{I,23}(u_2, u_3)}{1 - u_3} \right) \cdot (1 - u_3) \\ &= C_{I,12} \left( \frac{u_1 u_3 + \theta u_1 u_3 (1 - u_1)(1 - u_3)}{1 - u_3}, \frac{u_2 u_3 + \theta u_2 u_3 (1 - u_2)(1 - u_3)}{1 - u_3} \right) \cdot (1 - u_3) \\ &= u_1 u_2 u_3^2 \left\{ \frac{[1 + \theta(1 - u_1)(1 - u_3)] [1 + \theta(1 - u_2)(1 - u_3)]}{1 - u_3} \right\} \times \\ &\quad \left\{ 1 + \theta \left( 1 - \frac{u_1 u_3 [1 + \theta(1 - u_1)(1 - u_3)]}{1 - u_3} \right) \left( 1 - \frac{u_2 u_3 [1 + \theta(1 - u_2)(1 - u_3)]}{1 - u_3} \right) \right\} \end{aligned}$$

where  $0 \leq u_1, u_2, u_3 \leq 1$  and  $-1 \leq \theta \leq 1$ .

- Survival FGM Type Modification Copula Type II:

$$\begin{aligned}
& \bar{C}_{II}(u_1, u_2, u_3) \\
= & \bar{C}_{II,12} \left( \frac{\bar{C}_{II,13}(u_1, u_3)}{u_3}, \frac{\bar{C}_{II,23}(u_2, u_3)}{u_3} \right) \cdot (1 - u_3) \\
= & C_{III,12} \left( \frac{C_{III,13}(u_1, u_3)}{1 - u_3}, \frac{C_{III,23}(u_2, u_3)}{1 - u_3} \right) \cdot (1 - u_3) \\
= & C_{III,12} \left( \frac{u_1 u_3 + \theta u_1^\alpha u_3^\beta (1 - u_1)(1 - u_3)}{1 - u_3}, \frac{u_2 u_3 + \theta u_2^\alpha u_3^\beta (1 - u_2)(1 - u_3)}{1 - u_3} \right) \cdot (1 - u_3) \\
= & u_1 u_2 u_3^2 \left\{ \frac{\left[ \frac{1 + \theta u_1^{\alpha-1} u_3^{\beta-1} (1 - u_1)(1 - u_3)}{1 - u_3} \right] \left[ \frac{1 + \theta u_2^{\alpha-1} u_3^{\beta-1} (1 - u_2)(1 - u_3)}{1 - u_3} \right]}{1 - u_3} \right\} \\
& + \theta u_1^{\alpha-1} u_2^{\beta-1} u_3^{\alpha+\beta-2} \left\{ \frac{1 + \theta u_1^{\alpha-1} u_3^{\beta-1} (1 - u_1)(1 - u_3)}{1 - u_3} \right\}^\alpha \left\{ \frac{1 + \theta u_2^{\alpha-1} u_3^{\beta-1} (1 - u_2)(1 - u_3)}{1 - u_3} \right\}^\beta \\
& \times \left\{ \frac{\left[ 1 - u_3 - u_1 u_3 - \theta u_1^\alpha u_3^\beta (1 - u_1)(1 - u_3) \right] \left[ 1 - u_3 - u_2 u_3 - \theta u_2^\alpha u_3^\beta (1 - u_2)(1 - u_3) \right]}{1 - u_3} \right\}
\end{aligned}$$

where  $\alpha \geq 1$ ;  $\beta \geq 1$ ;  $0 \leq u_1, u_2, u_3 \leq 1$  and the admissible range of  $\theta$  can be derived from Type II in (5).

- Survival FGM Type Modification Copula Type III:

$$\begin{aligned}
& \bar{C}_{III}(u_1, u_2, u_3) \\
= & \bar{C}_{III,12} \left( \frac{\bar{C}_{III,13}(u_1, u_3)}{u_3}, \frac{\bar{C}_{III,23}(u_2, u_3)}{u_3} \right) \cdot (1 - u_3) \\
= & C_{II,12} \left( \frac{C_{II,13}(u_1, u_3)}{1 - u_3}, \frac{C_{II,23}(u_2, u_3)}{1 - u_3} \right) \cdot (1 - u_3) \\
= & C_{II,12} \left( \frac{u_1 u_3 + \theta u_1 u_3 (1 - u_1)^\alpha (1 - u_3)^\beta}{1 - u_3}, \frac{u_2 u_3 + \theta u_2 u_3 (1 - u_2)^\alpha (1 - u_3)^\beta}{1 - u_3} \right) \cdot (1 - u_3) \\
= & u_1 u_2 u_3^2 \left\{ \frac{\left[ 1 + \theta (1 - u_1)^\alpha (1 - u_3)^\beta \right] \left[ 1 + \theta (1 - u_2)^\alpha (1 - u_3)^\beta \right]}{1 - u_3} \right\} \times \\
& \left\{ 1 + \theta \left( 1 - \frac{u_1 u_3 \left[ 1 + \theta (1 - u_1)^\alpha (1 - u_3)^\beta \right]}{1 - u_3} \right)^\alpha \left( 1 - \frac{u_2 u_3 \left[ 1 + \theta (1 - u_2)^\alpha (1 - u_3)^\beta \right]}{1 - u_3} \right)^\beta \right\}
\end{aligned}$$

where  $\alpha \geq 1$ ;  $\beta \geq 1$ ;  $0 \leq u_1, u_2, u_3 \leq 1$  and the admissible range of  $\theta$  can be derived from Type III in (5).

For more details of the admissible range of  $\theta$  about the generalized Farlie-Gumbel-Morgenstern distributions the reader is referred to the articles of Bairamov and Eryilmaz (2004), Bairamov, Kotz, and Bekci (2000), Bairamov and Kotz (2002), Bairamov and Kotz (2003), Kim et al. (2008) and Lai and Xie (2000).

The conditional joint distribution of  $U_1$  and  $U_2$  under the condition  $\{U_3 \geq a\}$  is

$$\begin{aligned}\bar{C}(U_1, U_2|U_3 \geq a) &= P(U_1 \geq u_1, U_2 \geq u_2|U_3 \geq a) \\ &= \frac{P(U_1 \geq u_1, U_2 \geq u_2, U_3 \geq a)}{P(U_3 \geq a)} \\ &= \frac{\bar{C}(u_1, u_2, a)}{1-a}.\end{aligned}$$

By the equi-dependence structure, the  $\bar{C}(U_1, U_2|U_3 \geq a)$  at each survival truncated FGM type modification copula type is as follows:

- Survival Truncated FGM Type Modification Copula Type I:

$$\begin{aligned}\bar{C}_I(u_1, u_2|a) &= \frac{1}{1-a} \left[ u_1 u_2 a^2 \left\{ \frac{[1 + \theta(1-u_1)(1-a)][1 + \theta(1-u_2)(1-a)]}{1-a} \right\} \times \right. \\ &\quad \left. \left\{ 1 + \theta \left( 1 - \frac{u_1 a [1 + \theta(1-u_1)(1-a)]}{1-a} \right) \left( 1 - \frac{u_2 a [1 + \theta(1-u_2)(1-a)]}{1-a} \right) \right\} \right] \\ &\quad \text{where } 0 \leq u_1, u_2, u_3 \leq 1.\end{aligned}$$

- Survival Truncated FGM Type Modification Copula Type II:

$$\begin{aligned}\bar{C}_{II}(u_1, u_2|a) &= \frac{1}{1-a} \left[ u_1 u_2 a^2 \left[ \left\{ \frac{[1 + \theta u_1^{\alpha-1} a^{\beta-1} (1-u_1)(1-a)][1 + \theta u_2^{\alpha-1} a^{\beta-1} (1-u_2)(1-a)]}{1-a} \right\} \right. \right. \\ &\quad \left. \left. + \theta u_1^{\alpha-1} u_2^{\beta-1} a^{\alpha+\beta-2} \left\{ \frac{1 + \theta u_1^{\alpha-1} a^{\beta-1} (1-u_1)(1-a)}{1-a} \right\}^\alpha \left\{ \frac{1 + \theta u_2^{\alpha-1} a^{\beta-1} (1-u_2)(1-a)}{1-a} \right\}^\beta \right. \right. \\ &\quad \left. \left. \times \left\{ \frac{[1-a-u_1 a - \theta u_1^\alpha a^\beta (1-u_1)(1-a)][1-a-u_2 a - \theta u_2^\alpha a^\beta (1-u_2)(1-a)]}{1-a} \right\} \right] \right] \\ &\quad \text{where } \alpha \geq 1; \beta \geq 1; 0 \leq u_1, u_2, u_3 \leq 1.\end{aligned}$$



- Survival Truncated FGM Type Modification Copula Type III:

$$\begin{aligned} & \bar{C}_{III}(u_1, u_2|a) \\ = & \frac{1}{1-a} \left[ u_1 u_2 a^2 \left\{ \frac{[1 + \theta(1-u_1)^\alpha(1-a)^\beta][1 + \theta(1-u_2)^\alpha(1-a)^\beta]}{1-a} \right\} \times \right. \\ & \left. \left\{ 1 + \theta \left( 1 - \frac{u_1 a [1 + \theta(1-u_1)^\alpha(1-a)^\beta]}{1-a} \right)^\alpha \left( 1 - \frac{u_2 a [1 + \theta(1-u_2)^\alpha(1-a)^\beta]}{1-a} \right)^\beta \right\} \right] \\ & \text{where } \alpha \geq 1; \beta \geq 1; 0 \leq u_1, u_2, u_3 \leq 1. \end{aligned}$$

The directional dependencies proposed by Sungur (2005) for the direction  $U_1$  to  $U_2$  and for the direction  $U_2$  to  $U_1$  under truncation  $U_3 \geq a$  are defined as

$$\begin{aligned} r_{U_1|U_2,a}(u_2) &= E[U_1|U_2 = u_2, U_3 \geq a] \\ &= 1 - \int_0^1 \frac{\partial \bar{C}(u_1, u_2, a)}{\partial u_2} du_1 = 1 - \frac{\partial}{\partial u_2} \int_0^1 \bar{C}(u_1, u_2, a) du_1, \end{aligned} \quad (7)$$

and

$$\begin{aligned} r_{U_2|U_1,a}(u_1) &= E[U_2|U_1 = u_1, U_3 \geq a] \\ &= 1 - \int_0^1 \frac{\partial \bar{C}(u_1, u_2, a)}{\partial u_1} du_2 = 1 - \frac{\partial}{\partial u_1} \int_0^1 \bar{C}(u_1, u_2, a) du_2. \end{aligned} \quad (8)$$

Table 3 shows the measures of dependence of two variables  $u_1$  and  $u_2$  under the truncated variable  $u_3 \geq a$  and the figure 1 shows the plots of directional dependence of for three different survival truncated FGM type modification copulas (Type I, Type II and Type III) under the truncated value  $a = 0.3$  using (7) and (8).

## 4 Using Survival Truncated FGM Type Modification Copulas to Analyze Directional Dependence in Gene Expression Datasets

The ability to recognize and quantify directionality in gene dependence relationships will enhance the ability of researchers working in the fields of genomics and proteomics to extract information from large sets of gene expression data, and thereby to better understand the complex mechanisms by which genes and proteins interact. The microarray dataset used in this analysis is from a previous

study on yeast cell-cycle regulation (Spellman et al., 1998) and is publicly available (<http://cellcycle-www.stanford.edu/>). The dataset is composed of measurements on 6221 genes observed at 80 time points. 800 genes were identified. We selected one group of genes with known interaction patterns (note that known interactions are still incomplete at present). The group includes eight histone genes-HHT1, HHT2, HHF1, HHF2, HTA1, HTA2, HTB1 and HTB2. These eight genes encode four histones: H2A, H2B, H3 and H4. Histones are proteins which bind tightly to DNA, helping to 'package' the genetic material in chromosomes. Because chromosomes are replicated during cell division, expression of histone genes must be tightly regulated in order for replication to proceed. Table 2 provides basic descriptive statistics of the expression data for the eight histone genes used in this study. From Table 2, we can see that the skewness and the kurtosis of each histone gene are almost zero. We also notice that the maximum of *HHT2* is much lower than the other maxima. The maxima of *HHF1* and *HHF2* are close to each other. But the minima of *HHF1* and *HHF2* are significantly different.

Table 2: Basic Statistics of Empirical Transformed Data

	Mean	Median	Minimum	Maximum	St.D	Skewness	Kurtosis
HHT1	-0.01450	0.03	-2.25	1.82	0.88255	-5.903e-47	9.862e-62
HHT2	-0.15556	-0.17	-2.32	1.69	0.82572	6.009e-47	1.009e-61
HHF2	-0.24212	-0.12	-2.84	1.99	0.99243	6.078e-48	4.759e-63
HHF1	-0.24303	-0.15	-3.18	2.00	1.02157	-7.982e-47	1.474e-61
HTB1	-0.18278	-0.10	-2.00	1.83	0.89974	-8.163e-49	3.273e-64
HTA2	-0.30008	-0.22	-3.06	1.76	0.93989	1.118e-49	2.311e-65
HTA1	-0.32810	-0.34	-2.84	1.86	0.98060	1.702e-47	1.879e-62
HTB2	-0.20398	-0.07	-2.47	1.83	0.89841	-3.456e-48	2.242e-63

Next, we estimate a parameter  $\theta$  in each survival truncated FGM type modification distribution. We define  $U_i := F_X(X_i)$  and  $V_i := F_Y(Y_i)$  for the continuous empirical marginal distribution function  $F_X$  and  $F_Y$ . We assume that  $U_i$  and  $V_i$  have uniform distribution  $U(0, 1)$ . Hence we can reduce our empirical likelihood function to

$$L(\theta; U, V) = \prod_{i=1}^n c(U_i, V_i) \quad (9)$$

The estimation of the parameter  $\theta$  is determined by maximizing the likelihood (9) for the real data

set. For a computational convenience of a MLE of  $\theta$ , the logarithmic form of (9) is as follows:

$$\hat{\theta} = \underset{\theta \in R}{\operatorname{argmax}} \sum_{i=1}^n \log L(\theta; U_i, V_i) \quad (10)$$

where  $R$  is the set of all possible  $\theta$ 's.

We employed Akaike's Information Criterion (AIC) to evaluate our copula model. Akaike's Information Criterion is defined as follows:

$$AIC = -2 \log(L(\theta; U, V)) + 2t$$

where  $t$  is the number of parameters of the model. The value of AIC is an indicator of which estimator fits better: the lower the AIC, the better the model. Table 4 shows the values of the AIC for the survival truncated FGM type modification copula models (Type I, Type II, and Type III). In this paper, to estimate the parameter  $\theta$ , we consider equation (9) because it is a simpler form of the likelihood function than the logarithmic form. But we found that there is a difficulty in estimating  $\theta$  by using the method of the maximum likelihood estimation. The problem is  $\frac{\partial L(\theta; U, V)}{\partial \theta}$  (or  $\frac{\partial \log L(\theta; U, V)}{\partial \theta}$ ), which is not a function of  $\theta$ . Therefore, we cannot estimate parameter  $\theta$  using the likelihood function. Instead, we used a numerical method to find a value of  $\theta$  to maximize the copula function. The procedures for finding an estimate of  $\theta$  are as follows:

**Step 0** Select a model which is Type I, II, or III.

**Step 1** Find the range of  $\theta$  from (5) under  $a = 0.3$ ,  $\alpha = 2$ , and  $\beta = 2$  for Type II, and under  $a = 0.3$ ,  $\alpha = 2$ , and  $\beta = 3$  for Type III.

**Step 2** Define a grid unit and move grid from the minimum to the maximum in the range of  $\theta$  for Type I, II, or III. (grid unit in data analysis, Type I: 0.001, Type II & III: 0.005.)

**Step 3** Calculate values of  $\theta$  from (10) as the grid increases.

**Step 4** Select the value of  $\theta$  maximizing (10).

Table 3 shows the measures of dependence of two variables  $u_1$  and  $u_2$  under the truncated variable  $u_3 \geq a$  for three different survival truncated FGM type modification copulas. In Figure 1, we see that the mean plots of  $r_{U|V, a=0.3}$  and  $r_{V|U, a=0.3}$  for Type I are almost identical, whereas the

mean plots of  $r_{U|V,a=0.3}$  and  $r_{V|U,a=0.3}$  for Type II ( $\alpha = 2$  and  $\beta = 2$ ) are nonidentical even if we set Type II ( $\alpha = 2$  and  $\beta = 2$ ) to be symmetric. This indicates that Type II has more direction dependence response, depending on the truncated value  $a$  and a parameter  $\theta$ . Also the mean plots of  $r_{U|V,a=0.3}$  and  $r_{V|U,a=0.3}$  for Type III ( $\alpha = 2$  and  $\beta = 3$ ) show more different than those of Type I because we set Type II ( $\alpha = 2$  and  $\beta = 3$ ) to be nonsymmetric.

The values of  $r_{U|V,a}$  and  $r_{V|U,a}$  for Type I, Type II ( $\alpha = 2$  and  $\beta = 2$ ), and Type III ( $\alpha = 2$  and  $\beta = 3$ ) are shown in Table 3. The values of  $r_{U|V,a}$  and  $r_{V|U,a}$  for Type I are almost identical. Therefore, we can conclude that Type I has no directional dependence property and Type II and III have directional dependence, unlike Type I. In addition, figure 1 provides evidence that Type II and III are able to recognize directional dependence between any two histone genes. In terms of the goodness of fit using gene data, the AIC values of Type I, Type II ( $\alpha = 2$  and  $\beta = 2$ ), and Type III ( $\alpha = 2$  and  $\beta = 3$ ) in Table 4 and Figure 2 indicate that the Type I model is better than Type II and Type III models because the values of AIC using the survival truncated FGM type modification copula Type I are much smaller than those using the the survival truncated FGM type modification copula Type II and Type II. In addition, the Type II ( $\alpha = 2$  and  $\beta = 2$ ) model is better than the Type III ( $\alpha = 2$  and  $\beta = 3$ ) model because the AIC values of Type II ( $\alpha = 2$  and  $\beta = 2$ ) are slightly smaller than those of Type III ( $\alpha = 2$  and  $\beta = 3$ ).

## 5 Conclusions

Dependence properties and measures of association between two or more variables can be investigated in terms of various copulas. In this paper, we have presented a flexible new class of survival truncated FGM type modification copulas. Using the described methods, we showed that the survival truncated FGM type modification copulas (Type II and Type III) have directional dependence. Therefore, the survival truncated FGM type modification copula models (Type II and Type III) can help to determine directional dependence among three different genes. Our methods were evaluated by employing AIC criterion. More work is needed, however, to refine our ability to analyze directional dependence within a three-variable framework, and to test these new tools with datasets where clearer evidence of unidirectional gene dependence is evident. In addition, we need to develop other goodness of fit criterion to select the best multivariate copula model for a given gene dataset.

## Acknowledgements

The authors are thankful to the Editor, Associate Editor and the two referees for valuable comments on the original version of this manuscript which lead to substantial improvement.

## References

- Bairamov I. and Eryilmaz S. (2004), Characterization of symmetry and exceedance models in multivariate FGM distributions. *Journal of Applied Statistical Science*, 13 (2), 87-99.
- Bairamov, I. , Kotz, S and Bekci, M (2000), New generalized F-G-M distributions and commitants of order statistics, *Journal of Applied Statistics*, 28, 521-536.
- Bairamov, I. and Kotz, S (2002), Dependence structure and symmetry of Huang-Kotz distributions and their extensions, *Metrika*, 56(1), 55-72.
- Bairamov, I and Kotz, S (2003), On a new family of positive quadrant dependent bivariate distributions, *International Mathematical Journal*, 3(11), 1247-1254.
- Cherubini, U., Luciano, E.,and Vecchiato, W. (2004), *Copula methods in finance*. Wiley Finance Series. John Wiley & Sons, Ltd., Chichester.
- Cook, R.D. and Johnson, M.E. (1981), A Family of Distributions for Modelling Non-Elliptically Symmetric Multivariate Data. *J. Roy. Statist. Soc. Ser. B* 43, 210-219.
- Jung, Y.-S., Kim, J.-M., and Sungur, E. A. (2007), "Directional Dependence of Truncation Invariant FGM Copula Functions: Application to Foreign Exchange Currency Data", unpublished manuscript.
- Kim, J.-M., Jung, Y., Sungur, E. A., Han, K., Park, C., and Sohn, I. (2008), A Copula Method for Modeling Directional Dependence of Genes, *BMC Bioinformatics* 9:225.
- Lai C.D. and Xie M. (2000), A new family of positive quadrant dependent bivariate distributions. *Statist. Probab. Letters*. 46, 359-364.
- Nelsen, R. B. (1999), *An introduction to copulas*. Springer-Verlag, New York.

- Rodríguez-Lallena, J. A., Úbeda-Flores, M. (2004), A new class of bivariate couplas. *Statist. Probab. Lett.* 66, 315-325.
- Sklar, A. (1959), Fonctions de repartition a n dimensions et leurs marges. (French) *Publ. Inst. Statist. Univ. Paris 8*, 229-231.
- Sklar, A. (1973), Random variables, joint distribution functions, and copulas. *Kybernetika (Prague)* 9, 449-460.
- Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., and Fucher, B. (1998), Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast *Saccharomyces cerevisiae* by Microarray Hybridization. *Molecular Biology of the Cell* 9, 3273-3297.
- Sungur, E. A. (1999), Truncation invariant dependence structures. *Comm. Statist. Theory Methods*, 28 (11), 2553-2568.
- Sungur, E. A. (2005), A Note on Directional Dependence in Regression Setting. *Comm. Statist. Theory Methods*, 34, 1957–1965.

Table 3: Measurement of Directional Dependence at Type I, II, III with Truncation 0.3

U vs V	Type I		Type II		Type III	
	$r_{U V,0.3}$	$r_{V U,0.3}$	$r_{U V,0.3}$	$r_{V U,0.3}$	$r_{U V,0.3}$	$r_{V U,0.3}$
1. HHT1 vs HHT2	0.91118675493	0.91116357356	0.91830094799	0.91828728055	1.23361636406	0.91524300810
2. HHT1 vs HHF2	0.91095599554	0.91093218469	0.91826322366	0.91828728055	1.24335198839	0.91486604955
3. HHT1 vs HHF1	0.91117488905	0.91116357356	0.91827828813	0.91828728055	1.23354869747	0.91524300810
4. HHT1 vs HTB1	0.91117518882	0.91116357356	0.91830960404	0.91828728055	1.24304825264	0.91486604955
5. HHT1 vs HTA2	0.91119595238	0.91116357356	0.91828573507	0.91828728055	1.22572998267	0.91555693422
6. HHT1 vs HTA1	0.91094335034	0.91093218469	0.91829158429	0.91828728055	1.24156671420	0.91492889451
7. HHT1 vs HTB2	0.91117923737	0.91116357356	0.91826222762	0.91828728055	1.21945072483	0.91580793826
8. HHT2 vs HHF2	0.91118709267	0.91118675493	0.91826322366	0.91830094799	1.23364022910	0.91526875320
9. HHT2 vs HHF1	0.91140593970	0.91141769097	0.91827828813	0.91830094799	1.21156652305	0.91614594240
10. HHT2 vs HTB1	0.91117518882	0.91118675493	0.91830960404	0.91830094799	1.22856000141	0.91545684638
11. HHT2 vs HTA2	0.91119595238	0.91118675493	0.91828573507	0.91830094799	1.22731286822	0.91551952922
12. HHT2 vs HTA1	0.91117461806	0.91118675493	0.91829158429	0.91830094799	1.22869991210	0.91545684638
13. HHT2 vs HTB2	0.91141021778	0.91141769097	0.91826222762	0.91830094799	1.20407581409	0.91645886138
14. HHF2 vs HHF1	0.91094363770	0.91095599554	0.91827828813	0.91826322366	1.24325880090	0.91492068138
15. HHF2 vs HTB1	0.91094390643	0.91095599554	0.91830960404	0.91826322366	1.24304825264	0.91492068138
16. HHF2 vs HTA2	0.91119595238	0.91118709267	0.91828573507	0.91826322366	1.22572998267	0.91560940556
17. HHF2 vs HTA1	0.91094335034	0.91095599554	0.91829158429	0.91826322366	1.24319687074	0.91492068138
18. HHF2 vs HTB2	0.91117923737	0.91118709267	0.91826222762	0.91826322366	1.22101447600	0.91579708515
19. HHF1 vs HTB1	0.91117518882	0.91117488905	0.91830960404	0.91827828813	1.23174579691	0.91532898860
20. HHF1 vs HTA2	0.91119595238	0.91117488905	0.91828573507	0.91827828813	1.21944649096	0.91583045984
21. HHF1 vs HTA1	0.91117461806	0.91117488905	0.91829158429	0.91827828813	1.23509461493	0.91520354578
22. HHF1 vs HTB2	0.91117923737	0.91117488905	0.91826222762	0.91827828813	1.22731748314	0.91551709685
23. HTB1 vs HTA2	0.91119595238	0.91117518882	0.91828573507	0.91830960404	1.22731286822	0.91548538926
24. HTB1 vs HTA1	0.91094335034	0.91094390643	0.91829158429	0.91830960404	1.24647177789	0.91473082586
25. HTB1 vs HTB2	0.91117923737	0.91117518882	0.91826222762	0.91830960404	1.21478819056	0.91598780701
26. HTA2 vs HTA1	0.91117461806	0.91119595238	0.91829158429	0.91828573507	1.23029134987	0.91541952691
27. HTA2 vs HTB2	0.91117923737	0.91119595238	0.91826222762	0.91828573507	1.21324356844	0.91610766862
28. HTA1 vs HTB2	0.91117923737	0.91117461806	0.91826222762	0.91829158429	1.21633758724	0.91594366975

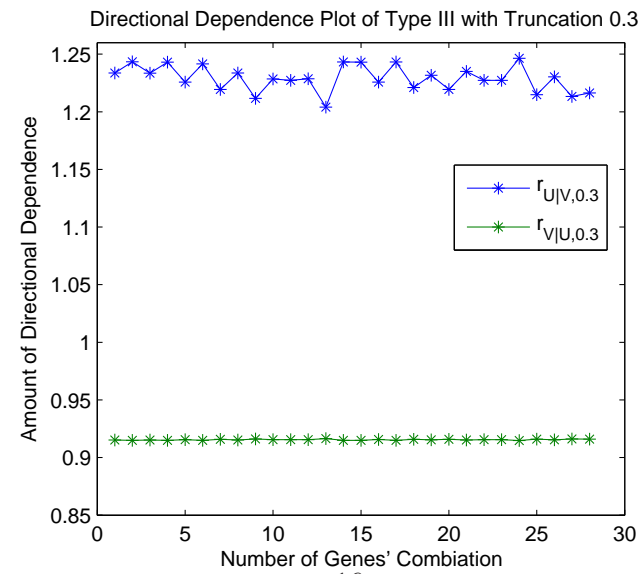
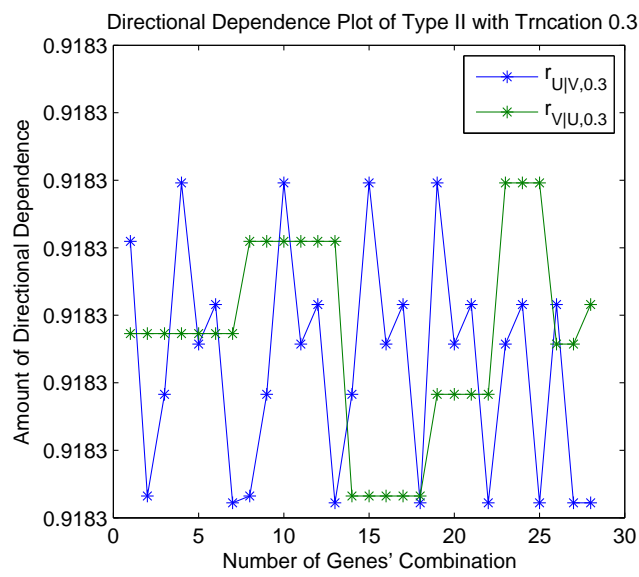
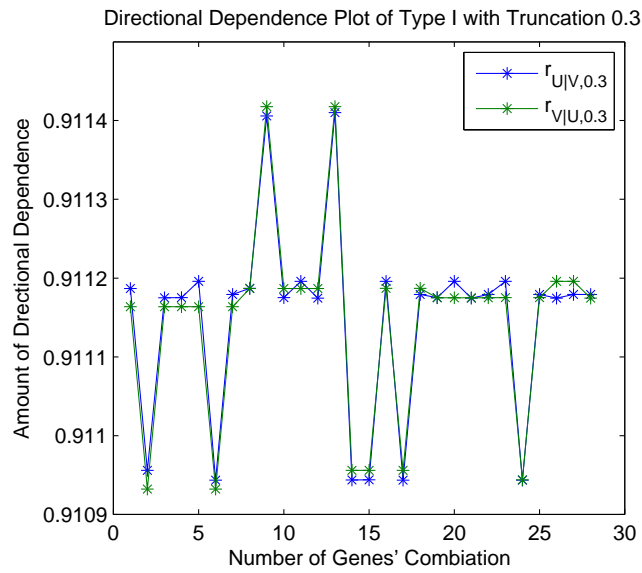


Figure 1: Plots of Directional dependence by Type under Truncation 0.3



Table 4: Akaike's Information Criterion (AIC) by Type

	Truncation: $a=0.3$		
	Type I	Type II	Type III
1. HHT1 vs HHT2	-70.8521157312	306.7020817934	307.9021568838
2. HHT1 vs HHF2	-70.9700265966	306.4619532701	307.2731127496
3. HHT1 vs HHF1	-70.8164804306	306.6831562838	308.0322387680
4. HHT1 vs HTB1	-70.9415181960	306.6310312178	307.2802644766
5. HHT1 vs HTA2	-70.8291759075	306.9770682957	308.3025855647
6. HHT1 vs HTA1	-70.9632626266	306.6072269198	307.2985869113
7. HHT1 vs HTB2	-70.6739916220	306.8164099732	308.9009928565
8. HHT2 vs HHF2	-70.8544692475	306.6585005289	307.9102452034
9. HHT2 vs HHF1	-70.4637443736	307.0405391701	309.6790153094
10. HHT2 vs HTB1	-70.7707748251	306.9070982775	308.1946917129
11. HHT2 vs HTA2	-70.6622613158	306.7766716690	308.4105863390
12. HHT2 vs HTA1	-70.7521009098	306.8107610485	308.2462265023
13. HHT2 vs HTB2	-70.3385426103	306.9931918476	310.1698264737
14. HHF2 vs HHF1	-70.9302011373	306.3746847371	307.3203687860
15. HHF2 vs HTB1	-70.9998160819	306.5793843911	307.1255020018
16. HHF2 vs HTA2	-70.7133726549	306.7510930792	308.4622706962
17. HHF2 vs HTA1	-70.9374070629	306.4528045756	307.2797123517
18. HHF2 vs HTB2	-70.6311723271	306.6846689390	308.9294755331
19. HHF1 vs HTB1	-70.7835084043	306.8059150562	308.1215389426
20. HHF1 vs HTA2	-70.6480622539	306.9413047371	308.9076500977
21. HHF1 vs HTA1	-70.7953054175	306.6641396677	307.9774620809
22. HHF1 vs HTB2	-70.6726502437	306.5861748581	308.5997668282
23. HTB1 vs HTA2	-70.7230954133	306.9193072923	308.3459880729
24. HTB1 vs HTA1	-70.9668295265	306.6117003327	307.0611720721
25. HTB1 vs HTB2	-70.5924158111	306.9426720800	309.1881762701
26. HTA2 vs HTA1	-70.8385432282	306.8058250092	307.9749384984
27. HTA2 vs HTB2	-70.5788268299	306.8644176372	309.2637324120
28. HTA1 vs HTB2	-70.5859402194	306.9014941480	309.1650709023

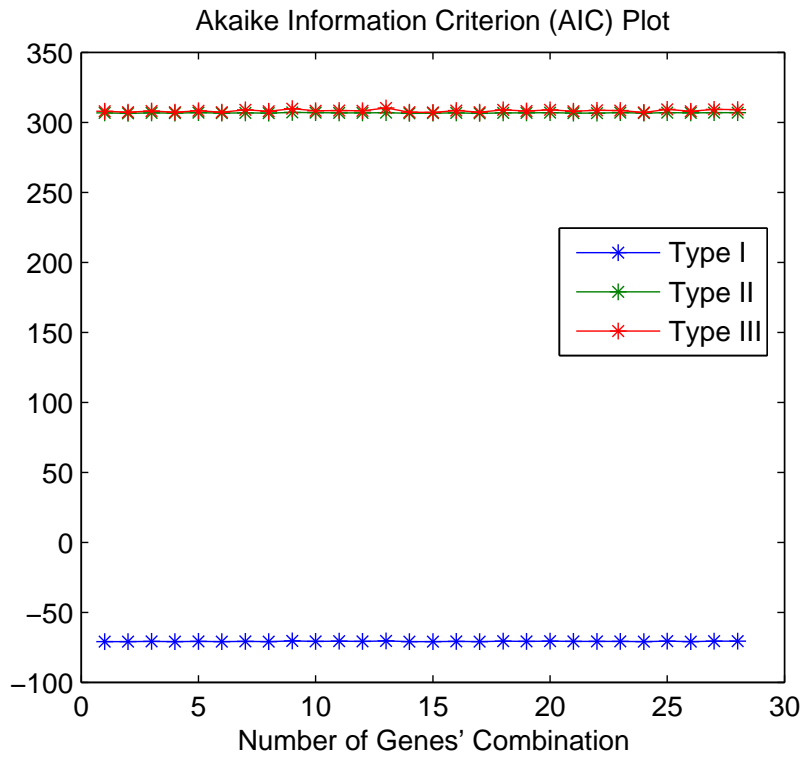


Figure 2: *Plots of Akaike's Information Criteria (AIC) by Type under Truncation 0.3*