1 BIVARIATE ARCHIMEDEAN COPULAS FOR CONTINGENCY TABLES WITH ORDERED CATEGORIES *

Vartan Choulakian (1) and Jules de Tibeiro (2)

ABSTRACT. Usually, either Goodman's uniform association model or Plackett's model are applied for the analysis of association in contingency tables with ordered categories. In some cases, these two well known methods do not fit well the data set. We propose, in addition to these two methods, the use of three one-parameter bivariate Archimedean copulas. The parameter of association in these copulas is estimated by maximum likelihood, which at the same time helps us in the choice of a good model. Two examples are provided.

Key words : Bivariate distribution ; multinomial distribution ; positive quadrant dependence ; measure of association ; maximum likelihood ; Goodman's U.

1. INTRODUCTION

We assume in this paper that the ordered categories in a contingency table are not independent, and the ordinality of variables dictates a kind of association between them, in most cases the nature of association being positive. That is, one would expect as one variable becomes ordinally larger (or smaller), then the other variable is more likely to be ordinally larger (or smaller).

2. BIVARIATE MODELS

First, we introduce some notation. Under the hypothesis that there is some kind of association between X and Y, we designate the bivariate density function by $h_{i\ j}$, and the cumulative distribution function by $H_{i\ j}$. Five bivariate models will be presented in this section, and all of them define PQD between the two

If $\varphi > 0$, then X and Y are PQD; if $\varphi < 0$, then X and Y are NQD and if $\varphi = 0$, then X and Y are independant, see Goodman (1979). Goodman's uniform association model is also a log-linear model which has one more parameters than the log-linear model under the independence hypothesis.

3. ARCHIMEDEAN COPULAS

3.1 Gumbel copula

$$H_{i\ j} = exp[-\{(-logF_i)^{\varphi} + (-logG_j)^{\varphi}\}^{1/\varphi}] \quad pour \ \ \varphi \geq 1 \tag{1}$$

$s(v) = (\ -\ log\ v)^{\varphi}$see for instance, Gumbel (1980) and Hougaard (1986). It is aslo named Gumbel-Hougaard copula by Hutchinson and Lai (1990). For $\varphi > 1$ in (1), X and Y are PQD and if $\varphi = 1$, X and Y are independent. Note that the Gumbel copula can not be a NQD.

3.2 Frank Copula

$$H_{i\ j} = [log\{1 + (\varphi^{F_i} - 1)(\varphi^{G_j} - 1)/(\varphi - 1)\}]/\ log\ \varphi \quad pour\ \varphi > 0\ \ and\ \varphi \neq 1 \tag{2}$$
$s(v) = log[(1 - \varphi^{v}\ )/(1 - \varphi)]$
and by taking the limit

$$H_{i\ j} = F_i G_j\ and\ s(v) = log\ v \ \ if \quad \varphi = 1$$

see for instance, Frank (1979) and Genest (1987). If $0 < \varphi < 1$ in (2), then X and Y are PQD; if $\varphi = 1$, X and Y are independent, and if $\varphi > 1$, X and Y are NQD. This is the only Archimedean copula that satisfies

$H_{i\ j} = Pr(X\ > i, Y > j)$, obtained from (2) by replacing F_i and G_j by $1 - F_i$ and $1 - G_j$ respectively; which means that (2) is still valid if one reverses the order categories of the variables X and Y.

3.3 Clayton Copula

$$H_{i\ j} = (F_i^{1-\varphi} + G_j^{1-\varphi} - 1)^{(1-\varphi)^{-1}} \quad pour\ \ \varphi > 0\ \ et\ \ \ \varphi \neq 1 \tag{3}$$

$s(v) = (v^{1-\varphi} - 1)/(\varphi - 1)$

and by taking the limit

$H_{i\ j} = F_i G_j\ and\ \ s(v) = log\ v \quad if \quad \varphi = 1$

4. REMARKS

We note that for Gumbel and Clayton copulas one can model either the bivariate distribution function H or the bivariate survival function $\bar{H}$.The three other models produce identical results if we reverse the order of the categories of X and Y variables.

5. DATA ANALYSIS

Before proposing any model for the data set, we have applied a Correspondence Analysis (C.A) with data set to visualize and describe the relationships between the variables.

6. CONCLUSION

We presented a collection of models, five in number, and this number can be increased, for the analysis of association in contingency tables with ordered categories. The estimation of the association parameter was done by maximum likelihood, which provided, at the same time, an efficient model selection technique to choose a good one. Other models can be added, for example, to model the parameter of association φ by a bilinear function,

$\varphi_{i\ j} = \sum_{k=1}^{m} \mu_i \nu_j$.

References

CLAYTON D.G. (1978), A model for association in bivariate life tables and its application in epidemiological studies of familial in chronic disease incidence, *Biometrika*, 65, 141-151.

FRANK M.J. (1979), On the simultaneous associativity of *F(x, y)* and *x+y-F(x, y)*, *Aequationes Mathematicae*, 19, 194-226.

GOODMAN L.A. (1979), Simple models for the analysis of cross-classifications having ordered categories, *Journal of the American Statistical Association*, 74, 537-552.

GOODMAN L.A. (1981), Association models and the bivariate normal for contingency tables with ordered categories, *Biometrika*, 68, 347-355.

GUMBEL E.J. (1960), Bivariate exponential distributions, *Journal of the American Statistical Association, 55, 698-707.*

PLACKETT R.L. (1965), A class of bivariate distributions, *Journal of The American Statistical Society*, 60, 516-522.