Research Article

Open Access

Fodé Tounkara* and Louis-Paul Rivest

Some New Random Effect Models for Correlated Binary Responses

Abstract: Exchangeable copulas are used to model an extra-binomial variation in Bernoulli experiments with a variable number of trials. Maximum likelihood inference procedures for the intra-cluster correlation are constructed for several copula families. The selection of a particular model is carried out using the Akaike information criterion (AIC). Profile likelihood confidence intervals for the intra-cluster correlation are constructed and their performance are assessed in a simulation experiment. The sensitivity of the inference to the specification of the copula family is also investigated through simulations. Numerical examples are presented.

Keywords: Multivariate exchangeable copulas, Exchangeable binary data, Profile interval, Maximum likelihood

MSC: Primary 62H10, secondary 62H12

1 Introduction

Data in the form of clustered binary responses arise in many fields of study. For example, in ophthalmology the two eyes of a patient are a cluster. In toxicological studies [8], a litter is a cluster of newborn animals. In education, the school and the classroom are clusters of students. Units within a cluster are more likely to be similar than units from different clusters and the study variables often exhibit an intra-cluster correlation. Standard statistical methods that ignore such a correlation provide a poor fit of data and can lead to erroneous inferences. Such an extra-binomial variation can be accommodated by specifying a random cluster effect: the responses in a cluster are conditionally independent given the cluster effect. A common model is the beta-binomial, see [23], where the conditional probability of success in a cluster has a beta distribution. This model has been criticized for its lack of flexibility when the cluster sizes are variable [6]. Other random effects models for clustered data include the logistic-normal-binomial model described by [24], and the probit-normal-binomial, see [15].

Several measures of association for binary traits have been proposed in the literature. These include the Kappa coefficient [5, Chap. 8], the odds ratio [3, 11]. This work focusses on the estimation of the "uncorrected" intra-cluster correlation (ICC). The ICC is a measure of the similitude of observations taken in the same cluster. It can be interpreted as an index of familial aggregation and as measure of inter-rater agreement, depending on the situation. When planning a cluster sampling or a cluster randomization trial, the clusters are the statistical units and the precision of the results depend on the ICC. Nowadays cluster randomized trials are widely used to assess various modes of community interventions. They involve assigning randomly clusters of per-

*Corresponding Author: Fodé Tounkara:** Department of Mathematics and Statistics, Université Laval„ 1045 av. de la Médecine, Québec (Québec) G1V 0A6 Canada, E-mail: fode.tounkara.1@ulaval.ca
Louis-Paul Rivest: Department of Mathematics and Statistics, Université Laval„ 1045 av. de la Médecine, Québec (Québec) G1V 0A6 Canada, E-mail: Louis-Paul.Rivest@mat.ulaval.ca

sons, often belonging to the same neighborhoodf or group, to the arms of a trial. In such situations the ICC plays a crucial role in the sample size determination, as it measures the homogeneity within clusters [5]. ICC estimates are now published in the epidemiological literature for various types of intervention, see [16] for a recent example. It is therefore important to have precise ICC estimators with reliable methods for constructing confidence intervals.

The estimation of the ICC for binary data has been considered by many authors. [17] compare more than 20 estimators of the ICC, derived using either maximum likelihood or a modified method of moments, in a Monte Carlo study. They conclude that the estimator of [7] can be used as an omnibus estimator as it generally has good sampling properties. Recently, the construction of confidence intervals for the ICC has been investigated by [4, 18, 25]. The first modeled the extra binomial variation using [12] generalized binomial distribution, [4] use a z-transform and a distribution-free large sample variance to construct an ICC confidence interval while [18] assumes a beta-binomial distribution. [18] also compared six methods for constructing confidence intervals for the ICC and found that, for beta-binomial data, the best procedure was a profile likelihood confidence interval (PLCI). Note that inference about the ICC using a Bayesian framework have also been investigated, see [1, 21, 22].

This work investigates the construction of PLCI for the ICC using a new class of probability models for exchangeable clustered binary data, associated with exchangeable Archimedean copulas [13, chap. 2]. This family contains several specifications for the extra binomial variation. The procedure investigated in this work consists in (i) selecting a model for the random cluster effect and (ii) constructing a profile confidence interval for the ICC using the model selected at step 1.

Section 2 introduces a class of model for an extra binomial variation associated with multivariate Archimedean copulas. Graphical models' comparisons are provided. Section 3 discusses model selection, the calculation of maximum likelihood estimates for the ICC, and the construction of profile likelihood confidence intervals. These new statistical methods are investigated through simulations in Section 4. Two numerical examples are used to illustrate the proposed approach in Section 5.

2 Within cluster dependency modeling for binary exchangeable data

2.1 Model formulation

Consider a random sample of K clusters of size n_i, $i = 1, 2, \ldots, K$. Let $Y_{ij}, j = 1, 2, \ldots, n_i$ be a set of n_i exchangeable Bernoulli random variables in cluster i, where $Y_{ij} = 1$ is a success and 0 is a failure. These random variables are identically distributed and possibly dependent. The total number of successes in the ith cluster is $Y_i = \sum_{j=1}^{n_i} Y_{ij}$. A standard model for the within cluster dependency assumes that the probability of success p_i is random and varies from one cluster to the next. The models considered here associate to cluster i a positive random variable θ_i whose Laplace transform, $\psi_\alpha(t) = E(e^{-t\theta_i})$, depends on a positive dependency parameter α is a such way that $P(\theta_i = 1) = 1$ when $\alpha = 0$. Let $\pi \in (0, 1)$ stands for the marginal probability of success; the proposed model is

$$p_i = e^{-\theta_i \psi_\alpha^{-1}(\pi)} \quad i = 1, \ldots, K, \tag{2.1}$$

where $\psi_\alpha^{-1}(\pi)$ denotes the function inverse of the Laplace transform $\psi_\alpha(\cdot)$ evaluated at π. Under model (2.1), $p_i \in (0, 1)$ and the marginal probability of success is $E(p_i) = \pi$. This model has a clear separation between π and the random variable for the dependency, θ_i. Indeed, $\psi_\alpha^{-1}(\pi)$ is a mere scale parameter for the distribution of $-\log p_i$. Model (2.1) assumes exchangeability within each cluster. This assumption could be questionable for familial data where the mother-child association could be stronger than that between two siblings. Table 1 gives the inverse of the Laplace transforms $\psi_\alpha^{-1}(t)$ and the densities $f_\alpha(\theta)$ of the random effects θ_i for four models commonly used in the copula literature. For Gumbel (G) and Clayton (C) families, f_α is a density defined on $\mathcal{R}^+$ while, for the models of Frank (F) and Joe (J), f_α is a probability mass function defined on the positive

Table 1: Inverse of Laplace transforms, the densities of the random effects θ_i, and moments for four copula families, Clayton (C), Frank (F), Gumbel (G), Joe (J).

Families	$\psi_\alpha^{-1}(t)$	$f_\alpha(\theta)$	λ_k
C	$(t^{-\alpha}-1)/\alpha$	$\dfrac{\theta^{1/\alpha-1}\exp(-\theta/\alpha)}{\alpha^{1/\alpha}\Gamma(1/\alpha)}$	$\left\{k\pi^{-\alpha}-k+1\right\}^{-\frac{1}{\alpha}}$
F	$-\log\left(\dfrac{e^{-\alpha t}-1}{e^{-\alpha}-1}\right)$	$(1-e^{-\alpha})^{\theta}/(\theta\alpha)$	$-\dfrac{1}{\alpha}\log\left\{1+\dfrac{(e^{-\alpha\pi}-1)^k}{(e^{-\alpha}-1)^{k-1}}\right\}$
G	$\left\{-\log(t)\right\}^{\alpha+1}$	Positive Stable	$\pi^{k^{\frac{1}{1+\alpha}}}$
J	$-\log\{1-(1-t)^{\alpha+1}\}$	$\left(\dfrac{1}{\theta}\right)(-1)^{\theta-1}$	$1-\{1-(1-(1-\pi)^{1+\alpha})^k\}^{\frac{1}{1+\alpha}}$

integers $\mathbb{N}^+$. For the models in (2.1), the k-moments $E(p_i^k)$ have the following form, $\lambda_k = \psi_\alpha\{k\psi_\alpha^{-1}(\pi)\}$; explicit expressions for these moments are given in Table 1. It follows that, given π and the association parameter α, the moment based definition of the intra-cluster correlation (ICC) ρ under models (2.1) is given by

$$\rho = \frac{E(p_i^2)-\{E(p_i)\}^2}{E(p_i)\{1-E(p_i)\}} = \frac{\psi_\alpha(2\psi_\alpha^{-1}(\pi))-\pi^2}{\pi(1-\pi)}. \tag{2.2}$$

The ICC is the correlation between different Bernoulli variables, Y_{is} and $Y_{i\ell}$, of cluster i. It is an increasing function of α, with $\rho = 0$ when $\alpha = 0$ for all models in Table 1. Further interpretations of this coefficient can be found in [5, Chap.8].

For Gumbel's model, the random effect θ_i has a positive stable distribution with parameter $1/(\alpha + 1)$ and $f_\alpha(\theta)$ does not have a closed form expression. For this model, the moments λ_k are equal to those for the q model of [8]. This provides an additional motivation for Kuk's proposal as it can be derived using stable random effects in (2.1). Consider now Clayton's family. When $\alpha = 1$, θ_i has a negative exponential distribution with parameter 1, and p_i has a beta distribution with parameters $(\frac{1}{\psi_\alpha^{-1}(\pi)}, 1)$. Therefore Clayton's model with $\alpha = 1$ gives the beta-binomial distribution with parameters $(\frac{1}{\psi_\alpha^{-1}(\pi)}, 1)$. For Frank's and Joe's model, θ_i is an integer valued random variable defined on $\mathbb{N}^+$. For Frank's copula, θ_i follows a logarithmic distribution, while for Joe's family it is the Sibuya distribution. The later distribution is heavy tailed as no moment exists. See [13] for more information about these distributions. It might seem unrealistic to have latent cluster effects distributed according to a positive stable law or to a discrete distribution. These assumptions are made to get, within (2.1), tractable models with probability mass function (pmf) having closed form expressions.

Swapping the successes and the failures, (2.1) gives new models for an extra binomial variation. The probability of success is then expressed as

$$p_i = 1 - e^{-\theta_i\psi_\alpha^{-1}(1-\pi)} \quad i = 1,\dots,K. \tag{2.3}$$

This duality was noted by [8] whose p model can be expressed as (2.3) where θ_i has a positive stable distribution. In the sequel, models obtained with (2.3) will be called dmodel since, as will be seen in the next section, they arise when the joint distribution of the Bernoulli variables Y_{ij} is expressed using a copula. In this paper, Kuk's p model is labeled by the dGumbel model. For (2.3), the intra cluster correlation is $\rho = \{\psi_\alpha(2\psi_\alpha^{-1}(1-\pi))-(1-\pi)^2\}/\{\pi(1-\pi)\}$.

Two approaches are available to calculate the pmf of Y_i. The first method evaluates alternating sums directly using the moments λ_k given in Table 1:

$$P(Y_i = k) = \binom{n_i}{k}E\{p_i^k(1-p_i)^{n_i-k}\} \tag{2.4}$$

$$= \binom{n_i}{k}\sum_{j=0}^{n_i-k}(-1)^j\binom{n_i-k}{j}\lambda_{k+j}. \tag{2.5}$$

For large clusters, say larger than 30, the evaluation of the alternating sums (2.5) is numerically unstable and yields negative probabilities. The second method to calculate the pmf of Y_i is to evaluate (2.4) directly by integrating over the random effects distribution. For the random effects having gamma distributions of

Clayton's family, we tried using the Gauss-Laguerre quadrature method. It gave poor approximations for large values of the dependence parameter α. Integrating over the random effect is possible for the models of Joe and Frank as they have discrete distributions. For these models,

$$P(Y_i = k) \approx \begin{cases} \left[1 - \sum_{\theta=1}^{N}\{1 - (1 - e^{-\theta\psi_\alpha^{-1}(\pi)})^{n_i}\}f_\alpha(\theta)\right] & \text{if } k = 0 \\ \binom{n_i}{k}\sum_{\theta=1}^{N}e^{-\theta\psi_\alpha^{-1}(\pi)k}\{1 - e^{-\theta\psi_\alpha^{-1}(\pi)}\}^{n_i-k}f_\alpha(\theta) & \text{if } 0 < k \le n_i \end{cases} \tag{2.6}$$

where N is a positive integer large enough for these approximations to be accurate. The values $N = 100$ and $N = 400$ yielded nearly identical results in the small cluster simulations that are presented in Section 4. We used $N = 400$ in the large cluster simulations presented in the next section.

Note that binomial mixture models constructed in terms of Laplace transforms can be found in [2, 8]. The innovative aspect of our paper is to parameterize these models in terms of a marginal probability π and the ICC ρ and to investigate this large class of models in a systematic way.

2.2 A copula formulation of the model

This section derives an alternative formulation for (2.1), in terms of the joint survival distribution of the Bernoulli random variables Y_{ij}, $j = 1, 2, \ldots, n_i$ for the units within a cluster. Dropping subscript i, let $\{y_j\}$ be a sequence of n zeros and ones. Under (2.1) the joint survival distribution of the Y_j's within a cluster is

$$P(Y_1 \ge y_1, \ldots, Y_n \ge y_n) = C_{\alpha,n}\{\bar{F}(y_1), \ldots, \bar{F}(y_n)\}, \tag{2.7}$$

where $\bar{F}(y)$ is the marginal survival function or Y_j, $\bar{F}(y) = P(Y \ge y) = \pi^y$ for $y = 0, 1$, and $C_{\alpha,n}(u_1, \ldots, u_n) = \psi_\alpha\{\sum_{j=1}^{n}\psi_\alpha^{-1}(u_j)\}$ is, for $u_j \in (0, 1)$ $j = 1, \ldots, n$, an n dimensional Archimedean copula for the dependency between the variables, see [13, Chap. 2] for more discussion. Note that the inverse $\psi_\alpha^{-1}(\cdot)$ of ψ_α is a decreasing function defined on $(0, 1)$ such that $\psi_\alpha^{-1}(1) = 0$. Under (2.7), $\lambda_k = P(Y_1 \ge 1, \ldots, Y_k \ge 1) = C_{\alpha,k}\{\pi, \ldots, \pi\} = \psi_\alpha\{k\psi_\alpha^{-1}(\pi)\}$.

Equation (2.7) is proved by noticing that, assuming (2.1),

$$\begin{aligned} P(Y_1 \ge y_1, \ldots, Y_n \ge y_n) &= E\{P(Y_1 \ge y_1, \ldots, Y_n \ge y_n|\theta)\} \\ &= E[\{\exp\{-\theta\sum y_j\psi_\alpha^{-1}(\pi)\}] \\ &= \psi_\alpha\{\sum y_j\psi_\alpha^{-1}(\pi)\} \\ &= \psi_\alpha[\sum \psi_\alpha^{-1}\{\bar{F}(y_j)\}] \end{aligned}$$

since, by construction, $y_j\psi_\alpha^{-1}(\pi) = \psi_\alpha^{-1}\{\bar{F}(y_j)\}$.

The value $\alpha = 0$ gives the independence copula, $C_{0,n}(u_1, \ldots, u_n) = u_1 \times \ldots \times u_n$. As α goes to ∞ the copulas for all the models of Table 1 converge to what is known as the Fréchet upper bound,

$$\lim_{\alpha\to\infty}\psi_\alpha\{\sum\psi_\alpha^{-1}(u_j)\} = \min_j u_j.$$

This corresponds to the perfect correlation where all the Bernoulli variables within a cluster takes the same value, e. g. $P(Y_1 = \ldots = Y_n) = 1$. An attractive feature of all models in Table 1 is that (π, ρ) has the full range $]0, 1[\times]0, 1[$, as $\pi \in (0, 1)$ and $\alpha \in (0, \infty)$. Moreover, ρ is a monotone function of α satisfying the following properties

$$\alpha \longrightarrow \infty \Longleftrightarrow \rho \longrightarrow 1 \text{ and } \alpha = 0 \Longleftrightarrow \rho = 0.$$

There is a copula behind most models for clustered binary data. Consider for instance the generalized binomial distribution, see [12]. Under this model the joint survival distribution within a cluster can be expressed as (2.7) with

$$C_\alpha(u_1, \ldots, u_n) = (1 - \alpha)\prod_{j=1}^{n}u_j + \alpha\min u_j, \quad \alpha \in (0, 1).$$

This expression highlights that the generalized binomial distribution is a discrete mixture model featuring two latent classes with respective probabilities $1 - \alpha$ and α. The clusters in the first class are made of independent units while in the second class all the units of a cluster have the same Y values. Such a model is rather unlikely in practice and one would expect (2.1) to represent better the within cluster association. Despite its unrealistic description of the dependency, the generalized binomial has been widely used to investigate statistical procedures for cluster binary data, see [17, 25].

Finally note that the estimation of the intra cluster correlation for discrete data using copulas was investigated by [19] in a different context. They were concerned with split-cluster designs, where individuals within a cluster are randomized to one of the two arms of a trial. They use bivariate copulas to model the association between the mean responses for the treated and for the untreated individuals in a cluster.

2.3 Graphical comparisons

A total of 8 Archimedean copulas models are available to account for an extra binomial variation. They were compared using exploratory analysis techniques such as correspondence analysis. The findings of these comparisons are now briefly presented. First the beta-binomial, the Clayton and the d-Clayton models are, for all practical purposes, equal. This is illustrated in (a), (b), and (c) of Figure 1 where the pmf for these three models obtained when $n = 15$, $\pi = 0.1, 0.2, 0.3$ and $\rho = 0.1$ are presented. Similar results are found for $\rho = 0.3$; they are presented in the supplementary material section. This means that when the beta-binomial is a candidate model for a data set, there is no need to include the Clayton and dClayton model as they provide fits that are very similar.

Graphs (d), (e) and (f) of Figure 1 compare the pmf for Joe (J), dJoe (dJ), Gumbel (G) and dGumbel (dG) models. The dJoe and dGumbel pmfs have nearly identical unimodal shapes while of the Gumbel and the Joe pmfs are similar with a bimodal structure. The pmfs of Figure 1 have a wide range of values for $Pr(Y_i = 0)$, when $\pi = 0.2$ it goes from 0.05 to 0.25. Additional comparisons, with $\rho = 0.3$ are presented in the supplementary material section. In all comparisons, the Joe and the dJoe pmf presented the most extreme shapes, while the beta-binomial pmf has an average shape, between these two extremes. Therefore, in the small cluster simulations presented in the next section, only these three models are considered as being representative of all the models of Table 1.

Figure 2 presents some pmf comparisons for clusters of size $n = 40$, with $\rho = 0.05$ and $\pi = 0.1, 0.2, 0.3$. Such small values of ρ are typical in community intervention studies, see (Légaré et al., 2011). The models presented are the beta-binomial (BB), Frank (F), dFrank (dF), Joe (J) and dJoe (dJ). Figures 2 shows that the pmf for all the models are different, except possibly for the dFrank and dJoe models, whose shape are very similar. The pmf shows a unimodal structure for the models of dFrank and dJoe, and a bimodal structure under F's and J's models. Joe's pmf assigns a large probability to zero as compared with the others models and the beta-binomial pmf has an average shape between that of the Joe and the dJoe models. Graphs for $\rho = 0.1$ are presented in the supplementary material section.

3 Maximum likelihood inference for ICC(ρ)

This section shows how to calculate maximum likelihood estimates for ρ and π using data $\{(n_i, y_i) : i = 1, \ldots, K\}$. It also discusses the construction of profile likelihood confidence intervals for ρ. Inference on ρ is carried out in a frequentist rather than a Bayesian set-up as this is commonly used in research on clustered randomized trials (Eldridge & Kerry, 2012, p.193).
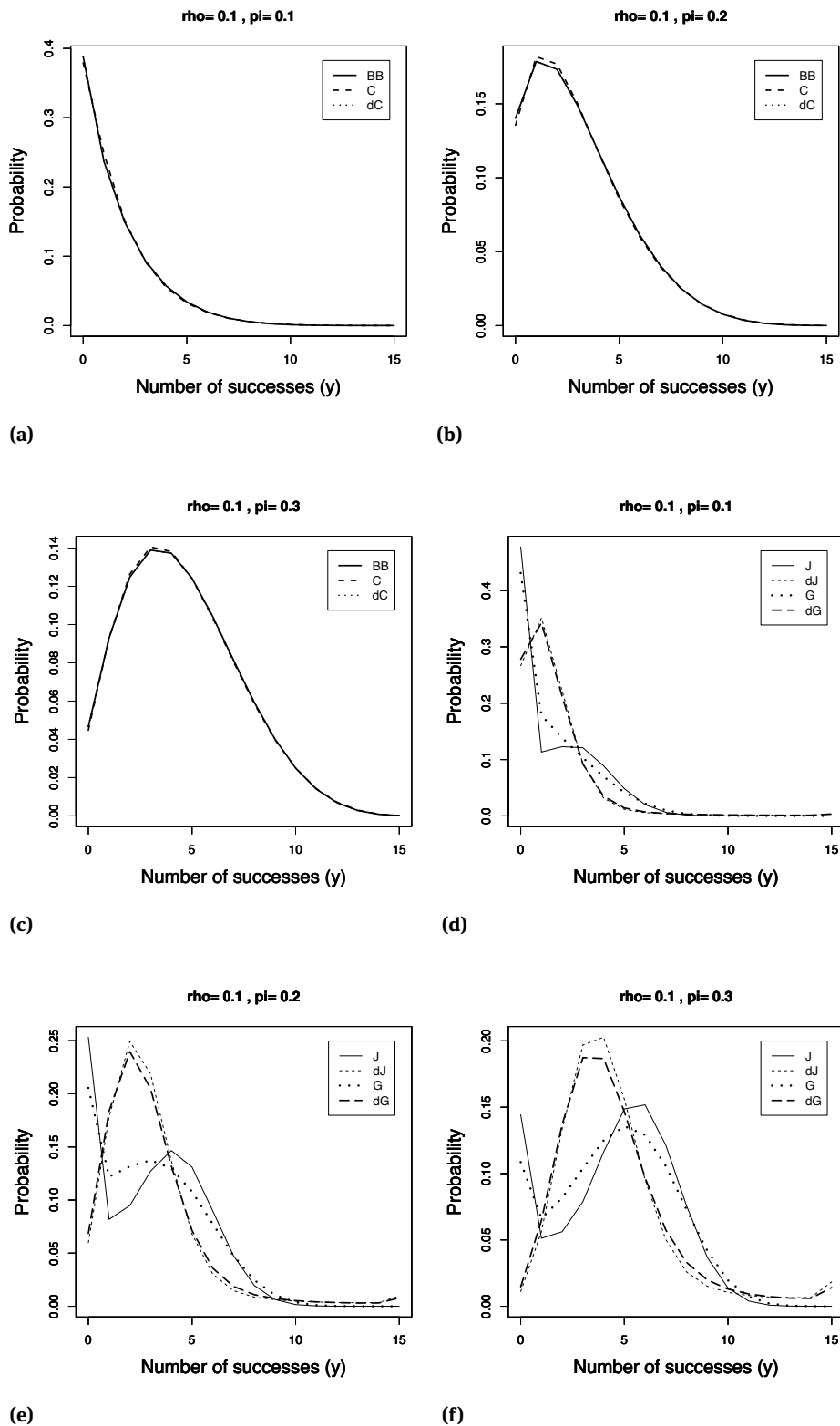
Figure 1: A comparison of the probability mass function for cluster size 15 under some small clusters models, for $\rho = 0.1$ and $\pi = 0.1, 0.2, 0.3$
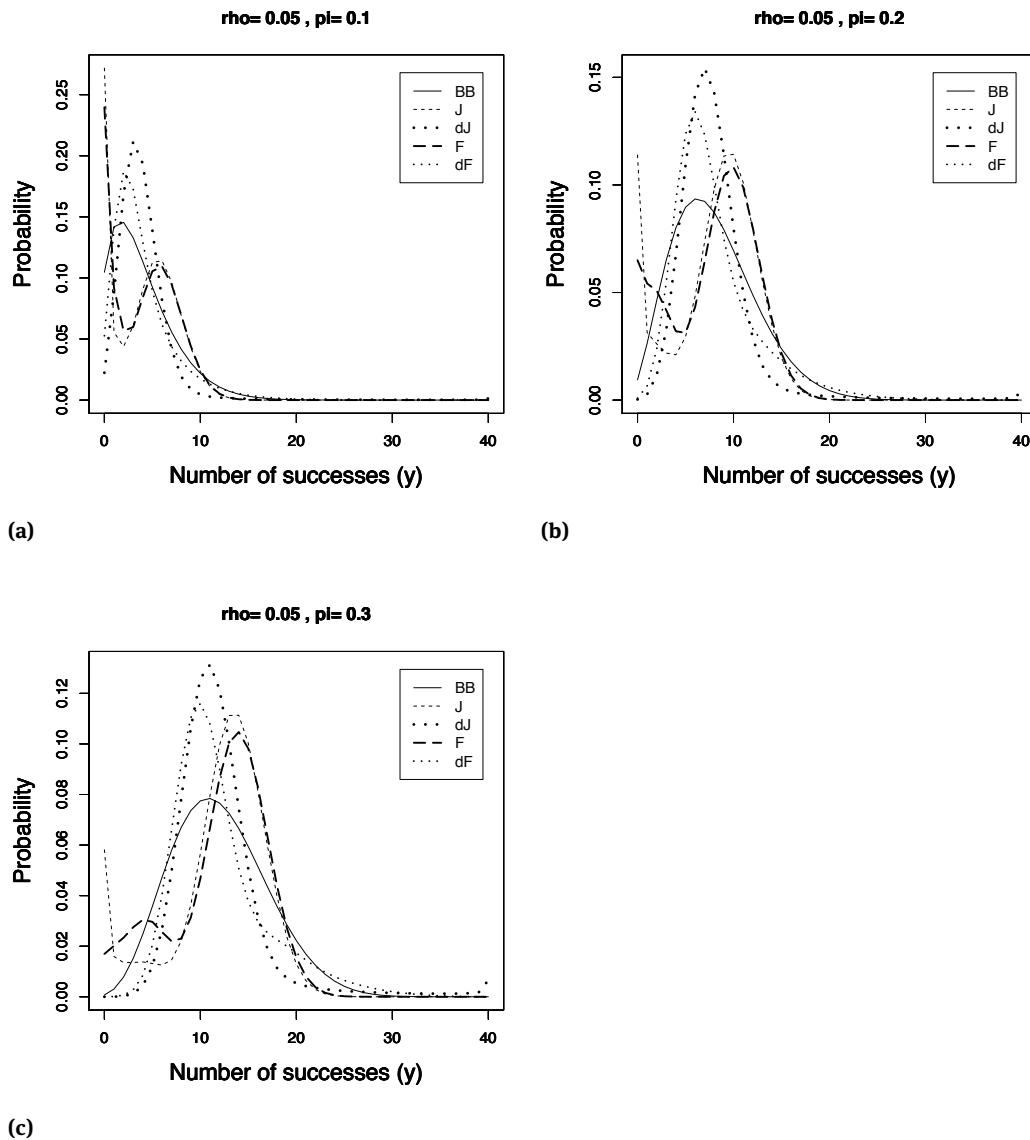
Figure 2: A comparison of the probability mass function for clusters of size 40 under the beta-binomial (BB), Joe (J), dJoe (dJ), Frank (F) and d-Frank (dIG), for $\rho = 0.05$ and $\pi = 0.1(a), 0.2(b), 0.3(b)$

3.1 Maximun Likelihood Estimates

The parameters of interest are ρ and π, thus α needs to be expressed in terms of these two parameters. The only model for which equation (2.2) leads to an explicit expression for α is Gumbel's families for which

$$\alpha = \frac{\log(2)}{\log\left[1 + \frac{\log\{\rho(1-\pi)+\pi\}}{\log \pi}\right]} - 1.$$

For all the other models of Table 1, solving (2.2) numerically is needed to calculate α; this can be done using the R function `uniroot`. The log-likelihood is

$$\log L(\pi, \rho) = \log\left\{\prod_{i=1}^{K} \binom{n_i}{y_i} E\{p_i^{y_i}(1 - p_i)^{n_i-y_i}\}\right\}. \tag{3.1}$$

It can be evaluated using one of the two methods presented in Section 2.1.

The maximum likelihood estimates are the values $\hat{\pi}$ and $\hat{\rho}$ that maximize the log-likelihood defined in (3.1). They were calculated using R. First an R function evaluating (3.1) in terms of ρ and π was written where α was evaluated in terms of ρ and π by solving (2.2) using `uniroot`. The search interval for α was set to $(10^{-5}, 10^{10})$; for some copulas (2.2) cannot be evaluated for α values as large as 10^{10} and this generates a `uniroot` warning. This had no impact on the result since `uniroot` automatically lowers the upper bound of the search interval when this happens. Then the values of π and ρ that maximized the likelihood were obtained using the function `optim` of R. We used as starting values, the sample mean $\hat{\pi}_0 = \sum y_i / \sum n_i$ and $\hat{\rho}_{FC}$, the Fleiss Cuzick moment based estimate of ρ defined in the next section. The log-likelihood is maximized in the square $[1e^{-5}, .99999] \times [1e^{-5}, .99999]$. The Fisher information matrix was then estimated as minus the Hessian of the likelihood function; this yields standard errors, $s.e.$, for the two estimates. Negative estimates of ρ are not possible. Negative estimates of ρ have little practical interest, indeed [5, Chap. 8] recommend setting them to zero and proceeding as the data with cluster was independent.

3.2 Model selection criteria

A model's fit can be assessed using a deviance, equal to the likelihood ratio statistics comparing the copula model to that of a saturated model where each cluster has its own fixed probability of success p_i,

$$Dev = 2\left[\log\left\{\sum_{i=1}^{K}\binom{n_i}{y_i} y_i^{y_i}(n_i - y_i)^{n_i - y_i}/n_i^{n_i}\right\} - \log L(\hat{\pi}, \hat{\rho})\right], \tag{3.2}$$

Note however that this deviance does not have an asymptotic χ^2_{K-2} distribution when the copula model fits well as the null copula model is not a special case of the saturated model with cluster specific p_i when K is fixed.

To select a specific model to work with, we used the Akaike Information Criterion (AIC), defined by

$$AIC = -2\log L(\hat{\pi}, \hat{\rho}) + 2m, \tag{3.3}$$

where m, the number of parameters, is equal to 2 for all the copula models considered in this work. The preferred model is the one with the minimum AIC value. We did not use the model averaging approach of [14] as this does not permit the calculation of profile confidence intervals.

3.3 Profile Likelihood Confidence Interval (PLCI)

Wald confidence intervals for ρ, with a $100(1 - \tau)$ confidence level, are given by $\hat{\rho} \pm z_{1-\tau/2} s.e.(\hat{\rho})$, where $z_{1-\tau/2}$ is a normal quantile. This procedure may perform poorly for small to moderate sample sizes when estimating ρ, see Donner and Eliasziw (1992). This section reviews the construction of PLCI for ρ when using copula models associated to (2.1). Such intervals are generally considered to be more accurate than Wald confidence intervals. The profile likelihood for ρ is given by $L_P(\rho) = \max_\pi L(\pi, \rho)$, $\rho \in (1e^{-5}, .99999)$. We used the R function `optimize` for its evaluation; it involved a `uniroot` function to evaluate α as a function of ρ and π. It may happen that $L_P(\rho)$ cannot be numerically evaluated for large values of ρ; to calculate the profile confidence interval it is important to identify a value ρ_M such that $L_P(\rho)$ can be evaluated for $\rho \in (1e^{-5}, \rho_M)$. The $100(1 - \tau)\%$ PLCI for ρ is defined as the set

$$CI = \left\{\rho: \quad \log L_P(\rho) > \log L(\hat{\rho}, \hat{\pi}) - \frac{1}{2}\chi_1^2(1 - \tau)\right\}, \tag{3.4}$$

where $\chi_1^2(1 - \tau)$ is the $1 - \tau$ quantile of a chi-squared distribution with one degree of freedom. The set CI is typically an interval and is denoted $(\hat{\rho}_L, \hat{\rho}_U)$. We computed the two limits by finding two solutions to the equation $\log L_P(\rho) = \log L(\hat{\rho}, \hat{\pi}) - \frac{1}{2}\chi_1^2(1 - \tau)$, one in the interval $(0, \hat{\rho})$ and the other in the interval $(\hat{\rho}, \rho_M)$,

using the R function `uniroot`. When $\log L(\hat{\rho}, \hat{\pi}) - \frac{1}{2}\chi_1^2(1 - \tau) < \log L_P(1e^{-5})$, the interval lower bound was set at $\hat{\rho}_L = 0$. If ρ_M is not specified correctly then $\log L_P(\rho_M)$ cannot be evaluated; in such cases `uniroot` returns $\hat{\rho}$ as the value of the confidence interval upper bound. This is not a legitimate upper bound and samples for which $\hat{\rho}_U = \hat{\rho}$ were rejected in the simulations; such unacceptable upper bounds occurred only in the large cluster simulations. For a quick evaluation of $(\hat{\rho}_L, \hat{\rho}_U)$ the R function `plkhci` could also be used. When the true value of ρ is close to 0, the lower bound $\hat{\rho}_L$ is likely to be 0 and in this case the true coverage of the confidence might be slightly above its nominal level.

PLCI for ρ, constructed using the beta-binomial distribution, were investigated by [18]. He concluded through Monte Carlo simulations that the profile confidence interval performs well over a wide range of parameter values. A goal of this work is to investigate whether the beta-binomial profile confidence interval is an omnibus procedure applicable in all circumstances or whether the fit of the beta-binomial model should first be ascertained.

4 Simulation Study

The purpose of this simulation study is to assess whether the AIC criterion (3.3) of Section 3.2 provides a reliable model selection criterion. Its goal is also to assess the performance of the profile likelihood confidence intervals for ρ constructed for the models of Table 1. Two sets of simulations were carried out. The first featured small clusters, with sizes ranging between 1 and 15 with a mean of 4. The second one had large clusters with n_i between 27 to 65 with a mean of 38. Data was generated from the beta-binomial distribution, the Joe, dJoe, frank and dFrank models. For J's and dJ's models, cluster specific θ_i were generated from the Sibuya distribution while for the Frank and dFrank model, the logarithmic distribution was used. The accuracy of the AIC model selection criterion was measured using the proportions of times that model M_0 is selected when the data was simulated from model M_1:

$$minAIC(M_0, M_1) = \frac{\sum_{i=1}^{1000} I_i}{1000} \times 100,$$

where the indicator I_i equal to 1 if model M_0 provides the smallest AIC for sample i and 0 otherwise. The simulations also investigated 95% two-sided confidence intervals for ρ. Their performance was measured using the coverage (cov) and the average length (CIL) defined as

$$Cov = \frac{\sum_{i=1}^{1000} A_i}{1000} \times 100; \qquad\qquad CIL = \frac{\sum_{i=1}^{1000}(\hat{\rho}_{U_i} - \hat{\rho}_{L_i})}{1000},$$

where A_i equal to 1 if the true value ρ is in $(\hat{\rho}_{L_i}, \hat{\rho}_{U_i})$, and 0 otherwise. All programming was done in R. Samples with $y_i = 0$ or $y_i = n_i$ for all clusters were rejected.

4.1 Small cluster simulations

Three models are investigated in the small cluster simulations, beta-binomial, Joe and dJoe. Two values of ρ and π, namely 0.1 and 0.3, and three values of K, 100, 250, and 500, were considered. Table 2 summarizes the proportion of times that a model is selected as a function of the model from which the data is simulated. The relatively high values on the diagonals of each quadrant of Table 2 show that the models are identifiable. This is more pronounced, as either K, π or ρ increases. When a data set is generated from Joe's model, dJoe is nearly never selected and vice-versa. These results are in line with the findings of the graphical comparisons of Figures 1 where these two models were very different. Table 2 also confirms the middle ground nature of the beta-binomial since this model has the smallest number of correct identifications.

Table 3 reports results on PLCI that assume that the model is selected correctly. It evaluates PLCI for ρ under the beta-binomial, the Joe and the dJoe models. The PLCI has generally good statistical properties; its coverage is close to nominal 95% level and its average length is relatively small. As expected, the expected

Table 2: Small cluster simulations: The proportion of times that a model is selected by the AIC criterion as a function of the model used to simulate the data.

| | | | $\pi = 0.1$ | | | $\pi = 0.3$ | | |
| | | | Data | | | Data | | |
ρ	K	Models	BB	J	dJ	BB	J	dJ
0.1	100	BB	42.50	21.30	20.90	55.50	18.40	15.60
		J	43.50	74.20	14.70	27.90	79.80	5.90
		dJ	14.00	4.50	64.40	16.60	1.80	78.50
	250	BB	58.60	16.80	13.90	79.50	11.10	7.00
		J	33.60	82.80	4.90	15.00	88.90	1.00
		dJ	7.80	0.40	81.20	5.50	0.00	92.00
	500	BB	75.50	13.20	7.70	92.40	4.80	2.00
		J	22.20	86.80	0.80	5.40	95.20	0.00
		dJ	2.30	0.00	91.50	2.20	0.00	98.00
0.3	100	BB	54.70	16.60	15.70	79.00	8.50	9.50
		J	33.30	83.00	3.40	14.40	91.50	0.20
		dJ	12.00	0.40	80.90	6.60	0.00	90.30
	250	BB	79.10	10.10	6.70	94.80	2.60	0.80
		J	16.60	89.90	0.10	3.80	97.40	0.00
		dJ	4.30	0.00	93.20	1.40	0.00	99.20
	500	BB	93.20	4.70	1.30	99.80	0.10	0.00
		J	5.90	95.30	0.00	0.20	99.90	0.00
		dJ	0.90	0.00	98.70	0.00	0.00	100.00

confidence length decreases as the number of clusters increases. The PLCI for Joe's model is always shorter than the other two, with a very good empirical coverage. Thus PLCI constructed using the three models investigated in this section are reliable statistical techniques.

Additional simulation results are presented in the Supplementary Material Section. First the coverage and the average length of [25] ICC confidence interval constructed using the moment estimator of [7],

$$\hat{\rho}_{FC} = 1 - \frac{\sum y_i(n_i - y_i)/n_i}{(\sum n_i - K)\hat{\pi}(1 - \hat{\pi})}$$

are reported together with the performance of the beta-binomial PLCI when the model is misspecified, that is when the data comes from the Joe and the dJoe model (see Web Table 1 in the "Web Appendix B"). The Zou & Donner confidence interval is very conservative as it is always much wider than the PLCI. These simulations also show that the PLCI for the beta-binomial is sensitive to a model misidentification. For instance for dJ data, $K = 250$, $\pi = 0.1$, and $\rho = 0.1$, the estimated coverage of the beta-binomial PLCI, 76.70%, is much smaller than the nominal value of 95%. However, when data are from J's model this method in general has a tendency to provide the observed coverage probabilities that are slightly larger than the nominal level.

Table 3: Small cluster simulations: Confidence interval length (CIL) and empirical coverage (COV) of two-sided PLCI for ρ with a nominal 95% confidence level under beta-binomial, Joe, and dJoe models

| | | | $\rho = 0.1$ | | | | $\rho = 0.3$ | | | |
| | | | $\pi = 0.1$ | | $\pi = 0.3$ | | $\pi = 0.1$ | | $\pi = 0.3$ | |
	Data	Models	CIL	Cov	CIL	Cov	CIL	Cov	CIL	Cov
$K = 100$	BB	BB	0.22	95.30	0.17	93.90	0.33	95.20	0.23	95.60
	J	J	0.18	93.80	0.15	95.40	0.27	93.90	0.21	94.00
	dJ	dJ	0.27	95.90	0.19	93.70	0.38	95.29	0.26	95.40
$K = 250$	BB	BB	0.14	95.90	0.12	95.00	0.21	94.20	0.15	93.00
	J	J	0.11	94.50	0.10	94.40	0.17	95.10	0.13	94.10
	dJ	dJ	0.19	94.60	0.13	94.30	0.26	94.70	0.17	96.20

4.2 Large cluster simulations

We ran large cluster simulations with K =10, 25, and 50 clusters. Two values of ρ, 0.05 and 0.1, and of π, 0.1 and 0.3, were investigated. Besides the beta-binomial model, the models of Frank, dFrank, Joe, and dJoe were considered. Table 4 shows that the AIC has a hard time identifying the models when K = 10. The proximity, in Figure 2, of the pmf for Joe's and Frank's model shows up in Table 4 where the AIC has difficulties distinguishing one from the other. This improves as K increases and at K = 50 most models are identifiable. The identifiability increases with π. Results in Tables 5 show that, when the model is correctly identified, the PLCI

Table 4: Large cluster simulations: The proportion of times that a model is selected by the AIC criterion as a function of the model used to simulate the data.

ρ	K	Models	$\pi = 0.1$ Data BB	J	dJ	F	dF	$\pi = 0.3$ Data BB	J	dJ	F	dF
0.05	10	BB	40.85	5.90	32.86	8.77	34.38	54.25	11.58	28.93	19.22	27.78
		J	10.58	63.07	11.43	43.93	6.16	0.94	50.60	0.00	14.41	0.23
		dJ	1.48	0.00	20.20	0.10	8.01	1.15	0.00	30.91	0.00	6.94
		F	19.68	30.62	4.69	45.67	6.97	17.94	36.16	4.46	64.92	3.12
		dF	27.41	0.41	30.82	1.53	44.48	25.71	1.67	35.70	1.45	61.92
	25	BB	62.53	1.50	20.76	4.70	33.40	74.15	2.13	14.80	4.74	15.92
		J	3.71	68.20	4.89	36.64	1.42	0.40	72.44	0.00	19.17	0.00
		dJ	0.60	0.00	39.24	0.00	9.24	0.50	0.00	54.50	0.00	8.16
		F	11.02	30.30	4.58	58.66	1.93	9.62	25.33	1.60	75.98	0.41
		dF	22.14	0.00	30.53	0.00	54.01	15.33	0.10	29.10	0.10	75.51
	50	BB	80.10	0.00	12.77	1.00	24.95	90.40	0.10	5.41	1.30	8.50
		J	0.20	72.70	2.81	28.30	0.00	0.20	85.90	0.11	9.10	0.00
		dJ	0.10	0.00	55.81	0.00	5.01	0.10	0.00	73.41	0.00	5.40
		F	3.40	27.30	3.07	70.70	0.10	2.20	14.00	0.97	89.60	0.00
		dF	16.20	0.00	25.54	0.00	69.94	7.10	0.00	20.11	0.00	86.10
0.1	10	BB	38.60	4.10	23.95	7.11	32.42	48.64	4.66	18.16	12.29	24.01
		J	11.45	63.20	8.41	42.28	4.00	1.51	60.10	0.00	22.46	0.00
		dJ	3.14	0.00	30.10	0.30	11.05	1.01	0.00	41.84	0.00	8.73
		F	23.51	31.90	4.85	48.00	11.05	20.95	34.72	2.37	63.24	2.81
		dF	23.30	0.80	32.69	2.30	41.47	27.90	0.52	37.63	2.01	64.45
	25	BB	67.20	0.90	13.83	3.70	30.33	79.70	0.50	5.24	3.60	12.14
		J	1.20	67.00	2.15	35.60	0.10	0.00	80.10	0.00	13.70	0.00
		dJ	0.60	0.00	51.97	0.00	9.01	0.00	0.00	67.43	0.00	6.32
		F	10.80	32.10	2.15	60.40	1.00	7.30	19.40	0.21	82.70	0.10
		dF	20.20	0.00	29.92	0.30	59.56	13.00	0.00	27.12	0.00	81.44
	50	BB	82.20	0.00	5.31	0.90	19.80	93.50	0.00	0.81	0.80	4.60
		J	0.10	70.90	1.06	28.80	0.10	0.00	87.90	0.00	7.50	0.00
		dJ	0.10	0.00	71.44	0.00	3.80	0.00	0.00	83.35	0.00	2.40
		F	3.00	29.10	0.32	70.30	0.00	0.80	12.10	0.10	91.70	0.00
		dF	14.60	0.00	21.87	0.00	76.30	5.70	0.00	15.74	0.00	93.00

constructed with the beta-binomial, the Joe, the dJoe, the Frank and the dFrank models have good coverage properties, even with a relatively low (K = 10) number of clusters. As expected, the average confidence interval length decreases with the number of clusters. In general, the Joe and the Frank PLCI provide smaller expected lengths and an empirical coverage close to the nominal values

In Web Table 2 of "Web Appendix B", we reported confidence interval length (CIL) and empirical coverage (COV) of Zou and Donner's interval and of the beta-binomial PLCI, when the data comes from J's, dJ's F's and dF's distributions. We can see that Zou and Donner's method is very conservative. On the other hand, the PLCI based on beta-binomial model provides coverage below the nominal level when the model is misspecified. Thus using an omnibus PLCI based on the beta-binomial distribution is not a robust statistical procedure.

Table 5: Larger cluster simulations: Confidence interval length (CIL) and empirical coverage (COV) of two-sided PLCI for ρ with a nominal 95 confidence level for data generated from 5 models

| | | | $\rho = 0.05$ | | | | $\rho = 0.1$ | | | |
| | | | $\pi = 0.1$ | | $\pi = 0.3$ | | $\pi = 0.1$ | | $\pi = 0.3$ | |
	Data	Models	CIL	Cov	CIL	Cov	CIL	Cov	CIL	Cov
$K = 10$	BB	BB	0.19	98.48	0.15	97.81	0.27	97.87	0.21	98.03
	J	J	0.11	96.60	0.17	92.92	0.14	95.97	0.19	97.24
	dJ	dJ	0.46	93.96	0.31	96.30	0.49	97.18	0.33	98.28
	F	F	0.10	96.24	0.14	96.08	0.14	94.81	0.17	97.36
	dF	dF	0.25	97.82	0.20	96.47	0.32	98.57	0.24	98.60
$K = 25$	BB	BB	0.10	96.03	0.08	97.74	0.15	95.67	0.12	96.46
	J	J	0.06	96.43	0.10	94.22	0.08	95.19	0.12	96.64
	dJ	dJ	0.29	95.20	0.18	97.04	0.31	97.06	0.21	97.39
	F	F	0.06	96.80	0.09	95.63	0.08	93.59	0.10	95.97
	dF	dF	0.13	97.22	0.11	96.79	0.19	95.26	0.15	95.51
$K = 50$	BB	BB	0.06	94.16	0.06	94.93	0.11	94.10	0.09	95.60
	J	J	0.04	95.60	0.07	95.12	0.06	94.50	0.09	96.76
	dJ	dJ	0.19	97.54	0.12	97.53	0.22	98.22	0.15	94.86
	F	F	0.04	95.63	0.06	95.30	0.06	94.50	0.08	97.15
	dF	dF	0.08	96.45	0.07	94.74	0.13	95.90	0.10	94.78

5 Examples

In this section, we present two real life examples to illustrate the performance of estimators obtained under (2.1) and (2.3).

5.1 A pilot trial about shared decision making

This example is concerned with the impact of training physicians in shared decision making on their patients' involvement in the decision making process, see [9] for more details. The data in Table 7 presents the number of patients y_i reporting an active role in the decision about taking an antibiotics treatment for an acute respiratory infection, and the total patient enrollment n_i in $K = 9$ community health services of the Quebec region.

Table 6: Number of patients reporting an active role in the decision making process for 9 community health services

y_i	3	8	10	7	2	3	13	8	8
n_i	38	39	49	29	49	35	51	42	27

The ICC ρ enters the sample size calculation in these trials. Getting a reliable estimate is needed for planning purposes. Its value is typically small in community intervention trials and the upper limit of a confidence interval is useful to calculate a conservative sample size.

The deviance of the independence model is 19.366 for 9 degrees of freedom; the p-value of the independence test is $\Pr(\chi_9^2 > 19.366) = 1.3\%$. This suggests that $\rho > 0$. This is a large cluster data set and only six sets of estimates are available. The values of $\hat\rho$ reported in Table 7 differ little, except for those obtained with the dJ and dF models which are null. The corresponding AIC are larger than that for the independence model, suggesting a poor fit for these models. The best fitting model is that of Frank . Thus Frank's model is plausible for this data and the 95% profile confidence interval for ρ of (0.001,0.097) appears to be reliable.

As an application of this result, suppose that one were to design a cluster randomized trial to compare the efficiency of a method for training physicians in share decision making, with respect to no training, on the

Table 7: Six set of estimators for a community intervention trial.

Models	MLEs		95% CI for ρ		Model selection
	$\hat{\pi}$	$\hat{\rho}$	Lower	Upper	AIC
ZD	0.17	0.023	0.000	0.541	–
Binomial	0.17	–	–	–	52.36
BB	0.17	0.029	0.000	0.128	51.32
F	0.18	0.033	0.001	0.097	50.08
dF	0.17	0.000	0.000	0.277	54.36
J	0.18	0.027	0.000	0.104	52.22
dJ	0.17	0.000	0.000	0.293	54.36

proportion of patients engaging in share decision making. The test level is set at 5%, the power at 80% and the two proportions of patients engaging in shared decision making are 50% and 70% for the untreated and the treated community health services. The unit are patients and suppose that a community health service contributes 10 patients to the trial. How many community health services are needed, for each arm of the trial, to achieve the planned power of 80%? The result critically depends on the value of ρ. Using the standard sample size formula ([4, section 2.2], [5, Chap. 7]) with $\rho = 0.097$, 17 health community services are needed while this number drops to 12 if the calculations are done with the mle $\hat{\rho} = 0.033$.

5.2 Chronic Obstructive Pulmonary Disease (COPD) data

This data is presented in Example 3 of [10], and is considered by [20, 25]. In this data the familial aggregation of the COPD is used as a measure of how genetic and environmental factors may contribute to disease etiology. It involves 203 siblings from 100 families with sizes ranging from 1 to 6. The binary response here indicates whether a given sibling has impaired pulmonary functions.

In the Table 8 , we report estimates of π and ρ, profile confidence intervals for ρ, and deviances obtained for all the models considered in this work. Results derived with the nonparametric (NP) model of [20], with six parameters λ_k, $k = 1, \ldots, 6$, are also provided. As expected from graphical comparisons, the dG and dJ models have very similar fits. These models provide the best fits to the COPD data in terms of the AIC value. Moreover, the likelihood ratio test for comparing the fit of dGumbel model with that of the nonparametric model has $\chi^2_4 = 1.75$ p-value=0.78; this is not significant and the dGumbel fits well. The dGumbel 95% profile confidence interval for ρ is (0.036, 0.395); it is very close to the dJoe interval that has been shown to reliable in the small cluster simulations. Their length are 10% smaller than that of the Zou & Donner interval. As expected, the beta-binomial distribution, the Clayton and dClayton models give similar fits. Joe's and Frank's model do not fit. This agrees with graphical comparisons of Figure 1 where these two pmfs give much larger probabilities to 0 than the other model. Globally all profile likelihood confidence intervals are similar in this example, with differences less than 10% in confidence interval length. Given the small family size, the fact that the selection of a particular model has a limited impact on the analysis does not come as a surprise. The statistical methods proposed in this work are useful when dealing with larger clusters.

6 Discussion

This paper has suggested to model the intra cluster association for binary data using multivariate Archimedean copulas. This family contains a wide range of distributional shape for the extra binomial variation. It has been demonstrated through a simulation study that the AIC is a useful criterion for selecting a particular model for the extra binomial variation. An important conclusion of our study is that profile likelihood confidence intervals for Archimedean copulas models have good coverage properties, even when the number of clusters is small. The model selection step is important. An omnibus method, such as using the beta-binomial profile

Table 8: MLEs of π and ρ, and two-sided 95% confidence interval for ρ under 11 models for the (COPD) data. Degrees of freedom (*DF*), Deviance (*Dev*) and , AIC criterion of models are also reported.

	MLEs		95% CI for ρ			Goodness of fit		Model selection
Models	$\hat{\pi}$	$\hat{\rho}$	Lower	Upper	CL	*DF*	*Dev*	*AIC*
ZD	0.296	0.180	0.008	0.402	0.394	-	-	-
B	0.296	–	–	–	—	99	146.232	188.065
BB	0.282	0.213	0.059	0.401	0.342	98	137.406	181.239
G	0.283	0.207	0.052	0.388	0.336	98	138.760	182.593
dG	0.280	0.190	0.036	0.395	0.359	98	136.018	179.851
C	0.282	0.212	0.059	0.400	0.341	98	137.344	181.177
dC	0.282	0.213	0.059	0.403	0.344	98	137.407	181.240
J	0.286	0.191	0.038	0.368	0.329	98	139.848	183.681
dJ	0.281	0.180	0.031	0.390	0.359	98	136.101	179.934
F	0.286	0.201	0.045	0.382	0.336	98	139.342	183.175
dF	0.282	0.206	0.051	0.402	0.352	98	136.826	180.659
NP	0.284	0.200	–	–	–	94	134.266	186.099

likelihood confidence interval for the ICC regardless of whether this model fits well, may lead to confidence interval with poor coverage properties.

References

[1] Ahmed, M. and Shoukri, M. (2010). A bayesian estimator of the intracluster correlation coefficient from correlated binary responses. *J. Data Sci.*, 8:127–137.

[2] Alanko, T. and Duffy, J. C. (1996). Compound binomial distribution for modeling consumption data. *The Statistician*, 45:269–286.

[3] Ananth, C. V. and Preisser, J. S. (1999). Bivariate logistic regression: Modeling the association of small for gestational age births in twin gestations. *Stat. Med.*, 18:2011–2023.

[4] Chakraborty, H., Moore, J., Carlo, W. A., Hartwell, T. D., and Wright, L. L. (2009). A simulation based technique to estimate intracluster correlation for a binary variable. *Contemporary Clinical Trials*, 30:71–80.

[5] Eldridge, S. and Kerry, S. (2012). *A Practical Guide to Cluster Randomised Trials in Health Services Research*. Wiley, New York.

[6] Feng, Z. and Crizzle, J. E. (1992). Correlated binomial variates: Properties of estimators of intraclass correlation and its effect on sample size calculation. *Stat. Med.*, 11:1600–1614.

[7] Fleiss, J. L. and Cuzick, J. (1979). The reliability of dichotomous judgments: Unequal numbers of judges per subject. *Appl. Psychol. Meas.*, 3:537–542.

[8] Kuk, A. Y. C. (2004). A generalized estimating equation approach to modelling foetal response in developmental toxicity studies when the number of implants is dose dependent. *J. Roy. Statist. Soc. Ser. C*, 52:52–61.

[9] Légaré, F., Labrecque, M., LeBlanc, A., Njoya, M., Laurier, C., Côté, L., Godin, G., Thivierge, R. L., O'Connor, A., and S., S.-J. (2011). Training family physicians in shared decision making for the use of antibiotics for acute respiratory infections: a pilot clustered randomized controlled trial. *Health Expect.*, 1:96–110.

[10] Liang, K. Y., Qaqish, B., and Zeger, S. (1992). Multivariate regression analyses for categorical data. *J. Roy. Statist. Soc. Ser. B.*, 54:3–40.

[11] Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, 78:153–160.

[12] Madsen, R. W. (1993). Generalized binomial distribution. *Comm. Statist. Theory Methods*, 22:3065–3086.

[13] Mai, J.-M. and Scherer, M. (2012). *Simulating Copulas; Stochastic Models, Sampling Algorithms and Applications*. Series in Quantitative Finance: Volume 4. World Scientific Publishing Company, New York.

[14] Nikoloulopoulos, A. K. and Karlis, D. (2008). Multivariate logit copula model with an application to dental data. *Stat. Med.*, 27:6393–6406.

[15] Ochi, Y. and Prentice, R. L. (1984). Likelihood inference in a correlated probit regression model. *Biometrika*, 71:531–542.

[16] Pals, S. L., Beaty, B. L., Posner, S. F., and Bull, S. (2009). Estimates of intraclass correlation for variables related to behavioral hiv/std prevention in a predominantly african american and hispanic sample of young women. *Health Education & Behavior*, 36:182–194.

[17] Ridout, M. S., Demétrio, C. G. B., and Firth, D. (1999). Estimating intraclass correlation for binary data. *Biometrics*, 55:137–148.

[18] Saha, K. K. (2012). Profile likelihood-based confidence interval of the intraclass correlation for binary outcome data sampled from clusters. *Stat. Med.*, 31:3982–4002.

[19] Shoukri, M. M., Kumar, P., and Colak, D. (2011). Analyzing dependent proportions in cluster randomized trials: Modeling inter-cluster correlation via copula function. *Comput. Statist. Data Anal.*, 55:1226–1235.

[20] Stefanescu, C. and Turnbull, B. W. (2003). Likelihood inference for exchangeable binary data with varying cluster sizes. *Biometrics*, 59:18–24.

[21] Turner, R. M., Omar, R. Z., and Thompson, S. G. (2001). Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Stat. Med.*, 20:453–472.

[22] Turner, R. M., Omar, R. Z., and Thompson, S. G. (2006). Constructing intervals for the intracluster correlation coefficient using bayesian modelling, and application in cluster randomized trials. *Stat. Med.*, 25:1443–1456.

[23] Williams, D. A. (1975). The analysis of binary response from toxicological experiments involving reproduction and teratogenicity. *Biometrics*, 31:949–952.

[24] Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *J. Appl. Statist.*, 31:305–309.

[25] Zou, G. and Donner A., (2004). Confidence interval estimation of the intraclass correlation coefficient for binary outcome data. *Biometrics.*, 60:807–811.