



**The University of British Columbia
Vancouver, Canada**

**THE ESTIMATION METHOD OF
INFERENCE FUNCTIONS FOR MARGINS
FOR MULTIVARIATE MODELS**

by

Harry Joe and James J. Xu

Technical Report #166

**THE
DEPARTMENT
OF STATISTICS**

October 1996

ABSTRACT

An estimation approach is proposed for models for a multivariate (non-normal) response with covariates when each of the parameters (either a univariate or a dependence parameter) of the model can be associated with a marginal distribution. The approach consists of estimating univariate parameters from separately maximizing univariate likelihoods, and then estimating dependence parameters from separate bivariate likelihoods or from a multivariate likelihood. The analysis of this method is done through the theory of inference or estimating functions, and the jackknife method is proposed for obtaining standard errors of the parameters and functions of the parameters. The approach proposed here make a large contribution to the computational feasibility of carrying out inference with multivariate models. Examples illustrate the approach, and simulation results are used to indicate the efficiency.

Key words: estimating equations, inference functions, copula, jackknife for variance estimation, likelihood, multivariate non-normal response, marginal distributions.

1. Introduction.

This report is concerned with an estimation approach that can be applied to parametric multivariate models in which each parameter of the model can be associated with a lower-dimensional margin. By a multivariate model, we mean a model for a multivariate response vector $\mathbf{Y} = (Y_1, \dots, Y_d)$ with covariate vector \mathbf{x} . The parameters are either univariate parameters or dependence parameters, and the approach is aimed primarily at a multivariate non-normal response. For a sample of size n , the multivariate data has the form $(\mathbf{Y}_i, \mathbf{x}_i)$ for the i th subject, with $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{id})$, where d is the dimension of the response vector. If the Y_{ij} 's are repeated measures or observed sequentially in time, then more generally we could have a time-varying or margin-dependent covariate vector.

The set of parameters of the model are estimated through a (nonlinear) system of estimating equations, with each estimating equation being a score function (partial derivative of a log-likelihood) from some marginal distribution of the multivariate model. This method is called the *inference function for margins* (IFM) method because the theory of inference functions (Godambe 1960, 1976, 1991; McLeish and Small, 1988) imposes optimality criteria on the functions in the estimating equations rather than the estimators obtained from them.

We combine this method with the use of the jackknife method for estimation of the standard errors of the parameters and functions of the parameters. This eliminates the need for analytic derivatives to

obtain the inverse Godambe information matrix or asymptotic covariance matrix associated with the vector of parameter estimates. In addition, because each inference function derives from some log-likelihood of a marginal distribution, the inference or score functions do not have to be obtained explicitly (that is, derivatives are not needed) if the numerical optimizations for the log-likelihoods of the marginal distributions are done with a quasi-Newton routine (e.g., Nash, 1990). The idea of the jackknife for use with estimating equations is quite new; previously, Lipsitz, Dear and Zhao (1994) used a one-step jackknife with estimating equations in a context of clustered survival data.

The estimation method applies to models in which the univariate margins are separated from the dependence structure, for example, when the dependence is summarized through a copula. If F is a d -variate distribution or model, with univariate margins F_1, \dots, F_d , then there is a copula or multivariate $U(0, 1)$ distribution C such that

$$F(y_1, \dots, y_d) = C(F_1(y_1), \dots, F_d(y_d)). \quad (1.1)$$

(See Joe, 1993; Joe and Hu, 1996, for parametric families of copulas with nice properties.) This includes the case of multivariate extreme value distributions in which the univariate margins are each in the generalized extreme value family and the copula is in the multivariate extreme value class (C satisfies $C(u_1^t, \dots, u_d^t) = C^t(u_1, \dots, u_d)$ for all $t > 0$). Other examples are models in which the multivariate normal distribution is used as a latent vector distribution, for example, the multivariate probit model for multivariate binary or ordinal data (Ashford and Sowden, 1970; Anderson and Pemberton, 1985) and the multivariate Poisson-log normal distribution for count data (Aitchison and Ho, 1989).

The use of the IFM method with the jackknife for estimation of standard errors makes many more multivariate models computationally feasible to work with for data analysis and inference. The previous lack of this method may explain why models such as the multivariate probit and multivariate Poisson-log normal models have not been much used (see Lesaffre and Molensbergh, 1991, on the former model).

The details of the IFM method and some models for which the method applies are given in Section 2. To concentrate on the ideas of the method and to avoid cumbersome notation, we do not present it in its most general possible form. Section 3 summarizes some advantages of the IFM method. Section 4 contains asymptotic results associated with the IFM method, including two ways for handling covariates. Section 5 has some simulation results showing the efficiency of the estimates from the IFM method compared with the classical maximum likelihood estimation method. Section 6 has a couple of brief examples involving data sets to illustrate the use of the IFM method. The paper concludes with a discussion in Section 7 which includes extensions of the concepts in Section 2.

For notation, vectors are usually assumed to be row vectors unless otherwise stated.

2. Inference method for margins and jackknife.

The inference function for margin (IFM) method is useful for models with the closure property of parameters associated with or being expressed in lower-dimensional margins (Xu, 1996). Here to concentrate on the ideas, we demonstrate the main details for a parametric model of the form (1.1), first without covariates and then with covariates. This is followed by its application to a few models that have appeared in the statistical literature. From this, we expect that the reader can see the general applicability of the method.

Consider a copula-based parametric model for the random vector \mathbf{Y} , with cumulative distribution function (cdf)

$$F(\mathbf{y}; \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d, \boldsymbol{\theta}) = C(F_1(y_1; \boldsymbol{\alpha}_1), \dots, F_d(y_d; \boldsymbol{\alpha}_d); \boldsymbol{\theta}) \quad (2.1)$$

where F_1, \dots, F_d are univariate cdfs with respective (vector) parameters $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d$, and C is a family of copulas parametrized by a (vector) parameter $\boldsymbol{\theta}$. We assume that C has a density c (mixed derivative of order d). The vector \mathbf{Y} could be discrete or continuous. In the former case, the joint probability mass function $f(\cdot; \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d, \boldsymbol{\theta})$ for \mathbf{Y} can be derived from the cdf in (2.1), and we let the univariate marginal probability mass functions be denoted by f_1, \dots, f_d ; in the latter case, we assume that F_j has density f_j for $j = 1, \dots, d$, and that \mathbf{Y} has density

$$f(\mathbf{y}; \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d, \boldsymbol{\theta}) = c(F_1(y_1; \boldsymbol{\alpha}_1), \dots, F_d(y_d; \boldsymbol{\alpha}_d); \boldsymbol{\theta}) \prod_{j=1}^d f_j(y_j; \boldsymbol{\alpha}_j).$$

For a sample of size n , with observed random vectors $\mathbf{y}_1, \dots, \mathbf{y}_n$, we can consider the d log-likelihood functions for the univariate margins,

$$L_j(\boldsymbol{\alpha}_j) = \sum_{i=1}^n \log f_j(y_{ij}; \boldsymbol{\alpha}_j), \quad j = 1, \dots, d,$$

and the log-likelihood function for the joint distribution,

$$L(\boldsymbol{\theta}, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d) = \sum_{i=1}^n \log f(\mathbf{y}_i; \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d, \boldsymbol{\theta}). \quad (2.2)$$

A simple case of the IFM method consists of doing d separate optimizations of the univariate likelihoods, followed by an optimization of the multivariate likelihood as a function of the dependence parameter vector. More specifically,

- a. the log-likelihoods L_j of the d univariate margins are separately maximized to get estimates $\tilde{\boldsymbol{\alpha}}_1, \dots, \tilde{\boldsymbol{\alpha}}_d$;
- b. the function $L(\boldsymbol{\theta}, \tilde{\boldsymbol{\alpha}}_1, \dots, \tilde{\boldsymbol{\alpha}}_d)$ is maximized over $\boldsymbol{\theta}$ to get $\tilde{\boldsymbol{\theta}}$.

That is, under regularity conditions, $(\tilde{\boldsymbol{\alpha}}_1, \dots, \tilde{\boldsymbol{\alpha}}_d, \tilde{\boldsymbol{\theta}})$ is the solution of

$$(\partial L_1 / \partial \boldsymbol{\alpha}_1, \dots, \partial L_d / \partial \boldsymbol{\alpha}_d, \partial L / \partial \boldsymbol{\theta}) = \mathbf{0}. \quad (2.3)$$

This procedure is computationally simpler than estimating all parameters $\alpha_1, \dots, \alpha_d, \theta$ simultaneously from L in (2.2). A numerical optimization with many parameters is much more time-consuming compared with several numerical optimizations, each with fewer parameters.

If the copula model (2.1) has further structure such as a parameter associated with each bivariate margin, simplifications of the second step (b) can be made, so that no numerical optimization with a large number of parameters is needed. For example, with a multivariate normal latent distribution, there is a simplification as shown in the example given later in this section; the correlation parameters can be estimated from separate likelihoods of the bivariate margins.

If it is possible to maximize L to get estimates $\hat{\alpha}_1, \dots, \hat{\alpha}_d, \hat{\theta}$, then one could compare these with the estimates $\tilde{\alpha}_1, \dots, \tilde{\alpha}_d, \tilde{\theta}$ as an estimation consistency check to evaluate the adequacy of the copula. For comparison with the IFM method, the maximum likelihood estimate (MLE) refers to $(\hat{\alpha}_1, \dots, \hat{\alpha}_d, \hat{\theta})$. Under regularity conditions, this comes from solving

$$(\partial L / \partial \alpha_1, \dots, \partial L / \partial \alpha_d, \partial L / \partial \theta) = \mathbf{0};$$

contrast with (2.3). Note that for multivariate normal (MVN) distributions, consisting of the MVN copula with correlation matrix $\theta = R$ and $N(\mu_j, \sigma_j^2)$ univariate margins $[\alpha_j = (\mu_j, \sigma_j^2)]$, that $\hat{\alpha}_j = \tilde{\alpha}_j$, $j = 1, \dots, d$, and $\hat{\theta} = \tilde{\theta}$. The equivalence of the estimators generally does not hold. Possibly because the MVN distribution is dominant in multivariate statistics, attention has not been given to variations of maximum likelihood estimation for multivariate models.

Since it is computationally easier to obtain $(\tilde{\alpha}_1, \dots, \tilde{\alpha}_d, \tilde{\theta})$, a natural question is its asymptotical efficiency compared with $(\hat{\alpha}_1, \dots, \hat{\alpha}_d, \hat{\theta})$. Outside of efficiency considerations, the former set of estimates provide a good starting point for the latter if one needs and can get $(\hat{\alpha}_1, \dots, \hat{\alpha}_d, \hat{\theta})$. Approximations leading to the asymptotic distribution of $(\tilde{\alpha}_1, \dots, \tilde{\alpha}_d, \tilde{\theta})$ are given in Section 4. From this, one can (numerically) compare the asymptotic covariance matrices of $(\tilde{\alpha}_1, \dots, \tilde{\alpha}_d, \tilde{\theta})$ and $(\hat{\alpha}_1, \dots, \hat{\alpha}_d, \hat{\theta})$. Also an estimate of the asymptotic covariance matrix of $(\tilde{\alpha}_1, \dots, \tilde{\alpha}_d, \tilde{\theta})$ can be obtained. The theory is a special case of using a set of estimating equations to estimate a vector of parameters. Some results on the efficiency for some specific models are given in Section 5.

Extensions to covariates is straightforward if one allows some or all of the parameters to be functions of the covariates. There are standard ways to handle the inclusion of covariates for univariate parameters, but not so for the inclusion of covariates for dependence parameters. As before, one has the likelihoods L_1, \dots, L_d and L but they are functions of more parameters, such as regression parameters for the covariates. Results for asymptotics of the parameter estimates with covariates are given in Section 4. This is less straightforward than the independent and identically distributed (iid) case.

We next give some examples to make the above discussion more concrete, and then conclude the section with the use of the jackknife method for estimation of standard errors.

Example 2.1. (Multivariate probit model with no covariates.) The multivariate probit model for the multivariate binary response vector \mathbf{Y} has stochastic representation: $Y_j = I(Z_j \leq \alpha_j)$, $j = 1, \dots, d$, where $\mathbf{Z} \sim N_d(\mathbf{0}, R)$, and $\boldsymbol{\theta} = R = (\rho_{jk})$ is a correlation matrix. Let the data be $\mathbf{y}_i = (y_{i1}, \dots, y_{id})$, $i = 1, \dots, n$. For $j = 1, \dots, d$, let $N_j(0)$ and $N_j(1)$ be the number of 0's and 1's among the y_{ij} 's. For $1 \leq j < k \leq d$, let $N_{jk}(i_1, i_2)$ be the frequency of (i_1, i_2) among the pairs (y_{ij}, y_{ik}) , for (i_1, i_2) equal $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$. From the j th univariate likelihood

$$L_j^*(\alpha_j) = [\Phi(\alpha_j)]^{N_j(1)} [1 - \Phi(\alpha_j)]^{N_j(0)},$$

$\tilde{\alpha}_j = \Phi^{-1}(N_j(1)/n)$, $j = 1, \dots, d$. For this example, one can estimate the dependence parameters from separate bivariate likelihoods $L_{jk}^*(\rho_{jk}) = L_{jk}^*(\rho_{jk}, \alpha_j, \alpha_k)$ rather than using the d -variate log-likelihood L in (2.2). Let Φ_ρ be the bivariate standard normal cdf with correlation ρ . Then

$$\begin{aligned} L_{jk}^*(\rho_{jk}, \alpha_j, \alpha_k) &= [\Phi_{\rho_{jk}}(\alpha_j, \alpha_k)]^{N_{jk}(11)} [\Phi(\alpha_j) - \Phi_{\rho_{jk}}(\alpha_j, \alpha_k)]^{N_{jk}(10)} [\Phi(\alpha_k) - \Phi_{\rho_{jk}}(\alpha_j, \alpha_k)]^{N_{jk}(01)} \\ &\quad \cdot [1 - \Phi(\alpha_j) - \Phi(\alpha_k) + \Phi_{\rho_{jk}}(\alpha_j, \alpha_k)]^{N_{jk}(00)}, \end{aligned}$$

and $\tilde{\rho}_{jk}$ is the root ρ of $N_{jk}(11)/n = \Phi_\rho(\tilde{\alpha}_j, \tilde{\alpha}_k)$. \square

Example 2.2. (Multivariate probit model with covariates.) Let the data be $(\mathbf{Y}_i, \mathbf{x}_i)$, $i = 1, \dots, n$, where \mathbf{Y}_i is a d -variate response vector and \mathbf{x}_i is a covariate vector. The multivariate probit model has stochastic representation $Y_j = I(Z_{ij} \leq \beta_{j0} + \boldsymbol{\beta}_j \mathbf{x}_i^T)$, $j = 1, \dots, d$, $i = 1, \dots, n$, where $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{id})$ are iid with distribution $N_d(\mathbf{0}, R)$, and $R = (\rho_{jk})$. In the usual multivariate probit model, the dependence parameters ρ_{jk} are not functions of covariates. However the IFM method works if they are (compatible) functions of the covariates. With $\boldsymbol{\gamma}_j = (\beta_{j0}, \boldsymbol{\beta}_j)$, the j th univariate likelihood is

$$L_j^*(\boldsymbol{\gamma}_j) = \prod_{i=1}^n [\Phi(\beta_{j0} + \boldsymbol{\beta}_j \mathbf{x}_i^T)]^{y_{ij}} [1 - \Phi(\beta_{j0} + \boldsymbol{\beta}_j \mathbf{x}_i^T)]^{1-y_{ij}}.$$

For $1 \leq j < k \leq d$, the (j, k) bivariate likelihood is:

$$\begin{aligned} L_{jk}^*(\rho_{jk}, \boldsymbol{\gamma}_j, \boldsymbol{\gamma}_k) &= \prod_{i=1}^n \{ [\Phi_{\rho_{jk}}(\beta_{j0} + \boldsymbol{\beta}_j \mathbf{x}_i^T, \beta_{k0} + \boldsymbol{\beta}_k \mathbf{x}_i^T)]^{I(y_{ij}=y_{ik}=1)} \\ &\quad \cdot [\Phi(\beta_{j0} + \boldsymbol{\beta}_j \mathbf{x}_i^T) - \Phi_{\rho_{jk}}(\beta_{j0} + \boldsymbol{\beta}_j \mathbf{x}_i^T, \beta_{k0} + \boldsymbol{\beta}_k \mathbf{x}_i^T)]^{I(y_{ij}=1, y_{ik}=0)} \\ &\quad \cdot [\Phi(\beta_{k0} + \boldsymbol{\beta}_k \mathbf{x}_i^T) - \Phi_{\rho_{jk}}(\beta_{j0} + \boldsymbol{\beta}_j \mathbf{x}_i^T, \beta_{k0} + \boldsymbol{\beta}_k \mathbf{x}_i^T)]^{I(y_{ij}=0, y_{ik}=1)} \\ &\quad \cdot [1 - \Phi(\beta_{j0} + \boldsymbol{\beta}_j \mathbf{x}_i^T) - \Phi(\beta_{k0} + \boldsymbol{\beta}_k \mathbf{x}_i^T) + \Phi_{\rho_{jk}}(\beta_{j0} + \boldsymbol{\beta}_j \mathbf{x}_i^T, \beta_{k0} + \boldsymbol{\beta}_k \mathbf{x}_i^T)]^{I(y_{ij}=y_{ik}=0)} \}. \end{aligned}$$

For $j = 1, \dots, d$, let $\tilde{\boldsymbol{\gamma}}_j$ be the IFM estimate from maximizing L_j^* . Then for $1 \leq j < k \leq d$, $\tilde{\rho}_{jk}$ is the IFM estimate from maximizing $L_{jk}^*(\rho_{jk}, \tilde{\boldsymbol{\gamma}}_j, \tilde{\boldsymbol{\gamma}}_k)$ as a function of ρ_{jk} . \square

Examples 2.1 and 2.2 extend to multivariate probit models for ordinal data, in which there are more univariate cutoff parameters. With (Z_1, \dots, Z_d) having $N_d(\mathbf{0}, R)$ distribution, the stochastic representation for (Y_1, \dots, Y_d) is: $Y_j = m$ if $\beta_{j,m-1} < Z_j \leq \beta_{j,m}$, $m = 1, \dots, r_j$, r_j is the number of categories of the j th variable, $j = 1, \dots, d$. (Without loss of generality, assume $\beta_{j,0} = -\infty$ and $\beta_{j,r_j} = \infty$ for all j .) There are $\sum_{j=1}^d (r_j - 1)$ univariate parameters or cutoff points. If there is a covariate vector \mathbf{x} , then the parameters $\beta_{j,m}$ and ρ_{jk} can depend on \mathbf{x} , with constraints such as $\beta_{j,m-1}(\mathbf{x}) < \beta_{j,m}(\mathbf{x})$.

Example 2.3. (Multivariate Poisson–lognormal distribution.) Suppose we have a random sample of iid random vectors \mathbf{y}_i , $i = 1, \dots, n$, from the probability mass function:

$$f(\mathbf{y}; \boldsymbol{\mu}, \Sigma) = \int_{[0, \infty)^d} \prod_{j=1}^d p(y_j; \lambda_j) g(\boldsymbol{\lambda}; \boldsymbol{\mu}, \Sigma) d\lambda_1 \cdots d\lambda_d, \quad y_j = 0, 1, 2, \dots,$$

where $p(y; \lambda) = e^{-\lambda} \lambda^y / y!$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$, $\log \boldsymbol{\lambda} = (\log \lambda_1, \dots, \log \lambda_d)$,

$$g(\boldsymbol{\lambda}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-d/2} (\lambda_1 \cdots \lambda_d)^{-1} |\Sigma|^{-1/2} \exp\{-0.5(\log \boldsymbol{\lambda} - \boldsymbol{\mu}) \Sigma^{-1} (\log \boldsymbol{\lambda} - \boldsymbol{\mu})^T\},$$

$\lambda_j > 0$, $j = 1, \dots, d$, is a d -variate lognormal density, with mean vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)$ and covariance matrix $\Sigma = (\sigma_{ij})$. This model uses the multivariate normal distribution for a latent vector, so that parameters can be estimated from log-likelihoods of univariate and bivariate margins, after reparametrizing Σ into the vector of variances $(\sigma_{11}, \dots, \sigma_{dd})$ and a correlation matrix $R = (\rho_{jk})$. Let $\boldsymbol{\alpha}_j = (\mu_j, \sigma_{jj})$, $j = 1, \dots, d$.

The j th univariate marginal density is

$$f_j(y_j; \boldsymbol{\alpha}_j) = \int_0^\infty \frac{e^{-\lambda_j} \lambda_j^{y_j}}{y_j! \sqrt{2\pi \sigma_{jj} \lambda_j}} \exp\{-0.5(\log \lambda_j - \mu_j)^2 / \sigma_{jj}\} d\lambda_j.$$

The (j, k) bivariate marginal density is

$$\begin{aligned} f_{jk}(y_j, y_k; \boldsymbol{\alpha}_j, \boldsymbol{\alpha}_k, \rho_{jk}) &= \int_0^\infty \int_0^\infty \frac{e^{-\lambda_j} \lambda_j^{y_j} e^{-\lambda_k} \lambda_k^{y_k}}{y_j! y_k! 2\pi \lambda_j \lambda_k \sqrt{\sigma_{jj} \sigma_{kk} - \sigma_{jk}^2}} \\ &\cdot \exp\{-0.5(1 - \rho_{jk}^2)^{-1}[(\log \lambda_j - \mu_j)^2 / \sigma_{jj} + (\log \lambda_k - \mu_k)^2 / \sigma_{kk} \\ &\quad - 2\rho_{jk}(\log \lambda_j - \mu_j)(\log \lambda_k - \mu_k) / \sqrt{\sigma_{jj} \sigma_{kk}}]\} d\lambda_j d\lambda_k, \end{aligned}$$

where $\rho_{jk} = \sigma_{jk} / \sqrt{\sigma_{jj} \sigma_{kk}}$. Hence $\tilde{\boldsymbol{\alpha}}_j$ obtains from maximizing $L_j(\boldsymbol{\alpha}_j) = \sum_{i=1}^n \log f_j(y_{ij}; \boldsymbol{\alpha}_j)$, and $\tilde{\rho}_{jk}$ obtains from maximizing $L_{jk}(\rho_{jk}, \tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\alpha}}_k) = \sum_{i=1}^n \log f_{jk}(y_{ij}, y_{ik}; \tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\alpha}}_k, \rho_{jk})$ as a function of ρ_{jk} .

A simple extension to include covariates is to make the parameters μ_j linear in the covariates for $j = 1, \dots, d$. \square

Example 2.4. (Multivariate extreme value models.) In this example, we indicate how the IFM method applies to some parametric multivariate extreme value models. Other than one simple exchangeable dependence copula model, we do not state multivariate extreme value models here in order to save space (see Joe

1994 and references therein for details). The multivariate extreme value models have generalized extreme value distributions as univariate margins, that is,

$$F_j(y_j; \boldsymbol{\alpha}_j) = \exp\{-(1 + \gamma_j[y_j - \mu_j]/\sigma_j)_+^{-1/\gamma_j}\}, \quad -\infty < x_j < \infty, \quad -\infty < \gamma_j, \mu_j < \infty, \sigma_j > 0,$$

where $\boldsymbol{\alpha}_j = (\gamma_j, \mu_j, \sigma_j)$ and $x_+ = \max\{0, x\}$. In Tawn (1990) and Joe (1994), the initial data analysis consisted of fitting the univariate margins separately to get $\tilde{\boldsymbol{\alpha}}_j$, $j = 1, \dots, d$, followed by comparison of the models for the dependence structure with the univariate margins fixed at the estimated parameter values $\tilde{\boldsymbol{\alpha}}_j$. One way to compare various models is from the AIC (Akaike information criterion) values associated with the log-likelihoods $L(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\alpha}}_1, \dots, \tilde{\boldsymbol{\alpha}}_d)$. However if d is large, one can estimate the dependence parameters in $\boldsymbol{\theta}$ more easily with the IFM method.

The sequence of log-likelihoods of marginal distributions is indicated for several families. For a multivariate extreme value copula C with one (exchangeable) dependence parameter θ , such as

$$C(u_1, \dots, u_d; \theta) = \exp\{-[\sum_j (-\log u_j)^\theta]^{1/\theta}\}, \quad \theta \geq 1,$$

one can estimate θ from maximizing $L(\theta, \tilde{\boldsymbol{\alpha}}_1, \dots, \tilde{\boldsymbol{\alpha}}_d)$ as a function of θ . For other copulas with more dependence parameters, one can use log-likelihoods of marginal distributions. Examples are:

1. The Hüsler and Reiss (1989) model, which derives from a non-standard extreme value limit of the multivariate normal distribution, has a dependence parameter associated with each bivariate margin, and it is closed under taking of margins. That is, $\boldsymbol{\theta} = (\theta_{jk} : 1 \leq j < k \leq d)$. For $1 \leq j < k \leq d$, the parameter θ_{jk} can be estimated by maximizing the (j, k) bivariate log-likelihood $L_{jk}(\theta_{jk}, \tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\alpha}}_k)$ as a function of θ_{jk} .
2. Joe (1994) has a class of models which have a dependence parameter associated with each bivariate margin, but some parameters are interpreted as conditional dependence parameters. In the d -dimensional case, the estimation of $\boldsymbol{\theta} = (\theta_{jk} : 1 \leq j < k \leq d)$ is most easily obtained from maximizing the following sequence of bivariate log-likelihoods, each in one parameter:

$$\begin{aligned} L_{j,j+1}(\theta_{j,j+1}, \tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\alpha}}_{j+1}), \quad j = 1, \dots, d-1, \\ L_{j,j+2}(\theta_{j,j+2}, \tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\alpha}}_{j+2}, \tilde{\theta}_{j,j+1}, \tilde{\theta}_{j+1,j+2}), \quad j = 1, \dots, d-2, \\ \dots \\ L_{1d}(\theta_{1,d}, \tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_d, \tilde{\theta}_{jk}, (j, k) \neq (1, d)). \end{aligned}$$

□

There are many other models for which the IFM method can be used (Xu 1996). However the above examples should illustrate the general usefulness for parametric multivariate models that have some closure properties of margins and nice dependence interpretations for the multivariate parameters. Two examples with data sets that have appeared in the literature are given in Section 6.

Under some regularity conditions, the parameter estimate $\tilde{\boldsymbol{\eta}} = (\tilde{\boldsymbol{\alpha}}_1, \dots, \tilde{\boldsymbol{\alpha}}_d, \tilde{\boldsymbol{\theta}})$ can be shown to be asymptotically multivariate normal, using the theory of inference functions or estimating equations (Godambe 1960, 1976, 1991). The asymptotic covariance matrix $n^{-1}V$ involves derivatives of the inference functions (which for us are score functions from some log-likelihood functions). To avoid the tedious taking and coding of derivatives (especially if symbolic manipulation software cannot be used), we propose the use of the jackknife method to estimate the asymptotic covariance matrix and to obtain standard errors for functions of the parameters. (These functions include probabilities of being in some category or probabilities of exceedances). Hence with the jackknife and the use of the quasi-Newton method for numerical optimization, only the log-likelihood functions of some marginal distributions must be coded.

We present the jackknife in the case of no covariates. However it extends easily to the case of covariates. For the delete-one jackknife, let $\tilde{\boldsymbol{\eta}}^{(i)}$ be the estimator of $\boldsymbol{\eta} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d, \boldsymbol{\theta})$ with the i th observation \mathbf{Y}_i deleted, $i = 1, \dots, n$. Assuming $\tilde{\boldsymbol{\eta}}$ and $\tilde{\boldsymbol{\eta}}^{(i)}$'s are row vectors, the jackknife estimator of $n^{-1}V$ is

$$\sum_{i=1}^n (\tilde{\boldsymbol{\eta}}^{(i)} - \tilde{\boldsymbol{\eta}})^T (\tilde{\boldsymbol{\eta}}^{(i)} - \tilde{\boldsymbol{\eta}}).$$

For large samples in which the delete-one jackknife would be computationally too time-consuming, the jackknife can be modified into estimates from deletions of more than one observation at a time. Suppose $n = gm$, with g groups or blocks of m ; g estimators can be obtained with the k th estimate based on $n - m$ observations after deleting the m observations in the k th block. (It is probably best to randomize the n observations into the m blocks. The disadvantage is that the jackknife estimates depends on the randomization. However reasonable estimates for standard errors should obtain and there are approximations in estimating standard errors for any method.) Now let $\tilde{\boldsymbol{\eta}}^{(k)}$ be the estimator of $\boldsymbol{\eta}$ with the k th block deleted, $k = 1, \dots, g$. We think of m as fixed with $g \rightarrow \infty$. The jackknife estimator of $n^{-1}V$, which is asymptotically consistent, is

$$V_{n,g} = \sum_{k=1}^g (\tilde{\boldsymbol{\eta}}^{(k)} - \tilde{\boldsymbol{\eta}})^T (\tilde{\boldsymbol{\eta}}^{(k)} - \tilde{\boldsymbol{\eta}}). \quad (2.5)$$

The jackknife method can also be used for estimates of functions of parameters. The delta or Taylor method requires partial derivatives (of the function with respect to the parameters) and the jackknife method eliminates the need for this. As above, let $\tilde{\boldsymbol{\eta}}^{(k)}$ be the estimator of $\boldsymbol{\eta}$ with the k th block deleted, $k = 1, \dots, g$, and let $\tilde{\boldsymbol{\eta}}$ be the estimator based on the entire sample. Let $b(\boldsymbol{\eta})$ be a (real-valued) quantity of interest. From each subsample, compute an estimate of $b(\boldsymbol{\eta})$; i.e., $b(\tilde{\boldsymbol{\eta}}^{(k)})$, $k = 1, \dots, g$, and $b(\tilde{\boldsymbol{\eta}})$ are obtained. The jackknife

estimate of the standard error of $b(\tilde{\boldsymbol{\eta}})$ is

$$[V_{n,g}(b)]^{1/2} = \left\{ \sum_{k=1}^g [b(\tilde{\boldsymbol{\eta}}^{(k)}) - b(\tilde{\boldsymbol{\eta}})]^2 \right\}^{1/2}. \quad (2.6)$$

What the above means for the computing sequence for the jackknife method is that one should maintain a table of the parameter estimates for the full sample and each jackknife subsample. Then one can use this table for computing estimates of one or more functions of the parameters, together with the corresponding standard errors.

The use of the jackknife for functions of parameters also gets around the problem of the ‘optimal’ parametrization for speed of convergence to asymptotic normality (and hence estimate of the asymptotic covariance matrix for the delta method).

3. Some advantages of IFM method. In this brief section, we summarize some advantages of the IFM method. In the discussion below, the ML method refers to maximum likelihood of all parameters simultaneously, as in (2.2).

1. The IFM method makes inference for many multivariate models computationally feasible.
2. More statistical models exist for univariate and bivariate distributions, so that it allows one to do inference and modelling starting with univariate and lower-dimensional margins.
3. It allows one to compare models for the dependence structure and do a sensitivity analysis of the models for predictions and inferences. There is some robustness against misspecification of the dependence structure. Also there should be more robustness against outliers or perturbations of the data, compared with the ML method.
4. Sparse multivariate data can create problems for ML method. But the IFM method avoids the sparseness problem to a certain degree, especially if parameters can all be estimated from univariate and bivariate likelihoods; this could be a major advantage in a smaller sample situation.

4. Summary of asymptotic results.

In this section, we present asymptotic results for the case of iid response vectors with no covariates and then look at two approaches for the extension to include covariates. In the last subsection, results on the consistency of the jackknife are given.

When we refer to a density, it means a density with respect to an appropriate measure space. The density is a probability mass function in the discrete case, but we use a common notation and terminology to cover both the cases of continuous and discrete response vectors.

4.1. Independent and identically distributed case.

Throughout this subsection, we assume that the usual regularity conditions (see for example, Serfling 1980) for asymptotic maximum likelihood theory hold for the multivariate model as well as all of its margins. A unification of the variations of the IFM method is through inference or estimating functions in which each function is a score function or the (partial) derivative of a log-likelihood of some marginal density. Let $\boldsymbol{\eta} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d, \boldsymbol{\theta})$ be the row vector of parameters and let $\boldsymbol{\psi}$ be a row vector of inference functions of the same dimension as $\boldsymbol{\eta}$.

Let $\mathbf{Y}, \mathbf{Y}_1, \dots, \mathbf{Y}_n$ be iid with the density $f(\cdot; \boldsymbol{\eta})$. Suppose the row vector of estimating equations for the estimator $\tilde{\boldsymbol{\eta}}$ is

$$\sum_{i=1}^n \boldsymbol{\psi}(\mathbf{Y}_i, \tilde{\boldsymbol{\eta}}) = \mathbf{0}. \quad (4.1)$$

Let $\partial \boldsymbol{\psi}^T / \partial \boldsymbol{\eta}$ be the matrix with (j, k) component $\partial \psi_j(\mathbf{y}, \boldsymbol{\eta}) / \partial \eta_k$, where ψ_j is the j th component of $\boldsymbol{\psi}$ and η_k is the k th component of $\boldsymbol{\eta}$. From an expansion of (4.1) similar to the derivation of the asymptotic distribution of a MLE, under the regularity conditions, the asymptotic distribution of $n^{1/2}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta})^T$ is equivalent to that of

$$\{-E[\partial \boldsymbol{\psi}^T(\mathbf{Y}, \boldsymbol{\eta}) / \partial \boldsymbol{\eta}]\}^{-1} n^{-1/2} \sum_{i=1}^n \boldsymbol{\psi}^T(Y_i, \boldsymbol{\eta}).$$

Hence it has the same asymptotic distribution as $\{-E[\partial \boldsymbol{\psi}^T(\mathbf{Y}, \boldsymbol{\eta}) / \partial \boldsymbol{\eta}]\}^{-1} \mathbf{Z}^T$, where $\mathbf{Z} \sim N(\mathbf{0}, \text{Cov}(\boldsymbol{\psi}(\mathbf{Y}; \boldsymbol{\eta})))$. That is, the asymptotic covariance matrix of $n^{1/2}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta})^T$, called the inverse Godambe information matrix, is $V = D_{\boldsymbol{\psi}}^{-1} M_{\boldsymbol{\psi}} (D_{\boldsymbol{\psi}}^{-1})^T$, where $D_{\boldsymbol{\psi}} = E[\partial \boldsymbol{\psi}^T(\mathbf{Y}, \boldsymbol{\eta}) / \partial \boldsymbol{\eta}]$, $M_{\boldsymbol{\psi}} = E[\boldsymbol{\psi}^T(\mathbf{Y}, \boldsymbol{\eta}) \boldsymbol{\psi}(\mathbf{Y}, \boldsymbol{\eta})]$.

4.2. Inclusion of covariates: approach 1

Now, we assume that we have independent, non-identically distributed random vectors \mathbf{Y}_i , $i = 1, \dots, n$, with \mathbf{Y}_i having density $f(\cdot; \boldsymbol{\eta}_i)$, $\boldsymbol{\eta}_i = \boldsymbol{\eta}(\mathbf{x}_i, \boldsymbol{\gamma})$ for a function $\boldsymbol{\eta}$ and a parameter function $\boldsymbol{\gamma}$. The check of the necessary conditions depends somewhat on the specific models. However we indicate the general types of conditions that must hold.

We assume that each component of $\boldsymbol{\eta} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d, \boldsymbol{\theta})$ is a function of \mathbf{x} , more specifically, $\boldsymbol{\alpha}_j = \mathbf{a}_j(\mathbf{x}, \boldsymbol{\gamma}_j)$, $j = 1, \dots, d$, and $\boldsymbol{\theta} = \mathbf{t}(\mathbf{x}, \boldsymbol{\gamma}_{d+1})$, with $\mathbf{a}_1, \dots, \mathbf{a}_d, \mathbf{t}$ having components that are each functions of linear combinations of the functions of the components of \mathbf{x} (that is, there are link functions linking \mathbf{x} to the parameters). We assume that the inference function vector $\boldsymbol{\psi}$ has a component for each parameter in $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_d, \boldsymbol{\gamma}_{d+1})$.

We explain the notation here for Examples 2.1 and 2.2. With no covariates, $\boldsymbol{\eta} = (\alpha_1, \dots, \alpha_d, \boldsymbol{\theta})$, where α_j is the cutoff point for the j th univariate margin, and $\boldsymbol{\theta} = R = (\rho_{jk})$ is a correlation matrix. With covariates, $\alpha_j = a_j(\mathbf{x}, \beta_{j0}, \boldsymbol{\beta}_j) = \beta_{j0} + \boldsymbol{\beta}_j \mathbf{x}^T$, $j = 1, \dots, d$, and $t(\mathbf{x}, \boldsymbol{\theta}) = \boldsymbol{\theta}$, so that $\boldsymbol{\gamma} = (\beta_{10}, \boldsymbol{\beta}_1, \dots, \beta_{d0}, \boldsymbol{\beta}_d, \boldsymbol{\theta})$ with $\boldsymbol{\gamma}_j = (\beta_{j0}, \boldsymbol{\beta}_j)$ and $\boldsymbol{\gamma}_{d+1} = \boldsymbol{\theta}$.

In place of $f(\mathbf{y}; \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d, \boldsymbol{\theta})$ and $f_j(y_j, \boldsymbol{\alpha}_j)$ in the case of no covariates, we now have the densities:

$$f_{\mathbf{Y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}; \boldsymbol{\gamma}) \stackrel{\text{def}}{=} f(\mathbf{y}; \mathbf{a}_1(\mathbf{x}, \boldsymbol{\gamma}_1), \dots, \mathbf{a}_d(\mathbf{x}, \boldsymbol{\gamma}_d), \mathbf{t}(\mathbf{x}, \boldsymbol{\gamma}_{d+1}))$$

and

$$f_{Y_j|\mathbf{x}}(y_j|\mathbf{x}; \boldsymbol{\gamma}_j) \stackrel{\text{def}}{=} f_j(y_j; \mathbf{a}_j(\mathbf{x}, \boldsymbol{\gamma}_j)), \quad j = 1, \dots, d.$$

In a simple case, the estimate $\tilde{\boldsymbol{\gamma}}$ from the IFM method has component $\tilde{\boldsymbol{\gamma}}_j$ coming from the maximization of

$$L_j(\boldsymbol{\gamma}_j) = \sum_{i=1}^n \log f_{Y_j|\mathbf{x}}(y_{ij}|\mathbf{x}_i; \boldsymbol{\gamma}_j), \quad (4.2)$$

$j = 1, \dots, d$, and $\tilde{\boldsymbol{\gamma}}_{d+1}$ comes from the maximization of $L(\tilde{\boldsymbol{\gamma}}_1, \dots, \tilde{\boldsymbol{\gamma}}_d, \boldsymbol{\gamma}_{d+1})$ in $\boldsymbol{\gamma}_{d+1}$, where

$$L(\boldsymbol{\gamma}) = \sum_{i=1}^n \log f_{\mathbf{Y}|\mathbf{x}}(\mathbf{y}_i|\mathbf{x}_i; \boldsymbol{\gamma}). \quad (4.3)$$

Alternatively, the components of $\boldsymbol{\gamma}_{d+1}$ may be estimated from log-likelihoods of lower-dimensional margins such as in Example 2.2. In any case, let $\boldsymbol{\psi}$ be the vector of inference functions from partial derivatives of log-likelihood functions of margins.

Conditions for the preceding asymptotic results to hold are of the following sense (we do not take up space to write out the numerous equations):

- a. mixed derivatives of $\boldsymbol{\psi}$ of first and second order are dominated by integrable functions;
- b. products of these derivatives are uniformly integrable;
- c. the link functions are twice continuously differentiable with first and second order derivatives bounded away from zero;
- d. covariates are uniformly bounded, the sample covariance matrix of the covariates \mathbf{x}_i is strictly positive definite;
- e. a Lindeberg–Feller type condition holds.

References for these types of conditions and proofs of asymptotic normality are in, for example, Bradley and Gart (1962) and Hoadley (1971).

Assuming that the conditions hold, then the asymptotic normality result has the form:

$$n^{-1/2} V_n^{-1/2} (\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^T \rightarrow_d N(\mathbf{0}, I),$$

where $V_n = D_n^{-1} M_n (D_n^{-1})^T$ with

$$D_n = n^{-1} \sum_{i=1}^n \mathbb{E} [\partial \boldsymbol{\psi}^T(\mathbf{Y}_i, \boldsymbol{\gamma}) / \partial \boldsymbol{\gamma}] \quad \text{and} \quad M_n = n^{-1} \sum_{i=1}^n \mathbb{E} [\boldsymbol{\psi}^T(\mathbf{Y}_i, \boldsymbol{\gamma}) \boldsymbol{\psi}(\mathbf{Y}_i, \boldsymbol{\gamma})].$$

Details for the case of multivariate discrete models are given in Xu (1996); the results generalize to the continuous case when the assumptions hold.

4.3. Inclusion of covariates: approach 2.

A second approach for asymptotics allows for parameters to be more general functions of the covariates, and treats the covariates as realizations of random vectors. This approach assumes a joint distribution for response vector and covariates, with the parameters for the marginal distribution of the covariate vector treated as nuisance parameters. This assumption might be reasonable for a random sample of subjects in which \mathbf{x}_i and \mathbf{Y}_i are observed together (no control of \mathbf{x}_i before observing \mathbf{Y}_i).

Similar to the preceding subsection, we write the conditional density

$$f_{\mathbf{Y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}; \boldsymbol{\gamma}) = f(\mathbf{y}; \boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\gamma})).$$

Let $\mathbf{Z}_i = (\mathbf{Y}_i, \mathbf{x}_i)$, $i = 1, \dots, n$. These are treated as iid random vectors from the density

$$f_{\mathbf{Z}}(\mathbf{z}; \boldsymbol{\gamma}) = f_{\mathbf{Y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}; \boldsymbol{\gamma}) \cdot f_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\omega}). \quad (4.4)$$

For inference, we are interested in $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}(\mathbf{x}, \boldsymbol{\gamma})$, and not in $\boldsymbol{\omega}$. Marginal distributions of (4.4) for the IFM method are:

$$f_{Y_j|\mathbf{x}}(y_j|\mathbf{x}; \boldsymbol{\gamma}_j) \cdot f_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\omega}), \quad j = 1, \dots, d.$$

If $\boldsymbol{\omega}$ is treated as a nuisance parameter, then the log-likelihood in $\boldsymbol{\gamma}$ from (4.5) below is essentially the same as that in Approach 1.

Let $\boldsymbol{\gamma}, \boldsymbol{\alpha}_j, \mathbf{a}_j, \boldsymbol{\theta}, \mathbf{t}$ be the same as in the preceding subsection, except that \mathbf{a}_j , $j = 1, \dots, d$, and \mathbf{t} could be more general functions of the covariate vector \mathbf{x} . The vector estimate from the IFM method has component $\tilde{\boldsymbol{\gamma}}_j$ coming from the maximization of

$$L_j(\boldsymbol{\gamma}_j) = \sum_{i=1}^n \log[f_{Y_j|\mathbf{x}}(y_{ij}|\mathbf{x}_i; \boldsymbol{\gamma}_j) f_{\mathbf{x}}(\mathbf{x}_i; \boldsymbol{\omega})], \quad (4.5)$$

$j = 1, \dots, d$, and $\tilde{\boldsymbol{\gamma}}_{d+1}$ coming from the maximization of $L(\tilde{\boldsymbol{\gamma}}_1, \dots, \tilde{\boldsymbol{\gamma}}_d, \boldsymbol{\gamma}_{d+1})$ in $\boldsymbol{\gamma}_{d+1}$, where

$$L(\boldsymbol{\gamma}) = \sum_{i=1}^n \log[f_{\mathbf{Y}|\mathbf{x}}(\mathbf{y}_i|\mathbf{x}_i; \boldsymbol{\gamma}) f_{\mathbf{x}}(\mathbf{x}_i; \boldsymbol{\omega})]. \quad (4.6)$$

Note that optimization of (4.2) and (4.5), and of (4.3) and (4.6) are the same. Alternatively, the components of $\boldsymbol{\gamma}_{d+1}$ may be estimated from log-likelihoods of lower-dimensional margins. In any case, let $\boldsymbol{\psi}$ be the vector of inference functions from partial derivatives of log-likelihoods functions of margins.

Assume that the standard regularity conditions hold for $f_{\mathbf{Z}}$ and its margins, then the asymptotic theory from the iid case holds for the estimates using the IFM method in subsection 4.1. The asymptotic normality result is

$$n^{-1/2}(\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^T \rightarrow_d N(\mathbf{0}, V),$$

where $V = D_{\boldsymbol{\psi}}^{-1} M_{\boldsymbol{\psi}} (D_{\boldsymbol{\psi}}^{-1})^T$, with $D_{\boldsymbol{\psi}} = E[\partial \boldsymbol{\psi}^T(\mathbf{Y}, \mathbf{x}, \boldsymbol{\gamma}) / \partial \boldsymbol{\gamma}]$, $M_{\boldsymbol{\psi}} = E[\boldsymbol{\psi}^T(\mathbf{Y}, \mathbf{x}, \boldsymbol{\gamma}) \boldsymbol{\psi}(\mathbf{Y}, \mathbf{x}, \boldsymbol{\gamma})]$.

This approach for asymptotics with covariates appears to be new. The asymptotic covariance matrices for $n^{-1/2}(\tilde{\gamma} - \gamma)$ are different in the two asymptotic approaches. However the inference functions are the same in the two approaches. The use of the empirical distribution function to estimate the inverse Godambe information matrix or the use of the jackknife would lead to the same standard error estimates in the two approaches.

4.4. Consistency of jackknife.

Under the assumptions of the subsection 4.1, the jackknife estimators in (2.5) and (2.6) are consistent as $n \rightarrow \infty$ or $g \rightarrow \infty$ with m fixed. That is, $nV_{n,g} - V \rightarrow_p 0$ and

$$nV_{n,g}(b) - \frac{\partial b}{\partial \boldsymbol{\eta}} V \frac{\partial b}{\partial \boldsymbol{\eta}^T} \rightarrow_p 0.$$

Under the assumptions of the subsections 4.2 and 4.3, the jackknife estimator with covariates which is similar to that in (2.5) and (2.6) is also consistent as $n \rightarrow \infty$ with m fixed.

Standard techniques are used to prove these results. Details are given in Xu (1996).

5. Efficiency and simulation results. Efficiency comparisons of the IFM method with the ML method is a difficult task because of the general intractability of the asymptotic covariance matrices, and computational time required to obtain the MLE. Nevertheless, comparisons of various types are made in Xu (1996) for a number of multivariate models. One comparison is that of the asymptotic covariance matrices of the MLE and the IFM estimator, over the range of the parameters, in cases where the terms of the covariances matrices can be obtained. More generally, Monte Carlo simulations are required to compare the two estimators. All of the comparisons that were done suggest that the IFM method is highly efficient. Intuitively, we expect the IFM method to be quite efficient because it depends heavily on maximum likelihood, albeit from likelihoods of marginal distributions.

This section summarizes one of the many efficiency comparisons that were made, and an assessment of the jackknife for standard error estimation.

In Xu (1996), simulation results are obtained for dimensions $d = 3, 4$, for the multivariate probit model for binary or ordinal data with covariates, and for other copula-based models for multivariate discrete data. In almost all of the simulations, the relative efficiency, as measured by the ratio of the mean square errors of the IFM estimator to the MLE is close to 1. Typical simulation summaries are given in Tables 1 and 2. Table 1 is from Example 2.1 with trivariate binary data ($d = 3$) with true parameters $\alpha_j = 0$, $j = 1, 2, 3$, and (i) $\rho_{jk} = 0.6$, $1 \leq j < k \leq 3$, and (ii) $\rho_{12} = \rho_{23} = 0.8$, $\rho_{13} = 0.64$. In the parametrization for the estimation and as shown in the Table 1, $\theta_{jk} = (\exp(\rho_{jk}) - 1)/(\exp(\rho_{jk}) + 1)$, $j < k$, were used. Table 2 is from Example 2.2; the same parameters as (i) and (ii) were used for the correlation matrix, but there is one covariate. The parameters used were: $\beta_{10} = 0.7$, $\beta_{20} = 0.5$, $\beta_{30} = 0.3$ and $\beta_{11} = \beta_{21} = \beta_{31} = 0.5$, with x_i iid $N(0, 1/4)$.

Table 1: Parameter estimates, square root mean square errors and efficiency ratios for the multivariate probit model for binary data; $d = 3$. True parameters $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

n	margin parameters	1 α_1	2 α_2	3 α_3	(1,2) θ_{12}	(1,3) θ_{13}	(2,3) θ_{23}
$\theta_{12} = \theta_{13} = \theta_{23} = 1.3863$							
100	IFM	0.003	-0.002	0.005	1.442	1.426	1.420
		(0.131)	(0.121)	(0.128)	(0.376)	(0.380)	(0.378)
	MLE	0.002	-0.003	0.004	1.441	1.426	1.420
		(0.131)	(0.121)	(0.128)	(0.376)	(0.380)	(0.378)
	r	0.998	0.999	0.999	0.999	0.999	0.999
1000	IFM	-0.0006	-0.0016	-0.0008	1.3924	1.3897	1.3906
		(0.040)	(0.038)	(0.039)	(0.114)	(0.114)	(0.113)
	MLE	-0.0018	-0.0028	-0.0019	1.3919	1.3893	1.3902
		(0.040)	(0.038)	(0.039)	(0.114)	(0.114)	(0.113)
	r	0.997	0.997	0.997	1.000	1.001	1.000
$\theta_{12} = \theta_{23} = 2.1972, \theta_{13} = 1.5163$							
100	IFM	0.0027	-0.0006	0.0003	2.2664	1.5571	2.2586
		(0.131)	(0.123)	(0.130)	(0.454)	(0.377)	(0.453)
	MLE	0.0015	-0.0020	-0.0012	2.2646	1.5552	2.2579
		(0.131)	(0.123)	(0.131)	(0.453)	(0.377)	(0.452)
	r	0.999	1.000	0.999	1.001	1.001	1.002
1000	IFM	-0.0006	-0.0001	-0.0005	2.2009	1.5174	2.2043
		(0.040)	(0.038)	(0.039)	(0.135)	(0.118)	(0.136)
	MLE	-0.0023	-0.0020	-0.0022	2.2003	1.5166	2.2036
		(0.040)	(0.038)	(0.039)	(0.135)	(0.118)	(0.137)
	r	0.996	1.000	0.996	0.999	1.000	1.000

Two sample sizes $n = 100$ and $n = 1000$ are used. The number of simulations in each case is 1000. For each case, the entries of the Tables 1 and 2 are: for the IFM estimates and then the MLE, the average and square root mean square error for each parameter, and the ratio r of the first square root mean square error to the second.

In Xu (1996), simulations were also used to assess the adequacy of the jackknife estimator of the standard error, with the comparison to the estimator of the standard error from the Godambe information matrix in some cases where the latter was possible to obtain. The conclusion is that the jackknife estimator does very well. A typical simulation summary is given below.

To simplify the computations, we use a trivariate probit model, and estimate only the dependence parameters $\theta_{12}, \theta_{13}, \theta_{23}$ from the bivariate likelihoods with the univariate parameters fixed. For the s th simulation, let the estimates be denoted as $\tilde{\theta}_{12}^{(s)}, \tilde{\theta}_{13}^{(s)}, \tilde{\theta}_{23}^{(s)}$. We then compute the jackknife estimate of variance (with g groups of size m such that $g \times m = n$) for $\tilde{\theta}_{12}^{(s)}, \tilde{\theta}_{13}^{(s)}, \tilde{\theta}_{23}^{(s)}$. Let these variance estimates be $v_{12}^{(s)}, v_{13}^{(s)}, v_{23}^{(s)}$ and let the asymptotic variance estimate of $\tilde{\theta}_{12}, \tilde{\theta}_{13}, \tilde{\theta}_{23}$ based on Godambe information matrix be v_{12}, v_{13}, v_{23} . We compare the following three variance estimates measures:

Table 2: Parameter estimates, square root mean square errors and efficiency ratios for the multivariate probit model for binary data with a continuous covariate; $d = 3$. True parameters are $\beta_{10} = 0.7$, $\beta_{20} = 0.5$, $\beta_{30} = 0.3$ and $\beta_{11} = \beta_{21} = \beta_{31} = 0.5$, and $x_i \sim N(0, 1/4)$.

n	margin parameters	1		2		3		(1,2)	(1,3)	(2,3)
		β_{10}	β_{11}	β_{20}	β_{21}	β_{30}	β_{31}	θ_{12}	θ_{13}	θ_{23}
$\theta_{12} = \theta_{13} = \theta_{23} = 1.3863$										
100	IFM	0.722 (0.136)	0.529 (0.326)	0.488 (0.144)	0.520 (0.278)	0.312 (0.137)	0.524 (0.310)	1.453 (0.398)	1.403 (0.402)	1.473 (0.401)
	MLE	0.722 (0.137)	0.532 (0.320)	0.486 (0.144)	0.519 (0.278)	0.311 (0.138)	0.522 (0.308)	1.458 (0.402)	1.407 (0.412)	1.476 (0.406)
	r	0.999	1.019	0.999	1.002	0.993	1.005	0.990	0.976	0.989
1000	IFM	0.704 (0.042)	0.495 (0.089)	0.501 (0.046)	0.504 (0.084)	0.306 (0.041)	0.504 (0.093)	1.413 (0.140)	1.380 (0.109)	1.391 (0.124)
	MLE	0.703 (0.042)	0.494 (0.090)	0.500 (0.045)	0.503 (0.084)	0.305 (0.040)	0.503 (0.093)	1.415 (0.139)	1.381 (0.109)	1.393 (0.124)
	r	1.004	0.988	1.004	0.993	1.007	1.001	1.000	1.006	1.000
$\theta_{12} = \theta_{23} = 2.1972, \theta_{13} = 1.5163$										
100	IFM	0.722 (0.136)	0.529 (0.326)	0.490 (0.133)	0.543 (0.272)	0.300 (0.131)	0.544 (0.309)	2.303 (0.494)	1.556 (0.362)	2.303 (0.525)
	MLE	0.721 (0.136)	0.532 (0.317)	0.488 (0.134)	0.542 (0.279)	0.298 (0.131)	0.538 (0.306)	2.318 (0.504)	1.550 (0.371)	2.310 (0.533)
	r	1.000	1.028	0.993	0.976	1.002	1.010	0.981	0.978	0.985
1000	IFM	0.704 (0.042)	0.495 (0.089)	0.502 (0.045)	0.500 (0.076)	0.303 (0.041)	0.506 (0.091)	2.220 (0.155)	1.541 (0.123)	2.213 (0.142)
	MLE	0.703 (0.042)	0.494 (0.089)	0.499 (0.045)	0.498 (0.075)	0.301 (0.041)	0.505 (0.091)	2.222 (0.156)	1.541 (0.124)	2.215 (0.142)
	r	1.011	1.002	1.007	1.015	1.010	1.010	0.991	0.997	0.997

Table 3: Comparison of estimates of standard errors, (i) true, (ii) Godambe, (iii) jackknife with g groups; $n = 500$.

	approach	MSE(θ_{12})	MSE(θ_{13})	MSE(θ_{23})
$\boldsymbol{\alpha} = (0.0, 0.7, 0.0)', \boldsymbol{\theta} = (-0.5, 0.5, -0.5)$				
(g, m) (500, 1) (250, 2) (125, 4) (50, 10)	(i)	0.00416	0.00314	0.00426
	(ii)	0.00402	0.00329	0.00402
	(iii)			
		0.00408	0.00331	0.00410
		0.00407	0.00333	0.00412
		0.00405	0.00333	0.00412
		0.00411	0.00340	0.00418
$\boldsymbol{\alpha} = (0.7, 0.7, 0.7)', \boldsymbol{\theta} = (0.9, 0.7, 0.5)$				
(g, m) (500, 1) (250, 2) (125, 4) (50, 10)	(i)	0.00063	0.00269	0.00452
	(ii)	0.00059	0.00248	0.00437
	(iii)			
		0.00061	0.00250	0.00441
		0.00061	0.00251	0.00442
		0.00062	0.00253	0.00447
		0.00062	0.00254	0.00450
$\boldsymbol{\alpha} = (1.0, 0.5, 0.0)', \boldsymbol{\theta} = (0.8, 0.6, 0.8)$				
(g, m) (500, 1) (250, 2) (125, 4) (50, 10)	(i)	0.00163	0.00385	0.00141
	(ii)	0.00174	0.00418	0.00133
	(iii)			
		0.00182	0.00440	0.00136
		0.00184	0.00441	0.00137
		0.00185	0.00443	0.00136
		0.00188	0.00448	0.00139

(i) mean square error (MSE): $\frac{1}{N} \sum_{s=1}^N (\tilde{\theta}_{12}^{(s)} - \theta_{12})^2$, $\frac{1}{N} \sum_{s=1}^N (\tilde{\theta}_{13}^{(s)} - \theta_{13})^2$, $\frac{1}{N} \sum_{s=1}^N (\tilde{\theta}_{23}^{(s)} - \theta_{23})^2$,

(ii) estimate from Godambe matrix: v_{12} , v_{13} , v_{23} ,

(iii) estimate from jackknife: $\frac{1}{N} \sum_{s=1}^N v_{12}^{(s)}$, $\frac{1}{N} \sum_{s=1}^N v_{13}^{(s)}$, $\frac{1}{N} \sum_{s=1}^N v_{23}^{(s)}$,

The MSE in (i) should be considered as the true variance of the parameter estimate, and (ii) and (iii) should be compared with each other and with (i). Table 3 summarizes the numerical computation of the variance estimates of $\tilde{\theta}_{12}, \tilde{\theta}_{13}, \tilde{\theta}_{23}$ based on approaches (i), (ii) and (iii); the sample size is $n = 500$ and there were $N = 500$ simulations in each case. For (iii) the jackknife method, different combinations of (g, m) were used. Three cases with different marginal parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and different dependence parameters $\boldsymbol{\theta} = (\theta_{12}, \theta_{13}, \theta_{23})$ were used. The parameter values are reported in the table. In Table 3, the three measures are very close to each other. This is also the cases in simulations in other models.

6. Data examples.

6.1 Example 1.

Table 4: Trivariate count data: some summary statistics.

margin	mean	variance	max	Q3	med	Q1	min	margin	correlation
1	4.7	15.07	22	6	4	2	1	1,2	0.0192
2	6.5	13.64	15	9	6	4	0	1,3	-0.1666
3	6.6	32.61	30	8	6	3	0	2,3	-0.3667

Table 5: Trivariate count data: estimates of parameters for the multivariate Poisson–lognormal model

margin	$\tilde{\mu}$ (SE)	$\tilde{\sigma}$ (SE)
1	1.388 (0.098)	0.551 (0.122)
2	1.784 (0.098)	0.425 (0.090)
3	1.660 (0.120)	0.672 (0.121)
margin	$\tilde{\rho}_{jk}$ (SE)	
1,2	0.059 (0.315)	
1,3	-0.260 (0.208)	
2,3	-0.605 (0.206)	

This example consists of a data set in Aitchison and Ho (1989) of trivariate counts, the counts of pathogenic bacteria at 50 different sterile locations measured by three different air samplers. One of the objective of the study is to investigate the relative effectiveness of three different air samplers to detect pathogenic bacteria.

Summary statistics (means, variance, quartiles, maxima, minima and pairwise correlations) are given in Table 4. The initial data analysis indicate that there are some extra–Poisson variation as the variance to mean ratio for each margin (or sampler) ranges from 2 to 5, with the sampler 3 more spread out than the other two samplers.

We fit the multivariate Poisson–lognormal model (see Example 2.3) to the data and estimate the parameters using the IFM method, with the (delete–one) jackknife method for standard errors. Table 5 contains the estimates and SEs of the parameters. The estimated means, variances and correlations, computed as functions of the parameters based on the fitted model, are quite close to the empirical means, variances, correlations in Table 4. Various diagnostics showed that the multivariate Poisson–lognormal model to be a more appropriate model for this data set than other models. The analysis indicate that the sampler 3 tends to be negatively correlated with sampler 2 and sampler 1. Similar results were obtained in Aitchison and Ho (1989). Using an approximation to the multivariate log–likelihood, they report estimates from a final model with $\sigma_j = \sigma$ for all j , and $\rho_{jk} = \rho$ for all $j \neq k$, and got estimates (standard errors) of $\hat{\mu}_1 = 1.39$ (0.11), $\hat{\mu}_2 = 1.75$ (0.10), $\hat{\mu}_3 = 1.70$ (0.10), $\hat{\sigma} = 0.56$ (0.05), and $\hat{\rho} = -0.28$ (0.10). These are quite close to our estimates from the IFM method for this simplified model: $\tilde{\mu}_1 = 1.388$ (0.098), $\tilde{\mu}_2 = 1.784$ (0.098), $\tilde{\mu}_3 = 1.660$ (0.120), $\tilde{\sigma} = 0.523$ (0.062), and $\tilde{\rho} = -0.347$ (0.133).

Table 6: Multivariate ordinal data: univariate marginal (and relative) frequencies.

Outcomes	margin			
	1979	1980	1981	1982
< 5mi.				
L	14 (0.122)	18 (0.157)	14 (0.122)	18 (0.157)
M	69 (0.600)	73 (0.635)	72 (0.626)	70 (0.609)
H	32 (0.278)	24 (0.209)	29 (0.252)	27 (0.235)
> 5mi.				
L	9 (0.059)	35 (0.229)	17 (0.111)	23 (0.150)
M	110 (0.719)	93 (0.608)	117 (0.765)	110 (0.719)
H	34 (0.222)	25 (0.163)	19 (0.124)	20 (0.131)
all				
L	23 (0.086)	53 (0.198)	31 (0.116)	41 (0.153)
M	179 (0.668)	166 (0.619)	189 (0.705)	180 (0.672)
H	66 (0.246)	49 (0.183)	48 (0.179)	47 (0.175)

6.2 Example 2.

This example consists of multivariate (longitudinal) ordinal response data listed in Fienberg *et al.* (1985) and Conaway (1989). The data come from a study on the psychological effects of the accident at the Three Mile Island nuclear power plant in 1979. We use a multivariate probit model for each level of a binary covariate. This model highlights different features of the data compared with the models and analyses in cited articles. More detailed data analyses and comparisons of models are given in Xu (1996).

The study focuses on the changes in levels of stress of mothers of young children living within 10 miles of the plant. Four waves of interviews were conducted in 1979, 1980, 1981, 1982, and one variable measured at each time point is the level of stress (categorized as low, medium, or high). Hence stress is treated as an ordinal response variable with three categories, now labelled as L, M, H. There were 268 mothers in the study, and they were stratified into two groups, those living within 5 miles of the plant, and those living between 5 and 10 miles from the plant. There were 115 mothers in the first group and 153 in the second group.

Over the four time points and three levels are the ordinal response variable, there were 81 possible four-tuples of the form LLLL to HHHH but many of these had a zero count. There was only one subject with a big change in the stress level (L to H or H to L) from one year to another. 42% of the subjects were categorized into the same stress level in all four years. The frequencies by univariate margin (or by year) are given in Table 6. The medium stress category predominates and there is a higher relative frequency of subjects in the high stress category for the group within five miles of the plant compared with the the group exceeding five miles. From Table 6, there are not big changes over time, but there is a small trend towards lower stress levels, for the group exceeding five miles.

Table 7: Multivariate ordinal data: estimates of parameters (and SEs) for the multivariate probit model.

parameter	< 5 mi estimate (SE)	> 5 mi estimate (SE)
β_{11}	-1.16 (0.16)	-1.57 (0.17)
β_{12}	0.59 (0.13)	0.76 (0.11)
β_{21}	-1.01 (0.14)	-.74 (0.11)
β_{22}	0.81 (0.13)	0.98 (0.12)
β_{31}	-1.17 (0.16)	-1.22 (0.14)
β_{32}	0.67 (0.13)	1.15 (0.13)
β_{41}	-1.01 (0.14)	-1.04 (0.13)
β_{42}	0.73 (0.13)	1.12 (0.13)
ρ_{12}	0.785 (0.060)	0.678 (0.079)
ρ_{13}	0.696 (0.067)	0.463 (0.111)
ρ_{14}	0.653 (0.086)	0.436 (0.108)
ρ_{23}	0.806 (0.059)	0.750 (0.066)
ρ_{24}	0.636 (0.096)	0.510 (0.104)
ρ_{34}	0.844 (0.052)	0.562 (0.116)

Table 7 has estimates and standard errors of the multivariate probit model by the distance group. As would be expected, the dependence parameters for consecutive years are larger. In comparisons of the two groups (< 5 mi and > 5 mi), the dependence parameters are larger for the first group. This means that the mothers in the first group are probably more consistent in their responses over time.

7. Discussion.

The IFM method presented in this report applies to multivariate models with closure properties of the parameters, for example, parameters are associated with or are expressed in lower-dimensional marginal distributions. Together with the jackknife, it makes inference and data analysis more computationally feasible for many multivariate models.

There are modifications of the method when a parameter is common to more than one margin, and these will be presented in a separate report. One example in which a parameter can appear in more than one margin is when the copula has a special dependence structure; for the exchangeable dependence structure, a parameter is common to all bivariate margins, and for the autoregressive of order one dependence structure for a latent MVN distribution, each bivariate margin has a parameter which is the power of the lag-one correlation parameter. Other examples arise when different univariate margins have a common parameter; for example, in a repeated measures study with short time series over a short period of time, it may be reasonable to assume common regression coefficients for the different time points.

Besides working with higher-dimensional margins, which may be computationally difficult in some cases, there are two ways in which the IFM method can extend. They are:

1. average or weight the estimators from the likelihoods of the margins with the common parameter;
2. create inference functions or estimating equations from the sum of log-likelihoods of the margins that have the common parameter.

For both of these methods, the jackknife is still valid for obtaining standard errors of estimates of parameters.

Details on the use of variations of the IFM method for comparisons of models, with sensitivity analyses, are given in Xu (1996) for several data sets.

Acknowledgements. This research has been supported by a NSERC Canada grant.

References.

- Aitchison, J. and Ho, C.H. (1989). The multivariate Poisson-log normal distribution. *Biometrika*, **76**, 643–653.
- Anderson, J.A. and Pemberton, J.D. (1985). The grouped continuous model for multivariate ordered categorical variables and covariate adjustment. *Biometrics*, **41**, 875–885.
- Ashford, J.R. and Sowden, R.R. (1970). Multivariate probit analysis. *Biometrics*, **26**, 535–546.
- Bradley, R. A. and Gart, J. J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika*, **49**, 205–213.
- Conaway, M.R. (1989). Analysis of repeated categorical measurements with conditional likelihood methods. *J. Amer. Statist. Assoc.*, **84**, 53–62.
- Fienberg, S.E., Bromet, E.J., Follman, D., Lambert, D., and May, S.M. (1985). Longitudinal analysis of categorical epidemiological data: a study of Three Mile Island. *Environ. Health Perspectives*, **63**, 241–248.
- Godambe, V. P. (1960). An optimal property of regular maximum likelihood estimation. *Ann. Math. Statist.*, **31**, 1208–1211.
- Godambe, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika*, **63**, 277–284.
- Godambe, V.P. [editor] (1991). *Estimating Functions*. Oxford University Press, Oxford.
- Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *Ann. Math. Statist.*, **42**, 1977–1991.
- Hüsler, J. and Reiss, R.–D. (1989). Maxima of normal random vectors between independence and complete dependence. *Statist. Probab. Letters*, **7**, 283–286.
- Joe, H. (1993). Parametric family of multivariate distributions with given margins. *J. Mult. Anal.*, **46**, 262–282.
- Joe, H. (1994). Multivariate extreme value distributions and applications to environmental data. *Canad. J. Statist.*, **22**, 47–64.

- Joe, H. and Hu, T. (1996). Multivariate distributions from mixtures of max-infinately divisible distributions. *J. Mult. Anal.*, **57**, 240–265.
- Lesaffre, E. and Molenberghs, G. (1991). Multivariate probit analysis: a neglected procedure in medical statistics. *Statist. in Medicine*, **10**, 1391–1403.
- Lipsitz, S. R., Dear, K. B. G., Zhao, L. (1994). Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data. *Biometrics*, **50**, 842–846.
- McLeish, D. L. and Small, C. G. (1988). *The Theory and Applications of Statistical Inference Functions*, Lecture Notes in Statistics 44, Springer-Verlag, New York.
- Nash, J.C. (1990). *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*, 2nd edition. Hilger, New York.
- Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- Tawn, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika*, **77**, 397–415.
- Xu, J. J. (1996). *Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data*. Ph.D. thesis. Department of Statistics, University of British Columbia.