Statistica Applicata Vol. 4, n. 4, 1992

KNOT SELECTION IN FUZZY NON LINEAR MULTIVARIATE ANALYSIS

J.L.A. van Rijckevorsel *

Department of Statistics, TNO Institute of Preventive Health Care, Leiden, The Netherlands.

J. de Leeuw

Department of Psychology and Mathematics, UCLA, Los Angeles, USA.

This paper studies the effect of different types of discretization on the approximation of the first eigenvalue of the class of bivariate distributions. Different sample sizes are taken into account. By a reverse process the discretization is corrected for by nonlinear transformation of the discretized variables. Optimal nonlinear transformations are approximated by zero and first degree B–splines, which can be interpreted as fuzzy coding functions. The sampling error and discretization error are evaluated by a Monte Carlo design with three sample sizes, four different discretizations, using the bootstrap. The main result is that the discretization effect is small in relation to the sample effect. This result has a wider bearing on knot selection in general.

Keywords: Knot sequence, splines, correspondence analysis, bivariate distribution, non linear transformation, fuzzy coding

1. INTRODUCTION

I A SHORT REVIEW OF FUZZY NON LINEAR MULTI–VARIATE ANALYSIS (NLMVA)

Classical multivariate techniques come in two subclasses. There are multinomal techniques, intended for continuous multivariables and multinomali techniques for categorical variables. The assumption of multinormality is almost never true, it produces low dimensional sufficient stable statistics like covariances. Multinomial techniques have high dimensional insufficient statistics that are unstable, like cell counts but their nonparametric assumptions are almost always true. This situation gave rise to many so called nonlinear techniques. The review of NLMVA by Coppi and Di Ciaccio in this volume covers approximately the whole spectrum of non linear techniques assemled under that label. Here we like to refrain of two classes of techniques that generalize classical MVA by nonlinear transformations with knot

* *While at leave at the Department of Statistics, University of Naples Federico II, Italy.*

selection to reduce complexity of relationships and noise: categorical and continuous NLMVA.

Say the domain of a real valued random variable h, bounded by two values *a* and *b*, *a*<*b*, is partitioned into a number of intervals between *a* and *b*. Each interval is bounded by two interval points, t and t_{q+1}, called knots or discretization points. Assume that the knot sequence *{t}* consists of an increasing sequence of knots on [*a*,*b*]:

$$a \leq t_1 < t_2 < \ldots\ldots < t_{q-1} < t_q \leq b$$

The polynomial spline consists of several polynomial pieces of order *v*, defined on intervals between at least two subsequent knots. If $v \geq 2$ pieces are connected e.g. at the knots, such that together they constitute a continuous function on the domain of the variable. Furthermore when $v \geq 3$ connections are such that i-th left and right derivation are equal, i = 1,.., *v*–2. As a consequence the function is smooth. Piecewise polynomials called spline functions are a flexible set of transformation functions. Two types of spline functions are important for NLMVA, the regression spline and the smoothing spline. There exist many more kinds of spline functions, that we will not discuss.

The regression spline. A very simple spline function is the step function. A dummy variable consists of a step function that codes data points into a category of a nominal variable. This creates an immediate link with categorical multivariate data analysis. A linear transformation, the most elementary of transformations exclusively linked with multinormal techniques, is equal to a piecewise linear spline function defined on just one interval (i.e. with two external knots), which creates a link with linear MVA. Higher degree spline functions are smooth non linear transformation functions with more degrees of freedom. Spline functions make it possible to deal with the MVA of categorical, categorized and continuous data in one framework. Such splines are called regression splines (Ramsay, 1988) and they invariably have less coefficients than data points. This means that decisions have to be made about the quantity and placement of the knots.

Categorical NLMVA as advocated by Gifi (1990) and his coworkers quantifies categorical data in such a way that multinormal criteria like the multiple correlation or the sum of eigenvalues of the quantified data are maximized. Which means that canonical correlation, multiple regression and related techniques are made available for categorical data. The quantification process is called optimal scaling and De Leeuw (1988) discusses the optimality properties. Optimal scaling is a way of "transforming" categorical data using stepfunctions with or without additional restrictions. Van Rijckevorsel (1987, 1988) generalizes the coding by stepfunctions

in order to get a system of fuzzy coding that includes Basis splines as a special case. The french analogon is called "codage flou" (Martin, 1988). In categorical data analysis knot selection is a prior, that is defined in the stages that precede data analysis. Like the definition of the scale values of a rating scale, or the categories of a nominal variable. As soon as a continuous variable is categorized into classes or intervals, which is customary to make age intervals or income brackets for instance, this categorization is not necessarly a prior any more. It has become a parameter that can be optimized, which bring us to knot selection. The need to slice a variable is often caused by the desire to expose non linear relationships in the data. Categorical NLMVA techniques with optimal scaling have however no provision for smooth transformations; albeit it is conceptually only a small step to use regression splines as a way of optimal scaling. Exceptions are the PRINQUAL technique incorporated in SAS (1988) generalizing PCA (=Principal Components Analysis) and the SPLINALS technique by Van Rijckevorsel (1987) generalizing MCA (=Multiple Correpondance Analysis) & PCA.

This state of affairs is regrettably so because conceptually as technically, it is quite feasible to generalize OVERALS (=the canonical analysis of k–sets), REDUNDALS (=redundancy analysis), SERIALS (=time series), MORALS (=multiple regression), ... etc. to analyze continuous data in a non parametric way using piecewise polynomials as a kind of optimal scaling in their backfitting algorithm (and thus introducing the problem of knot selection meanwhile).

The smoothing spline. If two objectives are considered simultaneously by transforming a variable, like maximizing the model fit and penalizing its roughness meanwhile, we are dealing with another type of spline functions, called the interpolating or smoothing spline (Ramsay 1988). The smoothing spline has as many knots as distinct data points and initially no knot selection problem exists. There is however the problem of balancing the amount of smoothness against the model fit. This is neatly combined in one transformation function with a single smoothing parameter called lambda. Here a decision has to be made about the value of lambda. The model in the smoothing spline is the regression model, or more generally, an additive model. The smoothing spline is by definition more appropriate for continuous NLMVA. Continuous NLMVA, advocated by Breiman and Friedman (1985), Hastie and Tibshirani (1990), and Friedman (1991) uses non parametric uni– or multivariate transformations of continuous data that optimize multinormal criteria. Excellent reviews of continuous NLMVA are Hastie and Tibshirani (1990) and Friedman (1991). For the more or less independent french approach in PCA see Besse (1989).

II THE STATE OF THE ART OF KNOT SELECTION IN NLMVA

Historically the term knot placement is linked with the use of piecewise polynomials (spline functions). The local functional pieces are defined on intervals that are bordered by knots. Nowadays the knot selection is to be seen in the much wider context of window selection with a non linear smoother of any type. (Breiman, 1988).

Knot selection when using regression splines in NLMVA can be solved automatically. The "optimal" selection of knots is a combinatorial non linear optimization problem that in general is not (yet) easily combined with popular statistical methods like least squares estimation or maximum likelihood estimation. For various methods and their drawbacks see Hastie et al. (1990). There exist a vivid dispute whether knot placement can be optimized or not. And whether optimization is worth the effort, because with the signal to noise ratio's one encounters in NLMVA, the shape of the transformation function does not depend heavily on the placement or number of knots and knot selection might as well be solved by rules of the thumb that work satisfactorily. The comments on Ramsay (1988) are most enlightening. See also Hastie et al. (1990, p.247–254).

Another closely related problem is the automatic setting of the smoothness parameter when using smoothing splines. This is mostly solved by generalized cross validation (=GCV). See Gu and Wahba (1991). Conceptually it is a related problem to knot selection in regression splines because there the smoothness of the transformation function depends partly on the knot selection. A more complicated problem is the automatic simultaneous setting of the degree and selecting a multivariate knot placement when using multivariate regression splines. This introduces a new knot selection like problem: the transformation function cannot be defined on combinations of intervals of different variables that are (nearly) empty. This is called the curse of dimensionality. See Hastie et al. (1990, p.259–) or Friedman (1991).

Knot selection is such a general problem that it has to be restricted to a manageable form that can be solved analytically and computationally. The subproblem that we discuss is the estimation of the true averaged correlation coefficient, after slicing bivariately distributed variables in different ways. Maximizing the correlation coefficient is a classical MVA criterion. The influence of various knot sequences on the estimation of the correlation coefficient for different sample sizes should tell us ideally something about the error caused by the knot placement. The transformation function used is a regression spline, based on B–splines of degrees 0,1 with a least squares criterion. (De Boor,1978 p. 249)

2. THEORY

The optimization problem

The notation used is as follows: E, R and V are expectation, correlation and variance operators. $\underline{h}$ is a random variable. $R^{(s)}$ is the s–th correlation matrix (m x m), $j=1,...,m$ is the running index for variables and $s=1,....\infty$ is the running index for the dimensions of the space L_j of nonlinear transformation functions of variable j. $\underline{x}$ is a random variable to be estimated. δ_{st} is Kronecker's delta and $r_{jl}^{(s)}$ is the correlation between j and l in dimension s. $\sigma(*,*) = \min\{\sigma(y;z|y;z)\}$ is the minimum of $\sigma(y;z)$, while $\sigma(y;z)$ is the loss depending on y and z.

The optimization problem is to find a nonlinear transformation of a multivariate random variable, which is a piecewise function on the discretization points and which maximizes the averaged squared correlation ratio between the transformed variables. The correlation ratio is defined as the variance of the conditional expectation of $\underline{x}$, given $\underline{h}_j$, divided by the variance of $\underline{x}$. Much of the necessary theory can be found in Gifi, (1990), Van Rijckevorsel et al., (1985).

Say we want to find weights $a_1,....,a_m$ and a random variable $\underline{x}$ such that

$$\min\ \sigma(\underline{x};a) = \frac{1}{m}\sum_{j=1}^{m} V\left(\underline{x} - a_j\underline{h}_j\right)$$

is minimized with normalization $V(\underline{x}) = 1$ or $\sum_{j=1}^{m} V\left(a_j\underline{h}_j\right) = m$. For either normalization this leads to $\sigma(*,*) = 1 - \lambda_+$, while λ_+ is equal to the dominant eigenvalue of the matrix $R^{(1)}$ with elements $m^{-1}R(a_j\underline{h}_j, a_l\underline{h}_l$. This is a fairly trivial result in linear multivariate analysis. The first eigenvalue of a matrix of correlation coefficients is equal to the averaged squared correlation.

The nonlinear case

Consider the non linear transformation

$$\phi_j \in L_j = \{\ \phi : E[\ \phi(\underline{h}_j)] = 0;\ E[\ \phi(\underline{h}_j)^2\ ] < \infty\ \}$$

and instead over the linear weights $a_1....a_m$, we minimize σ over $\phi = \{\phi_1,...,.\phi_m\}$ and over $\underline{x}$

$$\sigma(\underline{x};\phi) = \frac{1}{m}\sum_{j=1}^{m} V\left(\underline{x} - \phi_j\left(\underline{h}_j\right)\right),$$

with normalization $V(\underline{x}) = 1$ or $\sum_{j=1}^{m} V\left(\phi_j\left(\underline{h}_j\right)\right) = m..$

Define a complete orthonormal basis G_{js}, s=1,2, ..., for L_j such that

$\phi_j(\underline{h}_j) = \sum_{S=1}^{\infty} a_s G_{js}(\underline{h}_j)$, then minimizing σ ($\underline{x}$;ϕ) has the characteristic function

$$\sum_{t=1}^{\infty} R^{(st)} a_s = \lambda_+ a_s,$$

while $R^{(st)}$ is equal to an *m* by *m* correlation matrix for dimensions *s* and *t*. This minimization leads to σ (*,*) = 1 – λ_+ and $\lambda_+$is equal to the dominant eigenvalue of a supermatrix with submatrices $R^{(st)}$ with elements $m^{-1}R(\phi(\underline{h}_j),\phi(\underline{h}_l))$. For a space L_j of non linear transformation functions we are dealing with the eigenvalue of an infinitely large supermatrix that consists of a infinite number of *m* x *m* correlation matrices. The largest eigenvalue of the correlation matrix of non linearly transformed variables is equal to the maximum averaged correlation ratio times the number of variables.

For the multivariate case we restrict the basis of non linear transformations in such a way that the correlations between dimensions are equal to zero:

$$R(g_{js}(\underline{h}_j),g_{lt}(\underline{h}_l)) = \delta_{st} r_{jl}^{(s)}.$$

There are many bivariate distributions whose canonical functions (= eigenvectors) are orthogonal polynomials. For an extensive discussion with references see Schriever (1986). Schriever (1986, p. 51–52) proves that relevant mathematical properties concerning canonical functions, carry over to the discretized distributions. Meaning that the canonical functions of such discretized bivariate distributions are also orthogonal functions, providing the discretization is fine enough. The best known example is the Mehler expansion of the bivariate standard normal, where we choose Hermite–Chebyshev polynomials as an infinite basis, approximated by a finite basis of stepfunctions. In this case the nonlinear transformation of bivariate normal variables that maximizes the first eigenvalue is a linear transformation. See also Kendall & Stuart (1979, p.600). Other bivariate distributions whose canonical functions are orthogonal polynomials are the bivariate binomial, the bivariate Poisson, the bivariate Gamma and many others (Lancaster 1980, Schriever 1986). The eigenproblem is thus:

$$\sigma (*,*) = 1 - \lambda_+$$

while $\phi_j(\underline{h}_j) = a_j\underline{h}_j$ and λ_+ is equal to the dominant eigenvalue of the correlation matrix with elements $m^{-1}R(a_j\underline{h}_j,a_l\underline{h}_l)$.

Approximation by a finite basis

Orthogonal polynomials are approximated arbitrarily well by a finite basis of stepfunctions. Define discretization as the non linear transformation $\psi_j(\underline{h}_j)$. If $\underline{e}_j$ = $\psi_j(\underline{h}_j,.....,\underline{h}_m)$ are random bivariate in the above sense and ϕ_j is a monotone non

linear transformation then

$$\phi_j(\underline{e}_j) = a_j\psi_j^{-1}(\underline{e}_j) = a_j\underline{h}_j.$$

The nonlinear transformation of the discretized bivariately distributed variables is a linear transformation of the original (non discretized) variables. The usual discretization into categories is a nonlinear transformation by stepfunctions. This can be generalized by considering other non–crisp i.e. fuzzy codes on the same sequence of knots. A smooth and regular way of fuzzy coding consists of using B-splines. While a step function is equal to a B–spline of degree 0. Higher degree B-splines are linear combinations of positive quantities starting with the stepfunction. A B–spline is defined by de Boor (1978, p.131) as

$$G_{q,v}(\underline{h}) = (\underline{h} - t_q)(t_{q+v-1} - t_q)^{-1}G_{q,v-1}(\underline{h}) + (t_{q+v} - \underline{h})(t_{q+v} - t_{q+1})^{-1}G_{q+v,v-1}(h)$$

and

$$G_{q,1}(\underline{h}) = \begin{cases} 1, t_q \le \underline{h} < t_{q+1} \\ 0, \text{elsewhere.} \end{cases}$$

A polynomial spline function ϕ is defined as a linear combination of basis splines:

$$\phi(\underline{h}) = \sum_{q=1}^{W} a_q G_{q,v}(\underline{h}).$$

The corresponding notation is: r equals the number of interior knots or discretization points; v is the order of the basis; t_q is the knot q; q is the running index of the number of dimensions and $w = r + v$ is the dimension of the basis, while the number of intervals $\neq 0$ equals v, so $G_{q,v}(\underline{h})$ is the q–th basis–function of order v (i.e. degree $v+1$) for the variable $\underline{h}$.

The first eigenvalue of the correlation matrix of non linearly transformed variables based on stepfunctions is found by multiple correspondence analysis or homogeneity analysis. Fuzzy homogeneity analysis finds the analogue based on higher degree B–splines. See De Leeuw et al. (1988). Fuzzy and crisp multiple correspondence analysis can be interpreted as an approximation of continuous nonlinear analysis as defined by Dauxois and Pousse (1976) or Mallet (1982).

Let us resume this section. The averaged squared correlation between bivariately distributed variables is approximated by the averaged squared correlation between non linear transformations of their discretizations, and these transformations are linear transformations of the original variables because of distributional properties. For a fixed knot sequence the approximation of the original variables by a first degree spline function will be a linear function of these variables and the

approximation by stepfunctions will be linear in its steps, independent of the knot placement or number of knots, loosely speaking. The approximation by a stepfunction with a fixed knot sequence will therefore always be inferior to the polynomial spline function of degree one in this case.

3. METHOD

Three samples of sizes 100, 1000, 10000 of nine bivariate normal random variables are drawn with a fixed correlation coefficient equal to .5. We use four different knot sequences of four knots each: Optimal (1), U–shaped (2), Uniform (3) and Skew (4). The knot sequences are named after the shape of the histogram of the proportions by which they partition the surface under the normal curve (see figure 1). The first knot sequence optimizes a loss function that actually describes the loss with respect to the normal distribution. This is why we call this knot sequence optimal. The loss is caused by the discretization using stepfunctions and minimizing this loss is not equivalent to maximizing the first eigenvalue for further references see Gifi (1990, p. 405).. Because of the ease of computation and the familiarity of many results we selected the random bivariate normal distribution as an example.

The dominant eigenvalue of the continuous distribution with all mutual correlations equal to .5 is $\lambda_+ = 45/81 = .5556$. For each knot sequence there exists an approximation of the first eigenvalue of the continuous distribution, given a finite approximation with a limited number of stepfunctions. The first eigenvalue of a correspondence analysis of two discretized variables is proportional to the maximal correlation between them. The result is due to Hirschfeld in 1935, see Kendall and Stuart (1979). Due to symmetry we can translate the results to the case of nine variables. These theoretically maximal values for the respective knot sequences are .5222 (Optimal), .4938 (U–shaped), .5160 (Uniform) and .4981 (Skew). For each particular knot sequence the best approximation by stepfunctions should come near to the corresponding value. The fact that these theoretical values differ from the eigenvalue of the continuous variables is caused by the intrinsic discretization error. Each knot sequence has its own discretization error. For 12 combinations of knot sequences and sample sizes these correlations are approximated by stepfunctions.

From each sample and knot sequence 10 bootstrap samples were taken (i.e. with replacement), resulting in 120 bootstrap samples. 120 analyses using stepfunctions were executed. The bootstrap study only concerns stepfunctions.

The first eigenvalue of the continuous distribution is approximated directly by

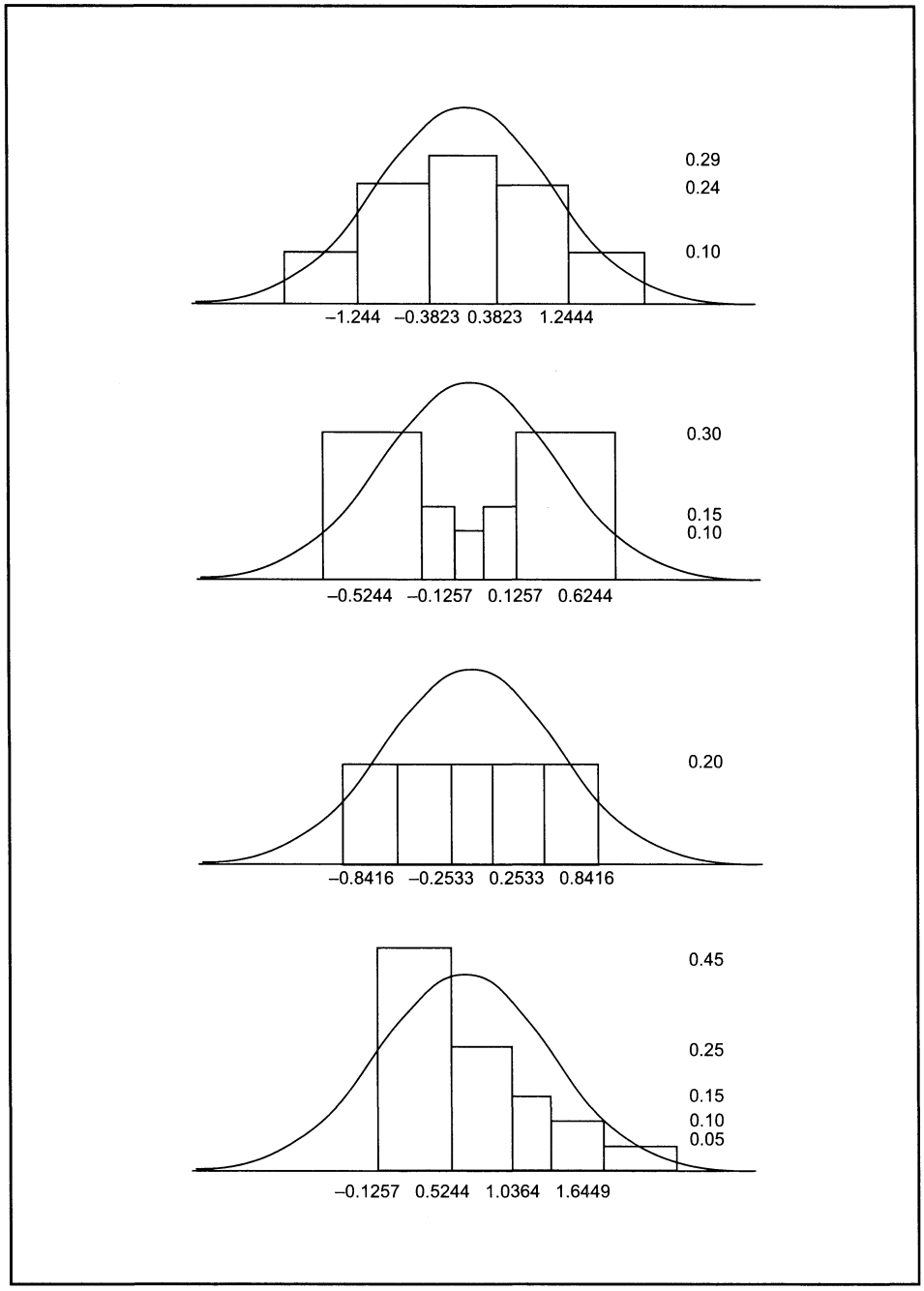

Fig. 1: The 'optimal', 'U-shaped', 'uniform' and 'skew' discretization of the normal distribution with knots on the x-axis and the proportions of surface under the curve on the y-axis.

transformations based on first degree splines, because the discretization error is smoothed away by the global smoothness of the resulting transformation, that perfectly coincides with a first degree polynomial. This is a property that a stepfunction only asymptotically can achieve.

The spline approximations are computed for the same sample sizes and knot sequences. The approximations are computed by fuzzy homogeneity analysis using zero and first degree B–splines, where only the first eigenvalue is considered.

For the computation of the first degree spline transformations we used the SAS routine TRANSREG.

4. RESULTS

The result is that the averaged squares correlations between the original variables are nicely approximated by the averaged squared correlations between non linear transformations of the discretized variables. Transformations based on stepfunctions approximate the theoretical eigenvalues for the respective

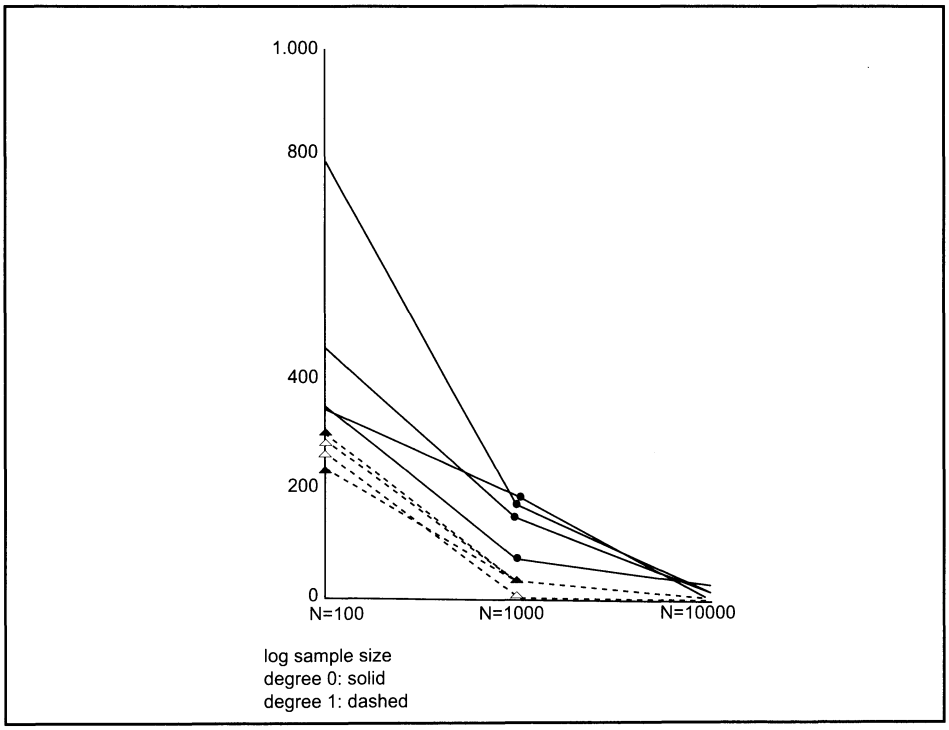

Fig. 2: The log approximation error * 1000 of zero and first degree B–splines versus log sample size for four different knot sequences

discretizations, see figure 2.

The approximation is better with increasing sample size irrespective of the knot sequence for both types of transformation. The variation in bootstrap means per sample size does hardly vary over the different knot sequences. See table I. It doesn't matter, so to say, which knot sequence you select, they all approximate their respective theoretical value just as good. The approximation by first degree B-splines is superior to the approximation by stepfunctions. This is correct because a piecewise linear is a better approximation to a linear function for a fixed knot sequence than a piecewise stepfunction.

5. DISCUSSION

Where you put the knots is of minor importance because whatever you do, you are not far wrong. See tables I & II.

This means that a knot optimizer like Friedman's recursive partitioning on discretized bivariate distributions with maximally 4 knots would add little to the absolute information we have already. When using first degree B–splines and/or larger samples ($N \geq 1000$) there is nothing to gain by knot optimization anyhow in this case. For stepfunctions we would see that the approximation will be near to .5222 because for the "optimal" knot sequence asymptotic approximations are already very good for small numbers of discretization points. See Gifi (1990, p. 405). This means that the actual space for improvement with a fixed number of knots is approx. 5%. And additionally with a larger number of knots the population value is approximated, with an absolute gain of maximally 9%. The surprising effect of our research reported here, is that for such a limited number of knots and for such "weird" positioning of knots, the approximation of the first eigenvalue is surprisingly close to the real thing. And because the approximation is based on a kind of optimal scaling combined with the popular criterion of the first eigenvalue, this has a wider bearing on other forms of NLMVA. Our results confirm Gifi's conjecture that the sample error is larger than the discretization error for a fixed number of knots. See table I. Because this argument has a general character it also refers to the sometimes hot blooded discussions about optimal knot placement in spline transformations. A.o. Wold and Ramsay are advocating robust knot placement with equal numbers of measurements per interval and few knots. For further references see Van Rijckevorsel (1988) and Ramsay (1988). This study confirms these rules of the thumb and adds the additional caveat to beware of small samples combined with stepfunctions. From the point of stability and approximation first degree B–

splines are to be preferred over stepfunctions in analyzing various discretized bivariate distributions. The results reported in this paper if restricted to the first eigenvalue, pertain to the general case because many bivariate distributions are associated with orthogonal polynomials.

Tab. I: The approximation of the first eigenvalue by the nonlinear transformations based on step functions for four different knot placements

	Theoretical value	N	Bootstrap mean	Bootstrap variance
Optimal	.	100	.4762	$9.6*10^{-4}$
	.5222	1000	.5369	$9.6*10^{-5}$
		10000	.5208	$1.2*10^{-5}$
		100	.4582	$1.1*10^{-3}$
U–shaped	.4938	1000	.5013	$1.6*10^{-4}$
		10000	.4964	$4.6*10^{-6}$
		100	.4359	$1.0*10^{-3}$
Uniform	.5160	1000	.4988	$1.0*10^{-4}$
		10000	.5174	$8.9*10^{-6}$
		100	.5327	$1.3*10^{-3}$
Skew	.4981	1000	.5169	$1.8*10^{4}$
		10000	.4984	$2.0*10^{-5}$

Tab. II: The approximation of the first eigenvalue of the continuous multinormal distribution (=.5556) by first degree B–splines.

	N = 100	N = 1000	N = 10000
Optimal	.5857	.5592	.5560
U–shaped	.5793	.5590	.5559
Uniform	.5819	.5596	.5559
Skew	.5846	.5595	.5560

REFERENCES

Besse Ph., 1989, *Approximation spline et optimalité en analyse en composantes principales.* Thèse d'Etat. Toulouse: Université de Toulouse III.

Breiman L., 1988, Comment on "Monotone splines in action" J. Ramsay. *Statist. Sci.* 3, 446–50.

Breiman L., Friedman J.H., 1985, Estimating optimal transformations for multiple regression and correlation (with discussion). *J. Amer. Statist. Assoc.* 80, 580–619.

Coppi R. and Di Ciaccio A., 1992, The methodological impact of nonlinear analysis. *Statistica Applicata,* this issue.

Dauxois J. and Pousse A., 1976, *Les analyses factorielles en calcul de probabilité et en statistique: Essai d'ètude synthetique.* Thèse d'Etat. Toulouse: Université de Toulouse III.

De Boor C., 1978, *A practical guide to Splines.* Springer, New York.

De Leeuw J., 1988, Multivariate analysis with linearizable regressions. *Psychometrika* 53 4 437–454.

De Leeuw J., Van Rijckevorsel J.L.A., 1988, Beyond Homogeneity Analysis. In: J.L.A. van Rijckevorsel and J. de Leeuw (Eds.), *Component and Correspondence Analysis,* Wiley, New York.

Friedman J.H., 1991, Multivariate Adaptive Regression Splines (with discussion). *Ann. Statist.* 19, 1–141.

Gifi A., 1990, *Nonlinear Multivariate Analysis.* Wiley, New York.

Gu, C., Wahba G., 1991, Minimizing GVC/GML scores with multiple smoothing parameters via the Newton method. *J. Sci. Statist. Comput. 12.*

Hastie T., Tibshirani R., 1990, *Generalized Additive Models.* London, Chapman and Hall.

Kendall M., Stuart A., 1979, *The Advanced Theory of Statistics,* Vol. 2, Charles Griffin, London.

Lancaster H.O., 1980, Orthogonal models for contingency tables. In: P.R. Krisnaiah (ed.) *Developments in Statistics 3.* Academic Press, New York.

Mallet J.L., 1982, Propositions for fuzzy characteristic functions in data analysis. In: H. Caussinus *et al.*. (Eds.), *Compstat 82. Proceedings in Computational Statistics.* Physica–Verlag, Vienna.

Martin J–F., 1988, On probability coding. In: J.L.A. van Rijckevorsel and J. de Leeuw (Eds.), *Component and Correspondence Analysis,* Wiley, New York.

Ramsay J.O., 1988, Monotone regression splines in action. *Statistical Science.,* Vol. 3, 4, 425–461.

SAS Additional SAS/STAT Procedures, Release 6.03. *SAS Technical Report: P–179,* SAS Institute Inc., Cary, NC.

Van Rijckevorsel J.L.A., 1987, *The application of Horseshoes and Fuzzy Coding in Multiple Correspondence Analysis.* DSWO Press, Leiden.

Van Rijckevorsel J.L.A., 1988, Fuzzy coding and B–splines. In: J.L.A. van Rijckevorsel and J. de Leeuw (Eds.), *Component and Correspondance Analysis,* Wiley, New York.

Van Rijckevorsel J.L.A., De Leeuw J., Bettonvil B., 1985, *Recovery and stability in nonlinear principal components analysis.* RR–85–21, Department of Datatheory, University of Leiden.

Schriever B.F., 1986, *Order Dependence.* CWI–Tract 20. Centre for Mathematics and Computers Science, Amsterdam.

RIASSUNTO

Questo articolo studia l'effetto di differenti tipi di discretizzazione sull'approssimazione del primo autovalore della classe di distribuzioni bivariate. Campioni di diversa ampiezza sono presi in considerazione. La discretizzazione viene corretta attraverso un procedimento inverso di trasformazione non lineare di variabili discretizzate. Le trasformazioni non lineari ottimali sono approssimate per mezzo di B–splines di grado zero e uno, che possono a loro volta essere interpretate come funzioni di codifica fuzzy. L'errore di campionamento e l'errore di discretizzazione sono stimati per mezzo di un disegno Monte Carlo con tre campioni di differente ampiezza e quattro differenti discretizzazioni, usando il metodo bootstrap. Il risultato più importante è che l'effetto della discretizzazione è limitato rispetto all'effetto prodotto dalla differente ampiezza dei campioni. Questo risultato fornisce un contributo di più ampia portata al problema della selezione dei nodi.