Acta Universitatis Matthiae Belii

Series Mathematics
Number 18

Matej Bel University, Banská Bystrica, Slovakia
2011

Acta Universitatis Matthiae Belii, Series Mathematics

Managing editor of the issue: **Petr Hliněný**

Faculty of Informatics, Masaryk University,
Botanická 68a,
602 00 Brno,
Czech Republic
E-mail: hlineny@fi.muni.cz

Scope of Journal: Acta Universitatis Matthiae Belii, Series Mathematics publishes original research articles and survey papers in all areas of mathematics and theoretical computer science. Occasionally valuable papers dealing with applications of mathematics in other fields can be published.

Manuscripts: Manuscripts submitted for publication must be written in English. Please send the source LaTeX file (use please `amsart` document class) and figures needed to compile the output form as an attachment by e-mail to the address `actamath@savbb.sk`.

Subscriptions: Acta Universitatis Matthiae Belii series Mathematics is published once a year. The journal is available for exchange. Information on exchange is available from the Editorial Office. On-line version of the journal is available at `http://actamath.savbb.sk`.

Matej Bel University

Acta Universitatis Matthiae Belii
series Mathematics
No. 18 (2011)

Published by
Science and Research Institute
Matej Bel University, Banská Bystrica

Proceedings of ISCAMI 2010 Conference

edited by

Radko Mesiar and Daniel Ševčovič

Editorial

International Student Conference on Applied Mathematics and Informatics is a series of traditional student conferences organised in turn by Slovak Technical University and University of Ostrava. In 2010, ISCAMI was organised by the Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Slovak University of Technology, in collaboration with IRAFM, University of Ostrava. The conference was held in May 20-23, 2010, in Bratislava. It took over a tradition and has extended its scope by subjects in informatics and applications in mathematical economics and finance. The main purpose of the conference series is to bring together young researchers and students and to give them an opportunity to present their achievements and ideas in the area of applied mathematics, informatics and various applications.

The next five papers are full versions of peer reviewed contributions presented at the 11th ISCAMI 2010. This special issue was edited by R. Mesiar and D. Ševčovič. The next conference is scheduled for May 6-8, 2011 in Malenovice, Czech Republic.

R. Mesiar and D. Ševčovič

ACTA UNIVERSITATIS MATTHIAE BELII, series MATHEMATICS 18 (2011), 7–16
Also available at http://actamath.savbb.sk
ISBN 978-80-557-0160-8

A Confidence Interval for the Probability Difference of Overall Treatment Effect – Simulation Study

Pavla Dokoupilová
Department of Mathematics and Statistics
Faculty of Science, Masaryk University
Kotlářská 2, 611 37 Brno, Czech republic
E-mail: pavla@krajickova.cz

Abstract

One of the main aims of the meta-analysis of clinical trials is the determination of the effectivity of a new type of treatment. The effectivity is determined by the difference of the effectivity of a standard treatment and the new treatment. In the case of binary data the difference can be measured by a probability difference. This paper presents the construction of the confidence interval for the probability difference of overall treatment effects in the meta-analysis based on multicentre trials. For the construction of the confidence interval the procedures of Wimmer & Witkovský (2004) and Kenward & Roger (1997) have been used. The second part of this paper is a simulation study which presents properties of the proposed confidence interval.

Keywords multicenter trial, confidence interval, probability of success, linear model with random effects.

MSC (2010) 62P10.

Received: July 20, 2010; Revised: October 6, 2010; Accepted: October 7, 2010.

1 The model

Let us consider a clinical trial performed in I centers. Suppose that the number of subjects included in the trial in the ith center is $n_{T,i} + n_{C,i}$ for $i = 1, 2, \ldots, I$ where $n_{T,i}$ is the number of patients in the treated group and $n_{C,i}$ is the number of patients in the control group. Patients in the treated group in the ith center succeed with probability $p_{T,i}$ and patients in the control group in the ith center succeed with probability $p_{C,i}$ for $i = 1, 2 \ldots, I$. All subjects are consider to be independent.

Number of successes in the treated group in the ith center is denoted by random variable $X_{T,i}$ and number of successes in the control group in the ith center is denoted by random variable $X_{C,i}$. Then $X_{T,i} \sim Bi(n_{T,i}, p_{T,i})$ and $X_{C,i} \sim Bi(n_{C,i}, p_{C,i})$. $X_{l,i} \sim Bi(n_{l,i}, p_{l,i})$ for $l \in \{T, C\}$ means that $X_{l,i}$ has binomial distribution with the sample of size $n_{l,i}$ and the probability of success $p_{l,i}$. Random variables $X_{T,1}, \ldots, X_{T,I}$, $X_{C,1}, \ldots, X_{C,I}$ are stochastic independent. We will next work with random variables

$$Y_{l,i} = \frac{X_{l,i}}{n_{l,i}}, \text{ for } l \in \{T, C\} \text{ and } i = 1, \ldots, I. \tag{1.1}$$

Suppose that the true probabilities of success in the ith center $p_{T,i}$ and $p_{C,i}$, randomly fluctuate around common probabilities of success p_T and p_C. We want to estimate the probability difference $p_T - p_C$. So

$$p_{l,i} = p_l + b_{l,i}, \text{ for } l \in \{T, C\} \text{ and } i = 1, \ldots, I. \tag{1.2}$$

where $b_{l,i}$ is a random effect of the ith center and suppose that $b_{l,i} \sim N(0, \sigma^2_{l,0})$ [1] which means $b_{l,i}$ is normally distributed with the mean 0 and the variance $\sigma^2_{l,0}$.

The final situation can be represented by linear model with random effects

$$Y_{l,i} = p_l + b_{l,i} + \varepsilon_{l,i} \text{ for } l \in \{T, C\} \text{ and } i = 1, \ldots, I \tag{1.3}$$

where $\varepsilon_{l,i}$ are error terms and $\varepsilon_{l,i} \sim N(0, \sigma^2_{l,i}/n_{l,i})$. In matrix notation we get

$$\begin{aligned} \mathbf{Y} = \begin{pmatrix} \mathbf{Y_T} \\ \mathbf{Y_C} \end{pmatrix} \approx & \quad N\left(\begin{pmatrix} \mathbf{1}_{I\times 1} & \mathbf{0}_{I\times 1} \\ \mathbf{0}_{I\times 1} & \mathbf{1}_{I\times 1} \end{pmatrix} \begin{pmatrix} p_T \\ p_C \end{pmatrix}, \mathbf{\Sigma} = \sigma^2_{T,0} \begin{pmatrix} \mathbf{I}_{I\times I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \right. \\ + & \quad \left. \sigma^2_{C,0} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{I\times I} \end{pmatrix} + \textstyle\sum_{i=1}^{I} \sigma^2_{T,i} \mathbf{G}_i + \sum_{j=1}^{I} \sigma^2_{C,j} \mathbf{H}_j \right), \end{aligned} \tag{1.4}$$

where for $i, j = 1, \ldots, I$

$$\mathbf{G}_i = \left(\begin{array}{ccc|c} 0 & \cdots & 0 & \\ & \ddots & & \\ \vdots & \frac{1}{n_{T,i}} & \vdots & \mathbf{0}_{I\times I} \\ & \ddots & & \\ 0 & \cdots & 0 & \\ \hline \multicolumn{3}{c|}{\mathbf{0}_{I\times I}} & \mathbf{0}_{I\times I} \end{array}\right) \quad \mathbf{H}_j = \left(\begin{array}{c|ccc} \mathbf{0}_{I\times I} & \multicolumn{3}{c}{\mathbf{0}_{I\times I}} \\ \hline & 0 & \cdots & 0 \\ & & \ddots & \\ \mathbf{0}_{I\times I} & \vdots & \frac{1}{n_{C,j}} & \vdots \\ & & & \ddots \\ & 0 & \cdots & 0 \end{array}\right).$$

Notation $\mathbf{Y} \approx N(\boldsymbol{\mu}, \mathbf{\Sigma})$ means that $\mathbf{Y}$ has approximately normal distribution with mean $\boldsymbol{\mu}$ and variance matrix $\mathbf{\Sigma}$.

2 Point estimator of the vector of common probabilities of success

If we know the variance components $\sigma^2_{l,0}$ and $\sigma^2_{l,i}$ for $l \in \{T, C\}$ and $i = 1, \ldots, I$, the optimal estimator of the vector of the common probability of successful treatment would be

[1] Of course it is supposed that $\sigma^2_{l,0}$ is such that "practically" $0 < p_l + b_{l,i} < 1$. In simulations it is ensured with a proper choice of $\sigma^2_{l,0}$. In the case mentioned in section 4, it is $\sigma^2_{l,0} \in \left\{0, \frac{1}{4}\left(\frac{p_l}{3}\right)^2, \frac{1}{2}\left(\frac{p_l}{3}\right)^2, \frac{3}{4}\left(\frac{p_l}{3}\right)^2, \left(\frac{p_l}{3}\right)^2\right\}$ for $p_l \leq 0.5$ and $\sigma^2_{l,0} \in \left\{0, \frac{1}{4}\left(\frac{1-p_l}{3}\right)^2, \frac{1}{2}\left(\frac{1-p_l}{3}\right)^2, \frac{3}{4}\left(\frac{1-p_l}{3}\right)^2, \left(\frac{1-p_l}{3}\right)^2\right\}$ for $p_l > 0.5$. The unacceptable situations happened in case of $\sigma^2_{l,0} = \left(\frac{p_l}{3}\right)^2$ in 0.14 %, $\sigma^2_{l,0} = \frac{3}{2}\left(\frac{p_l}{3}\right)^2$ in 0.03 % and $\sigma^2_{l,0} = \frac{1}{2}\left(\frac{p_l}{3}\right)^2$ in 0.001 % from 100000 replications.

$$\begin{pmatrix}\hat{p}_T\\ \hat{p}_C\end{pmatrix} = \left(\begin{pmatrix}\mathbf{1}_{1\times I} & \mathbf{0}_{1\times I}\\ \mathbf{0}_{1\times I} & \mathbf{1}_{1\times I}\end{pmatrix}\boldsymbol{\Sigma}^{-1}\begin{pmatrix}\mathbf{1}_{I\times 1} & \mathbf{0}_{I\times 1}\\ \mathbf{0}_{I\times 1} & \mathbf{1}_{I\times 1}\end{pmatrix}\right)^{-1}\begin{pmatrix}\mathbf{1}_{1\times I} & \mathbf{0}_{1\times I}\\ \mathbf{0}_{1\times I} & \mathbf{1}_{1\times I}\end{pmatrix}\boldsymbol{\Sigma}^{-1}\mathbf{Y}. \quad (2.1)$$

So we replace unknown covariance matrix $\boldsymbol{\Sigma}$ by its estimator $\hat{\boldsymbol{\Sigma}}$ which we get if we replace the unknown variance components $\sigma^2_{l,0}$ and $\sigma^2_{l,i}$ by their estimators $\hat{\sigma}^2_{l,0}$ and $\hat{\sigma}^2_{l,i}$ for $l \in \{T, C\}$ and $i = 1, \ldots, I$. The estimators $\hat{\sigma}^2_{l,0}$ and $\hat{\sigma}^2_{l,i}$ we derive as follows.

From (1.1) we get

$$var(Y_{l,i}) = \frac{p_{l,i}(1-p_{l,i})}{n_{l,i}}$$

and using notation from (1.3) we obtain

$$\sigma^2_{l,i} = p_{l,i}(1-p_{l,i})$$

for $l = \{T, C\}$ and $i = 1, \ldots, I$. Now consider an estimator of $\sigma^2_{l,i}$ which was suggested by Agresti & Caffo (2000) as

$$\hat{\sigma}^2_{l,i} = \tilde{p}_{l,i}(1-\tilde{p}_{l,i}), \text{ where } \tilde{p}_{l,i} = \frac{X_{l,i}+2}{n_{l,i}+4}.$$

Than we can write

$$\hat{\sigma}^2_{l,i} = \frac{X_{l,i}+2}{n_{l,i}+4}\left(1 - \frac{X_{l,i}+2}{n_{l,i}+4}\right) \text{ for } l \in \{T, C\} \text{ and } i = 1, \ldots, I.$$

For estimation of $\sigma^2_{l,0}$ we use procedure suggested by Mandel & Paule (1982). The estimator $\hat{\sigma}^2_{l,0}$ for $l \in \{T, C\}$ we obtain as iterative solution of the following equations

$$\hat{\mu}^{MP}_l = \frac{\sum_{i=1}^{I} \frac{X_{l,i}}{n_{l,i}\hat{\sigma}^2_{l,0}+\hat{\sigma}^2_{l,i}}}{\sum_{j=1}^{I} \frac{n_{l,j}}{n_{l,j}\hat{\sigma}^2_{l,0}+\hat{\sigma}^2_{l,j}}}$$

$$\sum_{i=1}^{I} \frac{\left(\frac{X_{l,i}}{n_{l,i}} - \hat{\mu}^{MP}_l\right)^2}{\hat{\sigma}^2_{l,0} + \frac{\hat{\sigma}^2_{l,i}}{n_{l,i}}} = I - 1.$$

Finally we obtain a point estimator of the vector of the common probabilities of successful treatment

$$\hat{\mathbf{p}} = \begin{pmatrix}\hat{p}_T\\ \hat{p}_C\end{pmatrix} = \begin{pmatrix} \frac{\sum_{i=1}^{I} \frac{X_{T,i}}{n_{T,i}\hat{\sigma}^2_{T,0}+\hat{\sigma}^2_{T,i}}}{\sum_{j=1}^{I} \frac{n_{T,j}}{n_{T,j}\hat{\sigma}^2_{T,0}+\hat{\sigma}^2_{T,j}}} \\ \\ \frac{\sum_{i=1}^{I} \frac{X_{C,i}}{n_{C,i}\hat{\sigma}^2_{C,0}+\hat{\sigma}^2_{C,i}}}{\sum_{j=1}^{I} \frac{n_{C,j}}{n_{C,j}\hat{\sigma}^2_{C,0}+\hat{\sigma}^2_{C,j}}} \end{pmatrix}. \quad (2.2)$$

3 Interval estimator of the probability difference

As an estimatior of the covariance matrix of $\hat{\mathbf{p}}$ is commonly used

$$\hat{\mathbf{\Phi}} = \left(\begin{pmatrix} \mathbf{1}_{1\times I} & \mathbf{0}_{1\times I} \\ \mathbf{0}_{1\times I} & \mathbf{1}_{1\times I} \end{pmatrix} \hat{\mathbf{\Sigma}}^{-1} \begin{pmatrix} \mathbf{1}_{I\times 1} & \mathbf{0}_{I\times 1} \\ \mathbf{0}_{I\times 1} & \mathbf{1}_{I\times 1} \end{pmatrix} \right)^{-1}.$$

Kenward & Roger (1997) suggested an adjusted estimator $\hat{\mathbf{\Phi}}_A$

$$\hat{\mathbf{\Phi}}_A = \hat{\mathbf{\Phi}} + 2\hat{\mathbf{\Lambda}}, \tag{3.1}$$

where

$$\hat{\mathbf{\Lambda}} = \hat{\mathbf{\Phi}} \left\{ \sum_{k=1}^{2I+2} \sum_{l=1}^{2I+2} \hat{W}_{kl} (\hat{\mathbf{Q}}_{kl} - \hat{\mathbf{P}}_k \hat{\mathbf{\Phi}} \hat{\mathbf{P}}_l) \right\} \hat{\mathbf{\Phi}}$$

and $\hat{W}_{kl}$ is the (k,l)th element of estimator of the covariance matrix of the variance components $\sigma^2_{l,0}$ and $\sigma^2_{l,i}$ for $l \in \{T, C\}$ and $i = 1, \ldots, I$. The covariance matrix $\mathbf{W}$ can be obtained as inversion of the expected information matrix of the variance components REML estimators.

$$\mathbf{W}(\sigma^2_{T0}, \sigma^2_{T1}, \ldots, \sigma^2_{TI}, \sigma^2_{C0}, \sigma^2_{C1}, \ldots, \sigma^2_{CI}) = \mathbf{I}_F^{-1}(\sigma^2_{T0}, \sigma^2_{T1}, \ldots, \sigma^2_{TI}, \sigma^2_{C0}, \sigma^2_{C1}, \ldots, \sigma^2_{CI}).$$

Elements of $\mathbf{I}_F$ we get from

$$\{\mathbf{I}_F\}_{kl} = \frac{1}{2} \left[\{\mathbf{S}\}_{kl} - \mathrm{Tr}(2\mathbf{\Phi}\mathbf{Q}_{kl} - \mathbf{\Phi}\mathbf{P}_k\mathbf{\Phi}\mathbf{P}_l \right] \text{ for } k, l \in \{1, 2, \ldots, 2I+2\}.$$

And next using Kenward & Roger's procedure we have for $i = 1, \ldots, I$, $j = 1, \ldots, I$, $k = 1, \ldots, I$, $i \neq k$ and $j \neq k$ (in the same notation as in Kenward & Roger (1997))

$$\mathbf{P}_{T,0} = \begin{pmatrix} -\sum_{i=1}^{I} \left(\frac{n_{T,i}}{n_{T,i}\sigma^2_{T,0} + \sigma^2_{T,i}} \right)^2 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{P}_{C,0} = \begin{pmatrix} 0 & 0 \\ 0 & -\sum_{j=1}^{I} \left(\frac{n_{C,j}}{n_{C,j}\sigma^2_{C,0} + \sigma^2_{C,j}} \right)^2 \end{pmatrix},$$

$$\mathbf{P}_{T,i} = \begin{pmatrix} -\frac{n_{T,i}}{(n_{T,i}\sigma^2_{T,0} + \sigma^2_{T,i})^2} & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{P}_{C,j} = \begin{pmatrix} 0 & 0 \\ 0 & -\frac{n_{C,j}}{(n_{C,j}\sigma^2_{C,0} + \sigma^2_{C,j})^2} \end{pmatrix},$$

$$\mathbf{Q}_{T,0;T,0} = \begin{pmatrix} \sum_{i=1}^{I} \left(\frac{n_{T,i}}{n_{T,i}\sigma^2_{T,0} + \sigma^2_{T,i}} \right)^3 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{Q}_{C,0;C,0} = \begin{pmatrix} 0 & 0 \\ 0 & \sum_{j=1}^{I} \left(\frac{n_{C,j}}{n_{C,j}\sigma^2_{C,0} + \sigma^2_{C,j}} \right)^3 \end{pmatrix},$$

$$\mathbf{Q}_{T,0;T,i} = \begin{pmatrix} \frac{n^2_{T,i}}{(n_{T,i}\sigma^2_{T,0} + \sigma^2_{T,i})^3} & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{Q}_{C,0;C,j} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{n^2_{C,j}}{(n_{C,j}\sigma^2_{C,0} + \sigma^2_{C,j})^3} \end{pmatrix},$$

$$\mathbf{Q}_{T,i;T,i} = \begin{pmatrix} \frac{n_{T,i}}{(n_{T,i}\sigma^2_{T,0} + \sigma^2_{T,i})^3} & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{Q}_{C,j;C,j} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{n_{C,j}}{(n_{C,j}\sigma^2_{C,0} + \sigma^2_{C,j})^3} \end{pmatrix},$$

$$\mathbf{Q}_{T,0;C,0} = \mathbf{Q}_{T,0;C,j} = \mathbf{Q}_{T,i;C,0} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\mathbf{Q}_{T,i;T,k} = \mathbf{Q}_{C,j;C,k} = \mathbf{Q}_{T,i;C,j} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and nonzero elements of $\mathbf{S}$

$$\{\mathbf{S}\}_{T,0;T,0} = \sum_{i=1}^{I} \left(\frac{n_{T,i}}{n_{T,i}\sigma_{T,0}^2 + \sigma_{T,i}^2} \right)^2, \quad \{\mathbf{S}\}_{C,0;C,0} = \sum_{j=1}^{I} \left(\frac{n_{C,j}}{n_{C,j}\sigma_{C,0}^2 + \sigma_{C,j}^2} \right)^2,$$

$$\{\mathbf{S}\}_{T,0;T,i} = \frac{n_{T,i}}{\left(n_{T,i}\sigma_{T,0}^2 + \sigma_{T,i}^2\right)^2}, \qquad \{\mathbf{S}\}_{C,0;C,j} = \frac{n_{C,j}}{\left(n_{C,j}\sigma_{C,0}^2 + \sigma_{C,j}^2\right)^2},$$

$$\{\mathbf{S}\}_{T,i;T,i} = \frac{1}{\left(n_{T,i}\sigma_{T,0}^2 + \sigma_{T,i}^2\right)^2}, \qquad \{\mathbf{S}\}_{C,j;C,j} = \frac{1}{\left(n_{C,j}\sigma_{C,0}^2 + \sigma_{C,j}^2\right)^2}.$$

The matrices $\hat{\mathbf{W}}$, $\hat{\mathbf{Q}}$ and $\hat{\mathbf{P}}$ are estimators of $\mathbf{W}$, $\mathbf{Q}$ and $\mathbf{P}$ which we obtain by replacing unknown variance components $\sigma_{T,0}^2, \sigma_{T,1}^2, \ldots, \sigma_{T,I}^2, \sigma_{C,0}^2, \sigma_{C,1}^2, \ldots, \sigma_{C,I}^2$ by their estimators $\hat{\sigma}_{T,0}^2, \hat{\sigma}_{T,1}^2, \ldots, \hat{\sigma}_{T,I}^2, \hat{\sigma}_{C,0}^2, \hat{\sigma}_{C,1}^2, \ldots, \hat{\sigma}_{C,I}^2$.

Kenward & Roger (1997) also suggested an approximation of the random variable

$$\lambda F = \lambda(\hat{p_T} - \hat{p_C} - (p_T - p_C))^2 \left(\begin{pmatrix} 1 & -1 \end{pmatrix} \hat{\mathbf{\Phi}}_A \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right)^{-1}$$

by Fisher-Snedecor distribution with 1 and m degrees of freedom where

$$\lambda = \frac{m}{E^*(m-2)} \quad \text{and} \quad m = 4 + \frac{3}{\rho - 1}. \tag{3.2}$$

Also in the same notation as in Kenward & Roger (1997) all necessary quantities we get as

$$\rho = \frac{V^*}{2(E^*)^2}, \qquad E^* = \frac{1}{1 - A_2}, \qquad V^* = 2\left[\frac{1 + c_1 B}{(1 - c_2 B)^2 (1 - c_3 B)}\right],$$

$$c_1 = \frac{g}{3 + 2(1-g)}, \quad c_2 = \frac{1-g}{3 + 2(1-g)}, \quad c_3 = \frac{3-g}{3 + 2(1-g)},$$

$$g = \frac{2A_1 - 5A_2}{3A_2}, \qquad B = \frac{1}{2}(A_1 + 6A_2), \qquad \mathbf{\Theta} = \mathbf{L}(\mathbf{L}^T \hat{\mathbf{\Phi}} \mathbf{L})^{-1} \mathbf{L}^T, \quad \mathbf{L} = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

$$A_1 = \sum_{k=1}^{2I+2} \sum_{l=1}^{2I+2} W_{kl} \operatorname{Tr}(\mathbf{\Theta\Phi P}_k \mathbf{\Phi}) \operatorname{Tr}(\mathbf{\Theta\Phi P}_l \mathbf{\Phi}),$$

$$A_2 = \sum_{k=1}^{2I+2} \sum_{l=1}^{2I+2} W_{kl} \operatorname{Tr}(\mathbf{\Theta\Phi P}_k \mathbf{\Phi\Theta\Phi P}_l \mathbf{\Phi}).$$

Finally we get the $100 \times (1 - \alpha)$ confidence interval for the difference of probabilities of overall treatment effects $p_T - p_C$ in the form

$$\left\langle \hat{p}_T - \hat{p}_C - \sqrt{\lambda^{-1} \left(\begin{pmatrix} 1 & -1 \end{pmatrix} \hat{\Phi}_A \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right) F_{1,m}(\alpha)}\,, \right.$$

$$\left. \hat{p}_T - \hat{p}_C + \sqrt{\lambda^{-1} \left(\begin{pmatrix} 1 & -1 \end{pmatrix} \hat{\Phi}_A \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right) F_{1,m}(\alpha)} \right\rangle \cap \langle -1, 1 \rangle \tag{3.3}$$

where $F_{1,m}(\alpha)$ is the critical α-value of the Fisher-Snedecor distribution with 1 and m degrees of freedom, $\hat{p}_T$ and $\hat{p}_C$ are given in (2.2), $\hat{\Phi}_A$ is given in (3.1), λ and m can be obtained from (3.2).

4 Simulation results

The simulation study was focused on empirical coverage probabilities of the 95% interval estimator. To explore the behavior of the confidence interval for the difference of probabilities of overall treatment effects, the simulations were conducted for four main different settings. In all four settings the values of unknown parameters I, $n_{T,i}$ and $n_{C,i}$, $\sigma^2_{l,0}$ where $l \in \{T, C\}$, p_T and p_C were following. The number of center $I \in \{5, 10, 15, 20\}$, the number of subjects $n_{Ti}, n_{Ci} \in \{100, 50, 30, 15, 10\}$, the variance of random effects $\sigma^2_{l,0} \in \left\{0, \frac{1}{4}\left(\frac{p_l}{3}\right)^2, \frac{1}{2}\left(\frac{p_l}{3}\right)^2, \frac{3}{4}\left(\frac{p_l}{3}\right)^2, \left(\frac{p_l}{3}\right)^2\right\}$ for $p_l \leq 0.5$, $\sigma^2_{l,0} \in \left\{0, \frac{1}{4}\left(\frac{1-p_l}{3}\right)^2, \frac{1}{2}\left(\frac{1-p_l}{3}\right)^2, \frac{3}{4}\left(\frac{1-p_l}{3}\right)^2, \left(\frac{1-p_l}{3}\right)^2\right\}$ for $p_l > 0.5$ and both true probabilities of success $p_T, p_C \in \{0.05, 0.15, \ldots, 0.85, 0.95\}$. For each situation 5000 replications were made.

Except 95% confidence interval from (3.3) (CI) the simulations were also conducted for modified 95% confidence interval (MCI) according Wimmer & Witkovsky (2004). The modifications was made only for diagonal elements of $\mathbf{S}$ which were replaced by following expressions for $i = 1, \ldots, I$ and $j = 1, \ldots, I$

$$\{\mathbf{S}\}_{T,i;T,i} = \frac{n_{T,i}}{\sigma^4_{T,i}}\left(1 - \frac{2\sigma^2_{T,0}}{n_{T,i}\sigma^2_{T,0} + \sigma^2_{T,i}} + \frac{n_{T,i}\sigma^4_{T,0}}{(n_{T,i}\sigma^2_{T,0} + \sigma^2_{T,i})^2}\right),$$
$$\{\mathbf{S}\}_{C,j;C,j} = \frac{n_{C,j}}{\sigma^4_{C,j}}\left(1 - \frac{2\sigma^2_{C,0}}{n_{C,j}\sigma^2_{C,0} + \sigma^2_{C,j}} + \frac{n_{C,j}\sigma^4_{C,0}}{(n_{C,j}\sigma^2_{C,0} + \sigma^2_{C,j})^2}\right).$$

The empirical coverage probabilities are displayed using contour lines. The doted line is contour line matching 95% level. In all situations described below the empirical coverage probabilities of CI weren't below the nominal 95% level. However they weren't lower than 99.5% level, as is illustrated by the Figure 1. The white places in the graph means the empirical coverage probabilities were 1. This is mainly due to large width of CI for small numbers of subjects.

4.1 *Balanced situation across the trial*

In balanced situation across the trial the number of subjects in the i center in treated group $n_{T,i}$ is the same as the number of subjects in the i center in control group $n_{C,i}$ and is also the same as the number of subjects in the j center in treated group $n_{T,j}$ and control group $n_{C,j}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, I$. That is

$$n_{T,i} = n_{T,j} = n_{C,i} = n_{C,j} \text{ for } i, j = 1, \ldots, I.$$

For the MCI one can observe two areas with lower empirical coverage probability with cores at $p_T = 85\%$ and $p_C = 15\%$ and wise versa for $n_{T,i} = n_{C,i} = 15$ and $\sigma^2_{T,0} = \sigma^2_{C,0} = 0$. When the value of $n_{T,i}$ or $n_{C,i}$ is increased the cores move to the lower right corner and upper left corner (Figure 2). With growing value of I the area around cores grow too. One can also observe that with growing $\sigma^2_{T,0}$ or $\sigma^2_{C,0}$ these areas with lower empirical coverage probability fast grow too (Figure 3). The influence of growing I and $\sigma^2_{T,0}$ or $\sigma^2_{C,0}$ is approximately same in all considered situations.

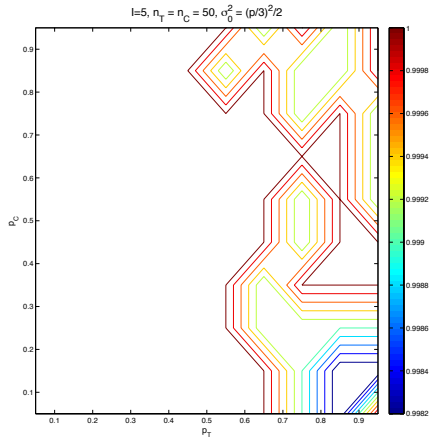


Figure 1: Contour lines of CI

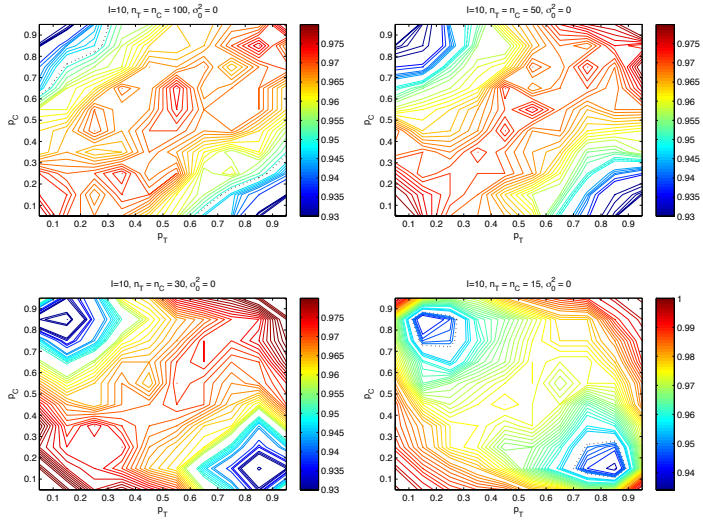


Figure 2: Contour lines of MCI in balanced situation

4.2 *Balanced situations across centers*

In balanced situation across centers the number of subjects in the i center in treated group $n_{T,i}$ is different from the number of subjects in the i center in control group $n_{C,i}$, but is the same as the number of subjects in the j center in treated group $n_{T,j}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, I$. That is

$$n_{T,i} = n_{T,j} \neq n_{C,i} = n_{C,j} \text{ for } i, j = 1, \ldots, I.$$

As is illustrated in Figure 4 the second simulated situation showed similar results, only cores of areas move according to the difference between $n_{T,i}$ and $n_{C,i}$.

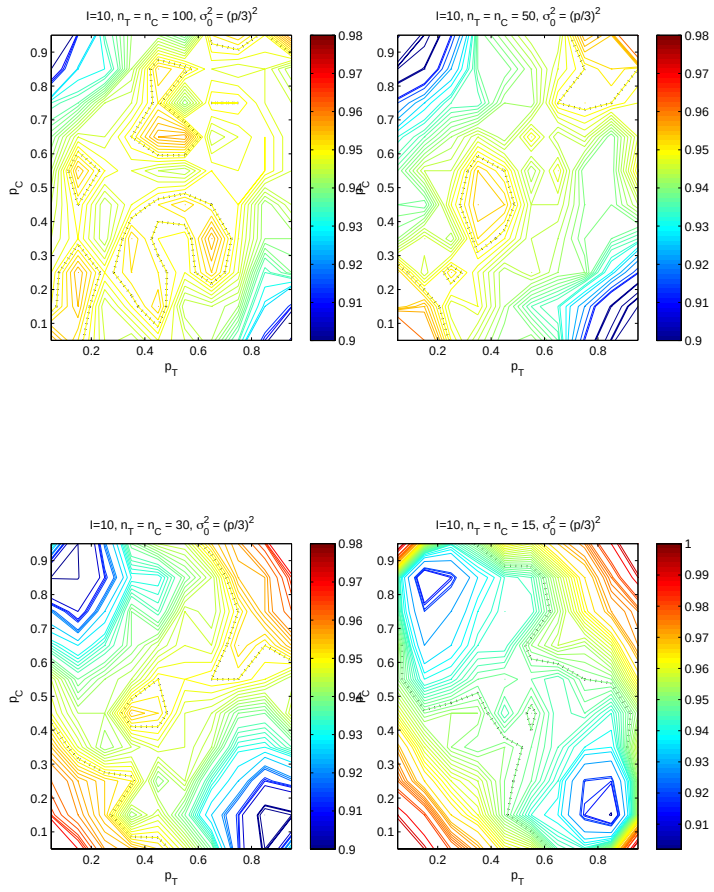


Figure 3: Contour lines of MCI in balanced situation and $\sigma^2_{l,0} = \left(\frac{p_l}{3}\right)^2$, $\sigma^2_{l,0} = \left(\frac{1-p_l}{3}\right)^2$

4.3 *Balanced situations across groups*

In balanced situation across groups the number of subjects in the i center in treated group $n_{T,i}$ is the same as the number of subjects in the i center in control group $n_{C,i}$, but is different from the number of subjects in the j center in treated group $n_{T,j}$ and control group $n_{C,j}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, I$. That is

$$n_{T,i} \neq n_{T,j}, n_{C,i} \neq n_{C,j} \wedge n_{T,i} = n_{C,i} \text{ for } i, j = 1, \ldots, I.$$

The third situation does not have big influence on results in compare to previous situations. The results depended mainly on the highest value between $n_{T,i}$ or $n_{C,i}$ (Figure 5).

4.4 *Unbalanced situations*

In unbalanced situation the number of subjects in the i center in treated group $n_{T,i}$ can be different from the number of subjects in the i center in control group $n_{C,i}$ and

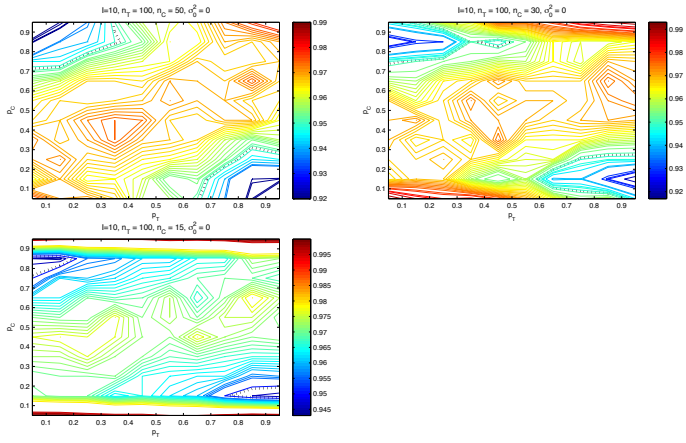


Figure 4: Contour lines of MCI in balanced situations across centers

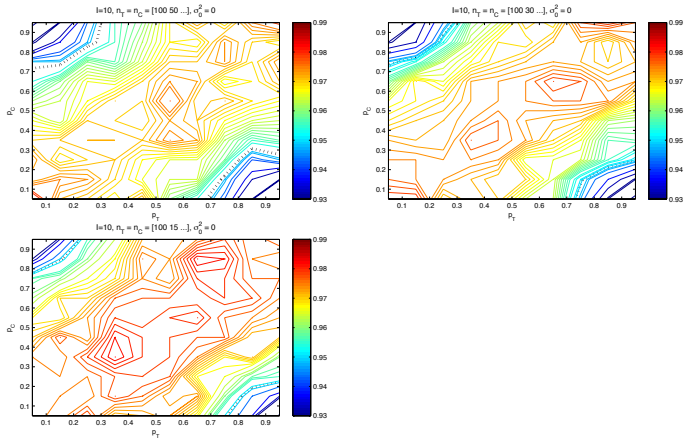


Figure 5: Contour lines of MCI in balanced situations across groups

can also be different from the number of subjects in the j center in treated group $n_{T,j}$ and control group $n_{C,j}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, I$. As expected this situation combine results of two previous situations.

In all situations the width of CI was approximately two times the width of MCI. In further work the comparing will be extended to higher values of I, because the width of CI seems to be getting smaller with growing number of centers in which is trial conducted along with the empirical coverage probability above the nominal 95% level. The reason for conducting simulations for $\sigma^2_{C,0} = 0$ was a possibility of comparison these results with GLMM approach mentioned in section 4.5.

4.5 *Comments on GLMM approach*
The standard approach to the presented problem which can be found in Whitehead (2002) is based on generalized linear mixed model (GLMM). Let us consider random variables Z_{ij} which have Bernoulli distribution with success probability p_{ij} for $i = 1, \ldots, I$ and $j = 1, \ldots, n_i$ ($n_i = n_{T,i} + n_{C,i}$). Suppose the following GLMM model

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha + \beta_{0i} + \beta_1 U_{ij} + \nu_{1i} U_{ij} \tag{4.1}$$

where $U_{ij} = 0$ for the control group, $U_{ij} = 1$ for the treated group, ν_{1i} is a random effect of the ith center and suppose that $\nu_{1i} \sim N(0, \sigma^2_{T,0})$. In this model there is only one random effect of the treatment and no random effect for the control group, i.e. $\sigma^2_{C,0}$. The overall treatment effect is in this model measured by a log odds ratio β_1.

According to our opinion the disadvantage of this approach is its computational behavior for small numbers of subjects n_i and lower probabilities. In these cases the calculation of β_1 do not converge. The greater number of centers I is the greater numbers of subjects n_i have to be. The comparison of the our and the GLMM approach will be subject of further work.

References

[1] A. Agresti and B. Caffo, *Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures*, The American Statistician **54** (2000).

[2] M. G. Kenward and J. H. Roger, *Small sample inference for fixed effects from restricted maximum likelihood*, Biometrics **53** (1997).

[3] J. Mandel and R. C. Paule, *Consensus values and weighting factors*, Journal of Research of the National Bureau of Standarts **87** (1982).

[4] A. Whitehead, "Meta-Analysis of Controlled Clinical Trials", John Wiley & Sons Ltd., Chichester, 2002.

[5] Wimmer G., Witkovský V.: *Konfidenčné intervaly pre efekt ošetrenia v klinických pokusoch*, In: "ROBUST 2004. Sborník prací 13. letní školy JČMF ROBUST 2004 uspořádané Jednotou českých matematiku a fyziku za podpory KPMS 2004 v Třešti", JČMF, Praha, 2004.

ACTA UNIVERSITATIS MATTHIAE BELII, series MATHEMATICS 18 (2011), 17–27
Also available at http://actamath.savbb.sk
ISBN 978-80-557-0160-8

A Distance-based Extension of the Majority Judgement Voting System

Edurne Falcó
PRESAD Research Group, Dept. of Applied Economics, University of Valladolid, Spain
E-mail: edurne@eco.uva.es

José Luis García-Lapresta
PRESAD Research Group, Dept. of Applied Economics, University of Valladolid, Spain
E-mail: lapresta@eco.uva.es

Abstract

It is common knowledge that the political voting systems suffer inconsistencies and paradoxes such that Arrow has shown in his well-known Impossibility Theorem. Recently Balinski and Laraki have introduced a new voting system called Majority Judgement (MJ) which tries to solve some of these limitations. In MJ voters have to asses the candidates through linguistic terms belonging to a common language. From this information, MJ assigns as the collective assessment the lower median of the individual assessments and it considers a sequential tie-breaking method for ranking the candidates. The present paper provides an extension of MJ focused to reduce some of the drawbacks that have been detected in MJ by several authors. The model assigns as the collective assessment a label that minimizes the distance to the individual assessments. In addition, we propose a new tie-breaking method also based on distances.

Keywords voting systems, Majority Judgement, distances.

MSC (2010) 91B12, 91B14.

Received: August 30, 2010; Revised: August 30, 2010; Accepted: November 2, 2010.

1 Introduction

Social Choice Theory shows that there does not exist a completely acceptable voting system for electing and ranking alternatives. The well-known Arrow Impossibility Theorem [1] proves with mathematic certainty that no voting system simultaneously fulfills certain desirable properties[1].

Recently Balinski and Laraki [2, 4, 5] have proposed a voting system called Majority Judgement (MJ) which tries to avoid these unsatisfactory results and allows the voters to assess the alternatives through linguistic labels, as *Excellent*, *Very good*, *Good*, ..., instead of rank order the alternatives. Among all the individual assessments

[1] Any voting rule that generates a collective weak order from every profile of weak orders, and satisfies independence of irrelevant alternatives and unanimity is necessarily dictatorial, insofar as there are at least three alternatives and three voters.

given by the voters, MJ chooses the median as the collective assessment. Balinski and Laraki also describe a tie-breaking process which compares the number of labels above the collective assessment and those below of it.

These authors also have an experimental analysis of MJ [3] carried out in Orsay during the 2007 French presidential election. In that paper the authors show some interesting properties of MJ and they advocate that this voting system is easily implemented and that it avoids the necessity for a second round of voting.

Desirable properties and advantages have been attributed to MJ against the classical Arrow framework of preferences' aggregation. Among them are the possibility that voters show more faithfully and properly their opinions than in the conventional voting systems, anonymity, neutrality, independence of irrelevant alternatives, etc. However, some authors (see Felsenthal and Machover [6], García-Lapresta and Martínez-Panero [7] and Smith [9]) have shown several paradoxes and inconsistencies of MJ.

In this paper we propose an extension of MJ which diminishes some of the MJ inconveniences. The approach of the paper is distance-based, both for generating a collective assessment of each alternative and in the tie-breaking process that provides a weak order on the set of alternatives. As in MJ we consider that individuals assess the alternatives through linguistic labels and we propose as the collective assessment a label that minimizes the distance to the individual assessments. These distances between linguistic labels are induced by a metric of the parameterized Minkowski family. Depending on the specific metric we use, the discrepancies between the collective and the individual assessments are weighted in a different manner, and the corresponding outcome can be different.

The paper is organized as follows. In Section 2, the MJ voting system is formally explained. Section 3 introduces our proposal, within a distance-based approach. Specifically, the election of the collective assessment for each alternative and the tie-breaking method are introduced. In Section 4 we include two illustrative examples showing the influence of the metric used in the proposed method and its differences with respect to MJ and Range Voting (Smith [9]). Finally, in Section 5 we collect some conclusions.

2 Majority Judgement

We consider[2] a finite set of voters $V = \{1, \ldots, m\}$, with $m \geq 2$, who evaluate a finite set of alternatives $X = \{x_1, \ldots, x_n\}$, with $n \geq 2$. Each alternative is assessed by each voter through a linguistic term belonging to an ordered finite scale $L = \{l_1, \ldots, l_g\}$, with $l_1 < \cdots < l_g$ and granularity $g \geq 2$. Each voter assesses the alternatives in an independent way and these assessments are collected by a matrix $\left(v_j^i\right)$, where $v_j^i \in L$ is the assessment that the voter i gives to the alternative x_j.

MJ chooses for each alternative the median of the individual assessment as the collective assessment. To be precise, the single median when the number of voters is odd and the lower median in the case that the number of voters is even. We denote with $l(x_j)$ the collective assessment of the alternative x_j. Given that several alternatives might share the same collective assessment, Balinski and Laraki [2] propose a sequential tie-breaking process. This can be described through the following terms (see García-Lapresta and Martínez-Panero [7]):

[2]The current notation is similar to the one introduced by García-Lapresta and Martínez-Panero [7]. This allows us to describe the MJ process, presented by Balinski and Laraki [2], in a more precise way.

$$N^+(x_j) = \#\{i \in V \mid v_j^i > l(x_j)\}, \quad N^-(x_j) = \#\{i \in V \mid v_j^i < l(x_j)\}$$

and

$$t(x_j) = \begin{cases} -1, & \text{if } N^+(x_j) < N^-(x_j), \\ 0, & \text{if } N^+(x_j) = N^-(x_j), \\ 1, & \text{if } N^+(x_j) > N^-(x_j). \end{cases}$$

Taking into account the collective assessments and the previous indices, we define a weak order[3] $\succeq$ on X in the following way: $x_j \succeq x_k$ if and only if one of the following conditions hold:

1. $l(x_j) > l(x_k)$.
2. $l(x_j) = l(x_k)$ and $t(x_j) > t(x_k)$.
3. $l(x_j) = l(x_k)$, $t(x_j) = t(x_k) = 1$ and $N^+(x_j) > N^+(x_k)$.
4. $l(x_j) = l(x_k)$, $t(x_j) = t(x_k) = 1$, $N^+(x_j) = N^+(x_k)$ and $N^-(x_j) \leq N^-(x_k)$.
5. $l(x_j) = l(x_k)$, $t(x_j) = t(x_k) = 0$ and $m - N^+(x_j) - N^-(x_j) \geq m - N^+(x_k) - N^-(x_k)$.
6. $l(x_j) = l(x_k)$, $t(x_j) = t(x_k) = -1$ and $N^-(x_j) < N^-(x_k)$.
7. $l(x_j) = l(x_k)$, $t(x_j) = t(x_k) = -1$, $N^-(x_j) = N^-(x_k)$ and $N^+(x_j) \geq N^+(x_k)$.

The asymmetric and symmetric parts of $\succeq$ are defined in the usual way:

$$x_j \succ x_k \Leftrightarrow \text{not } x_k \succeq x_j$$
$$x_j \sim x_k \Leftrightarrow (x_j \succeq x_k \text{ and } x_k \succeq x_j).$$

Next example of how MJ works is shown.

Example 2.1. Consider three alternatives x_1, x_2 and x_3 that are evaluated by seven voters through a set of six linguistic terms $L = \{l_1, \ldots, l_6\}$, the same set used in MJ [3], whose meaning is shown in Table 1. The assessments obtained for each alternative

l_1	l_2	l_3	l_4	l_5	l_6
To reject	*Poor*	*Acceptable*	*Good*	*Very good*	*Excellent*

Table 1: Meaning of the linguistic terms

are collected and ranked from the lowest to the highest in Table 2. For ranking the three alternatives, first we take the median of the individual assessments, that will be the collective assessment for each one of the mentioned alternatives: $l(x_1) = l_5$, $l(x_2) = l_4$ and $l(x_3) = l_4$. Given that x_1 has the best collective assessment, it will be the one ranked in first place. However, the alternatives x_2 and x_3 share the same

[3]A *weak order* (or *complete preorder)* is a complete and transitive binary relation.

x_1	l_1	l_1	l_3	l_5	l_5	l_5	l_6
x_2	l_1	l_4	l_4	l_4	l_4	l_5	l_6
x_3	l_1	l_3	l_4	l_4	l_5	l_5	l_5

Table 2: Assessments of Example 2.1

collective assessment, we need to turn to the tie-breaking process, where we obtain $N^+(x_2) = 2$, $N^-(x_2) = 1$ and $t(x_2) = 1$; $N^+(x_3) = 3$, $N^-(x_3) = 2$ and $t(x_3) = 1$. Since both alternatives have the same t ($t(x_2) = t(x_3) = 1$), we should compare their N^+: $N^+(x_2) = 2 < 3 = N^+(x_3)$. Therefore, the alternative x_3 defeats the alternative x_2, and the final order is $x_1 \succ x_3 \succ x_2$.

3 Distance-based method

In this section the alternative method to MJ that we propose through a distance-based approach is introduced. The first step for ranking the alternatives is to assign a collective assessment $l(x_j) \in L$ to each alternative $x_j \in X$. For its calculation, the vectors $(v_j^1, \ldots, v_j^m)$ that collect all the individual assessments for each alternative $x_j \in X$ are taken into account.

The proposal, that is detailed below, involves how to choose a $l(x_j) \in L$ that minimizes the distance between the vector of individual assessments $(v_j^1, \ldots, v_j^m)$ and the vector $(l(x_j), \ldots, l(x_j)) \in L^m$. The election of that term is performed in an independent way for each alternative. This guarantees the fulfillment of the *independence of irrelevant alternatives principle*[4].

Once a collective assessment $l(x_j)$ has been associated with each alternative $x_j \in X$, we rank the alternatives according to the ordering of L. Given the possible existence of ties, we also propose a sequential tie-breaking process based on the difference between the distance of $l(x_j)$ to the assessments higher than $l(x_j)$ and the distance of $l(x_j)$ to the assessments lower than $l(x_j)$.

3.1 Distances

A *distance* or *metric* on $\mathbb{R}^m$ is a mapping $d : \mathbb{R}^m \times \mathbb{R}^m \longrightarrow \mathbb{R}$ that fulfills the following conditions for all $(a_1, \ldots, a_m), (b_1, \ldots, b_m), (c_1, \ldots, c_m) \in \mathbb{R}^m$:

1. $d((a_1, \ldots, a_m), (b_1, \ldots, b_m)) \geq 0$.
2. $d((a_1, \ldots, a_m), (b_1, \ldots, b_m)) = 0 \Leftrightarrow (a_1, \ldots, a_m) = (b_1, \ldots, b_m)$.
3. $d((a_1, \ldots, a_m), (b_1, \ldots, b_m)) = d((b_1, \ldots, b_m), (a_1, \ldots, a_m))$.
4. $d((a_1, \ldots, a_m), (b_1, \ldots, b_m)) \leq$
   $d((a_1, \ldots, a_m), (c_1, \ldots, c_m)) + d((c_1, \ldots, c_m), (b_1, \ldots, b_m))$.

Given a distance $d : \mathbb{R}^m \times \mathbb{R}^m \longrightarrow \mathbb{R}$, the *distance on L^m induced by d* is the mapping $\bar{d} : L^m \times L^m \longrightarrow \mathbb{R}$ defined by

$$\bar{d}((l_{a_1}, \ldots, l_{a_m}), ((l_{b_1}, \ldots, l_{b_m})) = d((a_1, \ldots, a_m), (b_1, \ldots, b_m)).$$

[4]This principle says that the relative ranking between two alternatives would only depend on the preference or assessments on these alternatives and must not be affected by other alternatives, that must be irrelevant on that comparison.

An important class of distances in $\mathbb{R}^m$ is constituted by the family of *Minkowski distances* $\{d_p \mid p \geq 1\}$, which are defined by

$$d_p((a_1,\ldots,a_m),(b_1,\ldots,b_m)) = \left(\sum_{i=1}^{m} |a_i - b_i|^p\right)^{\frac{1}{p}},$$

for all $(a_1,\ldots,a_m)$, $(b_1,\ldots,b_m) \in \mathbb{R}^m$.

We choose this family due to the fact that it is parameterized and it includes from the well-known *Manhattan* ($p = 1$) and *Euclidean* ($p = 2$) distances, to the limit case, the *Chebyshev* distance ($p = \infty$). The possibility of choosing among different values of $p \in (1,\infty)$ gives us a very flexible method, and we can choose the most appropriate p according to the objectives we want to achieve with the election.

Given a Minkowski distance on $\mathbb{R}^m$, we consider the induced distance on L^m which works with the assessments vector through the subindexes of the corresponding labels:

$$\bar{d}_p((l_{a_1},\ldots,l_{a_m}),(l_{b_1},\ldots,l_{b_m})) = d_p((a_1,\ldots,a_m),(b_1,\ldots,b_m)).$$

Remark 3.1. The ordinal scale of linguistic terms we use, L, is just a finite scale whose consecutive terms are equidistant. Following Balinski and Laraki [2], each term of the scale has associated a linguistic label. What matters is not the name of the label but the position of the label in the ordinal scale. This is the reason we consider the number of changes we need for going from a term to another one[5]. In this sense, the distance between two labels' vectors is based on the number of positions that we need to cover to go from one to another, in each of its components. To move from l_{a_i} to l_{b_i} we need to cover $|a_i - b_i|$ positions. For instance between l_5 and l_2 we need to cover $|5-2| = 3$ positions: from l_5 to l_4, from l_4 to l_3 and from l_3 to l_2.

3.2 Election of a collective assessment for each alternative

Our proposal is divided into several stages. First we assign a collective assessment $l(x_j) \in L$ to each alternative $x_j \in X$ which minimizes the distance between the vector of the individual assessments, $(v_j^1,\ldots,v_j^m) \in L^m$, and the vector of m replicas of the desired collective assessment, $(l(x_j),\ldots,l(x_j)) \in L^m$.

For this, first we establish the set $L(x_j)$ of all the labels $l_k \in L$ satisfying

$$\bar{d}_p((v_j^1,\ldots,v_j^m),(l_k,\ldots,l_k)) \leq \bar{d}_p((v_j^1,\ldots,v_j^m),(l_h,\ldots,l_h)),$$

for each $l_h \in L$, where $(l_h,\ldots,l_h)$ and $(l_k,\ldots,l_k)$ are the vectors of m replicas of l_h and l_k, respectively. Thus, $L(x_j)$ consists of those labels that minimize the distance to the vector of individual assessments. Notice that $L(x_j) = \{l_r,\ldots,l_{r+s}\}$ is always an interval, because it contains all the terms from l_r to l_{r+s}, where $r \in \{1,\ldots,g\}$ and $0 \leq s \leq g - r$. Two different cases are possible:

1. If $s = 0$, then $L(x_j)$ contains a single label, which will automatically be the collective assessment $l(x_j)$ of the alternative x_j.
2. If $s > 0$, then $L(x_j)$ has more than one label. In order to select the most suitable label of $L(x_j)$, we now introduce $L^*(x_j)$, the set of all the labels $l_k \in L(x_j)$ that fulfill

$$\bar{d}_p((l_k,\ldots,l_k),(l_r,\ldots,l_{r+s})) \leq \bar{d}_p((l_h,\ldots,l_h),(l_r,\ldots,l_{r+s})),$$

[5]This is not exactly the same that identifying each linguistic label with a number.

for all $l_h \in L(x_j)$, where $(l_k, \ldots, l_k)$ and $(l_h, \ldots, l_h)$ are the vectors of $s+1$ replicas of l_k and l_h, respectively.

(a) If the cardinality of $L(x_j)$ is odd, then $L^*(x_j)$ has a unique label, the median term, that will be the collective assessment $l(x_j)$.

(b) If the cardinality of $L(x_j)$ is even, then $L^*(x_j)$ has two different labels, the two median terms. In this case, similarly to the proposal of Balinski and Laraki [2], we consider the lowest label in $L^*(x_j)$ as the collective assessment $l(x_j)$.

It is worth pointing out two different cases when we are using induced Minkowski distances.

1. If $p = 1$, we obtain the same collective assessments that those given by MJ, the median[6] of the individual assessments. However, the final results are not necessarily the same that in MJ because we use a different tie-breaking process, as is shown later.
2. If $p = 2$, each collective assessment is the closest label to the "mean" of the individual assessments[7], which is the one chosen by the *Range Voting* (RV) method[8] (see Smith [9]).

It is interesting to note that when we choose $p \in (1, 2)$, we find situations where the collective assessment is located between the median and the "mean". This allows us to avoid some of the problems associated with MJ and RV. See García-Lapresta and Martínez-Panero [7] for a different proposal based on centered OWA operators (Yager [10]).

3.3 Tie-breaking method

Usually there exist more alternatives than linguistic terms, so it is very common to find several alternatives sharing the same collective assessment. But irrespectively of the number of alternatives, it is clear that some of them may share the same collective assessment, even when the individual assessments are very different. For these reasons it is necessary to introduce a tie-breaking method that takes into account not only the number of individual assessments above or below the obtained collective assessment (as in MJ), but the positions of these individual assessments in the ordered scale associated with L.

As mentioned above, we will calculate the difference between two distances: one between $l(x_j)$ and the assessments higher than $l(x_j)$ and another one between $l(x_j)$ and the assessments lower than the $l(x_j)$. Let $\boldsymbol{v}_j^+$ and $\boldsymbol{v}_j^-$ the vectors composed by the assessments v_j^i from $\left(v_j^1, \ldots, v_j^m\right)$ higher and lower than the term $l(x_j)$, respectively. First we calculate the two following distances:

$$D^+(x_j) = \bar{d}_p\left(\boldsymbol{v}_j^+, (l(x_j), \ldots, l(x_j))\right),$$
$$D^-(x_j) = \bar{d}_p\left(\boldsymbol{v}_j^-, (l(x_j), \ldots, l(x_j))\right),$$

[6] It is more precise to speak about the interval of medians, because if the assessments' vector has an even number of components, then there are more than one median. See Monjardet [8].

[7] The chosen label is not exactly the arithmetic mean of the individual assessments, because we are working with a discrete spectrum of linguistic terms and not in the continuous one of the set of real numbers.

[8] RV works with a finite scale given by equidistant real numbers, and it ranks the alternatives according to the arithmetic mean of the individual assessments.

where the number of components of $(l(x_j), \ldots, l(x_j))$ is the same that in $\boldsymbol{v}_j^+$ and in $\boldsymbol{v}_j^-$, respectively (obviously, the number of components of $\boldsymbol{v}_j^+$ and $\boldsymbol{v}_j^-$ can be different).

Once these distances have been determined, a new index $D(x_j) \in \mathbb{R}$ is calculated for each alternative $x_j \in X$: the difference between the two previous distances:

$$D(x_j) = D^+(x_j) - D^-(x_j).$$

By means of this index, we provide a kind of compensation between the individual assessments that are bigger and smaller than the collective assessment, taking into account the position of each assessment in the ordered scale associated with L.

For introducing our tie-breaking process, we finally need the distance between the individual assessments and the collective one:

$$E(x_j) = \bar{d}_p\left((v_j^1, \ldots, v_j^m), (l(x_j), \ldots, l(x_j))\right).$$

Notice that for each alternative $x_j \in X$, $E(x_j)$ minimizes the distance between the vector of individual assessments and the linguistic labels in L, such as has been considered above in the definition of $L(x_j)$.

The use of the index $E(\cdot)$ is important in the tie-breaking process because if two alternatives share the same couple $(l(\cdot), D(\cdot))$, the alternative with the lower $E(\cdot)$ is the alternative whose individual assessments are more concentrated around the collective assessment, i.e., the consensus is higher.

Summarizing, for ranking the alternatives we will consider the following triplet

$$T(x_j) = (l(x_j), D(x_j), E(x_j)) \in L \times \mathbb{R} \times [0, \infty)$$

for each alternative $x_j \in X$.

The sequential process works in the following way:

1. We rank the alternatives through the collective assessments $l(\cdot)$. The alternatives with higher collective assessments will be preferred to those with lower collective assessments.
2. If several alternatives share the same collective assessment, then we break the ties through the $D(\cdot)$ index. The alternatives with a higher $D(\cdot)$ will be preferred.
3. If there are still ties, we break them through the $E(\cdot)$ index, in such a way such that the alternatives with a lower $E(\cdot)$ will be preferred.

Formally, the sequential process can be introduced by means of the lexicographic weak order $\succeq$ on X defined by $x_j \succeq x_k$ if and only if

1. $l(x_j) \geq l(x_k)$ or
2. $l(x_j) = l(x_k)$ and $D(x_j) > D(x_k)$ or
3. $l(x_j) = l(x_k)$, $D(x_j) = D(x_k)$ and $E(x_j) \leq E(x_k)$.

Remark 3.2. Although it is possible that ties still exist, whenever two or more alternatives share $T(\cdot)$, these cases are very unusual when considering metrics with $p > 1$.[9] For instance, consider seven voters that assess two alternatives x_1 and x_2

x_1	l_2	l_2	l_2	l_2	l_4	l_4	l_6
x_2	l_2	l_2	l_2	l_2	l_3	l_5	l_6

Table 3: Individual assessments

by means of the set of linguistic terms given in Table 1. Table 3 includes these assessments arranged from the lowest to the highest labels.

It is easy to see that for $p = 1$ we have $T(x_1) = T(x_2) = (l_2, 8, 8)$, then $x_1 \sim x_2$ (notice that MJ and RV also provide a tie). However, if $p > 1$, the tie disappears. So, we have $x_2 \succ x_1$, excepting for $p \in (1.179, 1.203)$, where $x_1 \succ x_2$.

4 Two illustrative examples

This section focus on how the election of the parameter p is relevant in the final ranking of the alternatives. We show this fact through two different examples. The first one considers a case where the median of the individual assessments is the same for all the alternatives. And the second one considers a situation where the mean of the individual assessments' subindexes is the same for all the alternatives. In both examples we use the set of six linguistic terms $L = \{l_1, \ldots, l_6\}$ whose meaning is shown in Table 1.

As mentioned above, the sequential process for ranking the alternatives is based on the triplet $T(x_j) = (l(x_j), D(x_j), E(x_j))$ for each alternative $x_j \in X$. However, by simplicity, in the following examples we only show the first two components, $(l(x_j), D(x_j))$. In these examples we also obtain the outcomes provided by MJ and RV.

Example 4.1. Table 4 includes the assessments given by six voters to four alternatives x_1, x_2, x_3 and x_4 arranged from the lowest to the highest labels.

x_1	l_1	l_2	l_4	l_4	l_4	l_6
x_2	l_1	l_1	l_3	l_4	l_6	l_6
x_3	l_2	l_2	l_2	l_4	l_5	l_6
x_4	l_1	l_1	l_4	l_5	l_5	l_5

Table 4: Assessments in Example 4.1

Notice that the mean of the individual assessments' subindexes is the same for the four alternatives, $\frac{21}{6} = 3.5$. Since RV ranks the alternatives according to this mean, it produces a tie $x_1 \sim x_2 \sim x_3 \sim x_4$. However, it is clear that this outcome might not seem reasonable, and that other rankings could be justified. Using MJ, where $l(x_1) = l(x_4) = l_4 > l_3 = l(x_2) > l_2 = l(x_3)$ and, according to the MJ tie-breaking process, we have $t(x_1) = -1 < 1 = t(x_4)$. Thus, MJ produces the outcome $x_4 \succ x_1 \succ x_2 \succ x_3$.

We now consider the distance-based procedure for seven values of p. In Table 5 we can see the influence of these values on $(l(x_j), D(x_j))$, for $j = 1, 2, 3$. The corresponding rankings are included in Table 6.

For $p = 1$, we have $T(x_1) = (l_4, -3, 7)$, $T(x_2) = (l_3, 10, 11)$, $T(x_3) = (l_2, 9, 9)$ and $T(x_4) = (l_4, -3, 9)$. Then, we obtain the ranking $x_1 \succ x_4 \succ x_2 \succ x_3$, a different

[9]The Manhattan metric ($p = 1$) produces more ties than the other metrics in the Minkowski family because of the simplicity of its calculations.

	$p=1$	$p=1.25$	$p=1.5$	$p=1.75$	$p=2$	$p=5$	$p=10$
x_1	$(l_4,-3)$	$(l_4,-2.375)$	$(l_4,-2.008)$	$(l_4,-1.770)$	$(l_3,1.228)$	$(l_3,0.995)$	$(l_3,0.000)$
x_2	$(l_3,10)$	$(l_3,2.264)$	$(l_3,1.888)$	$(l_3,1.669)$	$(l_3,1.530)$	$(l_3,1.150)$	$(l_3,1.072)$
x_3	$(l_2,9)$	$(l_3,2.511)$	$(l_3,2.254)$	$(l_3,2.104)$	$(l_3,2.010)$	$(l_4,-0.479)$	$(l_4,-0.232)$
x_4	$(l_4,-3)$	$(l_4,-2.815)$	$(l_4,-2.682)$	$(l_4,-2.585)$	$(l_3,0.777)$	$(l_3,0.199)$	$(l_3,0.089)$

Table 5: $(l(x_j), D(x_j))$ in Example 4.1

MJ	$p=1$	$p=1.25$	$p=1.5$	$p=1.75$	$p=2$	$p=5$	$p=10$
x_4	x_1	x_1	x_1	x_1	x_3	x_3	x_3
x_1	x_4	x_4	x_4	x_4	x_2	x_2	x_2
x_2	x_2	x_3	x_3	x_3	x_1	x_1	x_1
x_3	x_3	x_2	x_2	x_2	x_4	x_4	x_4

Table 6: Rankings in Example 4.1

outcome than obtained using MJ. For $p=1.25$, $p=1.5$ and $p=1.75$, we obtain $x_1 \succ x_4 \succ x_3 \succ x_2$; and for $p=2$, $p=5$ and $p=10$, we have $x_3 \succ x_2 \succ x_1 \succ x_4$.

Example 4.2. Similarly to the previous example, Table 7 includes the assessments given by seven voters to three alternatives x_1, x_2 and x_3 arranged from the lowest to the highest labels.

x_1	l_1	l_1	l_2	l_3	l_6	l_6	l_6
x_2	l_2	l_3	l_3	l_3	l_6	l_6	l_6
x_3	l_3	l_3	l_3	l_3	l_4	l_4	l_4

Table 7: Assessments in Example 4.2

Clearly, the individual assessments of the three alternatives share the same median, l_3. According to the MJ tie-breaking process, we have

$$\begin{aligned} &t(x_1) = 0 < 1 = t(x_2) = t(x_3) \\ &N^+(x_1) = N^+(x_2) = N^+(x_3) = 3 \\ &N^-(x_3) = 0 < 1 = N^-(x_2) < 3 = N^-(x_1). \end{aligned}$$

Thus, MJ produces the outcome $x_3 \succ x_2 \succ x_1$.

This outcome does not seem logical, because x_2 has a clear advantage over x_3. On the other hand, RV ranks order the alternatives as follows: $x_2 \succ x_1 \succ x_3$, since the mean of the individual assessments' subindexes are 3.571, 4.143 and 3.429 for x_1, x_2 and x_3, respectively.

We now consider the distance-based procedure for seven values of p, the same considered in the previous example. Table 8 shows the influence of these values on $(l(x_j), D(x_j))$, for $j=1,2,3$.

Notice that in this example the same ranking is obtained for all the considered values of p: $x_2 \succ x_1 \succ x_3$. This outcome coincides with RV, and it is more consistent than that obtained by MJ.

	$p=1$	$p=1.25$	$p=1.5$	$p=1.75$	$p=2$	$p=5$	$p=10$
x_1	$(l_3, 4)$	$(l_3, 3.168)$	$(l_3, 2.702)$	$(l_4, -1.475)$	$(l_4, -1.332)$	$(l_4, -1.000)$	$(l_4, -0.986)$
x_2	$(l_3, 8)$	$(l_4, 0.975)$	$(l_4, 0.922)$	$(l_4, 0.868)$	$(l_4, 0.818)$	$(l_4, 0.455)$	$(l_4, 0.235)$
x_3	$(l_3, 3)$	$(l_3, 2.408)$	$(l_3, 2.080)$	$(l_3, 1.873)$	$(l_3, 1.732)$	$(l_3, 1.246)$	$(l_3, 1.116)$

Table 8: $(l(x_j), D(x_j))$ in Example 4.2

5 Concluding remarks

In this paper we have presented an extension of the Majority Judgement voting system developed by Balinski and Laraki [2, 3, 4, 5]. This extension is based on a distance approach but it also uses linguistic labels to evaluate the alternatives. We choose as the collective assessment for each alternative a label that minimizes the distance to the individual assessments. It is important to note that our proposal coincides in this aspect with Majority Judgement whenever the Manhattan metric is used.

We also provide a tie-breaking process through the distances between the individual assessments higher and lower than the collective one. This process is richer than the one provided by Majority Judgement, which only counts the number of alternatives above or below the collective assessment, irrespectively of what they are. We also note that our tie-breaking process is essentially different to Majority Judgement even when the Manhattan metric is considered.

It is important to note that using the distance-based approach we pay attention to all the individual assessments that have not been chosen as the collective assessment. With the election of a specific metric of the Minkowski family we are deciding how to evaluate these other assessments. We may distinguish four cases:

1. If $p = 1$, the collective assessment is just the median label and no other individual assessment is relevant in this stage. However, in the tie-breaking process, all the individual assessments are taken into account, each of them with the same weight or importance.
2. If $p = 2$, the collective assessment is a kind of "mean" of the individual assessments because it minimizes the Euclidean distance to the individual assessments. In this stage all the voters have the same importance. However, in the tie-breaking process we are giving more importance to the assessments that are further to the collective assessment than to those labels that are closer to the collective assessment.
3. If $p \in (1, 2)$, we are moving between the two previous cases. The collective assessment gives less importance to the median of the individual assessments and more to the other assessments whenever p increases. In the tie-breaking process higher values of p give more importance to extreme individual assessments and the smaller p, the more egalitarian the procedure will be (with the individual assessments).
4. If $p \in (2, \infty)$, the collective assessment depends on the extreme assessments more than on the central ones, the higher p, the more intense this dependency will be. If they are balanced in both sides, this has no effect in the final outcome. But if one of the sides has more extreme opinions, the collective label will go close to them. The tie-breaking process gives also more weight to the extreme opinions.

These aspects provide flexibility to our extension and it allows to devise a wide class of voting systems that may avoid some of the drawbacks related to Majority Judgement and Range Voting without losing their good features. This becomes specially interesting when the value of the parameter p in the Minkowski family belongs to the open interval $(1, 2)$, since $p = 1$ and $p = 2$ correspond to the Manhattan and the Euclidean metrics, respectively, just the metrics used in Majority Judgement and Range Voting. For instance, the election of $p = 1.5$ allows us to have a kind of compromise between both methods.

As shown in the previous examples, when the value of parameter p increases, the distance-based procedure focuses more and more on the extreme assessments. However, if the individual assessments are well balanced on both sides, the outcome is not very affected by the parameter p.

In further research we will analyze the properties of the presented extension of Majority Judgement within the Social Choice framework.

Acknowledgments

The authors are grateful to the members of the PRESAD Research Group, Jorge Alcalde-Unzu and two anonymous referees for their comments and suggestions. The financial support of the Spanish Ministerio de Ciencia e Innovación (projects ECO2009-07332, ECO2009-12836 and ECO2008-03204-E/ECON), ERDF and Junta de Castilla y León (Consejería de Educación, Projects VA092A08 and GR99) are also gratefully acknowledged.

References

[1] K. J. Arrow, "Social Choice and Individual Values". 2nd ed., John Wiley and Sons, New York, 1963.

[2] M. Balinski and R. Laraki, *A theory of measuring, electing and ranking*, Proceedings of the National Academy of Sciences of the United States of America **104** (2007), 8720-8725.

[3] M. Balinski and R. Laraki, *Election by Majority Judgement: Experimental evidence.* Ecole Polytechnique – Centre National de la Recherche Scientifique, Cahier no. 2007-28, 2007.

[4] M. Balinski and R. Laraki, The Majority Judgement, `https://sites.google.com/site/ridalaraki/majority-judgment`, 2007.

[5] M. Balinski and R. Laraki, "Majority Judgment. Measuring, Ranking, and Electing". The MIT Press, Cambridge MA, 2011.

[6] D. S. Felsenthal and M. Machover, *The Majority Judgement voting procedure: a critical evaluation*, Homo Oeconomicus **25** (2008), 319-334.

[7] J. L. García-Lapresta and M. Martínez-Panero, *Linguistic-based voting through centered OWA operators*, Fuzzy Optimization and Decision Making **8** (2009), 381-393.

[8] B. Monjardet, *Mathèmatique Sociale and Mathematics. A case study: Condorcet's effect and medians*, Electronic Journal for History of Probability and Statistics **4** (2008).

[9] W. D. Smith, On Balinski & Laraki's Majority Judgement median-based range-like voting scheme, `http://rangevoting.org/MedianVrange.html`, 2007.

[10] R. R. Yager, *Centered OWA operators*, Soft Computing **11** (2007), 631-639.

ACTA UNIVERSITATIS MATTHIAE BELII, series MATHEMATICS 18 (2011), 29–33
Also available at http://actamath.savbb.sk
ISBN 978-80-557-0160-8

$S(2,1)$-labeling of graphs with cyclic structure

Karina Chudá
Faculty of Mathematics, Physics and Informatics,
Comenius University in Bratislava, Mlynská dolina,
842 48 Bratislava, Slovakia
E-mail: karina_chuda@medusa.sk

Abstract

We present the values of the σ-number of two infinite classes of graphs with cyclic structure, namely prisms and the Isaacs graphs, depending on their order.

Keywords Channel assignment, $S(2,1)$-labeling, σ-number, action of cyclic group, prism, Isaacs snark.

MSC (2010) 05C15, 05C78.

Received: August 5, 2010; Revised: October 2, 2010; Accepted: October 3, 2010.

1 Introduction

In this paper, we give a brief review of our work concerning the σ-number of two infinite classes of graphs with cyclic structure – prisms and Isaacs graphs. The paper also contains several research announcements of results to be published in the extended version of the paper containing detailed proofs.

The $S(2,1)$-labeling problem is a variation of the $L(2,1)$-labeling problem or, more general, of the $L(d_1,d_2)$-labeling problem – a survey on the $L(d_1,d_2)$-labeling problem is given by Calamoneri in [1]. An *r-$S(2,1)$-labeling* of a graph G is a mapping from the vertex-set of G to the cyclic group $\mathbb{Z}_r$ such that every pair of vertices adjacent in G has labels at least 2 apart in $\mathbb{Z}_r$ and simultaneously every pair of vertices at distance 2 in G has distinct labels in $\mathbb{Z}_r$. The *σ-number* of a graph G is the smallest r such that G admits an r-$S(2,1)$-labeling. A survey on the σ-number is presented by Yeh in [5].

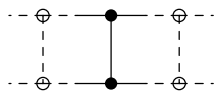


Figure 1: Prism

Although a prism can be regarded as the Cartesian product of a cycle and of the complete graph of order 2, a different equivalent description is more suitable for us. A

prism Y_m, for $m \geq 3$, is a graph consisting of m segments isomorphic to the complete graph K_2 arranged into a cycle, where both vertices of a given segment are connected to the corresponding vertices of the preceding and of the succeeding segments – a plain given segment and its incidence with the dashed preceding and the dashed succeeding segments are shown in Figure 1. The Isaacs graphs form a superclass of the Isaacs snarks constructed by Isaacs in [4], as only odd members of the class are snarks. *The Isaacs graph* J_m, for $m \geq 3$, is a graph consisting of m segments isomorphic to the claw $K_{1,3}$ arranged into a cycle, where the leaves of a given segment are connected to the leaves of the preceding and of the succeeding segments in the manner indicated in Figure 2 – the given segment is plain while the preceding and the succeeding segments are dashed.

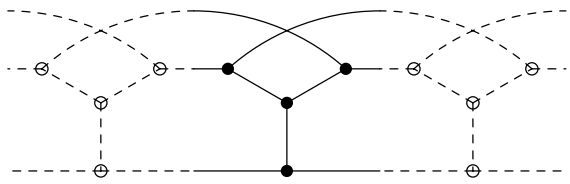


Figure 2: The Isaacs graph

2 Strategy

In order to determine the σ-number of a graph, we investigate whether or not the graph admits an r-$S(2,1)$-labeling. It follows from the definition of the σ-number that if a graph does admit an r-$S(2,1)$-labeling, then its σ-number is at most r and, conversely, if a graph does not admit an r-$S(2,1)$-labeling, then its σ-number is at least $r+1$.

To find out whether a graph G admits an r-$S(2,1)$-labeling, we proceed in two steps. First, we cover G with overlapping subgraphs in such a way that for every pair of vertices adjacent in G, as well as for every pair of vertices at distance 2 in G, there is at least one of the covering subgraphs in which the vertices have the same distances as in G. Second, we take such r-$S(2,1)$-labelings of the subgraphs that the labels of vertices in the overlapping parts match, so that they naturally form an r-$S(2,1)$-labeling of G. In the second step, we obtain an r-$S(2,1)$-labeling of G because of the choice of the covering subgraphs in the first step. Conversely, if no such r-$S(2,1)$-labelings of the subgraphs exist that the labels of vertices in the overlapping parts match, then no r-$S(2,1)$-labeling of G exists.

In the following, let G_m stand for Y_m or J_m. Due to the cyclic structure of G_m, we can choose m isomorphic subgraphs to cover G_m. To fulfil the conditions for the cover, we take m subgraphs isomorphic to the graph Y_C shown in Figure 3 to cover Y_m and m subgraphs isomorphic to the graph J_C shown in Figure 4 to cover J_m. Now, let G_C stand for Y_C whenever $G_m = Y_m$ and let G_C stand for J_C whenever $G_m = J_m$. We cover the i-th, the $(i+1)$-st and the $(i+2)$-nd segment of G_m with the i-th copy of G_C; throughout, indices are taken modulo m. Now, we have to determine all r-$S(2,1)$-labelings of G_C and to find out whether there are m-tuples of these r-$S(2,1)$-labelings which can be assigned to the copies of G_C in such a way that

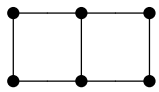


Figure 3: The graph Y_C

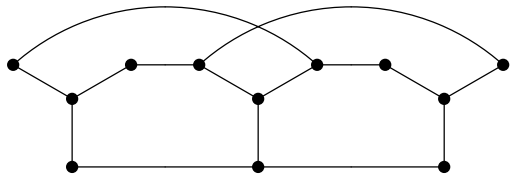


Figure 4: The graph J_C

the labels of vertices in the overlapping parts match. We call an r-$S(2,1)$-labeling l_i of G_C *concatenable* with the r-$S(2,1)$-labeling l_{i+1} of G_C if l_{i+1} labels the vertices of the left and of the central segments of G_C with the same labels as l_i labels the corresponding vertices of the central and of the right segments of G_C. Furthermore, we call a cyclic m-tuple of r-$S(2,1)$-labelings of G_C *concatenable* if the i-th r-$S(2,1)$-labeling of G_C is concatenable with the $(i+1)$-st r-$S(2,1)$-labeling of G_C for every i. Since there is a one-to-one correspondence between the r-$S(2,1)$-labelings of G_m and the concatenable cyclic m-tuples of r-$S(2,1)$-labelings of G_C, the existence of a concatenable cyclic m-tuple of r-$S(2,1)$-labelings of G_C is a sufficient and a necessary condition for the existence of an r-$S(2,1)$-labeling of G_m. We define a directed graph $D_r(G_C)$ whose vertex-set is formed by the r-$S(2,1)$-labelings of G_C and which has an arc from one r-$S(2,1)$-labeling of G_C to an another precisely when the former r-$S(2,1)$-labeling of G_C is concatenable with the latter. It follows from the definition of $D_r(G_C)$ that the existence of a concatenable cyclic m-tuple of r-$S(2,1)$-labelings of G_C is equivalent to the existence of a closed walk of length m in $D_r(G_C)$.

Since the number of r-$S(2,1)$-labelings of G_C might be very large for both $G_C = Y_C$ and $G_C = J_C$, we use the action of translations and reflections of $\mathbb{Z}_r$ and the action of the automorphism group of G_C to partition the r-$S(2,1)$-labelings of G_C into orbits. Subsequently, we only consider the representatives of these orbits. Having determined all representatives, we can reconstruct all r-$S(2,1)$-labelings of G_C by applying automorphisms of G_C and translations and reflections of $\mathbb{Z}_r$.

Since G_m is a cubic graph, it contains the claw $K_{1,3}$ as a subgraph. Observe that the σ-number of $K_{1,3}$ is equal to 6. Since the σ-number of a subgraph does not exceed the σ-number of the supergraph, we conclude that the σ-number of G_m is at least 6. Thus, we only have to investigate the r-$S(2,1)$-labelings of G_m and G_C for r at least 6.

Due to the large number of distinct r-$S(2,1)$-labelings of G_C the proofs are rather involved, although straightforward. We omit them by referring to the author's PhD.

Thesis [2].

3 Prisms

For $r = 6$, we have twelve distinct r-$S(2,1)$-labelings of Y_C. It is easy to see that in the directed graph $D_6(Y_C)$, there are closed walks of length m if and only if $m \equiv 0 \pmod 3$. Therefore $\sigma(Y_m) = 6$ for $m \equiv 0 \pmod 3$ and $\sigma(Y_m) \geq 7$ for $m \not\equiv 0 \pmod 3$.

For $r = 7$, we have 196 distinct r-$S(2,1)$-labelings of Y_C. From the directed graph $D_7(Y_C)$, we construct a directed voltage graph D' of order 28 with voltages in $\mathbb{Z}_7$ such that there exists a closed walk of length m in $D_7(Y_C)$ if and only if there exists a closed walk of length m with net voltage of 0 in D'. More details on voltage graphs can be found in [3] by Gross and Tucker. It can be shown that in D', there are closed walks of length m with net voltage of 0 if and only if $m \notin \{4,5,8,11\}$. Therefore $\sigma(Y_m) = 7$ for $m \not\equiv 0 \pmod 3$ and $m \notin \{4,5,8,11\}$ and $\sigma(Y_m) \geq 8$ for $m \in \{4,5,8,11\}$.

For $r = 8$, we can find r-$S(2,1)$-labelings of Y_C that can form a concatenable cyclic m-tuple of r-$S(2,1)$-labelings of Y_C for every $m \geq 3$. Therefore $\sigma(Y_m) = 8$ for $m \in \{4,5,8,11\}$.

Summarizing previous results, we obtain the following theorem.

Theorem 3.1. *[2, Theorem 3.1] Let Y_m be a prism of order $2m$, for $m \geq 3$. Then*

$$\sigma(Y_m) = \begin{cases} 6 & \text{for } m \equiv 0 \pmod 3 \\ 7 & \text{for } m \not\equiv 0 \pmod 3 \text{ and } m \notin \{4,5,8,11\} \\ 8 & \text{for } m \in \{4,5,8,11\}. \end{cases}$$

4 The Isaacs graphs

For $r = 6$, we have no r-$S(2,1)$-labelings of J_C. Therefore $\sigma(J_m) \geq 7$ for every $m \geq 3$.

For $r = 7$, we have 1176 distinct r-$S(2,1)$-labelings of J_C. From the directed graph $D_7(J_C)$, we construct its adjacency matrix A. Observe that a closed walk of length m in $D_7(J_C)$ exists if and only if there exists a non-zero diagonal element in A^m. By calculating the powers of A, we can see that there are non-zero diagonal elements in A^m if and only if $m \notin \{3,4,5,7,8,9,11\}$. Therefore $\sigma(J_m) = 7$ for $m \notin \{3,4,5,7,8,9,11\}$ and $\sigma(J_m) \geq 8$ for $m \in \{3,4,5,7,8,9,11\}$.

For $r = 8$, we can find r-$S(2,1)$-labelings of J_C that can form a concatenable cyclic m-tuple of r-$S(2,1)$-labelings of J_C for every $m \geq 3$. Therefore $\sigma(J_m) = 8$ for $m \in \{3,4,5,7,8,9,11\}$.

Summarizing previous results, we obtain the following theorem.

Theorem 4.1. *[2, Theorem 4.1] Let J_m be an Isaacs graph of order $4m$, for $m \geq 3$. Then*

$$\sigma(J_m) = \begin{cases} 7 & \text{for } m \notin \{3,4,5,7,8,9,11\} \\ 8 & \text{for } m \in \{3,4,5,7,8,9,11\}. \end{cases}$$

5 Remarks

The presented strategy can be used to calculate the σ-number of other graphs with cyclic structure. Besides this, after minor modifications, it can be used to determine also the σ-number of graphs with nearly cyclic structure – for instance the σ-number of the generalized Blanuša snarks investigated in a subsequent paper.

References

[1] T. Calamoneri, *The $L(h,k)$-Labelling Problem: A Survey and Annotated Bibliography*, The Computer Journal **49** (2006), 585-608.

[2] K. Chudá, "Labeling Graphs with a Condition at Distance 2", PhD. Thesis, Comenius University in Bratislava, 2010.

[3] J. L. Gross and T. W. Tucker, "Topological Graph Theory", John Wiley & Sons, New York, 1987.

[4] R. Isaacs, *Infinite Families of Nontrivial Trivalent Graphs Which Are Not Tait Colorable*, Amer. Math. Monthly **28** (1975), 221-239.

[5] R. K. Yeh, *A Survey on Labeling Graphs with a Condition at Distance Two*, Discrete Math. **306** (2006), 1217-1231.

ACTA UNIVERSITATIS MATTHIAE BELII, series MATHEMATICS 18 (2011), 35–45
Also available at http://actamath.savbb.sk
ISBN 978-80-557-0160-8

Central limit theorem on MV-algebras

Jana Kelemenová
Faculty of Natural Sciences, Matej Bel University
Department of Mathematics Tajovského 40
974 01 Banská Bystrica, Slovakia
E-mail: kelemen@fpv.umb.sk

Mária Kuková
Faculty of Natural Sciences, Matej Bel University
Department of Mathematics Tajovského 40
974 01 Banská Bystrica, Slovakia
E-mail: kukova@fpv.umb.sk

Abstract

The aim is to approve the Central limit theorem on MV-algebras by the new approach, presented by Riečan in [1]. We use the observable as a distribution function instead of the σ-homomorphism. The main idea is in local representation of σ-algebras.
The following theorem is proved: Let M be a σ-complete MV-algebra with product, $m : M \longrightarrow \langle 0, 1\rangle$ be a σ- additive state, $(x_n)_n$ be a sequence of independent, equally distributed, square integrable strong observables. Hence $E[x_1] = E[x_2] = ... = a$, $\sigma(x_1) = \sigma(x_2) = ... = \sigma$. Then for any t $\in R$

$$\lim_{n\to\infty} m\left(\frac{\frac{1}{n}\sum_{i=1}^{n} x_i - a}{\frac{\sigma}{\sqrt{n}}}((-\infty, t))\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{u^2}{2}}du$$

Keywords MV-algebra, central limit theorem.

MSC (2010) 60F05, 60A86, 08A72.

Received: July 5, 2010; Revised: November 2, 2010; Accepted: November 2, 2010.

Seeing that the paper [1] haven't been published yet, let us summarize the main ideas of Riečan's new approach to probability on MV-algebras. We will do it in the first two paragraphs. In Paragraph 3 will we present our main result - a proof of the central limit theorem.

1 MV-algebras

We shall use an excellent characterization of MV-algebras given by D . Mundici ([3]) by the help of l-groups. An l-group is an algebraic system

$$(G, +, \leq)$$

such that

$(G, +)$ is an Abelian group,

$(G, \leq)$ is a partially ordered set being a lattice,

$a \leq b \Longrightarrow a + c \leq b + c$ for and a, b, c in G.

A typical example of an MV-algebra is the unit interval $\langle 0, 1\rangle$ of real numbers with two binary operations $\oplus, \odot$,

$$a \oplus b = (a + b) \wedge 1$$
$$a \odot b = (a + b - 1) \vee 0$$

with the partial ordering $\leq$, and two fixed elements $0, 1$ (0 is the least element of $\leq$, 1 is the greatest element of $\leq$).

Generally the situation is analogous.

Definition 1.1. An MV-algebra is an algebraic system

$$(M, \oplus, \odot, \leq, 0, u),$$

where

$$M = \langle 0, u\rangle$$

is an interval in an l-group $G = (G, +, \leq), 0$ is the neutral element of G (i.e. $a+0 = a$ for any $a \in G$), u is the strong unit of G (i.e. to any $a \in G$ there exists $n \in N$ such that $a \leq u + u + ... + u(n\text{-times})$),

$$\begin{aligned} a \oplus b &= (a + b) \wedge u, \\ a \odot b &= (a + b - u) \vee 0. \end{aligned}$$

In two-valued logic Boolean algebras can be represented e.g. by charateristic functions $\chi_A : \Omega \to \{0, 1\}$ where,

$$\chi_A(v) = \begin{cases} 1, & \text{if } v \in A, \\ 0, & \text{if } v \notin A. \end{cases}$$

In multi-valued logic MV-algebras instead of two-valued functions

$$\chi_A : \Omega \to \{0, 1\},$$

multivalued functions

$$\mu_A : \Omega \to \langle 0, 1\rangle$$

are considered. Evidently

$$\chi_A \oplus \chi_B = \chi_{A\cup B}, \chi_A \odot \chi_B = \chi_{A\cap B}.$$

Hence $\mu_A \odot \mu_B$ can be considered as the conjunction, $\mu_A \oplus \mu_B$ as the disjunction, $1 - \mu_A$ as the negation.

Unlike in [4], Riečan uses more general definition of a state and an observable in [1].

Definition 1.2. A state on an MV-algebra M is a mapping $m : M \to \langle 0, 1\rangle$ satisfying the following conditions:

(i) $m(u) = 1, m(0) = 0$;

(ii) $a_n \nearrow a \Longrightarrow m(a_n) \nearrow m(a)$;

(iii) (iii) $a_n \searrow a \Longrightarrow m(a_n) \searrow m(a)$.

Definition 1.3. Let $\mathcal{J} = \{(-\infty, t); t \in R\}$. An observable on M is any mapping $x : \mathcal{J} \to M$ satisfying the conditions:

(i) $t_n \nearrow \infty \Longrightarrow x((-\infty, t_n)) \nearrow u$;

(ii) $t_n \searrow -\infty \Longrightarrow x\left((-\infty, t_n)\right) \searrow 0$;

(iii) $t_n \nearrow t \Longrightarrow x((-\infty, t_n)) \nearrow x((-\infty, t))$.

Thus, a distribution function can be created as a composition of the observable and the state (see also [1]).

Theorem 1.4. *Let $m : M \to \langle 0, 1\rangle$ be a state, $x : \mathcal{J} \to M$ be an observable. Define $F : R \to \langle 0, 1\rangle$ by the formula*

$$F(t) = m(x((-\infty, t))), t \in R$$

Then F has the following properties:

(i) F is non-decreasing;

(ii) $\lim_{t\to\infty} F(t) = 1$;

(iii) $\lim_{t\to-\infty} F(t) = 0$;

(iv) F is left continuous in any point $t \in R$.

Proof is straightforward.

Recall that in the Kolmogorov theory the mean value $E(\xi)$ of a random variable $\xi : (\Omega, \mathcal{S}, P) \to R$ is defined as an integral

$$E(\xi) = \int_\Omega \xi dP.$$

Let $g : R \to R$ be a Borel measurable function. The transformation formula states

$$E(g \circ \xi) = \int_R g(t) dF(t),$$

where F is the distribution function of ξ. Therefore

$$\begin{aligned} E(\xi) &= \int_R t dF(t), \\ E(\xi^2) &= \int_R t^2 dF(t), \\ \sigma(\xi)^2 &= E(\xi^2) - E(\xi)^2 = \int_R t^2 dF(t) - (\int_R t dF(t))^2. \end{aligned}$$

Definition 1.5. An observable $x : \mathcal{J} \to M$ is called to be integrable if there exists

$$E(x) = \int_R t dF(t),$$

where $F : R \to \langle 0, 1\rangle$ is the distribution function of the observable x. The observable x is square integrable, if there exists

$$\int_R t^2 dF(t).$$

2 MV algebras with product

Similarly as in [2] we shall work with a further binary operation called product (see also [4] and [5]). Recall that an MV algebra in an interval is an l-group $(G, +, \leq)$ and denote by $-$ the inverse group operation, i.e. $a - a = 0$ for any $a \in G$.

Definition 2.1. An MV-algebra with product is a pair$(M, \cdot)$, where M is an MV-algebra and $\cdot$ is a commutative and associative binary operation on M satisfying the following conditions:

(i) $u \cdot a = a$ for any $a \in M$;

(ii) $a \cdot ((b - c) \vee 0) = (a \cdot b - a \cdot c) \vee 0$ for any $a, b, c \in M$.

The first problem we shall solve is the construction of sums of observables. In the classical case

$$(\xi + \eta) = g \circ T, \text{ where } T = (\xi, \eta)\,;\ \Omega \longrightarrow R^2,\ g : R^2 \longrightarrow R,\ g(u, v) = u + v.$$

Therefore

$$(\xi + \eta)^{-1} = T^{-1} \circ g^{-1}.$$

Of course,

$$g^{-1}\left((-\infty, t)\right) = \{(u, v)\,; u + v < t\} = D_t.$$

Hence instead of T^{-1} we shall construct a mapping

$$h : \{D_t; t \in R\} \longrightarrow M.$$

We will use the notation

$$x\left(\langle a, b)\right) = x\left((-\infty, b)\right) - x\left((-\infty, a)\right)$$

Definition 2.2. For $t \in R$ put

$$h\left(D_t\right) = \bigvee_{n=1}^{\infty} \bigoplus_{i=-\infty}^{\infty} \left[x\left(\langle \frac{i-1}{2^n}, \frac{i}{2^n} \right) \right) \ . \ y\left(\left(-\infty, t - \frac{i}{2^n} \right) \right) \right]$$

Theorem 2.3. *Let M be a σ-complete MV-algebra with product, $x, y : \mathcal{J} \longrightarrow M$ be observables, then the mapping $z : \mathcal{J} \longrightarrow M$ defined by*

$$z\left((-\infty, t)\right) = h\left(D_t\right),$$

is an observable.

Proof. The properties (i) and (ii) of Definition 1.3 follows by the inequalities

$$x\left((-\infty,k)\right).y\left((-\infty,k)\right)\leq z\left((-\infty,2k)\right)\leq x\left((-\infty,2k)\right)\vee y\left((-\infty,k)\right)$$

holding for any $k\in Z$. By [4, Proposition 3.2] we have

$$c.\left(a\vee b\right)=\left(c.a\right)\vee\left(c.b\right)$$

for any $a,b,c\in M$. Therefore, if $t_k\nearrow t$, then

$$\bigoplus_{k=1}^{\infty}z\left((-\infty,t_k)\right)=\bigoplus_{k=1}^{\infty}\bigvee_{n=1}^{\infty}\bigoplus_{i=-\infty}^{\infty}x\left(\langle\frac{i-1}{2^n},\frac{i}{2^n}\right)\right).y\left(\left(-\infty,t_k-\frac{i}{2^n}\right)\right)=$$

$$=\bigvee_{n=1}^{\infty}\bigoplus_{i=-\infty}^{\infty}x\left(\langle\frac{i-1}{2^n},\frac{i}{2^n}\right)\right)\cdot\left(\bigoplus_{k=1}^{\infty}y\left(\left(-\infty,t_k-\frac{i}{2^n}\right)\right)\right)=z\left((-\infty,t)\right).$$

□

Definition 2.4. Let M be a σ-complete MV-algebra with product, $x,y:\mathcal{J}\longrightarrow M$ be observables. Then its sum is defined by the formula

$$\left(x+y\right)\left((-\infty,t)\right)=h\left(D_t\right)=h\left(g^{-1}\left((-\infty,t)\right)\right).$$

Remark 2.5. Evidently Theorem 2.3 and Definition 2.4 can be generalized for n summs and $x_1,...,x_n:\mathcal{J}\longrightarrow M$:

$$D_t^n=\left\{(m_1,...,m_n);m_1+...+m_n<t\right\},M_n=\left\{D_t^n;\,t\in R\right\},$$

$$g_n\left(m_1,...,m_n\right)=m_1+...+m_n$$

$$h_n:M_n\longrightarrow M,x_1+...+x_n\left((-\infty,t)\right)=h_n\left(D_t^n\right)=h_n\left(g_n^{-1}\left((-\infty,t)\right)\right),$$

hence

$$\left(\sum_{i=1}^{n}x_i\right)\left((-\infty,t)\right)=h_n\left(g_n^{-1}\left((-\infty,t)\right)\right),$$

$$\sum_{i=1}^{n}x_i=h_n\circ g_n^{-1}.$$

The second problem solved in the section is a characterization of independence of observables in MV-algebras.

Definition 2.6. Observables $x_1,...,x_n$ are independent, if for any $t_1,...,t_n\in R$

$$m\left(h_n\left((-\infty,t_1)\times(-\infty,t_2)\times...\times(-\infty,t_n)\right)\right)=$$

$$=m\left(x_1\left((-\infty,t_1)\right)\right).m\left(x_2\left((-\infty,t_2)\right)\right).\,...\,.m\left(x_n\left((-\infty,t_n)\right)\right).$$

Remark 2.7. In the Kolmogorov theory also the following two assertions hold:

1. If $F : R \longrightarrow \langle 0,1\rangle$ is a distribution function, then there exists exactly one probability measure $\lambda_F : \mathcal{B}(R) \longrightarrow \langle 0,1\rangle$ such that

$$\lambda_F(\langle a,b)) = F(b) - F(a)$$

for any $a, b \in R,\ a \leq b$

2. If $F_1, ..., F_n : R \longrightarrow \langle 0,1\rangle$ are distribution functions, then there exists exactly one probability measure $\lambda_F : \mathcal{B}(R^n) \longrightarrow \langle 0,1\rangle$ such that

$$\lambda_F(A_1 \times A_1 \times ... \times A_n) = \lambda_{F_1}(A_1).\lambda_{F_2}(A_2)....\lambda_{F_n}(A_n)$$

for any $A_1, A_2, ..., A_n \in \mathcal{B}(R)$. Notation $\lambda_F = \lambda_{F_1} \times \lambda_{F_2} \times ... \times \lambda_{F_n}$.

Definition 2.8. An observable $x : \mathcal{J} \longrightarrow M$ is called strong, if

$$\langle a,b) \cap \langle c,d) = \emptyset \Longrightarrow (x(\langle a,b)).\alpha).(x(\langle c,d)).\beta) = 0$$

for any $\alpha, \beta \in M$.

Definition 2.9. A state $m : M \longrightarrow \langle 0,1\rangle$ is called σ-additive, if

$$m\left(\bigoplus_{n=1}^{\infty} a_n\right) = \sum_{n=1}^{\infty} m(a_n)$$

whenever $a_n \wedge a_m = 0\ \ (n \neq m), a_n \in M$.

Theorem 2.10. *Let M be σ- complete MV-algebra with product, m be a σ-additive state, $x_1, \ldots, x_n$ independent strong observables. Then*

$$\lambda_{F_1} \times ... \times \lambda_{F_n}(D_t^n) = m(h(D_t^n))$$

for any $t \in R$.

Proof. We have

$$\begin{aligned}
D_t &= \bigcup_{n=1}^{\infty}\bigcup_{i=-\infty}^{\infty} \langle\frac{i-1}{2^n}, \frac{i}{2^n}\rangle \times \left(-\infty, t - \frac{i}{2^n}\right)\\
h(D_t) &= \bigvee_n \bigoplus_i x_1\left(\langle\frac{i-1}{2^n}, \frac{i}{2^n}\right)\right) \cdot x_2\left(\left(-\infty, t - \frac{i}{2^n}\right)\right)\\
\lambda_{F_1}(\langle a,b)) &= F_1(b) - F_1(a)\\
F_1(b) &= m(x_1((-\infty, b)))
\end{aligned}$$

We shall present it for $n = 2$. Of course, since x_1 is strong,

$$x_1\left(\langle\frac{i-1}{2^n}, \frac{i}{2^n}\right)\right) \cdot x_2\left(\left(-\infty, t - \frac{i}{2^n}\right)\right) . x_1\left(\langle\frac{j-1}{2^n}, \frac{j}{2^n}\right)\right) \cdot x_2\left(\left(-\infty, t - \frac{j}{2^n}\right)\right) =$$

$$= x_1(\langle a,b)).\alpha.x_1(\langle c,d)) \cdot \beta = 0$$

for $i \neq j$. Therefore

$$m\left(\bigoplus_{i=-\infty}^{\infty} x_1\left(\left\langle \frac{i-1}{2^n}, \frac{i}{2^n}\right)\right) \cdot x_2\left(\left(-\infty, t-\frac{i}{2^n}\right)\right)\right) =$$

$$= \sum_{i=-\infty}^{\infty} m\left(x_1\left(\langle \frac{i-1}{2^n}, \frac{i}{2^n}\right)\right) \cdot x_2\left(\left(-\infty, t-\frac{i}{2^n}\right)\right)$$

hence

$$\lambda_{F_1} \times \lambda_{F_2}(D_t) = \lambda_{F_1} \times \lambda_{F_2}\left(\bigcup_{n=1}^{\infty}\bigcup_{i=-\infty}^{\infty}\left(\langle \frac{i-1}{2^n}, \frac{i}{2^n}\right) \times \left(-\infty, t-\frac{i}{2^n}\right)\right) =$$

$$= \lim_{n\to\infty} \sum_{i=-\infty}^{\infty} \lambda_{F_1} \times \lambda_{F_2}\left(\langle \frac{i-1}{2^n}, \frac{i}{2^n}\right) \times \left(-\infty, t-\frac{i}{2^n}\right)\right) =$$

$$= \lim_{n\to\infty} \sum_{i=-\infty}^{\infty} \lambda_{F_1}\left(\langle \frac{i-1}{2^n}, \frac{i}{2^n}\right)\right) . \lambda_{F_2}\left(\left(-\infty, t-\frac{i}{2^n}\right)\right) =$$

$$= \lim_{n\to\infty} \sum_{i=-\infty}^{\infty} m\left(x_1\left(\langle \frac{i-1}{2^n}, \frac{i}{2^n}\right)\right)\right) . m\left(x_2\left(\left(-\infty, t-\frac{i}{2^n}\right)\right)\right) =$$

$$= \lim_{n\to\infty} \sum_{i=-\infty}^{\infty} m\left(x_1\left(\langle \frac{i-1}{2^n}, \frac{i}{2^n}\right)\right) . x_2\left(\left(-\infty, t-\frac{i}{2^n}\right)\right)\right).$$

$$= m\left(\bigvee_{n=1}^{\infty}\bigoplus_{i=-\infty}^{\infty} x\left(\langle \frac{i-1}{2^n}, \frac{i}{2^n}\right)\right) \cdot \left(\bigoplus_{k=1}^{\infty} y\left(\left(-\infty, t_k-\frac{i}{2^n}\right)\right)\right)\right) = z\left((-\infty, t)\right)$$

$$= m\left(h_2\left(D_t^n\right)\right).$$

□

3 Central limit theorem

We have founded our reasoning on the Kolmogorov theory.

Theorem 3.1. *Let $(\Omega, \mathcal{S}, P)$ be a probability space, $(\xi_n)_n$ be a sequence of independent, square integrable, equally distributed random variables. Let $E(\xi_1) = E(\xi_2) = a, \sigma(\xi_1) = \sigma(\xi_2) = ... = \sigma$. Then for any $t \in R$*

$$\lim_{n\to\infty} P\left(\frac{(\frac{1}{n}\sum_{i=1}^{n}\xi_i) - a}{\frac{\sigma}{\sqrt{n}}} < t\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{u^2}{2}} du.$$

Let $(x_n)_n$ be a sequence of independent observables. We have already defined

$$\sum_{i=1}^{n} x_i((-\infty, t)) = h_n \circ g_n^{-1}((-\infty, t)),$$

where

$$g_n(u_1, ..., u_n) = u_1 + ... + u_n.$$

And we have to define

$$\frac{(\frac{1}{n}\sum_{i=1}^{n}\xi_i - a)}{\frac{\sigma}{\sqrt{n}}}((-\infty, t)).$$

Where

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\xi_i - a}{\frac{\sigma}{\sqrt{n}}} < t$$

if and only if

$$\sum_{i=1}^{n}\xi_i < \left(t\frac{\sigma}{\sqrt{n}} + a\right)n.$$

Then we can define

Definition 3.2. Let M be a σ-complete MV-algebra with product, $x_1, ..., x_n$ be independent observables. Then we define

$$\frac{(\frac{1}{n}\sum_{i=1}^{n}x_i - a)}{\frac{\sigma}{\sqrt{n}}}((-\infty, t)) = (\sum_{i=1}^{n}x_i)\left(\left(-\infty, (t\frac{\sigma}{\sqrt{n}} + a)n\right)\right)$$

$$= h_n\left(g_n^{-1}\left(\left(-\infty, (t\frac{\sigma}{\sqrt{n}} + a)n\right)\right)\right) = h_n \circ g_n^{-1}\left(\left(-\infty, (t\frac{\sigma}{\sqrt{n}} + a)n\right)\right).$$

Theorem 3.3. *Let M be a σ-complete MV-algebra with product, $m : M \to \langle 0, 1\rangle$ be a σ-additive state, $(x_n)_n$ be a sequence of independent, equally distributed, square integrable strong observables. Denote $E[x_1] = E[x_2] = ... = a$, $\sigma(x_1) = \sigma(x_2) = ... = \sigma$. Then for any $t \in R$*

$$\lim_{n\to\infty} m\left(\frac{\frac{1}{n}\sum_{i=1}^{n}x_i - a}{\frac{\sigma}{\sqrt{n}}}((-\infty, t))\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{u^2}{2}}du$$

Proof. Now we have to construct Kolmogorov probability space (depending on the sequence $(x_n)_n$). For any $n \in N$ define

$$\mu_n = \lambda_{F_1} \times \lambda_{F_2} \times ... \times \lambda_{F_n} : \mathcal{B}(R^n) \to \langle 0, 1\rangle.$$

Then $(\mu_n)_n$ presents a consistent system of probability measures, i.e.

$$\mu_{n+1}|\mathcal{B}(R^n) = \mu_n.$$

Let $\mathcal{C}$ be the family of all cylinders in R^N, i.e. all sets of the form

$$\{(u_n)_{n=1}^{\infty} \in R^N; u_{i_1} \in A_1, u_{i_2} \in A_2, ..., u_{i_k} \in A_k\}.$$

Then, using the Kolmogorov consistence theorem, there exists exactly one probability measure

$$P : \sigma(\mathcal{C}) \to \langle 0, 1\rangle$$

such that

$$P(\pi_n^{-1}(A)) = \mu_n(A)$$

for any $n \in N, A \in \mathcal{B}(R^n)$, where $\pi_n : R^N \to R^n$ is the projection

$$\pi_n((u_i)_{i=1}^{\infty}) = (u_1, ..., u_n).$$

Now we have obtained a probability space

$$(R^N, \sigma(\mathcal{C}), P).$$

Next, define

$$\xi_k : R^N \to R$$

by the formula

$$\xi_k((u_i)_{i=1}^{\infty}) = u_k.$$

Compute

$$P_{\xi_k}((-\infty, t)) = P(\{u_n \in R^n; u_k < t\}) =$$

$$= \lambda_{F_1} \times ... \times \lambda_{F_k}(R \times R \times ... \times R \times (-\infty, t)) =$$

$$= \lambda_{F_k}((-\infty, t)).$$

hence $\xi_k : R^N \to R$ and $x_k : \mathcal{J} \to M$ have the same distribution function. Especially

$$E(\xi_k) = \int_R t dF_k(t) = E(x_k).$$

Moreover, put

$$\eta_n = \frac{\frac{1}{n}\sum_{i=1}^{n} \xi_i - a}{\frac{\sigma}{\sqrt{n}}}$$

and

$$y_n = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i - a}{\frac{\sigma}{\sqrt{n}}}$$

i.e.

$$\eta_n^{-1}((-\infty, t)) = \pi_n^{-1}\left(g_n^{-1}\left(\left(-\infty, (t\frac{\sigma}{\sqrt{n}} + a)n\right)\right)\right)$$

and

$$y_n\left((-\infty, t)\right) = \left(\frac{\frac{1}{n}\sum_{i=1}^{n} x_i - a}{\frac{\sigma}{\sqrt{n}}}\right)((-\infty, t)) =$$

$$= h_n\left(g_n^{-1}\left(\left(-\infty, (t\frac{\sigma}{\sqrt{n}} + a)n\right)\right)\right)$$

Then

$$P\left(\eta_n^{-1}\left((-\infty, t)\right)\right) = P\left(\pi_n^{-1}\left(g_n^{-1}\left(\left(-\infty, (t\frac{\sigma}{\sqrt{n}} + a)n\right)\right)\right)\right) =$$

$$= m\left(h_n\left(g_n^{-1}\left(\left(-\infty, (t\frac{\sigma}{\sqrt{n}} + a)n\right)\right)\right)\right) = m\left(y_n\left((-\infty, t)\right)\right).$$

Now we shall prove that $(\xi_n)_n$ are independent:

$$P\left(\xi_1^{-1}((-\infty, t_1)) \cap \xi_2^{-1}((-\infty, t_2)) \cap ... \cap \xi_n^{-1}((-\infty, t_n))\right) =$$

$$= P \circ \pi_n^{-1}\left((-\infty, t_1) \times (-\infty, t_2) \times ... \times (-\infty, t_n)\right) =$$

$$= m \circ h_n\left((-\infty, t_1) \times (-\infty, t_2) \times ... \times (-\infty, t_n)\right)) =$$

$$= m\left(h_n\left((-\infty, t_1) \times (-\infty, t_2) \times ... \times (-\infty, t_n)\right)\right) =$$

$$= m\left(x_1((-\infty, t_1))\right).m\left(x_2((-\infty, t_2))\right). .... .m\left(x_n((-\infty, t_n))\right) =$$

$$= P\left(\xi_1^{-1}((-\infty, t_1))\right).P\left(\xi_2^{-1}((-\infty, t_2))\right). .... .P\left(\xi_n^{-1}((-\infty, t_n))\right).$$

Therefore by Theorem 3.1, Theorem 2.10 and the statement

$$P(\eta_n^{-1}((-\infty, t))) = m\left(y_n((-\infty, t))\right),$$

for $a = E\left(\xi_1\right) = E\left(\xi_2\right) = ... = E\left(\xi_n\right)$ we have

$$\lim_{n\to\infty} P\left(\frac{\frac{1}{n}\sum_{i=1}^{n}\xi_i - a}{\frac{\sigma}{\sqrt{n}}} < t\right) = \lim_{n\to\infty} P\left(\eta_n^{-1}\left((-\infty, t)\right)\right) =$$

$$= \lim_{n\to\infty} P\left(\pi_n^{-1}\left(g_n^{-1}\left(\left(-\infty, (t\frac{\sigma}{\sqrt{n}} + a)n\right)\right)\right)\right) =$$

$$= \lim_{n\to\infty} P \circ \pi_n^{-1}\left(g_n^{-1}\left(\left(-\infty, (t\frac{\sigma}{\sqrt{n}} + a)n\right)\right)\right) =$$

$$= \lim_{n\to\infty} m\left(h_n\left(g_n^{-1}\left(\left(-\infty, (t\frac{\sigma}{\sqrt{n}} + a)n\right)\right)\right)\right) =$$

$$= \lim_{n\to\infty} m\left(y_n((-\infty, t))\right) = \lim_{n\to\infty} m\left(\frac{\frac{1}{n}\sum_{i=1}^{n}x_i - a}{\frac{\sigma}{\sqrt{n}}}\left((-\infty, t)\right)\right) =$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{u^2}{2}}du.$$

□

Remark 3.4. We have shown, that on σ-complete MV-algebra with product, each observable corresponds to a random variable with the same distribution, and a couple of independent observables corresponds to a couple of independent random variables. This way all version of central limit theorems can be translated into the languange of σ-complete MV-algebras with product. However, that was not the aim of this paper.

References

[1] B. Riečan, *On a new approach to probability theory on MV-algebras*, to appear.

[2] B. Riečan, *On the product MV-algebras. Tatra Mt. Math. Publ.* **16** (1999), 143 – 149.

[3] B. Riečan and L. Lašová, *On the probability theory on the Kôpka D-posets*, to appear.

[4] B. Riečan and D. Mundici, *Probability on MV algebras*, In: "Handbook of Measue Theory", Elsevier, Amsterdam 2002, 869 – 909.

[5] F. Montagna, *An algebraic approach to propositional fuzzy logic*, J. Logic Lang. Inf. (2000), 91-124.

ACTA UNIVERSITATIS MATTHIAE BELII, series MATHEMATICS 18 (2011), 47–53
Also available at http://actamath.savbb.sk
ISBN 978-80-557-0160-8

Kernel choice with respect to the bandwidth in kernel density estimates*

Kamila Vopatová
Department of Mathematics and Statistics,
Faculty of Science, Masaryk University, Kotlářská 2
637 11 Brno, Czech Republic
E-mail: vopatova@mail.muni.cz

Abstract

Kernel density estimates belong to the most popular nonparametric density estimates. It is well known that these estimates depend on a bandwidth, which controls the smoothness of the estimate, and on a kernel, which plays a role of weight function.
We focus on the kernel function choice, especially on kernels with bounded supports. Our aim is to study the kernel optimality with respect to the bandwidth choice. In a simulation we show a comparison of the kernels. We propose that the cosine kernel may be a good alternative to the frequently used Epanechnikov kernel.

Keywords Kernel, bounded support, density estimation, cross-validation.

MSC (2010) 62G07, 30C40

Received: July 29, 2010; Revised: August 28, 2010; Accepted: August 30, 2010.

1 Kernel density estimation

The concept of nonparametric estimates was introduced in the fifties and sixties, see e.g. [3, 5] and references therein. After years, there are still problems to be solved or to be improved.

Let a d-variate random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ come from distribution with a density f. The kernel density estimator $\hat{f}$ is defined as a weighted average of observations

$$\hat{f}(\mathbf{x}, H) = \frac{1}{n}\sum_{i=1}^{n} K_H(\mathbf{x} - \mathbf{X}_i) = \frac{1}{n}|H|^{-1/2}\sum_{i=1}^{n} K\big(H^{-1/2}(\mathbf{x} - \mathbf{X}_i)\big).$$

K is a d-variate kernel function, which is often taken to be a probability density function satisfying $\int_{\mathbb{R}^d} K(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$, where we omit the subscript $\mathbb{R}^d$ in the rest of the text. H is a symmetric positive definite $d \times d$ matrix called a bandwidth matrix and $\mathbf{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$ is a generic vector.

Transfer of the kernel estimation from univariate settings to the multivariate settings brings a few new issues. The first problem is to find a kernel and the second one is to find an optimal bandwidth. There are two common ways to create the multivariate kernel K from the univariate kernel k: the product kernel $K^P(\mathbf{x}) = \prod_{i=1}^{d} k(x_i)$

*This research was supported by Masaryk University under the Student Project Grant MUNI/A/1001/2009.

and the spherically symmetric kernel $K^S(\mathbf{x}) = c_k^{-1} k((\mathbf{x}^T\mathbf{x})^{1/2})$, $c_k = \int k(\sqrt{\mathbf{x}^T\mathbf{x}})\ \mathrm{d}\mathbf{x}$. Let us denote the class of symmetric positive definite $d \times d$ matrices as $\mathcal{H}_{\mathcal{F}}$. One needs to choose $d(d+1)/2$ distinct entries of matrix $H \in \mathcal{H}_{\mathcal{F}}$, which is computationally intensive. On the other hand, using a single parameter simplification, i.e. $H = h^2 \cdot I_d$ (I_d is a $d \times d$ identity matrix), is not advised for data which have different dispersions in the co-ordinate directions, see [4]. Thus, the diagonal matrix class $\mathcal{H}_{\mathcal{D}}$ ($H = \mathrm{diag}(h_1^2, \ldots, h_d^2)$) seems to be a compromise between computational speed and sufficient flexibility.

Mean integrated square error (MISE) quantifies the performance of a multivariate kernel density estimator. MISE can be rewritten as a sum of an integrated variance and an integrated square bias

$$\mathrm{MISE}(H) = E \int \left[\hat{f}(\mathbf{x}, H) - f(\mathbf{x})\right]^2 \mathrm{d}\mathbf{x} = \int \mathrm{Var}\, \hat{f}(\mathbf{x}, H)\ \mathrm{d}\mathbf{x} + \int \mathrm{Bias}^2\, \hat{f}(\mathbf{x}, H)\ \mathrm{d}\mathbf{x}.$$

It is easy to see, that finding the bandwidth matrix H_{MISE}, which minimizes this error, is very difficult. Wand and Jones [5] derived, under some assumptions on the density f, the kernel function K, and the bandwidth matrix H, the asymptotic mean integrated square error

$$\mathrm{AMISE}(H) = n^{-1}|H|^{-1/2}V(K) + \tfrac{1}{4}\beta_2(K)^2(\mathrm{vech}\, H)^T \Psi_4(\mathrm{vech}\, H).$$

$V(K) = \int K^2(\mathbf{x})\ \mathrm{d}\mathbf{x}$, $\beta_2(K) = \int x_i^2 K(\mathbf{x})\ \mathrm{d}\mathbf{x}$ is independent of i and vech is a vector half operator, i.e. for a matrix M, vech M is a $d(d+1)/2 \times 1$ vector of stacked columns of the lower triangular matrix of M. The matrix Ψ_4 includes entries depending on the unknown density f. For a d-variate function g and for a vector $\mathbf{r} = (r_1, \ldots, r_d)$ of non-negative integers, $g^{(\mathbf{r})}$ is defined

$$g^{(\mathbf{r})}(\mathbf{x}) = \frac{\partial^{|\mathbf{r}|}}{\partial x_1^{r_1} \cdots \partial x_d^{r_d}} g(\mathbf{x})$$

assuming that the derivative exists. The notation $|\mathbf{r}|$ is for the sum of the components of the vector $\mathbf{r}$. Each entry of Ψ_4 can be written, under sufficient conditions, in the form $\psi_{\mathbf{r}} = \int f^{(\mathbf{r})}(\mathbf{x}) f(\mathbf{x})\ \mathrm{d}\mathbf{x}$, where $|\mathbf{r}|$ is even (see Section 4.3 in [5]).

2 Simulation study

2.1 Cross-validation methods

There is a wide range of methods for the optimal bandwidth choice, thus we aimed our study to cross-validation (CV) methods.

The most widely used cross-validation method is the least square cross-validation method (LSCV) [5].

$$\mathrm{LSCV}(H) = \int \left(\hat{f}(\mathbf{x}, H)\right)^2 \mathrm{d}\mathbf{x} - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(\mathbf{X}_i, H),$$

is the LSCV objective function, where $\hat{f}_{-i}(\mathbf{X}_i, H) = (n-1)^{-1}\sum_{j=1, j\neq i}^{n} K_H(\mathbf{X}_i - \mathbf{X}_j)$ is a leave-one-out estimator of f. LSCV function can be written in terms of convolutions $(f * g)(x) = \int_{\mathbb{R}} f(t) g(x-t)\ \mathrm{d}t$ (see e.g. [1])

$$\mathrm{LSCV}(H) = n^{-1}(n-1)^{-1}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} (K_H * K_H - 2K_H)(\mathbf{X}_i - \mathbf{X}_j) + n^{-1}V(K)|H|^{-1/2}.$$

LSCV is also called the unbiased cross-validation, because the LSCV(H) function is unbiased in the sense that $E[\mathrm{LSCV}(H)] = \mathrm{MISE}(H) - \int f^2(\mathbf{x})\,\mathrm{d}\mathbf{x}$.

Biased cross-validation (BCV) method estimates AMISE, i.e. BCV is a biased estimate of MISE. There are two types of BCV, depending on the way of estimating functionals $\psi_{\mathbf{r}}$ [1]:

$$\mathrm{BCV}_1(H) = n^{-1}V(K)|H|^{-1/2} + \frac{1}{4}\beta_2(K)^2(\mathrm{vech}\,H)^T\hat{\Psi}_4(\mathrm{vech}\,H),$$

where

$$\hat{\psi}_{\mathbf{r}} = n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1,\\ j\neq i}}^{n}\big(K_H^{(\mathbf{r})} * K_H\big)(\mathbf{X}_i - \mathbf{X}_j).$$

The latter is defined by the function

$$\mathrm{BCV}_2(H) = n^{-1}V(K)|H|^{-1/2} + \frac{1}{4}\beta_2(K)^2(\mathrm{vech}\,H)^T\tilde{\Psi}_4(\mathrm{vech}\,H),$$

with

$$\tilde{\psi}_{\mathbf{r}} = n^{-1}\sum_{i=1}^{n}\hat{f}_{-i}^{(\mathbf{r})}(\mathbf{X}_i, H) = n^{-1}(n-1)^{-1}\sum_{i=1}^{n}\sum_{\substack{j=1,\\ j\neq i}}^{n}K_H^{(\mathbf{r})}(\mathbf{X}_i - \mathbf{X}_j).$$

2.2 Kernels

We focused on studying the product kernels with bounded supports. Due to a wide range of the class of the kernels with bounded supports $\mathcal{D}$: $\{[x_1, x_2] \in \mathbb{R}^2 : |x_1| \leq 1 \wedge |x_2| \leq 1\}$, we selected the easiest five two-dimensional kernels listed in Table 1. Their one-dimensional representations are displayed in Figure 1.

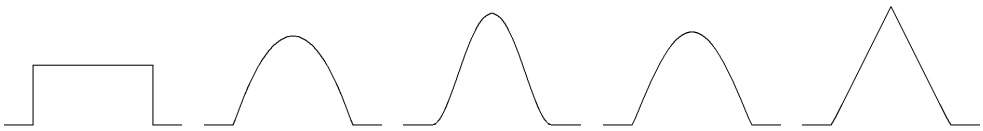
	$K(x_1, x_2)$	$V(K)$	$\beta_2(K)$
Uniform	$1/4$	$1/4$	$1/3$
Epanechnikov	$9/16(1-x_1^2)(1-x_2^2)$	$9/25$	$1/5$
Biweight	$225/256(1-x_1^2)^2(1-x_2^2)^2$	$25/49$	$1/7$
Cosine	$\pi^2/16\cos(x_1\pi/2)\cos(x_2\pi/2)$	$(\pi/4)^4$	$1-8/\pi^2$
Triangular	$(1-\lvert x_1\rvert)(1-\lvert x_2\rvert)$	$4/9$	$1/6$

Table 1: Two-dimensional product kernels with bounded supports.

Remark 2.1. Epanechnikov kernel is the optimal kernel in the AMISE sense, i.e. Epanechnikov kernel minimizes the functional $T(K) = [V(K)\beta_2(K)]^{2/3}$ (see [5]). From $\mathrm{eff}(K) = [T(K)/T(K_{Epan})]^{3/2}$ is apparent that the cosine kernel is a convenient alternative to Epanechnikov kernel.

2.3 Densities

We drew samples of the size $n = 50$ and $n = 100$ from densities listed in Table 2. Contour plots of target densities are displayed in Figure 2. One hundred replications for each of the sample sizes and for each of the densities were generated.



(a) Uniform (b) Epanechnikov (c) Biweight (d) Cosine (e) Triangular

Figure 1: One-dimensional kernels with bounded supports.

(A)	Normal	$N_2(0,0;4,1,0)$
(B)	Student	$t_2(4)$
(C)	Weibull	$W(2,2)\cdot W(2,4)$
(D)	Exponential	$Exp(2)\cdot Exp(1)$
(E)	Gamma Student	$Gamma(2,1)\cdot t(5)$
(F)	Lognormal	$LN_2(0,0;1,1,0)$

Table 2: Target densities.

2.4 Comparative criteria

In the case of the diagonal AMISE-optimal bandwidth matrix $H_{\mathrm{AMISE}} = \mathrm{diag}(h_{1,\mathrm{A}}^2, h_{2,\mathrm{A}}^2)$, we can express its entries [5]

$$h_{1,\mathrm{A}} = \left[\frac{\psi_{04}^{3/4} V(K)}{\beta_2(K)^2 \psi_{40}^{3/4}(\psi_{22} + \psi_{04}^{1/2}\psi_{40}^{1/2})n}\right]^{1/6},$$
$$h_{2,\mathrm{A}} = \left[\frac{\psi_{40}^{3/4} V(K)}{\beta_2(K)^2 \psi_{04}^{3/4}(\psi_{22} + \psi_{04}^{1/2}\psi_{40}^{1/2})n}\right]^{1/6}.$$

We used two criteria to decide which kernel fits best the chosen method: the average of squared Euclidean norm of difference vectors

$$\mathrm{ED} = \mathrm{avg}_H \|(\hat{h}_1 - h_{1,\mathrm{A}}, \hat{h}_2 - h_{2,\mathrm{A}})^T\|_2^2$$

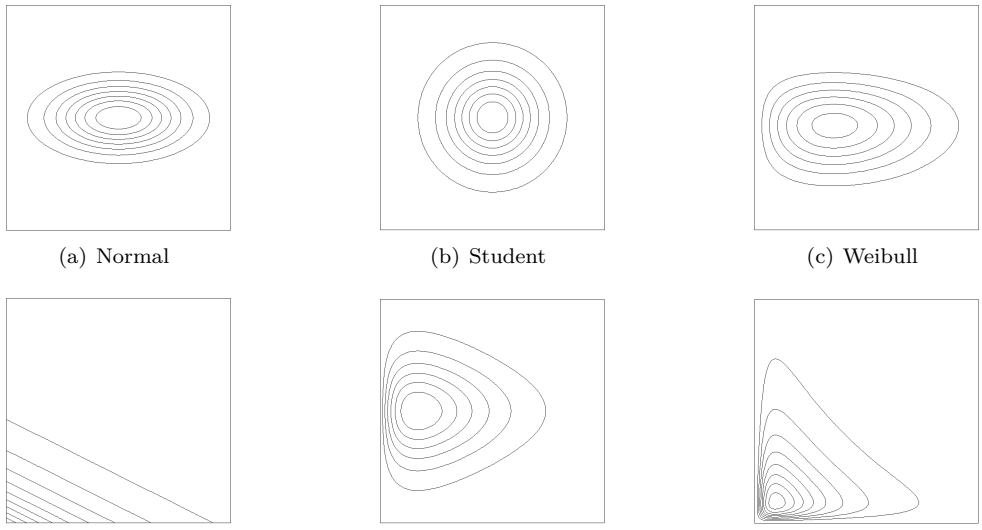
and the average of integrated square errors

$$\mathrm{ISE} = \mathrm{avg}_H \int \left[\hat{f}(\mathbf{x}, H) - f(\mathbf{x})\right]^2 \mathrm{d}\mathbf{x},$$

where the average is taken over simulated realizations. ED can be considered as a visual criterion and ISE, which was computed numerically, can be viewed as a numerical criterion.

3 Results

We compared a performance of kernels within each of the mentioned cross-validation methods. We selected bandwidth matrices — we computed values of $\min_{H\in\mathcal{H}_\mathcal{D}} \mathrm{LSCV}(H)$, $\min_{H\in\mathcal{H}_\mathcal{D}} \mathrm{BCV}_1(H)$, and $\min_{H\in\mathcal{H}_\mathcal{D}} \mathrm{BCV}_2(H)$ on a dense grid numerically. For a faster computation, we used ideas by Horová et al. [2]. Tables 3 and 4 summarize results of ED criterion and Tables 5 and 6 summarize results of ISE criterion.



(a) Normal (b) Student (c) Weibull

(d) Exponential (e) Gamma Student (f) Lognormal

Figure 2: Contour plots of target densities.

ED		LSCV				
density	n	Uni	Epan	Biw	Cos	Tri
(A)	50	0.76(0.07)	0.69(0.05)	0.67(0.06)	1.82(0.16)	4.24(0.28)
	100	0.85(0.05)	0.76(0.04)	0.58(0.04)	2.01(0.12)	4.97(0.21)
(B)	50	0.86(0.04)	0.88(0.05)	0.82(0.05)	1.21(0.08)	2.75(0.17)
	100	0.92(0.02)	0.90(0.02)	0.81(0.02)	1.16(0.05)	2.61(0.09)
(C)	50	0.80(0.02)	0.81(0.02)	0.75(0.02)	1.62(0.07)	3.32(0.12)
	100	0.95(0.02)	1.00(0.02)	0.95(0.02)	1.84(0.05)	3.28(0.10)
(D)	50	0.69(0.08)	0.69(0.09)	0.63(0.09)	1.62(0.21)	4.32(0.42)
	100	0.62(0.05)	0.42(0.04)	0.31(0.03)	1.34(0.14)	3.82(0.27)
(E)	50	1.05(0.05)	1.04(0.05)	1.04(0.06)	1.85(0.12)	3.78(0.21)
	100	1.13(0.03)	1.13(0.02)	1.07(0.02)	1.66(0.07)	3.55(0.12)
(F)	50	2.05(0.06)	2.26(0.05)	2.43(0.06)	2.55(0.09)	5.10(0.23)
	100	2.11(0.05)	2.19(0.04)	2.35(0.04)	2.56(0.08)	4.96(0.14)

Table 3: LSCV method: an average of Euclidean norm with a standard error.

One can use each of the kernels for the least square cross-validation method. The bias cross-validation methods require some smoothness conditions to be satisfied. In this case, we can use only Epanechnikov, the biweight, and the cosine kernel for BCV_1 and the biweight and the cosine kernel for BCV_2.

The biweight kernel seems to be the best choice for the least square cross-validation method. According to ED-criterion, Epanechnikov and the uniform kernel are also good choice. By contrast, the uniform kernel is the least suitable choice according to ISE-criterion. The second best choice is Epanechnikov and the cosine kernel. On the other hand, LSCV-optimal bandwidths suffer from a large variability (see [2]). For all kernels this variability is approximately the same.

In the case of the BCV_1, we can choose between using Epanechnikov and the

ED		BCV_1			BCV_2	
density	n	Epan	Biw	Cos	Biw	Cos
(A)	50	3.79(0.02)	6.29(0.03)	4.15(0.02)	2.62(0.06)	1.12(0.03)
	100	3.43(0.01)	5.44(0.01)	3.74(0.01)	2.31(0.03)	1.07(0.02)
(B)	50	1.19(0.01)	2.12(0.01)	1.34(0.01)	0.54(0.01)	0.15(∗)
	100	1.14(∗)	1.91(0.01)	1.26(∗)	0.53(0.01)	0.16(∗)
(C)	50	0.63(∗)	1.16(0.01)	0.71(0.01)	0.22(0.01)	0.04(∗)
	100	0.62(∗)	1.05(∗)	0.69(∗)	0.23(0.01)	0.05(∗)
(D)	50	1.90(0.01)	3.27(0.02)	2.11(0.02)	1.05(0.03)	0.35(0.01)
	100	1.79(0.01)	2.88(0.01)	1.97(0.01)	1.02(0.02)	0.40(0.01)
(E)	50	1.14(0.01)	1.98(0.01)	1.26(0.01)	0.66(0.02)	0.34(0.02)
	100	1.06(∗)	1.75(0.01)	1.17(∗)	0.56(0.01)	0.25(0.01)
(F)	50	0.01(∗∗)	0.04(∗)	0.01(∗∗)	0.32(0.02)	0.57(0.02)
	100	0.02(∗∗)	0.06(∗∗)	0.02(∗∗)	0.15(∗)	0.31(∗)

Table 4: BCV_1 and BCV_2 methods: an average of Euclidean norm with a standard error (∗ stands for the standard error less than 0.005 and ∗∗ for the standard error less than 0.001).

$100 \times$ ISE		LSCV				
density	n	Uni	Epan	Biw	Cos	Tri
(A)	50	0.48(0.02)	0.39(0.02)	0.38(0.02)	0.43(0.02)	0.48(0.02)
	100	0.37(0.01)	0.27(0.01)	0.24(0.01)	0.30(0.01)	0.38(0.01)
(B)	50	1.11(0.04)	0.86(0.03)	0.77(0.03)	0.90(0.04)	1.10(0.04)
	100	0.97(0.02)	0.69(0.02)	0.58(0.02)	0.70(0.02)	0.91(0.03)
(C)	50	3.54(0.07)	2.47(0.07)	2.03(0.07)	2.48(0.07)	3.06(0.07)
	100	3.52(0.05)	2.47(0.05)	1.97(0.05)	2.37(0.05)	2.84(0.06)
(D)	50	5.23(0.07)	4.43(0.08)	4.05(0.08)	4.28(0.08)	4.62(0.08)
	100	5.04(0.06)	4.12(0.06)	3.67(0.06)	3.92(0.06)	4.30(0.06)
(E)	50	1.03(0.03)	0.79(0.03)	0.73(0.03)	0.82(0.04)	0.94(0.04)
	100	0.97(0.02)	0.71(0.02)	0.61(0.02)	0.68(0.02)	0.83(0.02)
(F)	50	5.32(0.08)	4.25(0.08)	3.76(0.08)	4.10(0.08)	4.57(0.08)
	100	5.32(0.06)	4.12(0.06)	3.59(0.06)	3.97(0.06)	4.47(0.06)

Table 5: LSCV method: an average of the integrated square error with a standard error.

cosine kernel, because their performance is comparable regarding both critera. But the BCV_1 method gives quite underestimated values of the optimal bandwidth H.

Concerning BCV_2, it is obvious that the cosine kernel performs better than the biweight kernel. Density (F) is an exception, but there is influence of boundary effects. The main advantage of the cosine kernel over the biweight kernel is that its fourth derivative, needed for calculation of BCV_2, is not a constant function, as it is the case for the biweight kernel.

4 Conclusion

In this paper we compared several two-dimensional kernels with a bounded support with respect to the method for finding optimal bandwidth. We propose that the biweight kernel is the best choice for the LSCV method and the cosine kernel is the best choice for the BCV_2. In the case of BCV_1, one can decide whether to use the

$100 \times$ ISE		BCV_1			BCV_2	
density	n	Epan	Biw	Cos	Biw	Cos
(A)	50	4.79(0.12)	9.48(0.32)	5.51(0.15)	1.24(0.05)	0.73(0.03)
	100	4.80(0.12)	9.08(0.28)	5.54(0.15)	0.90(0.02)	0.54(0.01)
(B)	50	5.82(0.28)	11.35(0.58)	6.81(0.33)	1.48(0.07)	0.94(0.04)
	100	5.81(0.23)	11.17(0.51)	6.63(0.28)	1.07(0.04)	0.67(0.02)
(C)	50	9.87(0.22)	18.90(0.42)	11.18(0.25)	2.44(0.10)	1.55(0.07)
	100	9.81(0.19)	18.80(0.45)	11.26(0.23)	1.72(0.05)	1.10(0.04)
(D)	50	7.43(0.22)	14.10(0.41)	8.43(0.25)	3.37(0.10)	3.34(0.08)
	100	7.65(0.17)	14.38(0.37)	8.98(0.23)	2.55(0.06)	2.55(0.06)
(E)	50	5.81(0.18)	11.42(0.40)	6.68(0.21)	1.47(0.05)	0.94(0.03)
	100	6.16(0.17)	11.63(0.42)	7.04(0.22)	1.12(0.03)	0.71(0.02)
(F)	50	7.90(0.25)	15.94(0.73)	9.09(0.32)	2.97(0.08)	2.96(0.08)
	100	8.95(0.28)	17.51(0.66)	10.48(0.35)	2.37(0.05)	2.31(0.05)

Table 6: BCV_1 and BCV_2 methods: an average of the integrated square error with a standard error.

cosine kernel or Epanechnikov kernel.

In the future, we want to extend this study to a larger group of kernels and also to a higher dimension.

References

[1] T. Duong and M. L. Hazelton, *Cross-validation Bandwidth Matrices for Multivariate Kernel Density Estimation*, Scand. J. Statist. **32** (2005) 485–506.

[2] I. Horová, J. Koláček, J. Zelinka and K. Vopatová, *Bandwidth choice for kernel density estimates*, Proc. IASC (2008) 542–551.

[3] D.W. Scott, "Multivariate Density Estimation: Theory, Practice, and Visualization", John Wiley & Sons, New York, 1992.

[4] M.P. Wand and M.C. Jones, *Comparison of Smoothing Parameterizations in Bivariate Kernel Density Estimation*, J. Amer. Statist. Assoc. **88** (1993), 520–528.

[5] M.P. Wand and M.C. Jones: "Kernel Smoothing", Chapman & Hall, London, 1995.

Title	Acta Universitatis Matthiae Belii Natural Science Series Series Mathematics
Pages	53
Number of copies	110
First Edition	
Size	B5
Typeset with LaTeX	
Design and layout	Ján Karabáš
Published By	Science and Research Institute, Matej Bel University
Printed by	Bratia Sabovci, s. r. o.
ISBN	978-80-557-0160-8

Acta Universitatis Matthiae Belii

Series Mathematics Number 18 2011

Contents