Fuzzy Bayesian Networks with Likert Scales

Jan Mrógala[1], Irina Perfilieva[1], and Jiří Vomlel*[1,2]

[1]Institute for Research and Applications of Fuzzy Modeling
University of Ostrava
30. dubna 22, 701 03 Ostrava 1, Czechia

[2]Institute of Information Theory and Automation
Czech Academy of Sciences
Pod Vodárenskou věží 4, 182 00, Prague 8, Czechia
vomlel@utia.cas.cz

Abstract

Our work is motivated by the applications of probabilistic models in the social sciences, in which surveys and questionnaires are typically used to collect respondents' opinions via a Likert scale. The dividing lines between the states on the Likert scale are vague, so it is natural to interpret them using fuzzy numbers instead of integers. We treat the true model variables as hidden continuous variables, the values of which are observed only through their fuzzified counterparts. This approach seems more conceptually appropriate in the context of surveys and questionnaires, since the modeled variables are continuous by nature but are only observed on a fuzzy, discrete scale. Probabilistic inference with continuous variables is challenging when the assumption of normality of the variables' distribution is violated, which is particularly true for variables modeling polarizing issues. We approximate continuous, multidimensional probability distributions using an F-transform composed of basic functions with central points, called nodes, at a multidimensional grid. We illustrate the suggested approach using a small Bayesian network model of data from the survey "Dividing Lines in Czech Society."

1 Introduction

Uncertainty can manifest in multiple ways. One aspect is the likelihood of a variable's state, typically expressed as a probability. Another aspect is the vagueness of a variable's

*Corresponding author

state, which can be represented by a fuzzy set.

Our work is motivated by the application of probabilistic models in the social sciences, where surveys and questionnaires are commonly used to collect respondents' opinions via a Likert scale. Likert scales consist of several states, which are usually represented by a set of consecutive integers. A seven-point Likert scale, for example, has the following states: (1) strongly disagree, (2) disagree, (3) somewhat disagree, (4) neutral, (5) somewhat agree, (6) agree, and (7) strongly agree. The dividing lines between these states are vague, so it is natural to interpret them using fuzzy sets.

Bayesian networks (Pearl, 1988; Jensen and Nielsen, 2007; Koller and Friedman, 2009)) are probabilistic graphical models in which directed acyclic graphs are used to model relations between random variables. Pan and Liu introduced fuzzy Bayesian networks in (Pan and Liu, 2000). Their formalism fuzzifies a Bayesian network by replacing continuous variables with discrete ones. The mapping between the two is approximated by a conditional Gaussian distribution. The second form replaces continuous variables with discrete partners only when necessary and uses conditional Gaussian regression for continuous dependencies.

We treat continuous variables as hidden variables, observing their values only through their fuzzified counterparts. We extend Bayesian networks with additional variables. Each additional variable represents a fuzzy observation of the corresponding unobserved continuous variable. In the context of surveys and questionnaires, this approach seems more conceptually appropriate, since the modeled variables are continuous by nature but are observed on a fuzzy, discrete scale. Unfortunately, this proposal makes inference challenging because it must be performed with continuous variables. Furthermore, the assumption of normality of the variables' distribution is often significantly violated. Quite often, probability distributions have several modes. This is particularly true of polarizing issues. Therefore, we approximate the continuous, multidimensional probability distributions of our fuzzy Bayesian network using the F-transform (**?**)perfilieva-2006, perfilieva-2008), which is composed of basic functions with central points (called nodes) at a multidimensional grid. We illustrate the suggested approach using data from the survey "Dividing Lines in Czech Society", discussed in (Buchtík, 2023).

The paper is organized as follows: In Section 2 we introduce the necessary terminology from the fuzzy set theory and present three types of membership functions used in this paper. These membership functions form the basis of F-transform and its inverse and they are described in Section 3. Finally, in Section 4 the F-transform using Gaussian basic functions is applied to inference in a small Bayesian network. We conclude the paper with a discussion of future research directions.

2 Fuzzy sets and fuzzy partitions

In this paper we will treat the studied variables as continuous random variables that are hidden and their values are observed only through their fuzzified counterparts with their values from a Likert scale[1]. This idea is not entirely new since some authors have

[1]In this paper we consider the Likert scale with seven values, i.e., $n = 7$.

already suggested to identify each Likert response category with a fuzzy set (Gil and González-Rodríguez, 2012).

A fuzzy set is a pair (U, A) where U is a set (often required to be non-empty) and A is a mapping from universe U to $[0, 1]$ referred as the membership function. The fuzzy set is often identified with its membership function. An observation $X = k$ of a random variable X, where k is a value from the Likert scale $\{1, \ldots, n\}$ will be interpreted as a fuzzy set $A_k : [1, n] \to [0, 1]$.

A family of fuzzy sets $\{A_k, k = 1, \ldots, n\}$ where all A_k are defined on the same universe U is called fuzzy partition. In this paper the universe U will be an interval of real numbers $[a, b]$. In Figure 1 we present examples of two fuzzy partitions that we discuss in more detail later in this section. In Perfilieva (2006) a fuzzy partition $A_1, \ldots, A_n$ of $[a, b]$ satisfying so called Ruspini conditions was introduced. Let $x_1 < \ldots < x_n$ be fixed nodes from the interval of real numbers $[a, b]$ such that $x_1 = a$, $x_n = b$ and $n \geq 2$.

Definition 1 (Ruspini partition)**.** We say that the fuzzy sets $A_1, \ldots, A_n$, identified with their membership functions defined on $[a, b]$, establish a Ruspini partition of $[a, b]$ if they fulfill the following conditions for $k = 1, \ldots, n$:

1. $A_k : [a, b] \to [0, 1]$, $A_k(x_k) = 1$;
2. $A_k(x) = 0$ if $x \notin (x_{k-1}, x_{k+1})$, where for uniformity of notation, we set $x_0 = a$ and $x_{n+1} = b$;
3. $A_k(x)$ is continuous;
4. $A_k(x)$, for $k = 2, \ldots, n$, strictly increases on $[x_{k-1}, x_k]$ and $A_k(x)$, for $k = 1, \ldots, n-1$, strictly decreases on $[x_k, x_{k+1}]$;
5. for all $x \in [a, b]$,
$$\sum_{k=1}^{n} A_k(x) = 1.$$

The membership functions $A_1, \ldots, A_n$ are called *basic functions*. A point $x \in [a, b]$ is *covered* by basic function A_k if $A_k(x) > 0$. The shape of the basic functions is not predetermined, so it can be selected based on additional requirements (such as smoothness).

The first basic function considered in this paper is the Gaussian function defined for $x \in [a, b]$ as:

$$A_k(x) \quad = \quad \frac{1}{\sigma_X^k \cdot \sqrt{2\pi}} \exp\left(-\frac{(x - c_X^k)^2}{2 \cdot (\sigma_X^k)^2}\right)$$

where parameters are the values σ_X^k of the standard deviation and the means c_X^k specify centers of the fuzzy sets[2]. In the following experiments, we always positioned the centers

[2]In all our experiments we assume that standard deviations are equal, i.e. $\sigma_X^1 = \ldots = \sigma_X^7$.

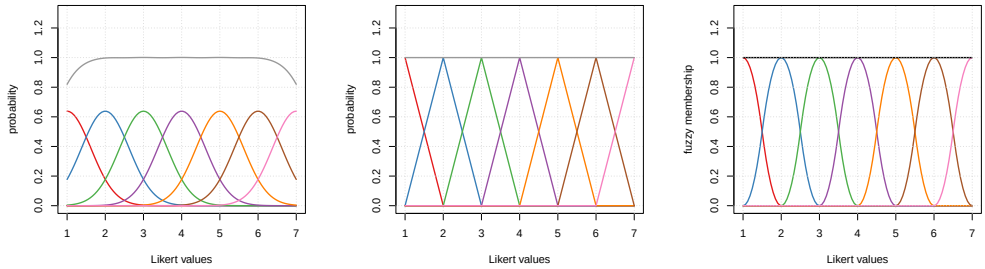

Figure 1: Fuzzy partitions using Gaussian (left), triangular (middle), and Π-curves (right) basic functions.

at certain points of the considered grid[3]. This fuzzy partition does not satisfy the conditions of the Ruspini partition. See the right hand side of Figure 1 where the top black curve corresponds to the sum of values of all seven basic functions.

The second option we consider are triangular functions. Each basic function T_k is specified by its center c_X^k and its width w_X^k for $x \in [a, b]$ as:

$$T_k(x) \quad = \quad \max\left\{0, 1 - \frac{2|x - c_X^k|}{w_X^k}\right\}$$

Another option, commonly used in the fuzzy sets applications are Π-curves. They are named after their shape that for certain parametrizations resemble the letter Π. They are composed from the so called S-functions defined for $x, u, v \in [a, b]$, $u < v$ as:

$$S(x, u, v) \quad = \quad \begin{cases} 0 & \text{if } x \leq u \\ 2\left(\frac{x-u}{v-u}\right)^2 & \text{if } u < x \leq \frac{u+v}{2} \\ 1 - 2\left(\frac{x-v}{v-u}\right)^2 & \text{if } \frac{u+v}{2} < x \leq v \\ 1 & \text{if } v < x \end{cases}$$

In this paper we will use Π-curves defined for $x \in [1, 7]$ as:

$$\Pi_k(x) \quad = \quad S(x, k - s, k) - S(x, k, k + s) \ ,$$

where k are coordinates of the centers of individual basic functions and s is their span.[4]. The advantage of this fuzzy transformation is that for the above parametrization satisfies the Ruspini condition.

[3]The position of centers could be considered as another parameter to be specified with the help of a domain expert or by an optimization algorithm.

[4]In this paper we use $k \in \{1, \ldots, 7\}$ with $s = 1$ or $k \in \{1, 4, 7\}$ with $s = 2$.

Since the fuzzy transformations of two variables X and Y are independent then an observation $X = i, Y = j$, where i, j are the Likert values in data can be interpreted as a two-dimensional fuzzy set with the basic function

$$\begin{aligned} f_{i,j}(x, y|c_X, c_Y, \sigma_X, \sigma_Y) &= f_i(x|c_X, \sigma_X) \cdot f_i(y|c_Y, \sigma_Y) \\ &= \frac{1}{2 \cdot \pi \cdot \sigma_X \cdot \sigma_Y} \exp -\frac{1}{2} \cdot \left(\frac{(x - c_X)^2}{\sigma_X^2} + \frac{(y - c_Y)^2}{\sigma_Y^2} \right) \end{aligned}$$

where the parameters are the standard deviations σ_X and σ_Y of X and Y, respectively and the mean values c_X, c_Y that specifies the center of the fuzzy set[5]. Similarly, for the Π-curves we get

$$\Pi_{i,j}(x, y) = \Pi_i(x) \cdot \Pi_j(y) \ .$$

3 F-transform

In this section, we briefly introduce F-transform - a technique that we will use to approximate multidimensional probability distributions. In this paper, F-transform is used to represent a discrete probability distribution by a finite set of its F-transform components. We present the definition of F-transform (Perfilieva, 2006) for a discrete probability distribution of two variables[6] defined on the grid

$$\mathbb{X} = \mathbb{X}_1 \times \mathbb{X}_2 = \{1, \ldots, N\} \times \{1, \ldots, M\} \ .$$

Let $A_1, \ldots, A_n$ and $B_1, \ldots, B_m$ be basic functions. Further, let $A_1, \ldots, A_n : [1, N] \to [0, 1]$ and $B_1, \ldots, B_m : [1, M] \to [0, 1]$ be fuzzy partitions of $[1, N]$ and $[1, M]$, respectively. Assume that the grid $\mathbb{X}$ is *sufficiently dense with respect to the chosen partitions*, which means that $(\forall k)(\exists x \in \mathbb{X}_1) : A_k(x) > 0$, and $(\forall l)(\exists y \in \mathbb{X}_2 : B_l(y) > 0$.

Definition 2. A $n \times m$ matrix U of real numbers is called *F-transform* of P with respect to $\{A_1, \ldots, A_n\}$ and $\{B_1, \ldots, B_m\}$ if for all $k = 1, \ldots, n,\ l = 1, \ldots, m$,

$$U_{kl} = \frac{\sum_{j=1}^{M} \sum_{i=1}^{N} P(X = x_i, Y = y_j) \cdot A_k(x_i) \cdot B_l(y_j)}{\sum_{j=1}^{M} \sum_{i=1}^{N} A_k(x_i) \cdot B_l(y_j)} \ . \tag{1}$$

The elements U_{kl} are called *components of F-transform*.

In Figure 2 we illustrate F-transform computation for a single variable X (Referendum about EU) resulting in the vector

$$(U_1, U_2, U_3, U_4, U_5, U_6, U_7) = (\ 0.168,\ 0.101,\ 0.101,\ 0.144,\ 0.087,\ 0.107,\ 0.197\ ) \ .$$

We can try to reconstruct the original probability distribution P and in this way get an approximation $\hat{P}$ of P by the application of the inverse F-transform to its F-transform represented by matrix U.

[5] In all our experiments we assume that standard deviations are equal, i.e. $\sigma_X^1 = \sigma_Y^1 = \ldots = \sigma_X^7 = \sigma_Y^7$ and also the correlation $\rho_{X,Y}$ between X and Y in the basic function $f_{i,j}$ is zero.

[6] The generalization to more dimensional probability distributions is straightforward.

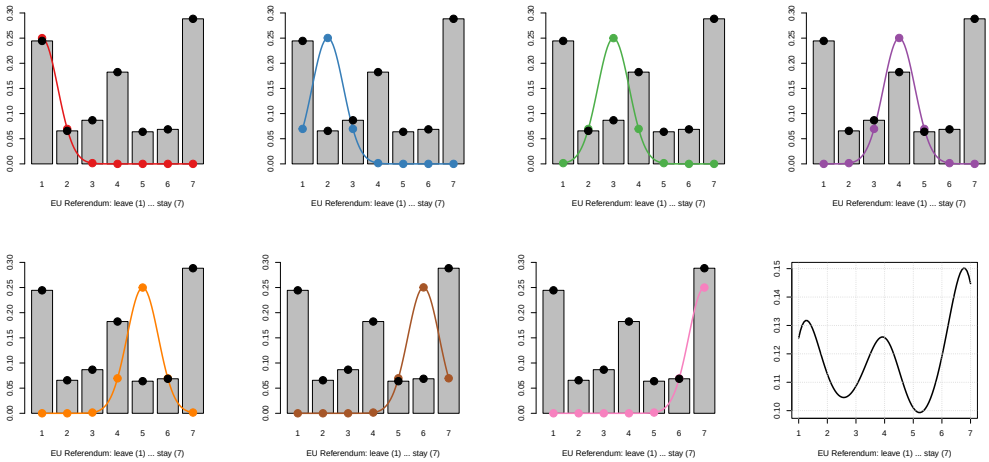

Figure 2: Illustration of the F-transform computations using Gaussian basic functions. The values denoted by dots from the Gaussian curve and from the corresponding columns are multiplied and all products are summed up. The last plot on the bottom right is the result of the inverse F-transform for the full interval $[1, 7]$.

Definition 3. The *inverse F-transform* maps U to $\hat{P} : \mathbb{X} \to [0, 1]$ with respect to $\{A_1, \ldots, A_n\}$ and $\{B_1, \ldots, B_m\}$ and it is defined as follows:

$$\hat{P}(X = x_i, Y = y_j) \quad = \quad \frac{\sum_{k=1}^{n} \sum_{l=1}^{m} U_{kl} \cdot A_k(i) \cdot B_l(j)}{\sum_{k=1}^{n} \sum_{l=1}^{m} A_k(i) \cdot B_l(j)} \quad . \tag{2}$$

In Figure 3 we compare the square root error of the F-transform approximations on a regular 7x7 and 3x3 grids with respect to the original distribution for two variables (Geopolitcs and Referendum about EU), both reported using the seven-value Likert scale. It is not surprising that the F-transform on the 7x7 grid is able to fit perfectly the original table if the fuzziness is very low. More interesting and useful are approximations on the sparser grid 3x3. In this way we achieve a compression of the original table and since the application of F-transform corresponds to smoothing we can expect that these compressed probability tables represent well the actual values. In our experiment we use triangular and Gaussian fuzzy membership functions.

In Figure 4 we can compare the original probability distribution having 7x7 values with its approximations using F-transform on a regular 3x3 grid. The Gaussian membership function has both sigma values equal to 0.8. The triangular fuzzy membership function has both width values equal to 3. The Π-curves membership function has the parameter $s = 1.8$. The Gaussian and Π-curves membership functions seem preferable for the consequent computations since their better correspond to the original table.

With the increasing value of the standard deviation σ the number of modes of the probability distribution is decreasing. We note that rather than an optimization task to

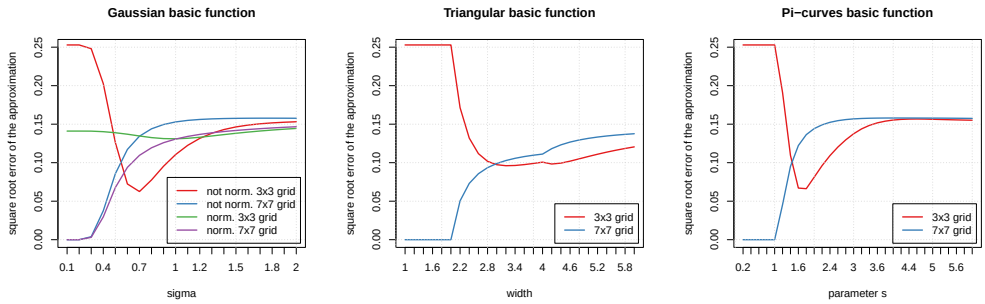

Figure 3: Square root error of the F-transform approximations on a regular 7x7 and 3x3 grids using Gaussian (left), triangular (middle), and Π-curves (right) basic functions.

be solved it is a task for a domain expert to decide how the variables relate, e.g., how many modes the probability distribution is expected to have.

4 BN inference using F-transform

To illustrate the application of F-transform for probabilistic inference in Bayesian networks we use the dataset from the survey "Dividing Lines in Czech Society" (Buchtík, 2023). This survey contained 30 pairs of contradictory statements related to Czech society, denoted $A_1, \ldots, A_{30}$, 14 pairs of contradictory statements that relate to the respondents personally, denoted $D_1, \ldots, D_{14}$, and answers to several demographic questions. The dataset contains answers of 1661 respondents. We performed a projection on the original dataset, retaining only four variables:

A_{15} Geopolitics: integral part of Western Europe (1) vs. neutral bridge (7),

A_{18} Referendum about EU: leave (1) vs. stay (7),

A_{19} Czech Society after 1989: the right direction (1) vs. the wrong direction (7), and

D_{11} Personal benefits from the EU membership: beneficial (1) vs. non-beneficial (7).

In Figure 5, we present the Bayesian network structure representing the relationships between these four variables learned by optimizing the BIC score.

For the illustration of the suggested method we will use a simple but still interesting scenario. Assume two of three child variables were observed and we want to compute the conditional probability of the parent variable and the remaining child. Without loss of generality assume that $A_i = a$ and $A_j = b$ was observed and we want to compute conditional probability distributions $P(D|A_i = a, A_j = b)$ and $P(A_k|A_i = a, A_j = b)$. We

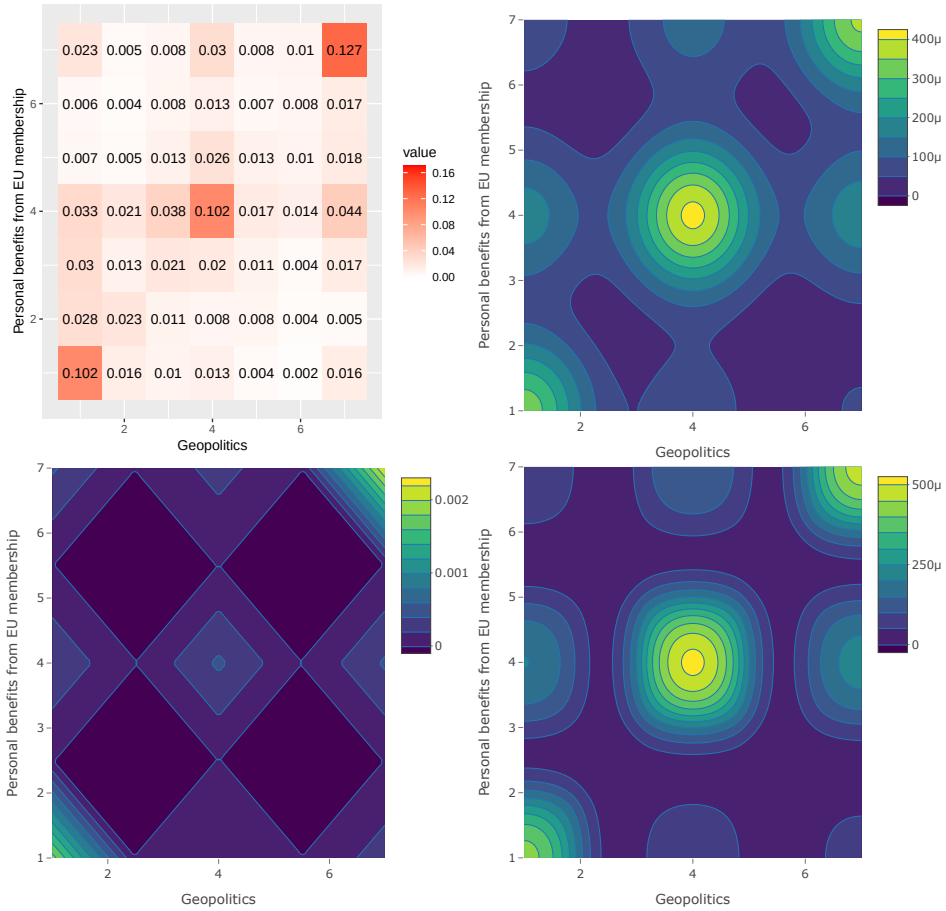

Figure 4: Comparison of the original probability table (top) and its F-transform approximations on a regular 3x3 grid using Gaussian (topright), triangular (bottomleft) and Π-curves (bottomright) membership functions.

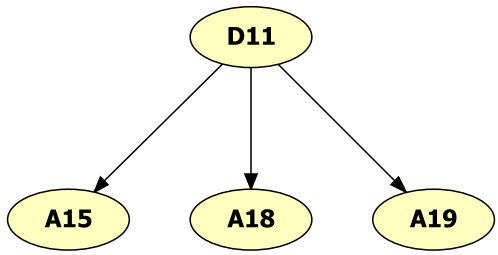

Figure 5: Bayesian network representing relations between four variables from the "Dividing Lines in Czech Society" survey.

perform the following computations for all combinations of values c of A_k and d of D:

$$\psi(D=d) = P(A_i=a|D=d)\cdot P(A_j=b|D=d)\cdot P(D=d) \quad (3)$$

$$p = \sum_d \psi(D=d) \quad (4)$$

$$P(D=d|A_i=a,A_j=b) = \frac{1}{p}\cdot\psi(D=d) \quad (5)$$

$$P(A_k=c|A_i=a,A_j=b) = \sum_d P(A_k=c|D=d)\cdot P(D=d|A_i=a,A_j=b) \ , \quad (6)$$

where p denotes the probability of evidence $P(A_i=a,A_j=b)$ and serves as the normalization constant. In the context of our example, assume $A_i=A_{15}$, $A_j=A_{18}$, $A_k=A_{19}$, and $D=D_{11}$. Then the formulas (3), (4), and (5) correspond to the multiplication of one row of probability table $P(A_{15},D_{11})$ corresponding to $A_{15}=a$ with one row of $P(A_{18},D_{11})$ corresponding to $A_{18}=b$ and normalizing the resulting vector. The formula (6) is multiplication of each column of probability table $P(A_{19},D_{11})$ with corresponding element of the resulting vector and computing the marginal sums over all columns.

All two-dimensional distributions used in the above computations are parametrized on the $3x3$ grid. Note that instead of $3\cdot 7\cdot 7=147$ parameters required for the original discrete model only $3\cdot 3\cdot 3=27$ parameters are needed. Generally, for complex models the savings using F-transform can be substantially larger and may correspond to a shift from an intractable to a tractable model. in Figure 7 (left) the mean square error of the approximation is presented as a function of σ of the Gaussian basic function for the two versions of F-transform (not-normalized and normalized). We can see that the lowest error is achieved by the not-normalized version with $\sigma=0.8$. On the right hand side of this figure an example of the comparison of a conditional probability distribution in the original model and its approximation is presented. In the plot $P(A_{19}|A_{15}=1,A_{18}=3)$ is presented. Note that the observation $A_{18}=3$ is not on the 3x3 grid but it does not pose any problem in the inference algorithm since using the inverse F-transform the probability distribution given any value of A_{18} from interval $[1,7]$ can be estimated.

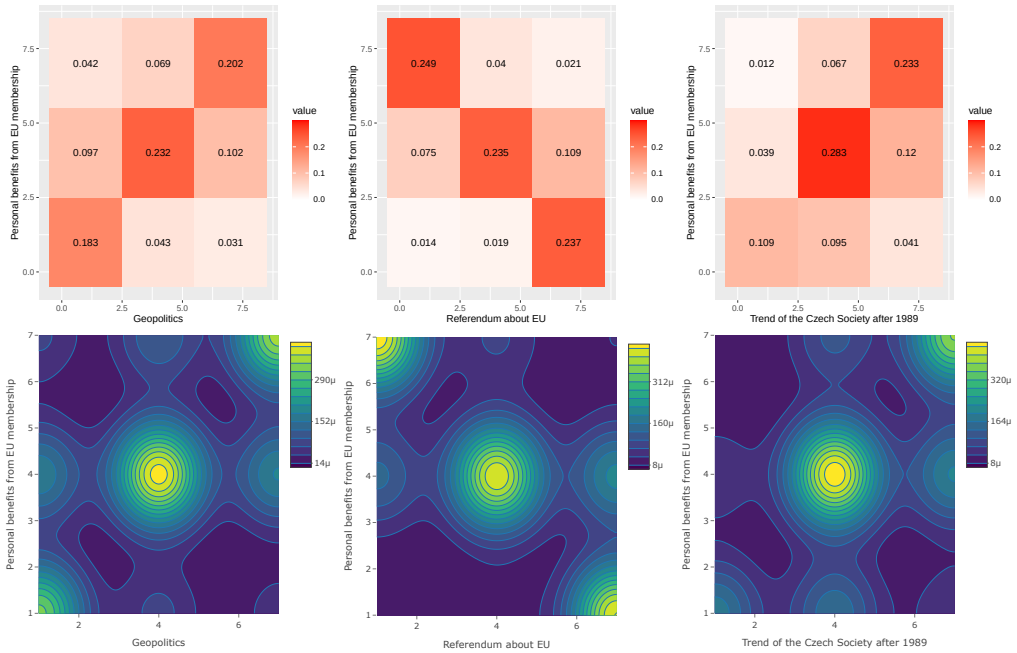

Figure 6: Matrices of the F-transform on a 3x3 grid (top) and their corresponding approximations after the inverse F-transform (bottom).

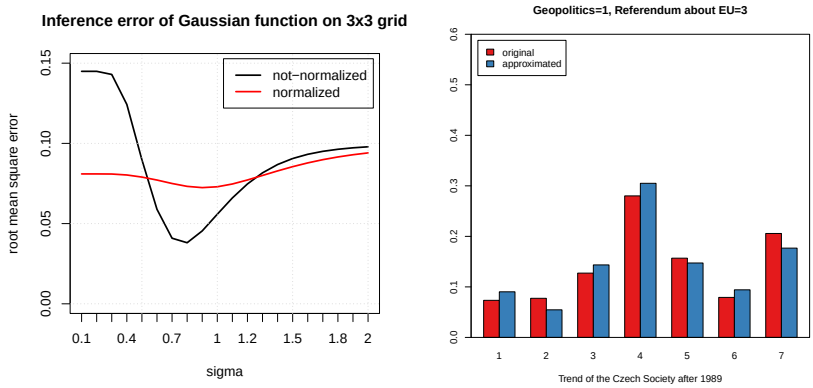

Figure 7: The root mean square error of the approximation as a function of σ of the Gaussian basic function (left) and the comparison of a conditional probability distribution in the original model and its approximation (right).

5 Conclusions

We have argued that Likert scale can be naturally interpreted using fuzzy sets. F-transform can be used to approximate probability tables in Bayesian networks, which can lead to computationally more efficient probabilistic inference. However, this paper represents a starting point and future developments in several directions are necessary.

Efficient inference procedure is still to be described and implemented. We would like to use keypoints instead of regular grids selected so that they represent well the modelled probability distribution with least memory requirements. A procedure for the combination of two probability tables with different keypoints is to be devised. Another dimension to be explored are more general shapes of membership functions in more dimensional transformations, e.g. using general covariance matrix instead of just a diagonal one.

The F-transform representation of the probability distribution resembles Kernel Density Estimates (Rosenblatt, 1956; Silverman, 1986). They differ in that the Kernel Density Estimates place basic functions at every datapoint in the dataset. However, some of the procedures used there may be useful also for the F-transform representation.

Acknowledgment

Developed within the project of the University of Ostrava: Social Dimension of New Technologies in the Energy Sector in the Ostrava Metropolitan Area (reg. number CZ.02.01.01/00/23_021/000859), with financial support from the European Union through the Jan Amos Komenský Operational Programme.

References

M. Buchtík. *Různá vyprávění o jedné společnosti*. Friedrich-Ebert-Stiftung and Masarykova demokratická akademie Praha, 2023.

M. A. Gil and G. González-Rodríguez. Fuzzy vs. likert scale in statistics. In E. Trillas, P. P. Bonissone, L. Magdalena, and J. Kacprzyk, editors, *Combining Experimentation and Theory: A Hommage to Abe Mamdani*, pages 407–420. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-24666-1. URL `https://doi.org/10.1007/978-3-642-24666-1_27`.

F. V. Jensen and T. D. Nielsen. *Bayesian Networks and Decision Graphs*. Information Science and Statistics. Springer New York, NY, 2 edition, 2007. doi: 10.1007/978-0-387-68282-2.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

H. Pan and L. Liu. Fuzzy Bayesian networks—a general formalism for representation, inference and learning with hybrid Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(07):941–962, 2000.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

I. Perfilieva. Fuzzy transforms: Theory and applications. *Fuzzy Sets and Systems*, 157: 993–1023, 2006. URL https://doi.org/10.1016/j.fss.2005.11.012.

M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956. URL https://doi.org/10.1214/aoms/1177728190.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*, volume 26 of *Monographs on Statistics and Applied Probability*. Chapman and Hall/CRC, London, 1986. ISBN 978-0412246203. URL https://doi.org/10.1201/9781315140919.