Cooperation via sharing of probabilistic information

Miroslav Kárný* and Tatiana V. Guy

Institute of Information Theory and Automation,
Academy of Sciences of the Czech Republic,
P.O. Box 18, 182 08 Prague, Czech Republic
E-mail: school@utia.cas.cz
E-mail: guy@utia.cas.cz
*Corresponding author

Antonella Bodini and Fabrizio Ruggeri

Institute of Applied Mathematics and Information Technology,
National Research Council,
via E. Bassini 15, I-20133 Milan, Italy
E-mail: anto@mi.imati.cnr.it
E-mail: fabrizio@mi.imati.cnr.it

Abstract: The paper concerns a cooperation problem in multiple participant decision making (DM). A fully scalable cooperation model with individual participants being Bayesian decision makers who use fully probabilistic design of the optimal decision strategy is presented. The solution suggests a flat structure of cooperation, where each participant interacts with several 'neighbours'. The cooperation consists in providing probabilistic distributions a participant uses for its DM. The group DM is then determined by a way of exploitation of the offered non-standard (probabilistic) fragmental information.

The paper proposes a systematic procedure by formulating and solving the exploitation problem in a Bayesian way.

Keywords: Bayesian decision making; multiple participant decision making; sharing of probabilistic information; cooperation.

Reference to this paper should be made as follows: Kárný, M., Guy, T.V., Bodini, A. and Ruggeri, F. (2009) 'Cooperation via sharing of probabilistic information', *Int. J. Computational Intelligence Studies*, Vol. 1, No. 2, pp.139–162.

Biographical notes: Miroslav Kárný is a Senior Researcher at the Institute of Information Theory and Automation (UTIA), Academy of Sciences of the Czech Republic, Prague, Czech Republic. He obtained his PhD in 1977 and DrSc in 1990 from Academy of Sciences of the Czech Republic. He is with UTIA since 1973, from 1990 as the Head of Department of Adaptive Systems. His research interests cover various theoretical and application aspects of dynamic Bayesian decision making under uncertainty.

Tatiana V. Guy received her DiplEng from Kiev Polytechnic Institute, USSR in 1991 and PhD from the Czech Technical University, Prague in 1999. She is a Researcher at Department of Adaptive Systems, Institute of Information Theory and Automation. Her research interests include modelling, adaptive control and decision making.

Antonella Bodini is a Researcher at the Institute of Applied Mathematics and Information Technology (IMATI), National Research Council, Milan, Italy. She obtained her PhD in 2002 from University of Pavia (Italy). She is with IMATI since 2001. Her research interests cover various theoretical aspects of statistical inference, modelling uncertainty and their application to ecological and climatic problems mainly.

Fabrizio Ruggeri received his BSc in Milan, MSc in Carnegie Mellon and PhD in Duke. He is the Research Director at CNR and Faculty in a PhD programme in Pavia. He is the former ENBIS President. He is the Editor in Chief of *Applied Stochastic Models in Business and Industry* and *Encyclopaedia of Statistics in Quality and Reliability*. He is an author of more than 100 papers and Editor of two books. He has been involved in many conferences and summer schools, mostly in Bayesian or industrial statistics.

1 Introduction

Decision making (DM) theory provides a good framework for formalisation and conceptual solution of many problems arising in such complex areas as estimation, prediction, pattern recognition, fault detection, control, etc. There is a number of sophisticated and well-elaborated approaches and techniques developed for DM, (DeGroot, 1970; Bell et al., 1988) proven to be successful in many applications (see e.g., Dyer et al., 1992; Quinn et al., 2003). However none of the approaches can serve as a universal one to be applied to the above-mentioned diversity of problems.

One of the reasons is that to cover such a broad spectrum of problems the successful DM theory should be able to cope with:

1 uncertainty, which is (almost) always present

2 incomplete information available

3 multiple aims

4 always limited abilities to sense, act and evaluate.

The standard Bayesian DM theory (DeGroot, 1970; Berger, 1985) meets the requirements 1 and 2, and the theory of fully probabilistic DM (Kárný and Guy, 2006; Kárný, 2008), seems to satisfy the requirement 3. The requirement 4 can be generally met only via a distributed DM. In spite of the numerous attempts, e.g., Wolpert and Tumer (1999, 2001), Vlassis (2003), Yiming et al. (2003), Stirling (2004), etc., there is no systematic, widely accepted, solution.

A group of decision makers (called here *participants*), acting within a flat structure (neither coordinator nor hierarchy is assumed), is considered in the paper. Each participant is characterised by limited ability to perceive, to model and to evaluate and can extend this ability only mildly; for a discussion see Section 2. The cooperation and interaction structure of a group of participants is assumed to be fixed.

The participant solves DM task to achieve its particular DM aim with respect to its *environment*, which often contains other participants. The participant is Bayesian and applies fully probabilistic design (FPD) (Kárný and Guy, 2006), to generate its DM strategy. Participant's operation domain has non-empty intersection with operation domains of a few other participants, so-called *neighbours* that have different DM aims and solve own DM tasks. DM activity of neighbours has non-negligible (sometimes rather strong) influence on the participant's DM and the participant tends to cooperate with the neighbours, in order to improve the effectiveness of its DM.

How does the cooperation work? Each participant employing FPD operates with several DM elements, expressed in probabilistic form, namely: model of its environment, DM aim and DM strategy. The *cooperation* in the present context implies that cooperating participants exchange/share their DM elements. As the operation domains of the cooperating participants do not overlap completely, the DM elements offered for exchange represent *partial* and *imprecise* information pieces about the uncertain and random environment. The participant then needs a methodology, which effectively solves the problem *how to combine (merge) fragmental information pieces* provided by its neighbours. This problem has been addressed repeatedly, for instance, in connection with probabilistic expert systems (Cowell et al., 2003), knowledge elicitation, (O'Hagan et al., 2006), cooperation of participants (Andrýsek et al., 2007; Kárný et al., 2007); etc. However, none of the existing solutions seems to be complete and automatic enough to be applied in the cooperation task concerned.

The paper proposes a merging methodology applicable to the problem formulated above. The basic idea '*take all offered information pieces as outputs of noisy information channels and try to estimate parameters of the underlying source*' is not new, for example see, Oakley and O'Hagan (2005) and references there. Unlike the attempts made elsewhere, the proposed approach is not case-specific and covers the simple methodological line:

1. consider the uncorrupted information source behind as a measuring device described by an unknown parameter
2. consider the available information pieces as data about the parameter
3. model the relation of data and parameter, i.e., choose suitable likelihood function and prior distribution
4. perform the standard Bayesian point estimation to learn the parameter.

Section 2 outlines decision-making scenario in FPD setting as well as introduces necessary terms and assumptions. Section 3 and Section 4 form the core of the paper and describe the methodology proposed. Section 5 illustrates the results methodology yields. The design of the methodology dominates over technical and computational aspects of the problem. Consequently, the number of open problems is large, see Section 6.

2 The cooperation scenario

The section introduces the considered decision-making scenario together with the basic assumptions. The text below relies on the participant's description and the cooperation structure outlined in Introduction.

The term participant (decision maker) generally refers either to a single decision maker or to a group of decision makers acting as a whole. Both an electronic decision maker and the human being as well as their combination can be considered.

The following basic notation is adopted throughout the paper. $k^* = \{k_1, k_2, \ldots\}$ stands for a set of ks. The cardinality of the set k^* is denoted $\mathring{k}$ and provides the number of elements of k^*. If k is a vector, k' denotes the transpose of k. $f(x)$ is a probability density function (pdf) of a random variable $x, f(\cdot|\cdot)$ stands for a conditional pdf. In Section 3, where relations of pdfs' values are modelled, symbols $G(\cdot), G(\cdot|\cdot)$ are used for pdfs describing this 'higher level' model. $\hat{x}$ denotes an estimate of x.

Note that generally decisions are made repeatedly throughout time course. For the solution of the considered merging problem, time is not important and hence the time index is omitted throughout the paper.

2.1 Participant and its DM

The participant interacts with its environment in a *closed decision-making loop* – formed by coupling of the participant with its environment – in order to influence the environment. To solve this task, the participant deals with a finite collection of random variables describing the behaviour of the decision-making loop.

- Actions selected by the participant and influencing the environment.
- Available observations that inform the participant about the state of the closed decision-making loop.
- Internal states that participant considers but does not observe directly.

The distribution of the *behaviour's realisation* $\mathfrak{b} \in \mathfrak{b}^*$ is described by a (regular) pdf $f(\mathfrak{b})$ defined with respect to a dominating measure $d\mathfrak{b}$ (typically, Lebesgue or counting one).

The participant designs and applies a causal decision rule, which maps a realisation $\mathfrak{p} \in \mathfrak{p}^*$ of the random *past* history on a realisation $\mathfrak{a} \in \mathfrak{a}^*$ of the participant's admissible action. By definition, the realisation of the past history $\mathfrak{p}$ is *known* when an action $\mathfrak{a}$ is chosen. The participant makes its decisions under uncertainty about unknown realisations $\mathfrak{f} \in \mathfrak{f}^*$ of the *future* behaviour.

Note that formally all considered but unobserved variables are included into the future behaviour.

Let us decompose a realisation of the closed-loop behaviour into three components as follows:

$$\mathfrak{b} = (\mathfrak{f}, \mathfrak{a}, \mathfrak{p}). \tag{1}$$

The uncertainty of the participant is then expressed by non-voidness of $\mathfrak{f}$ in (1).

For any (1), the pdf $f(\mathfrak{b})$ can be factorised:

$$f(\mathfrak{b}) \equiv f(\mathfrak{f}, \mathfrak{a}, \mathfrak{p}) = f(\mathfrak{f}|\mathfrak{a}, \mathfrak{p}) f(\mathfrak{a}|\mathfrak{p}) f(\mathfrak{p}). \tag{2}$$

In (2) $f(\mathfrak{f}|\mathfrak{a}, \mathfrak{p})$ models the participant's environment, $f(\mathfrak{a}|\mathfrak{p})$ represents a randomised decision rule of the participant and $f(\mathfrak{p})$ describes the past history.

The participant tries to achieve its DM aim with respect to its environment under given constraints. Bayesian nature of the participant and employment of FPD for design of DM rule yield that the participant:

- relies on a chosen environment model $f(\mathfrak{f}|\mathfrak{a},\mathfrak{p})$
- expresses its DM aim and existing constraints via so-called *ideal pdf*, ${}^{I}f(\mathfrak{b})$, which describes desirable realisation of the closed-loop behaviour $\mathfrak{b}$
- selects a randomised decision rule $f(\mathfrak{a}|\mathfrak{p})$ as a minimiser of a proximity measure of $f(\mathfrak{b})$ and ${}^{I}f(\mathfrak{b})$.

Kárný (2008) provides arguments why FPD uses the Kullback-Liebler divergence

$$D=\left(f(\mathfrak{b})\middle\|{}^{I}f(\mathfrak{b})\right)=\int f(\mathfrak{b})\ln\left(\frac{f(\mathfrak{b})}{{}^{I}f(\mathfrak{b})}\right)\mathrm{d}\mathfrak{b} \tag{3}$$

as a measure of proximity of $f(\mathfrak{b})$ and ${}^{I}f(\mathfrak{b})$. In (3), $\int\cdot\,\mathrm{d}\mathfrak{b}$ means a definite multivariate integration over the integrand's domain. The optimal decision rule is then

$${}^{o}f(\mathfrak{a}|\mathfrak{p})\in\operatorname{Arg}\min_{\{f(\mathfrak{a}|\mathfrak{p})\}^{*}} D=\left(f(\mathfrak{b})\middle\|{}^{I}f(\mathfrak{b})\right). \tag{4}$$

The following proposition summarises the solution of FPD.

Proposition 1: (Solution of FPD). The optimal decision rule (4) is given by the formula

$${}^{o}f(\mathfrak{a}|\mathfrak{p})\quad\propto\quad {}^{I}f(\mathfrak{a}|\mathfrak{p})\exp\left[-\omega(\mathfrak{a}|\mathfrak{p})\right],\quad \omega(\mathfrak{a}|\mathfrak{p})\equiv\int f(\mathfrak{f}|\mathfrak{a},\mathfrak{p})\ln\left(\frac{f(\mathfrak{f}|\mathfrak{a},\mathfrak{p})}{{}^{I}f(\mathfrak{f}|\mathfrak{a},\mathfrak{p})}\right)\mathrm{d}\mathfrak{f}, \tag{5}$$

where $\propto$ denotes an equality without specifying a unique normalising factor.

Proof: The proof follows from the following string of equalities and from the fact that the Kullback-Leibler divergence $D=(f\|h)$ reaches its smallest zero value for $f=h$, Vajda (1989). In the string, the respective $f(\cdot)$ and their ideal counterparts ${}^{I}f(\cdot)$ are regular pdfs with respect to the measures $\mathrm{d}\mathfrak{b},\mathrm{d}\mathfrak{p},\mathrm{d}\mathfrak{a},\mathrm{d}\mathfrak{f}$.

$$D=\left(f\middle\|{}^{I}f\right)=\int f(\mathfrak{b})\ln\left(\frac{f(\mathfrak{b})}{{}^{I}f(\mathfrak{b})}\right)\mathrm{d}\mathfrak{b}=\int\int\int f(\mathfrak{f},\mathfrak{a},\mathfrak{p})\ln\left(\frac{f(\mathfrak{f},\mathfrak{a},\mathfrak{p})}{{}^{I}f(\mathfrak{f},\mathfrak{a},\mathfrak{p})}\right)\mathrm{d}\mathfrak{f}\,\mathrm{d}\mathfrak{a}\,\mathrm{d}\mathfrak{p}$$

$$\underbrace{=}_{(*)}\int f(\mathfrak{p})\left\{\ln\left(\frac{f(\mathfrak{p})}{{}^{I}f(\mathfrak{p})}\right)+\int f(\mathfrak{a}|\mathfrak{p})\left[\ln\left(\frac{f(\mathfrak{a}|\mathfrak{p})}{{}^{I}f(\mathfrak{a}|\mathfrak{p})}\right)+\underbrace{\int f(\mathfrak{f}|\mathfrak{a},\mathfrak{p})\ln\left(\frac{f(\mathfrak{f}|\mathfrak{a},\mathfrak{p})}{{}^{I}f(\mathfrak{f}|\mathfrak{a},\mathfrak{p})}\right)\mathrm{d}\mathfrak{f}}_{\equiv\omega(\mathfrak{a}|\mathfrak{p})}\right]\mathrm{d}\mathfrak{a}\right\}\mathrm{d}\mathfrak{p}$$

$$=\int f(\mathfrak{p})\ln\left(\frac{f(\mathfrak{p})}{{}^{I}f(\mathfrak{p})}\right)\mathrm{d}\mathfrak{p}-\ln\left\{\int {}^{I}f(\mathfrak{a}|\mathfrak{p})\exp\left[-\omega(\mathfrak{a}|\mathfrak{p})\right]\mathrm{d}\mathfrak{a}\right\}+\int f(\mathfrak{p})D\left(f(\mathfrak{a}|\mathfrak{p})\middle\|{}^{o}f(\mathfrak{a}|\mathfrak{p})\right)\mathrm{d}\mathfrak{p}$$

The validity of (*) is given by Fubini theorem. The sum of the first and the second term in the last equality is independent of the optimised $f(\mathfrak{a}|\mathfrak{p})$.

Remark 1:

- The ideal pdf ${}^{I}f(\mathfrak{b})$ describes constraints on actions as the support of the optimal decision rule (5) is contained in the support of its ideal counterpart.
- The explicit formula of the optimal decision rule is obtained even when designing decision strategy, i.e., a sequence of decision rules, Kárný and Guy (2006).

2.2 Cooperating participant

A number of participants acting within a fixed cooperation and interaction structure are assumed. Neither coordinator nor hierarchy is supported by the structure. All participants are Bayesian participants and use the FPD, see Section 2.1.

Assumption 1 summarises additional assumptions about the participant that are necessary to build (relatively) realistic, normative and scalable cooperation scheme.

Assumption 1 The participant of given abilities

A The participant, labelled by $k \in \{1,2,3,\ldots\}$, has fixed DM elements consisting of:

$\mathfrak{b}_k^* \equiv$ the set of behaviour's realisations considered by the kth participant

$\mathfrak{p}_k^* \equiv$ the set of past's realisations considered by the kth participant

$\mathfrak{a}_k^* \equiv$ the set of action's realisations considered by the kth participant

$\mathfrak{f}_k^* \equiv$ the set of future's realisations considered by the kth participant

$f_k(\mathfrak{f}_k|\mathfrak{a}_k, \mathfrak{p}_k) \equiv$ the pdf chosen by the kth participant for modelling its environment, (2)

$f_k(\mathfrak{a}_k|\mathfrak{p}_k) \equiv$ the pdf describing the decision rule considered by the kth participant, (2)

$f_k(\mathfrak{p}_k) \equiv$ the pdf describing the past related to the kth participant, (2)

${}^{I}f_k(\mathfrak{b}_k) \equiv$ the ideal pdf expressing DM aims and constraints of the kth participant.

B Any participant *k* has a few fixed neighbours, i.e., participants having non-empty intersection of behaviours' sets with $\mathfrak{b}_k^*$.

C All participants are selfish. It means that the only participant can choose DM elements according to its DM task. None of its neighbours can directly influence or change the participant's DM elements.

The success of DM heavily depends on quality of modelling of the participant's environment, which includes active neighbours. The reachable quality of modelling is restricted by the natural limitations on the participant's ability to model and evaluate. To impose on cooperation, the proposed methodology should respect these limitations and suppose no extra demands on the participant's ability.

Assumption 2 defines an overall mechanism of the participant's cooperation based upon sharing DM elements with its neighbours.

Assumption 2 Cooperation via sharing DM elements

A Each participant *k* offers some projections (marginal or conditional pdfs) of $f_k(\mathfrak{b}_k)$ and ${}^I f_k(\mathfrak{b}_k)$ to its neighbours while they offer theirs.

This ensures involvement of yet selfish, but cooperating participants.

B Each participant allocates a 'reasonable' part of its evaluation abilities to modify its pdfs $f_k(\mathfrak{b}_k)$ and ${}^I f_k(\mathfrak{b}_k)$ by the offered probabilistic information.

The participant is assumed to be provided with a universal tool realising the modification methodology. The adjective 'reasonable' anticipates that the participant is able to perform the proposed evaluations. This excludes generic use of the theory of incomplete (Bayesian) games (Harsanyi, 2004), which force the participant to model not only its environment but its neighbours too.

C There is no mechanism either to distinguish reliability of information offered or to judge/compare the importance of information offered.

This is the most restrictive assumption and models insufficient reasons for making distinctions. This assumption can be relaxed in refined cooperation schemes, where the participant extends its actions by an active weighting of the offered information pieces.

3 Cooperation methodology

A cooperating participant is equipped by a special tool, which automatically processes the information offered by its neighbours. Hence, the cooperating participant gets an opportunity to improve quality of its DM via exploiting information provided by the neighbours.

Let us consider the kth participant, $k \in k^* \equiv \left\{1, 2, \ldots, \mathring{k}\right\}$. It is selfish but willing to cooperate. A fixed finite set k^* of its neighbours (see Assumption 1 (B)) acts on the union of their behaviours' sets, i.e.

$$\mathfrak{b}^* \equiv \cup_{k \in k^*} \mathfrak{b}_k^*. \tag{6}$$

The combination of information pieces related to the closed decision loop models, $f_k(\mathfrak{b}_k), k \in k^*$, is performed in the same way as those concerned with the ideal pdfs, ${}^I f_k(\mathfrak{b}_k), k \in k^*$. An instance of informational pieces related to models $f_k(\mathfrak{b}_k)$ is treated below. Processing of the information related to ${}^I f_k(\mathfrak{b}_k)$ can be done similarly.

3.1 Combination of pdfs as Bayesian estimation

Let us consider the collection of neighbours as a *multivariate participant* dealing with a *global* closed decision loop model $f(\mathfrak{b})$. The action realisation of this multivariate participant is $\mathfrak{a} = (\mathfrak{a}_1, \ldots, \mathfrak{a}_k)$, while its past $\mathfrak{p}$ and future $\mathfrak{f}$ realisations are composed of $\{\mathfrak{p}_k\}_{k \in k^*}$ and $\{\mathfrak{f}_k\}_{k \in k^*}$ in a complex way whose details are unimportant within the considered context. The global pdf $f(\mathfrak{b})$ is unknown to particular neighbours and the projections of imprecise (noisy) versions $f_k(\mathfrak{b})$ of the global pdf $f(\mathfrak{b}), \mathfrak{b} \in \mathfrak{b}^*$ are processed.

For a fixed $k \in k^*$, let us decompose $\mathfrak{b}$ in a slightly different way than before. This decomposition reflects a complement of $\mathfrak{b}_k$ to $\mathfrak{b}$ by introducing the part $\mathfrak{u}_k$ unconsidered by the kth participant:

$$\mathfrak{b} \equiv (\mathfrak{f}_k, \mathfrak{a}_k, \mathfrak{p}_k, \mathfrak{u}_k) \equiv (\text{future, action, past, unconsidered variables}) \tag{7}$$

Then, the general form of the projections to environments' models is

$$f_k(\mathfrak{f}_k | \mathfrak{a}_k, \mathfrak{p}_k) = \frac{\int f_k(\mathfrak{f}_k, \mathfrak{a}_k, \mathfrak{p}_k, \mathfrak{u}_k) \mathrm{d}\mathfrak{u}_k}{\int f_k(\mathfrak{f}_k, \mathfrak{a}_k, \mathfrak{p}_k, \mathfrak{u}_k) \mathrm{d}\mathfrak{f}_k \mathrm{d}\mathfrak{u}_k}. \tag{8}$$

To get an environment model employed by the kth participant, the joint pdf $f_k(\mathfrak{b})$ is processed in the non-linear but deterministic way (8). Exploitation of this fact helps to reach (relatively) low computational complexity of the proposed combination of pdfs: the formula (8) is applied to the final combination only.

The combination is constructed supposing the joint pdfs $f_k(\mathfrak{b}), k \in k^*$, are available for all $\mathfrak{b} \in \mathfrak{b}^*$. The general case, when projections $f_k(\mathfrak{f}_k | \mathfrak{a}_k, \mathfrak{p}_k)$ of the joint pdfs $f_k(\mathfrak{b})$ are only available, is treated below.

3.2 Point estimate of the global pdf

The considered construction of the pdfs' combination (merging) casts the problem into the standard Bayesian framework.

Let us consider an unknown global pdf $f(\mathfrak{b}), \mathfrak{b} \in \mathfrak{b}^*$ as an unknown parameter. Then, the pdfs $f_k(\mathfrak{b})$, $\mathfrak{b} \in \mathfrak{b}^*$, $k \in k^*$, represent noisy observations (data) of that parameter by particular observing devices (i.e., participants). The point estimate $\hat{f}(\mathfrak{b})$ of the unknown parameter $f(\mathfrak{b})$ based on the observations $f_k(\mathfrak{b}), \mathfrak{b} \in \mathfrak{b}^*, k \in k^*$, is the 'natural' Bayesian merger searched for.

A point estimate $f\left(\mathfrak{b}\middle|\hat{\mathcal{V}}\right)$ of $f(\mathfrak{b}), \mathfrak{b} \in \mathfrak{b}^*$, within a class of estimators $\left\{f\left(\mathfrak{b}\middle|\mathcal{V}\right)\right\}_{\mathcal{V}\in\mathcal{V}^*}$ given by the statistic $\mathcal{V} \in \mathcal{V}^*$ is searched for.

Assumption 3 'Localness principle'. The choice of $f\left(\mathfrak{b}\middle|\hat{\mathcal{V}}\right)$ and estimation of $f(\mathfrak{b})$ at a particular $\underline{\mathfrak{b}} \in \mathfrak{b}^*$ depend only on the relation of values $f(\underline{\mathfrak{b}})$ and $f_k(\underline{\mathfrak{b}})$, $k \in k^*$.

If we stay within Bayesian framework, consider proper smooth loss functions and adopt Assumption 3, then the best point estimate $f\left(\mathfrak{b}\middle|\hat{\mathcal{V}}\right)$ of $f(\mathfrak{b})$ is the minimiser of the expected value of the Kullback-Leibler divergence (Bernardo, 1979). Its minimisation is equivalent to the minimisation of the expected value of the Kerridge inaccuracy, (Kerridge, 1961). It gives the rule for selecting the best $\hat{\mathcal{V}}$

$$\hat{\mathcal{V}} \in \text{Arg}\min_{\mathcal{V}\in\mathcal{V}^*} -\int_{\mathfrak{b}\in\mathfrak{b}^*} \mathsf{E}\left[f(\mathfrak{b})\middle|f_k(\mathfrak{b}), k \in k^*\right] \ln\left(f\left(\mathfrak{b}\middle|\mathcal{V}\right)\right)\mathrm{d}\mathfrak{b}, \tag{9}$$

where the expectation is taken over the uncertainty about values of the global pdf $f(\mathfrak{b})$.

Thus, the construction of $f\left(\mathfrak{b}\middle|\hat{\mathcal{V}}\right)$ relies on the evaluation of the conditional expectation

$$\hat{f}(\mathfrak{b}) \equiv \mathsf{E}\left[f(\mathfrak{b})\middle|f_k(\mathfrak{b}), k \in k^*\right], \forall \mathfrak{b} \in \mathfrak{b}^*. \tag{10}$$

3.3 Knowledge on modelled relation

In order to obtain the conditional expectation (10), we have to relate the data, i.e., $f_k(\mathfrak{b})$, $\mathfrak{b} \in \mathfrak{b}^*$, $k \in k^*$, to the unknown parameter, i.e., global pdf, $f(\mathfrak{b})$, $\mathfrak{b} \in \mathfrak{b}^*$. The model can hardly be unique. Two unsatisfactory and one promising variant are presented below. Their choice was driven by the necessity to have feasible estimation, see Assumption 2 (B).

Let temporarily assume that the treated pdfs have the common domain $\mathfrak{b}^* = \mathfrak{b}_k^*, \forall k \in k^*$, (6). This assumption is relaxed in Section 4.

The following notations are used further in the text:

$$\begin{aligned} &v_k - \text{a } \textit{value} \text{ of the pdf } f_k(\mathfrak{b}) \text{ at fixed } \underline{\mathfrak{b}} \in \mathfrak{b}^* \\ &v - \text{an unknown estimated } \textit{value} \text{ of the global pdf } f(\mathfrak{b}) \text{ at the same } \underline{\mathfrak{b}}. \end{aligned} \tag{11}$$

In (11), v_k denotes a particular data provided by the kth participant for a fixed realisation of behaviour, while v stands for an estimated value of the unknown parameter for the same realisation.

Under the notation (11) the modelled relationship 'pdf offered by the kth participant' – 'unknown global pdf' can be written as $G\left(v_k \middle| v,\cdot\right)$, $G \in G^*$, where G^* is a set of possible models, i.e., pdfs relating v_k to v. As a meaningful model has to respect the nature of the processed data (the pdfs $f_k(\mathfrak{b})$, $k \in k^*$) and of the estimated parameter (the global pdf $f(\mathfrak{b})$), the set G^* generally should relate non-negative functions having the integral equal to unity. The latter condition can be neglected as it is met by normalisation in the final merging.

To select a desired model from G^*, a relevant prior knowledge complemented by additional assumptions on modelling relationships is used. The adopted assumptions are summarised below.

Assumption 4 Conditions on pdf $G\left(v_1,\ldots,v_{\mathring{k}} \middle| v,\cdot\right)$

A There exist finite-dimensional functions $\Psi_1,\ldots,\Psi_{\mathring{k}}$ and $\Omega_1,\ldots,\Omega_{\mathring{k}}$, defining 'generalised moments' such that

$$\begin{aligned}&\mathsf{E}\left[\Psi_k\left(v_k\right)\middle| v,\Psi_1,\ldots,\Psi_{\mathring{k}},\Omega_1,\ldots,\Omega_{\mathring{k}}\right]\equiv\\&\int_{v_k^*}\Psi_k\left(v_k\right)G\left(v_k\middle| v,\Psi_1,\ldots,\Psi_{\mathring{k}},\Omega_1,\ldots,\Omega_{\mathring{k}}\right)\mathrm{d}v_k=\Omega_k\left(v\right),\ k\in k^*,\end{aligned}\tag{12}$$

i.e., the expectation $\mathsf{E}\left[\cdot\middle| v,\Psi_1,\ldots,\Omega_1,\ldots\right]$ is made over the random mechanism generating the values $v_k = f_k(\mathfrak{b})$, $k \in k^*$.

The equalities (12) relate the particular pdfs provided by participants to the global pdf describing the set of neighbours. Respective models, discussed below, differ in the choice of functions Ψ and Ω in (12). The finite-dimensionality requirement respects Assumption 2 (B).

B Values $v_k, k \in k^*$, provided by different participants are conditionally independent for the given global pdf $f(\mathfrak{b})$, i.e., $G\left(v_1,\ldots,v_k \middle| v\right)=\prod_{k\in k^*}G\left(v_k\middle| v\right)$. This assumption postulates independence of personal deviations, which is practically acceptable.

C Values $v_k \equiv f_k(\mathfrak{b})$ of the pdf $f_k(\mathfrak{b})$ for different realisations $\mathfrak{b}_i \neq \mathfrak{b}_j$ of the behaviour are independent when $f(\mathfrak{b}),\Psi_k,\Omega_k$ are given. Hence $G\left(v_k(\mathfrak{b}_1),v_k(\mathfrak{b}_2),\ldots\middle| v\right)=\prod_i G\left(v_k(\mathfrak{b}_i)\middle| v\right)$.

This independence assumption admits significant changes of values v_k provided by the participant k even for slightly differing behaviours.

D According to the maximum entropy principle, the desired model $G\left(v_1,\dots,v_{\mathring{k}} \middle| v,\cdot\right)$, relating the pdfs values $v_1,\dots,v_{\mathring{k}}$ offered by neighbours $k\in k^*$, should possess the highest entropy among those meeting the Assumptions A, B and C.

This assumption guarantees that from the set G^* of possible models, determined by the restrictive assumptions, we choose the pdf $G(\cdot|\cdot)$ with the given 'generalised moments' $\Omega_k(v)$, $k\in k^*$, (12) but sufficiently flat not to forbid large deviations of $\Psi_k(v_k)$ from $\Omega_k(v)$.

Both data v_k, $k\in k^*$, and estimated parameter $f(\mathfrak{b})$ are generally infinite-dimensional and should be treated as random processes. Such a treatment can be avoided as an estimate $\hat{f}(\mathfrak{b})$ of $f(\mathfrak{b})$ specified for *each* realisation of behaviour $\mathfrak{b}\in\mathfrak{b}^*$ (Assumption 3) is searched for.

Let us suppose $\mathring{k}$ participants operating on a finite collection of different behaviours $(\mathfrak{b},\mathfrak{b}_1,\mathfrak{b}_2,\dots,\mathfrak{b}_n)$, $n<\infty$.. Each participant provides $\left(f_k(\mathfrak{b}),f_k(\mathfrak{b}_1),\dots,f_k(\mathfrak{b}_n)\right)$, $k\in k^*$ about the unknown global pdf $f(\mathfrak{b})$. Assumption 4 (B) and Assumption 4 (C) restrict the search of the model to the finite-dimensional distributions of the random vector

$$\underbrace{f_1(\mathfrak{b}),\dots,f_{\mathring{k}}(\mathfrak{b})}_{(*)},\underbrace{f_1(\mathfrak{b}_1),\dots,f_{\mathring{k}}(\mathfrak{b}_1),\dots,f_1(\mathfrak{b}_n),\dots,f_{\mathring{k}}(\mathfrak{b}_n)}_{(**)} \tag{13}$$

with conditionally independent entries. Thus, the distribution of the part of interest (*) in (13) is uninfluenced by the distribution of the extendable part (**). Besides, the constructed parametric model is fully determined by the marginal pdfs $G\left(v_k\middle|f(\mathfrak{b}),\Psi_k,\Omega_k\right)$ on values $v_k=f_k(\mathfrak{b}),k\in k^*$, of the modelled pdfs given by the value $v=f(\mathfrak{b})$ of the estimated global pdf. It holds

$$G\left(v_1,\dots,v_{\mathring{k}}\middle|f(\mathfrak{b}),\Psi_1,\dots,\Psi_{\mathring{k}},\Omega_1,\dots,\Omega_{\mathring{k}}\right)=\prod_{k\in k^*}G\left(v_k\middle|f(\mathfrak{b}),\Psi_k,\Omega_k\right).$$

Assumption 4 (D) defines $G\left(v_k\middle|f(\mathfrak{b}),\Psi_k,\Omega_k\right)$ as an entropy maximiser restricted by linear-in-G constraint implied by Assumption 4 (A). This convex optimisation problem has the solution (Cover and Thomas, 1991),

$$G\left(v_k \left| f(\mathfrak{b}), \Psi_k, \Omega_k\right.\right) = \frac{\exp\left[-\lambda_k'\left(f(\mathfrak{b})\right)\Psi_k\left(v_k\right)\right]}{\int_0^\infty \exp\left[-\lambda_k'\left(f(\mathfrak{b})\right)\Psi_k\left(v_k\right)\right]\mathrm{d}v_k}, \tag{14}$$

with the vector function of Lagrangian multipliers $\lambda_k\left(f(\mathfrak{b})\right)$ solving the equations, $k \in k^*$, $\mathfrak{b} \in \mathfrak{b}^*$,

$$\int \Psi_k\left(v_k\right) \frac{\exp\left[-\lambda_k'\left(f(\mathfrak{b})\right)\Psi_k\left(v_k\right)\right]}{\int_0^\infty \exp\left[-\lambda_k'\left(f(\mathfrak{b})\right)\Psi_k\left(v_k\right)\right]\mathrm{d}v_k}\mathrm{d}v_k = \Omega_k\left(f(\mathfrak{b})\right). \tag{15}$$

The choice of the functions Ψ, Ω has to guarantee that integral in (14) is finite and (15) is solvable. The subsequent sections discuss candidates of this type.

3.4 Exponential model

In the cooperative context, it is 'natural' to assume that values of the pdfs offered by participants $v_k = f_k(\mathfrak{b})$, $k \in k^*$, are unbiased guesses of the values of the unknown global pdf $v = f(\mathfrak{b})$, $\mathfrak{b} \in \mathfrak{b}^*$, i.e.,

$$\mathsf{E}\left[v_k \left| v\right.\right] = v, \forall k \in k^*. \tag{16}$$

When considering this, i.e., choosing Ψ_k, Ω_k in (12) as one-dimensional identities, the constructed parametric model (14), (15) is the exponential one with $\lambda_k = \lambda = 1/v$

$$G\left(v_k \left| v, (\Psi,\Omega) \equiv identities\right.\right) = G\left(v_k \left| v\right.\right) = \chi\left(v_k \geq 0\right)\frac{1}{v}\exp\left[-\frac{v_k}{v}\right], \tag{17}$$

with $\chi(\cdot)$ denoting the indicator function of the set in its argument.

The Bayesian parameter estimation with the model (17) is easy. For the completely flat improper prior distribution on (non-negative) values v of the global pdf $f(\mathfrak{b})$, it has posterior Gamma distribution with $\mathring{k}$ degrees of freedom and expectation equal to the arithmetic mean of the measurements, i.e.,

$$\mathsf{E}\left[v \left| v_k, k \in k^*\right.\right] \equiv \frac{1}{\mathring{k}} \sum_{k \in k^*} v_k. \tag{18}$$

This merging is known as arithmetic pooling of opinions (O'Hagan et al., 2006). It is widely used. But, criticised for its sensitivity to outlying opinions, which make it spread too much. More important, the exponential model neglects the fact that relative errors of small estimated values are usually larger than that of large values. Indeed, the exponential model has the property

$$\frac{\text{standard deviation}}{\text{expectation}} = 1. \tag{19}$$

All that discards exponential model from the further consideration.

3.5 Truncated normal model

To avoid (19), the assumption of unbiasedness (16) is complemented by an explicit requirement on the second moment determined by an additional parameter ρ

$$\mathsf{E}\left[v_k^2 \middle| v, \rho\right] = \rho + v^2, \rho \equiv \rho(\mathfrak{b}) > 0, \forall k \in k^* \tag{20}$$

i.e., $\Psi_k(v_k) = \left[v_k, v_k^2\right]$, $\Omega_k(v) = \left[v, \rho + v^2\right]$.

In this case, the parametric distribution (14) becomes the normal one, truncated on the non-negative domain of the modelled values v_k

$$G\left(v_k \middle| v, \rho\right) = \chi(v_k \geq 0) \frac{\mathcal{N}_{v_k}(\mu, r)}{\int_0^\infty \mathcal{N}_{v_k}(\mu, r) \mathrm{d}v_k}, \ \mathcal{N}_v(\mu, r) \equiv \frac{\exp\left[-\frac{(v-\mu)^2}{2r}\right]}{\sqrt{2\pi r}},$$

where the parameters $\mu \equiv \mu(v, \rho)$, $r \equiv r(v, \rho)$ are related to the given $v = f(\mathfrak{b})$ and variance $\rho(\mathfrak{b})$ in the highly non-linear way. It holds

$$v = \mu + \sqrt{r} \frac{\mathcal{N}_{v=0}(\mu, r)}{\int_0^\infty \mathcal{N}_v(\mu, r) \mathrm{d}v}.$$

The relation of the variance ρ to μ, r (20) is even more complex (Greene, 2003). Consequently, the evaluation of the posterior expectation of v given by v_k, $k \in k^*$ is to be done numerically. The evaluation complexity motivates to search for an alternative model $G(\cdot|\cdot)$.

3.6 Log-normal model

Here, we again require unbiasedness and allow variance ρ of deviations $(v_k - v)$, $k \in k^*$ to be tuned independently of *v*. The non-negativity of processed data is, however, respected by relating $\ln(v_k)$ to $\ln(v)$. This can be done under:

Assumption 5 Partial compatibility of pdfs. The pdf values $v_k = f_k(\mathfrak{b})$, $k \in k^*$ and $v = f(\mathfrak{b})$ are strictly positive on $\mathfrak{b}^*$.

Under Assumption 5 and modelling Assumption 4 with

$$\mathsf{E}\left[\ln\left(v_k\right)\middle|v,\rho\right]=\ln\left(\mu\right),\ \mathsf{E}\left[\ln^2\left(v_k\right)\middle|v,\rho\right]=\rho+\mu^2$$

we get $G\left(v_k\middle|v,\rho\right)$ (14) in the log-normal form. The assumption on unbiasedness, $\mathsf{E}\left[v_k\middle|v,\rho\right]=v$ and the formula for the mean of the log-normal distribution lead to the definition

$$\ln\left(\mu\right)\equiv\ln\left(v\right)-0.5\rho.$$

Let us select a completely flat prior pdf on $\mu>0$. To get a finite conditional expectation $\hat{v}\equiv\hat{f}\left(\mathfrak{b}\right)=\mathsf{E}\left[v=f\left(\mathfrak{b}\right)\middle|v_k=f_k\left(\mathfrak{b}\right),k\in k^*\right]$ (10), the prior pdf on $\rho>0$ has to fall to the infinity faster than $\exp\left(-\frac{2\rho}{\mathring{k}}\right)$. This leads to the prior pdf

$$G\left(v,\rho\right)\propto\exp\left(-\frac{\left(2+\varepsilon\right)\rho}{\mathring{k}}\right),\ \varepsilon>0,\ \varepsilon\text{ small.}\tag{21}$$

It gives the posterior pdf $G\left(v,\rho\middle|v_k,k\in k^*\right)$ of the form

$$G\left(v,\rho\middle|v_k,k\in k^*\right)\propto\rho^{-\frac{\mathring{k}}{2}}\exp\left(-\frac{\left(2+\varepsilon\right)\rho}{\mathring{k}}\right)\exp\left\{-\frac{\mathring{k}}{2\rho}\left[\left(\ln\left(\frac{v}{\overline{v}}\right)-0.5\rho\right)^2+\hat{\rho}\right]\right\}\tag{22}$$

$$\begin{aligned}&\overline{v}\equiv\left[\prod_{k\in k^*}v_k\right]^{\frac{1}{\mathring{k}}}\left(\text{geometric mean}\right),\\&\hat{\rho}\equiv\frac{1}{\mathring{k}}\sum_{k\in k^*}\ln^2\left(v_k\right)-\overline{v}^2\left(\text{the normalised least-squares remainder}\right).\end{aligned}\tag{23}$$

To find normalisation of the posterior pdf (22) and to compute approximately its first moment, an evaluation of the integrals $E_i,\ i=0,1,$

$$E_i\equiv\int_0^\infty v^i\exp\left\{-\frac{\mathring{k}}{2\rho}\left[\left(\ln\left(\frac{v}{\overline{v}}\right)-0.5\rho\right)^2\right]\right\}\mathrm{d}v\tag{24}$$

should be performed. By using the following substitution

$$\left\{x\equiv\ln\left(\frac{v}{\overline{v}}\right)-0.5\rho,\ v=\overline{v}\exp\left(x+0.5\rho\right),\ \mathrm{d}v=v\mathrm{d}x\right\},$$

the integrals (24) read

$$E_i = \overline{v}^{i+1} \exp\left(0.5(i+1)\rho\right) \int_{-\infty}^{\infty} \exp\left\{(i+1)x - \frac{\mathring{k} x^2}{2\rho}\right\} \mathrm{d}x = \left(2\pi / \mathring{k}\rho\right)^{0.5} \overline{v}^{i+1} \exp\left[\frac{(i+1)^2 \rho}{2\mathring{k}}\right].$$

This leads to the factorised form of the posterior pdf in which the last factor is proportional to the normalised posterior pdf of v conditioned on $v_k,\ k \in k^*$ and on ρ

$$G\left(v, \rho \middle| v_k, k \in k^*\right) \propto \rho^{-\frac{\mathring{k}-1}{2}} \exp\left\{-\frac{\mathring{k}\hat{\rho}}{2\rho} - \frac{(3+\varepsilon)\rho}{2\mathring{k}}\right\} \frac{\exp\left\{-\frac{\mathring{k}}{2\rho}\left(\ln\left(\frac{v}{\overline{v}}\right) - 0.5\rho\right)^2\right\}}{\rho^{0.5}\overline{v} \exp\left(\frac{\rho}{2\mathring{k}}\right)}.$$

It gives the posterior expectation of v conditioned on $v_k,\ k \in k^*,$ and on ρ

$$\mathsf{E}\left[v \middle| v_k, k \in k^*, \rho\right] = E_1 / E_0 = \overline{v} \exp\left(\frac{3}{2\mathring{k}}\rho\right).$$

The marginal posterior pdf of ρ reads

$$G\left(\rho \middle| v_k, k \in k^*\right) \quad \propto \quad \rho^{-\frac{\mathring{k}-1}{2}} \exp\left\{-\frac{\mathring{k}\hat{\rho}}{2\rho} - \frac{(3+\varepsilon)\rho}{2\mathring{k}}\right\}.$$

This is a proper pdf and provides the finite expectation of $\exp\left(\frac{3\rho}{\mathring{k}}\right)$ needed for the evaluation of

$$\hat{v} \equiv \hat{f}(\mathfrak{b}) \equiv \mathsf{E}\left[v \middle| v_k, k \in k^*\right] = \mathsf{E}\left[\mathsf{E}\left[v \middle| v_k, k \in k^*, \rho\right] \middle| v_k, k \in k^*\right]. \quad (25)$$

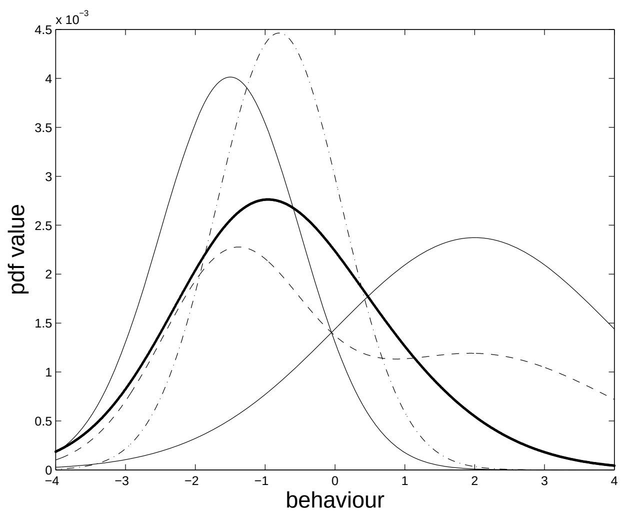
The evaluation of the 'outer' expectation has no closed form. We approximate it by inserting maximum a posteriori probability estimate $\hat{\rho}$ of ρ into the 'inner' expectation (25). For $\varepsilon \to 0,$ (21), we get

$$\hat{v} \equiv \hat{f}(\mathfrak{b}) \quad \equiv \quad \mathsf{E}\left[v \middle| v_k, k \in k^*\right] \approx \overline{v} \exp\left[\frac{\mathring{k}}{\mathring{k}-1} \frac{2\hat{\rho}}{1 + \sqrt{1 + \frac{12\hat{\rho}}{\left(\mathring{k}-1\right)^2}}}\right]. \quad (26)$$

Note that performed numerical experiments indicate that the approximation quality is high enough.

The formula (26) increases the geometric mean (23) whenever the normalised least squares remainder $\hat{\rho}$ (23) is large. An illustration of this advantageous property can be seen on Figure 1 presenting the merging results according to the formula (26). The pdfs provided by respective participants concern scalar behaviour $\mathfrak{b}$ and are normal, $f_1(\mathfrak{b}) = \mathcal{N}_{\mathfrak{b}}(-1.5,1)$ and $f_2(\mathfrak{b}) = \mathcal{N}_{\mathfrak{b}}(2,4)$. The proposed merging is a compromise between the often used geometric merging and the arithmetic merging.

Figure 1 Comparison of various merging methods



Notes: The thin solid line denotes processed pdfs $f_1(\mathfrak{b})$ (left) and $f_2(\mathfrak{b})$ (right). The dashed line denotes arithmetic merging (18). The dashed-dot line indicates geometric merging (23). The thick solid line marks the proposed merging.

4 Processing fragmental information pieces

The simplicity of the merging described in the previous section stems from the fact that the inherent complexity of the projecting operators (8) is avoided. Thus there is a need to solve general case, in which respective participants provide some marginal and conditional pdfs, by extending the processed pdfs to the joint pdfs acting on the space of all behaviours $\mathfrak{b}^*$ and then combining these extensions according to (26).

This idea is elaborated here. As the second part is already solved, the further text provides a solution on how to construct an extension of the processed pdfs.

The kth participant deals with the behaviour, cf. (7),

$$\mathfrak{b}_k \equiv (\mathfrak{f}_k, \mathfrak{a}_k, \mathfrak{p}_k) \equiv (\text{modelled, conditioning}) \text{ variables considered by } k\text{th participant,}$$

where at least $\mathfrak{f}_k$ is non-void.

Let us construct an extension ${}^{e}f_k(\mathfrak{b})$ of the given pdf $f_k(\mathfrak{f}_k|\mathfrak{a}_k,\mathfrak{p}_k)$ to the global space $\mathfrak{b}^*$ of behaviours $\mathfrak{b}$. This extension should be close to the global pdf $f(\mathfrak{b})$. The adopted Bayesian view point, Assumption 3 and the result presented in Bernardo (1979), imply that the extension should minimise the expected Kerridge inaccuracy $-\int \hat{f}(\mathfrak{b})\ln\left({}^{e}f_k(\mathfrak{b})\right)\mathrm{d}\mathfrak{b}$. At the same time, the constructed extension ${}^{e}f_k(\mathfrak{b})$ should preserve the pdf $f_k(\mathfrak{f}_k|\mathfrak{a}_k,\mathfrak{p}_k)$ expressing the information offered by the participant.

The Proposition 2 and its proof rely on decompositions (1), (7) and factorisation (2).

Proposition 2: (Optimal extension): The pdf

$$ {}^{e}f_k(\mathfrak{b}_k) = \hat{f}(\mathfrak{u}_k|\mathfrak{f}_k,\mathfrak{a}_k,\mathfrak{p}_k)\, f_k(\mathfrak{f}_k|\mathfrak{a}_k,\mathfrak{p}_k)\, \hat{f}(\mathfrak{a}_k,\mathfrak{p}_k) \tag{27} $$

is the unique minimiser of $-\int \hat{f}(\mathfrak{b})\ln\left({}^{e}f_k(\mathfrak{b})\right)\mathrm{d}\mathfrak{b}$ with the a priori given factor $f_k(\mathfrak{f}_k|\mathfrak{a}_k,\mathfrak{p}_k)$. In the formula (27) $\hat{f}(\cdot|\cdot)$ are appropriate factors of $\hat{f}(\mathfrak{b})$, $\mathfrak{u}_k$ is the part of behaviour $\mathfrak{b}$ unconsidered by the kth participant, cf. (7).

Proof: Fubini theorem, the chain rule for pdfs and the following factorisation of the given $\hat{f}(\mathfrak{b})$ and the constructed ${}^{e}f_k(\mathfrak{b}_k)$

$$ \begin{aligned} \hat{f}(\mathfrak{b}) &= \hat{f}(\mathfrak{u}_k|\mathfrak{f}_k,\mathfrak{a}_k,\mathfrak{p}_k)\,\hat{f}(\mathfrak{f}_k|\mathfrak{a}_k,\mathfrak{p}_k)\,\hat{f}(\mathfrak{a}_k,\mathfrak{p}_k), \\ {}^{e}f(\mathfrak{b}) &= {}^{e}f(\mathfrak{u}_k|\mathfrak{f}_k,\mathfrak{a}_k,\mathfrak{p}_k)\,{}^{e}f(\mathfrak{f}_k|\mathfrak{a}_k,\mathfrak{p}_k)\,{}^{e}f(\mathfrak{a}_k,\mathfrak{p}_k) \end{aligned} $$

imply

$$ \begin{aligned} -\int_{\mathfrak{b}^*} \hat{f}(\mathfrak{b})\ln\left({}^{e}f_k(\mathfrak{b})\right)\mathrm{d}\mathfrak{b} = &-\int_{(\mathfrak{f}_k,\mathfrak{a}_k,\mathfrak{p}_k)^*} \hat{f}(\mathfrak{f}_k,\mathfrak{a}_k,\mathfrak{p}_k) \\ &\left[\int_{\mathfrak{u}_k^*} \hat{f}(\mathfrak{u}_k|\mathfrak{f}_k,\mathfrak{a}_k,\mathfrak{p}_k)\ln\left({}^{e}f_k(\mathfrak{u}_k|\mathfrak{f}_k,\mathfrak{a}_k,\mathfrak{p}_k)\right)\mathrm{d}\mathfrak{u}_k\right]\mathrm{d}(\mathfrak{f}_k,\mathfrak{a}_k,\mathfrak{p}_k) \\ &-\int_{(\mathfrak{u}_k,\mathfrak{a}_k,\mathfrak{p}_k)^*} \hat{f}(\mathfrak{u}_k|\mathfrak{f}_k,\mathfrak{a}_k,\mathfrak{p}_k)\,\hat{f}(\mathfrak{a}_k,\mathfrak{p}_k) \\ &\left[\int_{\mathfrak{f}_k^*} \hat{f}(\mathfrak{f}_k|\mathfrak{a}_k,\mathfrak{p}_k)\ln\left({}^{e}f_k(\mathfrak{f}_k|\mathfrak{a}_k,\mathfrak{p}_k)\right)\mathrm{d}\mathfrak{f}_k\right]\mathrm{d}(\mathfrak{u}_k,\mathfrak{a}_k,\mathfrak{p}_k) \\ &-\int_{(\mathfrak{u}_k,\mathfrak{f}_k)^*} \hat{f}(\mathfrak{u}_k|\mathfrak{f}_k,\mathfrak{a}_k,\mathfrak{p}_k)\,\hat{f}(\mathfrak{f}_k|\mathfrak{a}_k,\mathfrak{p}_k) \\ &\left[\int_{(\mathfrak{a}_k,\mathfrak{f}_k)^*} \hat{f}(\mathfrak{a}_k,\mathfrak{p}_k)\ln\left({}^{e}f_k(\mathfrak{a}_k,\mathfrak{p}_k)\right)\mathrm{d}(\mathfrak{a}_k,\mathfrak{p}_k)\right]\mathrm{d}(\mathfrak{u}_k,\mathfrak{f}_k). \end{aligned} $$

The factor ${}^{e}f_k(\mathfrak{f}_k|\mathfrak{a}_k,\mathfrak{p}_k) = f_k(\mathfrak{f}_k|\mathfrak{a}_k,\mathfrak{p}_k)$ is a priori given so that the second term cannot be influenced. The expressions in brackets [ ] in the first and the third term are conditional versions of Kerridge inaccuracy that are, for an arbitrary condition, uniquely minimised by ${}^{e}f_k\left(\mathfrak{u}_k \middle| \mathfrak{f}_k, \mathfrak{a}_k, \mathfrak{p}_k\right) = \hat{f}\left(\mathfrak{u}_k \middle| \mathfrak{f}_k, \mathfrak{a}_k, \mathfrak{p}_k\right)$ and ${}^{e}f_k\left(\mathfrak{a}_k, \mathfrak{p}_k\right) = \hat{f}\left(\mathfrak{a}_k, \mathfrak{p}_k\right)$, respectively. They minimise whole terms as these minima are multiplied by non-negative pdfs and integrated over conditions. □

Now, it remains to insert the obtained extensions ${}^{e}f_k(\mathfrak{b})$ into the formula (26) determining the conditional expectation $\hat{f}(\mathfrak{b}) = \mathsf{E}\left[f(\mathfrak{b}) \middle| f_k\left(\mathfrak{f}_k \middle| \mathfrak{a}_k, \mathfrak{p}_k\right), k \in k^*\right]$ of the global pdf $f(\mathfrak{b})$. This provides the *final* non-linear equation for merging $\hat{f}(\mathfrak{b})$

$$\hat{f}(\mathfrak{b}) \propto {}^{e}\overline{f}(\mathfrak{b}) \exp\left[\frac{\mathring{k}}{\mathring{k}-1} \frac{2\,{}^{e}\hat{\rho}(\mathfrak{b})}{1 + \sqrt{1 + \frac{12\,{}^{e}\hat{\rho}(\mathfrak{b})}{\left(\mathring{k}-1\right)^2}}}\right] \tag{28}$$

The left superscript ^{e} stresses that the geometric mean (23) and the normalised least-squares remainder (23) are evaluated from the pdfs ${}^{e}f_k(\mathfrak{b})$, $k \in k^*$, (27) that depend on factors derived from the constructed $\hat{f}(\mathfrak{b})$, see (27).

The explicit solvability of (28) can hardly be expected but the equation is 'naturally' prepared for solution by successive approximations. Experiments support conjecture that a solution of (28) always exists.

After finding $\hat{f}(\mathfrak{b})$, its k-specific projections $\hat{f}_k\left(\mathfrak{f}_k \middle| \mathfrak{a}_k, \mathfrak{p}_k\right)$ (8) provide the constructed merging. Their uniqueness is conjectured while existence of multiple solutions of (28) cannot be excluded.

5 Illustrative examples

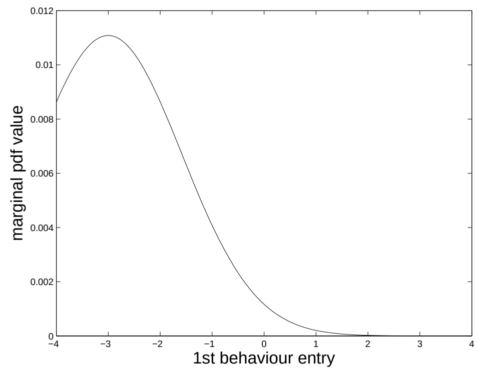
5.1 Merging of fragmental information

The example demonstrates the proposed merging (28) in a simple case of a two-dimensional behaviour $\mathfrak{b} = (\mathfrak{b}_1, \mathfrak{b}_2)$. The first participant provides only marginal pdf on $\mathfrak{b}_1$, $f_1(\mathfrak{b}_1) = \mathcal{N}_{\mathfrak{b}_1}(-3, 2)$, while the second participant offers the joint pdf

$$f_2(\mathfrak{b}_1, \mathfrak{b}_2) = \mathcal{N}_{(\mathfrak{b}_1, \mathfrak{b}_2)}\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1.50 & 0.45 \\ 0.45 & 1.50 \end{bmatrix}\right)$$

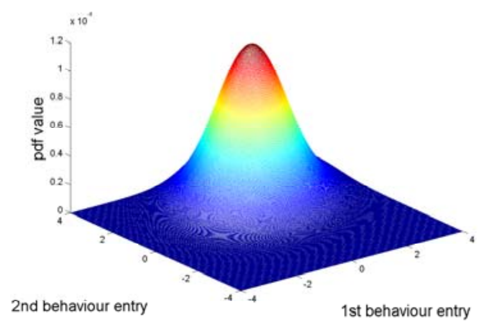
The obtained merging results are presented in Figure 2 to Figure 6. The results were obtained after eight iterations when the norm between two successive iterations dropped below the threshold $1e-5$.

Figure 2 Merging of fragmental information



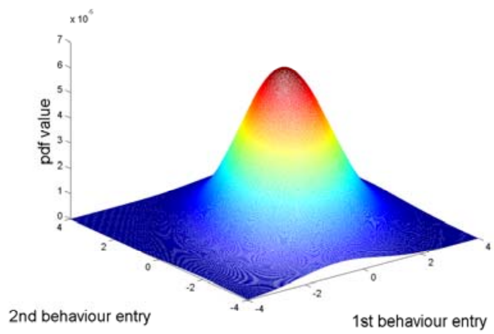
Note: Pdf $f_1(\mathfrak{b}_1)$ offered by the first participant.

Figure 3 Merging of fragmental information (see online version for colours)



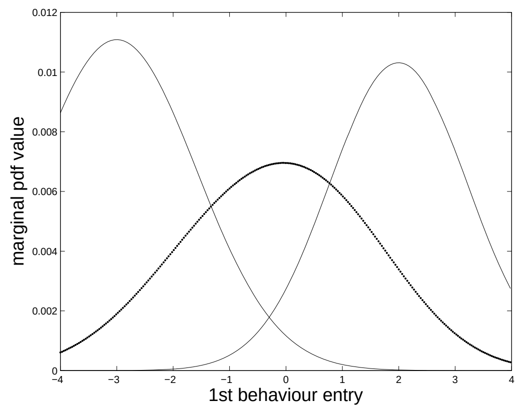
Note: Pdf $f_2(\mathfrak{b}_1, \mathfrak{b}_2)$ offered by the second participant.

Figure 4 Merging of fragmental information (see online version for colours)



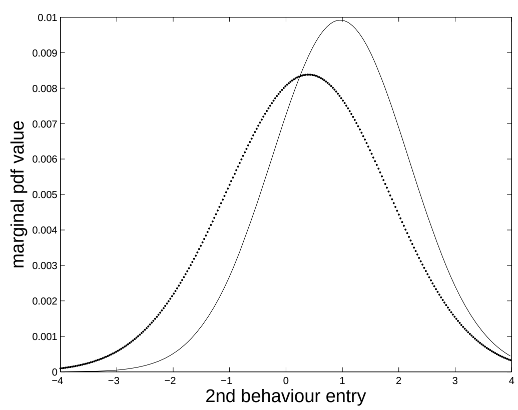
Note: The merging result $\hat{f}(\mathfrak{b}_1, \mathfrak{b}_2)$.

Figure 5 Merging of fragmental information



Notes: The pdfs processed are depicted by thin line: left hill is $f_1(\mathfrak{b}_1)$, right hill is $f_2(\mathfrak{b}_1) = \int f_2(\mathfrak{b}_1, \mathfrak{b}_2)\, d\mathfrak{b}_2$. The thick line marks marginal pdf of the merging $\hat{f}(\mathfrak{b}_1) = \int \hat{f}(\mathfrak{b}_1, \mathfrak{b}_2)\, d\mathfrak{b}_2$.

Figure 6 Merging of fragmental information



Notes: Thin line denotes pdf processed, $f_2(\mathfrak{b}_2) = \int f_2(\mathfrak{b}_1, \mathfrak{b}_2)\, d\mathfrak{b}_1$. Pdf $f_1(\mathfrak{b}_2)$ is not offered. The dotted line marks marginal pdf of the merging $\hat{f}(\mathfrak{b}_2) = \int \hat{f}(\mathfrak{b}_1, \mathfrak{b}_2)\, d\mathfrak{b}_1$.

Note that the results are strongly influenced by the correlation between $\mathfrak{b}$-entries considered by the second participant: for higher correlations, the shifts of the marginal pdf describing $\mathfrak{b}_2$ are more pronounced.

5.2 Elicitation of prior knowledge

The example illustrates the use of the developed methodology for knowledge elicitation.

Let us assume that the participant $k=1$ has a prior pdf $f_1(\mathfrak{b})$ on behaviour $\mathfrak{b}$ and the participant $k=2$ offers partial information in the form of a (generalised) moment

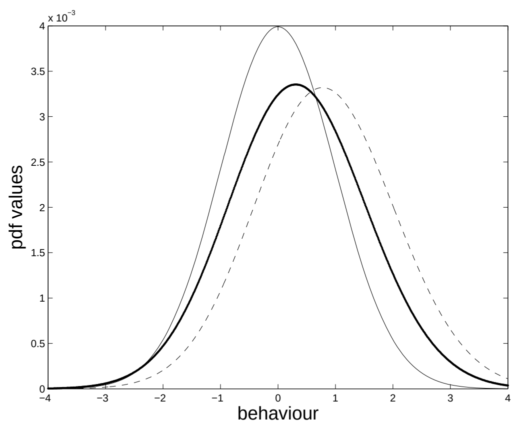
$$\mu=\int \phi(\mathfrak{b}) f_2(\mathfrak{b})\,\mathrm{d}\mathfrak{b}, \tag{29}$$

where the finite-dimensional vector μ and the vector function $\phi(\mathfrak{b})$ are the offered probabilistic elements. Following the proposed methodology, the information (29) is extended to the pdf ${}^{e}f_2(\mathfrak{b})$. The extension ${}^{e}f_2(\mathfrak{b})$ is then merged with $f_1(\mathfrak{b})$ according to (28). The extension ${}^{e}f_2(\mathfrak{b})$ is again taken as the pdf nearest to the constructed result of merging $\hat{f}(\mathfrak{b})$ in Kerridge-inaccuracy sense under the constraint (29). The extension has the form

$$ {}^{e}f_2(\mathfrak{b}) \propto \hat{f}(\mathfrak{b}) \exp\left[-\zeta'\phi(\mathfrak{b})\right] \tag{30}$$

with the vector ζ chosen so that $f_2(\mathfrak{b})={}^{e}f_2(\mathfrak{b})$ in (30) meets (29).

Figure 7 Elicitation of knowledge about mean value

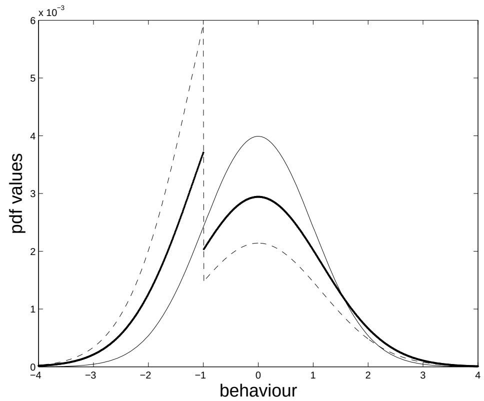


Notes: Original pdf $f_1(\mathfrak{b})=\mathcal{N}_{\mathfrak{b}}(0,1)$ (thin line) modified by information about mean value $\mathsf{E}_2[\mathfrak{b}]=\int \mathfrak{b}\, f_2(\mathfrak{b})\mathrm{d}\mathfrak{b}=0.8$. Dashed line denotes extension of the last to the pdf ${}^{e}f_2(\mathfrak{b})$. Thick line marks the merging result $\hat{f}(\mathfrak{b})$.

Figure 7 and Figure 8 present the results of such a merging with scalar $\mathfrak{b}$, μ and $\phi(\mathfrak{b})$. The pdf $f_1(\mathfrak{b})=\mathcal{N}_{\mathfrak{b}}(0,1)$ was modified by the information elicited from the second participant. Both results were obtained after 11 iterations of successive approximations.

Figure 7 describes the case with $\mu = 0.8$ and $\phi(\mathfrak{b}) = \mathfrak{b}$, i.e., for the offered information on expected value of $\mathfrak{b}$. Figure 8 shows the result for $\mu = 0.5$ and $\phi(\mathfrak{b}) = \chi(\mathfrak{b} > -1)$, i.e., for the processed information on the median of $\mathfrak{b}$.

Figure 8 Elicitation of knowledge about median



Notes: Original pdf $f_1(\mathfrak{b}) = \mathcal{N}_{\mathfrak{b}}(0,1)$ (thin line) modified by information about median $\int_{-1}^{\infty} f_2(\mathfrak{b})\mathrm{d}\mathfrak{b} = 0.5$. Dashed line denotes extension of the last to the pdf ${}^{e}f_2(\mathfrak{b})$. Thick line marks the merging result $\hat{f}(\mathfrak{b})$.

The examples illustrate plausible properties of the proposed methodology. To get practically useful result, the approximation task (9) is to be solved within an appropriate class $f(\mathfrak{b}|\mathcal{V})$, $\mathcal{V} \in \mathcal{V}^*$, containing typically more regular functions than the obtained $\hat{f}(\mathfrak{b})$.

6 Concluding remarks

The paper addresses a merging problem arose in multiple participant DM. The merging is treated as a Bayesian estimation problem. The presented solution is more complete compared to the previous works and leads to feasible algorithms covering infinite-dimensional cases. The case of different observation spaces was not treated in this way before.

The important methodological shift consists in a clear separation of non-linear deterministic mapping of joint distributions to the supplied conditional or marginal pdfs from handling the influence of 'personal noise'. The proposed modelling and processing order: combine noisy joint distributions and then map the results on observed lower-dimensional pdfs are computationally simpler and new. The application of this approach to fragmental information pieces, given by noisy projections of the underlying global pdf is enabled by extending them to a full space in a unique, well-justified, way.

The paper proposes a unified methodology. It does not mean that there is no ambiguity in solving a specific problem. However, the modelling of the relations of the unknown merging model (parameter) to respective processed pdfs (data) is the only optional step. Thus, modelling can only be blamed if the results are not satisfactory enough.

The algorithm based on log-normal model provides a practically applicable algorithm, which can be viewed as a compromise between arithmetic and geometric merging.

The following open problems remain to be solved.

- Uniqueness of the participant-specific projections of (28) is conjectured but not proved.
- General applicability of successive approximations for solving (28) is not verified.
- The important case of offered sample pdfs is not covered due to the required non-singularity of processed pdfs (the only restriction on compatibility of the processed information pieces). Ideally, the merging should reduce to the Bayes rule when the sample pdf is processed.
- Conversion of the offered parametric distributions and generalised moments to non-parametric pdfs has to be solved systematically. It will represent a sort of communication protocol studied and used in connection with closely related multi-agent systems (MAS, Vlassis, 2003).
- Actions controlling communications have to be explicitly considered and their optimisation designed. These extensions have to respect the considered flat structure and will create counterparts of negotiation, bargaining and conflict resolution strategies studied in MAS.
- Study and exploitation of the extensive overlap with a range of existing techniques originating in knowledge elicitation, probabilistic expert systems, MAS, etc., should be performed.

Acknowledgements

This research was supported by a common project IMATI CNR Milano and ÚTIA AV ČR, GAČR 102/08/0567 as well as by MŠMT ČR 2C06001.

References

Andrýsek, J., Bodini, A., Kárný, M., Kracík, J. and Ruggeri, F. (2007) 'On combining partial and incompatible information in e-negotiation and e-arbitration', *Group Decision and Negotiation*, Vol. 10, pp.1–10.

Bell, D.E., Raiffa, H. and Tversky, A. (1988) *Decision Making: Descriptive, Normative and Prescriptive Interactions*, ISBN 0521368510, 9780521368513, Cambridge University Press.

Berger, J. (1985) *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York.

Bernardo, J.M. (1979) 'Expected information as expected utility', *The Annals of Statistics*, Vol. 7, No. 3, pp.686–690.

Cover, T. and Thomas, J. (1991) *Elements of Information Theory*, 2nd ed., Wiley, Bratislava.

Cowell, R., Dawid, A., Lauritzen, S.L. and Spiegelhalter, D. (2003) *Probabilistic Networks and Expert Systems*, 2nd ed., Springer.

DeGroot, M. (1970) *Optimal Statistical Decisions*, McGraw-Hill, New York.

Dyer, J.S., Fishburn, P.C., Steuer, R.E., Wallenius, J. and Zionts, S. (1992) 'Multiple criteria decision making, multi attribute utility theory: the next ten years', *Management Sciences*, Vol. 38, No. 5, pp.645–654.

Greene, W. (2003) *Econometric Analysis*, 5th ed., Prentice Hall.

Harsanyi, J. (2004) 'Games with incomplete information played by Bayesian players, I–III', *Management Science*, Supplement, Vol. 50, No. 12.

Kárný, M. (2008) 'Bayesian paradigm and fully probabilistic design', Preprints of the *17th World Congress of the International Federation of Automatic Control*, IFAC, Seoul.

Kárný, M. and Guy, T.V. (2006) 'Fully probabilistic control design', *Systems & Control Letters*, Vol. 55, No. 4, pp.259–265.

Kárný, M., Kracík, J. and Guy, T. (2007) 'Cooperative decision making without facilitator', in Fradkov, A.L. and Andrievsky B.R. (Eds.): *IFAC Workshop 'Adaptation and Learning in Control and Signal Processing'* /9./. IFAC.

Kerridge, D. (1961) 'Inaccuracy and inference', *Journal of Royal Statistical Society B*, Vol. 23, pp.284–294.

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D.J., Oakley, J. and Rakow, T. (2006) *Uncertain Judgement: Eliciting Experts' Probabilities*, John Wiley & Sons.

Oakley, J. and O'Hagan, A. (2005) *Uncertainty in Prior Elicitation: A Non-Parametric Approach*, Technical report, Department of Probability and Statistics, University of Shefield, Shefield. Revised version of research report No. 521/02.

Quinn, A., Ettler, P., Jirsa, L., Nagy, I. and Nedoma, P. (2003) 'Probabilistic advisory systems for data-intensive applications', *International Journal of Adaptive Control and Signal Processing*, Vol. 17, No. 2, pp.133–148.

Stirling, W. (2004) 'Satisfying games for multiple-participant coordinated decision-making', in J. Andrýsek et al. (Eds.): *Mutliple Participant Decision Making*, pp.3–16, Advanced Knowledge International, Maggil Adelaide.

Vajda, I. (1989) *Theory of Statistical Inference and Information*, Kluwer Academic Publishers, Dordrecht.

Vlassis, N. (2003) *A Concise Introduction to Multi-Agent Systems and Distributed AI*, Informatics Institute, University of Amsterdam, 2003, available at http://www.sciece.uva.nl/vlassis/cimasdai.

Wolpert, D. and Tumer, K. (1999) 'Collective intelligence for optimization', in J. Boyan (Ed.): *Sixteenth International Joint Conference on Artificial Intelligence, Workshop ML-1, Statistical Machine Learning for Large-Scale Optimization*, Stockholm, pp.121–128.

Wolpert, D. and Tumer, K. (2001) 'Optimal payoff functions for members of collectives', *Advances in Complex Systems*, Vol. 4, Nos. 2/3, pp.265–279.

Yiming, Y., Xun, Y. and Lee, J. (2003) 'Performance and attention in multiagent object search team', *IEEE Transactions on Systems Man and Cybernetics, Part A – Systems and Humans*, Vol. 33, No. 2, pp.257–263.