A Note on Approximation of Shenoy's Expectation Operator Using Probabilistic Transforms

R. Jiroušek[a,b] , V. Kratochvíl[a,b] and J. Rauh[c]

[a]Faculty of Management, University of Economics, Jindřichův Hradec, CZ;
[b]Czech Academy of Sciences, Institute of Information Theory and Automation, Prague, CZ;
[c]Max Planck Institute for Mathematics in the Sciences, Leipzig , D

ARTICLE HISTORY
Compiled April 24, 2020

ABSTRACT
Recently, a new way of computing an expected value in the Dempster-Shafer theory of evidence was introduced by Prakash P. Shenoy. Up to now, when they needed the expected value of a utility function in D-S theory, the authors usually did it indirectly: first, they found a probability measure corresponding to the considered belief function, and then computed the classical probabilistic expectation using this probability measure. To the best of our knowledge, Shenoy's operator of expectation is the first approach that takes into account all the information included in the respective belief function. Its only drawback is its exponential computational complexity. This is why, in this paper, we compare five different approaches defining probabilistic representatives of belief function from the point of view, which of them yields the best approximations of Shenoy's expected values of utility functions.

KEYWORDS
Expectation, belief function, probabilistic transform, commonality function, utility, ambiguity, Choquet integral.

1. Introduction

Criteria for finding optimal decisions are usually based on a maximum expected utility principle. As Glenn Shafer [Shafer(1986)] wrote already in 1986:

> The controversy raised by this book *(here he meant Savage's book [Savage(1954)])* and Savage's subsequent writings are now part of the past. Many statisticians now use Savage's idea of personal probability in their practical and theoretical work. [...] To do otherwise is to violate a canon of rationality.

This reflects the fact that the maximum expected utility principle is often used not only when the knowledge from the respective field of application is embodied in a probabilistic model but also when the applied model is built within the framework of belief functions. Nevertheless, to compute the necessary value of expected utility, the respective belief function is usually transformed into an appropriate probability measure. For this, several procedures were designed - we call them *probability transforms*

CONTACT R. Jiroušek. Email: radim@utia.cas.cz
CONTACT V. Kratochvíl. Email: velorex@utia.cas.cz
CONTACT J. Rauh. Email:: jrauh@mis.mpg.de

in this paper. As advocated by Cobb and Shenoy, the only one, which is compatible with the Dempster-Shafer theory of belief functions is the plausibility transform [Cobb and Shenoy(2006)]. Other transforms are more likely compatible with the theory of belief functions interpreted as *generalized probability* [Halpern and Fagin(1992)]. This interpretation reflects the fact that a belief function specifies a convex set of probability measures, which is called a *credal set*. In this paper we consider widely used *pignistic transform* advocated by Philippe Smets [Smets(2005)], two others that are usually omitted in the context of belief function: *maximum entropy* and *Perez' barycenter* [Perez(1985)], and a new transform, which is a convex combination of belief and plausibility functions.

In our best knowledge, the first idea of how to compute an expected value for a belief function directly, i.e., avoiding its transformation into a probability measure, is due to Prakash P. Shenoy [Shenoy(2018)]. From the theoretical point of view, it is a concept deserving a further investigation. As we will see in the following paragraph, it is defined with the help of commonality functions, which means that it suffers from a great computational complexity. Though there exists a probability transform yielding exactly the same expectations as Shenoy's operator (see Section 3), the computation of the respective probability measure is time demanding, too, so that it is for practical problems intractable. Therefore, in this paper, we study a problem whether any of the probability transforms, which can be easily computed from the respective probability assignment, reasonably approximates the results yielded by Shenoy's new operator.

To achieve this goal, the rest of the paper is organized as follows. The next section is devoted to the introduction of all the necessary notions including Shenoy's expectation operator. The notation used in this paper is adopted from [Shenoy(2018), Shenoy(2019)]. In Section 3, we show that Shenoy's expectation, if applied to a utility function, can be computed in an alternative way. Nevertheless, since this alternative way is also of exponential complexity, Sections 5 and 6 are devoted to the comparison of five approximation processes based on five different probabilistic transforms. In Sections 5, we study the behavior of these approaches when applied to a specific class of belief function models describing situations in which human decision-makers evince their ambiguity aversion, and in Section 6, the comparison is made with the help of randomly generated basic assignments.

2. Notation

Suppose X is a random variable with a finite state space Ω_X. Let 2^{Ω_X} denote the set of all *non-empty*[1] subsets of Ω_X. A *basic probability assignment* (basic assignment for short) m for X is a function $m : 2^{\Omega_X} \to [0,1]$ such that

$$\sum_{\mathsf{a} \in 2^{\Omega_X}} m(\mathsf{a}) = 1. \tag{1}$$

The subsets $\mathsf{a} \in 2^{\Omega_X}$ such that $m(\mathsf{a}) > 0$ are called *focal* elements of m. If m has only one focal element (it means that $m(\mathsf{a}) = 1$ for some $\mathsf{a} \in 2^{\Omega_X}$) then m is said to be *deterministic*. Among them, a special position is held by a *vacuous* basic assignment denoted by ι_X, for which $\iota_X(\Omega_X) = 1$. This basic assignment represents a total

[1] Notice that, in correspondence with Shenoy [Shenoy(2018)], we consider only normal basic assignments (i.e., basic assignments for which Equality 1 holds true), and therefore the exclusion of the empty set from 2^{Ω_X} simplifies some of the formulas.

ignorance.

If all focal elements of m are singletons (one-element subsets) of Ω_X, then we say m is *Bayesian*. In this case, m corresponds to a probability measure. If m is a convex combination of a Bayesian basic assignment with a vacuous basic assignment, then we say m is *quasi-Bayesian*. In this case, focal elements are singletons and the whole Ω_X.

Alternatively, a basic assignment m can be equivalently specified by any of the following three other functions. First two are the well-known *belief* and *plausibility* functions Bel_m and Pl_m that are defined

$$Bel_m(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} : \mathsf{b} \subseteq \mathsf{a}} m(\mathsf{b}), \qquad Pl_m(\mathbf{a}) = \sum_{\mathbf{b} \in 2^{\Omega} : \mathbf{b} \cap \mathbf{a} \neq \emptyset} m(\mathbf{b}),$$

for all $\mathsf{a} \in 2^{\Omega_X}$. In this paper, we need also a *commonality function* for m, which is defined for all $\mathsf{a} \in 2^{\Omega_X}$

$$Q_m(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} : \mathsf{b} \supseteq \mathsf{a}} m(\mathsf{b}).$$

It is obvious directly from their definitions that for all $\mathbf{a} \in 2^{\Omega_X}$, $Bel(\mathbf{a}) \leq Pl(\mathbf{a})$. For singletons, commonality and plausibility functions coincide:

$$Q_m(\{x\}) = Pl_m(\{x\}).$$

Since we consider only normal basic assignments for which $\sum_{\mathsf{a} \in 2^{\Omega_X}} m(\mathsf{a}) = 1$, it is known ([Shafer(1976)]) that

$$\sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} Q_m(\mathsf{a}) = 1. \tag{2}$$

As said above, these types of representation are equivalent to each other in the sense that if knowledge is represented by any of these functions, one can uniquely compute the remaining three using the following inverse transformations:

$$Bel_m(\mathsf{a}) = 1 - Pl_m(\Omega_X \setminus \mathsf{a});$$

$$m(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} : \mathsf{b} \subseteq \mathsf{a}} (-1)^{|\mathsf{a} \setminus \mathsf{b}|} Bel_m(\mathsf{b});$$

$$m(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} : \mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} Q_m(\mathsf{b}).$$

A *credal set* interpretation of belief functions is based on the fact that basic assignment m specifies the following convex set of probability measures P on Ω_X ($\mathcal{P}_{\Omega_X}$ denote the set of all probability measures on Ω_X):

$$\mathcal{P}(m) = \left\{ P \in \mathcal{P}_{\Omega_X} : \sum_{x \in \mathsf{a}} P(x) \geq Bel_m(\mathsf{a}) \text{ for } \forall \mathsf{a} \in 2^{\Omega_X} \right\}.$$

From this, one can easily deduce that for all $P \in \mathcal{P}(m)$, and any $\mathsf{a} \in 2^{\Omega_X}$

$$Bel_m(\mathsf{a}) \leq P(\mathsf{a}) \leq Pl_m(\mathsf{a}).$$

If m is Bayesian, then $\mathcal{P}(m)$ contains just one probability measure.

One can see directly from the definitions that belief function Bel_m (plausibility function Pl_m) is superadditive (subadditive), which means that for disjoint $\mathsf{a}, \mathsf{b} \in 2^{\Omega_X}$, $Bel_m(\mathsf{a} \cup \mathsf{b}) \geq Bel_m(\mathsf{a}) + Bel_m(\mathsf{b})$ ($Pl_m(\mathsf{a} \cup \mathsf{b}) \leq Pl_m(\mathsf{a}) + Pl_m(\mathsf{b})$, respectively).

Similarly to a probability measure on Ω_X, basic assignment (or any other above-introduced alternative function) expresses knowledge about chances that $x \in \Omega_X$ occurs. Therefore, knowing a real-valued utility function[2] $u : \Omega_X \to \mathbb{R}$, one should be able to compute the expected value of this utility under the knowledge represented by a basic assignment m. It has been proposed to compute such an expected value with the help of the *Choquet integral* [Choquet(1953)], which is known to have some advantageous properties especially for superadditive and subadditive capacities [Gilboa and Schmeidler(1994)]. In particular for getting the respective upper and lower limits, one can use the Choquet integral of the utility function with respect to the corresponding plausibility and belief functions [Coletti, Petturiti, and Vantaggi(2019)]. To compute the Choquet integral of utility function u with respect to the belief function, consider the set $\{u(x) : x \in \Omega_X\}$ of all values of the considered utility function u and order them, so that

$$\{u(x) : x \in \Omega_X\} = \{\alpha_1, \alpha_2, \ldots, \alpha_\ell\},$$

and $\alpha_1 < \alpha_2 < \ldots < \alpha_\ell$. Then

$$\oint u \; dBel_m = \sum_{i=1}^{\ell} \alpha_i (Bel_m(\mathsf{a}_i) - Bel_m(\mathsf{a}_{i+1})),$$

where $\mathsf{a}_i = \{x \in \Omega_X : u(x) \geq \alpha_i\}$ for $i = 1, 2, \ldots, \ell$ ($\mathsf{a}_{\ell+1} = \emptyset$). For its application to decision making see, e.g., [Smets(1981)] and [Coletti, Petturiti, and Vantaggi(2015)]. Its main disadvantage is that the Choquet integral is not linear in its integrand.

In this paper, we accept as a proper expected value of a real-valued function a result obtained by the application of a new expectation operator proposed by Shenoy in [Shenoy(2018)], and more deeply studied in [Shenoy(2019)].

For a real-valued function $g : 2^{\Omega_X} \longrightarrow \mathbb{R}$, Shenoy's expected value of g with respect to m is defined by the formula

$$E_m(g) = \sum_{\mathsf{a} \in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1} g(\mathsf{a}) Q_m(\mathsf{a}). \tag{3}$$

To be able to apply Formula (3) for computation of expected utility for $u : \Omega_X \to \mathbb{R}$, we should extend this utility function defined on Ω_X to a function defined on 2^{Ω_X}. We denote this extension $\hat{u}$. Let us stress that to keep the validity of properties of the expected values proven by Shenoy in [Shenoy(2019)] we have to follow the idea he uses when introducing his function v_m, which is nothing else than an "extension" of his real-valued state space Ω_X to all subsets of Ω_X (assigning a real value to each

[2] $\mathbb{R}$ denotes the set of real numbers.

subset of Ω_X). Therefore, we define for all $\mathsf{a} \in 2^{\Omega_X}$

$$\hat{u}(\mathsf{a}) = \frac{\sum\limits_{x\in\mathsf{a}} u(x)Q_m(\{x\})}{\sum\limits_{x\in\mathsf{a}} Q_m(\{x\})} \tag{4}$$

(in case that $\sum_{x\in\mathsf{a}} Q_m(\{x\}) = 0$ the value $\hat{u}(\mathsf{a})$ does not influence the resulting expected value of u and therefore we can choose any reasonable value; for example $\hat{u}(\mathsf{a}) = (\min_{x\in\mathsf{a}}\{u(x)\} + \max_{x\in\mathsf{a}}\{u(x)\})/2$). Thus, in this paper, by Shenoy's expected utility u with respect to m we understand the value

$$E_m(u) = \sum_{\mathsf{a}\in 2^{\Omega_X}} (-1)^{|\mathsf{a}|+1}\hat{u}(\mathsf{a})Q_m(\mathsf{a}), \tag{5}$$

where, let us stress once more, $\hat{u}$ is an extension of u according to Formula (4).

There is a long list of properties of Shenoy's expected utility proven in Section 4.3 of [Shenoy(2019)]. In this paper, we need just the following two[3]:

(S1) (*Linearity of expected value*) Suppose that for all $x \in \Omega_X$

$$\hat{u}(x) = \alpha u_1(x) + \beta u_2(x) + \delta,$$

where α, β and δ are real constants. Then, for any basic assignment m

$$E_m(\hat{u}) = \alpha E_m(u_1) + \beta E_m(u_2) + \delta.$$

(S2) (*Bounds on expected value*)

$$\min\{u(x) : x \in \Omega_X\} \leq E_m(u) \leq \max\{u(x) : x \in \Omega_X\}.$$

3. Probabilistic representation of Shenoy's expectation

When computing Shenoy's expected utility $E_m(u)$ according to Formula (3), one can easily get into troubles. Let X be a vector of six binary variables, i.e., Ω_X is a state space of six-dimensional boolean vectors. Thus, $|\Omega_X| = 2^6 = 64$, which means that the extension $\hat{u}$ of utility function u must be computed for $|2^{\Omega_X}| = 2^{64} - 1$ nonempty subsets of Ω_X, which is hopelessly intractable. Therefore, an opposite approach comes into consideration. Instead of extending the considered utility function from Ω_X to 2^{Ω_X} and computing Shenoy's expected utility $E_m(u)$ using Eq. (5), one can find a probability measure Sh_P_m, for which

$$E_m(u) = \sum_{x\in\Omega_X} u(x) Sh\_P_m(x).$$

[3]Though these assertions are not exactly in this wording among those proven in Section 4.3 of [Shenoy(2019)], one can easily get them from statements 3 (*Expected value of a function of* X) and 7 (*Bounds on expected value*), respectively, using the following simple modification: Shenoy considers real-valued state space, i.e., $\Omega_X \subset \mathbb{R}$. Therefore, to show the validity of the above-presented statements (1) and (2) it is enough to consider $\hat{\Omega}_X = \{g(x) : x \in \Omega\}$ with $\hat{m}(y) = \sum_{x\in\Omega_X : g(x)=y} m(x)$. Recall that this transformation is correct because we extend the considered utility function for all subsets of Ω_X in the same way as Shenoy defines his function v_m.

The existence of such a measure is guaranteed by the following assertion.

Proposition 3.1. *Consider a basic assignment m for X with a finite state space Ω_X. Define a probability measure Sh_P_m for all $x \in \Omega_X$ by $Sh\_P_m(x) = E_m(w_x)$, where the real functions $w_x : \Omega_X \to \mathbb{R}$ are defined*

$$w_x(y) = \begin{cases} 1, & \textit{if } y = x, \\ 0, & \textit{otherwise.} \end{cases} \tag{6}$$

Then for any utility function $u : \Omega_X \to \mathbb{R}$ the following equality holds

$$E_m(u) = \sum_{x \in \Omega_X} u(x) Sh\_P_m(x).$$

Proof. Notice that the considered utility function can be expressed as a weighted sum of functions w_x: $u(x) = \sum_{x \in \Omega_X} u(x) w_x$. Therefore, due to Statement (S1) (*Linearity of expected value*) from the preceding section

$$E_m(u) = E_m\left(\sum_{x \in \Omega_X} u(x) w_x\right) = \sum_{x \in \Omega_X} u(x) E_m(w_x) = \sum_{x \in \Omega_X} u(x) Sh\_P_m(x).$$

Thus, to prove the proposition, it remains to prove that $Sh\_P_m(x)$ is a probability measure. $Sh\_P_m(x)$ is nonnegative because it is an expected value of nonnegative "utility" function w_x – see Statement (S2) (*Bounds on expected value*). Taking the constant utility function $u_{\mathbf{1}}(x) = 1$, one gets

$$E_m(u_{\mathbf{1}}) = E_m\left(\sum_{x \in \Omega_X} 1 w_x\right) = \sum_{x \in \Omega_X} E_m(w_x) = \sum_{x \in \Omega_X} Sh\_P_m(x),$$

and, simultaneously, using Statement (S2) (*Bounds on expected value*), $E_m(u_{\mathbf{1}}) = 1$, which finishes the proof. □

The presented Proposition 3.1 theoretically gives instructions on how to find a probability measure with the help of which one can compute expected values for utility functions even for state spaces of high dimensions. However, the process of computation of the probability measure is computationally too time and space demanding even for small dimensions. Therefore, the rest of the paper is devoted to the comparison of several probability transforms that are for this purpose used. We will concentrate on their ability to approximate Shenoy's operator of expectation.

4. Probability transforms

In this paper, we study the properties of the following five mappings that assign a probability measure to each basic assignment. For other probability transforms see, e.g., [Cuzzolin(2012)]. Perhaps, the most famous is a *pignistic transform*, defined for all $x \in \Omega_X$ by the formula

$$Bet\_P_m(x) = \sum_{\mathbf{a} \in 2^\Omega : x \in \mathbf{a}} \frac{m(\mathbf{a})}{|\mathbf{a}|}.$$

Another transform is so called plausibility transform, which is a normalized plausibility function on singletons. Formally it is defined for all $x \in \Omega_X$

$$Pl\_P_m(x) = \frac{Pl_m(\{x\})}{\sum_{y \in \Omega_X} Pl_m(\{y\})}.$$

The main reason why we take this transform into consideration is that, as showed in [Cobb and Shenoy(2006)], it is the only transform compatible with the Dempster's rule of combination.

The other three probability transforms select a specific representative from the corresponding credal set. One is the *Maximum entropy* element of $\mathcal{P}(m)$, i.e.,

$$Me\_P_m(x) = \arg \max_{P \in \mathcal{P}(m)} H(P),$$

where $H(P)$ is the Shannon entropy of probability measure P

$$H(P) = -\sum_{x \in \Omega_X} P(x) \log_2 P(x).$$

The second is the Perez' barycenter [Perez(1985)] that has undeservedly fallen into oblivion:

$$Bac\_P_m(x) = \arg \min_{P \in \mathcal{P}(m)} \max_{Q \in \mathcal{P}(m)} Div(Q; P),$$

where $Div(Q; P)$ denote the well-known relative entropy (also called Kullback-Leibler divergence)

$$Div(Q; P) = \begin{cases} +\infty, & \text{if } \exists\ x \in \Omega_X : P(x) > 0 = Q(x); \\ \sum_{x \in \Omega_X} P(x) \log\left(\frac{P(x)}{Q(x)}\right), & \text{otherwise}^3. \end{cases}$$

The third one is the element of the credal set that can be expressed as a convex combination of belief and plausibility functions

$$Cs\_P_m(x) = \delta Bel_m(\{x\}) + (1 - \delta) Pl_m(\{x\}),$$

where

$$\delta = \frac{\left(\sum_{x \in \Omega_X} Pl_m(\{x\})\right) - 1}{\left(\sum_{x \in \Omega_X} Pl_m(\{x\})\right) - \left(\sum_{x \in \Omega_X} Bel_m(\{x\})\right)}.$$

[3] We always take $0 \log\left(\frac{0}{0}\right) = 0$.

In a way, it is surprising that we have not found this probability transform in the literature, though, in our opinion, it suggests itself by its simplicity. Its interesting property is expressed in the following assertion.

Proposition 4.1. *For any quasi-Bayesian basic assignment m for X (with a finite state space Ω_X)*

$$Cs\_P_m(x) = Bet\_P_m(x).$$

Proof. For quasi-Bayesian m, obviously, $Bel_m(\{x\}) = m(\{x\})$, and $Pl_m(\{x\}) = m(\{x\}) + m(\Omega_X) = Bel_m(\{x\}) + m(\Omega_X)$. Therefore,

$$\begin{aligned} Cs\_P_m(x) &= \delta Bel_m(\{x\}) + (1-\delta)Pl_m(\{x\}) \\ &= \delta Bel_m(\{x\}) + (1-\delta)(Bel_m(\{x\}) + m(\Omega_X)) \\ &= Bel_m(\{x\}) + (1-\delta)m(\Omega_X). \end{aligned}$$

It means that the mass $m(\Omega_X)$ is uniformly divided among the all elements of Ω_X. This is also the property of the pignistic transform, which, for quasi-Bayesian m, can be rewritten into the form

$$Bet\_P_m(x) = Bel_m(\{x\}) + \frac{m(\Omega_X)}{|\Omega_X|}.$$

□

As said above, the next two sections are devoted to the comparison of Shenoy's expected values with the approximations computed as probabilistic expected values considering the above-introduced five probability transforms. First, we do it for basic assignments describing situations under which human decision-makers evince the ambiguity aversion (like the famous Ellsberg's paradox [Ellsberg(1961)]).

5. Basic assignments with strong ambiguity

The examples presented in this section describe situations under which psychologists study human decision making under ambiguity. We consider situations when a color ball is drawn from an urn. We consider $\Omega_X = \{r, b, y, g, w\}$, and the random variable X achieves its value in correspondence whether the color of a drawn ball is red, blue, yellow, green, or white.

Though quite uninteresting from the point of view of this paper (we will see later why it appears uninteresting), we cannot avoid the vacuous basic assignment ι_X representing a total ignorance. In this case, we do not have any other information about the balls in the urn but

- there is at least one ball in the urn ($\emptyset$ is excluded from 2^{Ω_X});
- the urn contains only balls of the specified colors.

We will also consider a situation described by the famous Ellsberg's example [Ellsberg(1961)]. He considers the urn containing ninety balls, thirty of them are red, the remaining balls are either blue or yellow with unknown proportion. It may even happen that all of the remaining sixty balls are of the same color – blue or yellow. This situation is well described by a basic assignment m_e with two focal elements:

$m_e(\{r\}) = \frac{1}{3}$ and $m_e(\{b, y\}) = \frac{2}{3}$. (For a totally different treatment of this problem see [Pfeifer and Pankka(2017)].)

Like the Ellsberg's example, a one-red-ball example [Jiroušek and Shenoy(2017)] describes a situation in which the behavior of human decision-makers is considered paradoxical. In this example we know the total number of balls in the urn (it equals n) and that one and only one ball is red. The proportion of the remaining colors in the urn is unknown. The situation is depicted by basic assignment $m_{r,n}$ with two focal elements: $m_{r,n}(\{r\}) = \frac{1}{n}$ and $m_{r,n}(\{b, y, g, w\}) = \frac{n-1}{n}$. In this section we consider several such basic assignments with different total numbers of balls. Thus, e.g., for $n = 5$ we consider $m_{r,5}(\{r\}) = \frac{1}{5}$ and $m_{r,5}(\{b, y, g, w\}) = \frac{4}{5}$.

An interesting situation is got when we consider a basic assignment expressing the knowledge that, like in the Ellsberg's example, only balls of three colors (red, blue, and yellow) are in the urn, and we know that at least 20 % of them are red and not more than 30 % are yellow. This knowledge is expressed by the following basic assignment m_q: $m_q(\{r\}) = 0.2$, $m_q(\{r, b\}) = 0.5$, $m_q(\{r, b, y\}) = 0.3$. Notice that in this case the focal elements of m_q are nested ($\{r\} \subseteq \{r, b\} \subseteq \{r, b, y\}$) , and therefore the corresponding belief function is known to be a *possibilistic* measure.

Another possibilistic measure is the following basic assignment m_p, for which: $m_p(\{r\}) = 0.1$, $m_p(\{r, b\}) = 0.2$, $m_p(\{r, b, y\}) = 0.3$, $m_p(\{r, b, y, g\}) = 0.2$, $m_p(\Omega) = 0.2$.

The last basic assignment considered in this section is *pseudo-Bayesain*, i.e., the focal elements are only singletons and the whole space Ω_X: $m_b(\{r\}) = 0.5$, $m_b(\{b\}) = 0.05$, $m_b(\{y\}) = 0.05$, $m_b(\{g\}) = 0.05$, $m_b(\{w\}) = 0.05$, $m_b(\Omega) = 0.3$.

Table 1.: Basic assignments

denotation	values of all focal elements
ι_X	$\iota_X(\Omega) = 1$
m_e	$m_e(\{r\}) = \frac{1}{3}$, $m_e(\{b, y\}) = \frac{2}{3}$
$m_{r,3}$	$m_{r,3}(\{r\}) = \frac{1}{3}$, $m_{r,n}(\{b, y, g, w\}) = \frac{2}{3}$
$m_{r,5}$	$m_{r,5}(\{r\}) = \frac{1}{5}$, $m_{r,n}(\{b, y, g, w\}) = \frac{4}{5}$
$m_{r,15}$	$m_{r,15}(\{r\}) = \frac{1}{15}$, $m_{r,n}(\{b, y, g, w\}) = \frac{14}{15}$
m_q	$m_q(\{r\}) = 0.2$, $m_q(\{r, b\}) = 0.5$, $m_q(\{r, b, y\}) = 0.3$
m_p	$m_p(\{r\}) = 0.1$, $m_p(\{r, b\}) = 0.2$, $m_p(\{r, b, y\}) = 0.3$, $m_p(\{r, b, y, g\}) = 0.2$, $m_p(\Omega) = 0.2$
m_a	$m_a(\{r, b\}) = 0.2$, $m_a(\{y, g, w\}) = 0.3$, $m_a(\Omega) = 0.5$
m_b	$m_b(\{r\}) = 0.5$, $m_b(\{b\}) = 0.05$, $m_b(\{y\}) = 0.05$, $m_b(\{g\}) = 0.05$, $m_b(\{w\}) = 0.05$, $m_b(\Omega) = 0.3$

For a survey of all basic assignments, which are considered in this section see Table 1. In this table, only focal elements are presented. In other words, if a set $\mathsf{a} \in 2^{\Omega}$ does not explicitly appear in the table, it means that its corresponding basic assignment equals 0. The corresponding probability transforms are in Table 2.

For the purpose of this section, we use just eight utility functions; see Table 3. Notice that the first four utility functions correspond to the Ellsberg's example.

In this section, we describe results obtained from the experimental computations. For each pair, a basic assignment from Table 1 and a utility function from Table 3 we compute six values:

- Shenoy's expected utility value;
- expected utility value computed using the pignistic transform;
- expected utility value computed using the plausibility transform;
- expected utility value computed using the maximum entropy transform;
- expected utility value computed using the Perez' barycenter transform.
- expected utility value computed using the convex combination transform.

Each expected utility value computed using a probability transform is then compared with the corresponding Shenoy's expected utility value. Thus, for each probability transform, we receive $8 \times 9 = 72$ matrix of values expressing the difference between the results achieved with the help of the corresponding probability transform and those achieved by Shenoy's operator. This difference is expressed in percentage: by how many percents the expected value computed with the help of the respective probability transform differs from Shenoy's expected value. To make it visually more attractive, we depict in Figure 1 each such matrix by an 8×9 table, where each difference corresponds to one box. The darker the box, the greater the corresponding difference.

We see that the first row in all tables corresponding to ι_X is empty. It means that under the condition of total ignorance all the considered approaches yield the same expected utility (all probability transforms give the uniform probability measure). On the other side, one can immediately see that none of the considered probability transforms yields the same expected utility values as Shenoy's expectation operator for m_q, m_p, m_a and m_b. From a brief look, one can guess that the convex combination and pignistic transforms may yield the best approximations and that the behavior of the plausibility transform is from this point of view the least felicitous.

Another (for some readers a more appropriate) comparison of the five considered probability transforms may be got with the help of the Kullback-Leibler divergence. In Table 4, rows show results achieved for different basic assignment from Table 1. Recall that values of divergence $Div(Q; Sh\_P_m)$ express the dissimilarity of Sh_P_m and Q, where Q stands for probability measures received from the respective basic assignment m by the considered five probability transforms, and Sh_P_m denote the probability measure received by the process described in Proposition 3.1. The higher these values, the greater the difference between Sh_P_m and Q. The respective probability measures are in Table 2. The comparison of values from Table 4 confirms the conclusions mentioned in the previous paragraph: in general, the convex combination and pignistic transforms yield the best approximations, the plausibility transform is from this point of view the worst. Nevertheless, one has to take these simple conclusions with great care. A more detailed look at Table 4 shows that for m_q the best approximation is given by the plausibility transform.

Table 2.: Probability measures corresponding to basic assignments

basic assignment	type of transform	corresponding probability measure				
		r	b	y	g	w
ι_X	Sh_P_m	0.2000	0.2000	0.2000	0.2000	0.2000
	Bet_P_m	0.2000	0.2000	0.2000	0.2000	0.2000
	Pl_P_m	0.2000	0.2000	0.2000	0.2000	0.2000
	Me_P_m	0.2000	0.2000	0.2000	0.2000	0.2000
	Bac_P_m	0.2000	0.2000	0.2000	0.2000	0.2000
	Cs_P_m	0.2000	0.2000	0.2000	0.2000	0.2000
m_e	Sh_P_m	0.3333	0.3333	0.3333	0.0000	0.0000
	Bet_P_m	0.3333	0.3333	0.3333	0.0000	0.0000
	Pl_P_m	0.2000	0.4000	0.4000	0.0000	0.0000
	Me_P_m	0.3333	0.3333	0.3333	0.0000	0.0000
	Bac_P_m	0.3333	0.3333	0.3333	0.0000	0.0000
	Cs_P_m	0.3333	0.3333	0.3333	0.0000	0.0000
$m_{r,3}$	Sh_P_m	0.3333	0.1666	0.1666	0.1666	0.1666
	Bet_P_m	0.3333	0.1666	0.1666	0.1666	0.1666
	Pl_P_m	0.1111	0.2222	0.2222	0.2222	0.2222
	Me_P_m	0.3333	0.1666	0.1666	0.1666	0.1666
	Bac_P_m	0.3333	0.1666	0.1666	0.1666	0.1666
	Cs_P_m	0.3333	0.1666	0.1666	0.1666	0.1666
m_q	Sh_P_m	0.4676	0.3405	0.1918	0.0000	0.0000
	Bet_P_m	0.5500	0.3500	0.1000	0.0000	0.0000
	Pl_P_m	0.4762	0.3810	0.1429	0.0000	0.0000
	Me_P_m	0.3500	0.3500	0.3000	0.0000	0.0000
	Bac_P_m	0.6004	0.3212	0.0784	0.0000	0.0000
	Cs_P_m	0.5368	0.3368	0.1263	0.0000	0.0000
m_p	Sh_P_m	0.3121	0.2421	0.1923	0.1430	0.1106
	Bet_P_m	0.3900	0.2900	0.1900	0.0900	0.0400
	Pl_P_m	0.3125	0.2812	0.2188	0.1250	0.0625
	Me_P_m	0.2000	0.2000	0.2000	0.2000	0.2000
	Bac_P_m	0.4206	0.2924	0.1752	0.0777	0.0341
	Cs_P_m	0.3613	0.2613	0.2032	0.1161	0.0581
m_a	Sh_P_m	0.2130	0.2130	0.1913	0.1913	0.1913
	Bet_P_m	0.2000	0.2000	0.2000	0.2000	0.2000
	Pl_P_m	0.1842	0.1842	0.2105	0.2105	0.2105
	Me_P_m	0.2000	0.2000	0.2000	0.2000	0.2000
	Bac_P_m	0.1742	0.1742	0.2172	0.2172	0.2172
	Cs_P_m	0.1842	0.1842	0.2105	0.2105	0.2105
m_b	Sh_P_m	0.5154	0.1212	0.1212	0.1212	0.1212
	Bet_P_m	0.5600	0.1100	0.1100	0.1100	0.1100
	Pl_P_m	0.3636	0.1591	0.1591	0.1591	0.1591
	Me_P_m	0.5000	0.1250	0.1250	0.1250	0.1250
	Bac_P_m	0.5393	0.1152	0.1152	0.1152	0.1152
	Cs_P_m	0.5600	0.1100	0.1100	0.1100	0.1100

Table 3.: Utility functions

	r	b	y	g	w
u_1	100	0	0	0	0
u_2	0	100	0	0	0
u_3	100	0	100	0	0
u_4	0	100	100	0	0
u_5	0	100	200	300	0
u_6	0	100	0	200	0
u_7	100	0	0	200	100
u_8	50	150	70	220	30

Table 4.: Kullback-Leibler divergences

basic	$Div(\cdot\,; Sh\_P_m)$				
assignment	Bet_P_m	Pl_P_m	Me_P_m	Bac_P_m	Cs_P_m
ι_X	0	0	0	0	0
m_e	0	0.0437	0	0	0
$m_{r,3}$	0	0.1336	0	0	0
$m_{r,5}$	0	0.0810	0	0	0
$m_{r,15}$	0	0.0270	0	0	0
m_q	0.0337	0.0092	0.0423	0.0611	0.0162
m_p	0.0547	0.0184	0.0663	0.0770	0.0234
m_a	0.0014	0.0069	0.0014	0.0126	0.0009
m_b	0.0040	0.0465	0.0005	0.0011	0.0009
average	0.0104	0.0407	0.0123	0.0169	0.0046

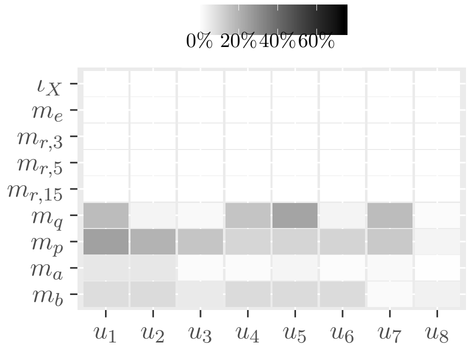
6. Comparison on randomly generated basic assignments

In addition to the computational experiments described in the previous Section, we also realized comparisons with randomly generated basic assignments. Let us stress at the very beginning that, because of the exponential complexity of computation Sh_P_m, all the comparisons presented were done for five-element state space Ω_X only.

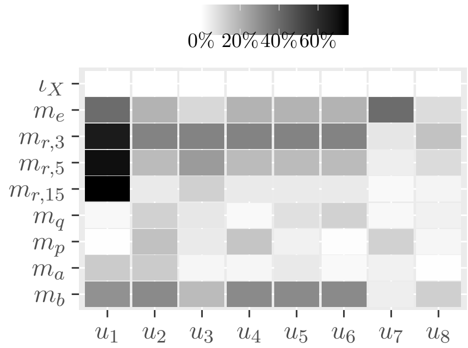
Because of its relative simplicity following (among others) from Proposition 4.1, and because of its popularity in applications, we start comparisons with a class of quasi-Bayesian basic assignments. We randomly generated 100 quasi-Bayesian basic assignments, for each of them, we found the corresponding Sh_P_m and computed the respective probability transforms. The average divergences expressing dissimilarity of Sh_P_m and the computed probability transforms are depicted in Table 5. In this table, we present not only the average Kullback-Leibler divergence but also the average *total variance*

$$Var(Q; P) = \sum_{x\in\Omega_X} |q(x) - p(x)|,$$

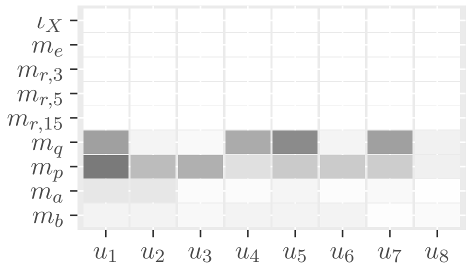
which is preferred by some authors. From this table, one can see that, in this very special situation, Perez' barycenter is the best approximation of Sh_P_m. It holds not only on average; for 97 quasi-Bayesian basic assignments (from 100 randomly generated)



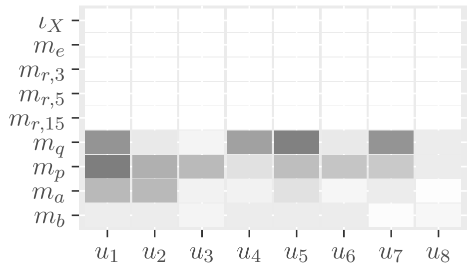
(a) Pignistic transform Bet_P_m



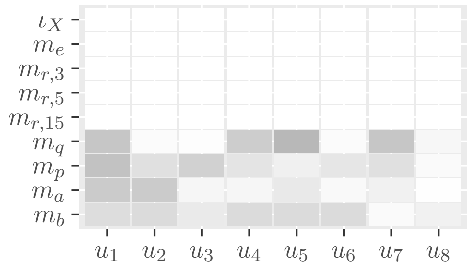
(b) Plausibility transform Pl_P_m



(c) Maximum entropy Me_P_m



(d) Perez'barycenter Bac_P_m



(e) Convex combination Cs_P_m

Figure 1.: Relative differences between Shenoy's expected utility values and those computed using probability transforms

Perez' barycenter approximated Sh_P_m best.

However, quite different results were achieved in a general case. We randomly generated 300 basic assignments for $|\Omega_X| = 5$. Reflecting the fact that in practical applications one usually defines basic assignments with a limited number of focal elements, we put on the randomly generated basic assignments the following restrictions: 100 of them did not have more than 3 focal elements, 100 of them did not have more than 6 focal elements and 100 of the did not have more than 11 focal elements. Since we did not find an interesting difference concerning the number of focal elements, in what follows we refer to the whole group of 300 basic assignments.

Again, for each randomly generated basic assignment, we computed the corresponding Sh_P_m, and all the five considered probability transforms: pignistic and plausibility transforms, the maximum entropy representative, Perez' barycenter and the convex combination transform. The corresponding average values of the Kullback-Leibler divergence and the total variance for the whole group of 300 randomly generated basic

Table 5.: Average of divergences for 100 randomly generated quasi-Bayesian basic assignments

	Div(· ;Sh_Pm)	Var(· ;Sh_Pm)
$Bet\_P_m = Cs\_P_m$	0.0041	0.0549
Pl_P_m	0.0070	0.0858
Me_P_m	0.0131	0.1087
Bac_P_m	0.0004	0.0152

assignments are shown in Table 6. In contrast to the results concerning quasi-Bayesian basic assignment, we do not see a strictly"dominant" probability transform. The average divergences from Table 6 are neither too reflective of the numbers, how many times the individual probability transforms approximated the probability measure Sh_P_m best, which is reported in the last column of Table 6. I.e., pignistic transform yielded 53x (out of 300 basic assignments) the best approximation of the probability measure Sh_P_m, plausibility transform" 74x, and so on (notice the sum of these numbers exceeds 300 because it happens quite often that two or more probability transforms yield the same probability measure). In a way, the numbers from Table 6 may slightly support the observation from the previous Section: the convex combination and the pignistic transforms may yield better approximations than other probability transforms.

Table 6.: Average of divergences for 300 randomly generated basic assignments

	Div(· ;Sh_Pm)	Var(· ;Sh_Pm)	
Bet_P_m	0.0245	0.1655	53
Pl_P_m	0.0315	0.1994	74
Me_P_m	0.0357	0.1868	104
Bac_P_m	0.0309	0.1835	44
Cs_P_m	0.0226	0.1597	92

7. Conclusions

We showed that Shenoy's expected utility can be computed in two ways: either extending the utility function to whole 2^{Ω_X} and computing its expected value using Shenoy's Formula (3), or finding the respective probability measure by the process described in Proposition 3.1, and computing the respective probabilistic expected value. However, because of their great computational complexity, both these procedures can be used only for small examples. Therefore, we considered five probability transforms that were, in our opinion, the best candidates from all the probability transforms described in the literature, and made a number of computational experiments trying to answer the question whether any of them is better than the others regarding their ability to approximate Shenoy's expectation. Based on specifically selected nine basic assignments described in Section 5, it seemed that, though not for all situations, usually the best approximation could be got with the help of convex combination and/or pignistic transforms. Let us recall that the pignistic transform was strongly advocated for decision-making applications by Philippe Smets [Smets(1989), Smets(2005)].

Quite different results were achieved with randomly generated quasi-Bayesian basic assignments. For this class of basic assignments, which is popular in applications because of their simple interpretability, the best approximations were undoubtedly yielded by Perez' barycenter transform. Unfortunately, such an unquestionable conclusion holds just for quasi-Bayesian basic assignments. In a general case, the achieved results are not convincing, they just slightly support observations achieved for specifically selected basic assignments described in Section 5.

Eventually, it is important to pinpoint the fact that all the commented results are based on the computations with a small state space ($|\Omega_X| = 5$), and that one can hardly make analogous computational experiments with cardinalities of a practical size because of the computational complexity connected with Shenoy's expected utility.

Acknowledgement

The research was supported by a grant from GA ČR No. 19-06569S (the first two authors).

References

[Choquet(1953)] Choquet, G. 1953. "1954." *Theory of Capacities." Annales de l'Institut Fourier5* 131–295.

[Cobb and Shenoy(2006)] Cobb, B. R., and P. P. Shenoy. 2006. "On the plausibility transformation method for translating belief function models to probability models." *International Journal of Approximate Reasoning* 41 (3): 314–340.

[Coletti, Petturiti, and Vantaggi(2015)] Coletti, Giulianella, Davide Petturiti, and Barbara Vantaggi. 2015. "Rationality principles for preferences on belief functions." *Kybernetika* 51 (3): 486–507.

[Coletti, Petturiti, and Vantaggi(2019)] Coletti, Giulianella, Davide Petturiti, and Barbara Vantaggi. 2019. "Models for pessimistic or optimistic decisions under different uncertain scenarios." *International Journal of Approximate Reasoning* 105: 305–326.

[Cuzzolin(2012)] Cuzzolin, F. 2012. "On the relative belief transform." *International Journal of Approximate Reasoning* 53 (5): 786–804.

[Ellsberg(1961)] Ellsberg, D. 1961. "Risk, ambiguity and the Savage axioms." *Quarterly Journal of Economics* 75 (2): 643–669.

[Gilboa and Schmeidler(1994)] Gilboa, Itzhak, and David Schmeidler. 1994. "Additive representations of non-additive measures and the Choquet integral." *Annals of Operations Research* 52 (1): 43–65.

[Halpern and Fagin(1992)] Halpern, J. Y., and R. Fagin. 1992. "Two views of belief: Belief as generalized probability and belief as evidence." *Artificial Intelligence* 54: 3.

[Jiroušek and Shenoy(2017)] Jiroušek, Radim, and Prakash P Shenoy. 2017. "Ambiguity aversion and a decision-theoretic framework using belief functions." In *Computational Intelligence (SSCI), 2017 IEEE Symposium Series on*, 1–7. IEEE.

[Perez(1985)] Perez, A. 1985. "The barycenter concept of a set of probability measures as a tool in statistical decision." In *4th Int. Vilnius Conf. on Probability Theory and Math. Statistics*, 226–228.

[Pfeifer and Pankka(2017)] Pfeifer, N., and H Pankka. 2017. "Modeling the Ellsberg paradox by argument strength." In *Proceedings of the 39th Cognitive Science So-*

ciety Meeting, edited by Howes A. Tenbrink T. Gunzelmann, G. and E. Davelaar, 925–930.

[Savage(1954)] Savage, L. J. 1954. *Foundations of Statistics.* New York: Wiley.

[Shafer(1976)] Shafer, G. 1976. *A Mathematical Theory of Evidence.* Princeton University Press.

[Shafer(1986)] Shafer, G. 1986. "Savage-revisited." *Statistical Science* 1 (4): 463–501.

[Shenoy(2018)] Shenoy, P. Prakash. 2018. "An Expectation Operator for Belief Functions in the Dempster-Shafer Theory." In *Proceedings of the 11th Workshop on Uncertainty Processing*, edited by V. Kratochvíl and J. Vejnarová, 165–176. MatfyzPress, Praha, Czech Republic.

[Shenoy(2019)] Shenoy, P. Prakash. 2019. "An Expectation Operator for Belief Functions in the Dempster-Shafer Theory." *International Journal of General Systems* Submitted.

[Smets(2005)] Smets, P. 2005. "Decision making in the TBM: the necessity of the pignistic transformation." *International Journal of Approximate Reasoning* 38 (2): 133–147.

[Smets(1981)] Smets, Philippe. 1981. "The degree of belief in a fuzzy event." *Information sciences* 25 (1): 1–19.

[Smets(1989)] Smets, Philippe. 1989. "Constructing the Pignistic Probability Function in a Context of Uncertainty." In *UAI*, Vol. 89, 29–40.