Comparison of Shenoy's Expectation Operator with Probabilistic Transforms and Perez' Barycenter

Radim Jiroušek
Faculty of Management
University of Economics, Prague
radim@utia.cas.cz

Václav Kratochvíl
ÚTIA
Czech Academy of Sciences
velorex@utia.cas.cz

Abstract

Shenoy's paper published in this Proceedings of WUPES 2018 introduces an operator that gives instructions how to compute an expected value in the Dempster-Shafer theory of evidence. Up to now, there was no direct way to get the expected value of a utility function in D-S theory. If needed, one had to find a probability mass function corresponding to the considered belief function, and then - using this probability mass function - to compute the classical probabilistic expectation.

In this paper, we take four different approaches to defining probabilistic representatives of a belief function and compare which one yields to the best approximations of Shenoy's expected values of various utility functions. The achieved results support our conjecture that there does not exist a probabilistic representative of a belief function that would yield the same expectations as the Shenoy's new operator.

1 Introduction

Criteria for finding optimal decisions are usually based on a maximum expected utility principle. As Glenn Shafer [8] wrote already in 1986: *The controversy raised by this book* (here he meant the Savage's book [7]) *and Savage's subsequent writings is 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 function theory. Nevertheless, to compute the necessary value of expected utility, the respective belief function is usually transformed into an appropriate probability distribution. For this, several procedures were designed - we call them *probability transforms* 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 [1]. The other transforms are more likely compatible with the theory of belief functions interpreted as *generalized probability* [4]. This interpretation reflects the fact that a belief function specifies a convex set of probability distributions, which is called a credal set. In this paper we consider widely used *pignistic transform* advocated by Philippe Smets [10], and two others that are usually omitted in the context of belief function: *maximum entropy* and *Perez' barycenter* [6].

To our best knowledge, the first idea how to compute an expected value for a belief function directly, i.e., avoiding its transformation into a probability distribution, is due to Prakash Shenoy [9]. From the theoretical point of view, it is a concept deserving a deep 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. If new computational procedures (avoiding the calculation of a commonality function and subsequent summation over all nonempty subsets of a state space) are not found, the application of this approach in practical problems will be limited. Though we conjecture that there does not exist a probability transform that would yield the same expectations as the Shenoy's operator, there arises an interesting problem: find a probability transform, which approximates the results of the new operator best. And it is the goal of this paper to compare the above-mentioned four probability transforms from this point of view.

To achieve this goal, the rest of the paper is organized as follows. Section 2 recalls basic concepts of belief function theory and introduces the necessary notation. In Section 3, four selected probability transforms are formally introduced. A battery of basic assignments, as well as a set of utility functions used for comparison are presented in Section 4. The main result of this paper (the comparison of the computed expected values) is presented in Section 5. The paper is concluded by Section 6, where the further research is proposed.

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

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

The subsets $\mathsf{a} \in 2^{\Omega_X}$ such that $m(\mathsf{a}) > 0$ are called *focal* elements of m. An important example is the vacuous basic assignment for X, denoted by ι_X, such that $\iota_X(\Omega_X) = 1$. It corresponds to a total ignorance. If all focal elements of m

[1]Notice that we exclude the empty set from 2^{Ω_X} in this paper.

are singletons (one-element subsets) of Ω_X, then we say m is *Bayesian*. In this case, m is equivalent to a probability distribution.

The information in a basic assignment m can be equivalently represented by corresponding *belief* and *plausibility* functions Bel_m and Pl_m, respectively that are defined as

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

for all $\mathsf{a} \in 2^{\Omega_X}$. In this paper we need also the fourth possibility of expressing a belief function. A *commonality function* for m is defined for all $\mathsf{a} \in 2^{\Omega_X}$

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

Notice that it is obvious that for all $\mathbf{a} \in 2^{\Omega}$, $Bel(\mathbf{a}) \leq Pl(\mathbf{a})$. For singletons (one-element subsets of Ω_X) commonality and plausibility functions coincide:

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

for all $x \in \Omega_X$. Since we consider only normal basic assignments for which $\sum_{\mathsf{a} \in 2^{\Omega_X}} m(\mathsf{a}) = 1$, it can be shown that

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

For a basic assignment m on Ω_X and the corresponding commonality function Q_m, Shenoy proposes a new operator computing the expected value of a general function[2] $g : 2^{\Omega_X} \longrightarrow \mathbb{R}$ [9]. Let us adopt his approach to the computation of an expected value of utility function $u : \Omega_X \longrightarrow \mathbb{R}$. First we need to extend the utility function from Ω_X to the whole 2^{Ω_X} (we denote the extension $\hat{u}$) in the way that for all $\mathsf{a} \in 2^{\Omega_X}$

$$\min_{x \in \mathsf{a}} \{u(x)\} \leq \hat{u}(\mathsf{a}) \leq \max_{x \in \mathsf{a}} \{u(x)\}.$$

Following Shenoy's idea we take the weighted average

$$\hat{u}(\mathsf{a}) = \frac{\sum_{x \in \mathsf{a}} u(x) Q_m(\{x\})}{\sum_{x \in \mathsf{a}} Q_m(\{x\})}$$

(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 value from the above specified interval; for example $\hat{u}(\mathsf{a}) = (\min_{x \in \mathsf{a}} \{u(x)\} + \max_{x \in \mathsf{a}} \{u(x)\})/2)$. Then Shenoy defines the expected value of u with respect to m as follows:

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

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

The last notion introduced in this section was already mentioned in Introduction. Basic assignment m specifies the following convex set of probability distributions P on Ω ($\mathcal{P}_\Omega$ denote the set of all probability distributions on Ω):

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

$\mathcal{P}(m)$ is called a *credal set* of basic assignment m. If m is Bayesian, then $\mathcal{P}(m)$ contains just one probability distribution.

3 Probability transforms

In this paper, we study properties of the following four mappings that assign a probability distribution to each basic assignment. For other probability transforms see e.g. [2]. Perhaps, the most famous is *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 the so-called *plausibility transform*, which is the respective plausibility function normalized on singletons. Formally it is defined for all $x \in \Omega_X$

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

The other two 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 distribution P

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

The other is the Perez' barycenter [6] 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 (called also Kullback-Leibler divergence in the literature)

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

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

4 Basic assignments and utility functions

All the examples presented in this paper correspond to a situation 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 it later), 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 balls of the specified colors only.

We will also consider a situation described by the famous Ellsberg's example [3]. He considers the situation when the urn contains 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 $m_e(\{b, y\}) = \frac{2}{3}$.

Like the Ellsberg's example, a one-red-ball example [5] 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 the next section we will consider several such basic assignments with different total numbers of balls. Thus, e.g., for $n = 5$ we will 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 50 % 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: $m_p(\{r\}) = 0.1$, $m_p(\{r, b\}) = 0.2$, $m_p(\{r, b, y\}) = 0.3$, $m_p(\{r, b, y, g\}) = 0.2$, $m_p(\Omega) = 0.2$.

For a survey of all basic assignments considered in the following 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.

For the purpose of this paper, we used just eight utility function. Naturally, to make a really serious comparison of probability functions we expect to use much larger batteries of basic assignments and utility functions, as well as we expect

Table 1: Basic assignments

denotation	values of all focal elements
ι_X	$\iota_X(\Omega) = 1$
m_e	$m_e(\{r\}) = \frac{1}{3}$, $m_e(\{b, y\}) = \frac{2}{3}$
$m_{r,n}$	$m_{r,n}(\{r\}) = \frac{1}{n}$, $m_{r,n}(\{b, y, g, w\}) = \frac{n-1}{n}$
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	$m_a(\{r, b\}) = 0.2$, $m_a(\{y, g, w\}) = 0.3$, $m_a(\Omega) = 0.5$

to widen also the set of the compared probability transforms. For the considered utility functions see Table 2. Notice that the first four utility functions correspond to the Ellsberg's example.

Table 2: Utility functions

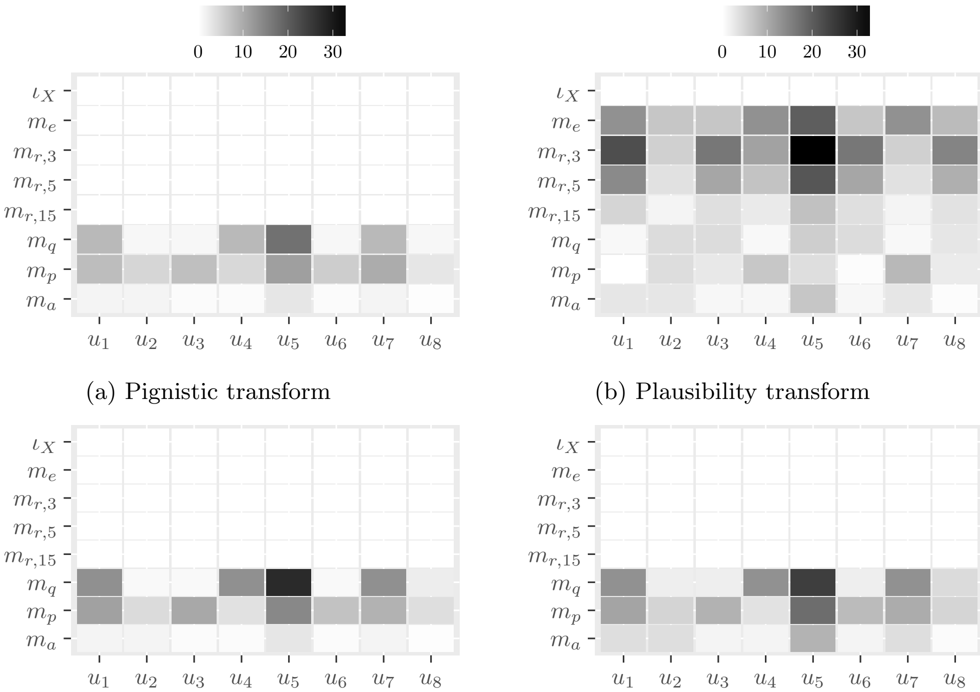
	r	b	y	g	w
u_1	100	0	0	0	0
u_2	0	100	0	0	0
u_3	100	0	100	0	0
u_4	0	100	100	0	0
u_5	0	100	200	300	0
u_6	0	100	0	200	0
u_7	100	0	0	200	100
u_8	50	150	70	220	30

5 Computations

In this section, we describe results obtained from the experimental computations. For each pair, a basic assignment from Table 1 (we considered three basic assignments corresponding to one-red-ball example: $m_{r,3}$, $m_{r,5}$, and $m_{r,15}$, i.e., 8 basic assignments in total) and a utility function from Table 2 we compute five values:

- Shenoy's expected utility value;
- expected utility value computed using pignistic transform;
- expected utility value computed using plausibility transform;

- expected utility value computed using maximum entropy transform;
- expected utility value computed using Perez' barycenter transform.



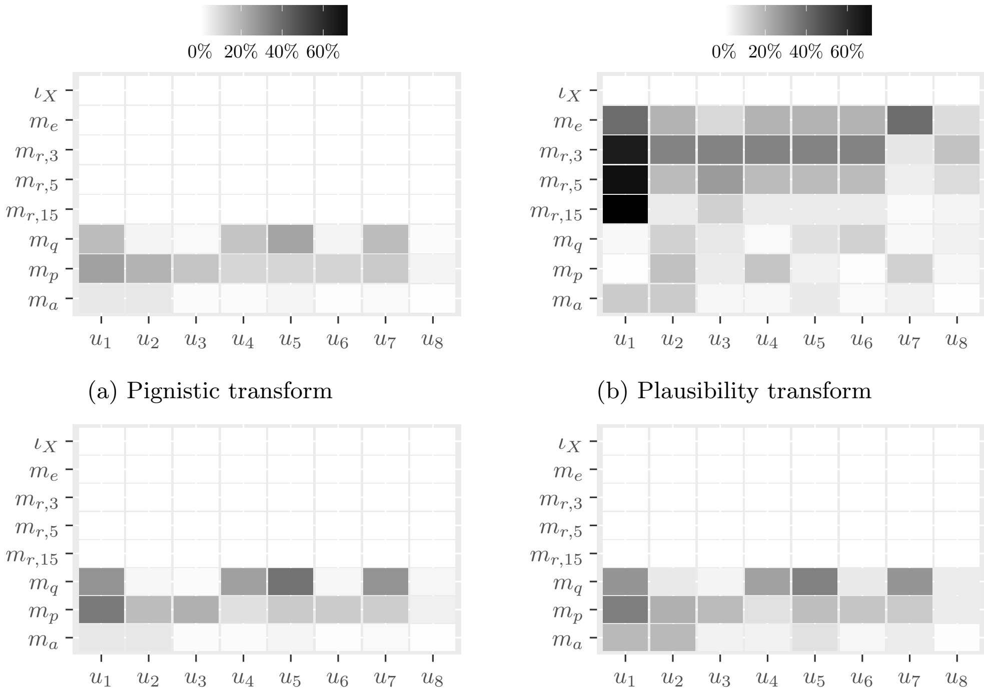
(a) Pignistic transform (b) Plausibility transform

(c) Maximum entropy (d) Perez'barycenter

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

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 $8 \times 8 = 64$ matrix of values (absolute values of the differences) expressing the difference between the results achieved with the help of the corresponding probability transform and those achieved by the new operator. To make it visually attractive, we depict each such matrix by a 8×8 table, where each difference corresponds to one box. The darker the box, the higher the corresponding difference. Figure 1 depicts the corresponding differences, Figure 2 depicts by how many percent the expected value computed with the help of the respective probability transform differs from the Shenoy's expected value.

We see that the first row in all tables corresponding to ι_X is empty meaning that under the condition of total ignorance all the considered approaches yield



(a) Pignistic transform

(b) Plausibility transform

(c) Maximum entropy

(d) Perez'barycenter

Figure 2: Relative difference between Shenoy's expected utility values and those computed using probability transforms

the same expected utility (all probability transforms give the uniform probability distribution).

6 Conclusions

Though the results achieved in this study should be considered preliminary, they give a hint that the plausibility transform, regardless it is considered by Cobb and Shenoy the only one corresponding to Dempster-Shafer theory of evidence, is quite unsuitable for estimating the expected utility. The question is whether there is any positive result that can be concluded from the described simple study. The achieved results may support the Smets' conviction that the pignistic transform is the best one for decision-making. The results may also suggest that for the situations described by simple basic assignments, the pignistic transform yield the same results as the maximum entropy principle and the Perez' barycenter.

In any case, this study is a starting milestone for further research. From the theoretical viewpoint, it would be interesting to know whether our conjecture about the nonexistence of a probability transform yielding the same expected values as Shenoy's operator is true or not. From the practical point of view, because of a great computational complexity of the new expectation operator, it is interesting to perform a study similar to the one presented in this paper, but with much greater the number of basic assignments and a higher the number of utility functions.

Acknowledgement

This work was supported by the Czech Science Foundation (project 16-12010S), and by the Czech Academy of Sciences (project MOST-18-04).

References

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

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

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

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

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

[6] A. Perez. 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*, pages 226–228, 1985.

[7] L. J. Savage. *Foundations of Statistics.* Wiley, New York, 1954.

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

[9] P. P. Shenoy. An expectation operator for belief functions in the Dempster-Shafer theory. to appear in Proceedings of WUPES'18.

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