How Sir Harold Jeffreys would create a belief function based on data

Milan Daniel[1], Radim Jiroušek[2], and Václav Kratochvíl[2]

[1]Institute of Computer Sciences, Czech Academy of Sciences
[1]*milan.daniel@cs.cas.cz*
[2]Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czech Republic
[2] {*radim;velorex*}*@utia.cas.cz*

Abstract

Not all normalized nonnegative monotone set functions are belief functions. This paper investigates ways to modify them to obtain a belief function that preserves some of their properties. The problem is motivated by an approach to data-based learning of belief function models. The approach is based on the idea that classical methods of mathematical statistics can provide estimates of lower bounds for unknown probabilities. Thus, methods of mathematical statistics can be used to obtain a reasonable rough estimate, which is further elaborated to obtain a desired belief function model.

1 Introduction

When learning a probabilistic model from data, you need to determine a large number of parameters, i.e., find estimates for many probabilities. But you are not completely sure about any of them especially if you have a limited amount of data. Using Bayesian statistics, you get a posterior distribution of the considered probability (usually suggesting to accept the most probable value), and if you want to be careful when applying the frequentist approach, you should consider some confidence intervals of the needed parameters. Thus, when learning models from data, it may seem more natural to do so in the framework of belief function theory (Shafer, 1976) than in the framework of probability theory. However, even this approach is not trivial. Although it has some similar properties, the function whose values are estimates of the lower bounds of the confidence intervals for each of the probabilities considered does not form a belief function. This happens only in small examples. Otherwise, we get more general monotonic set functions (usually called capacities), which need to be processed in order to be transformed into belief functions. And the possible ways to do this are the subject of this paper.

Thus, in this paper we explore the possibilities of searching for a belief function that would be obtained as a modification of the values of the statistical estimates of confidence intervals. After introducing the necessary terminology and notation, we begin to study the properties of functions whose values correspond to the statistical estimates of the lower bounds of confidence intervals. In this paper, we study only two types of their approximation by belief functions. In Section 3 we propose a procedure for finding approximations that preserve all the information extracted from the data by the statistical estimates, and in Sections 4 and 5 we study the lower approximations that do not add to the model any information that is not encoded in the considered statistical estimates.

2 Necessary notions from theory of belief functions

In the whole paper, Ω denotes a finite *frame of discernment*. A non-negative set function $f : 2^\Omega \to \mathbb{R}^+$ is called *pseudo-belief function* (PBF) if $f(\emptyset) = 0$, and it is

monotonic for $a \subset b \subseteq \Omega : f(a) \leq f(b)$, and

normalized $f(\Omega) = 1$.

Each PBF f is connected with a set of probability distributions defined on Ω. A *credal set* of PBF f is the following set of probability distributions.

$$\mathcal{P}(f) = \left\{\pi \text{ defined on } \Omega : \left(\forall a \subseteq \Omega : \pi(a) \geq f(a)\right)\right\}. \tag{1}$$

There is a natural partial ordering for PBFs. $f \leq g$ means that $f(a) \leq g(a)$ for all $a \subseteq \Omega$. $f < g$ denote that $f \leq g$ and $f \neq g$. Notice also that $f \leq g$ is equivalent with $\mathcal{P}(f) \supseteq \mathcal{P}(g)$.

We say a function $f : 2^\Omega \to \mathbb{R}^+$ is a *belief function* (BF) if it is a PBF and for all non-empty $a \subseteq \Omega$

$$\sum_{b \subseteq a} -1^{|a \setminus b|} f(b) \geq 0.$$

The set of all belief functions defined on given Ω is denoted by $\mathcal{BF}$.

Thus, for each BF f one can define non-negative function m_f on 2^Ω called *basic probability assignment* (BPA) m_f by the following expression

$$m_f(a) = \sum_{b \subseteq a} -1^{|a \setminus b|} f(b), \tag{2}$$

for which

$$f(a) = \sum_{b \subseteq a} m_f(b). \tag{3}$$

The sets $a \subseteq \Omega$ for which $m_f(a) > 0$ are called *focal elements* of f. Notice that if all focal elements of BF f are singletons ($|a| = 1$), then also $\mathcal{P}$ is a singleton. These BFs are called *Bayesian*. The representation of BF f using its BPA m_f is often preferred. It makes

the introduction of some notions, such as the following notion of pignistic transform and simple specification, more intuitive.

Pignistic transform of BF f is a specific element of its credal set $\mathcal{P}(f)$ (Dubois and Prade, 1982). It is a probability distribution defined for all $\omega \in \Omega$

$$\pi_f(\omega) = \sum_{a \subseteq \Omega : \omega \in a} \frac{m_f(a)}{|a|}.$$

Using Equation (3), it is trivial to show that $\pi_f(a) \geq f(a)$. It means that $\mathcal{P}(f) \neq \emptyset$ for all BFs f. Note that it does not hold for all PBFs. As a trivial example consider $\Omega = \{\omega_1, \omega_2\}$, and PBF g defined $g(\{\omega_1\} = g(\{\omega_2\} = 0.6$, $g(\Omega) = 1$, which obviously complies with the definition of PBFs. As a little bit more sophisticated example consider an arbitrary Ω, and PBF g defined

$$g(a) = \begin{cases} 0 & \text{if } |a| < |\Omega| - 1, \\ 1 - \frac{1}{|\Omega|+1} & \text{if } |a| = |\Omega| - 1, \\ 1 & \text{if } a = \Omega. \end{cases}$$

It means that for $\pi \in \mathcal{P}(g)$, $\pi(\omega) \leq \frac{1}{|\Omega|+1}$ for each $\omega \in \Omega$, which cannot hold for any probability distribution π.

The following trivial assertion holds.

Lemma 1 *Let g be a PBF on Ω. $\mathcal{P}(g) \neq \emptyset$ if and only if there exists BF f, for which $f \geq g$.*

Proof. If $\mathcal{P}(g) \neq \emptyset$ denote $\pi \in \mathcal{P}(g)$, and define Bayesian BF f through its BPA $m_f(\{\omega\}) = \pi(\omega)$. Since $\pi(a) \geq g(a)$ for all $a \subseteq \Omega$, the also $f \geq g$.

The opposite part of the equivalence is even simpler. The credal set $\mathcal{P}(f)$ of BF f is always nonempty, and therefore $\mathcal{P}(g) \supseteq \mathcal{P}(f)$ must also be nonempty. □

In (Jiroušek and Kratochvíl, 2025), the following notion was defined for BFs. We say that BF f is a *simple specification* of BF g if m_f is created from m_g by shifting some of its mass from some focal element to its subset; more precisely, there exist subsets $a \subset b \subseteq \Omega$ such that $m_f(a) = m_g(a) + \varepsilon$, and $m_f(b) = m_g(b) - \varepsilon$, and all the remaining focal elements of m_f are the copies of the focal elements of m_g, i.e., for all $c \in \Omega \setminus \{a, b\}$, $m_f(c) = m_g(c)$. Thus, this operation means that $f(c) = g(c) + \varepsilon$ for all $c \subset \Omega$, for which $(a \subseteq c) \setminus (b \subseteq c)$, and for all remaining c, $f(c) = g(c)$. So we see that $f > g$. In this paper we will apply this notion to all PBFs with the same effect. We will also generalize this notion in the sense that we will consider $\varepsilon < 0$. The reader immediately sees that for negative values of ε, $f < g$, and thus we will call this modification *simple generalization.*

Recall that in (Jiroušek and Kratochvíl, 2025) we proved the following assertion stating that if $f > g$, then f can be obtained from g by a sequence of simple specifications.

Lemma 2 *Let BFs f and g are defined on Ω. If $g < f$, then there exists a finite sequence of BFs $g = h_1, h_2, \ldots, h_k = f$ such that each h_{i+1} is a simple specification of h_i.*

In the computational procedures introduced below we will use *mass redistribution*, which consists of several simple specifications (generalizations) performed simultaneously. By this new term we understand the process, when masses assigned to several subsets are changed. We change PBF g to PBF f by *redistributing* ε from $b \subseteq \Omega$ to r its proper subsets $a_1, \ldots, a_r$ if

(i) for all $\ell = 1, \ldots, r$, $a_\ell \subsetneq b$;

(ii) $\varepsilon_1, \ldots, \varepsilon_r$ are such that $\sum_{\ell=1}^{r} \varepsilon_\ell = \varepsilon$, and for all ℓ, $\frac{\varepsilon_\ell}{\varepsilon} > 0$;

(iii) $m_f(b) = m_g(b) - \varepsilon$;

(iv) for all $\ell = 1, \ldots, r$, $m_f(a_\ell) = m_g(a_\ell) + \varepsilon_\ell$;

(v) for all the remaining $c \subseteq \Omega \setminus \{b, a_1, \ldots, a_r\}$, $m_f(c) = m_g(c)$.

Notice that if $\varepsilon > 0$, f is a specification of g (i.e., $f > g$), if $\varepsilon < 0$, f is a generalization of g (i.e., $f < g$) because the condition (ii) guarantees that all ε_ℓ are of the same sign; they are all positive or negative.

Each PBF g splits the whole set of BFs $\mathcal{BF}$ into three disjoint parts: *inner* (*upper*) *envelop* of g

$$\overline{\mathcal{B}}(g) = \{f \in \mathcal{BF} : f \geq g\},$$

outer (*lower*) *envelop* of g

$$\underline{\mathcal{B}}(g) = \{f \in \mathcal{BF} : f \leq g\},$$

and the set of BFs, which are *incomparable* with g, i.e., $\mathcal{BF} \setminus \left(\underline{\mathcal{B}}(g) \cup \overline{\mathcal{B}}(g)\right)$.

3 Upper Approximations of Pseudo-Belief Functions

Consider a general PBF $g : 2^\Omega \to [0, 1]$. Let us explore ways to find a suitable approximation of g with some BF. There are several possible ways to do this. In this paper, we will only consider approximations by BFs either from $\underline{\mathcal{B}}(g)$ or $\overline{\mathcal{B}}(g)$. Approximations from $\underline{\mathcal{B}}(g)$ are supported by the fact that they do not add any information that is not contained in g. But the only task we can solve optimally is to look for the solution in $\overline{\mathcal{B}}(g)$. In this case the optimal solution is obtained by the following *Upper approximation procedure* presented here in the form that produces both the BF f and the corresponding BPA m_f (it is a trivial application of Formulas (2) and (3) to show that the procedure produces a consistent pair of functions).

Upper Approximation Procedure
For $k = 1, ..., |\Omega|$
For $a \subseteq \Omega : |a| = k$
$f(a) := \max\left[g(a),\ \sum_{b \subsetneq a} m_f(b)\right]$;
$m_f(a) := \max\left[0,\ g(a) - \sum_{b \subsetneq a} m_f(b)\right]$;
If $f(\Omega) > 1$ **Then Fail;**

If the procedure does not fail, then

- f is a BF;
- $f \geq g$;
- f is the lowest element of $\overline{\mathcal{B}}(g)$.

The first two properties are obvious from the procedure. The last one is proved in the following assertions.

Lemma 3 *Let BF f be an output of the Upper approximation procedure applied to PBF g (it means the procedure does not fail). Then there does not exist $\hat{f} \in \overline{\mathcal{B}}(g)$ for which $f > \hat{f}$.*

Proof. Assume the opposite. Let $\hat{f} \in \overline{\mathcal{B}}(g)$, and $f > \hat{f}$. By Lemma 2, there exists in $\overline{\mathcal{B}}(g)$ a BF for which f is its simple specification. Thus, without loss of generality, we can assume that f is a simple specification of $\hat{f}$. It means that there exists $a \subseteq \Omega$, for which $f(a) > \hat{f}(a)$. It follows from the definition of simple specification that for all $b \subseteq \Omega$, $|b| < |a|$, BFs f and $\hat{f}$ coincide, i.e., for all $b \subsetneq a$, $f(b) = \hat{f}(b)$ and $m_f(b) = m_{\hat{f}}(b)$. The latter equality yields that

$$\hat{f}(a) \geq \sum_{b \subsetneq a} m_f(b).$$

Since $\hat{f} \in \overline{\mathcal{B}}(g)$, we also know that $\hat{f}(a) \geq g(a)$. It means that

$$\hat{f}(a) \geq \max\left[g(a); \textstyle\sum_{b \subsetneq a} m_f(b)\right] = f(a),$$

which is in contradiction with our assumption that $f(a) > \hat{f}(a)$. □

In other words, Lemma 3 states that Upper approximation procedure finds a solution that is not dominated by another BF. The following theorem states the solution is unique.

Theorem 4 *Let BF f be an output of the Upper approximation procedure applied to PBF g (it means the procedure does not fail). Then $\overline{\mathcal{B}}(f) = \overline{\mathcal{B}}(g)$.*

Proof. Since $f \in \overline{\mathcal{B}}(g)$, it is obvious that $\overline{\mathcal{B}}(f) \subseteq \overline{\mathcal{B}}(g)$. Therefore, assuming the opposite, there must be $\overline{f} \in \overline{\mathcal{B}}(g) \setminus \overline{\mathcal{B}}(f)$ for which there exists at least one $a \subseteq \Omega$ such that $\overline{f}(a) < f(a)$. Consider such a with the smallest cardinality, which means that for all $b \subsetneq a$, $f(b) \geq \overline{f}(b)$. This choice guarantees that $\hat{f} : 2^\Omega \to [0,1]$, $\hat{f}(a) = \overline{f}(a)$, and $\hat{f}(c) = f(c)$ for all remaining $c \in 2^\Omega \setminus \{a\}$ is correctly defined BF (it is monotonic). However, $\hat{f} < f$, which contradicts Lemma 3. □

Thus, the set of optimal upper BF-approximations of g

$$\overline{\mathcal{A}}(g) = \left\{f \in \overline{\mathcal{B}} : \overline{\mathcal{B}}(g) \cap \underline{\mathcal{B}}(f) = \{f\}\right\}$$

is a one-point set. It means that if the procedure fails, then $\overline{\mathcal{B}}(g) = \emptyset$. There cannot exist BF $\hat{f}$, for which $\sum_{b \subsetneq a} m_{\hat{f}}(b) > g(\Omega) = 1$.

Cybersecurity data example - part I

Table 1: Features (variables) characterizing organizations

Notation	Variable	# of values
L	Legislation (regulated/unregulated by law)	2
S	Security knowledge status of decision makers	5
T	Total security score of organization	5
U	User experience	4
V	Volume of resources invested in cybersecurity	5
W	Willingness of CIOs to educate themselves	5

In the current section, we illustrate the above-presented ideas concerning the data-based learning of BF models using the *cybersecurity* data collected by Švadlenka (2025) (for a more detailed description, see his cited PhD thesis). The data describe six characteristics (variables – see Table 1) of fifty-two organizations (records).

For the sake of simplicity, we start with the simplest possible case, considering only two dichotomous variables. We consider variables L and W, the latter binarized as shown in Table 2. Thus, we consider $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, and the available data, which are summarized in the 2×2 contingency table (Table 2).

Table 2: 2×2 contingency table

	$W =\leq 3$	$W > 3$
$L = 1$	$\{\omega_1\}$	$\{\omega_2\}$
$L = 2$	$\{\omega_3\}$	$\{\omega_4\}$

	$W =\leq 3$	$W > 3$
$L = 1$	14	17
$L = 2$	15	6

The learning process is based on the idea that a belief function of an event a is a lower bound for the possible probability of that event. The lower bound of a binomial confidence interval has a similar property. Namely, there is only a small chance that an actual probability of event a is less than the lower bound of a confidence interval computed from the given data. As the title of this article suggests, we are considering Jeffreys confidence intervals. We started with $\alpha = 0.05$ (Lee, 1989), although other levels and other estimates of confidence intervals can (and will) be used. The lower bounds of these intervals for $n = 52$ are tabulated in Table 3. The values are calculated using the R command (Crawley, 2012)

```
for(x in 0:52){print(c(x,qbeta(0.05/2, x+0.5, 52-x+0.5)))}.
```

Using this approach, we get a function $g : 2^\Omega \to [0, 1]$ that is monotone. If we also set its value for Ω to one, we get a PBF that fits the given data. This function is tabulated

Table 3: Lower limits of Jeffreys confidence intervals with $\alpha = 0.05$, and $n = 52$

	$0*$	$1*$	$2*$	$3*$	$4*$	$5*$
$*0$	0	0.1033921	0.2615272	0.4415067	0.6422402	0.9136317
$*1$	0.0020829	0.1179209	0.2786434	0.460575	0.6637672	0.9136317
$*2$	0.0080773	0.1327931	0.2959618	0.4798459	0.6856412	0.9530633
$*3$	0.0165147	0.1479794	0.3134786	0.4993247	0.7079003	
$*4$	0.0265221	0.1634559	0.3311906	0.5190176	0.7305931	
$*5$	0.0376467	0.1792032	0.3490958	0.5389325	0.7537837	
$*6$	0.0496248	0.1952054	0.367193	0.5590787	0.7775591	
$*7$	0.06228799	0.2114491	0.3854819	0.5794675	0.8020423	
$*8$	0.0755216	0.2279231	0.4039627	0.6001125	0.8274177	
$*9$	0.08924337	0.2446184	0.422637	0.6210301	0.8539855	

in Table 4. Since it is not a BF, we let the Upper approximation process modify it. The whole process is recorded in the table. Looking at the resulting function f, we see that it is not a BF, it is not normalized. This means that, due to Theorem 4, no BF dominates g. However, this does not mean that there is no way to find a BF corresponding to the given data with the properties of the upper approximation. Two ways to modify the function g are presented here, both related to the reliability of the process by which g was constructed. The alternative way — using a generalization of Shafer's discounting — is analyzed in our next contribution in these proceedings Daniel et al. (2025).

Table 4: Application of the Upper approximation process with $\alpha = 0.05$

a	# occurrences	$g(a)$	$\sum_{b \subsetneq a} m_f(b)$	$f(a)$	$m_f(a)$
$\{\omega_1\}$	14	0.163	0	0.163	0.163
$\{\omega_2\}$	17	0.211	0	0.211	0.211
$\{\omega_3\}$	15	0.179	0	0.179	0.179
$\{\omega_4\}$	6	0.050	0	0.050	0.050
$\{\omega_1, \omega_2\}$	31	0.461	0.375	0.461	0.086
$\{\omega_1, \omega_3\}$	29	0.423	0.343	0.423	0.080
$\{\omega_1, \omega_4\}$	20	0.262	0.213	0.262	0.048
$\{\omega_2, \omega_3\}$	32	0.480	0.391	0.480	0.089
$\{\omega_2, \omega_4\}$	23	0.313	0.261	0.313	0.052
$\{\omega_3, \omega_4\}$	21	0.279	0.229	0.279	0.050
$\{\omega_1, \omega_2, \omega_3\}$	46	0.778	0.809	0.809	0
$\{\omega_1, \omega_2, \omega_4\}$	37	0.579	0.611	0.611	0
$\{\omega_1, \omega_3, \omega_4\}$	35	0.539	0.571	0.571	0
$\{\omega_2, \omega_3, \omega_4\}$	38	0.600	0.632	0.632	0
Ω	52	1.000	1.009	1.009	0

We know that with $\alpha = 0.05$ it can happen that $\pi(a) < g(a)$ with probability 0.025.

Therefore, the probability that all 15 probabilities considered are greater than the respective values of the function g is only 0.975^{15}, which is less than 0.7. To deal with such unreliability, we can either decrease α, or apply the idea of Shafer (1976) to *discount* PBF g. The latter approach is to follow the idea of Shafer (1976) proposed for inaccurate sources: take a coefficient of *discount rate* δ and recompute values of g for all proper subsets $a \subsetneq \Omega$ as $g(a) := (1-\delta)g(a)$.

In the following, we consider the first case and lower the level of the confidence intervals used. First, we computed the function g based on the estimates of the Jeffreys confidence intervals with $\alpha = 0.03$. It turned out that this PBF did not have a non-empty credal set either. We succeeded in obtaining a suitable PBF by considering the estimates of Jeffreys confidence intervals with $\alpha = 0.02$. The corresponding PBF g and a log of the corresponding calculation are presented in Table 5. It is worth noting that a very similar result can be obtained by applying the same algorithm to the original g presented in Table 4, discounted at the discount rate of $\delta = 0.9$.

Table 5: Process of computation of the Upper approximation; $\alpha = 0.02$

a	# occurrences	$g(a)$	$\sum_{b \subsetneq a} m_f(a)$	$f(a)$	$m_f(a)$
$\{\omega_1\}$	14	0.147	0	0.147	0.147
$\{\omega_2\}$	17	0.192	0	0.192	0.192
$\{\omega_3\}$	15	0.162	0	0.162	0.162
$\{\omega_4\}$	6	0.041	0	0.041	0.041
$\{\omega_1, \omega_2\}$	31	0.436	0.339	0.436	0.097
$\{\omega_1, \omega_3\}$	29	0.398	0.308	0.398	0.090
$\{\omega_1, \omega_4\}$	20	0.241	0.187	0.241	0.053
$\{\omega_2, \omega_3\}$	32	0.455	0.354	0.455	0.101
$\{\omega_2, \omega_4\}$	23	0.291	0.233	0.291	0.058
$\{\omega_3, \omega_4\}$	21	0.257	0.202	0.257	0.055
$\{\omega_1, \omega_2, \omega_3\}$	46	0.755	0.788	0.788	0
$\{\omega_1, \omega_2, \omega_4\}$	37	0.554	0.587	0.587	0
$\{\omega_1, \omega_3, \omega_4\}$	35	0.513	0.547	0.547	0
$\{\omega_2, \omega_3, \omega_4\}$	38	0.575	0.608	0.608	0
Ω	52	1.000	0.994	1.000	0.006

It is perhaps unnecessary to say that decreasing the level of confidence α (or discount rate δ) finally always results in finding a PBF g with a nonempty credal set.

In this example, we were interested in finding a BF on Ω, whose cardinality was 4. To have a simple example, we had to binarize variable W. When considering non-simplified two-dimensional contingency tables of the considered cybersecurity example, the cardinality of the considered space of discernment increases up to 25. Not to speak of considering three-dimensional contingency tables, where the cardinality of Ω can grow up to 125. Thus, we easily go beyond the capacities of current computers, and the open question is whether there are similar approaches that take into account the need to keep the complexity of the resulting BPs reasonable.

4 Lower Approximations of Pseudo-Belief Functions

As mentioned above, the lower approximations should be preferred if one does not want the approximation to contain any information not contained in PBF g. Thus, in this case, we consider the set of optimal outer approximations

$$\underline{\mathcal{A}}(g) = \left\{ f \in \underline{\mathcal{B}} : \underline{\mathcal{B}}(g) \cap \overline{\mathcal{B}}(f) = \{f\} \right\}.$$

It is a set of Pareto optimal outer approximations of g that are not dominated by any other outer approximation of g. The problem remains which one to choose and how to compute it. As in many other areas of research where one has to choose one solution from a set of Pareto optimal solutions, it depends on whether there is some additional information or some supporting criterion to take into account. This is all subject to further research.

A relatively simple way to find a lower approximation is to apply the following Easy lower approximation procedure. In contrast to the Upper approximation procedure introduced in Section 3, where the procedure was unambiguously described, the following pseudo-code uses a step that can be implemented in several different ways.

Easy Lower Approximation Procedure
For $k = 1, ..., |\Omega|$
While $A := \left\{ a \subseteq \Omega : |a| = k \;\; \& \;\; g(a) < \sum_{b \subsetneq a} m_f(b) \right\} \neq \emptyset$
Choose any $a \in A$; $\varepsilon := g(a) - \sum_{b \subsetneq a} m_f(b)$;
Redistribute ε to $m_f(b_\ell), b_\ell \subsetneq a, \quad \ell = 1, \ldots, r$;
For $a \subseteq \Omega : |a| = k$
$m_f(a) := g(a) - \sum_{b \subsetneq a} m_f(b)$;
For $a \subseteq \Omega$
$f(a) := \sum_{b \subseteq a} m_f(b)$;

Recall that the step "*Redistribute* ε to $m_f(b_\ell), b_\ell \subsetneq a, \ell = 1, \ldots, r$;" means that you have to choose the system $b_\ell \subsetneq a, \ell = 1, \ldots, r$, and split ε into corresponding ε_ℓ so that all ε_ℓ are negative and $\varepsilon = \sum_{\ell=1}^{r} \varepsilon_\ell$. No matter how it is implemented, the process of redistributing a negative value ε to m_f always realizes several simple generalizations we discussed in Section 2. Thus, when the redistribution is finished, $g(a) = \sum_{b \subsetneq a} m_f(b)$. It can always be realized in such a way that all $m_f(b_\ell)$ are non-negative. In the example below, we will implement this step so that a third of ε is added to all $b \subset a : |b| = |a| - 1$. We take a third of ε because, in the case of the following example, it is applied when $|\{b \subset a : |b| = |a| - 1\}| = 3$.

As said, the discussed process of redistribution can always be done, because $\sum_{b \subsetneq a} m_f(b) > |\varepsilon|$. This can be implemented in many ways. It is also a topic for further research to study which of them is preferable. Note that if possible, one should redistribute ε to $m_f(b)$, for $b \subset a : |b| = |a| - 1$, and not to $m_f(b)$, for $b \subsetneq a$ with $|b| < |a| - 1$. This is because the simple generalization when ε_i is subtracted from $m_f(a)$ and added to $m_f(b)$ with $|b| = |a| - 2$ can be realized as two successive simple generalizations, first

from a to c (for $b \subset c \subset a$), and second from c to b, which would give a hint that the resulting BF does not belong to the Pareto optimal outer approximations.

There are even more open questions regarding the redistribution step. Although it is quite likely, we are not sure whether one can always redistribute ε only to sets b, for $b \subset a : |b| = |a| - 1$. However, the most important open question is whether there is an implementation that guarantees that the resulting approximation is Pareto optimal. Greater chances of producing Pareto optimal solutions have Advanced lower approximation procedure, which is described after an example in the following section. We will see that in the advanced procedure, the redistribution process is considered simultaneously for all subsets of the same cardinality, rather than separately in a cycle. However, this increases the computational complexity of the whole process.

Cybersecurity data example - part II

Let us get back to considering the cybersecurity data and the function g defined in Table 4, where its values correspond to the Jeffreys estimates of lower bounds of confidence intervals with $\alpha = 0.05$.

Table 6: Application of the Easy lower approximation process with $\alpha = 0.05$

		$k=1,2$	$k=3$									
a	$g(a)$	$m_f(a)$	ε	$m_f(a)$	ε	$m_f(a)$	ε	$m_f(a)$	ε	$m_f(a)$	$f(a)$	$f^\star(a)$
$\{\omega_1\}$	0.163	0.163		0.163		0.163		0.163		0.163	0.163	0.163
$\{\omega_2\}$	0.211	0.211		0.211		0.211		0.211		0.211	0.211	0.211
$\{\omega_3\}$	0.179	0.179		0.179		0.179		0.179		0.179	0.179	0.179
$\{\omega_4\}$	0.050	0.050		0.050		0.050		0.050		0.050	0.050	0.050
$\{\omega_1,\omega_2\}$	0.461	0.086		0.075		0.068		0.068		0.068	0.443	0.461
$\{\omega_1,\omega_3\}$	0.423	0.080		0.070		0.070		0.065		0.065	0.407	0.407
$\{\omega_1,\omega_4\}$	0.262	0.048		0.048		0.041		0.037		0.037	0.250	0.246
$\{\omega_2,\omega_3\}$	0.480	0.089		0.079		0.079		0.079		0.076	0.466	0.464
$\{\omega_2,\omega_4\}$	0.313	0.052		0.052		0.045		0.045		0.042	0.303	0.298
$\{\omega_3,\omega_4\}$	0.279	0.050		0.050		0.050		0.045		0.042	0.271	0.279
$\{\omega_1,\omega_2,\omega_3\}$	0.778		-0.031	0		0.007		0.012		0.015	0.778	0.778
$\{\omega_1,\omega_2,\omega_4\}$	0.579				-0.021	0		0.005		0.008	0.579	0.579
$\{\omega_1,\omega_3,\omega_4\}$	0.539						-0.014	0		0.003	0.539	0.539
$\{\omega_2,\omega_3,\omega_4\}$	0.600								-0.009	0	0.600	0.600
Ω	1									0.041	1	1

When applied to this function g, Easy lower approximation procedure skips the While cycle for $k = 1, 2$, because for these k, set $A = \left\{a \subseteq \Omega : |a| = k \ \ \& \ \ g(a) < \sum_{b \subsetneq a} m_f(b)\right\}$

is empty. A is nonempty only for $k = 3$. The whole calculation is shown in Table 6. The values of the resulting BF are in the column headed by f. We do not know if f is Pareto optimal or not. We do not know it even for the solution $f^\star$ in the last column, which was computed by the procedure described in the following section. The reader certainly noticed that two solutions f and $f^\star$ are incomparable.

5 Advanced Lower Approximations

The idea of this procedure, here called Advanced lower approximation, is based on the behavior of the Easy lower approximation procedure. The reader can see it in the example presented above. After applying the redistribution step to $a = \{\omega_1, \omega_2, \omega_3\} \in A$ (for $k = 3$) we get $g(a) = \sum_{b \subsetneq a} m_f(b)$, and therefore $m_f(a) = 0$. But in the next step of this cycle, the redistribution is applied to $\bar{a} = \{\omega_1, \omega_2, \omega_4\} \in A$. When this redistribution process is finished, we get, analogously, $g(\bar{a}) = \sum_{b \subsetneq \bar{a}} m_f(b)$, but for the preceding a we get $g(a) > \sum_{b \subsetneq a} m_f(b)$, and therefore $m_f(a) > 0$. Therefore, in the procedure we are going to describe, we leave the idea of redistributing negative ε's in a cycle, we want to redistribute all values simultaneously to get $m_f(a) = 0$ for all $a \in A$.

Perhaps the first idea could be to set up a system to be solved by linear programming methods. Unknown variables are all potential shifts of mass functions. Assume a fixed k, and that we only want to move masses to subsets whose cardinality is one less than the considered k. Denote $\ell_k = |\{b \subset a : |b|+1 = |a| = k\}|$. Then we get $\ell_k.|A_k|$ unknown variables. The constraints are given by the equality that for all $a \in A_k$, $g(a) = \sum_{b \subsetneq a} m_f(b)$, and the sum of all shifts from the set a must equal ε_a, and that all new values of $m_f(b)$ for all b are nonnegative. Such an approach is possible, but we believe it is unnecessarily computationally expensive. Much simpler is the following iterative process.

Advanced Lower Approximation Procedure
For $k = 1, ..., |\Omega|$
Set $A_k = \left\{ a \subseteq \Omega : |a| = k \;\; \& \;\; g(a) < \sum_{b \subsetneq a} m_f(b) \right\} \neq \emptyset$;
If $A_k \neq \emptyset$ **Then**
Untill $\max[|\varepsilon_a|] < 10^{-8}$
For $a \in A_k \;\; \varepsilon_a := g(a) - \sum_{b \subsetneq a} m_f(b)$;
Choose $a \in A_k : \max[|\varepsilon_a|]$;
Redistribute ε_a to $m_f(b_\ell), b_\ell \subsetneq a, \ell = 1, \ldots, r$;
For $a \subseteq \Omega : |a| = k$
$m_f(a) := g(a) - \sum_{b \subsetneq a} m_f(b)$;
For $a \subseteq \Omega$
$f(a) := \sum_{b \subseteq a} m_f(b)$;

Let us add two comments to the iterative process introduced above (realized in the Untill cycle). First, the criterion that ends the cycle means that you are willing to accept that all differences are so small that they can be considered zero. Second, realize that

the values of ε_a reach both negative and positive values. When splitting this value for Redistribution, all of its parts must be of the same sign. This is also why we choose a according to the absolute value of ε_a.

This procedure yielded the BF $f^\star$ from Table 6 Let us also remark that, up to now, we have always got along with the simplest possible realization of the Redistribution step by one simple cycle:

For $b \subset a : |b| + 1 = |a|$

$$m_f(b) := m_f(b) + \frac{\varepsilon_a}{|\{b \subset a : |b| + 1 = |a|\}|}.$$

6 Conclusion

The paper introduces two types of approximations of PBFs by BFs. The research is motivated by the idea that data-based learning of belief function models can start with a normalized monotone set function whose values are defined by lower bounds of the corresponding confidence intervals. Although the paper presents simple algorithms for obtaining both lower and upper bounds of PBFs, it raises more questions than it answers. The main questions concern the optimality of the proposed lower approximations. There is also more room for computational experiments and the design of heuristic algorithms, because, as it is quite natural in the framework of belief functions, all computational processes are of very high computational complexity. Thus, the paper provides a good basis for future research in several directions.

References

M. J. Crawley. *The R book*. John Wiley & Sons, 2012.

M. Daniel, R. Jiroušek, and V. Kratochvíl. Discounting or optimizing? different approaches to pseudo-belief function correction. In *Proceedings of the 13th Workshop on Uncertainty Processing (WUPES'25)*, pages 104–115, 2025.

D. Dubois and H. Prade. On several representations of an uncertain body of evidence. In *IFAC Symposium on Theory and Application of Digital Control (IFAC 1982)*, pages 167–181. North Holland, 1982.

R. Jiroušek and V. Kratochvíl. About twenty-five naughty entropies in belief function theory: do they measure informativeness? *International Journal of Approximate Reasoning*, page 109454, 2025. doi: 10.1016/j.ijar.2025.109454.

P. M. Lee. *Bayesian statistics*. Oxford University Press London:, 1989.

G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976. ISBN 9780691100425.

R. Švadlenka. *Analýza implementace bezpečnostních opatření pod tíhou legislativních změn*. PhD thesis, Prague University of Economics and Business, 2025.