Learning Belief Functions from Data via Polyhedral Methods

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

[1]ÚTIA, Czech Academy of Sciences, Czechia
[2]Faculty of Management, Prague University of Economics and Business, Czechia

Abstract

We present a polyhedral framework for learning belief functions from data when empirical lower and upper probability bounds are obtained from Jeffreys' binomial confidence intervals. Such bounds, interpreted as empirical belief and plausibility values for all subsets of the outcome space, generally yield a *pseudo-belief function* that may not correspond to any valid basic probability assignment (BPA) satisfying the axioms of Dempster–Shafer theory.

Our approach formulates the correction problem as a system of linear constraints in the BPA space, where the feasible solutions form a convex polyhedron of belief functions consistent with the empirical bounds. We investigate several linear optimization criteria for selecting a representative BPA from this feasible set, including L_1-projection to the empirical lower bounds, Dubois–Prade entropy maximization, sparsity-oriented objectives, and cardinality-weighted allocations.

1 Introduction

Belief function theory, developed within the Dempster–Shafer framework, offers a flexible and expressive way of modeling uncertainty. Unlike probability theory, it allows us to represent partial, imprecise, or even conflicting information through belief and plausibility functions derived from basic probability assignments (BPAs). For example, it can assign probability mass to a set of outcomes rather than a single one, reflecting genuine ignorance.

However, learning belief functions from empirical data is challenging. Due to their combinatorial nature, belief functions lack simple statistical estimators, and naive frequency-based approaches often yield invalid or incoherent results. In particular, lower and upper estimates of event probabilities — such as those derived from confidence intervals — may not correspond to any valid belief function because they can violate monotonicity or fail to sum to a consistent BPA. We refer to such inconsistent estimates as *pseudo-belief functions*.

Several recent works address this issue from different perspectives. For example, Bordini et al. [2023] learn calibrated belief functions by regressing p-values from conformal prediction outputs, enabling direct generation of belief functions from data. Geng et al. [2024] propose an interpretable, incremental learning approach for belief rule models from labeled data, which is complementary to our polyhedral framework. On the methodological side, Zhang et al. [2023] introduce a belief correlation measure combining belief entropy and relative entropy, providing tools for comparing belief functions. Furthermore, in application domains such as medical image segmentation, Huang et al. [2022] provide a comprehensive review of data-driven belief function methods. However, none of these works explicitly address the problem of transforming statistically estimated bounds into a valid belief function while preserving as much of the original information as possible.

In this paper, we address the following specific problem: given empirical lower and upper bounds on event probabilities, find a valid belief function that is as close as possible to these bounds. Our primary source of bounds is the *Jeffreys' confidence intervals*, which offer a principled Bayesian estimation for binomial proportions under a noninformative prior. When computed for all subsets of the outcome space, the lower bound of the Jeffreys interval can be interpreted as an empirical estimate of belief value, and the upper bound as plausibility value. The resulting set function is, in general, only a pseudo-belief function.

Our contribution is to formulate the correction of such pseudo-belief functions as a linear optimization problem entirely in the space of BPAs. This has two main advantages: (i) the BPA constraints are directly enforced through non-negativity and normalization, and (ii) any additional modeling preference can be incorporated as a linear objective function. We explore several such objectives, including sparsity, cardinality-weighted allocation, and entropy maximization. We also introduce projection-based criteria, which directly minimize a chosen distance between the candidate belief function and the empirical pseudo-belief function derived from data.

The proposed approach is implemented using standard linear programming techniques and is applicable to moderate-sized frames of discernment. We provide only simple experiments on synthetic and empirical datasets to demonstrate how different optimization criteria influence the resulting belief structure.

This paper naturally continues the line of research initiated by Daniel et al. [2025b] and further developed in Daniel et al. [2025a].

2 Notation and Preliminaries

Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ be a finite set of mutually exclusive and exhaustive outcomes, called the *frame of discernment*. Let 2^Ω denote its power set.

A *basic probability assignment* (BPA) is a function $m : 2^\Omega \to [0, 1]$ satisfying:

$$m(\emptyset) = 0, \qquad \sum_{A \subseteq \Omega} m(A) = 1. \tag{1}$$

The subsets $A \subseteq \Omega$ for which $m(A) > 0$ are called *focal elements*.

From m, we derive the *belief function*:

$$\mathrm{bel}_m(A) = \sum_{B \subseteq A} m(B), \quad A \subseteq \Omega, \tag{2}$$

and the *plausibility function*:

$$\mathrm{pl}_m(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad A \subseteq \Omega. \tag{3}$$

These functions define for each subset A an interval $[\mathrm{bel}_m(A), \mathrm{pl}_m(A)]$ that bounds the probability of A under incomplete or imprecise information. Any probability distribution $\mathbb{P}$ that satisfies

$$\mathrm{bel}_m(A) \leq \mathbb{P}(A) \leq \mathrm{pl}_m(A), \quad \text{for all } A \subseteq \Omega,$$

is said to be consistent with the belief function m. The set of all such probability distributions is called the *credal set* induced by m. It is a closed, convex, and bounded subset of the probability simplex.

From a geometric perspective, it is a polyhedron. In the following, we will work with polyhedrons in $\mathbb{R}^{2^n}$ - a space of BPAs - where each BPA m is just a point. Therefore, let us introduce a few notions from the area of polyhedral geometry.

Polyhedra and their representations. A *polyhedron* is the intersection of finitely many closed halfspaces in $\mathbb{R}^d$. Its bounded case is called a *polytope*. Every polytope can be described in two equivalent ways:

- **H-representation (halfspace representation):** The polytope is given as the bounded solution set of a system of linear inequalities:

  $$\mathcal{P} = \{x \in \mathbb{R}^d : Mx \leq b\},$$

  where $M \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$. In our setting, x represents the vector of BPA components $m(A)$, the matrix M encodes non-negativity, normalization, and empirical bounds (Jeffreys' intervals), and b is the corresponding right-hand side.

- **V-representation (vertex representation):** The polytope is the convex hull of a finite set of *extreme points* (vertices):

  $$\mathcal{P} = \text{conv}\{v^{(1)}, v^{(2)}, \ldots, v^{(k)}\}.$$

  Each vertex is a point that satisfies all constraints and cannot be expressed as a nontrivial convex combination of other feasible BPAs.

The two representations are equivalent: starting from the H-representation, the V-representation can be obtained via a vertex enumeration algorithm (e.g., the double description method), and vice versa via convex hull algorithms. However, conversion between them can be computationally expensive, especially for high-dimensional problems such as belief functions over large frames.

In this work, we will deal with H-representation, because, as we will show later, our constraints can be naturally expressed as linear inequalities. Optimization over the polytope is then carried out directly in this representation using standard linear programming methods.

For background on belief function theory, see Shafer [1976]; for polyhedral theory, see Ziegler [1995], Grünbaum [2003]. The geometric analysis of belief functions is further explored in Cuzzolin [2020], which motivates our use of polyhedral techniques throughout this paper.

3 Polyhedral Construction in the BPA Space

We work in the space of *basic probability assignments* (BPAs), where each belief function is represented by a nonnegative vector

$$m : 2^\Omega \to [0, 1], \quad m(\emptyset) = 0, \quad \sum_{A \subseteq \Omega} m(A) = 1.$$

Although this space has exponential dimension in $|\Omega|$, it allows us to express belief and plausibility constraints as *linear inequalities* via the Möbius transform.

We interpret *Jeffreys' confidence intervals*

$$[\underline{p}(A), \overline{p}(A)]$$

as empirical bounds for $\text{bel}(A)$ and $\text{pl}(A)$ on all subsets $A \subseteq \Omega$. Our aim is to find a BPA m such that

$$\underline{p}(A) \leq \text{bel}_m(A), \qquad \text{pl}_m(A) \leq \overline{p}(A), \qquad \forall A \subseteq \Omega,$$

together with normalization and non-negativity laid on BPAs.

Using the definitions from equations (2) and (3), each of these bounds becomes a linear inequality in the components of m. Together with the normalization condition (1) and the non-negativity constraints, they form a polytope (it is bounded e.g. by a 0-1 hypercube)in $\mathbb{R}^{2^n}$. We denote this feasible set as:

$$\mathcal{M}_{\text{feas}} = \left\{ m \in \mathbb{R}^{2^n} \;\middle|\; \begin{array}{ll} m(A) \geq 0 \quad \forall A \subseteq \Omega, & \text{non-negativity} \\ \sum_{A \subseteq \Omega} m(A) = 1, & \text{normalization} \\ \underline{p}(A) \leq bel_m(A) = \sum_{B \subseteq A} m(B) \quad \forall A \subseteq \Omega, & \text{lower-bounds on beliefs} \\ \overline{p}(A) \geq pl_m(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad \forall A \subseteq \Omega, & \text{upper-bounds on plausibility} \end{array} \right\}. \tag{4}$$

This polyhedron represents all valid belief functions that remain consistent with the Jeffreys' intervals. Our task is then to explore this polyhedron and extract representative solutions.

Example 3.1. *Consider the binary frame* $\Omega = \{0, 1\}$ *with ten measurements: 6 times outcome* 0*, 3 times outcome* 1*, and 1 ambiguous observation. For Jeffreys' intervals, we discard the ambiguous case, so* $n = 9$ *observations remain. (Note that the ambiguous observation can be treated differently, based on the problem context.)*

The 95% *Jeffreys' lower/upper bounds for the relevant subsets are:*

$$\begin{aligned} \underline{p}(\{0\}) &= 0.35, & \overline{p}(\{0\}) &= 0.90, \\ \underline{p}(\{1\}) &= 0.10, & \overline{p}(\{1\}) &= 0.65, \\ \underline{p}(\{0,1\}) &= 0.76, & \overline{p}(\{0,1\}) &= 1.00. \end{aligned}$$

Let $x = m(\{0\})$*,* $y = m(\{1\})$*,* $z = m(\{0,1\})$*. Let us have a basic polytope of all BPAs in* $\mathbb{R}^3$ *with dimensions* $\{x, y, z\}$*:*

$$\begin{aligned} &\textit{(Normalization)} && x + y + z = 1, \\ &\textit{(Non-negativity)} && x, y, z \geq 0. \end{aligned}$$

From the Möbius definitions:

$$\begin{aligned} \text{bel}(\{0\}) &= x, \\ \text{bel}(\{1\}) &= y, \\ \text{bel}(\{0,1\}) &= x + y + z, \\ \text{pl}(\{0\}) &= x + z, \\ \text{pl}(\{1\}) &= y + z. \end{aligned}$$

Adding the Jeffreys' constraints gives[1]*:*

$$\begin{aligned} &\textit{(Lower bounds on belief)} && x \geq 0.35, \\ & && y \geq 0.10, \\ &\textit{(Upper bounds on plausibility)} && x + z \leq 0.90, \\ & && y + z \leq 0.65. \end{aligned}$$

[1]Note that we do *not* include inequalities such as $x + z \geq p(\{0,1\}) = 0.76$ for the full set $\{0,1\}$. This is because $\text{bel}(\{0,1\}) = x + y + z = 1$ by normalization, so the Jeffreys' lower bound 0.76 is automatically satisfied. Similarly, no upper bound for $\text{pl}(\{0,1\})$ is needed, as $\text{pl}(\{0,1\}) = 1$ for any BPA.

This defines a 2-dimensional polytope in $\mathbb{R}^3$, lying on the plane $x+y+z=1$, and constrained by the empirical bounds. The resulting feasible region represents all BPAs that are consistent with the observed data and respect the structure of belief functions. In the following section, we will analyze this polytope and explore candidate solutions based on various optimization criteria.

Within this feasible polytope, we can choose a representative BPA according to various *linear objectives*—e.g., zero objective (arbitrary vertex), cardinality-weighted allocation, Dubois–Prade entropy maximization, or relaxed sparsity. The LP formulation remains as follows, where $\mathcal{M}_{\text{feas}}$ is the polyhedron of feasible BPAs defined in (4), and c is the vector of some constants:

$$\min_{m \in \mathcal{M}_{\text{feas}}} \quad c^\top m$$

Why not stay in the probability space? Since the already mentioned concept of credal set is a polyhedron in a much simpler space (non-exponentional), one might consider remaining in this probability space and looking for a credal set directly as a polyhedron of probability distributions $\mathbb{P}$ satisfying the empirical Jeffreys' bounds - i.e. the biggest polyhedron that is (i) a credal set (a corresponding BPA exists), and, (ii) satisfying Jeffreys' bounds:

$$\underline{p}(A) \leq \mathbb{P}(A) \leq \overline{p}(A), \quad \forall A \subseteq \Omega.$$

However, to guarantee that such a set of probabilities corresponds to some valid belief function, one would still need to ensure the existence of a BPA m such that $\mathbb{P}$ lies inside the credal set induced by m. This requirement is equivalent to enforcing all belief function axioms, which involve an exponential number of constraints in $|\Omega|$. In practice, this means that the computational complexity remains essentially the same as in the BPA formulation, while the probability-space representation loses the convenient linearity of the Möbius transform.

Why not stay in the belief function bel **space?** One might consider working directly in the space of belief functions bel, since Jeffreys-type lower bounds could be imposed as simple inequalities $\text{bel}(A) \geq \underline{p}(A)$ and $\text{bel}(\Omega \setminus A) \geq 1 - \overline{p}(A)$. However, not every set function satisfying such bounds corresponds to a valid belief function. Validity requires that bel be an ∞-*monotone capacity* (totally monotone), meaning it satisfies the whole hierarchy of inclusion–exclusion inequalities for all $k \geq 2$ and all $A_1, \ldots, A_k \subseteq \Omega$:

$$\sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} \, \text{bel}\left(\bigcap_{i \in I} A_i \right) \geq 0,$$

which means nothing else, than the respective inverse Möbius transform to BPA produces non-negative values. Supermodularity (the case $k = 2$) is just one of these necessary conditions, but higher-order constraints are also required. Enforcing full ∞-monotonicity in the belief space is combinatorially as demanding as the BPA formulation, yet without the advantage that BPA validity can be guaranteed by simple non-negativity and normalization. For this reason, we prefer the BPA space, where both empirical bounds and validity constraints remain linear and are directly compatible with LP solvers.

4 Selecting the final BPA from the feasible polyhedron

The procedure described so far yields the *feasible polyhedron*

$$\mathcal{M}_{\text{feas}} \subset \mathbb{R}^{2^{|\Omega|}}$$

of all basic probability assignments (BPAs) m that are consistent with the given Jeffreys' intervals, i.e., those whose belief and plausibility functions satisfy the corresponding lower and upper bounds. While any $m \in \mathcal{M}_{\text{feas}}$ represents a valid belief function in agreement with the data, in practical applications we typically require a *single* BPA for further use. This raises the natural question: *Which BPA should we choose?*

Our approach is to formulate this selection as a linear optimization problem over $\mathcal{M}_{\text{feas}}$. The choice of optimization criterion determines which property of the BPA is emphasized. Because linear programming is used, the objective must be a *linear* functional of m. This still allows for a wide range of criteria, such as:

- **Weighted sum of focal element masses:** $\sum_{A\subseteq\Omega} w_A m(A)$, where the weights w_A can be chosen to prefer, for example, small focal elements (encouraging specificity) or large focal elements (encouraging generality).
- **Closeness of belief to the lower estimate $\underline{p}$:** One can search for a BPA whose belief function bel_m is as close as possible to $\underline{p}$ in the sense of a weighted L_1 distance:

$$\min_{m\in\mathcal{M}_{\text{feas}}} \sum_{A\subseteq\Omega} w_A \left|\text{bel}_m(A) - \underline{p}(A)\right|,$$

  which, after rewriting in terms of m, becomes a linear program. This criterion can be seen as finding a BPA that *best matches the lower bound* implied by the Jeffreys' intervals.

In the following subsections, we first focus on the L_1-based projection to $\underline{p}$, and then turn to other linear criteria such as the objectives that emphasize sparsity, weight concentration, or other structural properties of the focal elements.

4.1 L_1 projection to the Jeffreys' lower bound

We seek a BPA m that stays as close as possible to these bounds in the L_1 sense:

$$\min_m \sum_{A\subseteq\Omega} w_A \left|\text{bel}_m(A) - \underline{p}(A)\right| \quad \text{s.t. } m \in \mathcal{M}_{\text{feas}}.$$

In our case, the feasibility constraints already impose $\text{bel}_m(A) \geq \underline{p}(A)$ for all A, so the absolute value simplifies to

$$\text{bel}_m(A) - \underline{p}(A),$$

making the objective a purely linear function of m. This avoids the need for auxiliary variables normally required for L_1 norms.

Recall that

$$\text{bel}_m(A) = \sum_{B\subseteq A} m(B).$$

Substituting into the L_1 objective with feasibility $\text{bel}_m(A) \geq \underline{p}(A)$, we obtain:

$$\min_m \sum_{A\subseteq\Omega} w_A \left(\sum_{B\subseteq A} m(B) - \underline{p}(A)\right) \quad \text{s.t.} \quad m(B) \geq 0,\ \sum_{B\subseteq\Omega} m(B) = 1.$$

Rearranging the double sum yields:

$$\min_m \sum_{B\subseteq\Omega} c_B\, m(B) - \sum_{A\subseteq\Omega} w_A\, \underline{p}(A),$$

where

$$c_B = \sum_{A \supseteq B} w_A.$$

The constant term $\sum_{A \subseteq \Omega} w_A \underline{p}(A)$ depends only on the data $\underline{p}$ and the weights w_A, but not on m. It is therefore constant across all feasible m and can be safely omitted from the optimization problem without affecting the location of the optimum. The problem then reduces to a purely linear program

$$\min_m \sum_B c_B \, m(B) \quad \text{s.t.} \quad m \in \mathcal{M}_{\text{feas}}.$$

We test two simple weighting schemes:

- *Uniform:* $w_A \equiv 1$, giving equal importance to each subset A. The coefficient for $m(B)$ in the objective is then $c_B = 2^{n-|B|} - 1$ corresponding to the number of supersets of B. The result of the optimization process is then denoted as $m^{L_1}_{\text{unit}}$
- *Layered:* $w_A = \binom{n}{|A|}^{-1}$, which equalizes the *total* weight of all subsets of the same cardinality. For example, all singletons together have the same total weight as all doubletons, regardless of how many such sets exist. This prevents layers with many subsets (e.g., around $|A| \approx n/2$) from dominating the objective. The coefficient for $m(B)$ is then, using [wik, a,b]

$$c_B = \sum_{k=|B|}^{n} \binom{n}{k}^{-1} \binom{n-|B|}{k-|B|} = \frac{n+1}{|B|+1}.$$

The result of the optimization process is then denoted as $m^{L_1}_{\text{layered}} =$

4.2 Other linear and nonlinear objectives

In the previous subsection, we introduced an L_1-based criterion that selects a BPA whose induced belief function lies as close as possible to the lower Jeffreys' bound $\underline{p}$ in terms of a weighted absolute deviation. While this approach has the advantage of a clear interpretation and a natural link to the underlying statistical estimation, it is by no means the only plausible way to select a representative BPA from the feasible polytope $\mathcal{M}_{\text{feas}}$.

Depending on the modeling goals, it may be desirable to choose a BPA with different structural properties, such as favoring large or small focal elements, maximizing some notion of entropy, or encouraging sparsity (i.e. to keep the number of focal elements as low as possible). Below, we present four alternative criteria. Each is optimized subject to the constraints defining $\mathcal{M}_{\text{feas}}$, and except for the Dubois–Prade entropy, they are linear in $m(A)$ and thus directly compatible with linear programming.

- **Zero Objective (ZO)**:

$$\text{Minimize } f_{ZO}(m) = 0.$$

This baseline criterion imposes no preference over feasible BPAs. The LP solver will simply return an arbitrary vertex of $\mathcal{M}_{\text{feas}}$, determined by its internal pivoting rules.

- **Cardinality-Weighted Minimization (CW)**:

$$\text{Minimize } f_{CW}(m) = \sum_{\substack{A \subseteq \Omega \\ A \neq \emptyset}} \frac{1}{|A|} \, m(A).$$

This penalizes mass on small subsets and favors BPAs with larger, more general focal elements. The weight $1/|A|$ decreases monotonically with set size, so solutions are biased toward high-cardinality focal elements.

- **Dubois–Prade Entropy (HD)**:

$$\text{Maximize } H_D(m) = \sum_{\substack{A \subseteq \Omega \\ A \neq \emptyset}} m(A) \log |A|.$$

Introduced by Dubois and Prade [1987], this linear[2] utility promotes high imprecision by rewarding mass on large sets. Its maximum is reached when all mass is assigned to Ω.

- **Relaxed Sparsity (rSP)**:

Sparse belief functions—those assigning mass to only a small number of focal elements—are often desirable. In expert-driven applications, sparsity reflects the focus on a few meaningful scenarios; in data-driven settings, it aids interpretability and mitigates overfitting.

Directly minimizing the number of nonzero masses, assuming δ to be Dirac delta function[3]

$$f_{SP}(m) = \sum_{\substack{A \subseteq \Omega \\ A \neq \emptyset}} \delta[m(A) > 0],$$

is non-linear and non-convex, making it incompatible with standard LP solvers. To retain LP compatibility, we introduce for each A an auxiliary variable $z_A \in [0, 1]$ and enforce additional constraints:

$$m(A) \leq M \cdot z_A,$$

with M a large constant (e.g., $M = 1$). The variable z_A acts as an indicator proxy: it must be positive whenever $m(A)$ is positive. The sum $\sum z_A$ then approximates the number of nonzero focal elements, which we want to minimize. I.e.:

$$\text{Minimize} \quad f_{\text{rSP}}(m, z) = \sum_{\substack{A \subseteq \Omega \\ A \neq \emptyset}} z_A.$$

This objective approximately counts the number of active focal elements, encouraging solutions with as few non-zero $m(A)$ as possible while staying in a purely linear framework. The cost is the doubling of the number of decision variables, but for moderate frame sizes this remains tractable and produces interpretable, compact BPAs.

These alternative criteria offer different trade-offs between specificity and imprecision, and can be combined or compared with the L_1-Jeffreys approach to assess robustness of the resulting BPA.

Example 4.1. *We now solve the polyhedral correction problem from the previous example using all criteria introduced so far, including the L_1 variants. The resulting mass assignments are summarized in Table 1. Red columns display the Jeffreys' lower bounds $\underline{p}$ and the corresponding pseudo-BPA $m_{\underline{p}}$ obtained by Möbius inversion; note that $m_{\underline{p}}$* does not *satisfy normalization (it does not sum to 1), hence it is not a valid BPA.*

[2]Note that $log(|A|)$ is just another constant since A is known - it is a defining property of the variable corresponding to $m(A)$.

[3]$\delta(x) = 0$ if $x = 0$ and $\delta(x) = 1$ otherwise

Table 1: Corrected BPAs under different optimization criteria. Red columns show data-derived quantities: the Jeffreys' lower bound $\underline{p}$ and the pseudo-BPA $m_{\underline{p}}$.

	$\underline{p}$	$m_{\underline{p}}$	m_{ZO}	m_{CW}	m_{HD}	m_{rSP}	$m^{L_1}_{\mathrm{unit}}$	$m^{L_1}_{\mathrm{layered}}$
$\{w_1\}$	0.35	0.35	0.90	0.35	0.35	0.90	0.35	0.35
$\{w_2\}$	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
$\{w_1, w_2\}$	0.76	0.31	0.00	0.55	0.55	0.00	0.55	0.55

Here, $m_{\underline{p}}$ is the pseudo-BPA obtained by Möbius inversion of $\underline{p}$. Its components sum to $0.76 = 0.35 + 0.10 + 0.31 \neq 1$, *confirming that $\underline{p}$ is not itself a belief function but a data-driven lower estimate. All corrected BPAs m in the table satisfy* $\mathrm{bel}_m(A) \geq \underline{p}(A)$ *and* $\mathrm{pl}_m(A) \leq \overline{p}(A)$ *for each A.*

The distribution of the remaining belief mass (i.e., the slack above $\underline{p}$) differs across criteria. In this binary domain, ***ZO*** *and* ***rSP*** *coincide, and likewise* ***CW***, ***HD***, *and the L_1 solutions $m^{L_1}_{\mathrm{unit}}$, $m^{L_1}_{\mathrm{layered}}$ coincide. This is not a surprise if you realize that $\mathcal{M}_{feas}$ is a triangle in the plane $x+y+z=1$, i.e. there are 3 extreme vertices only.*

5 Experiment

We illustrate our approach using a fragment of empirical data [4] collected by Švadlenka [2025] , where the frame of discernment is

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}.$$

The dataset contains 52 observations in total, with the following frequencies: ω_1 appears 14 times, ω_2 17 times, ω_3 15 times, and ω_4 6 times.

From these empirical frequencies, we construct Jeffreys' confidence intervals for all nonempty subsets of Ω, yielding lower and upper probability bounds $[\underline{p}, \overline{p}]$. These serve as constraints in the polyhedral formulation described in the previous section.

All computations were performed in R [R Core Team, 2023], using the Linear Programming (LP) solver `lpSolve` [Berkelaar, 2024] and the `rcdd` package [Geyer et al., 2023]. The resulting polytope $\mathcal{M}_{\mathrm{feas}}$ is highly nontrivial: its V-representation contains 13,856 vertices, illustrating that the space of feasible BPAs is both large and structurally rich.

To extract representative BPAs, we applied all optimization criteria introduced earlier:

- **ZO** — zero objective,
- **CW** — cardinality-weighted allocation,
- **HD** — Dubois–Prade entropy maximization,
- **rSP** — relaxed sparsity,
- L_1 **projection** to the Jeffreys' lower bound (uniform and layered weights).

Table 2 compares the obtained BPAs with $\underline{p}$ — the empirical Jeffreys' lower bound, and $m_{\underline{p}}$ — the pseudo-BPA obtained by Möbius inversion of $\underline{p}$ assuming that $\underline{p}$ should be a regular belief function bel and $m_{\underline{p}}$ the corresponding BPA (negative components highlighted in red). The last row shows the Dubois–Prade entropy H_D for each solution.

[4] See [Daniel et al., 2025b] to read about which fragment of collected data was used.

Table 2: Comparison of belief mass assignments under different objective functions. Red entries mark negative components in the pseudo-BPA $m_{\underline{p}}$.

	$\underline{p}$	$m_{\underline{p}}$	m_{ZO}	m_{CW}	m_{HD}	m_{rSP}	$m^{L_1}_{\mathrm{unit}}$	$m^{L_1}_{\mathrm{layered}}$
$\{\omega_1\}$	0.163	0.163	0.163	0.163	0.174	0.400	0.174	0.174
$\{\omega_2\}$	0.211	0.211	0.297	0.211	0.222	0.321	0.222	0.222
$\{\omega_3\}$	0.179	0.179	0.317	0.179	0.190	0.229	0.190	0.190
$\{\omega_4\}$	0.050	0.050	0.222	0.050	0.060	0.050	0.060	0.060
$\{\omega_1,\omega_2\}$	0.461	0.086	0	0.081	0.065	0	0.065	0.065
$\{\omega_1,\omega_3\}$	0.423	0.080	0	0.075	0.059	0	0.059	0.059
$\{\omega_1,\omega_4\}$	0.262	0.048	0	0.044	0.027	0	0.027	0.027
$\{\omega_2,\omega_3\}$	0.480	0.089	0	0.089	0.068	0	0.068	0.068
$\{\omega_2,\omega_4\}$	0.313	0.052	0	0.052	0.031	0	0.031	0.031
$\{\omega_3,\omega_4\}$	0.279	0.050	0	0.050	0.029	0	0.029	0.029
$\{\omega_1,\omega_2,\omega_3\}$	0.778	-0.031	0	0	0	0	0	0
$\{\omega_1,\omega_2,\omega_4\}$	0.579	-0.032	0	0	0	0	0	0
$\{\omega_1,\omega_3,\omega_4\}$	0.539	-0.032	0	0	0	0	0	0
$\{\omega_2,\omega_3,\omega_4\}$	0.600	-0.032	0	0	0	0	0	0
Ω	0.953	0.070	0	0	0.075	0	0.075	0.075
H_D		0	0	0.271	0.297	0	0.297	0.297

Discussion. Several points can be drawn from Table 2: The L_1 projections (unit and layered) yield exactly the same BPA, which coincides with the HD solution. This is extremely interesting, especially when realizing the number of extreme vertices of respective polyhedron. Not-suprisingly, HD/L_1 solutions achieve the highest possible value of entropy H_D under the constraints. Observe that as the Jeffreys' bounds require nonzero mass on smaller sets, preventing all mass from being assigned to Ω.

In all cases, the mass on each focal element meets or exceeds its Jeffreys' lower bound, ensuring that the LP-based correction fully respects the empirical constraints while allowing the choice of a representative BPA to be guided by different epistemic preferences. The convergence of HD and L_1-based solutions illustrates a degree of *stability* in the selection process: different criteria can lead to the same BPA when the feasible region is strongly shaped by the data. Conversely, criteria like rSP or CW explore alternative vertices of the feasible polytope, which may be more appropriate in applications that value sparsity or higher belief on larger sets.

6 Conclusion

We have presented a polyhedral approach to correcting pseudo-belief functions derived from Jeffreys' binomial confidence interval bounds, ensuring that the resulting models are valid belief functions. By formulating the problem as a set of linear constraints in the space of basic probability assignments, we have shown that the feasible solutions form a convex polytope representing all belief functions consistent with the empirical bounds. This geometric formulation allows the use of linear programming to select representative solutions according to different optimization criteria, reflecting distinct epistemic attitudes.

The experimental study demonstrates that the method reliably produces belief functions satisfying all lower-bound constraints, while retaining flexibility to adapt the distribution of mass among focal elements. Some optimization criteria, such as Dubois–Prade entropy maximization

and certain L_1-based objectives, yield identical results in highly constrained settings, whereas others produce markedly different mass allocations, offering alternative interpretations of the same empirical data. This highlights both the robustness of the correction process and its adaptability to the needs of specific applications.

The framework is general and can be extended to alternative sources of empirical constraints, different definitions of belief function optimality, and larger frames of discernment where computational efficiency becomes critical. We believe that this approach can serve as a practical tool for bridging the gap between statistical estimation and belief function modeling, particularly in decision-making contexts where uncertainty must be expressed in a credal form.

References

Binomial coefficient. https://en.wikipedia.org/wiki/Binomial_coefficient, a. accessed Aug. 11, 2025.

Hockey-stick identity. https://en.wikipedia.org/wiki/Hockey-stick_identity, b. accessed Aug. 11, 2025.

Michel Berkelaar. *lpSolve: Interface to lpSolve v. 5.5 to Solve Linear/Integer Programs*, 2024. URL https://CRAN.R-project.org/package=lpSolve. R package version 5.6.23.

Martin Bordini, Thibaut Dénoeux, and Eric Gaussier. Learning calibrated belief functions from conformal predictions. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 2626–2647. PMLR, 2023. URL https://proceedings.mlr.press/v202/martin-bordini23a.html.

Fabio Cuzzolin. *The geometry of uncertainty: the geometry of imprecise probabilities*. Springer Nature, 2020.

Milan Daniel, Radim Jiroušek, and Václav 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, 2025a.

Milan Daniel, Radim Jiroušek, and Václav Kratochvíl. How sir harold jeffreys would create a belief function based on data. In *Proceedings of the 13th Workshop on Uncertainty Processing (WUPES'25)*, pages 92–103, 2025b.

Didier Dubois and Henri Prade. Properties of measures of information in evidence and possibility theories. *Fuzzy sets and systems*, 24(2):161–182, 1987.

Zhiqiang Geng et al. Data- and knowledge-driven belief rule learning for hybrid models. *Information Sciences*, 670:119897, 2024. doi: 10.1016/j.ins.2024.119897. URL https://www.sciencedirect.com/science/article/abs/pii/S0020025524011150.

Charles J. Geyer, Glen D. Meeden, and incorporates code from cddlib written by Komei Fukuda. *rcdd: Computational Geometry*, 2023. URL https://CRAN.R-project.org/package=rcdd. R package version 1.6.

Branko Grünbaum. *Convex Polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2nd edition, 2003. ISBN 9780387004242. doi: 10.1007/978-1-4613-0019-9.

F. Huang, S. Ruan, and T. Dénoeux. Application of belief functions to medical image segmentation: A review. *Information Fusion*, 86:61–83, 2022. doi: 10.1016/j.inffus.2022.05.002. URL https://arxiv.org/abs/2205.01733.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL https://www.R-project.org/.

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

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

Z. Zhang et al. A new correlation measure for belief functions and their application. *Entropy*, 25(5):756, 2023. doi: 10.3390/e25050756. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC10297068/.

Günter M. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. ISBN 9780387943657. doi: 10.1007/978-1-4613-8431-1.