

Article

From Non-Parametric Predictive Inference to Evidence-Theoretic Uncertainty Representation in Artificial Intelligence

María Isabel A. Benítez, Serafín Moral-García and Joaquín Abellán * 

Department of Computer Science and Artificial Intelligence, University of Granada, 18071 Granada, Spain; miab@correo.ugr.es (M.I.A.B.); seramoral@decsai.ugr.es (S.M.-G.)

* Correspondence: jabellán@decsai.ugr.es

Abstract

Artificial intelligence systems that learn or reason from finite empirical data often require uncertainty representations that go beyond a single precise probability distribution. This is especially relevant when observations are scarce, incomplete or not reliable enough to support precise probabilistic assessments. In current data-driven AI tools, empirical information extracted from data must often be converted into a structured uncertainty model before it can be used for reasoning, learning or decision support. The singleton intervals induced by NPI-M and A-NPI-M provide such a representation, since they express the predictive information obtained from the observed data without introducing externally chosen cautiousness parameters. Evidence theory is useful in this context because it allows partial support to be assigned to sets of alternatives, making it suitable for representing imperfect knowledge in AI systems. This paper studies how Non-Parametric Predictive Inference for multinomial data (NPI-M) can be connected with evidence theory through reachable probability intervals. Since the exact NPI-M model does not directly define a credal set, we focus on its approximated version, A-NPI-M, which preserves the NPI-M singleton bounds and represents them through reachable probability intervals. We analyze whether the resulting credal set can be represented exactly by a belief function, showing that this is not possible in general, although exact representations may exist in particular cases. Motivated by this limitation, we construct a basic probability assignment whose belief and plausibility values reproduce the A-NPI-M singleton bounds. The resulting belief function preserves the marginal interval information of A-NPI-M while adding an evidential structure on composite events, and its associated set of compatible probability distributions is included in the A-NPI-M credal set. The construction is presented by cases, illustrated with numerical examples and compared with the belief-function representation of the Imprecise Dirichlet Model. The proposed model provides a theoretical representation layer that may support uncertainty-aware AI procedures by transforming empirical predictive information into structured imperfect knowledge before reasoning, learning or decision-support criteria are applied.



Academic Editor: Gianni D'Angelo

Received: 30 May 2026

Revised: 17 June 2026

Accepted: 23 June 2026

Published: 25 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: artificial intelligence; data-driven uncertainty modelling; evidence theory; imprecise probabilities; non-parametric predictive inference; reachable probability intervals

1. Introduction

Representing uncertainty from data is a central problem in artificial intelligence, statistical inference, machine learning and decision support systems. In many practical situations, the available information is finite, incomplete, noisy or not reliable enough to justify the

selection of a single precise probability distribution. This is especially common when information is extracted from databases, where observations may be affected by sampling limitations, missing values, measurement errors or other forms of partial reliability. In such cases, the imprecision associated with the extracted information should not be hidden by forcing a precise probabilistic assessment. Imprecise probability models provide a natural framework for these situations, allowing uncertainty to be represented by a set of compatible probability distributions rather than by a single probability measure [1]. In finite settings, this is often done through credal sets, that is, closed and convex sets of probability distributions [2].

In artificial intelligence, uncertainty representation is not only a theoretical issue but also a practical requirement for learning, reasoning and decision making from data. Modern AI systems increasingly transform empirical observations into models of information, knowledge or decision support. When the data supporting these models are scarce or only partially reliable, a single probability vector may lead to overconfident reasoning. Mathematical models based on probability intervals, credal sets or belief functions provide suitable tools for representing this imprecision, because they allow data-derived information to be used without assuming a level of precision that is not supported by the data. In this context, parameter-free or weakly parametrized uncertainty models are particularly valuable, because they reduce the dependence on externally chosen assumptions while preserving the imprecision induced by the data themselves. Recent work in AI has also emphasized the role of uncertainty-aware reasoning and data-driven uncertainty quantification in intelligent systems [3,4].

Several mathematical formalisms are closely related to this view. Evidence theory represents uncertainty by means of belief and plausibility functions induced by a basic probability assignment [5,6]. It allows partial support to be assigned to sets of alternatives rather than forcing all evidence to be distributed over single outcomes. Probability interval models describe uncertainty by assigning lower and upper probabilities to the elementary events of a finite space [7]. These models are especially useful in data-based AI applications, because the interval bounds can often be obtained directly from empirical frequencies. Among them, reachable probability intervals are particularly relevant, since their bounds are tight and do not introduce artificial imprecision.

Evidence theory has been widely used in artificial intelligence applications, especially in information fusion, uncertain reasoning, sensor-based intelligent systems and decision support [8–10]. Related ideas have also appeared in classification and ensemble learning, where belief and Dempster–Shafer theories have been used to manage classifier consensus and imperfect evidence [11]. This makes evidence theory a natural candidate for representing imperfect knowledge obtained from data, provided that the connection with the underlying data-generated uncertainty model is clearly specified.

Belief functions and reachable probability intervals are both connected with imprecise probabilities, but they are not equivalent formalisms in general. A belief function induces lower and upper probabilities for all events and, therefore, also probability intervals for singletons. However, a system of reachable probability intervals does not necessarily determine a belief function whose associated set of compatible probability distributions coincides with the credal set defined by those intervals [12]. This distinction is important when one wants to interpret interval models generated from data within evidence theory.

A known case in which such a connection can be established is the Imprecise Dirichlet Model (IDM), introduced for inference from multinomial data [2]. The IDM generates probability intervals from observed frequencies and a cautiousness parameter, and it has been widely used in classification and learning under uncertainty [13,14]. An important property of the IDM is that its induced probability intervals can also be represented by

a belief function [12]. In this representation, part of the mass is assigned to the observed singleton events and the remaining mass is assigned to the whole frame [12].

Non-Parametric Predictive Inference for Multinomial data (NPI-M) offers a different way to build uncertainty models from categorical data, avoiding the introduction of a parametric prior model or an externally fixed cautiousness parameter [15,16]. This feature is relevant for data-driven AI systems, where uncertainty representations are often required to be obtained directly from the observed data. In contrast to models that depend on a user-selected parameter, NPI-M aims to express predictive uncertainty from the empirical information itself. The exact NPI-M model provides lower and upper probabilities from a predictive interpretation of the observed data, but it does not directly define a credal set. For this reason, an approximated version, denoted A-NPI-M, has been used in classification problems. A-NPI-M preserves the singleton lower and upper probabilities induced by NPI-M and represents them through reachable probability intervals. This approximation has shown behaviour very close to the original model based on NPI in practical classification applications [17].

The relationship between A-NPI-M and belief functions is more delicate than in the IDM case. Since reachable probability intervals and belief functions are not equivalent in general, the credal set defined by the A-NPI-M probability intervals cannot be expected to admit an exact belief function representation in all cases. In this paper, we analyze this issue explicitly. We provide examples showing that such an exact representation is not possible in general, together with particular cases where it can be obtained.

Motivated by this limitation, we construct a basic probability assignment whose belief and plausibility values reproduce the A-NPI-M lower and upper probabilities for singleton events. This construction should not be understood as an exact belief function representation of the whole A-NPI-M credal set in full generality. Rather, it defines an evidential model associated with A-NPI-M that preserves its marginal interval information on elementary events. The associated belief function adds an evidential structure on composite events, and the corresponding set of compatible probability distributions is included in the credal set defined by the A-NPI-M probability intervals.

The construction is necessarily different from the known belief function representation of the IDM. In the IDM, all singleton intervals have the same width, and the remaining mass can be assigned to the whole frame. In A-NPI-M, the interval width depends on the observed frequency configuration: interior frequencies and boundary frequencies induce different widths. Consequently, the remaining mass must be distributed over suitable focal elements that are not singletons.

This connection makes it possible to compare A-NPI-M with the IDM from a common evidential perspective. In particular, we analyze the belief and plausibility intervals induced by both models and show that, for the usual choice $s = 1$, the IDM intervals are strictly contained in the A-NPI-M intervals for every frequency configuration. We also study the case $s = 2$, where equality is possible only in a particular boundary configuration.

The contribution of this paper is therefore threefold. First, it studies the possible exact equivalence between A-NPI-M probability intervals and belief functions, showing by examples that such an equivalence does not hold in general, although it may occur in particular cases. Second, it provides procedures and examples for constructing a belief function that reproduces the A-NPI-M bounds on singleton events, yielding an evidence-theoretic uncertainty representation at the singleton level. Third, it compares the resulting singleton intervals with those induced by the IDM, with special attention to the usual values $s = 1$ and $s = 2$. These results connect non-parametric predictive inference, imprecise probabilities and evidence theory within a common framework for AI systems that must

transform finite empirical data into structured models of imperfect knowledge without relying on arbitrary tuning parameters.

The remainder of this paper is organized as follows. Section 2 introduces the necessary background on credal sets, reachable probability intervals, belief functions, the IDM and A-NPI-M. Section 3 studies the relationship between A-NPI-M probability intervals and belief functions, showing by examples that an exact belief function representation is not possible in general and identifying cases where it can be obtained. Section 4 constructs a basic probability assignment whose belief and plausibility values reproduce the A-NPI-M singleton bounds. Section 5 compares A-NPI-M with the IDM from the point of view of the induced singleton intervals. Section 6 discusses the implications of the proposed construction for uncertainty representation in data-driven AI systems and for the use of information measures from evidence theory. Finally, Section 7 summarizes the main conclusions and outlines future work.

2. Background

2.1. Imprecise Probabilities and Credal Sets

Classical probability theory represents uncertainty by means of a single probability distribution. This representation is appropriate when the available information is sufficiently precise, but it may be too restrictive when data are scarce, incomplete, noisy or only partially reliable. In such situations, using a single probability distribution may fail to reflect the imprecision associated with the available information. A representation based on a set of compatible probability distributions is therefore more reliable for explicitly modelling this imprecision.

Imprecise probability theory provides a broader framework in which uncertainty is represented by a set of probability distributions rather than by a single one [1]. When this set is closed and convex, it is usually called a credal set. Let

$$X = \{x_1, \dots, x_K\}$$

be a finite space. A credal set $\mathcal{P}$ on X induces lower and upper probabilities for every event $A \subseteq X$, defined as

$$\underline{P}(A) = \inf_{P \in \mathcal{P}} P(A), \quad \overline{P}(A) = \sup_{P \in \mathcal{P}} P(A).$$

Thus, the available information is expressed by all probability distributions compatible with the constraints imposed by the model. This makes credal sets a natural formalism for representing epistemic uncertainty.

This type of representation is especially appropriate when the available information has been extracted from data sources that cannot be considered fully reliable or complete. Instead of summarizing such information by a single probability distribution, imprecise probability models retain a range of compatible assessments, making the limitations of the empirical information explicit.

Several mathematical theories can be interpreted as particular cases or structured sub-families within the general framework of imprecise probabilities. At one of the most general levels, arbitrary credal sets allow for closed and convex sets of probability distributions without imposing additional structural restrictions. More specific models include Choquet capacities of order 2, belief functions in Dempster–Shafer theory, reachable probability intervals and graded possibility models [18]. These theories differ in their expressive power, their computational properties and the type of information they are best suited to represent.

Evidence theory is based on belief and plausibility functions induced by a basic probability assignment [5,6]. It is especially useful when information can be assigned not only to singletons but also to subsets of the frame. Probability interval models, by contrast,

specify lower and upper probabilities for the elementary events. Among them, reachable probability intervals form an important class because their bounds are tight: the lower and upper value assigned to each singleton can actually be attained by some probability distribution in the associated credal set [7].

These formalisms are related, but they are not equivalent in general. Belief functions and reachable probability intervals are both contained in broader families of imprecise probability models, but neither of them contains the other as a whole [12]. Some belief functions cannot be represented only through singleton probability intervals, and some systems of reachable probability intervals do not correspond to belief functions. Nevertheless, there are relevant models generated from data that belong to both formalisms. This is the case of the probability intervals generated by the Imprecise Dirichlet Model (IDM), which are reachable and can also be represented by a belief function [12]. This position is important for the present paper, since our aim is to study whether a similar connection with evidence theory can be established for A-NPI-M.

From the perspective of data-driven artificial intelligence, these formalisms can be seen as different ways of transforming empirical information into structured uncertainty models. This step is relevant because many AI tools do not operate directly on raw observations but on summaries, local distributions, predictive scores or decision criteria obtained from data. When these quantities are based on finite or partially reliable samples, the uncertainty attached to the empirical information should be represented explicitly. Models based on credal sets, belief functions or reachable probability intervals provide such representations, each with a different balance between expressiveness and computational tractability.

Figure 1 provides a schematic view of the relationship among the main uncertainty formalisms considered in this paper. In particular, it recalls that belief functions and reachable probability intervals are not equivalent in general, although some interval models generated from data may admit a representation in both frameworks. This is the case of the IDM, and the question addressed later is whether an analogous representation can be obtained for A-NPI-M.

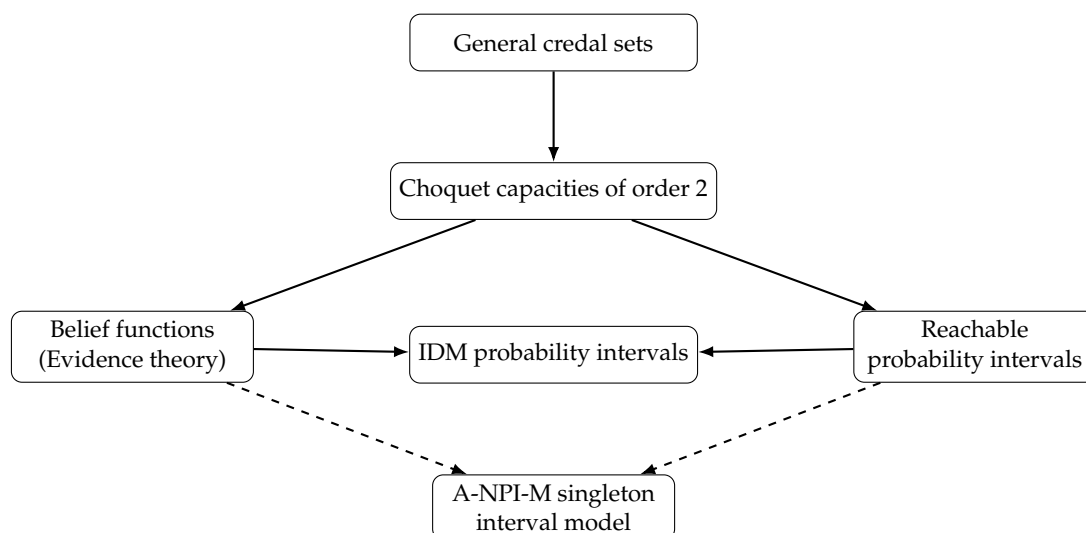


Figure 1. Schematic position of some imprecise probability models. General credal sets provide a general representation. Belief functions and reachable probability intervals are structured sub-families that are not equivalent in general. IDM probability intervals belong to both families. The dashed connection indicates the belief function representation of the A-NPI-M singleton interval model studied in this paper.

2.2. Evidence Theory

Evidence theory, also known as Dempster–Shafer theory, represents uncertainty through a basic probability assignment defined on subsets of the frame of discernment [5,6]. Let $X = \{x_1, \dots, x_K\}$ be a finite set, and let 2^X denote its power set, that is, the set of all subsets of X . A basic probability assignment is a function

$$m : 2^X \rightarrow [0, 1]$$

such that

$$m(\emptyset) = 0, \quad \sum_{A \subseteq X} m(A) = 1.$$

Subsets $A \subseteq X$ with $m(A) > 0$ are called focal elements. The corresponding belief and plausibility functions are given by

$$Bel(A) = \sum_{B \subseteq A} m(B), \quad Pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad A \subseteq X.$$

Conversely, if a lower probability function is given, its associated Möbius transform is defined by

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} Bel(B), \quad A \subseteq X.$$

A lower probability function is a belief function when this transform defines a valid basic probability assignment, that is, when all the values $m(A)$ are non-negative and sum to one [12].

For singleton events, the interval

$$[Bel(\{x_i\}), Pl(\{x_i\})]$$

can be interpreted as the lower and upper probability assigned to x_i .

From the viewpoint of imprecise probabilities, a belief function induces a set of compatible probability distributions. Let

$$\Delta_K = \left\{ P = (p_1, \dots, p_K) : p_i \geq 0, i = 1, \dots, K, \sum_{i=1}^K p_i = 1 \right\}$$

denote the set of all probability distributions on X , where $p_i = P(x_i)$. The set of probability distributions associated with m is defined as

$$\mathcal{P}(m) = \{P \in \Delta_K : Bel(A) \leq P(A) \leq Pl(A), A \subseteq X\}.$$

Equivalently, since belief and plausibility are conjugate functions, it is enough to impose the lower constraints

$$\mathcal{P}(m) = \{P \in \Delta_K : Bel(A) \leq P(A), A \subseteq X\}.$$

Thus, a belief function can be understood as a particular way of defining a credal set through lower probabilities induced by a basic probability assignment.

Evidence theory is particularly useful because it allows probability mass to be assigned not only to singletons but also to subsets that are not singletons. In this way, it can represent both support for precise outcomes and partial ignorance among several alternatives. This feature is central in the present work, because the belief functions considered later use focal elements that are not singletons to reproduce the singleton bounds induced by A-NPI-M.

2.3. Reachable Probability Intervals

A probability interval model assigns to each singleton $x_i \in X$ a lower and an upper probability,

$$[l_i, u_i], \quad i = 1, \dots, K.$$

The associated credal set is

$$\mathcal{P}(I) = \{P \in \Delta_K : l_i \leq P(x_i) \leq u_i, i = 1, \dots, K\},$$

where Δ_K denotes the set of all probability distributions on X .

Not every system of intervals is equally informative. Some interval bounds may be impossible to attain by any probability distribution in the associated credal set. In that case, the intervals contain artificial imprecision. To avoid this problem, the notion of reachability is used [7]. A system of probability intervals is reachable if, for each singleton x_i , every value in its interval can be attained by at least one probability distribution in the associated credal set. In particular, the lower and upper bounds are tight.

Equivalently, a system of probability intervals

$$I = \{[l_i, u_i]\}_{i=1}^K$$

is reachable if and only if the following two conditions hold for every $j = 1, \dots, K$:

$$\sum_{i \neq j} l_i + u_j \leq 1,$$

and

$$\sum_{i \neq j} u_i + l_j \geq 1.$$

The first condition guarantees that the upper bound u_j can be attained without violating the lower bounds of the remaining components. The second condition guarantees that the lower bound l_j can be attained without forcing the remaining components above their upper bounds.

Reachable probability intervals are especially useful in data-based models, because they provide a compact and computationally tractable representation of the uncertainty induced by empirical frequencies. They have also been used in classification and decision tree induction, where uncertainty measures are computed repeatedly from local class distributions [14]. In data-driven AI applications, this compactness is important because local empirical distributions must often be transformed into uncertainty models many times during learning, model selection or decision making.

The relationship between reachable probability intervals and belief functions is subtle. Both are important uncertainty formalisms within imprecise probabilities, but they are not equivalent in general, as shown in [12]. Some belief functions cannot be fully represented by singleton probability intervals, and some systems of reachable probability intervals cannot be represented by any belief function. Nevertheless, some relevant families of reachable probability intervals can also be represented by belief functions. A known example is given by the intervals generated by the Imprecise Dirichlet Model. This distinction is important in the present paper, because the A-NPI-M intervals studied later are reachable, but they do not necessarily define a credal set associated with a belief function.

2.4. The Imprecise Dirichlet Model

The Imprecise Dirichlet Model (IDM) was introduced by Walley for inference from multinomial data without committing to a single precise prior distribution [2]. Let n_i be the number of observations of class x_i , $i = 1, \dots, K$, with

$$\sum_{i=1}^K n_i = N.$$

For a cautiousness parameter $s > 0$, the IDM induces the probability intervals

$$[l_i^{IDM}, u_i^{IDM}] = \left[\frac{n_i}{N+s}, \frac{n_i+s}{N+s} \right], \quad i = 1, \dots, K.$$

The width of each interval is

$$u_i^{IDM} - l_i^{IDM} = \frac{s}{N+s}.$$

Thus, all classes receive the same amount of imprecision. For fixed s , this imprecision decreases as the sample size N increases. This behaviour is appropriate when imprecision is interpreted mainly as a consequence of limited data.

The IDM is one of the most widely used models for generating probability intervals from data. Walley [2] discussed $s = 1$ and $s = 2$ as relevant practical choices for the cautiousness parameter, and these values have been widely used in classification and related learning procedures [19–21].

A key property of the IDM is that its induced probability intervals are reachable and can be represented by a belief function. The corresponding basic probability assignment is

$$m_{IDM}(\{x_i\}) = \frac{n_i}{N+s}, \quad i = 1, \dots, K,$$

and

$$m_{IDM}(X) = \frac{s}{N+s}.$$

All other masses are zero. This construction has a simple interpretation: the observed empirical information is assigned to the singleton focal elements, while the remaining mass is assigned to the whole frame, representing global ignorance. The belief and plausibility values induced by this mass assignment satisfy

$$Bel_{IDM}(\{x_i\}) = l_i^{IDM}, \quad Pl_{IDM}(\{x_i\}) = u_i^{IDM}.$$

Therefore, IDM probability intervals provide a clear example of a reachable interval model generated from data that can also be interpreted within evidence theory.

2.5. Non-Parametric Predictive Inference and A-NPI-M

Non-Parametric Predictive Inference for Multinomial data (NPI-M) provides an alternative framework for predictive inference from categorical data [15,16,22]. Unlike the IDM, NPI-M does not require the selection of a cautiousness parameter. Its aim is to base inference directly on the observed data, avoiding prior specifications and parametric assumptions as far as possible. This feature is especially relevant in data-driven AI settings, where uncertainty models are often required to summarize the information obtained from data without introducing externally chosen tuning parameters. In this sense, the lower and upper probabilities induced by NPI-M can be interpreted as a direct representation of the predictive information supported by the empirical observations.

Let

$$X = \{x_1, \dots, x_K\}$$

be a finite set of categories, and let n_i denote the number of observations of category x_i , with

$$\sum_{i=1}^K n_i = N.$$

For any event $A \subseteq X$, denote

$$n_A = \sum_{x_i \in A} n_i.$$

The exact NPI-M model provides lower and upper probabilities for events, not only for singleton events. For a non-empty proper event $A \subset X$, the NPI-M lower and upper probabilities can be written as

$$\underline{P}_{NPI}(A) = \max\left\{\frac{n_A - 1}{N}, 0\right\},$$

and

$$\bar{P}_{NPI}(A) = \min\left\{\frac{n_A + 1}{N}, 1\right\}.$$

In addition,

$$\underline{P}_{NPI}(\emptyset) = \bar{P}_{NPI}(\emptyset) = 0, \quad \underline{P}_{NPI}(X) = \bar{P}_{NPI}(X) = 1.$$

This definition on events is important. The exact NPI-M model is not defined only through the marginal bounds on singletons; it also imposes constraints on composite events. These constraints express the predictive structure of NPI-M and are part of the exact model.

However, the set of probability distributions associated with the exact NPI-M constraints is not, in general, a closed and convex set of probability distributions [22]. This makes the exact model less convenient for learning procedures in which one needs to optimize uncertainty measures over a credal set. This is the case, for example, in classification and decision tree induction, where uncertainty values may need to be computed repeatedly from local class distributions.

For this reason, an approximated version, usually denoted by A-NPI-M, has been used in data mining applications [17]. A-NPI-M retains the singleton bounds induced by NPI-M and considers the closed and convex set of probability distributions determined by those singleton probability intervals. Thus, A-NPI-M replaces the exact NPI-M model defined on events by the credal set defined by the singleton probability intervals. This approximation is useful for AI learning tools because it converts the predictive information extracted from empirical frequencies into a tractable credal representation while preserving the NPI-M bounds on the elementary categories.

For each class x_i , the singleton bounds are

$$[l_i^A, u_i^A] = \left[\max\left\{\frac{n_i - 1}{N}, 0\right\}, \min\left\{\frac{n_i + 1}{N}, 1\right\} \right].$$

The associated A-NPI-M credal set is

$$\mathcal{P}_A = \left\{ P \in \Delta_K : l_i^A \leq P(x_i) \leq u_i^A, i = 1, \dots, K \right\},$$

where Δ_K denotes the set of all probability distributions on X .

Therefore, the passage from NPI-M to A-NPI-M changes the representation: NPI-M provides lower and upper probabilities for all events, whereas A-NPI-M works with the closed and convex set induced by the singleton probability intervals. This makes A-NPI-M easier to use in computational procedures that require optimization over a credal set.

When the observed frequency is interior, that is, when $0 < n_i < N$, the A-NPI-M interval becomes

$$[l_i^A, u_i^A] = \left[\frac{n_i - 1}{N}, \frac{n_i + 1}{N} \right],$$

whose width is

$$u_i^A - l_i^A = \frac{2}{N}.$$

In the boundary cases, the behaviour is different. If $n_i = 0$, then

$$[l_i^A, u_i^A] = \left[0, \frac{1}{N} \right],$$

whereas if $n_i = N$, then

$$[l_i^A, u_i^A] = \left[\frac{N - 1}{N}, 1 \right].$$

In both boundary cases, the interval width is $1/N$.

Thus, A-NPI-M is a model with vanishing imprecision: its interval widths decrease as the sample size increases. However, unlike the IDM, the width is not uniform across all frequency configurations. Interior classes have width $2/N$, whereas boundary classes have width $1/N$. This feature will be central for the belief function construction developed later in this paper, because the remaining mass cannot simply be assigned to the whole frame, as happens in the IDM.

A-NPI-M has been used in classification because it provides a non-parametric alternative based on data to the IDM while producing a closed and convex set of probability distributions. Previous studies have shown that the approximated model behaves practically like the original procedure based on NPI in decision tree classification while being easier to handle computationally [17]. Its main advantage is that it allows for the use of uncertainty measures based on optimization and credal learning methods, which require a closed and convex set of probability distributions.

It is important to note, however, that the passage from NPI-M to A-NPI-M changes the representation. NPI-M provides lower and upper probabilities for all events, whereas A-NPI-M works with the credal set induced by the singleton probability intervals. Therefore, A-NPI-M preserves the NPI-M bounds on elementary categories, but it may be less restrictive than the exact NPI-M model on composite events.

The previous discussion highlights the point addressed in the following sections. The IDM and A-NPI-M both generate reachable probability intervals from multinomial data, but they do so in different ways. The IDM is parametric and has a known belief function representation, whereas A-NPI-M is non-parametric, and its relationship with evidence theory is more delicate. Clarifying this relationship is useful for AI systems that rely on empirical information, because it allows the data-based predictive intervals of A-NPI-M to be interpreted within an evidence-theoretic framework. We first analyze whether the credal set defined by the A-NPI-M probability intervals can be represented exactly by a belief function. We then construct a basic probability assignment that preserves the A-NPI-M singleton bounds and use this construction to compare A-NPI-M with the IDM.

3. A-NPI-M Probability Intervals and Belief Functions

This section studies the relationship between the credal set defined by the A-NPI-M probability intervals and belief functions. From the perspective of data-driven artificial intelligence, this relationship is relevant because the A-NPI-M intervals provide a compact representation of predictive information obtained from empirical frequencies, whereas belief functions provide an evidence-theoretic representation of imperfect knowledge.

Connecting both formalisms would make it possible to interpret the data-based information provided by A-NPI-M within an evidential framework.

We first show, through a representative example, that an exact belief function representation is not possible in general. This motivates the subsequent construction of a basic probability assignment that preserves the A-NPI-M singleton bounds, even when the associated set of compatible probability distributions does not coincide with the whole A-NPI-M credal set.

Consider a frame with four elements,

$$X = \{x_1, x_2, x_3, x_4\},$$

and observed frequencies

$$(n_1, n_2, n_3, n_4) = (2, 2, 2, 2), \quad N = 8.$$

Since all frequencies are interior, A-NPI-M induces

$$[l_i^A, u_i^A] = \left[\frac{1}{8}, \frac{3}{8}\right], \quad i = 1, \dots, 4.$$

Thus,

$$\mathcal{P}_A = \left\{ P \in \Delta_4 : \frac{1}{8} \leq P(x_i) \leq \frac{3}{8}, i = 1, \dots, 4 \right\}.$$

The lower probability induced by a reachable system of probability intervals on an event $A \subseteq X$ is

$$\underline{P}_A(A) = \max \left\{ \sum_{x_i \in A} l_i^A, 1 - \sum_{x_i \notin A} u_i^A \right\}.$$

For singleton events, this gives

$$\underline{P}_A(\{x_i\}) = \frac{1}{8}, \quad i = 1, \dots, 4.$$

For any pair, for instance $\{x_1, x_2\}$, we obtain

$$\underline{P}_A(\{x_1, x_2\}) = \max \left\{ \frac{1}{8} + \frac{1}{8}, 1 - \frac{3}{8} - \frac{3}{8} \right\} = \frac{1}{4}.$$

For any subset with three elements, for instance $\{x_1, x_2, x_3\}$, we have

$$\underline{P}_A(\{x_1, x_2, x_3\}) = \max \left\{ \frac{1}{8} + \frac{1}{8} + \frac{1}{8}, 1 - \frac{3}{8} \right\} = \frac{5}{8}.$$

Assume now that there exists a belief function whose associated set of compatible probability distributions coincides with $\mathcal{P}_A$. Then the lower probability induced by that belief function would have to coincide with $\underline{P}_A$ on all events. Therefore, its Möbius transform would have to define a non-negative basic probability assignment.

For singleton focal elements, the Möbius transform gives

$$m(\{x_i\}) = \underline{P}_A(\{x_i\}) = \frac{1}{8}, \quad i = 1, \dots, 4.$$

For any pair $\{x_i, x_j\}$,

$$m(\{x_i, x_j\}) = \underline{P}_A(\{x_i, x_j\}) - m(\{x_i\}) - m(\{x_j\}) = \frac{1}{4} - \frac{1}{8} - \frac{1}{8} = 0.$$

For any subset with three elements $\{x_i, x_j, x_k\}$,

$$m(\{x_i, x_j, x_k\}) = \underline{P}_A(\{x_i, x_j, x_k\}) - m(\{x_i\}) - m(\{x_j\}) - m(\{x_k\}) = \frac{5}{8} - \frac{3}{8} = \frac{1}{4}.$$

There are four subsets with three elements, so the total mass assigned to them is

$$4 \cdot \frac{1}{4} = 1.$$

The total mass assigned to singletons is

$$4 \cdot \frac{1}{8} = \frac{1}{2}.$$

Since the masses of pairs are zero, the mass assigned to the whole frame would be

$$m(X) = 1 - \frac{1}{2} - 1 = -\frac{1}{2}.$$

This is not possible for a basic probability assignment, because all masses must be non-negative. Therefore, the credal set $\mathcal{P}_A$ defined by these A-NPI-M probability intervals cannot be represented exactly by a belief function.

This example shows that the credal set defined by A-NPI-M probability intervals is not, in general, the set of compatible probability distributions associated with a belief function. In other words, the predictive information represented by the A-NPI-M singleton intervals cannot always be transferred unchanged into evidence theory at the level of the whole credal set. This is an important point for AI systems that use uncertainty models derived from data: preserving empirical interval information does not automatically imply that the same model admits an exact evidential interpretation on all events.

Consequently, the construction developed below should not be interpreted as an exact belief function representation of the whole A-NPI-M credal set. Instead, we construct a basic probability assignment that preserves the A-NPI-M lower and upper probabilities on singleton events. This provides an evidence-theoretic model associated with A-NPI-M that keeps the marginal predictive information obtained from the data while adding an evidential structure on composite events.

4. A Belief Function Model Preserving the A-NPI-M Singleton Intervals

Let $X = \{x_1, \dots, x_K\}$ be a finite frame of discernment, and let n_i denote the observed frequency of class x_i , with

$$\sum_{i=1}^K n_i = N.$$

The A-NPI-M model induces the following reachable probability intervals on singletons:

$$[l_i^A, u_i^A] = \left[\max \left\{ \frac{n_i - 1}{N}, 0 \right\}, \min \left\{ \frac{n_i + 1}{N}, 1 \right\} \right], \quad i = 1, \dots, K.$$

Thus, when $0 < n_i < N$, the interval associated with x_i is

$$[l_i^A, u_i^A] = \left[\frac{n_i - 1}{N}, \frac{n_i + 1}{N} \right],$$

and its width is $2/N$. In the boundary cases $n_i = 0$ or $n_i = N$, the width is instead $1/N$.

These intervals summarize the predictive information obtained from the observed data. In data-driven AI systems, such local empirical information often needs to be transformed

into a structured uncertainty model before it can be used in reasoning, learning or decision-support procedures. The purpose of this section is to show how the singleton information provided by A-NPI-M can be preserved within an evidence-theoretic representation.

As discussed above, the credal set defined by the A-NPI-M probability intervals is not, in general, the set of compatible probability distributions associated with a belief function. Therefore, the aim of this section is more specific. We construct a basic probability assignment whose belief and plausibility values reproduce the A-NPI-M lower and upper probabilities on singleton events. In this sense, the construction defines a belief function model associated with A-NPI-M at the singleton level.

The construction is more involved than in the IDM. In the IDM, all singleton intervals have the same width, and the remaining mass can be assigned to the whole frame X . In A-NPI-M, by contrast, interior observed classes have interval width $2/N$, whereas boundary classes have interval width $1/N$. Hence, assigning a single mass to X would increase the plausibility of all singletons by the same amount and would not reproduce the different behaviour of observed and boundary classes. For this reason, the remaining mass must be distributed over suitable focal elements that are not singletons.

Let m_A denote the basic probability assignment associated with the A-NPI-M singleton bounds. The construction is given by cases. In each case, the mass assigned to singleton focal elements represents the direct empirical support provided by the observed frequencies, whereas the remaining mass is distributed over focal elements that are not singletons in order to reproduce the A-NPI-M plausibility bounds. This is the mechanism by which the predictive interval information obtained from data is translated into an evidential structure.

The three possible configurations of the construction are summarized schematically in Figure 2, which shows how the remaining mass is assigned to non-singleton focal elements in each case in order to recover the A-NPI-M singleton bounds.

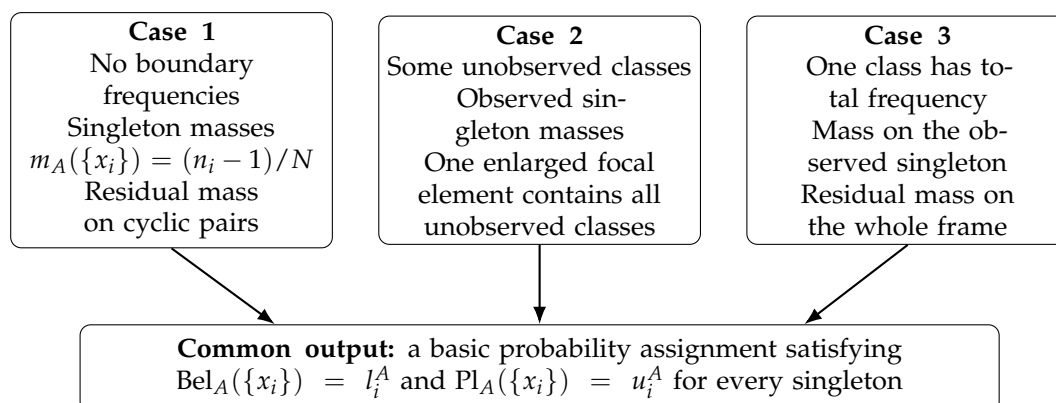


Figure 2. Schematic view of the three cases in the construction of the basic probability assignment m_A .

4.1. Case 1: No Boundary Frequencies

Assume first that

$$0 < n_i < N, \quad i = 1, \dots, K.$$

In this case, all A-NPI-M intervals have width $2/N$. Define

$$m_A(\{x_i\}) = \frac{n_i - 1}{N}, \quad i = 1, \dots, K.$$

The total mass assigned to singleton focal elements is

$$\sum_{i=1}^K m_A(\{x_i\}) = \sum_{i=1}^K \frac{n_i - 1}{N} = \frac{N - K}{N}.$$

Hence, the remaining mass is

$$1 - \sum_{i=1}^K m_A(\{x_i\}) = \frac{K}{N}.$$

This remaining mass is distributed over pairs of consecutive classes arranged in a cycle. More precisely, define

$$m_A(\{x_i, x_{i+1}\}) = \frac{1}{N}, \quad i = 1, \dots, K,$$

where $x_{K+1} = x_1$. All other masses are zero. If $K = 2$, the same focal element $\{x_1, x_2\}$ would appear twice in the cyclic notation. In that special case, it is enough to assign

$$m_A(\{x_1, x_2\}) = \frac{2}{N}.$$

The argument below remains unchanged, since each singleton receives an additional plausibility contribution equal to $2/N$.

Then, for every singleton $\{x_i\}$,

$$Bel_A(\{x_i\}) = m_A(\{x_i\}) = \frac{n_i - 1}{N}.$$

Moreover, each x_i appears in exactly two focal elements that are not singletons, namely $\{x_{i-1}, x_i\}$ and $\{x_i, x_{i+1}\}$, with the cyclic convention. Therefore,

$$Pl_A(\{x_i\}) = m_A(\{x_i\}) + \frac{1}{N} + \frac{1}{N},$$

and so

$$Pl_A(\{x_i\}) = \frac{n_i - 1}{N} + \frac{2}{N} = \frac{n_i + 1}{N}.$$

Consequently,

$$[Bel_A(\{x_i\}), Pl_A(\{x_i\})] = \left[\frac{n_i - 1}{N}, \frac{n_i + 1}{N} \right] = [l_i^A, u_i^A], \quad i = 1, \dots, K.$$

4.2. Case 2: Some Unobserved Classes and No Class with Total Frequency

Assume now that some classes have zero frequency, but no class has frequency N . Let

$$O = \{i : n_i > 0\}, \quad Z = \{i : n_i = 0\},$$

and suppose that $|O| = c \geq 2$. Let

$$X_Z = \{x_j : j \in Z\}$$

be the set of unobserved classes.

For observed classes, define

$$m_A(\{x_i\}) = \frac{n_i - 1}{N}, \quad i \in O,$$

whereas for unobserved classes define

$$m_A(\{x_j\}) = 0, \quad j \in Z.$$

The total mass assigned to observed singletons is

$$\sum_{i \in O} \frac{n_i - 1}{N} = \frac{N - c}{N},$$

because $\sum_{i \in O} n_i = N$. Hence, the remaining mass is

$$1 - \frac{N - c}{N} = \frac{c}{N}.$$

Order the observed classes as

$$O = \{i_1, \dots, i_c\}.$$

We assign the remaining mass to c focal elements of mass $1/N$. These focal elements are based on a cycle over the observed classes, but one of them is enlarged by including all unobserved classes. More precisely, define

$$m_A(\{x_{i_1}, x_{i_2}\} \cup X_Z) = \frac{1}{N},$$

and

$$m_A(\{x_{i_r}, x_{i_{r+1}}\}) = \frac{1}{N}, \quad r = 2, \dots, c - 1,$$

together with

$$m_A(\{x_{i_c}, x_{i_1}\}) = \frac{1}{N}.$$

All other masses are zero.

With this construction, each observed class belongs to exactly two focal elements that are not singletons. Therefore, for every $i \in O$,

$$Bel_A(\{x_i\}) = \frac{n_i - 1}{N},$$

and

$$Pl_A(\{x_i\}) = \frac{n_i - 1}{N} + \frac{2}{N} = \frac{n_i + 1}{N}.$$

On the other hand, every unobserved class x_j , with $j \in Z$, belongs only to the enlarged focal element

$$\{x_{i_1}, x_{i_2}\} \cup X_Z,$$

which has mass $1/N$. Thus,

$$Bel_A(\{x_j\}) = 0, \quad Pl_A(\{x_j\}) = \frac{1}{N}, \quad j \in Z.$$

Hence,

$$[Bel_A(\{x_j\}), Pl_A(\{x_j\})] = \left[0, \frac{1}{N}\right] = [l_j^A, u_j^A], \quad j \in Z.$$

Therefore, also in the presence of unobserved classes, provided no class has total frequency, the A-NPI-M singleton intervals are recovered as belief and plausibility intervals.

4.3. Case 3: One Class Has Total Frequency

Finally, assume that there exists a class x_a such that

$$n_a = N.$$

Then necessarily

$$n_j = 0, \quad j \neq a.$$

The A-NPI-M intervals are

$$[l_a^A, u_a^A] = \left[\frac{N-1}{N}, 1 \right],$$

for the observed class, and

$$[l_j^A, u_j^A] = \left[0, \frac{1}{N} \right], \quad j \neq a,$$

for all the remaining classes.

In this case, define

$$m_A(\{x_a\}) = \frac{N-1}{N},$$

and

$$m_A(X) = \frac{1}{N}.$$

All other masses are zero.

Then

$$Bel_A(\{x_a\}) = \frac{N-1}{N}, \quad Pl_A(\{x_a\}) = 1,$$

because both focal elements $\{x_a\}$ and X intersect $\{x_a\}$. For any $j \neq a$,

$$Bel_A(\{x_j\}) = 0,$$

and

$$Pl_A(\{x_j\}) = \frac{1}{N},$$

because the only focal element intersecting $\{x_j\}$ is X . Thus,

$$[Bel_A(\{x_i\}), Pl_A(\{x_i\})] = [l_i^A, u_i^A], \quad i = 1, \dots, K.$$

4.4. Properties of the Construction

The previous construction shows that the singleton intervals induced by A-NPI-M can always be reproduced within the belief function formalism. The result is summarized in the following proposition.

Proposition 1. *Let m_A be the basic probability assignment defined in the previous cases. Then, for every $i = 1, \dots, K$,*

$$Bel_A(\{x_i\}) = l_i^A, \quad Pl_A(\{x_i\}) = u_i^A.$$

Consequently, m_A reproduces the A-NPI-M lower and upper probabilities on singleton events.

Proof. The result follows directly from the construction in each frequency configuration. If all frequencies are interior, each singleton receives its lower mass and belongs to exactly two focal elements that are not singletons of mass $1/N$. If unobserved classes are present, observed classes again belong to two such focal elements, whereas each unobserved class belongs to only one enlarged focal element. If one class has total frequency, the focal elements are the observed singleton and the whole frame. In each case, the resulting belief and plausibility values coincide with the A-NPI-M singleton bounds. $\square$

This result is relevant for uncertainty representation in AI because it shows that the marginal predictive information produced by A-NPI-M can always be preserved inside a belief function. Therefore, even when the whole A-NPI-M credal set does not admit an

exact belief function representation, its singleton interval information can still be used to build an evidence-theoretic model directly from the observed data.

The construction differs from the corresponding representation of the IDM. In the IDM, the interval width is the same for every class and the remaining mass can therefore be assigned to the whole frame X . In A-NPI-M, by contrast, interior observed classes have interval width $2/N$, whereas boundary classes have interval width $1/N$. Consequently, the remaining mass must be distributed over suitable focal elements that are not singletons, typically pairs of observed classes, with special treatment for unobserved classes.

It is important to note that this construction guarantees only the preservation of the singleton bounds:

$$Bel_A(\{x_i\}) = l_i^A, \quad Pl_A(\{x_i\}) = u_i^A, \quad i = 1, \dots, K.$$

It does not imply, in general, that the set of compatible probability distributions associated with m_A coincides with the credal set defined by the A-NPI-M probability intervals. Since a belief function also induces lower and upper probabilities for composite events, the constructed m_A may impose additional constraints beyond those defined by the singleton probability intervals.

Let

$$\mathcal{P}(m_A) = \{P \in \Delta_K : Bel_A(B) \leq P(B) \leq Pl_A(B), B \subseteq X\}$$

be the set of compatible probability distributions associated with m_A , and let

$$\mathcal{P}_A = \left\{P \in \Delta_K : l_i^A \leq P(x_i) \leq u_i^A, i = 1, \dots, K\right\}$$

be the credal set defined by the A-NPI-M probability intervals. Since m_A reproduces the A-NPI-M singleton bounds, every distribution compatible with m_A satisfies the A-NPI-M interval constraints on the singletons. Hence,

$$\mathcal{P}(m_A) \subseteq \mathcal{P}_A.$$

The inclusion may be strict. Therefore, the proposed construction should be understood as a belief function model of A-NPI-M at the singleton level, rather than as an exact belief function representation of the whole A-NPI-M credal set. This distinction is important in data-driven AI applications: the construction preserves the empirical predictive bounds used by learning and decision procedures, while the additional evidential constraints on composite events provide a structured representation of imperfect knowledge.

4.5. Illustration of the Strict Inclusion

The possible strictness of the inclusion can be seen in a simple example. Consider again $X = \{x_1, x_2, x_3, x_4\}$ with observed frequencies

$$(n_1, n_2, n_3, n_4) = (2, 2, 2, 2), \quad N = 8.$$

The A-NPI-M singleton intervals are

$$[l_i^A, u_i^A] = \left[\frac{1}{8}, \frac{3}{8}\right], \quad i = 1, \dots, 4.$$

The associated interval credal set is therefore

$$\mathcal{P}_A = \left\{P \in \Delta_4 : \frac{1}{8} \leq P(x_i) \leq \frac{3}{8}, i = 1, \dots, 4\right\}.$$

Using the construction of Case 1, the belief function assigns mass $1/8$ to each singleton and mass $1/8$ to each cyclic pair

$$\{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_4\}, \{x_4, x_1\}.$$

Hence, for the composite event $A = \{x_1, x_2\}$, the constructed belief function gives

$$\text{Bel}_A(A) = m_A(\{x_1\}) + m_A(\{x_2\}) + m_A(\{x_1, x_2\}) = \frac{3}{8}.$$

However, the lower probability induced by the A-NPI-M singleton intervals on the same event is

$$\underline{P}_A(A) = \max\left\{\frac{1}{8} + \frac{1}{8}, 1 - \frac{3}{8} - \frac{3}{8}\right\} = \frac{1}{4}.$$

Thus the belief function imposes the additional constraint

$$P(x_1) + P(x_2) \geq \frac{3}{8},$$

whereas the A-NPI-M interval credal set only requires $P(x_1) + P(x_2) \geq 1/4$ for this event. From an uncertainty-representation perspective, these additional constraints can be interpreted as an increase in epistemic cautiousness at the level of composite events. The construction does not merely preserve the singleton bounds; it also prevents some combinations of probabilities that are compatible with the marginal intervals but less consistent with the evidential structure induced by the data. In AI systems based on finite local samples, this type of additional cautiousness may help avoid overconfident decisions based only on weak marginal information. For instance, the distribution

$$P = \left(\frac{1}{8}, \frac{1}{8}, \frac{3}{8}, \frac{3}{8}\right)$$

belongs to $\mathcal{P}_A$, because all its singleton probabilities satisfy the A-NPI-M intervals, but it does not belong to $\mathcal{P}(m_A)$, because $P(x_1) + P(x_2) = 1/4 < 3/8$. Therefore,

$$\mathcal{P}(m_A) \subsetneq \mathcal{P}_A$$

in this example.

This example clarifies the type of information modification introduced by the proposed transformation. The A-NPI-M singleton bounds are preserved exactly, but the resulting belief function may add structural constraints on composite events. In AI applications, this means that the proposed model should be interpreted as an evidential representation of the A-NPI-M singleton predictive information, not as a complete replacement of the whole A-NPI-M credal set. A general quantification of the excluded part of $\mathcal{P}_A$, for example by means of volume comparisons or optimization-based distances, is left for future work.

4.6. Numerical Examples

This subsection illustrates the previous belief function construction for A-NPI-M in the three cases considered above. In all examples, we verify that the belief and plausibility values induced by the basic probability assignment coincide with the A-NPI-M singleton bounds.

4.6.1. Case 1: No Boundary Frequencies

Consider a frame with three classes,

$$X = \{x_1, x_2, x_3\},$$

and observed frequencies

$$(n_1, n_2, n_3) = (4, 3, 3), \quad N = 10.$$

All frequencies are interior, since $0 < n_i < N$ for every i . Therefore, A-NPI-M gives

$$[l_i^A, u_i^A] = \left[\frac{n_i - 1}{N}, \frac{n_i + 1}{N} \right].$$

Hence,

$$[l_1^A, u_1^A] = [0.3, 0.5],$$

$$[l_2^A, u_2^A] = [0.2, 0.4],$$

and

$$[l_3^A, u_3^A] = [0.2, 0.4].$$

Following the construction, we first assign mass to the singleton focal elements:

$$m_A(\{x_1\}) = \frac{3}{10}, \quad m_A(\{x_2\}) = \frac{2}{10}, \quad m_A(\{x_3\}) = \frac{2}{10}.$$

The remaining mass is

$$1 - \left(\frac{3}{10} + \frac{2}{10} + \frac{2}{10} \right) = \frac{3}{10}.$$

This mass is distributed over the pairs of consecutive classes:

$$m_A(\{x_1, x_2\}) = \frac{1}{10},$$

$$m_A(\{x_2, x_3\}) = \frac{1}{10},$$

and

$$m_A(\{x_3, x_1\}) = \frac{1}{10}.$$

Therefore, the belief values on singletons are

$$Bel_A(\{x_1\}) = m_A(\{x_1\}) = \frac{3}{10},$$

$$Bel_A(\{x_2\}) = m_A(\{x_2\}) = \frac{2}{10},$$

and

$$Bel_A(\{x_3\}) = m_A(\{x_3\}) = \frac{2}{10}.$$

The plausibility values are obtained by adding the masses of all focal elements intersecting the corresponding singleton. Thus,

$$Pl_A(\{x_1\}) = m_A(\{x_1\}) + m_A(\{x_1, x_2\}) + m_A(\{x_3, x_1\}) = \frac{3}{10} + \frac{1}{10} + \frac{1}{10} = \frac{5}{10},$$

$$Pl_A(\{x_2\}) = m_A(\{x_2\}) + m_A(\{x_1, x_2\}) + m_A(\{x_2, x_3\}) = \frac{2}{10} + \frac{1}{10} + \frac{1}{10} = \frac{4}{10},$$

and

$$Pl_A(\{x_3\}) = m_A(\{x_3\}) + m_A(\{x_2, x_3\}) + m_A(\{x_3, x_1\}) = \frac{2}{10} + \frac{1}{10} + \frac{1}{10} = \frac{4}{10}.$$

Hence,

$$[Bel_A(\{x_1\}), Pl_A(\{x_1\})] = [0.3, 0.5],$$

$$[Bel_A(\{x_2\}), Pl_A(\{x_2\})] = [0.2, 0.4],$$

and

$$[Bel_A(\{x_3\}), Pl_A(\{x_3\})] = [0.2, 0.4],$$

which coincide with the A-NPI-M singleton intervals.

4.6.2. Case 2: Some Unobserved Classes and No Class with Total Frequency

Consider now a frame with four classes,

$$X = \{x_1, x_2, x_3, x_4\},$$

and observed frequencies

$$(n_1, n_2, n_3, n_4) = (5, 3, 2, 0), \quad N = 10.$$

Here x_4 is unobserved, but no class has frequency N . Thus,

$$O = \{1, 2, 3\}, \quad Z = \{4\}.$$

The A-NPI-M intervals are

$$[l_1^A, u_1^A] = \left[\frac{4}{10}, \frac{6}{10} \right] = [0.4, 0.6],$$

$$[l_2^A, u_2^A] = \left[\frac{2}{10}, \frac{4}{10} \right] = [0.2, 0.4],$$

$$[l_3^A, u_3^A] = \left[\frac{1}{10}, \frac{3}{10} \right] = [0.1, 0.3],$$

and

$$[l_4^A, u_4^A] = \left[0, \frac{1}{10} \right] = [0, 0.1].$$

We assign singleton masses to the observed classes:

$$m_A(\{x_1\}) = \frac{4}{10}, \quad m_A(\{x_2\}) = \frac{2}{10}, \quad m_A(\{x_3\}) = \frac{1}{10},$$

and

$$m_A(\{x_4\}) = 0.$$

The remaining mass is

$$1 - \left(\frac{4}{10} + \frac{2}{10} + \frac{1}{10} \right) = \frac{3}{10}.$$

It is assigned to three focal elements that are not singletons:

$$m_A(\{x_1, x_2, x_4\}) = \frac{1}{10},$$

$$m_A(\{x_2, x_3\}) = \frac{1}{10},$$

and

$$m_A(\{x_3, x_1\}) = \frac{1}{10}.$$

All other masses are zero.

Each observed class appears in exactly two focal elements that are not singletons:

$$x_1 \in \{x_1, x_2, x_4\}, \{x_3, x_1\},$$

$$x_2 \in \{x_1, x_2, x_4\}, \{x_2, x_3\},$$

and

$$x_3 \in \{x_2, x_3\}, \{x_3, x_1\}.$$

The unobserved class x_4 appears only in one focal element that is not a singleton, namely $\{x_1, x_2, x_4\}$.

Therefore,

$$Bel_A(\{x_1\}) = \frac{4}{10}, \quad Pl_A(\{x_1\}) = \frac{4}{10} + \frac{1}{10} + \frac{1}{10} = \frac{6}{10},$$

$$Bel_A(\{x_2\}) = \frac{2}{10}, \quad Pl_A(\{x_2\}) = \frac{2}{10} + \frac{1}{10} + \frac{1}{10} = \frac{4}{10},$$

$$Bel_A(\{x_3\}) = \frac{1}{10}, \quad Pl_A(\{x_3\}) = \frac{1}{10} + \frac{1}{10} + \frac{1}{10} = \frac{3}{10},$$

and

$$Bel_A(\{x_4\}) = 0, \quad Pl_A(\{x_4\}) = \frac{1}{10}.$$

Hence,

$$[Bel_A(\{x_1\}), Pl_A(\{x_1\})] = [0.4, 0.6],$$

$$[Bel_A(\{x_2\}), Pl_A(\{x_2\})] = [0.2, 0.4],$$

$$[Bel_A(\{x_3\}), Pl_A(\{x_3\})] = [0.1, 0.3],$$

and

$$[Bel_A(\{x_4\}), Pl_A(\{x_4\})] = [0, 0.1].$$

Again, these intervals coincide with the A-NPI-M singleton bounds.

4.6.3. Case 3: One Class Has Total Frequency

Finally, consider the frame

$$X = \{x_1, x_2, x_3, x_4\},$$

with observed frequencies

$$(n_1, n_2, n_3, n_4) = (10, 0, 0, 0), \quad N = 10.$$

In this case, class x_1 has total frequency. The A-NPI-M intervals are

$$[l_1^A, u_1^A] = \left[\frac{9}{10}, 1 \right] = [0.9, 1],$$

and, for $j = 2, 3, 4$,

$$[l_j^A, u_j^A] = \left[0, \frac{1}{10} \right] = [0, 0.1].$$

The corresponding basic probability assignment is

$$m_A(\{x_1\}) = \frac{9}{10},$$

and

$$m_A(X) = \frac{1}{10}.$$

All other masses are zero.

Then

$$Bel_A(\{x_1\}) = \frac{9}{10},$$

and

$$Pl_A(\{x_1\}) = m_A(\{x_1\}) + m_A(X) = \frac{9}{10} + \frac{1}{10} = 1.$$

For any $j = 2, 3, 4$,

$$Bel_A(\{x_j\}) = 0,$$

and

$$Pl_A(\{x_j\}) = m_A(X) = \frac{1}{10}.$$

Therefore,

$$[Bel_A(\{x_1\}), Pl_A(\{x_1\})] = [0.9, 1],$$

and

$$[Bel_A(\{x_j\}), Pl_A(\{x_j\})] = [0, 0.1], \quad j = 2, 3, 4.$$

Thus, the belief and plausibility intervals induced by m_A coincide exactly with the A-NPI-M singleton intervals also in the boundary case where one class has total frequency.

These examples illustrate that the belief function construction adapts to the frequency configuration. When all classes are interior, the remaining mass is distributed symmetrically over pairs of classes. When unobserved classes are present, they are included in only one focal element that is not a singleton so that their plausibility increases by $1/N$ instead of $2/N$. Finally, when one class has total frequency, the construction reduces to a simple mass assignment with one singleton focal element and the whole frame. From an AI perspective, these examples show how different empirical frequency configurations lead to different evidential structures while preserving the same A-NPI-M interpretation of the singleton bounds as data-based predictive information.

5. Comparison of the IDM with A-NPI-M

The previous section constructed a basic probability assignment m_A whose belief and plausibility values reproduce the A-NPI-M lower and upper probabilities on singleton events. This construction should not be interpreted as an exact belief function representation of the whole A-NPI-M credal set in general. Nevertheless, it provides a natural evidential model of A-NPI-M at the singleton level.

In this section, we compare these singleton intervals with those induced by the Imprecise Dirichlet Model (IDM). This comparison is relevant because the IDM has a known and simple belief function representation, whereas the belief function associated with A-NPI-M is constructed to preserve only the singleton bounds. From the viewpoint of uncertainty representation in AI, the comparison clarifies how two data-generated interval models encode different degrees of epistemic caution from the same empirical frequencies. Thus, the comparison is made at the level of the induced belief and plausibility intervals on singletons.

The IDM admits a simple belief function representation [12]. For a fixed value of the IDM parameter $s > 0$, its probability intervals are

$$[l_i^{IDM}, u_i^{IDM}] = \left[\frac{n_i}{N+s}, \frac{n_i+s}{N+s} \right], \quad i = 1, \dots, K.$$

Equivalently, they can be obtained from the basic probability assignment

$$m_{IDM}(\{x_i\}) = \frac{n_i}{N+s}, \quad i = 1, \dots, K,$$

and

$$m_{IDM}(X) = \frac{s}{N+s}.$$

Thus,

$$Bel_{IDM}(\{x_i\}) = \frac{n_i}{N+s}, \quad Pl_{IDM}(\{x_i\}) = \frac{n_i+s}{N+s}.$$

By contrast, A-NPI-M induces the singleton intervals

$$[l_i^A, u_i^A] = \left[\max\left\{ \frac{n_i-1}{N}, 0 \right\}, \min\left\{ \frac{n_i+1}{N}, 1 \right\} \right].$$

As shown in the previous section, these singleton intervals can be reproduced by a suitable basic probability assignment m_A , although the construction is not as simple as in the IDM. In particular, the remaining mass cannot in general be assigned only to the whole frame X , because A-NPI-M does not assign the same interval width to all classes: interior frequencies have width $2/N$, whereas boundary frequencies have width $1/N$.

Since equality between two belief functions would imply equality of their belief and plausibility values on all singletons, the comparison of singleton bounds provides a necessary condition for any possible equality between the IDM belief function and the belief function model associated with A-NPI-M.

5.1. The Case $s = 1$

The value $s = 1$ is one of the most common choices for the IDM. In this case, the IDM intervals are

$$[l_i^{IDM}, u_i^{IDM}] = \left[\frac{n_i}{N+1}, \frac{n_i+1}{N+1} \right].$$

We show that, for $s = 1$, the IDM interval is always strictly contained in the corresponding A-NPI-M interval.

First, assume that $0 < n_i < N$. Then A-NPI-M gives

$$[l_i^A, u_i^A] = \left[\frac{n_i-1}{N}, \frac{n_i+1}{N} \right].$$

For the lower bounds, we have

$$\frac{n_i}{N+1} - \frac{n_i-1}{N} = \frac{N+1-n_i}{N(N+1)}.$$

Since $n_i < N$, it follows that

$$\frac{n_i}{N+1} > \frac{n_i-1}{N}.$$

For the upper bounds,

$$\frac{n_i+1}{N} - \frac{n_i+1}{N+1} = \frac{n_i+1}{N(N+1)} > 0.$$

Therefore,

$$\frac{n_i + 1}{N + 1} < \frac{n_i + 1}{N}.$$

Hence,

$$[l_i^{IDM}, u_i^{IDM}] \subsetneq [l_i^A, u_i^A], \quad 0 < n_i < N.$$

The same strict inclusion holds in the boundary cases. If $n_i = 0$, then

$$[l_i^{IDM}, u_i^{IDM}] = \left[0, \frac{1}{N+1}\right],$$

whereas

$$[l_i^A, u_i^A] = \left[0, \frac{1}{N}\right].$$

Since

$$\frac{1}{N+1} < \frac{1}{N},$$

we obtain

$$[l_i^{IDM}, u_i^{IDM}] \subsetneq [l_i^A, u_i^A].$$

If $n_i = N$, then

$$[l_i^{IDM}, u_i^{IDM}] = \left[\frac{N}{N+1}, 1\right],$$

whereas

$$[l_i^A, u_i^A] = \left[\frac{N-1}{N}, 1\right].$$

Moreover,

$$\frac{N}{N+1} - \frac{N-1}{N} = \frac{1}{N(N+1)} > 0.$$

Thus,

$$\frac{N}{N+1} > \frac{N-1}{N},$$

and again

$$[l_i^{IDM}, u_i^{IDM}] \subsetneq [l_i^A, u_i^A].$$

Consequently, for $s = 1$,

$$[l_i^{IDM}, u_i^{IDM}] \subsetneq [l_i^A, u_i^A], \quad i = 1, \dots, K.$$

Therefore, the credal set induced by the IDM intervals is contained in the credal set defined by the A-NPI-M probability intervals:

$$\mathcal{P}_{IDM}^{s=1} \subsetneq \mathcal{P}_A.$$

Equivalently, A-NPI-M represents a more cautious assessment than the IDM with $s = 1$ at the level of singleton probability intervals. This is relevant when these interval models are used inside learning or decision procedures, because the choice between IDM and A-NPI-M affects the amount of epistemic uncertainty that is propagated from the empirical data to the subsequent reasoning step.

This also implies that the belief function representation of the IDM with $s = 1$ cannot coincide with the belief function model associated with A-NPI-M, because their belief and plausibility intervals on singletons are already different.

5.2. The Case $s = 2$

For $s = 2$, the IDM intervals become

$$[l_i^{IDM}, u_i^{IDM}] = \left[\frac{n_i}{N+2}, \frac{n_i+2}{N+2} \right].$$

In this case, the relationship with A-NPI-M is more delicate. The systematic strict inclusion obtained for $s = 1$ no longer holds for all frequency configurations.

First, consider an interior frequency, $0 < n_i < N$. Equality between the lower bounds would require

$$\frac{n_i}{N+2} = \frac{n_i-1}{N}.$$

This is equivalent to

$$Nn_i = (N+2)(n_i-1),$$

and therefore to

$$n_i = \frac{N+2}{2}.$$

Equality between the upper bounds would require

$$\frac{n_i+2}{N+2} = \frac{n_i+1}{N}.$$

This is equivalent to

$$N(n_i+2) = (N+2)(n_i+1),$$

and therefore to

$$n_i = \frac{N-2}{2}.$$

Both equalities cannot hold simultaneously. Hence, for an interior frequency, the IDM interval with $s = 2$ cannot coincide exactly with the A-NPI-M interval:

$$[l_i^{IDM}, u_i^{IDM}] \neq [l_i^A, u_i^A], \quad 0 < n_i < N.$$

The boundary cases are different. If $n_i = 0$, then

$$[l_i^{IDM}, u_i^{IDM}] = \left[0, \frac{2}{N+2} \right],$$

whereas

$$[l_i^A, u_i^A] = \left[0, \frac{1}{N} \right].$$

These two intervals coincide if and only if

$$\frac{2}{N+2} = \frac{1}{N}.$$

This gives

$$2N = N+2,$$

and hence

$$N = 2.$$

Similarly, if $n_i = N$, then

$$[l_i^{IDM}, u_i^{IDM}] = \left[\frac{N}{N+2}, 1 \right],$$

whereas

$$[l_i^A, u_i^A] = \left[\frac{N-1}{N}, 1 \right].$$

Equality holds if and only if

$$\frac{N}{N+2} = \frac{N-1}{N}.$$

This again gives

$$N = 2.$$

Therefore, for $s = 2$, equality between the IDM and A-NPI-M intervals is possible only in the extreme case $N = 2$, when one class has total frequency. For instance, if

$$(n_1, n_2, \dots, n_K) = (2, 0, \dots, 0),$$

then A-NPI-M gives

$$[l_1^A, u_1^A] = \left[\frac{1}{2}, 1 \right],$$

and, for $j \neq 1$,

$$[l_j^A, u_j^A] = \left[0, \frac{1}{2} \right].$$

The IDM with $s = 2$ gives exactly the same intervals:

$$[l_1^{IDM}, u_1^{IDM}] = \left[\frac{2}{4}, 1 \right] = \left[\frac{1}{2}, 1 \right],$$

and, for $j \neq 1$,

$$[l_j^{IDM}, u_j^{IDM}] = \left[0, \frac{2}{4} \right] = \left[0, \frac{1}{2} \right].$$

Thus,

$$I_{IDM}^{s=2} = I_A$$

in this particular boundary case.

5.3. Interpretation

The previous analysis shows that the relationship between the IDM and A-NPI-M depends strongly on the value of s and on the observed frequency configuration. This is important in AI applications based on local empirical frequencies, because the selected uncertainty model determines how much imprecision is retained before applying an information measure, a decision rule or a learning criterion.

For $s = 1$, the result is especially clear. The IDM intervals are strictly contained in the A-NPI-M intervals for every class, and therefore

$$\mathcal{P}_{IDM}^{s=1} \subsetneq \mathcal{P}_A.$$

Thus, under this common choice of the IDM parameter, A-NPI-M induces a less informative, and therefore more cautious, singleton interval model than the IDM. In particular, the belief function representation of the IDM and the belief function model associated with A-NPI-M cannot coincide at the singleton level.

For $s = 2$, exact equality is not possible for interior frequencies, but it can occur in the extreme boundary case $N = 2$, when all observations are assigned to a single class. Hence, unlike the case $s = 1$, the comparison is more dependent on the frequency configuration.

This comparison also reflects a structural difference between the two models. The IDM assigns a single residual mass $s/(N+s)$ to the whole frame, and therefore all singleton

intervals have the same width. A-NPI-M, by contrast, produces intervals whose width depends on the position of the observed frequency: interior classes have width $2/N$, whereas boundary classes have width $1/N$. This explains why the belief function model associated with A-NPI-M cannot be reduced to the simple IDM construction with mass only on singletons and on the whole frame. Therefore, A-NPI-M cannot be interpreted simply as an IDM with a particular choice of s ; it induces a different evidential structure and a different pattern of caution across observed and boundary classes.

6. Implications for Uncertainty Representation in Data-Driven AI Systems

The construction proposed in this paper has a natural interpretation for artificial intelligence systems that learn, reason or make decisions from finite empirical information. In many AI procedures, empirical frequencies observed in local regions of the data space must be transformed into an uncertainty model before reasoning, learning or decision criteria can be applied. This occurs, for instance, in tree-based methods, rule induction, local classification models, knowledge-based decision systems and other tools where decisions depend on class distributions estimated from finite samples. In such situations, the way in which empirical information is represented determines how much epistemic uncertainty is preserved in the subsequent AI process.

The contribution proposed here should be understood as a representation layer rather than as a complete learning algorithm. Its role is to transform the interval information obtained from A-NPI-M into a belief function, so that tools developed in evidence theory can be applied to data-generated predictive information. In this sense, the construction provides a theoretical basis for uncertainty-aware AI procedures, while empirical validation in full learning systems remains a subject for future work.

The IDM has been useful in this context because its reachable probability intervals admit a direct belief function representation. This makes it possible to apply uncertainty and information measures defined for belief functions to the information generated from data. The construction proposed here plays a similar role for A-NPI-M, although with an important difference: the resulting belief function preserves the A-NPI-M singleton bounds but does not necessarily represent the whole A-NPI-M credal set. Therefore, the proposed model should be understood as an evidence-theoretic representation associated with the marginal predictive information provided by A-NPI-M.

A direct application of the proposed model arises in credal classification and decision tree induction. In these methods, local class frequencies must be converted into an uncertainty model before computing split criteria, uncertainty indices or decision rules. The belief function associated with A-NPI-M provides an evidence-theoretic alternative to the IDM representation, preserving the non-parametric character of NPI-M while enabling the use of belief, plausibility and evidence-based uncertainty measures. In this way, empirical predictive information can be used within AI tools that require structured representations of imperfect knowledge rather than precise probability vectors.

This distinction is relevant when measures from evidence theory are used. Such measures quantify the uncertainty of the constructed belief function, not necessarily the uncertainty of the full A-NPI-M credal set. Nevertheless, the construction provides a coherent way to transfer the marginal information generated by A-NPI-M into the framework of evidence theory. This can be useful when different uncertainty models generated from data must be compared through belief-function tools, or when uncertainty indices developed in evidence theory are required in learning, reasoning or decision-support procedures.

Other data-driven evidential constructions are possible. For example, belief functions may be obtained from confidence intervals, bootstrap procedures, random sets or other

statistical models of evidence. These approaches are valuable, but they usually depend on additional modelling choices such as a confidence level, a resampling mechanism or a particular statistical assumption. The present construction has a different purpose: it is specifically tied to A-NPI-M and preserves exactly the singleton bounds induced by its reachable probability intervals. Therefore, it should be regarded as complementary to other evidential models from data, rather than as a universally preferable alternative.

Table 1 summarizes this position. The proposed model should be read as an evidential model associated with A-NPI-M at the singleton level. It preserves the same lower and upper probabilities for elementary classes while adding an evidential structure on composite events. From the viewpoint of data-driven AI, this provides a formal mechanism for transforming empirical predictive information into a structured representation of imperfect knowledge that can be used in reasoning, learning and decision-support tasks.

Table 1. Uncertainty models generated from data considered in this work.

Model	Parameter-Free	Representation	Information Preserved	Relation with Evidence Theory
IDM	No	Reachable intervals	IDM singleton bounds for a selected value of s	Exact belief function representation
NPI-M	Yes	Bounds defined on events	Predictive lower and upper probabilities on events	Does not define a credal set in general
A-NPI-M	Yes	Singleton reachable intervals	NPI-M singleton lower and upper probabilities	No exact belief function representation in general
Proposed m_A	Yes	Basic probability assignment	A-NPI-M singleton bounds	Evidence-theoretic model preserving singleton information

The proposed construction is therefore not intended to replace the original A-NPI-M model. Rather, it provides an evidence-theoretic layer associated with its singleton predictive bounds. This layer makes it possible to use tools from evidence theory while keeping the parameter-free, data-based nature of NPI-M at the level of elementary classes. This is especially relevant for AI systems in which uncertainty models must be obtained directly from finite data and then used in computational procedures that require tractable representations.

6.1. Computational Cost and Scalability of the Construction

A practical point concerns the size of the resulting belief function. General belief-function models may require a number of focal elements that grows exponentially with the number of classes. This is not the case for the construction proposed here. If all frequencies are interior, the model contains K singleton focal elements and K pairwise focal elements, and therefore at most $2K$ focal elements. If some classes are unobserved and c classes are observed, the construction contains c singleton focal elements and c focal elements that are not singletons, so the number of focal elements is at most $2K$. Finally, if one class has total frequency, only two focal elements are needed.

Thus, the number of focal elements grows linearly with the number of classes. The storage requirement of the proposed basic probability assignment is therefore $O(K)$, and the construction can be performed in $O(K)$ time once the class frequencies have been computed. The computation of the singleton belief and plausibility values is also linear. More demanding computations may arise when optimizing particular uncertainty measures over the compatible credal set, but this cost depends on the selected measure and on the optimization procedure rather than on the construction of m_A itself. This observation is important because it shows that the proposed evidential representation avoids the exponential growth usually associated with unrestricted belief-function models.

6.2. Illustrative Use in a Small Classification Setting

This subsection gives a small illustrative example of how the proposed representation can be used inside a simple classification pipeline. The purpose is not to provide a full experimental validation but to show how the A-NPI-M singleton intervals can be transformed into a belief function and then evaluated by an uncertainty measure from evidence theory.

In order to use a total uncertainty measure that is directly connected with credal sets, we consider the maximum of Shannon entropy over the set of probability distributions compatible with a belief function:

$$H^*(m) = \max_{P \in \mathcal{P}(m)} \left\{ - \sum_{i=1}^K p_i \log p_i \right\}.$$

This quantity has been used as a total uncertainty measure for credal sets and belief functions. In Dempster–Shafer theory, maximum entropy is also one of the most established total uncertainty measures, since it satisfies the essential requirements proposed for this type of measure [23]. It is especially appropriate here because both the IDM representation and the proposed A-NPI-M-based construction induce sets of compatible probability distributions [13,22]. It therefore allows both models to be compared with the same uncertainty criterion.

Consider a three-class problem with frame $X = \{x_1, x_2, x_3\}$. Suppose that two candidate binary splits of a node produce the following local class-frequency vectors:

$$S_1 : (5, 5, 0), (5, 0, 5), \quad S_2 : (4, 3, 3), (3, 4, 3),$$

where each child node contains $N = 10$ observations. We compare the IDM belief function with $s = 1$ and the proposed A-NPI-M-based belief function m_A by computing $H^*(m)$ for each child node.

For the IDM with $s = 1$, the frequency vector $(5, 5, 0)$ induces the lower bounds $p_1 \geq 5/11$, $p_2 \geq 5/11$ and $p_3 \geq 0$. The maximum entropy distribution is therefore $(5/11, 5/11, 1/11)$, giving

$$H^*(m_{IDM}) \simeq 0.9348.$$

The same value is obtained for the symmetric frequency vector $(5, 0, 5)$. For the frequency vector $(4, 3, 3)$, the maximum entropy distribution is $(4/11, 7/22, 7/22)$, and hence

$$H^*(m_{IDM}) \simeq 1.0966.$$

The same value is obtained for $(3, 4, 3)$ by symmetry.

For the proposed A-NPI-M-based belief function, the vector $(5, 5, 0)$ has one unobserved class. The construction preserves the singleton bounds and induces the maximum entropy distribution $(0.45, 0.45, 0.10)$, giving

$$H^*(m_A) \simeq 0.9489.$$

For the vector $(4, 3, 3)$, all frequencies are interior. In this case, the uniform distribution $(1/3, 1/3, 1/3)$ belongs to the credal set associated with the constructed belief function, and therefore

$$H^*(m_A) = \log 3 \simeq 1.0986.$$

The symmetric frequency vectors in the second child nodes give the same values.

Table 2 summarizes the maximum-entropy values obtained for the two candidate splits under the IDM representation and under the proposed A-NPI-M-based belief function.

Table 2. Illustrative comparison of IDM and the proposed A-NPI-M-based belief function using maximum entropy as a total uncertainty measure.

Split	Child Frequencies	$H^*(m_{IDM}), s = 1$	$H^*(m_A)$	Interpretation
S_1	(5,5,0) and (5,0,5)	0.9348 in each child	0.9489 in each child	One unobserved class in each child
S_2	(4,3,3) and (3,4,3)	1.0966 in each child	1.0986 in each child	All frequencies are interior

The example shows how the proposed construction can be inserted into a simple decision-tree-style evaluation. In this particular illustration, both representations assign lower total uncertainty to split S_1 than to split S_2 , but the numerical uncertainty values are not the same. This demonstrates that the same empirical frequencies can lead to different total uncertainty assessments once the A-NPI-M interval information is represented as a belief function. The example is not intended to show that the proposed model is better than the IDM in predictive terms. Rather, it illustrates that the proposed construction provides a way to apply well-established evidence-theoretic and credal-set uncertainty measures to data-generated reachable probability intervals.

7. Conclusions

This paper studied the relationship between A-NPI-M and belief functions in the context of uncertainty representation for artificial intelligence systems based on finite empirical information. A-NPI-M is obtained from NPI-M by preserving its singleton lower and upper probabilities and representing them through a credal set defined by reachable probability intervals. From a data-driven AI perspective, these intervals can be interpreted as a structured representation of the predictive information obtained from the observed data, without introducing externally chosen cautiousness parameters. This point is important because information extracted from databases is rarely perfectly reliable or complete. When empirical information is affected by noise, scarcity or partial reliability, representing it by a single probability vector may hide relevant epistemic uncertainty. Models based on probability intervals, credal sets and belief functions provide a more faithful mathematical representation of this imprecision.

The revised manuscript also emphasizes that the proposed construction is not an empirically validated classifier but a theoretical connection that enables the use of well-studied evidence-theoretic uncertainty measures on information originally represented by A-NPI-M reachable probability intervals. The computational discussion added in the revision also shows that the proposed basic probability assignment is sparse, since the number of focal elements grows linearly with the number of classes.

Since reachable probability intervals and belief functions are not equivalent formalisms in general, the credal set defined by the A-NPI-M probability intervals cannot be expected to admit an exact belief function representation in all cases. We first illustrated this limitation by showing an example where the A-NPI-M probability intervals do not lead to a valid basic probability assignment with the same set of compatible probability distributions. At the same time, particular cases exist where such an exact equivalence is possible. This distinction is important because it clarifies that the connection between A-NPI-M and evidence theory must be formulated with care.

The main contribution of the paper is the construction of a basic probability assignment whose induced belief and plausibility values reproduce the A-NPI-M bounds on singleton events. The construction is given by cases, depending on the observed frequency configuration. When all frequencies are interior, the remaining mass is assigned to pairs of classes arranged in a cycle. When unobserved classes are present, they must appear in only one focal element that is not a singleton so that their plausibility increases by $1/N$

rather than by $2/N$. Finally, when one class has total frequency, the construction reduces to a simple mass assignment with one singleton focal element and the whole frame.

The resulting belief function should therefore be understood as an evidential model associated with A-NPI-M at the singleton level. It preserves exactly the lower and upper probabilities assigned to the elementary classes, but it may impose additional constraints on composite events. Consequently, the associated set of compatible probability distributions is included in the credal set defined by the A-NPI-M probability intervals, with equality only in particular cases. Thus, the proposed construction does not replace the original A-NPI-M model. Rather, it provides an evidence-theoretic layer associated with its singleton predictive bounds.

The paper also compared A-NPI-M with the IDM. The IDM has a known and simple belief function representation, where the remaining mass is assigned to the whole frame. In contrast, A-NPI-M requires the remaining mass to be allocated over focal elements that are not singletons. This allocation depends on the frequency configuration, because interior and boundary frequencies induce intervals of different widths. For the usual value $s = 1$, the IDM singleton intervals are strictly contained in the A-NPI-M intervals for every frequency configuration. Hence, A-NPI-M represents a systematically more cautious singleton interval model under this setting. For $s = 2$, equality is not possible for interior frequencies, but it can occur in the special boundary case $N = 2$, when all observations belong to a single class.

These results provide a bridge between non-parametric predictive inference, reachable probability intervals and evidence theory. From the viewpoint of artificial intelligence, the proposed construction offers a formal mechanism for transforming finite empirical information into an evidence-theoretic representation of imperfect knowledge. This is relevant for data-driven AI tools where local empirical frequencies must be converted into uncertainty models before applying reasoning, learning or decision criteria, as occurs in credal classification, decision tree induction and knowledge-based decision systems.

The connection with evidence theory is also important from the point of view of information quantification. Once the A-NPI-M singleton information has been represented through a belief function, uncertainty and information measures developed for evidence theory can be applied to the resulting model. Such measures have been widely studied during the last few decades and provide tools for quantifying nonspecificity, conflict, total uncertainty or other aspects of imperfect information. Therefore, the proposed construction makes it possible to analyze data-generated predictive information from A-NPI-M using the well-established machinery of evidence-theoretic uncertainty measures. However, these measures should be interpreted as quantifying the uncertainty of the constructed belief function, not necessarily that of the full A-NPI-M credal set.

Several directions for future work follow from this study. The first direction is to quantify the amount of information lost, or the additional constraints introduced, when passing from the A-NPI-M credal set to the associated belief function model. A second direction is to analyze which uncertainty measures are preserved, approximated or systematically modified by this transformation. This is especially relevant for maximum entropy, nonspecificity measures and other information measures used in evidence theory. A third direction is to study the use of the proposed evidential model in AI procedures such as credal classification, decision tree induction, rule learning and decision-support systems. In these applications, the model could provide a parameter-free alternative for representing local predictive information while allowing for the use of evidence-theoretic uncertainty indices. A further important direction is the empirical evaluation of the proposed evidential representation in complete learning procedures. In particular, future work should implement the proposed A-NPI-M-based belief function and the IDM belief function within

a common classification framework, such as credal decision trees or rule-based classifiers, and compare their predictive behaviour on standard benchmark datasets. Special attention should be paid to data-scarce and imbalanced settings, including medical, credit-scoring or other UCI-type datasets, where empirical class frequencies may be limited or only partially reliable. Such experiments would make it possible to evaluate whether the additional epistemic cautiousness induced by the A-NPI-M-based representation leads to measurable differences in split selection, uncertainty calibration, predictive performance or decision robustness. Additional computational experiments may also study runtime and memory consumption for increasing numbers of classes and larger classification problems.

Author Contributions: Conceptualization, M.I.A.B., S.M.-G. and J.A.; methodology, M.I.A.B., S.M.-G. and J.A.; formal analysis, M.I.A.B., S.M.-G. and J.A.; writing—original draft preparation, M.I.A.B., S.M.-G. and J.A.; writing—review and editing, M.I.A.B., S.M.-G. and J.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by grant PID2024-159012NA-I00, funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No real datasets were created or analyzed in this study. The synthetic frequency configurations used in the illustrative examples are fully reported in the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
A-NPI-M	Approximated Non-Parametric Predictive Inference for Multinomial data
BPA	Basic Probability Assignment
IDM	Imprecise Dirichlet Model
NPI-M	Non-Parametric Predictive Inference for Multinomial data

References

- Walley, P. *Statistical Reasoning with Imprecise Probabilities*; Chapman & Hall/CRC Monographs on Statistics & Applied Probability; Taylor & Francis: Abingdon, UK, 1991; Volume 42.
- Walley, P. Inferences from multinomial data; learning about a bag of marbles (with discussion). *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 3–57. [[CrossRef](#)]
- Bikakis, A.; Caire, P. Contextual and Possibilistic Reasoning for Coalition Formation. *AI* **2020**, *1*, 389–417. [[CrossRef](#)]
- Erhan, C.; Ure, N.K. Reducing Annotation Effort in Semantic Segmentation Through Conformal Risk Controlled Active Learning. *AI* **2025**, *6*, 270. [[CrossRef](#)]
- Dempster, A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Ann. Math. Stat.* **1967**, *38*, 325–339. [[CrossRef](#)]
- Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976.
- De Campos, L.M.; Huete, J.F.; Moral, S. Probability intervals: A tool for uncertain reasoning. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **1994**, *02*, 167–196. [[CrossRef](#)]
- Zhao, K.; Wang, X.; Deng, Y. A Survey: Optimization and Applications of Evidence Fusion Algorithms Based on Dempster–Shafer Theory. *Appl. Soft Comput.* **2022**, *124*, 109075. [[CrossRef](#)]
- Bellenger, A.; Gatepaille, S. Dempster–Shafer Theory for Data Fusion Applications. *arXiv* **2011**, arXiv:1106.3876.
- Bezerra, E.D.C.; Teles, A.S.; Coutinho, L.R.; da Silva e Silva, F.J. Dempster–Shafer Theory for Modeling and Treating Uncertainty in IoT Applications Based on Complex Event Processing. *Sensors* **2021**, *21*, 1863. [[CrossRef](#)] [[PubMed](#)]
- Song, C.; Pons, A.; Yen, K. Sieve: An Ensemble Algorithm Using Global Consensus for Binary Classification. *AI* **2020**, *1*, 242–262. [[CrossRef](#)]
- Abellán, J. Uncertainty measures on probability intervals from the imprecise Dirichlet model. *Int. J. General. Syst.* **2006**, *35*, 509–528. [[CrossRef](#)]

13. Bernard, J.M. An introduction to the imprecise Dirichlet model for multinomial data. *Int. J. Approx. Reason.* **2005**, *39*, 123–150. [[CrossRef](#)]
14. Abellán, J.; Moral, S. Building Classification Trees Using the Total Uncertainty Criterion. *Int. J. Intell. Syst.* **2003**, *18*, 1215–1225. [[CrossRef](#)]
15. Coolen, F.P.A.; Augustin, T. Learning from multinomial data: A nonparametric predictive alternative to the Imprecise Dirichlet Model. In *Proceedings of the ISIPTA'05, Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications, Pittsburgh, PA, USA, 20–23 July 2005*; Cozman, F.G., Nau, R., Seidenfeld, T., Eds.; Carnegie Mellon University: Pittsburgh, PA, USA, 2005; pp. 125–134.
16. Coolen, F.; Augustin, T. A nonparametric predictive alternative to the Imprecise Dirichlet Model: The case of a known number of categories. *Int. J. Approx. Reason.* **2009**, *50*, 217–230. [[CrossRef](#)]
17. Abellán, J.; Baker, R.M.; Coolen, F.P.; Crossman, R.J.; Masegosa, A.R. Classification with decision trees from a nonparametric predictive inference perspective. *Comput. Stat. Data Anal.* **2014**, *71*, 789–802. [[CrossRef](#)]
18. Klir, G.J. *Uncertainty and Information: Foundations of Generalized Information Theory*; John Wiley And Sons, Inc.: Hoboken, NJ, USA, 2006. [[CrossRef](#)]
19. Abellán, J.; Castellano, J.G. Improving the Naive Bayes Classifier via a Quick Variable Selection Method Using Maximum of Entropy. *Entropy* **2017**, *19*, 247. [[CrossRef](#)]
20. Abellán, J. Ensembles of decision trees based on imprecise probabilities and uncertainty measures. *Inf. Fusion* **2013**, *14*, 423–430. [[CrossRef](#)]
21. Abellán, J.; Mantas, C.J. Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. *Expert Syst. Appl.* **2014**, *41*, 3825–3830. [[CrossRef](#)]
22. Abellán, J.; Baker, R.M.; Coolen, F.P. Maximising entropy on the nonparametric predictive inference model for multinomial data. *Eur. J. Oper. Res.* **2011**, *212*, 112–122. [[CrossRef](#)]
23. Abellán, J.; Masegosa, A. Requirements for total uncertainty measures in Dempster-Shafer theory of evidence. *Int. J. General. Syst.* **2008**, *37*, 733–747. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.