# Construction of predictive belief functions using a frequentist approach

**Thierry Denœux**
UMR CNRS 6599 Heudiasyc
Université de Technologie de Compiègne
BP 20529 - F-60205 Compiègne cedex - France
Thierry.Denoeux@hds.utc.fr

## Abstract

This paper addresses the problem of building belief functions from statistical data. We describe a method for quantifying an agent belief about the realization of a discrete random variable $X$ with unknown probability distribution $\mathbb{P}_X$, having observed a realization of an iid random sample with the same distribution. The proposed solution verifies two "reasonable" properties with respect to $\mathbb{P}_X$: it is less committed than $\mathbb{P}_X$ with some user-defined probability, and it converges towards $\mathbb{P}_X$ in probability as the size of the sample tends to infinity.

**Keywords:** Dempster-Shafer Theory, Evidence Theory, Transferable Belief Model, Multinomial Proportions, Confidence Intervals

## 1 Introduction

Since its foundation in the late 1960's and in the 1970's [2, 14, 18], the Dempster-Shafer theory of belief functions has been widely used as a conceptual framework for modeling partial knowledge and reasoning under uncertainty. In most applications, a significant part of the available information comes from statistical data, and it is crucial to be able to model such information in the belief functions framework. The first application of belief functions was indeed statistical inference about parametric models [2][3][4]. Shafer [15] describes several distinct approaches to this problem, among which the approach initially proposed by Dempster, based on pivotal quantities, the likelihood-based approach exposed in Shafer's book [14], and Smets' method based on the GBT. Principles of statistical inference within the theory of Hints, an interpretation of Dempster-Shafer theory closely related to Dempster's model, are exposed in [11, chapter 9].

The specific problem addressed in this paper is the following. We consider a population $\Omega$, each element $\omega$ of which is described by a discrete observable characteristic $x \in \mathcal{X} = \{\xi_1, \ldots, \xi_K\}$. Individuals are randomly sampled from $\Omega$ according to some probability measure $\mu$. The mapping $X : \omega \to x$ is thus a random variable, with unknown probability distribution $\mathbb{P}_X$ defined by $\mathbb{P}_X(A) = \mu(X^{-1}(A))$ for all $A \subseteq \mathcal{X}$. Having drawn $n$ elements with replacement (or without replacement in the case of an infinite population), we have observed a realization $x_1, \ldots, x_n$ of a random iid sample $X_1, \ldots, X_n$ with parent distribution $\mathbb{P}_X$. We want to assess our degrees of belief concerning the realization of $X$ that will be observed when we shortly draw an additional individual from $\Omega$. A classical paradigm for this problem is that of an urn containing balls of different colors. Having observed the colors of $n$ balls randomly taken from the urn with replacement, we want to assess our beliefs concerning the color of the next ball.

The conceptual framework adopted in this paper will be based on the Transferable Belief Model (TBM), a nonprobabilistic, subjectivist interpretation of the Dempster-Shafer theory of belief functions [18]. In this model, a belief function is not interpreted as the lower envelope of a family of probability distributions, a view that is known to be incompatible with Dempster's rule of combination and, consequently, with the TBM [16]. We shall, however, for the particular problem at hand, define a certain form of consistency between the belief function of interest and a lower probability measure, as will be shown below. Basic knowledge of the mathematics of belief functions and their interpretation in the TBM will be assumed in this paper. The reader is referred to Shafer's book [14] and presentations of the TBM [18] for complete coverings of these topics.

The problem of inference from binomial and, more generally, multinomial data was originally addressed in the belief function framework by Dempster [2, 4]. Dempster's solution was later recovered by Kohlas in the Hint Theory framework (an interpretation of Dempster-Shafer theory close to Dempster's model), and in the TBM framework [17]. This solution will first be briefly recalled in Section 2. Our method will then be introduced in Section 3, and an approximate analytical solution for the case of ordered data will be presented in Section 4. Finally, Section 5 will conclude the paper.

## 2 Review of previous work

### 2.1 Dempster's approach

Belief functions were introduced by Dempster as part of a statistical inference framework proposed as an alternative to Bayesian methods and to Fisher's fiducial method [2]. One of the first applications of this new approach concerned binomial sampling with a continuous parameter $p$, and general multinomial sampling with a finite number of contemplated hypotheses [2].

Here, only the main results concerning the binomial sampling model will first be summarized. Let $X_1, \dots, X_n$ be an iid sample with parent variable $X \in \mathcal{X} = \{0, 1\}$ following a Bernoulli distribution with parameter $p \in \mathcal{P} = [0, 1]$. A random variable $W_i$ uniformly distributed on $\mathcal{W} = [0, 1]$ is supposed to underlie each observation $X_i$, with

$$X_i = 1 \Leftrightarrow W_i \leq p \; . \qquad (1)$$

The uniform distribution of $W_i$ can be thought of as modeling random sampling from an infinite population assimilated to the interval $[0, 1]$.

Equation (1) defines a multivalued mapping from $\mathcal{W}$ to $\mathcal{X} \times \mathcal{P}$, which maps any $w \in \mathcal{W}$ to $\{1\} \times [w, 1] \cup \{0\} \times [0, w]$. This mapping constrains the possible values of the triplet $(X_i, W_i, p)$, and can alternatively be represented by a logical belief function (i.e., a belief function with a single focal set) $m_i^{\Theta}$ on the joint space $\Theta = \mathcal{X} \times \mathcal{W} \times \mathcal{P}$. Now, the uniform probability distribution of $W_i$ defines a Bayesian belief function $m_i^{\mathcal{W}}$. Having observed a realization $x_i$ of each $X_i$, a belief function on $\mathcal{P}$ can be obtained by combining each belief function in the model using Dempster's rule, conditioning by $X_i = x_i$ for each $i = 1, \dots, n$, and marginalizing on $p$.

The prediction problem can then be handled by defining a new variable $X \sim \mathcal{B}(p)$, with associated uniform random variable $W$. The marginal bba induced about $X$ is:

$$m^{\mathcal{X}}(\{1\}) = \frac{N}{n+1} = \frac{\widehat{p}}{1 + 1/n} \qquad (2)$$

$$m^{\mathcal{X}}(\{0\}) = \frac{n-N}{n+1} = \frac{1-\widehat{p}}{1 + 1/n} \qquad (3)$$

$$m^{\mathcal{X}}(\mathcal{X}) = \frac{1}{n+1}, \qquad (4)$$

where $N = \sum_{i=1}^{n} x_i$ and $\widehat{p} = N/n$.

In principle, the above approach can be extended to the general multinomial case. However, the calculations are now much more complex. Dempster studied the trinomial case in [4] (without providing explicitly the equivalent of (2)-(4)), and he presented some results pertaining to the general case in [5]. However, the application of these results to compute the marginal belief function of $X$ has proved, so far and to our knowledge, mathematically intractable.

### 2.2 Discussion

The approach outlined above seems well founded theoretically. It provides a usable solution at least in the binomial case, and maybe for $K = 3$ (although this solution does not seem to have been fully worked out in that case). However, this approach seems to become quickly analytically intractable for larger $K$, essentially because of the difficulty to manipulate belief functions over continuous multidimensional spaces.

The approach proposed in this paper follows a completely different route. First of all, our objective will be more limited, in that we shall only attempt to build a belief function regarding a future observation $X$, given past observations $X_1, \ldots, X_n$, without expliciting our beliefs on $\mathcal{P}$. Hence, belief functions will not be used as a tool for parametric inference (for which frequentist confidence regions will be employed), but as a tool for prediction.

As mentioned aboved, another feature of our approach is that it will be essentially based on frequentist analysis. Given an iid random sample $\mathbf{X}_n = (X_1, \ldots, X_n)$ with parent probability distribution $\mathbb{P}_X$, we want to produce a belief function on $\mathcal{X}$, noted $bel^{\mathcal{X}}[\mathbf{X}_n]$, in such a way that the inequality $bel^{\mathcal{X}}[\mathbf{X}_n] \leq \mathbb{P}_X$ will hold in the long run at least in $100(1-\alpha)$ % of cases (i.e., for a fraction $100(1-\alpha)$ of the samples). For a given realization $\mathbf{x}_n = (x_1, \ldots, x_n)$, we shall thus obtain a belief function $bel^{\mathcal{X}}[\mathbf{x}_n]$, which will be guaranteed to have been obtained by a method yielding a belief function less committed than the probability measure $\mathbb{P}_X$ in $100(1-\alpha)$ % of cases. As will be shown below, such a belief function can easily be computed from multinomial confidence regions, and it has a simple interpretation.

## 3 New approach

### 3.1 Basic principles

Let us assume that we have an urn with balls of different colors, noted $\xi_1, \ldots, \xi_K$. Let $X$ denote the color of a ball taken randomly from the urn. As before, the probability distribution of $X$ is noted $\mathbb{P}_X$. For each $A \subseteq \mathcal{X} = \{\xi_1, \ldots, \xi_K\}$, $\mathbb{P}_X(A)$ represents the long run frequency of the event $X \in A$, which is simply equal in this example to the proportion of balls with color in $A$ contained in the urn. This quantity is constant (it depends only on the experimental setting), but unknown.

Assume that we will shortly draw a ball from this urn, and we want to model our beliefs regarding its color by a belief function $bel^{\mathcal{X}}$. If we know the composition of the urn, and hence the underlying long run frequency distribution $\mathbb{P}_X$, it is reasonable to postulate $bel^{\mathcal{X}} = \mathbb{P}_X$. As remarked by Hacking [9], this "frequency principle" seems very natural.

Let us now assume that we do not know the composition of the urn, but we have drawn $n$ balls with replacement. We have thus observed a realization of an iid random sample $\mathbf{X}_n = (X_1, \ldots, X_n)$, with parent distribution $\mathbb{P}_X$. Let $bel^{\mathcal{X}}[\mathbf{X}_n]$ denote a belief function constructed using $\mathbf{X}_n$. Which properties should be satisfied by $bel^{\mathcal{X}}[\mathbf{X}_n]$ ?

First, it seems natural to impose that $bel^{\mathcal{X}}[\mathbf{X}_n]$ become closer to $\mathbb{P}_X$ as $n \to \infty$, which can be seen as a weak form of Hacking's frequency principle. Loosely speaking, a sample of infinite size is equivalent to knowledge of the distribution of $X$, hence the belief function should asymptotically become identical to $\mathbb{P}_X$. This translates to the following requirement:

**Requirement** $R_1$**:** For all $A \subseteq \mathcal{X}$,

$$bel^{\mathcal{X}}[\mathbf{X}_n](A) \xrightarrow{P} \mathbb{P}_X(A), \text{ as } n \to \infty, \quad (5)$$

For finite $n$, what kind of relationship should be imposed between $bel^{\mathcal{X}}[\mathbf{X}_n]$ and $\mathbb{P}_X$ ? Since we have less information than in the asymptotic case, it seems natural to impose that $bel^{\mathcal{X}}[\mathbf{X}_n]$ be *less committed* than $\mathbb{P}_X$, as a consequence of the Least Commitment Principle [18]. We should then have $bel^{\mathcal{X}}[\mathbf{X}_n] \leq \mathbb{P}_X$. This requirement, however, appears to be much too stringent. Having observed a positive count $n_k$ for a certain value $\xi_k$ of $X$, we can rule out 0 as a possible value for $p_k$, but any arbitrarily small value $\epsilon$ remains possible, unlikely as it may be. The above requirement would then lead to $bel^{\mathcal{X}}[\mathbf{X}_n](A) = 0$, for any strict subset $A$ of $\mathcal{X}$, i.e., to the vacuous belief function.

As a less stringent requirement, we propose to impose that the inequality $bel^{\mathcal{X}}[\mathbf{X}_n] \leq \mathbb{P}_X$ be satisfied only "in most cases". Assuming that the random experiment that consists of drawing $n$ balls from the urn is repeated indefinitely, we would like $bel^{\mathcal{X}}[\mathbf{X}_n]$ to be less committed than $\mathbb{P}$ "most of the time", i.e. with at least some prescribed long run frequency $1 - \alpha$, where $\alpha \in (0, 1)$ is an arbitrarily small positive number. More formally, this can be expressed by the following second requirement:

**Requirement $R_2$:**

$$\mathbb{P}(bel^{\mathcal{X}}[\mathbf{X}_n] \leq \mathbb{P}_X) \geq 1 - \alpha. \tag{6}$$

Equation (6) can alternatively be written:

$$\mathbb{P}\left(bel^{\mathcal{X}}[\mathbf{X}_n](A) \leq \mathbb{P}_X(A), \forall A \subset \mathcal{X}\right) \geq 1 - \alpha.$$

It should be quite clear that, in this expression, as in (5), $\mathbb{P}_X$ denotes the true probability distribution of $X$, which is constant but unknown. The quantity $bel^{\mathcal{X}}[\mathbf{X}_n](A)$ is random, as it is a function of the random sample $\mathbf{X}_n$.

A belief function satisfying requirements $R_1$ and $R_2$ will be called a *predictive belief function at confidence level* $1 - \alpha$.

In the following, we shall examine methods for deriving such belief functions from multinomial confidence regions. Some definitions and results regarding these confidence regions will first be recalled in the following section.

### 3.2 Multinomial confidence regions

The main building block of our approach to constructing belief functions is composed of methods for building confidence regions on multinomial parameters. Given an iid sample $X_1, \ldots, X_n$ of a discrete random variable $X$ taking values in $\mathcal{X} = \{\xi_1, \ldots, \xi_K\}$, let $N_k = \sum_{i=1}^n 1_{\xi_k}(X_i)$ denote the number of observations in category $\xi_k$. The random vector $\mathbf{N} = (N_1, \ldots, N_K)$ has a multinomial distribution with parameters $n$ and $\mathbf{p} = (p_1, \ldots, p_K)$, with $p_k = \mathbb{P}_X(\{\xi_k\})$.

Let $S(\mathbf{N})$ be a random subset of the parameter space $\mathcal{P} = \{\mathbf{p} = (p_1, \ldots, p_K) \in [0, 1]^K \mid \sum_{k=1}^K p_k = 1\}$. $S(\mathbf{N})$ is said to be a confidence region for $\mathbf{p}$ at confidence level $1 - \alpha$, if

$$\mathbb{P}(S(\mathbf{N}) \ni \mathbf{p}) \geq 1 - \alpha,$$

i.e., the random region $S(\mathbf{N})$ contains the constant parameter vector $\mathbf{p}$ with probability (long-run frequency) $1 - \alpha$. It is an asymptotic confidence region if the above inequality only holds in the limit as $n \to \infty$.

The problem of finding confidence regions for multinomial proportions has received considerable attention in the statistical literature from the 1960's [13] [8] up to these days [7]. Of particular interest are simultaneous confidence intervals, i.e., regions defined as a Cartesian product of intervals:

$$S(\mathbf{N}) = [P_1^-, P_1^+] \times \ldots \times [P_K^-, P_K^+],$$

which have easy interpretation. Such asymptotic confidence regions were proposed by Quesenberry and Hurst [13], and Goodman [8]. The first solution is defined as:

$$P_k^- = \frac{a + 2N_k - \sqrt{\Delta_k}}{2(n + a)} \tag{7}$$

$$P_k^+ = \frac{a + 2N_k + \sqrt{\Delta_k}}{2(n + a)}, \tag{8}$$

where $a$ is the quantile of order $1-\alpha$ of the chi-square distribution with one degree of freedom, and $\Delta_k = a\left(a + \frac{4N_k(n-N_k)}{n}\right)$. It can easily be checked that the classical confidence interval on binomial $p$ is recovered as a special case when $K = 2$. For $K > 2$, Goodman remarked that the above confidence region is too conservative, and showed that $a$ could be replaced by $b$, the quantile of order $1-\alpha/K$ of the chi-square distribution with one degree of freedom. Although more sophisticated methods may yield smaller regions, particularly for small sample sizes, Goodman's intervals have been found to be good enough in most practical applications [12].

### 3.3 From multinomial confidence regions to lower probabilities

A confidence region $S(\mathbf{N})$ for multinomial proportions such as reviewed in Section 3.2 is usually interpreted as defining a set of plausible values for the vector parameter $\mathbf{p}$. However, since each value of $\mathbf{p}$ specifies a unique probability measure of $\mathcal{X}$, it is clear that $S(\mathbf{N})$ can equivalently be seen as defining a family of probability measures. To keep the notation as simple as possible, the same symbol $S(\mathbf{N})$ will be used to denote both the set of parameter values $\mathbf{p}$ and the set of probability measures. Let $P^-$ and $P^+$ denote, respectively, the lower and upper envelopes of $S(\mathbf{N})$, defined as $P^-(A) = \min_{P \in S(\mathbf{N})} P(A)$ and $P^+(A) = \max_{P \in S(\mathbf{N})} P(A)$. They can be easily computed using the following proposition.

PROPOSITION 1 *For all strict nonempty subset $A$ of $\mathcal{X}$,*

$$P^-(A) = \max\left(\sum_{\xi_k \in A} P_k^-, 1 - \sum_{\xi_k \notin A} P_k^+\right) \quad (9)$$

$$P^+(A) = \min\left(\sum_{\xi_k \in A} P_k^+, 1 - \sum_{\xi_k \notin A} P_k^-\right). \quad (10)$$

*Proof.* The proof of this propositions, as well as other proofs of the main results presented here, can be found in a long version of this paper [6].

Note that we have, as a direct consequence of Proposition 1, $P^+(A) = 1 - P^-(\overline{A}), \quad \forall A \subseteq \mathcal{X}$. Hence, the lower probability measure $P^-$ is sufficient to characterize $S(\mathbf{N})$:

$$S(\mathbf{N}) = \{P \mid P^- \leq P\}.$$

By construction, we have

$$\mathbb{P}(\mathbb{P}_X \in S(\mathbf{N})) = \mathbb{P}(P^- \leq \mathbb{P}_X) \geq 1 - \alpha, \quad (11)$$

and it is clear that $P^-(A) \xrightarrow{P} \mathbb{P}_X(A)$ as $n \to \infty$, for all $A \subseteq \mathcal{X}$. Hence, $P^-$ verifies our two requirements $R_1$ and $R_2$. Unfortunately, $P^-$ is not, in general, a belief function when $K > 3$ (a counterexample can easily be found). However, it is a 2-monotone capacity (a proof is given in [6]).

### 3.4 From lower probabilities to predictive belief functions

**The case $K = 2$**

When $K = 2$, the lower probability measure $P^-$ defined above is actually a belief function. Its bba is simply equal to:

$$\begin{aligned} m^{\mathcal{X}}(\{\xi_1\}) &= P_1^- \\ m^{\mathcal{X}}(\{\xi_2\}) &= P_2^- \\ m^{\mathcal{X}}(\mathcal{X}) &= 1 - P_1^- - P_2^-, \end{aligned}$$

with $P_1^-$ and $P_2^-$ defined by (7). If we note $\widehat{p} = N_1/n$, it is easy to show that:

$$\begin{aligned} m^{\mathcal{X}}(\{\xi_1\}) &\sim \widehat{p} - u_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \\ m^{\mathcal{X}}(\{\xi_2\}) &\sim 1 - \widehat{p} - u_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \\ m^{\mathcal{X}}(\mathcal{X}) &\sim 2u_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \end{aligned}$$

where $\sim$ denotes asymptotic equivalence. It is interesting to compare these expressions with (2)-(4). We can see that the mass $m^{\mathcal{X}}(\mathcal{X})$ tends towards 0 as $n^{-1/2}$ in our approach, whereas it has the higher convergence rate of $n^{-1}$ in Dempster's solution. Our solution is thus more conservative, which seems to be the price to pay to statisfy requirement $R_2$.

**The case $K = 3$**

When $K = 3$, $P^-$ is again a belief function. Its Möbius inverse is

$$\begin{aligned} m^{\mathcal{X}}(\{\xi_k\}) &= P_k^-, \quad k = 1, 2, 3 \\ m^{\mathcal{X}}(\{\xi_1, \xi_2\}) &= 1 - P_3^+ - P_1^- - P_2^-, \\ m^{\mathcal{X}}(\{\xi_1, \xi_3\}) &= 1 - P_2^+ - P_1^- - P_3^-, \\ m^{\mathcal{X}}(\{\xi_2, \xi_3\}) &= 1 - P_1^+ - P_2^- - P_3^-, \\ m^{\mathcal{X}}(\mathcal{X}) &= \frac{b}{n + b}, \end{aligned}$$

where $b$ is, as before, the quantile of order $1-\alpha/3$ of the chi-square distribution with one degree of freedom. It can be shown that all these masses are positive [6]. Consequently, $P^-$ is a belief function.

**The case $K > 3$**

Let $\mathcal{B}^{\mathcal{X}}(P^-)$ denote the set of belief functions $bel^{\mathcal{X}}$ on $\mathcal{X}$ verifying $bel^{\mathcal{X}} \leq P^-$. As a consequence of (11), we have, for any $bel^{\mathcal{X}} \in \mathcal{B}^{\mathcal{X}}(P^-)$:

$$\mathbb{P}(bel^{\mathcal{X}} \leq \mathbb{P}_X) \geq \mathbb{P}(P^- \leq \mathbb{P}_X) \geq 1 - \alpha.$$

Every element of $\mathcal{B}^{\mathcal{X}}(P^-)$ thus complies with requirement $R_2$ expressed by (6). However, most elements of that set (such as, e.g., the vacuous belief function) will generally not be very informative, and it seems natural to concentrate on the most committed elements of $\mathcal{B}^{\mathcal{X}}(P^-)$. A way to find belief function is to maximize the sum of belief degrees[1] $bel^{\mathcal{X}}(A)$ for all $A \subseteq \mathcal{X}$, under the constraints $bel^{\mathcal{X}}(A) \leq \mathbb{P}_X(A)$, for all $A \subseteq \mathcal{X}$. Let $J(m^{\mathcal{X}})$ denote this criterion. We have

$$J(m^{\mathcal{X}}) = \sum_{A \subseteq \mathcal{X}} bel^{\mathcal{X}}(A) \quad (12)$$

$$= 2^K \sum_{B \subseteq \mathcal{X}} 2^{-|B|} m^{\mathcal{X}}(B) \;, \quad (13)$$

We then have to solve the following linear program:

$$\max_{m^{\mathcal{X}}} J(m^{\mathcal{X}}) \quad (14)$$

under the constraints:

$$\sum_{B \subseteq A} m^{\mathcal{X}}(B) \leq P^-(A), \quad \forall A \subset \mathcal{X}, \quad (15)$$

$$\sum_{A \subseteq \mathcal{X}} m^{\mathcal{X}}(A) = 1, \quad (16)$$

$$m^{\mathcal{X}}(A) \geq 0, \quad \forall A \subseteq \mathcal{X} \;. \quad (17)$$

Any belief function $bel_n^{\mathcal{X}*}$ solution to the above linear programming problem obviously satisfies requirement $R_2$. The following proposition (proved in [6]) states that it also satisfies $R_1$.

Proposition 2 *Let $bel_n^{\mathcal{X}*}, n = 1, \ldots, \infty$ be a sequence of solutions of linear program (14)-(17). We have:*

$$bel_n^{\mathcal{X}*} \xrightarrow{P} \mathbb{P}_X \text{ as } n \to \infty.$$

[1]A similar criterion was proposed by Baroni and Vicig [1] and by Hall and Lawry [10] for approximating a lower probability measure by a belief function.

Solving linear program (14)-(17) is thus a way to construct a predictive belief function, as illustrated by the next example. Note that the uniqueness of the solution is not guaranteed, which is not important in practice, since all the solutions may be regarded as equivalent.

Example 1 A sample of 220 psychiatric patients were categorized as either neurotic ($\xi_1$), depressed ($\xi_2$), schizophrenic ($\xi_3$) or having a personality disorder ($\xi_4$) [12]. The observed counts were $\mathbf{n} = (91, 49, 37, 43)$. The Goodman confidence intervals at confidence level $1 - \alpha = 0.95$ are $[0.33, 0.50]$, $[0.16, 0.30]$, $[0.11, 0.24]$, $[0.14, 0.27]$. Table 1 shows the corresponding belief and mass functions.

Table 1: Belief and mass functions, at confidence level 0.95, for the data of Example 1.

| $A$ | $P^-(A)$ | $bel^{\mathcal{X}*}(A)$ | $m^{\mathcal{X}*}(A)$ |
|---|---|---|---|
| $\{\xi_1\}$ | 0.33 | 0.33 | 0.33 |
| $\{\xi_2\}$ | 0.16 | 0.14 | 0.14 |
| $\{\xi_1, \xi_2\}$ | 0.50 | 0.50 | 0.021 |
| $\{\xi_3\}$ | 0.11 | 0.097 | 0.097 |
| $\{\xi_1, \xi_3\}$ | 0.45 | 0.45 | 0.020 |
| $\{\xi_2, \xi_3\}$ | 0.28 | 0.28 | 0.036 |
| $\{\xi_1, \xi_2, \xi_3\}$ | 0.73 | 0.69 | 0.040 |
| $\{\xi_4\}$ | 0.14 | 0.12 | 0.12 |
| $\{\xi_1, \xi_4\}$ | 0.47 | 0.47 | 0.02 |
| $\{\xi_2, \xi_4\}$ | 0.30 | 0.30 | 0.035 |
| $\{\xi_1, \xi_2, \xi_4\}$ | 0.76 | 0.72 | 0.045 |
| $\{\xi_3, \xi_4\}$ | 0.25 | 0.25 | 0.035 |
| $\{\xi_1, \xi_3, \xi_4\}$ | 0.70 | 0.66 | 0.038 |
| $\{\xi_2, \xi_3, \xi_4\}$ | 0.50 | 0.48 | 0.019 |
| $\mathcal{X}$ | 1 | 1 | 0 |

Note that the applicability of the method is obviously limited to moderate values of $K$ (up to 10-15), since both the number of variables and the number of constraints grow exponentially with $K$. For large $K$ or when computation speed is an issue, however, it may be sufficient to compute suboptimal solutions.

This can be done, for instance, using the Iterative Rescaling Method (IRM) described in [10], which heuristically transforms the Möbius inversion of $P^-$ into a bba, by replacing each negative mass $m(A) < 0$ by zero, and rescaling masses assigned to relevant subsets of $A$.

Although the IRM algorithm may allow to find good approximations for moderate values of $K$, its time and space complexity is still exponential as a function of $K$ (it involves a loop over the subsets of $\mathcal{X}$). Much more drastic approximations may be obtained by limiting the search to a restricted parametrized family of belief functions $\mathcal{B}_0^{\mathcal{X}}(P^-) \subset \mathcal{B}^{\mathcal{X}}(P^-)$. When the elements of $\mathcal{X}$ are ordered, it is quite natural to consider belief functions whose focal elements are intervals, since the corresponding basic belief masses can easily be represented and interpreted. In that case, the optimal solution has a simple analytical expression, as will be shown in the next section.

## 4 Approximation in the case of ordered data

We assume in this section that a meaningful ordering has been defined among the elements of $\mathcal{X}$. By convention, we shall assume that $\xi_1 < \ldots < \xi_K$.

Let $A_{k,r}$ denote the subset $\{\xi_k, \ldots, \xi_r\}$, for $1 \leq k \leq r \leq K$ and let $\mathcal{I}$ denote the set of intervals of $\mathcal{X}$: $\mathcal{I} = \{A_{k,r}, 1 \leq k \leq r \leq K\}$. By imposing that the focal sets of $m$ be taken in $\mathcal{I}$, one reduces the number of basic belief numbers from $2^K - 1$ to $K(K+1)/2$. Let $m^{\mathcal{X}*}$ be bba defined by

$$m^{\mathcal{X}*}(A_{k,k}) = P_k^-,$$

$$m^{\mathcal{X}*}(A_{k,k+1}) = P^-(A_{k,k+1}) - P^-(A_{k+1,k+1}) - P^-(A_{k,k}),$$

$$m^{\mathcal{X}*}(A_{k,r}) = P^-(A_{k,r}) - P^-(A_{k+1,r}) - P^-(A_{k,r-1}) + P^-(A_{k+1,r-1})$$

for $r > k+1$, and $m^{\mathcal{X}*}(B) = 0$, for all $B \notin \mathcal{I}$. In [6], we prove that $m^{\mathcal{X}*}$ is a valid bba, and that it is optimal according to criterion $J$ defined by (12), in the set of bbas with focal elements in $\mathcal{I}$.

Example 2 Table 2 shows categorized data concerning January precipitation in Arizona (in inches), recorded during the period 1895-2004, together with the estimated probabilities of each class, and Goodman simultaneous confidence intervals at confidence level 0.95. The masses $m^{\mathcal{X}*}(A_{k,r})$ are given in Table 3.

Table 2: Arizona January precipitation data, with simultaneous 95 % confidence intervals.

| class $\xi_k$ | $n_k$ | $n_k/n$ | $p_k^-$ | $p_k^+$ |
|---|---|---|---|---|
| $< 0.75$ | 48 | 0.44 | 0.32 | 0.56 |
| $[0.75, 1.25)$ | 17 | 0.15 | 0.085 | 0.27 |
| $[1.25, 1.75)$ | 19 | 0.17 | 0.098 | 0.29 |
| $[1.75, 2.25)$ | 11 | 0.10 | 0.047 | 0.20 |
| $[2.25, 2.75)$ | 6 | 0.055 | 0.020 | 0.14 |
| $\geq 2.75$ | 9 | 0.082 | 0.035 | 0.18 |

Table 3: Basic belief masses for the precipitation data. Masses are given to intervals $A_{k,r} = \{\xi_k, \ldots, \xi_r\}$ with $r \geq k$. Each cell at the intersection of row $k$ and colums $r$ contains $m(A_{k,r})$.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0.32 | 0 | 0 | 0.13 | 0.11 | 0 |
| 2 | - | 0.085 | 0 | 0 | 0.012 | 0.14 |
| 3 | - | - | 0.098 | 0 | 0 | 0 |
| 4 | - | - | - | 0.047 | 0 | 0 |
| 5 | - | - | - | - | 0.020 | 0 |
| 6 | - | - | - | - | - | 0.035 |

## 5 Conclusion

We have proposed a method for quantifying, in the belief functions framework, the uncertainty concerning a discrete random variable $X$ with unknown probability distribution $\mathbb{P}_X$, based on a realization of an iid sample from the same distribution. The proposed solution verifies two "reasonable" properties with respect to $\mathbb{P}_X$: it is less committed than $\mathbb{P}_X$ with some user-defined probability, and it converges towards $\mathbb{P}_X$ in probability as the size of the sample tends to infinity.

In this paper, only the case of a discrete random variable $X$ has been considered. The method can be applied to the continuous case by discretizing the sample values (which is a form of coarsening), and vacuously extending the obtained belief function. A specific method designed for the continuous case is under study.

## References


[1] P. Baroni and P. Vicig. An uncertainty interchange format with imprecise probabilities. *International Journal of Approximate Reasoning*, 40(3):147–180, 2005.

[2] A. P. Dempster. New methods of reasoning towards posterior distributions based on sample data. *Annals of Mathematical Statistics*, 37:355–374, 1966.

[3] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[4] A. P. Dempster. A generalization of Bayesian inference (with discussion). *J. R. Statistical Society B*, 30:205–247, 1968.

[5] A. P. Dempster. A class of random convex polytopes. *Annals of Mathematical Statistics*, 43(1):260–272, 1972.

[6] T. Denœux. Constructing belief functions from sample data using multinomial confidence regions. *International Journal of Approximate Reasoning (To appear)*, 2006.

[7] J. Glaz and C. P. Sison. Simultaneous confidence intervals for multinomial proportions. *Journal of Statistical Planning and Management*, 82:251–262, 1999.

[8] L. A. Goodman. On simultaneous confidence intervals for multinomial proportions. *Technometrics*, 7(2):247–254, 1965.

[9] I. Hacking. *Logic of Statistical Inference.* Cambridge University Press, Cambridge, 1965.

[10] J. W. Hall and J. Lawry. Generation, combination and extension of random set approximations to coherent lower and upper probabilities. *Reliability Engineering and System Safety*, (85):89–101, 2004.

[11] J. Kohlas and P.-A. Monney. *A Mathematical Theory of Hints. An Approach to the Dempster-Shafer Theory of Evidence.* Springer-Verlag, Berlin, 1995.

[12] W. L. May and W. D. Johnson. A SAS macro for constructing simultaneous confidence intervals for multinomial proportions. *Computer methods and Programs in Biomedicine*, 53:153–162, 1997.

[13] C. P. Quesenberry and D. C. Hurst. Large sample simultaneous confidence intervals for multinomial proportions. *Technometrics*, 6(2):191–195, 1964.

[14] G. Shafer. *A mathematical theory of evidence.* Princeton University Press, Princeton, N.J., 1976.

[15] G. Shafer. Belief functions and parametric models (with discussion). *J. Roy. Statist. Soc. Ser. B*, 44:322–352, 1982.

[16] Ph. Smets. Resolving misunderstandings about belief functions. *International Journal of Approximate Reasoning*, 6:321–344, 1990.

[17] Ph. Smets. Belief induced by the partial knowledge of the probabilities. In D. Heckerman *et al.*, editor, *Uncertainty in AI'94*, pages 523–530. Morgan Kaufmann, San Mateo, 1994.

[18] Ph. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66:191–243, 1994.