Inferring a possibility distribution from empirical data

Marie-Hélène Masson[1*†] and Thierry Denœux[†]

[*] Université de Picardie Jules Verne
[†] Université de Technologie de Compiègne
UMR CNRS 6599 Heudiasyc
BP 20529 - F-60205 Compiègne cedex - France

August 1, 2005

[1]Corresponding author. Email: mmasson@hds.utc.fr. Fax: (33) 3 44 23 44 77.

Abstract

Several transformations from probabilities to possibilities have been proposed. In particular, Dubois and Prade's procedure produces the most specific possibility distribution among the ones dominating a given probability distribution. In this paper, this method is generalized to the case where the probabilities are unknown, the only information being a data sample represented by a histogram. It is proposed to characterize the probabilities of the different classes by simultaneous confidence intervals with a given confidence level $1 - \alpha$. From this imprecise specification, a procedure for constructing a possibility distribution is described, insuring that the resulting possibility distribution will dominate the true probability distribution in at least $100(1 - \alpha)\%$ of the cases. Finally, a simple efficient algorithm is given which makes the computations tractable even if the number of classes is high.

Keywords: Probability-Possibility Transformation, Possibility theory, Statistics, Multinomial Confidence Regions, simultaneous confidence intervals.

1 Introduction

Uncertainty management is one of the major issues in complex decision system design. For a long time, the only tools for representing uncertainty were based on probability theory. Initiated by Zadeh [24], possibility theory has gained an increasing interest in recent years as it offers an alternative framework for uncertainty modeling and management. As pointed out by Dubois et al [12], probability theory can be viewed as a good model for representing randomness, while possibility theory is useful for representing partial knowledge or incompleteness. Although these two theories aim at representing different kinds of uncertainty, it is often desirable to move from one framework to another, for example to fuse heterogeneous information sources. Several transformations, in finite or continuous settings, have been proposed in the literature [9, 1, 3, 15, 16, 5], based on various principles such as consistency ("what is probable is possible") or information invariance.

A related, but distinct and more complex problem is to build a possibility distribution from empirical data. One approach, investigated in this paper, is to assume the (discrete) data to have been generated from an unknown probability (i.e., long-run frequency) distribution. If the sample is very large, the histogram of the data can be considered as a good approximation to the underlying probability distribution, and the above probability-possibility transformations can be applied. This approach, however, is clearly unsatisfactory in the case of limited data, where the empirical distribution can be very different from the true underlying distribution. We are then obviously not primarily interested in the empirical distribution, which is random and, to some extent, accidental, but rather in the underlying probability distribution, from which some properties can be inferred from the data.

This is, of course, a statistical inference problem, and it is natural to attack it using existing methods of statistical inference, whose results will need, however, to be interpreted in the possibilistic setting. A useful concept in classical statistics, which will be used in this

paper, is that of confidence region. A confidence region for a scalar of vector parameter is a random region in parameter space, defined as a function of the sample, which is guaranteed to contain the true parameter value with a certain prescribed probability, or confidence level $1-\alpha$. In this approach, it must be quite clear that the parameter is considered as unknown but constant, whereas the confidence region is random. Once the data has been observed, a realization of the confidence region can be computed. It is not known whether this particular region contains the true parameter value or not. It is only known that it was constructed using a method which, in the long run, yields a region containing the true parameter value for a fraction $1-\alpha$ of the samples.

In our case, the parameter is the vector $\mathbf{p} = (p_1, p_2, ..., p_K)$ of probabilities characterizing the unknown probability distribution of a random variable X on $\Omega = \{\omega_1, \ldots, \omega_K\}$. Let n_k denote the number of observations of class ω_k in a random sample of size N. Then, the random vector $\mathbf{n} = (\mathbf{n_1}, \ldots, \mathbf{n_K})$ as a multinomial distribution with parameter $\mathbf{p}$. A confidence region for $\mathbf{p}$ at level $1-\alpha$ can be computed using simultaneous confidence intervals. Such a confidence region can be considered as a set of probability distributions. In this paper, it is shown how to construct the most specific possibility distribution dominating all the probability distributions in that set. This procedure guarantees that the obtained possibility distribution will be consistent with the true unknown probability distribution, in the long run, in at least $100(1-\alpha)\%$ of the cases.

The paper is organized as follows. Section 2 present the necessary background on ordering relations and on Dubois and Prade's probability-possibility transformation. Our method is then exposed in Section 3, and an efficient computational procedure is introduced in Section 4. Finally, numerical experiments are presented in Section 5, and Section 6 concludes the paper.

2 Background

2.1 Ordering relations

Before developing our approach, we give some basic definitions about ordering relations. A binary relation on a set U is a subset of the Cartesian product U^2. The notation $u\mathcal{R}v$ is often used in place of $(u, v) \in \mathcal{R}$. A relation $\mathcal{R}$ is said to be:

- transitive if $\forall (u, v, w) \in U^3, \quad u\mathcal{R}v$ and $v\mathcal{R}w \Rightarrow u\mathcal{R}w$;
- antisymmetric if $\forall (u, v) \in U^2, \quad u\mathcal{R}v$ and $v\mathcal{R}u \Rightarrow u = v$;
- irreflexive if $\forall (u, v) \in U^2, u\mathcal{R}v \Rightarrow u \neq v$;
- complete if $\forall (u, v) \in U^2, \quad u \neq v \Rightarrow u\mathcal{R}v$ or $v\mathcal{R}u$.

A *partial order* is an antisymmetric, transitive relation. An irreflexive partial order is said to be strict. A *linear* order is a transitive, antisymmetric and complete order. A linear order $\mathcal{L}$ is compatible with a partial order $\mathcal{P}$ if and only if $\mathcal{P} \subseteq \mathcal{L}$. In this case, $\mathcal{L}$ is called a *linear extension* of $\mathcal{P}$. Each partial order can be associated to the set of its linear extensions.

2.2 Dubois and Prade's transformation

The problem of moving from probabilities to possibilities has received a lot of attention in the past [9, 1, 3, 15, 16, 5]. A consistency principle between probability and possibility was first stated by Zadeh [24] in an unformal way: what is probable should be possible. Dubois and Prade [6, 8] translated this requirement via the inequality

$$P(A) \leq \Pi(A) \quad \forall A \subseteq \Omega, \tag{1}$$

where P and Π are, respectively, a probability and a possibility measure on a domain $\Omega = \{\omega_1, \ldots, \omega_K\}$. In this case, Π is said to *dominate* P. Transforming a probability

measure into a possibilistic one then amounts to choosing a possibility measure in the set $\mathcal{F}(P)$ of possibility measures dominating P. Dubois et al [12, 4] proposed to add the following strong order preservation constraint, which insures the preservation of the shape of the distribution:

$$p_i < p_j \Leftrightarrow \pi_i < \pi_j \quad \forall i, j \in \{1, \ldots, K\}, \tag{2}$$

where $p_i = P(\{\omega_i\})$ and $\pi_i = \Pi(\{\omega_i\})$, for all $i \in \{1, \ldots, K\}$. It is then natural to search for the *most specific* possibility distribution verifying constraints (1) and (2) (we recall that possibility distribution π is *more specific* than π' if $\pi_i \leq \pi'_i, \forall i$).

Dubois and Prade [12, 4] showed that the solution to this problem exists and is unique. This solution can be described as follows. Assuming $p_i \neq p_j$ for all $i \neq j$, one can define a strict linear order $\mathcal{L}$ on $\Omega = \{\omega_1, ..., \omega_K\}$ such that:

$$(\omega_i, \omega_j) \in \mathcal{L} \Leftrightarrow p_i < p_j \ . \tag{3}$$

Let σ be the permutation of the indices $\{1, 2, ..., K\}$ associated to this linear order such that $p_{\sigma(1)} < p_{\sigma(2)} < ... < p_{\sigma(K)}$ or, equivalently:

$$\sigma(i) < \sigma(j) \Leftrightarrow (\omega_{\sigma(i)}, \omega_{\sigma(j)}) \in \mathcal{L} \ . \tag{4}$$

The permutation σ is a bijection and the reverse transformation σ^{-1} gives the rank of each p_i in the list of the probabilities sorted in the ascending order. The transformation of Dubois and Prade may then be expressed as:

$$\pi_i = \sum_{\{j | \sigma^{-1}(j) \leq \sigma^{-1}(i)\}} p_j \ . \tag{5}$$

EXAMPLE 1 Let $p_1 = 0.2$, $p_2 = 0.35$, $p_3 = 0.4$, and $p_4 = 0.05$. Then we have $\sigma(1) = 4$, $\sigma(2) = 1$, $\sigma(3) = 2$, $\sigma(4) = 3$ and $\sigma^{-1}(1) = 2$, $\sigma^{-1}(2) = 3$, $\sigma^{-1}(3) = 4$, $\sigma^{-1}(4) = 1$.

Transformation (5) gives the following possibility distribution :

$$\begin{aligned}
\pi_1 &= p_1 + p_4 = 0.2 + 0.05 = 0.25 \\
\pi_2 &= p_2 + p_1 + p_4 = 0.35 + 0.2 + 0.05 = 0.6 \\
\pi_3 &= p_3 + p_2 + p_1 + p_4 = 0.4 + 0.35 + 0.2 + 0.05 = 1 \\
\pi_4 &= p_4 = 0.05.
\end{aligned}$$

REMARK 1 Formulation (5) assumes the p_i's to be all different. If at least two probabilities are equal, (3) induces no longer a strict linear order, but a strict partial order $\mathcal{P}$ on Ω. This partial order can be represented by the set of its compatible linear extensions $\Lambda(\mathcal{P}) = \{\mathcal{L}_l, l = 1, L\}$. To each possible linear order $\mathcal{L}_l$ of $\Lambda(\mathcal{P})$, one can associate a permutation σ_l of the set $\{1, .., K\}$ such that:

$$\sigma_l(i) < \sigma_l(j) \Leftrightarrow (\omega_{\sigma_l(i)}, \omega_{\sigma_l(j)}) \in \mathcal{L}_l. \tag{6}$$

In that case, the most specific possibility distribution compatible with $\mathbf{p} = (p_1, ..., p_K)$ is obtained by taking the maximum over all possible permutations:

$$\pi_i = \max_{l=1,L} \sum_{\{j|\sigma_l^{-1}(j) \leq \sigma_l^{-1}(i)\}} p_j \ . \tag{7}$$

EXAMPLE 2 Let $p_1 = 0.2$, $p_2 = 0.5$, $p_3 = 0.2$, and $p_4 = 0.1$. We have thus two possible permutations $\sigma_1(1) = 4$, $\sigma_1(2) = 1$, $\sigma_1(3) = 3$, $\sigma_1(4) = 2$ and $\sigma_2(1) = 4$, $\sigma_2(2) = 3$, $\sigma_2(3) = 1$, $\sigma_2(4) = 2$. Transformation (7) gives the following possibility distribution :

$$\begin{aligned}
\pi_1 &= \max(p_4 + p_1, p_4 + p_3 + p_1) = \max(0.3, 0.5) = 0.5 \\
\pi_2 &= p_4 + p_1 + p_3 + p_2 = 1 \\
\pi_3 &= \max(p_4 + p_1 + p_3, p_4 + p_3) = \max(0.5, 0.3) = 0.5 \\
\pi_4 &= p_4 = 0.1.
\end{aligned}$$

It can be seen that $p_1 = p_3$ implies $\pi_1 = \pi_3$, a condition which is required for strong order preservation.

3 Inferring a possibility distribution from experimental data

3.1 Problem statement

We suppose that the available data consist of N observations $(X_1, X_2, ..., X_N)$ generated according to probability distribution $\mathbb{P}_X$ on $\Omega = \{\omega_1, \omega_2, ..., \omega_K\}$. Let n_i denote the number of observations falling in the ith class ω_i. The vector $\mathbf{n} = (n_1, n_2, ..., n_K)$ is a sample of a multinomial distribution with parameters $\mathbf{p} = (p_1, p_2, ..., p_K)$, where each $p_i = \mathbb{P}_X(\{\omega_i\}) > 0$ is the probability of occurence of the ith class and $\sum_{i=1}^{K} p_i = 1$. The classical approach for deriving a possibility distribution from this data would be to consider the vector of observed frequencies $\mathbf{f} = (f_1, f_2, ..., f_K)$ where $f_i = n_i/N$ as the true vector of probabilities $\mathbf{p}$ and to apply Dubois and Prade's transformation. However this approach does not take into account the uncertainty due to the sampling process, as shown in the following example:

Example 3 Figure 1 shows the possibility distributions computed with (5) using two samples drawn from a multinomial distribution with parameters $\mathbf{p} = [0.2; 0.35; 0.4; 0.05]^t$ and $N = 100$. It can be seen that the two distributions are very different although the samples are drawn from the same population. Moreover, one of the possibility distributions (left) does not dominate the true probability distribution (see Example 1).

Statistical inference provides some tools to draw conclusions about the characteristics of a population, based on information gathered from a sample. In particular, confidence intervals are a usual means to estimate the unknown parameters of a distribution. A confidence interval on a parameter at a given level α is a random interval (defined as a function of the observations) that contains the true value of the parameter with probability $1-\alpha$ (i.e., in $100(1-\alpha)\%$ of the cases when the data samples are repeatedly drawn from the unknown distribution). We propose in this paper, first, to estimate the p_i using confidence intervals on multinomial proportions, and then to derive a possibility distribution from

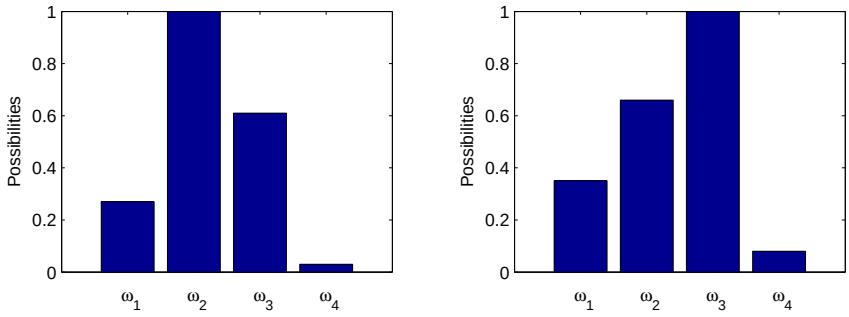

Figure 1: Two possibility distributions computed using samples from the same probability distribution.

these intervals. The procedure should insure that the possibility distribution will dominate the true probability distribution in at least $100(1-\alpha)\%$ of the cases, what can be formally stated as:

$$\mathbb{P}\left(\Pi(A) \geq \mathbb{P}_X(A), \forall A \subseteq \Omega\right) \geq 1-\alpha \ , \tag{8}$$

where $\mathbb{P}_X(A)$ is the unknown but constant probability of event A, and $\Pi(A)$ is a random variable defined as a function of the observations. Note that proposition (8) is equivalent to

$$\mathbb{P}\left(N(A) \leq \mathbb{P}_X(A), \forall A \subseteq \Omega\right) \geq 1-\alpha \ , \tag{9}$$

where N is the necessity measure associated to Π, defined as $N(A) = 1 - \Pi(\overline{A})$, $\forall A \subseteq \Omega$.

3.2 Confidence intervals for multinomial proportions

Making a guess about a population parameter on the basis of a sample is a classical question addressed by statistical inference. In point estimation, the estimate of an unknown parameter is a single value. In interval estimation, a scalar population parameter is typically estimated as a range of possible values, namely a confidence interval, with a given confidence level $1-\alpha$.

To construct confidence intervals for multinomial proportions, a first approach is to consider each n_i versus the remaining ones as the realization of a binomial distribution and

to make a set of confidence intervals independently from each other. This approach does not allow to control the overall confidence coefficient for the entire set of intervals. A better approach lies in constructing *simultaneous* confidence intervals with a joint confidence level $1-\alpha$. Finding simultaneous confidence intervals for multinomial proportions is an old problem and several methods have been proposed in the literature [19, 14, 13, 21]. All these methods attempt to find a confidence region $\mathcal{C}_{\mathbf{n}}$ in the parameter space $\{\mathbf{p} = (p_1, ...p_K) \in [0;1]^K | \sum_{i=1}^K p_i = 1\}$ as the Cartesian product of K intervals $[p_1^-, p_1^+] \times ... \times [p_K^-, p_K^+]$ such that

$$\mathbb{P}(\mathbf{p} \in \mathcal{C}_{\mathbf{n}}) \geq 1 - \alpha \tag{10}$$

This probability is called the coverage probability of the estimation. We have retained the solution proposed by Goodman [14] which has been tested using different simulations and shown to perform well in most practical situations [17]. The derivation of Goodman's formulation is explained in Appendix A. The main formulas are given in the sequel. Let

$$A = \chi^2(1 - \alpha/K, 1) + N, \tag{11}$$

where $\chi^2(1-\alpha/K, 1)$ denotes the quantile of order $1-\alpha/K$ of the chi-square distribution with one degree of freedom, and $N = \sum_{i=1}^K n_i$ denotes the size of the sample. Additionnally, let:

$$B_i = \chi^2(1 - \alpha/K, 1) + 2n_i, \tag{12}$$

$$C_i = \frac{n_i^2}{N}, \tag{13}$$

$$\Delta_i = B_i^2 - 4AC_i. \tag{14}$$

Then the bounds of the confidence intervals are defined as follows:

$$[p_i^-, p_i^+] = \left[\frac{B_i - \Delta_i^{\frac{1}{2}}}{2A}, \frac{B_i + \Delta_i^{\frac{1}{2}}}{2A} \right]. \tag{15}$$

Note that this formula relies on asymtotic approximations (see Appendix A). Based on simulation studies, May and Johnson [17] stated that Goodman's intervals perform

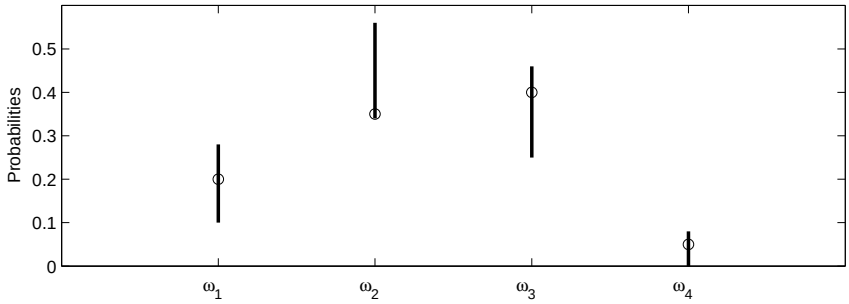

Figure 2: Confidence intervals of Example 4; the circles represent the true class probabilities.

Table 1: Confidence intervals of Example 4

i	1	2	3	4
p_i^-	0.10	0.34	0.25	0
p_i^+	0.28	0.56	0.46	0.08

well in terms of coverage probability and confidence region volume, provided that K, the number of classes, is greater than 2 and that each n_i is greater than 5. If the total sample size is small or the number of observation in a class is small, other methods of intervals construction should be used (see, for example, Sison and Glaz [21]).

EXAMPLE 4 Let the true probability be again $\mathbf{p} = [0.2; 0.35; 0.4; 0.05]^t$. We suppose that we observe a data sample of size 100 with the following distribution in the different classes: 18, 45, 35, and 2. Setting $\alpha = 0.1$, leads to the simultaneous confidence intervals given in Table 1 and represented in Figure 2.

3.3 Induced lower probability measure

A confidence region $\mathcal{C}_{\mathbf{n}}$ for multinomial proportions such as described above is usually interpreted as defining a set of plausible values for vector parameter $\mathbf{p}$. However, since each value of $\mathbf{p}$ specifies a unique probability measure on Ω, it is clear that $\mathcal{C}_{\mathbf{n}}$ can equivalently be seen as defining a family of probability measures. To keep the notation as simple as

possible, the same symbol $\mathcal{C}_{\mathbf{n}}$ will be used to denote the set of parameter values $\mathbf{p}$ and the set of probability measures. Let P^- and P^+ denote, respectively, the lower and upper enveloppes of $\mathcal{C}_{\mathbf{n}}$, defined as $P^-(A) = \min_{P \in \mathcal{C}_{\mathbf{n}}} P(A)$ and $P^+(A) = \max_{P \in \mathcal{C}_{\mathbf{n}}} P(A)$. They can be easily computed using the following proposition.

PROPOSITION 1

For all strict nonempty subset A of Ω,

$$P^-(A) = \max \left(\sum_{\omega_i \in A} p_i^-, 1 - \sum_{\omega_i \notin A} p_i^+ \right) \tag{16}$$

$$P^+(A) = \min \left(\sum_{\omega_i \in A} p_i^+, 1 - \sum_{\omega_i \notin A} p_i^- \right). \tag{17}$$

Proof. For all $A \subset \Omega$, $P^-(A)$ is the solution of the following linear program:

$$\min_{p_1,\ldots,p_K} \sum_{\omega_i \in A} p_i,$$

subject to the constraints $\sum_{i=1}^K p_i = 1$ and $p_i^- \leq p_i \leq p_i^+$, $i = 1, \ldots, K$. This problem is a special case of a family of linear programs studied by Dubois and Prade in the context of fuzzy arithmetics [7, 10]. Equation (16) can be derived from the general formula given in [10, page 55]. Equation (17) can be obtained in a similar way. □

Note that we have, as a direct consequence of Proposition 1:

$$P^+(A) = 1 - P^-(\overline{A}), \quad \forall A \subseteq \Omega.$$

Hence, the lower probability measure P^- is sufficient to characterize $\mathcal{C}_{\mathbf{n}}$:

$$\mathcal{C}_{\mathbf{n}} = \{P \mid P^- \leq P\}.$$

By construction, we have

$$\mathbb{P}(\mathbb{P}_X \in \mathcal{C}_{\mathbf{n}}) = \mathbb{P}(P^- \leq \mathbb{P}_X) \geq 1 - \alpha \tag{18}$$

and, equivalently,

$$\mathbb{P}(P^+ \geq \mathbb{P}_X) \geq 1 - \alpha \ . \tag{19}$$

Equations (18) and (19) are similar to (9) and (8), respectively. However, P^- is not a necessity measure. It is not even, in general, a belief function [20] when $K > 2$, as shown by Example 5 below. However, it can be shown to be a 2-monotone capacity, i.e., we have

$$P^-(A \cup B) \geq P^-(A) + P^-(B) - P^-(A \cap B), \quad \forall A, B \subseteq \Omega.$$

Consequently, P^- is a coherent lower probability measure [22].

EXAMPLE 5 Let us return to the data and confidence region computed in Example 4. The corresponding lower probabilities are shown in Table 2. As shown by Shafer [20], a mapping $f : 2^\Omega \to [0, 1]$ is a belief function iff its Möbius inverse, defined as:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B), \quad \forall A \subseteq \Omega,$$

is a basic belief assignment (i.e., if $m(A) \geq 0$ for all A, and $\sum_{A \subseteq \Omega} m(A) = 1$). The Möbius inverse of P^-, shown in Table 2, assigns a negative value to Ω. Consequently, P^- is not a belief function.

3.4 Generating a possibility distribution from interval-valued probabilities

We have seen that a set of simultaneous multinomial confidence intervals at confidence level $1 - \alpha$ can be seen as defining a set of probability measures that can be described exactly by its lower envelope P^-, which is a coherent lower probability, or, equivalently, by its upper envelope P^+.

We recall that our goal is to find a probability measure Π verifying (8). Now, it is clear that this property is satisfied for any possibility Π dominating P^+, since $\Pi(A) \geq P^+(A)$, $\forall A \subseteq \Omega$ and (19) implies (8).

The problem of finding the least specific possibility measure dominating a plausibility function has been addressed in [11], in which an algorithm is proposed for computing a

Table 2: Lower and upper probabilities induced by the confidence intervals of Table 1, and corresponding Möbius inverse.

A	$P^-(A)$	$P^+(A)$	$m(A)$
$\{\omega_1\}$	0.1038	0.2938	0.1038
$\{\omega_2\}$	0.3323	0.5735	0.3323
$\{\omega_1, \omega_2\}$	0.4362	0.7530	0
$\{\omega_3\}$	0.2429	0.4747	0.2429
$\{\omega_1, \omega_3\}$	0.3467	0.6636	0.0000
$\{\omega_2, \omega_3\}$	0.6139	0.8921	0.0387
$\{\omega_1, \omega_2, \omega_3\}$	0.9077	0.9959	0.1900
$\{\omega_4\}$	0.0041	0.0923	0.0041
$\{\omega_1, \omega_4\}$	0.1079	0.3861	0
$\{\omega_2, \omega_4\}$	0.3364	0.6533	0
$\{\omega_1, \omega_2, \omega_4\}$	0.5253	0.7571	0.0850
$\{\omega_3, \omega_4\}$	0.2470	0.5638	0
$\{\omega_1, \omega_3, \omega_4\}$	0.4265	0.6677	0.0757
$\{\omega_2, \omega_3, \omega_4\}$	0.7062	0.8962	0.0882
Ω	1.0000	1.0000	-0.1607

good (though not optimal) solution. However, this algorithm is not applicable here since, as we have shown in the previous section, P^- is not a belief function (or, equivalently, P^+ is not a plausibility function).

It is clear that Π dominates the upper probability measure P^+ if and only if it dominates all elements in the corresponding family of probability measures, i.e., all probability measures P such that $P \leq P^+$, or, equivalently, all probability distributions verifying $p_i^- \leq p_i \leq p_i^+$, $i = 1, \ldots, K$. We thus reformulate our goal as finding the most specific possibility distribution on Ω, dominating every probability distribution defined by $p_i \in [p_i^-, p_i^+] \quad \forall i$, or, equivalently, every possibility distribution induced by possible values of the p_i's, whatever their value in $[p_i^-, p_i^+]$.

To achieve this goal, our approach will be to use the transformation described in

Section 2.2, which allows to compute the most specific possibility distribution dominating any particular probability measure.

Let $\mathcal{P}$ denote the partial order induced by the intervals $[p_i] = [p_i^-, p_i^+]$:

$$(\omega_i, \omega_j) \in \mathcal{P} \Leftrightarrow p_i^+ < p_j^-. \tag{20}$$

As explained in Section 2.1, this partial order may be represented by the set of its compatible linear extensions $\Lambda(\mathcal{P}) = \{\mathcal{L}_l, l = 1, L\}$, or, equivalently, by the set of the corresponding permutations $\{\sigma_l, l = 1, L\}$.

Formally, the solution of our problem may be thus computed as follows:

1. For each possible permutation σ_l associated to each linear orders in $\Lambda(\mathcal{P})$, and each class ω_i, solve the following linear program (LP):

$$\pi_i^{\sigma_l} = \max_{p_1,\ldots,p_K} \sum_{\{j|\sigma_l^{-1}(j)\leq\sigma_l^{-1}(i)\}} p_j \tag{21}$$

under the constraints:

$$\begin{cases} \sum_{k=1}^{K} p_k = 1 \\ p_k^- \leq p_k \leq p_k^+ \quad \forall k \in \{1, ..., K\} \\ p_{\sigma_l(1)} \leq p_{\sigma_l(2)} \leq \cdots \leq p_{\sigma_l(K)} \end{cases} \tag{22}$$

2. Then, take the distribution dominating all the distributions π^{σ_l}:

$$\pi_i = \max_{l=1,L} \pi_i^{\sigma_l} \quad \forall i \in \{1, ..., K\}. \tag{23}$$

If the $[p_i]$ are simultaneous confidence intervals computed using (15) with a confidence level $1-\alpha$, this procedure ensures that the resulting possibility measure is the most specific one that dominates all the compatible probability measures, and, consequently, the upper probability P^+. As a consequence, it verifies property (8).

EXAMPLE 6 We consider four classes characterized by the probability intervals given in Table 1 and represented in Figure 2. The corresponding partial order of the classes is:

$$\mathcal{P} = \{(\omega_4, \omega_1), (\omega_1, \omega_2), (\omega_4, \omega_2), (\omega_4, \omega_3)\}.$$

There are three permutations compatible with $\mathcal{P}$: $\sigma_1 = (4, 1, 3, 2)$, $\sigma_2 = (4, 1, 2, 3)$, $\sigma_3 = (4, 3, 1, 2)$. The corresponding ranks are $\sigma_1^{-1} = (2, 4, 3, 1)$, $\sigma_2^{-1} = (2, 3, 4, 1)$ and $\sigma_3^{-1} = (3, 4, 2, 1)$. Table 3 gives the solutions of the different linear programs and the final possibility distribution. Classes ω_2 and ω_3 logically receive the greatest possibility, since both corresponding probabilities could be ranked in the last position. Class ω_4, always ranked in the first position, has a possibility degree corresponding to its maximum of probability. The value of 0.64 of possibility degree for the first class is obtained for the third permutation with an optimal value of $\mathbf{p}$ equal to [0.28,0.36,0.28,0.08].

Table 3: Solutions of the linear programs and derived optimal solution (Example 6).

l	$\pi_1^{\sigma_l}$	$\pi_2^{\sigma_l}$	$\pi_3^{\sigma_l}$	$\pi_4^{\sigma_l}$
1	0.36	1	0.66	0.08
2	0.32	0.66	1	0.08
3	0.64	1	0.36	0.08
$\max_l$	0.64	1	1	0.08

REMARK 2 Some of the probability distributions in $\mathcal{C}_n$ are such that $p_i = p_j$ for some $i \neq j$. As emphasized in Remark 1, the corresponding possibility distributions are built by considering all the linear extensions of the induced partial orders. This is consistent with our approach, which considers *all* the linear extensions of the partial order induced by the probability intervals.

REMARK 3 It is very important to note that the constraint set may be inconsistent for some problems defined by (21)-(22). As an example, consider the following probability intervals: $p_1 = [0; 0.9]$, $p_2 = [0.1; 0.3]$, $p_3 = [0; 0.8]$. A possible permutation is $\sigma(1) = 1$, $\sigma(2) = 3$, $\sigma(3) = 2$. However, the associated constraint set is not consistent, since the equality constraint can not be satisfied, p_1 and p_3 being limited by the upper value of p_2 which is 0.3. The final possibility distribution is thus derived by taking the maximum of the solutions over the feasible problems which gives here $\pi_1 = 1$, $\pi_2 = 0.6$ and $\pi_3 = 1$.

REMARK 4 The above procedure could be applied if the $[p_i]$ were obtained differently, for example, directly elicited from experts. However, property (8), insuring a coverage rate of a predefined level, which is our main concern in this paper, would not be satisfied. Confidence intervals may be also obtained by other procedures, like the one suggested by Walley [23] in the context of the theory of imprecise probabilities.

4 Computational procedure

4.1 Outline of the approach

To compute the possibility degrees of the different classes, the conceptually simplest approach is to generate all the linear extensions compatible with the partial order induced by the probability intervals, and then to solve the associated linear programs, as shown in the previous section. However, this approach is unfortunately limited to small values of K (say $K < 10$) due to the complexity of the algorithms generating linear extensions:

the fastest algorithm seems to be the algorithm of Pruesse and Ruskey [18], whose complexity is $O(L)$, where L is the number of linear extensions. Even for moderate values of K, L can be very large ($K!$ in the worst case) and generating all the linear extensions and solving the linear programs soon becomes intractable. We propose in this section a new formulation of the solution which will be shown to be equivalent to the first one and reduces the computations. This formulation will be explained in several steps:

- first, all the linear programs to be solved will be grouped in different subsets;
- then, an analytic expression for the best solution in each subset will be given;
- lastly, it will be shown that it is not necessary to evaluate the solution for every subset. A simple computational algorithm will be derived.

4.2 Step 1: Grouping the linear programs

For a given permutation σ, let S_i^σ denote the set of classes with a rank smaller than or equal to the one of ω_i in the permutation σ:

$$S_i^\sigma = \{j \mid \sigma^{-1}(j) \leq \sigma^{-1}(i)\}. \tag{24}$$

The set of all linear programs to be solved may be partitioned by remarking that several permutations can be associated to the same set S_i^σ. Let S denote such a set and let Σ_S be the set of all permutations σ such that $S_i^\sigma = S$, that is the set of permutations for which the probabilities smaller than p_i are the p_j, $j \in S$. Let π_i^S denote the maximum of the solutions found in Σ_S:

$$\pi_i^S = \max_{\sigma \in \Sigma_S} \pi_i^\sigma \tag{25}$$

The interest of such a grouping is that it is not necessary to solve $n_S = |\Sigma_S|$ linear programs to compute π_i^S. Only one linear program is needed, as expressed in the following proposition:

PROPOSITION 2

π_i^S is the solution of the following linear program:

$LP(S)$:

$$\pi_i^S = \max_{p_1,...,p_K} \sum_{j \in S} p_j \tag{26}$$

under the constraints:

$$\begin{cases} \sum_{k=1}^{K} p_k = 1 \\ p_k^- \leq p_k \leq p_k^+ \quad \forall k \in \{1, ..., K\} \\ p_k \leq p_i \quad \forall k \in S \end{cases} \tag{27}$$

Sketch of proof. In each subset Σ_S, all LP have the same objective function. Taking the maximum over the n_S solutions is equivalent to relax all the precedence constraints between the p_k.

It may be remarked that, if $\mathbf{p}^*$ denotes argmax π_i^S, we have necessarily $p_i^* = p_i^+$ since raising p_i to its largest value increases the objective function and in the same time relaxes the bounds of the constraints. This allows to reformulate constraints (27) as:

$$\begin{cases} \sum_{k=1}^{K} p_k = 1 \\ p_k^- \leq p_k \leq p_k^+ \quad \forall k \in \bar{S} \\ p_k^- \leq p_k \leq \min(p_i^+, p_k^+) \quad \forall k \in S, \end{cases} \tag{28}$$

4.3 Step 2: Analytic expression of π_i^S

Problem $LP(S)$ above consists in the maximization of a linear function under box constraints and one equality constraints. This is again a problem belonging to the class of problems studied in [7][10, page 55]. By applying the general formula given in [10, page55], we obtain the following analytical expression for the solution of $LP(S)$:

$$\pi_i^S = \min \left(\sum_{k \in S} \min(p_i^+, p_k^+), 1 - \sum_{k \in \bar{S}} p_k^- \right), \tag{29}$$

where $\bar{S}$ denote the complement of S.

Note that, to satisfy the normalization condition for probabilities, the sum of the upper bounds of the box contraints must be greater or equal to one, so that a solution exists if and only if:

$$\sum_{k \in S} \min(p_k^+, p_i^+) + \sum_{k \in \bar{S}} p_k^+ \geq 1. \tag{30}$$

4.4 Step 3: Search for the optimal S

Having expressed the solution for a given set S, it is now necessary to find the set S maximizing the possibility degree of ω_i. In fact, one obtains the overall solution for π_i by taking the maximum over all subsets S for which problem $LP(S)$ is feasible. The solution (29) may be expressed in another way by decomposing S as follows.

Let N_i denote the set of the indices of the classes with a rank necessarily smaller than the rank of ω_i, and let P_i denote the set of indices of the classes with a rank possibly, but not necessarily smaller than ω_i. Since $p_i^* = p_i^+$, it is clear that $p_j^+ < p_i^+$ implies $p_j^* < p_i^*$. Consequently, we have:

$$N_i = \{j / p_j^+ \leq p_i^+\}, \tag{31}$$

and

$$P_i = \{j \ | p_j^+ > p_i^+ \text{ and } p_j^- < p_i^+\}. \tag{32}$$

Note that N_i necessarily contains the singleton $\{i\}$. Any set of indices S for which a solution of problem $LP(S)$ exists satisfies $S = N_i \cup P_i^S$ for some $P_i^S \subseteq P_i$. Such a set will

be said to be *admissible*. Let S be an admissible set. Then, one obtains:

$$\begin{aligned}
\pi_i^S &= \min\left(\sum_{k\in N_i}\min(p_i^+,p_k^+)+\sum_{k\in P_i^S}\min(p_i^+,p_k^+), 1-\sum_{k\in\bar{S}}p_k^-\right)\\
&= \min\left(\sum_{k\in N_i}p_k^+ + \sum_{k\in P_i^S}p_i^+, 1-\sum_{k\in\bar{S}}p_k^-\right)\\
&= \min\left(\sum_{k\in N_i}p_k^+ + |P_i^S|p_i^+, 1-\sum_{k\in\bar{S}}p_k^-\right). \qquad (33)
\end{aligned}$$

We now prove that the evaluation of π_i^S for every admissible S is not necessary, using the following two propositions.

Proposition 3 (Monotonicity property)

Let S and S' two admissible sets. If $S \subseteq S'$ *then* $\pi_i^S \leq \pi_i^{S'}$.

Sketch of proof. The proposition is obvious: the left-hand part of the min in (33) is clearly increasing with the cardinality of S, and, in the same time, if $S \subset S'$ then, when going from π_i^S to $\pi_i^{S'}$ one removes one or more terms from the sum in the right-hand part.

Proposition 3 is very important as it gives a first way to drastically limit the search of the solution. We construct a search tree where the root is the set of indices with maximal cardinality $(N_i \cup P_i)$. The successors in the tree are constructed by successively removing one element of P_i from the current set of "candidates" so that the sets associated to any path in the tree consist in a family of nested subsets. By convention, level 0 (the root of the tree) corresponds to $P_i^S = P_i$; level 1 corresponds to the sets of indices where one element of P_i has been removed; at level 2, two elements have been removed, and so on. The monotonicity property helps to avoid an exhaustive search in the tree: whenever $LP(S)$ is found to be feasible at a given node of the tree (which can be checked by (30)), it is not necessary to evaluate the solution for its children nodes, so that the whole subtree of the current node may be cut off and many computations may be omitted. As we shall see, the following proposition allows to further simplify the computations.

PROPOSITION 4

Let S_1 and S_2 two admissible sets with corresponding levels in the tree $level_1$ and $level_2$, such that $LP(S_1)$ and $LP(S_2)$ are feasible. If $level_1 = level_2$ then $\pi_i^{S_1} = \pi_i^{S_2}$. Moreover if $level_1 > level_2$ then $\pi_i^{S_1} > \pi_i^{S_2}$.

Proof.

1. First, as already explained before, it is obvious that the left-hand part of the min in (33) only depends on the size of the set S considered, and thus on the level of the nodes in tree.

2. Second, the solution, except for the root, is never given by the right-hand part of the min, which can be shown as follows. Let $S = N_i \cup P_i^S$ a set of indices such that $LP(S)$ is found to be unfeasible and S' a child of S in the tree obtained by removing index j from P_i^S. Because $LP(S)$ is unfeasible, we have:

$$\begin{aligned} & \sum_{k\in S} \min(p_k^+, p_i^+) + \sum_{k\in\bar{S}} p_k^+ \leq 1 \\ \Leftrightarrow \quad & \sum_{k\in N_i} p_k^+ + |P_i^S| p_i^+ + \sum_{k\in\bar{S}} p_k^+ \leq 1 \\ \Leftrightarrow \quad & \sum_{k\in N_i} p_k^+ + (|P_i^S| - 1) p_i^+ \leq 1 - \sum_{k\in\bar{S}} p_k^+ - p_i^+ \end{aligned} \tag{34}$$

   Remembering that $j \in P_i^S$, so that $p_j^- < p_i^+$, (34) is thus equivalent to:

$$\begin{aligned} & \sum_{k\in N_i} p_k^+ + (|(P_i^S| - 1) p_i^+ \leq 1 - \sum_{k\in\bar{S}} p_k^+ - p_j^- \\ \Leftrightarrow \quad & \sum_{k\in N_i} p_k^+ + |P_i^{S'}| p_i^+ \leq 1 - \sum_{k\in\bar{S}} p_k^- - p_j^- \\ \Leftrightarrow \quad & \sum_{k\in N_i} p_k^+ + |P_i^{S'}| p_i^+ \leq 1 - \sum_{k\in\bar{S}'} p_k^- \\ \Leftrightarrow \quad & \pi_i^{S'} = \sum_{k\in N_i} p_k^+ + |P_i^{S'}| p_i^+. \end{aligned} \tag{35}$$

Propositions 3 and 4 now allow to propose an efficient computational algorithm: it starts from the root (level 0 of the tree). If the corresponding linear program is feasible,

it gives the solution. If not, the tree is explored using a breadth-first search strategy: as soon as a linear program is found to be feasible in a given level, it gives the solution and the search can be stopped. The proposed algorithm is detailed in Appendix B.

EXAMPLE 7 A small example is given to illustrate the algorithm. The size of the sample is $N = 250$. The observations are classified into $K = 5$ classes and the following distribution is observed: $\mathbf{n} = [39, 32, 64, 56, 56, 39]$. Setting $\alpha = 5\%$, we construct simultaneous confidence intervals yielding the following probability bounds:

$$\mathbf{p}^- = [0.1059, 0.0830, 0.1919, 0.1637, 0.1742], \tag{36}$$

$$\mathbf{p}^+ = [0.2239, 0.1922, 0.3328, 0.2986, 0.3114] \tag{37}$$

Table 4 gives the details of the computations involved by the proposed procedure. For a better understanding of the method, all possibilities at level 1 are detailed in Table 4 for $i = 2$, although the procedure normally stops as soon as a solution is found.

REMARK 5 During the revision of this paper, we became aware of reference [2], in which a heuristic procedure is given for generalizing the Dubois-Prade probability-possibility transformation to probability intervals. This procedure happens to be equivalent to applying formula (29) to the root of the tree, regardless of the existence of a solution at this level. Consequently, it is not optimal. In the above example, this method yields $\pi = [1, 0.9609, 1, 1, 1]$, whereas the optimal solution found by our algorithm is $\pi = [1, 0.7688, 1, 1, 1]$.

REMARK 6 Our algorithm essentially explores the tree using a breadth-first strategy, and terminates once a solution has been found. Although the worst case time complexity is the same as that of exhaustive search ($|P_i|!$ leaves in the tree), a solution was always found at level 0 or 1 in our simulations. This suggests that the actual complexity might actually be close to $O(|P_i|)$.

Table 4: Example 7: Details of computation.

i	level 0 (root)	level 1
$i=1$	$N_i=\{1,2\}\quad P_i^S=P_i=\{3,4,5\}$ feasibility=4*0.2239+0.1922=1.078 ≥ 1 $\pi_i=\min(1.078,1)=\mathbf{1}$	*not explored*
$i=2$	$N_i=\{2\}\quad P_i^S=P_i=\{1,3,4,5\}$ feasibility=5*0.1922=0.9609 ≤ 1 *unfeasible*	$N_i=\{2\}\quad P_i^S=\{3,4,5\}$ feasibility=4*0.1922+0.2239=0.9927 ≤ 1 *unfeasible*
		$N_i=\{2\}\quad P_i^S=\{1,4,5\}$ feasibility=4*0.1922+0.3328=1.1016 ≥ 1 $\pi_i=4*0.1922=\mathbf{0.7688}\rightarrow$ Exit
		$N_i=\{2\}\quad P_i^S=\{1,3,5\}$ feasibility=4*0.1922+0.2986=1.1016 ≥ 1 $\pi_i=4*0.1922=\mathbf{0.7688}$
		$N_i=\{2\}\quad P_i^S=\{1,3,4\}$ feasibility=4*0.1922+0.3114=1.0802 ≥ 1 $\pi_i=4*0.1922=\mathbf{0.7688}$
$i=3$	$N_i=\{1,2,3,4,5\}\quad P_i^S=P_i=\emptyset$ feasibility=0.2239+0.1922+0.3328 + 0.2986+0.3114=1.3589 ≥ 1 $\pi_i=\min(1.3589,1)=\mathbf{1}$	*not explored*
$i=1$	$N_i=\{1,2,4\}\quad P_i^S=P_i=\{3,5\}$ feasibility=3*0.2986+0.1922 +0.2239=1.3118 ≥ 1 $\pi_i=\min(1.3118,1)=\mathbf{1}$	*not explored*
$i=5$	$N_i=\{1,2,4,5\}\quad P_i^S=P_i=\{3\}$ feasibility=2*0.3114+0.2239+0.1922 +0.2986=1.3375 ≥ 1 $\pi_i=\min(1.3375,1)=\mathbf{1}$	*not explored*

5 Experiments

5.1 First experiment

Let us consider a histogram of continuous data composed of K=11 classes uniformly spaced from -2 to 2. The observed frequencies are assumed to be

$$f = [0.04, 0.02, 0.05, 0.09, 0.14, 0.27, 0.14, 0.14, 0.07, 0.02, 0.02],$$

leading to the possibility distribution computed from Dubois and Prade's transformation represented in Figure 3. One assumes that several samples of varying size (N=100,500,1000 and 10000) have led to the same frequency data. Goodman's confidence intervals, computed with $\alpha = 0.05$, are shown in Figure 4. The results of the proposed transformation in the four cases are shown in Figure 3. They confirm what was expected: our transformation always dominates Dubois and Prade's one and asymptotically converges to it as N goes to infinity.

5.2 Second experiment

In a second experiment, we studied the coverage probability of the proposed transformation (i.e., the probability that the possibility distribution dominates the true probability distribution). To do that, the following procedure was used:

- step 1: we considered Ω to be composed of $K = 5$ classes. Five proportions p_i were uniformly chosen in the interval [0,1] with the constraint $\sum_{i=1}^{K} p_i = 1$;
- step 2: the coverage probability was estimated using 100 samples of fixed size N, drawn from a multinomial distribution of parameters p_i; this estimation was done by computing, for each sample, a set of confidence intervals with a fixed α, transforming them into a possibility distribution using algorithm `Prob2poss` and checking whether:

$$\Pi(A) \geq P(A) \quad \forall A \subseteq \Omega$$

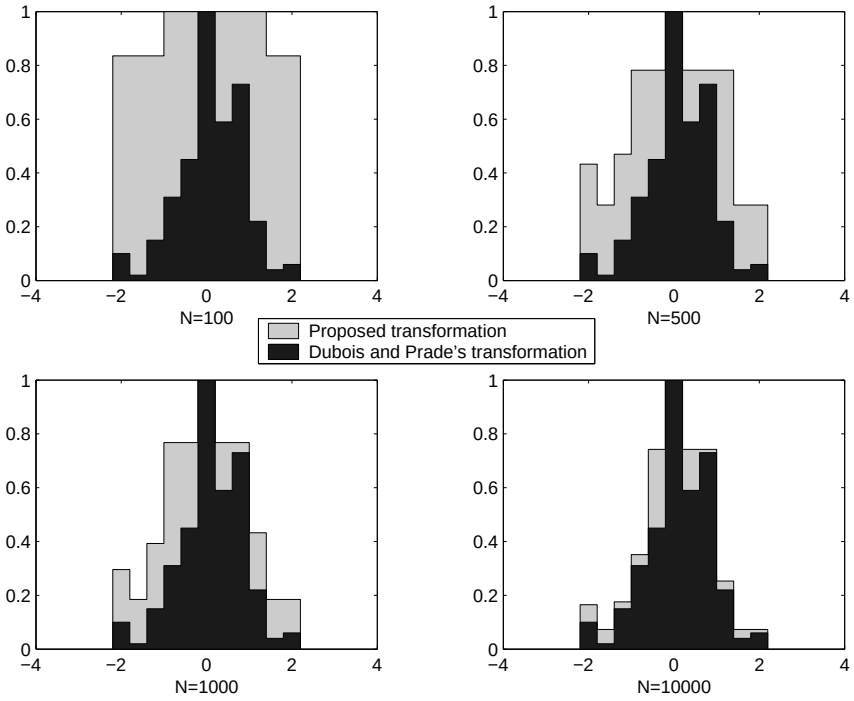

Figure 3: Experiment 1. Convergence of the proposed transformation to Dubois and Prade's one as $N \to \infty$.

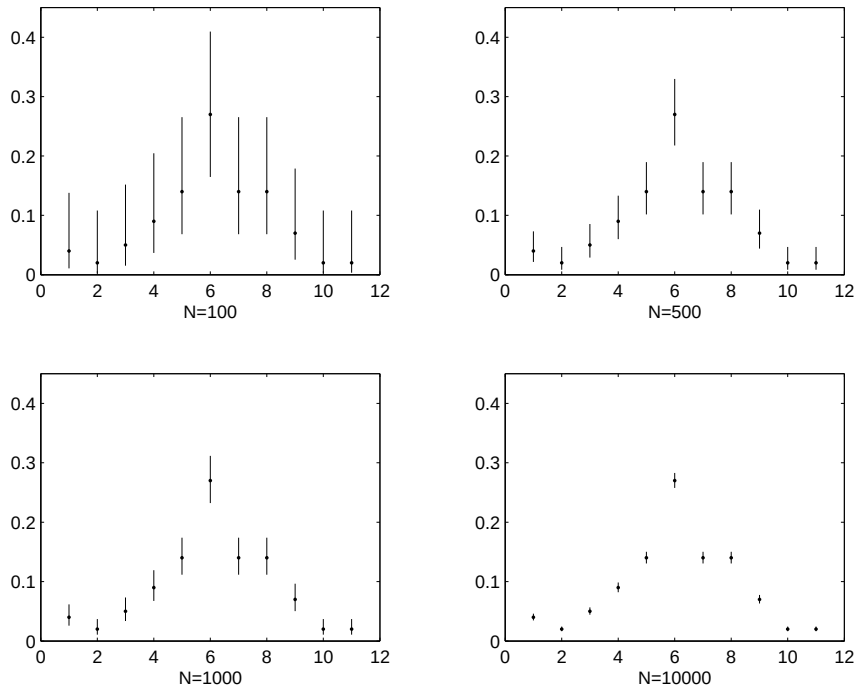

Figure 4: Experiment 1. Goodman's confidence intervals; the points represent the class frequencies.

(for each subset A of Ω, one checked if the maximum of the possibility degrees of the singletons in A was greater or equal to the sum of their probabilities).

- this estimated coverage probability was averaged over $nrep = 100$ replications of the previous experiment (steps 1 and 2).
- The whole experiment was repeated using different values of α (0.01,0.05,0.1,0.2,0.3,0.4) and N (100, 1000, 10000).

In Figure 5, is shown the coverage rate of the proposed transformation. It can be seen that the solution actually dominates the true probability distribution with a rate always much higher than $100(1-\alpha)\%$. Our approach is thus very cautious and conservative from this point of view. As a matter of comparison, the coverage rate of Goodman's intervals is shown in Figure 6. For the proposed transformation, the choice of α is not critical and whatever its value, a very good coverage rate is insured, for any sample size. The choice of α involves a trade-off between coverage probability and specificity of the resulting possibility distribution.

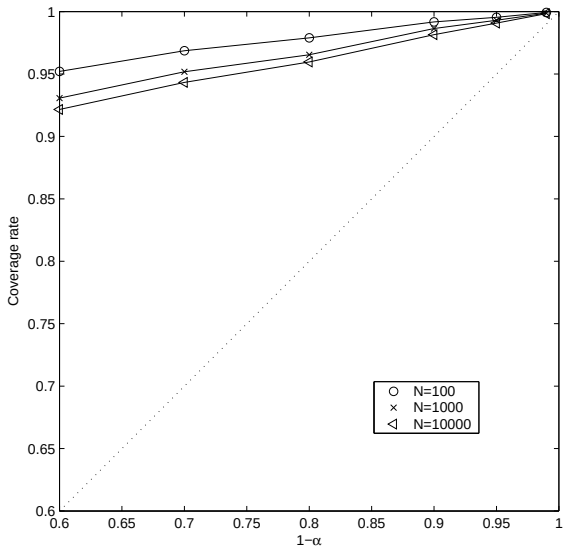

Figure 5: Experiment 2. Coverage rate of the proposed transformation.

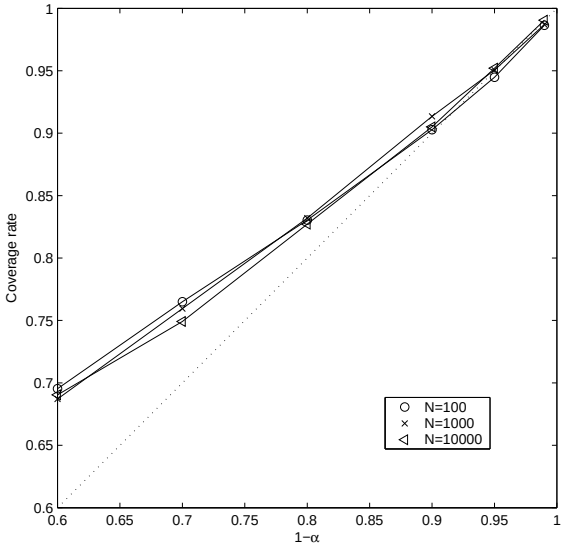

Figure 6: Experiment 2. Coverage rate of Goodman's confidence intervals.

6 Conclusion

A procedure has been proposed for constructing a possibility distribution from a discrete empirical frequency distribution. It is assumed that the data have been randomly generated from an unknown probability distribution. Based on this assumption, simultaneous confidence intervals for multinomial proportions are constructed, defining a set of probability distributions. We then construct the most specific possibility distribution which dominates all the probability distributions in that set. This procedure is guaranteed to yield a possibility measure that dominates the true generating probability measure (corresponding to long-run frequencies) in at least $100(1-\alpha)\%$ of the cases (i.e., for at least $100(1-\alpha)\%$ of the samples). We believe that this approach provides an interesting link between classical inferential statistics and Possibility Theory.

A Goodman's simultaneous confidence intervals

Before discussing the derivation of the confidence intervals of Goodman, a useful tool, referred to as the Bonferroni inequality, is explained. It may be stated as follows:

Bonferroni inequality Let $A_1, ..., A_K$ be K events of a sample space which each have a probability of $1-\alpha$ of occurring. Then the probability that they all occur simultaneously is $1 - K\alpha$.

Proof. We start from

$$P(A_1 \cap ... \cap A_K) = 1 - P(\bar{A_1} \cup ... \cup \bar{A_K}). \tag{38}$$

Using the fact that $P(\bar{A_i}) = \alpha$ for all i and that

$$P(\bar{A_1} \cup ... \cup \bar{A_K}) \leq P(\bar{A_1}) + ... + P(\bar{A_K}), \tag{39}$$

one easily obtains the following result:

$$P(A_1 \cap ... \cap A_k) \geq 1 - K\alpha. \tag{40}$$

To derive confidence intervals, Goodman first considers that each parameter p_i is the parameter of a binomial random variable. A normal approximation of a binomial random variable is then used: as N becomes large, $f_i = n_i/N$ becomes approximately normally distributed with mean p_i and variance $p_i(1-p_i)/N$ so that, the following random variable:

$$Z_i = \frac{\sqrt{N}(f_i - p_i)}{\sqrt{p_i(1-p_i)}} \tag{41}$$

is asymptotically normally distributed with 0 mean and variance 1. The square of each Z_i is thus a chi-square distribution with one degree of freedom. Confidence intervals for the p_i may be thus derived by finding the solutions of the following set of quadratic equations:

$$N(f_i - p_i)^2 = \chi^2(1-\alpha, 1)p_i(1-p_i) \quad \forall i = 1, K, \tag{42}$$

or, equivalently:

$$p_i^2(N + \chi^2(1-\alpha,1)) - (2n_i + \chi^2(1-\alpha,1)p_i + \frac{n_i^2}{N} = 0 \quad \forall i = 1, K \tag{43}$$

where $\chi^2(1-\alpha,1)$ denotes the quantile of order $1-\alpha$ of the chi-square distribution with one degree of freedom. Each equation has two solutions which define the lower and upper bounds of the confidence interval. Then, using a Bonferroni adjustement, Goodman replaces $\chi^2(1-\alpha,1)$ by $\chi^2(1-\alpha/K,1)$ in (43) , leading to the bounds expressed by equations (11)-(15).

B Procedure `Prob2Poss`

Algorithm 1 Probability-Possibility Transformation

1: **procedure** PROB2POSS($\mathbf{p}^-$,$\mathbf{p}^+$)

2: **for** i=1,K **do**

Initialization

3: $P_i^S = P_i$ using (32)

4: $S \leftarrow N_i \cup P_i^S$ using (31)

5: $level \leftarrow 0$

6: check for the feasibility of $LP(S)$ using (30)

7: **if** $LP(S)$ is feasible **then**

8: $\pi_i \leftarrow \pi_i^S$ computed from (33)

9: **Stop**,

10: **else**

11: **repeat**

12: $level \leftarrow level + 1$

13: Loop: remove one by one any combination C of $level$ elements from P_i, $P_i^S = P_i \backslash C$, $S \leftarrow N_i \cup P_i^S$, until one $LP(S)$ is found to be feasible and solve it using eq. (33). π_i is equal to this solution. Exit loop.

14: **until** a solution has been found

15: **end if**

16: **end for**

17: **end procedure**

References

[1] M.R. Civanlar and H. J. Trussel. Constructing membership functions using statistical data. *Fuzzy Sets and Systems*, 18:1–13, 1986.

[2] L.M. de Campos and J.F. Huete. Measurement of possibility distributions. *Int. J. General Systems*, 30(3):309–346, 2001.

[3] M. Delgado and S. Moral. On the concept of possibility-probability consistency. *Fuzzy Sets and Systems*, 21:311–318, 1987.

[4] D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets and probabilistic inequalities. *Reliable Computing*, 10:273–297, 2004.

[5] D. Dubois, H.T. Nguyen, and H. Prade. Possibility theory, probability and fuzzy sets: misunderstanding, bridges and gaps. In D. Dubois and H. Prade, editors, *Fundamental of fuzzy sets*. Kluwer Academic Publishers, Boston, 2000.

[6] D. Dubois and H. Prade. *Fuzzy sets and systems: theory and Applications*. Academic Press, New York, 1980.

[7] D. Dubois and H. Prade. Addition of interactive fuzzy numbers. *IEEE Transactions on Automatic Control*, 26(4):926–936, 1981.

[8] D. Dubois and H. Prade. Unfair coins and necessity measures, towards a possibilistic interpretation of histograms. *Fuzzy sets and Systems*, 10:15–20, 1983.

[9] D. Dubois and H. Prade. Fuzzy sets and statistical data. *European Journal of Operational Research*, pages 345–356, 1986.

[10] D. Dubois and H. Prade. *Possibility Theory: An approach to computerized processing of uncertainty*. Plenum Press, New-York, 1988.

[11] D. Dubois and H. Prade. Consonant approximations of belief measures. *International Journal of Approximate Reasoning*, 4:419–449, 1990.

[12] D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. In *Proceedings of the Fourth Int. Fuzzy Systems Association World Congress (IFSA'91), Brussels, Belgium*, pages 50–53, 1991.

[13] S. Fitzpatrick and A. Scott. Quick simultaneous confidence intervals for multinomial proportions. *Journal of the American Statistical Association*, 82(399):875–879, 1987.

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

[15] G.J. Klir. A principle of uncertainty and information invariance. *Int. J. of General Systems*, 17(2-3):249–275, 1990.

[16] V. Lasserre, G. Mauris, and L. Foulloy. A simple possibilistic modelisation of measurement uncertainty. In L.A. Zadeh Eds. B. Bouchon-Meunier, R.R. Yager, editor, *Uncertainty in Intelligent and Information Systems*, pages 58–69. World Scientific, 2000.

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

[18] G. Pruesse and F. Ruskey. Generating linear extensions fast. *SIAM Journal on Computing*, 23(2):373–386, 1994.

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

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

[21] C.P. Sison and J. Glaz. Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association*, 90(429):366–369, 1995.

[22] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[23] P. Walley. Inferences from multinomial data: learning about a bag of marbles (with discussion). *Journal of the Royal Statistical Society, Series B*, 58:3–57, 1996.

[24] L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.