PMLR: Proceedings of Machine Learning Research, vol. 62, 61-72, 2017 ISIPTA '17

Empirical Interpretation of Imprecise Probabilities

Marco E. G. V. Cattaneo M.CATTANEO@HULL.AC.UK
University of Hull
Kingston upon Hull (United Kingdom)

Abstract

This paper investigates the possibility of a frequentist interpretation of imprecise probabilities, by generalizing the approach of Bernoulli's *Ars Conjectandi*. That is, by studying, in the case of games of chance, under which assumptions imprecise probabilities can be satisfactorily estimated from data. In fact, estimability on the basis of finite amounts of data is a necessary condition for imprecise probabilities in order to have a clear empirical meaning. Unfortunately, imprecise probabilities can be estimated arbitrarily well from data only in very limited settings.

Keywords: imprecise probabilities; frequentist interpretation; empirical meaning; bag of marbles; strong estimability; consistent estimators; empirical recognizability.

1. Introduction

Imprecise probabilities mostly have a subjective, epistemic interpretation (Walley, 1991; Troffaes and de Cooman, 2014), while in this paper we will study the possibility of a frequentist, empirical interpretation for them. As regards precise probabilities, empirical interpretations are dominant in science and statistics. They are usually related to Bernoulli's law of large numbers, which connects the probabilities of events with the relative frequencies of the events' occurrence in sequences of independent repetitions of experiments.

This connection can be used asymptotically, by defining probabilities as limits of relative frequencies (Venn, 1866; von Mises, 1928, 1957; Reichenbach, 1935, 1949), but the empirical meaning of such probabilities for finite samples is then problematic. In order to have probabilities with a direct empirical meaning, the connection in Bernoulli's law of large numbers can be used in a finite-sample way, by defining probabilities as approximately equal to relative frequencies in large, but finite samples. The difficulty of this approach comes from the fact that the exact meaning of "approximately equal" is probabilistic, and therefore this definition of probability is circular.

A possible answer to this circularity consists in accepting it and interpreting probability as an abstract concept, whose meaning comes from the possibility of statistically falsifying probabilistic statements (Popper, 1935, 1959). An alternative, but related answer to the above circularity is the original approach of Bernoulli (1713, 2006): define probability only for games of chance (where the definition is unproblematic) and extend it to other fields by analogy. This analogy is empirically meaningful because Bernoulli's law of large numbers provides a way of estimating probabilities arbitrarily well (and thus also a way of statistically falsifying probabilistic statements).

In this paper, we will see if Bernoulli's approach can be extended to imprecise probabilities. That is, practically we will focus on games of chance: for example drawing colored marbles at random from a bag. In this situation, the precise probability of a certain color corresponds to the proportion of marbles of this color in the bag, and if we draw several marbles (with replacement) from the bag, we obtain probabilistic independence automatically from the noninteraction of the drawings. How should we interpret an imprecise probability in this setting? We will see that several different interpretations may be reasonable.

The spectacular achievement of Bernoulli was to prove, through his law of large numbers, that precise probabilities are estimable from finite amounts of data, and therefore have an empirical meaning. Analogously, a frequentist, empirical interpretation of imprecise probabilities is possible only if these are estimable from finite amounts of data. The core of the present paper consists of mathematical results about the estimability of imprecise probabilities, depending on their exact interpretation in the case of games of chance. These results are given in Section 3 (due to space limitations, proofs are omitted, and will appear only in an extended version of the paper), while the next section provides a quick overview of frequentist interpretations, and the last section concludes the paper and points to an open problem.

2. Interpretations of Imprecise Probabilities

The interpretations of (precise) probabilities can be roughly grouped in two main classes, often called subjective and frequentist (see for example Gillies, 2000). With a subjective (or epistemic, Bayesian, personalistic, . . .) interpretation, probabilistic statements are about the degrees of belief or knowledge of an individual. By contrast, with a frequentist (or empirical, objective, scientific, . . .) interpretation, probabilistic statements are about the material world. For this reason, frequentist interpretations of probabilities are the dominant interpretations in science and in statistics.

In particular, according to the subjective interpretation of de Finetti (1931, 1974–1975), a probability is an individual's fair price for a bet. This interpretation can quite naturally be extended to an interpretation of lower and upper probabilities as an individual's maximum buying price and minimum selling price for a bet (Williams, 1975, 2007; Walley, 1991; Troffaes and de Cooman, 2014). In fact, it can certainly be argued that with this subjective interpretation, imprecise probabilities are more natural than precise ones. However, the topic of the present paper is frequentist interpretations for imprecise probabilities, which, contrary to what happens for precise probabilities, are far less common than subjective ones.

Since usual imprecise probability measures correspond mathematically to sets of precise ones, they appear often in classical statistics, which is based on frequentist interpretations of probabilities. In particular, imprecise probabilities can be used to describe what has been learnt so far from data (see for example Cattaneo and Wiencierz, 2012; Antonucci et al., 2012), but in this case their interpretation is in reality epistemic, although more properly intersubjective than subjective. However, a truly frequentist interpretation is indeed obtained in classical statistics when imprecise probabilities do not describe what has been learnt, but what can potentially be learnt from infinite amounts of incomplete data (see for instance Manski, 2003; Dempster, 1967). Anyway, this frequentist interpretation of imprecise probabilities is limited to particular situations involving incomplete data, while we are looking for a generally valid interpretation.

In general, frequentist interpretations of precise probabilities are related to laws of large numbers implying that the relative frequency of an event's occurrence in a sequence of independent repetitions of an experiment converges to the probability of the event. Although laws of large numbers have been generalized to the case of imprecise probabilities (Walley and Fine, 1982; Cozman and Chrisman, 1997; Marinacci, 1999; de Cooman and Miranda, 2008; Peng, 2010; Chen and Wu, 2011), the generalization of frequentist interpretations is not straightforward.

If we would simply interpret the probability of an event as the limit of the relative frequency of its occurrence in an infinite sequence of independent repetitions of an experiment (Venn, 1866; von Mises, 1928, 1957; Reichenbach, 1935, 1949), then we could interpret lower and upper probabilities as limits inferior and superior of such a sequence, respectively. That is, the imprecise probability interpretation would extend the precise one to the case of nonconvergent sequences of relative frequencies, and also with this frequentist interpretation (besides the above subjective one) it could be argued that imprecise probabilities are more natural than precise ones. However, this interpretation is problematic for imprecise as well as precise probabilities, since no finite part of a sequence of relative frequencies has any connection at all with the limit of the sequence, and thus strictly speaking the interpretation has no empirical meaning.

In order to have an empirical meaning, a frequentist interpretation must make probabilistic statements falsifiable on the basis of finite amounts of data. Of course, probabilistic statements are in general not strictly falsifiable, but they can be methodologically falsifiable in the sense of Popper (1935, 1959) if they can be rejected through some reasonable statistical test with arbitrarily low significance level (see also Gillies, 1995, 2000). Such a test for the probability of an event could be based on Bernoulli's law of large numbers, which is a probabilistic statement connecting the probability of the event to the relative frequency of the event's occurrence in a finite sequence of independent repetitions. That is, we could consider frequentist probability as an abstract concept deriving its meaning from the theory surrounding it, which makes probabilistic statements (methodologically) falsifiable.

However, in the present paper we will follow a related, but more direct approach to frequentist probability, corresponding to the original interpretation of Bernoulli's law of large numbers in the *Ars Conjectandi* (Bernoulli, 1713, 2006). This book represents the starting point of modern probability theory, and interestingly also the (temporary) end point of imprecise probability (Shafer, 1978). Citing Sylla (2014): "before Bernoulli's work, there existed a mathematics of games of chance but that mathematics did not involve probability—not the Latin word *probabilis*, not relative frequencies and not degrees of certainty."

Bernoulli's law of large numbers is a theorem in the mathematics of games of chance. That is, a theorem about probabilities interpreted as ratios between the numbers of favorable and possible outcomes. Bernoulli extended the concept of probability to other fields by analogy with games of chance, an idea already present in the *Logique de Port-Royal* (Arnauld and Nicole, 1662, 1996). According to this approach, the probability of an event is interpreted through an analogy with a game of chance: for example as corresponding to the probability of drawing a black marble at random from a bag containing white and black marbles. Bernoulli's law of large numbers implies that it is possible to learn with arbitrarily high precision the probability of an event from the relative frequency of its occurrence in sufficiently many independent repetitions of an experiment.

3. Empirical Meaning of Imprecise Probabilities

The approach to frequentist probability of the *Ars Conjectandi* consists of two parts: the interpretation of probabilities by analogy with games of chance, and their estimability on the basis of finite amounts of data. In this section we will study how far this approach can be generalized to the case of imprecise probabilities. For the sake of simplicity, we will focus on a sequence of Bernoulli trials, whose outcomes are described by the binary random variables $X_1, X_2, \ldots \in \{0, 1\}$ (that is, we consider only the interpretation and estimability of imprecise probabilities of single events, not of whole imprecise probability measures on arbitrary sample spaces).

Each Bernoulli trial corresponds for instance to drawing a black or white marble (described by $X_i = 1$ or $X_i = 0$, respectively) at random from a bag containing only white and black marbles with a proportion $p_i \in [0, 1]$ of black ones (strictly speaking, all probabilities should be rational numbers, but for the sake of simplicity we will ignore this technical detail, since rational numbers are dense in the reals). The sequence of Bernoulli trials corresponds thus to drawings from a sequence of bags with possibly different proportions of black marbles. The noninteraction of the drawings corresponds to an assumption of independence of the random variables X_i, in the usual sense of (precise) probability theory (see also Chen and Wu, 2011; De Bock and de Cooman, 2012). We have a precise probability model when $p_i = p$ does not depend on i, and an imprecise one when $p_i \in [\underline{p}, \overline{p}]$ is not completely determined.

For example, in Section 2 we have considered two ways in which imprecise probability measures often appear in classical statistics. The first one, related to an intersubjective epistemic interpretation, is as descriptions of what has been learnt so far from data: this would be the case for instance if $[\underline{p}, \overline{p}]$ was obtained as a confidence interval for the precise probability p. The second one, related to a truly frequentist but limited interpretation, is as descriptions of what can potentially be learnt from infinite amounts of incomplete data: this would be the case for instance if $X_i = 1$ and $X_i = 0$ were observed with probabilities $\underline{p}$ and $1 - \overline{p}$, respectively, while with probability $\overline{p} - \underline{p}$ we would have a missing observation (independently of i). In this case, without making any assumptions about the noninformativity of the missing data, $[\underline{p}, \overline{p}]$ is the identification region of the precise probability p: that is, values of p in this interval cannot be discriminated on the basis of any amount of (incomplete) data.

In the general case without missing data, there are several possible interpretations of an imprecise probability $[\underline{p}, \overline{p}]$ with $0 \leq \underline{p} \leq \overline{p} \leq 1$ (where $\underline{p} = \overline{p}$ corresponds to the case of a degenerate interval representing a precise probability). In particular, Walley and Fine (1982) distinguish between an *ontological indeterminacy interpretation*, where

$$p_i \in [\underline{p}, \overline{p}] \tag{1}$$

is the only assumption about the sequence p_i, and an *epistemological indeterminacy interpretation*, where

$$p_i = p \in [\underline{p}, \overline{p}] \tag{2}$$

does not depend on i. The latter can also be seen as the special case in which the sequence of drawings (with replacement) is from the same bag, which contains a not completely determined proportion of black marbles. The interpretations (1) and (2) appear also in the theory of Markov chains with imprecise probabilities (which can be seen as generalizations of sequences of Bernoulli trials): for example in Hartfiel (1998) and Kozine and Utkin (2002), respectively. Moreover, the ontological indeterminacy interpretation (1) plays a prominent role in the theory of probabilistic graphical models with imprecise probabilities (which can be seen as further generalizations of Markov chains): see for instance Cozman (2005).

From the point of view of the estimability of the imprecise probability $[\underline{p}, \overline{p}]$, both interpretations (1) and (2) are problematic, because in general the sequence p_i does not determine the interval $[\underline{p}, \overline{p}]$. That is, with these interpretations the imprecise probability is only partially identified and therefore cannot in general be estimated with arbitrarily high precision. In order to make the imprecise probability identifiable, we can interpret it as allowing only the sequences $p_i \in [\underline{p}, \overline{p}]$ that determine in a certain sense the interval $[\underline{p}, \overline{p}]$. However, we would most likely betray the intuitive meaning of imprecise probabilities if we would exclude any starting sequence $p_1, \ldots, p_n \in [\underline{p}, \overline{p}]$. Similarly, assigning some kind of degree of plausibility to the starting sequences $p_1, \ldots, p_n \in [\underline{p}, \overline{p}]$ would also lead to a new model, different from the one of imprecise probabilities (such as the chaotic probability model of Fierens et al., 2009).

On the basis of these considerations, we obtain an *identifiable ontological indeterminacy interpretation*, where

$$p_i \in [\underline{\alpha}(p_1, p_2, \ldots),\, \overline{\alpha}(p_1, p_2, \ldots)] = [\underline{p}, \overline{p}] \tag{3}$$

is a condition on the sequence p_i, determined by two functions $\underline{\alpha}, \overline{\alpha} : [0,1]^{\mathbb{N}} \to [0,1]$ that do not depend on any finite number of their arguments (that is, each function would assign the same value to sequences differing only at a finite number of positions). These functions are considered to be fixed, but we do not need to further specify them in order to obtain the results of the present paper (that is, these results are valid for any particular choice of the above functions $\underline{\alpha}, \overline{\alpha}$). An example of such pairs of functions is the limits inferior and superior of the sequence p_i, implying that the whole width of the interval

$$[\underline{p}, \overline{p}] = \left[\liminf_{i\to\infty} p_i,\, \limsup_{i\to\infty} p_i\right] \tag{4}$$

is used by the sequence p_i, and infinitely many times (in the sense that the sequence gets infinitely many times arbitrarily close to both endpoints of the interval). A related example is the limits inferior and superior of the Cesàro means of the sequence p_i (that is, the limits inferior and superior of the averages of the starting sequences $p_1, \ldots, p_n$), implying that the whole width of the interval

$$[\underline{p}, \overline{p}] = \left[\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} p_i,\, \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} p_i\right] \tag{5}$$

is used by the sequence p_i, not only infinitely many times, but also not too rarely (in order to bring not only the sequence, but also its Cesàro means infinitely many times arbitrarily close to both endpoints of the interval).

However, it is intuitively clear that imprecise probabilities are not estimable in full generality, because for instance any finite amount of data from Bernoulli trials would always be perfectly compatible with the vacuous imprecise probability $[0,1]$, independently of the considered interpretation (1), (2), or (3). This difficulty in discriminating between the vacuous and other imprecise probabilities is related to the more general difficulty in comparing imprecise probability models with different degrees of imprecision (see also Seidenfeld et al., 2011; Cattaneo, 2013). Anyway, imprecise probabilities are estimable under additional assumptions about the possible intervals $[\underline{p}, \overline{p}]$. Let $\mathcal{I}$ be the set of all imprecise probabilities that are considered possible in a given situation: that is, let $\mathcal{I}$ be a nonempty set of intervals of the form $[\underline{p}, \overline{p}]$ with $0 \leq \underline{p} \leq \overline{p} \leq 1$.

Bernoulli's law of large numbers implies the *uniformly consistent estimability* of the precise probability $p_i = p \in [0,1]$ on the basis of the outcomes $X_1, X_2, \ldots$ of the Bernoulli trials. An estimator $\underline{\pi}_n : \{0,1\}^n \to [0,1]$ (or more precisely, a sequence of estimators $\underline{\pi}_n$) of $\underline{p}$ is said to be uniformly consistent when for all $\varepsilon > 0$ and all $\delta > 0$ there is an N such that

$$P\left(\left|\underline{\pi}_n(X_1, \ldots, X_n) - \underline{p}\right| > \varepsilon\right) \leq \delta \tag{6}$$

for all $n \geq N$, all $[\underline{p}, \overline{p}] \in \mathcal{I}$, and all (precise) probability measures P corresponding to the sequences p_i compatible with the imprecise probability $[\underline{p}, \overline{p}]$ according to the considered interpretation (1), (2), or (3) (since we are in the setting of games of chance, the interpretation of P is unproblematic: probabilities are ratios between the numbers of favorable and possible outcomes). The definition of a uniformly consistent estimator $\overline{\pi}_n$ of $\overline{p}$ is analogue, and $[\underline{p}, \overline{p}] \in \mathcal{I}$ is said to be uniformly consistently estimable when there are uniformly consistent estimators $\underline{\pi}_n$ and $\overline{\pi}_n$ of $\underline{p}$ and $\overline{p}$, respectively.

The uniform consistency of an estimator $\underline{\pi}_n$ of $\underline{p}$ is particularly important, because it implies that $[\underline{\pi}_N(X_1, \ldots, X_N) - \varepsilon,\, \underline{\pi}_N(X_1, \ldots, X_N) + \varepsilon]$ is a confidence interval for $\underline{p}$ with coverage probability at least $1 - \delta$ (an analogous result is implied by the uniform consistency of an estimator $\overline{\pi}_n$ of $\overline{p}$). That is, uniformly consistent estimators provide us with arbitrarily short confidence intervals of arbitrarily high confidence level, when we have a sufficiently large amount of data. In this sense, uniformly consistent estimability endows imprecise probabilities $[\underline{p}, \overline{p}] \in \mathcal{I}$ with a clear empirical meaning. However, the next theorem states that this is the case only when all nondegenerate intervals in $\mathcal{I}$ are isolated in $\mathcal{I}$. An interval $[\underline{p}, \overline{p}] \in \mathcal{I}$ is said to be nondegenerate when $\underline{p} < \overline{p}$, and it is said to be isolated in $\mathcal{I}$ when there is a $\gamma > 0$ such that $[\underline{p} - \gamma,\, \overline{p} + \gamma]$ does not intersect any other interval in $\mathcal{I}$ (degenerate or nondegenerate). If all nondegenerate intervals in $\mathcal{I}$ are isolated in $\mathcal{I}$, then all intervals in $\mathcal{I}$ (degenerate or nondegenerate) are pairwise disjoint, while the converse is not true. For example, if $\mathcal{I}$ consists of the nondegenerate interval $[0, \frac{1}{2}]$ and all the degenerate intervals $[p, p]$ with $\frac{1}{2} < p \leq 1$, then all elements of $\mathcal{I}$ are pairwise disjoint, but $[0, \frac{1}{2}]$ is not isolated in $\mathcal{I}$.

Theorem 1 *The following four statements are equivalent:*

(i) $[\underline{p}, \overline{p}] \in \mathcal{I}$ *is uniformly consistently estimable under the ontological indeterminacy interpretation (1),*

(ii) $[\underline{p}, \overline{p}] \in \mathcal{I}$ *is uniformly consistently estimable under the epistemological indeterminacy interpretation (2),*

(iii) $[\underline{p}, \overline{p}] \in \mathcal{I}$ *is uniformly consistently estimable under the identifiable ontological indeterminacy interpretation (3),*

(iv) all nondegenerate intervals in $\mathcal{I}$ *are isolated in* $\mathcal{I}$.

Theorem 1 implies in particular Bernoulli's law of large numbers, which corresponds to the case where $\mathcal{I}$ is the set of all degenerate intervals $[p, p]$ with $p \in [0, 1]$. More precisely, Bernoulli (1713, 2006) proved the result only in the case where $\mathcal{I}$ is the set of the $m + 1$ degenerate intervals $[p, p]$ such that $p \in [0, 1]$ is a rational number with (arbitrarily large) denominator m. For this case, he also provided an explicit way of calculating a value for the quantity N appearing in the definition of uniform consistency (6), thus obtaining a clear empirical meaning for precise probabilities through what we now call confidence intervals. Anyway, Theorem 1 shows that this is possible for imprecise probabilities only in very limited settings, independently of their exact interpretation.

In order to endow imprecise probabilities with a clear empirical meaning in more general settings, we can moderate our requirements for their estimability. In particular, Walley and Fine (1982) introduced the concept of *strong estimability*, which weakens uniformly consistent estimability (6) by allowing N to depend on the interval $[\underline{p}, \overline{p}]$, besides on ε and δ. When we weaken strong estimability further by allowing N to depend also on the probability measure P, we get the concept of *consistent estimability*. That is, strong estimability lies between consistent estimability and uniformly consistent estimability, and must not be confused with strongly consistent estimability (which corresponds to consistent estimability when convergence in probability is replaced by almost sure convergence). Anyway, strong estimability can also be interpreted as the generalization of consistent estimability to imprecise probabilities: in fact, strong estimability and consistent estimability are equivalent when all intervals in $\mathcal{I}$ are degenerate (that is, in the case of precise probabilities).

Theorem 2 *The following four statements are equivalent:*

(i) $[\underline{p}, \overline{p}] \in \mathcal{I}$ *is strongly estimable under the ontological indeterminacy interpretation (1),*

(ii) $[\underline{p}, \overline{p}] \in \mathcal{I}$ *is strongly estimable under the epistemological indeterminacy interpretation (2),*

(iii) $[\underline{p}, \overline{p}] \in \mathcal{I}$ *is strongly estimable under the identifiable ontological indeterminacy interpretation (3),*

(iv) all intervals in $\mathcal{I}$ *(degenerate or nondegenerate) are pairwise disjoint.*

Contrary to uniformly consistent estimability, strong estimability does not guarantee the existence of arbitrarily short confidence intervals of arbitrarily high confidence level for $\underline{p}$ and $\overline{p}$, but is nonetheless important because it is required in order for imprecise probabilities to be empirically recognizable, in the following sense. Given an imprecise probability $[\underline{p}, \overline{p}] \in \mathcal{I}$ and a desired level of precision for the estimators, we can choose n such that if the data $X_1, \ldots, X_n$ are generated according to $[\underline{p}, \overline{p}]$ (that is, according to any sequence p_i compatible with it), then $[\underline{p}, \overline{p}]$ can be estimated to the desired level of precision on the basis of $X_1, \ldots, X_n$ (in other words, an imprecise probability can be recognized arbitrarily well on the basis of finite amounts of data generated according to it). However, Theorem 2 shows that imprecise probabilities are empirically recognizable only in very limited settings, independently of their exact interpretation. In fact, requiring only strong estimability instead of uniformly consistent estimability as in Theorem 1 weakened only slightly the necessary and sufficient condition on $\mathcal{I}$. As a side result, the following corollary of Theorem 2 completes a basic result of Walley and Fine (1982) about the strong estimability of imprecise probability measures on finite sample spaces.

Corollary 3 *The necessary condition in Theorem 5.1 of Walley and Fine (1982) is sufficient as well, also in the case of infinitely many imprecise probability measures.*

Although consistent estimability (with respect to precise probability measures) is too weak to endow imprecise probabilities with a clear empirical meaning (in the sense that it does not guarantee their empirical recognizability), for completeness we can look at the consequences of requiring only this level of estimability. The next theorem shows that there is no difference between consistent estimability and strong estimability of imprecise probabilities, when only the interpretations (1) and (2) are considered. However, there is a difference when the identifiable ontological indeterminacy interpretation (3) is considered, as we will see in a moment.

Theorem 4 *The following three statements are equivalent:*

(i) $[\underline{p}, \overline{p}] \in \mathcal{I}$ *is consistently estimable under the ontological indeterminacy interpretation (1),*

(ii) $[\underline{p}, \overline{p}] \in \mathcal{I}$ *is consistently estimable under the epistemological indeterminacy interpretation (2),*

(iii) all intervals in $\mathcal{I}$ *(degenerate or nondegenerate) are pairwise disjoint.*

Theorems 1, 2, and 4 give necessary and sufficient conditions for the estimability of imprecise probabilities, but there is a difference between knowing that something is estimable and knowing how to estimate it. The next theorem closes this gap by explicitly giving examples of estimators with the required properties.

Theorem 5 *The following estimators of $\underline{p}$ and $\overline{p}$ satisfy all the properties considered in Theorems 1, 2, and 4, when the corresponding necessary and sufficient conditions on $\mathcal{I}$ are fulfilled:*

$$\underline{\pi}_n(x_1,\ldots,x_n) = \inf\left\{\underline{p} : [\underline{p},\overline{p}] \in \mathcal{I},\ \overline{p} + c_n > \frac{1}{n}\sum_{i=1}^n x_i\right\}, \tag{7}$$

$$\overline{\pi}_n(x_1,\ldots,x_n) = \sup\left\{\overline{p} : [\underline{p},\overline{p}] \in \mathcal{I},\ \underline{p} - c_n < \frac{1}{n}\sum_{i=1}^n x_i\right\}, \tag{8}$$

for all $x_1,\ldots,x_n \in \{0,1\}$, where c_n is any sequence of real numbers such that $\lim_{n\to\infty} c_n = 0$ and $\lim_{n\to\infty} \sqrt{n}\, c_n = +\infty$, while $\inf\varnothing$ and $\sup\varnothing$ can be defined arbitrarily.

The estimators (7) and (8) exploit the fact that the relative frequency $\frac{1}{n}\sum_{i=1}^n X_i$ of the occurrence of the event $X_i = 1$ will lie in $[\underline{p} - c_n,\, \overline{p} + c_n]$ with arbitrarily high probability when n is sufficiently large, independently of the considered interpretation (1), (2), or (3). Theorems 4 and 5 imply that when all intervals in $\mathcal{I}$ (degenerate or nondegenerate) are pairwise disjoint, the estimators (7) and (8) are also consistent under the identifiable ontological indeterminacy interpretation (3), since this property is weaker than the consistency under the ontological indeterminacy interpretation (1). However, the next theorem implies that the pairwise disjointness of the intervals in $\mathcal{I}$ is not a necessary condition for the consistent estimability of $[\underline{p},\overline{p}] \in \mathcal{I}$ under the identifiable ontological indeterminacy interpretation (3), because it is sufficient that all nondeterministic intervals in $\mathcal{I}$ (degenerate or nondegenerate) are pairwise disjoint. An interval $[\underline{p},\overline{p}] \in \mathcal{I}$ is said to be nondeterministic if it is not one of the two degenerate intervals $[0,0]$ and $[1,1]$.

Theorem 6 *A sufficient condition for $[\underline{p},\overline{p}] \in \mathcal{I}$ to be consistently estimable under the identifiable ontological indeterminacy interpretation (3) is that all nondeterministic intervals in $\mathcal{I}$ (degenerate or nondegenerate) are pairwise disjoint, while a necessary condition is that $\mathcal{I}$ does not contain at the same time the interval $[0,1]$ and another nondeterministic interval (degenerate or nondegenerate).*

The following estimators of $\underline{p}$ and $\overline{p}$ are consistent under the identifiable ontological indeterminacy interpretation (3), when the above sufficient condition on $\mathcal{I}$ is fulfilled:

$$\underline{\pi}'_n(x_1,\ldots,x_n) = \begin{cases} 1 & \text{if } x_1 = \cdots = x_n = 1, \\ \underline{\pi}_n(x_1,\ldots,x_n) & \text{otherwise,} \end{cases} \tag{9}$$

$$\overline{\pi}'_n(x_1,\ldots,x_n) = \begin{cases} 0 & \text{if } x_1 = \cdots = x_n = 0, \\ \overline{\pi}_n(x_1,\ldots,x_n) & \text{otherwise,} \end{cases} \tag{10}$$

for all $x_1,\ldots,x_n \in \{0,1\}$, where $\underline{\pi}_n$ and $\overline{\pi}_n$ are the estimators (7) and (8), respectively.

Theorems 1, 2, and 4 characterize three different levels of estimability of imprecise probabilities according to three different ways of interpreting them. Only one of the nine possible characterizations is missing: the one of consistent estimability under the identifiable ontological indeterminacy interpretation (3), because the necessary and sufficient conditions in Theorem 6 are different. In fact, this characterization seems to be much more difficult than the other eight, also because the exact meaning of the interpretation (3) depends on the functions $\underline{\alpha}, \overline{\alpha}$ considered. In particular, for the limits inferior and superior of the sequence p_i (4), it seems plausible that the sufficient condition of Theorem 6 is also necessary, but the proof does not seem to be straightforward.

In general, the results of the present section show that an empirical interpretation of imprecise probabilities is possible only in very limited settings, because imprecise probabilities cannot be estimated satisfactorily on the basis of finite amounts of data. This is hardly surprising when considering that imprecise probabilities are not identifiable in general under the interpretations (1) and (2), and only asymptotically identifiable under the interpretation (3). For these reasons, it can be interesting to study the estimability of the actual, finite-sample imprecise probabilities: that is, the estimability of $\min\{p_1, \ldots, p_n\}$ and $\max\{p_1, \ldots, p_n\}$ on the basis of the outcomes $X_1, \ldots, X_n$ of the corresponding Bernoulli trials.

The concepts of uniformly consistent estimability, strong estimability, and consistent estimability of the finite-sample imprecise probabilities $[\min\{p_1, \ldots, p_n\}, \max\{p_1, \ldots, p_n\}]$ can be obtained by replacing $\underline{p}$ with $\min\{p_1, \ldots, p_n\}$ in (6), and $\overline{p}$ with $\max\{p_1, \ldots, p_n\}$ in the analogue expression for $\overline{\pi}_n$ (the resulting concepts generalize the usual ones, since $\min\{p_1, \ldots, p_n\}$ and $\max\{p_1, \ldots, p_n\}$ are not necessarily constant). The next theorem implies that also the finite-sample imprecise probabilities have a very limited empirical meaning, since they can be estimated satisfactorily only when they are known to be precise.

Theorem 7 *The following six statements are equivalent:*

(i) $[\min\{p_1, \ldots, p_n\}, \max\{p_1, \ldots, p_n\}]$ *is uniformly consistently estimable under the ontological indeterminacy interpretation (1) of* $[\underline{p}, \overline{p}] \in \mathcal{I}$*,*

(ii) $[\min\{p_1, \ldots, p_n\}, \max\{p_1, \ldots, p_n\}]$ *is uniformly consistently estimable under the identifiable ontological indeterminacy interpretation (3) of* $[\underline{p}, \overline{p}] \in \mathcal{I}$*,*

(iii) $[\min\{p_1, \ldots, p_n\}, \max\{p_1, \ldots, p_n\}]$ *is strongly estimable under the ontological indeterminacy interpretation (1) of* $[\underline{p}, \overline{p}] \in \mathcal{I}$*,*

(iv) $[\min\{p_1, \ldots, p_n\}, \max\{p_1, \ldots, p_n\}]$ *is strongly estimable under the identifiable ontological indeterminacy interpretation (3) of* $[\underline{p}, \overline{p}] \in \mathcal{I}$*,*

(v) $[\min\{p_1, \ldots, p_n\}, \max\{p_1, \ldots, p_n\}]$ *is consistently estimable under the ontological indeterminacy interpretation (1) of* $[\underline{p}, \overline{p}] \in \mathcal{I}$*,*

(vi) all intervals in $\mathcal{I}$ *are degenerate.*

The estimability of $[\min\{p_1, \ldots, p_n\}, \max\{p_1, \ldots, p_n\}]$ under the epistemological indeterminacy interpretation (2) is uninteresting, since it corresponds to the estimability of precise probabilities, which is implied by Bernoulli's law of large numbers. Of the other six possible characterizations, the only one missing is again the one of consistent estimability under the identifiable ontological indeterminacy interpretation (3): in fact, it can be shown that under this interpretation, the consistent estimabilities of $[\min\{p_1, \ldots, p_n\}, \max\{p_1, \ldots, p_n\}]$ and $[\underline{p}, \overline{p}]$ are equivalent, and so we are back to the difficulties of Theorem 6.

4. Conclusion

We have seen that in particular situations involving incomplete data, imprecise probabilities can have a clear empirical meaning as identification regions of frequentist, precise probabilities. Unfortunately, such situations are exceptional, and imprecise probabilities do not have a generally valid, clear empirical meaning, in the sense discussed in this paper.

Imprecise probabilities can be interpreted in several ways in terms of precise probabilities, as done for example in the imprecise versions of the theories of Markov chains and probabilistic graphical models. However, all these interpretations have a very limited empirical meaning, since imprecise probabilities are strongly estimable (that is, empirically recognizable) only in situations in which they are known to belong to a given set of pairwise disjoint imprecise probabilities. These results get even worse when we consider the actual, finite-sample imprecise probabilities, instead of the virtual, asymptotic ones. Anyway, examples of estimators have been given explicitly in this paper for the cases in which imprecise probabilities are satisfactorily estimable.

A mathematically interesting open problem is the question for a necessary and sufficient condition on a set of possibly degenerate probability intervals $[\underline{p}, \overline{p}]$, in order for them to be consistently estimable on the basis of any sequence of independent Bernoulli trials with precise probabilities of success $p_i \in [\underline{p}, \overline{p}]$ such that the sequence p_i has $\underline{p}$ and $\overline{p}$ as limits inferior and superior, respectively.

Acknowledgments

The author wishes to thank the anonymous referees for their valuable comments and suggestions.

References

A. Antonucci, M. Cattaneo, and G. Corani. Likelihood-based robust classification with Bayesian networks. In S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. R. Yager, editors, *Advances in Computational Intelligence*, volume 3, pages 491–500. Springer, 2012.

A. Arnauld and P. Nicole. *La logique ou l'art de penser*. Savreux, 1662.

A. Arnauld and P. Nicole. *Logic or the Art of Thinking*. Cambridge University Press, 1996.

J. Bernoulli. *Ars Conjectandi*. Thurneysen Brothers, 1713.

J. Bernoulli. *The Art of Conjecturing*. Johns Hopkins University Press, 2006.

M. Cattaneo. On the robustness of imprecise probability methods. In F. G. Cozman, T. Denœux, S. Destercke, and T. Seidfenfeld, editors, *ISIPTA '13*, pages 33–41. SIPTA, 2013.

M. Cattaneo and A. Wiencierz. Likelihood-based Imprecise Regression. *International Journal of Approximate Reasoning*, 53:1137–1154, 2012.

Z. Chen and P. Wu. Strong laws of large numbers for Bernoulli experiments under ambiguity. In S. Li, X. Wang, Y. Okazaki, J. Kawabe, T. Murofushi, and L. Guan, editors, *Nonlinear Mathematics for Uncertainty and its Applications*, pages 19–30. Springer, 2011.

F. G. Cozman. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39:167–184, 2005.

F. G. Cozman and L. Chrisman. Learning convex sets of probability from data. Technical Report CMU-RI-TR 97-25, Robotics Institute, Carnegie Mellon University, 1997.

J. De Bock and G. de Cooman. Imprecise Bernoulli processes. In S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. R. Yager, editors, *Advances in Computational Intelligence*, volume 3, pages 400–409. Springer, 2012.

G. de Cooman and E. Miranda. Weak and strong laws of large numbers for coherent lower previsions. *Journal of Statistical Planning and Inference*, 138:2409–2432, 2008.

B. de Finetti. Sul significato soggettivo della probabilità. *Fundamenta Mathematicae*, 17:298–329, 1931.

B. de Finetti. *Theory of Probability. A Critical Introductory Treatment*, 2 volumes. Wiley, 1974–1975.

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

P. I. Fierens, L. C. Rêgo, and T. L. Fine. A frequentist understanding of sets of measures. *Journal of Statistical Planning and Inference*, 139:1879–1892, 2009.

D. Gillies. Popper's contribution to the philosophy of probability. In A. O'Hear, editor, *Karl Popper: Philosophy and Problems*, pages 103–120. Cambridge University Press, 1995.

D. Gillies. *Philosophical Theories of Probability*. Routledge, 2000.

D. J. Hartfiel. *Markov Set-Chains*. Springer, 1998.

I. O. Kozine and L. V. Utkin. Interval-valued finite Markov chains. *Reliable Computing*, 8:97–113, 2002.

C. F. Manski. *Partial Identification of Probability Distributions*. Springer, 2003.

M. Marinacci. Limit laws for non-additive probabilities and their frequentist interpretation. *Journal of Economic Theory*, 84:145–195, 1999.

S. Peng. Nonlinear expectations and stochastic calculus under uncertainty—with robust central limit theorem and G-Brownian motion. arXiv:1002.4546 [math.PR], 2010.

K. Popper. *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft*. Springer, 1935.

K. Popper. *The Logic of Scientific Discovery*. Hutchinson, 1959.

H. Reichenbach. *Wahrscheinlichkeitslehre. Eine Untersuchung über die logischen und mathematischen Grundlagen der Wahrscheinlichkeitsrechnung*. Luitingh-Sijthoff, 1935.

H. Reichenbach. *The Theory of Probability. An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. University of California Press, 1949.

T. Seidenfeld, M. J. Schervish, and J. B. Kadane. Forecasting with imprecise probabilities. In F. Coolen, G. de Cooman, T. Fetz, and M. Oberguggenberger, editors, *ISIPTA '11*, pages 317–326. SIPTA, 2011.

G. Shafer. Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences*, 19:309–370, 1978.

E. D. Sylla. Tercentenary of *Ars Conjectandi* (1713): Jacob Bernoulli and the founding of mathematical probability. *International Statistical Review*, 82:27–45, 2014.

M. C. M. Troffaes and G. de Cooman. *Lower Previsions*. Wiley, 2014.

J. Venn. *The Logic of Chance. An Essay on the Foundations and Province of the Theory of Probability, with Especial Reference to Its Application to Moral and Social Science*. Macmillan, 1866.

R. von Mises. *Wahrscheinlichkeit Statistik und Wahrheit*. Springer, 1928.

R. von Mises. *Probability, Statistics and Truth*. Allen & Unwin, 2nd edition, 1957.

P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, 1991.

P. Walley and T. L. Fine. Towards a frequentist theory of upper and lower probability. *The Annals of Statistics*, 10:741–761, 1982.

P. M. Williams. Notes on conditional previsions. Technical report, School of Mathematics and Physical Sciences, University of Sussex, 1975.

P. M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44:366–383, 2007.