Proceedings of Machine Learning Research 215:72–83, 2023 ISIPTA 2023

Dynamic Precise and Imprecise Probability Kinematics

Michele Caprio CAPRIO@SEAS.UPENN.EDU
PRECISE Center, Computer and Information Science, University of Pennsylvania, USA

Ruobin Gong RUOBIN.GONG@RUTGERS.EDU
Statistics, Rutgers University, USA

Abstract

We introduce dynamic probability kinematics (DPK), a method for an agent to mechanically update subjective beliefs in the presence of partial information. We then generalize DPK to dynamic imprecise probability kinematics (DIPK), which allows the agent to express their initial beliefs via a set of probabilities in order to further take ambiguity into account. Examples are provided to illustrate how the methods work.

Keywords: subjective probability, Jeffrey's updating, imprecise probabilities, probability kinematics, Bayes' rule

1. Introduction

Updating an opinion on the likelihood of an event when new data becomes available is one of the most natural tasks we perform daily. The goal of this paper is to introduce a method to update mechanically the subjective beliefs of an agent that faces ambiguity and who is only able to collect partial information.

With the former, we mean that a single probability measure is not enough to encapsulate the agent's initial beliefs, a very common and well documented situation [32, Section 1.1.4]; we inspect ambiguity in Section 1.1. Partial information means that the agent cannot collect crisp evidence; rather, they gather information whose nature is probabilistic. Our updating mechanism is based on probability kinematics (PK), an updating rule expressly conceived to deal with partial information. We inspect probability kinematics and its relation with the procedure we present in Section 1.2.

We call the method we propose dynamic imprecise probability kinematics (DIPK). It is framed within the credal sets theory paradigm [22]. In this field, a set of probability measures (called a credal set) is used to capture either the ambiguity initially faced by the agent, or inconsistency/imprecision in the process of collecting data. To derive DIPK, we first assume that the agent does not face ambiguity. We come up with a simpler updating technique that we call dynamic probability kinematics (DPK), and then we generalize it by requiring the agent to specify a set $\mathcal{P}$ of probability measures representing their initial beliefs. DIPK is especially useful because it allows the update to be performed mechanically: the agent only needs to specify $\mathcal{P}$. To the best of our knowledge, this is the first time a PK-rooted mechanical procedure to update subjective beliefs in the presence of ambiguity and partial information within the credal sets theory paradigm is presented.

1.1. Ambiguity

Precise probabilities are widely employed as the central vocabulary of many modes of uncertainty reasoning, nearly exclusively so in statistical inference, for example. In the subjective probability literature, the agent's initial beliefs about an event $A \subset \Omega$ are usually encapsulated in a single probability measure, that is then refined once new information in the form of data become available. As Walley points out in [32, Section 1.1.4], though, missing information and bounded rationality may prevent the agent from assessing probabilities precisely in practice, even if doing so is possible in principle. This may be due to the lack of information on how likely events of interest are, lack of computational time or ability, or because it is extremely difficult to analyze a complex body of evidence. We call this condition faced by the agent *ambiguity* [13]. Oftentimes agents do not realize they face ambiguity, as observed in [3] and in the de Finetti lecture delivered at ISBA 2021. There, Berger points out how most people tend to under-report variance; the folklore says by a factor of 3. People simply think that they know more than they actually do.

In the presence of ambiguity, the agent may only be able to specify a set $\mathcal{P}$ of probability measures that seem "plausible" or "fit" to express their initial opinion on the events of interest. Generally speaking, the farther apart (e.g. in the total variation distance) the "boundary elements" of $\mathcal{P}$ (i.e. its infimum and supremum), the higher the ambiguity faced by the agent. This way of proceeding, called the *sensitivity analysis approach*, is further examined in Section 6.1.

As Section 6 will discuss, the infima of the sets updated according to our DIPK procedure – that, as we shall see, are called lower probabilities – completely characterize the sets. That is why in [7, Section 7] the authors give lower and upper bounds for the updated lower and upper probabilities (the conjugate of lower probabilities), respectively, and in [7, Section 8] they study the behavior of the updated sets (contraction, dilation, sure loss) by giving sufficient conditions involving lower (and upper) probabilities.

1.2. Probability Kinematics

DPK and DIPK are rooted in probability kinematics (PK), also known as Jeffrey's rule of updating. PK can be seen as a generalization of Bayesian updating, the most famous and widely used technique to describe updating of beliefs. This latter prescribes the scholar to form an initial opinion on the plausibility of the event A of interest, where A is a subset of the state space Ω, and to express it by specifying a probability measure P, so that $P(A)$ can be quantified. Once some data E is collected, the Bayesian updating mechanism revises the initial opinion by applying the Bayes' rule $P^\star(A) \equiv P(A \mid E) = P(A \cap E)/P(E) = [P(E \mid A)P(A)]/P(E) \propto P(E \mid A)P(A)$, provided that $P(E) \neq 0$.[1] In [18, 19, 20], Richard Jeffrey makes a compelling case that Bayes' rule is not the only reasonable way of updating. Its use presupposes that both $P(E)$ and $P(A \cap E)$ have been quantified before event E takes place: this can be a very challenging task, for instance when E is not anticipated. Jeffrey points out that evidence is not always propositional (i.e. it may not be possible to represent it as a crisp subset); rather, it is oftentimes uncertain or partial.[2]

Bayes' rule is not well-suited for the agent who possesses partial information. The following example illustrates a situation in which Bayes' rule is not *directly* applicable to compute the updated probability of an event (we would need to enlarge the state space), but Jeffrey's rule is.

Example 1 *[11, Section 1.1] Three trials of a new surgical procedure are to be conducted at a hospital. Let* 1 *denote a successful outcome, and* 0 *an unsuccessful one. The state space has the form* $\Omega = \{000, 001, 010, 011, 100, 101, 110, 111\}$. *A colleague informs us that another hospital performed this type of procedure* 100 *times, registering* 80 *successful outcomes. This information is relevant and should influence our opinion about the outcome of the three trials, but it cannot be put in direct terms of the occurrence of an event in the original* Ω, *thus Bayes' rule is not directly applicable.*

Since the description contains no information about the order of the three trials, our initial opinion P *assumes that they are exchangeable. That is, consider the partition* $\{E_0, E_1, E_2, E_3\}$ *of* Ω *where* E_j *is the set of all outcomes with exactly* j *successes, exchangeability implies that we assign equal probabilities to atomic events within each partition. In other words,* $P(\{001\}) = P(\{100\}) = P(\{010\})$ *and* $P(\{110\}) = P(\{101\}) = P(\{011\})$.

The success rate at the other hospital informs our opinion over the partition $\{E_j\}$ *only, and nothing more. In relation to our old opinion* P, *our updated opinion* $P^\star$ *satisfies* $P(A \mid E_j) = P^\star(A \mid E_j)$ *for all* $A \subset \Omega$ *and all* $j \in \{0, 1, 2, 3\}$. *Upon specifying a new subjective assessment of the* $P^\star(E_j)$*'s, the updated probability measure* $P^\star$ *can be fully reassessed by the relation* $P^\star(A) = \sum_{j=0}^{3} P^\star(A \mid E_j)P^\star(E_j) = \sum_{j=0}^{3} P(A \mid E_j)P^\star(E_j)$. *It is within our liberty to reassess the* $P^\star(E_j)$*'s. We may, for example, regard the three trials as a random subsample of size three from those of the other hospital. This would equate* $P^\star(E_j)$ *to the probability of obtaining* j *successes from a Hypergeometric*$(100, 80, 3)$ *distribution.* △

The rule $P^\star(A) = \sum_{E_j \in \mathcal{E}} P(A \mid E_j)P^\star(E_j)$ is known as Jeffrey's rule of conditioning. It is valid when there is a partition $\mathcal{E}$ of the state space Ω such that $P^\star(A \mid E_j) = P(A \mid E_j)$, for all $A \subset \Omega$ and all $E_j \in \mathcal{E}$. Under this assumption, if $\mathcal{E}$ is finite Jeffrey's rule is a consequence of coherence [32, Section 6.11.8]. It is useful when new evidence cannot be identified with the occurrence of an event, but has the effect of changing the probabilities we assign to the events in partition $\mathcal{E}$. It has the practical advantage of reducing the assessment of $P^\star$ to the simpler task of assessing $P^\star(E_j)$, for all $E_j \in \mathcal{E}$. In the above example, instead of a full reassessment of probabilities on Ω, the agent only needs to deliberate new assessment of the four probabilities $P^\star(E_0)$ through $P^\star(E_3)$ based on the given information.

To see that Jeffrey's rule of conditioning is a generalization of Bayes' rule, consider partition $\{E, E^c\}$ for some $E \subset \Omega$. If $P^\star(E) = 1$, we have that $P^\star(A) = P(A \mid E)P^\star(E) + P(A \mid E^c)P^\star(E^c) = P(A \mid E)$, which is Bayes' rule. Unlike Bayesian conditionalization, however, if we are given the pair $\{P, P^\star\}$ of probability measures, we can always reconstruct a partition $\{E_j\}$ for which $\{P, P^\star\}$ could have arisen via Jeffrey's rule [11, Section 2].

Let us now discuss the relation between DPK and Jeffrey's updating. The three main tasks in PK are: (1) collecting a partition $\mathcal{E}$ of state space Ω; (2) subjectively assess the probability $P^\star(E)$ to attach to the elements E of partition $\mathcal{E}$; (3) compute the update $P^\star(A) = \sum_{E \in \mathcal{E}} P(A \mid E)P^\star(E)$. In DPK, we: (1') collect data points belonging to a generic set $\mathcal{X}$ that induce a partition $\mathcal{E}$ of state space Ω; (2') mechanically attach probabilities to the elements of the induced partition; (3') compute the update as in "regular" PK. We allow the evidence observed by the agent to belong to a general set $\mathcal{X}$; data points are regarded as the realization

[1] Conditioning on a zero probability event is technically possible, see e.g. literature on lexicographic probability [6] and layers of zero probabilities [10]. We will consider this eventuality in future work.

[2] Notice that when introducing PK, Jeffrey was not concerned about probabilities being precise: this is one of the main reasons why we introduce DIPK in Section 6.

of a random variable $X : \Omega \to \mathcal{X}$. Notice that if the distribution P_X of X were to be known, the elements of $\mathcal{X}$ would induce a unique partition $\mathcal{E} = \{E_j\}$ of Ω, where $E_j = \{\omega \in \Omega : X(\omega) = x_j\}$ and $P^\star(E_j) = P_X(\{x_j\})$, for all $x_j \in \mathcal{X}$. Instead, to further capture the idea of partial information, we consider the case where P_X is unknown. As we shall see, given data points $x_1, \dots, x_n \in \mathcal{X}$, they induce a partition $\mathcal{E} = \{E_j\}_{j=1}^{m+1}$, $m \leq n$, where m is the number of unique elements in $\{x_1, \dots, x_n\}$, $E_j = \{\omega \in \Omega : X(\omega) = x_j\}$ for $j \in \{1, \dots, m\}$, and $E_{m+1} = (\cup_{j=1}^m E_j)^c$. The relative frequency of $x_1, \dots, x_n$ will induce the probability that the agent assigns to the elements of $\mathcal{E}$, making the update from P to $P^\star$ mechanical.

1.3. Structure of the Paper

The paper is organized as follows. In Section 2, we discuss the connection between our work and three important papers. Further related literature is inspected in Appendix A. Sections 3 and 4 introduce dynamic probability kinematics (DPK). In Section 5, we explain how to subsequently update probability measure P as more and more data become available. Section 6 presents dynamic imprecise probability kinematics (DIPK), and Section 7 presents two examples that illustrate how to implement DPK and DIPK. Section 8 concludes, and Section 9 contains the proofs of our results.

2. Related Literature

In [26], the authors generalize Jeffrey's rule to credal sets theory; they introduce imaginary kinematics [26, Definition 7]. They combine Jeffrey's rule with Lewis' imaging [23] for credal sets to be able to update beliefs when possibly inconsistent probabilistic evidence is gathered. Evidence on some variables is called *inconsistent* when it contradicts certainty (or impossibility) in the agent's knowledge base. There are two main differences between our work and [26]. First, we consider an agent facing ambiguity who specifies a set of probability measures that encapsulates their initial beliefs, while [26] do not. Second, [26] considers the instance in which gathered evidence is partial and possibly inconsistent, while we only deal with the former. In the future we will generalize DIPK by relaxing the (tacit) assumption that the gathered evidence is consistent.

In [8] the authors provide an ergodic theory for the limit of a sequence of successive DIPK updates of a set representing the initial beliefs of an agent. As a consequence, they formulate a strong law of large numbers. These results are instrumental to increase the applicability of DIPK; for example, they underpin generalizations of classical Markov Chain Monte Carlo procedures that allow for DIPK updating.

We conclude this Section by mentioning the remarkable work by Diaconis and Zabell [11]. Their paper inspired our effort to derive a mechanical version of PK. In particular, on top of borrowing Example 1, we adapt their results to show that DPK is not commutative, and that DPK updates can be obtained by Bayesian updating in a larger space $\Omega' \supset \Omega$ (see Section 5).

3. A New Way of Updating Subjective Beliefs

In this and in the next Sections, we describe a new way of updating subjective beliefs based on Jeffrey's rule of conditioning [11, 18, 19, 20], which we call dynamic probability kinematics (DPK). Let Ω be the state space of interest, and assume it is at most countable. The version of DPK with uncountable Ω will be the subject of a future work. Suppose that P is a finitely additive probability measure on $(\Omega, \mathcal{F})$ representing an agent's initial beliefs around the elements of $\mathcal{F} = 2^\Omega$, and that we want to update it after collecting some data.[3] We work with finitely additive probabilities because, as we shall see in Section 6, when we introduce ambiguity into the picture they allow us to retrieve some interesting results from [32]. The agent observes data points $x_1, \dots, x_n$ that are realizations of a random quantity $X : \Omega \to \mathcal{X}$ whose distribution P_X is unknown.[4] Call F_X its cdf, and assume $\mathcal{X}$ is finite.[5] Notice that collecting $x_1, \dots, x_n$ is equivalent to observing $\omega_1, \dots, \omega_n \in \Omega$, and then computing $X(\omega_i) = x_i$. Consider now the collection $\mathcal{E}' := (E_i)_{i=1}^n$, where $E_i \equiv X^{-1}(x_i) := \{\omega \in \Omega : X(\omega) = x_i\}$. It induces partition $\mathcal{E} = \{E_j\}_{j=1}^{m+1}$ of Ω, $m \leq n$. Here, with a slight abuse of notation, we denote the unique elements of $\mathcal{E}'$ by $E_1, \dots, E_m$, and the complement of their union by $E_{m+1} = (\cup_{j=1}^m E_j)^c = \Omega \setminus \cup_{j=1}^m E_j$.

As an update to P, we propose $P_\mathcal{E} : \mathcal{F} \to [0,1]$, $A \mapsto P_\mathcal{E}(A) := \sum_{E_j \in \mathcal{E}} P(A \mid E_j) P_\mathcal{E}(E_j)$, such that $P_\mathcal{E}(E_j) \geq 0$, for all $E_j \in \mathcal{E}$, and $\sum_{E_j \in \mathcal{E}} P_\mathcal{E}(E_j) = 1$. We have the following.

Proposition 1 *$P_\mathcal{E}$ is a finitely additive probability measure.*[6]

In general, Jeffrey's rule of conditioning – as presented in [11, Equation 1.1] – is given by $P^\star(A) = \sum_j P(A \mid E_j) P^\star(E_j)$, where $P^\star$ is Jeffrey's posterior for P. It is valid when Jeffrey's condition is met, that is, when there is a given partition $\{E_j\}$ of the state space Ω such that

[3] We assume $\mathcal{F} = 2^\Omega$ to work with the richest possible sigma-algebra; all the results in this paper still hold if $\mathcal{F}$ is not the power set. Ω is assumed at most countable for simplicity: we want to focus on the updating mechanism and not on measure-theoretic complications.

[4] P_X is assumed to be a finitely additive probability measure on $\mathcal{X}$.

[5] This assumption is needed to ensure that partition $\tilde{\mathcal{E}}$ in Proposition 3 is finite.

[6] This proposition is true also in the countably additive case. That is, if P is countably additive, then $P_\mathcal{E}$ is also countably additive.

$P(A \mid E_j) = P^\star(A \mid E_j)$ is true for all $A \in \mathcal{F}$ and all j. Specifically, this condition is met by $P_\mathcal{E}$. Since $P_\mathcal{E}$ is a finitely additive probability measure by Proposition 1, it is true that, for all $A \in \mathcal{F}$, $P_\mathcal{E}(A) = \sum_{E_j \in \mathcal{E}} P_\mathcal{E}(A \mid E_j)P_\mathcal{E}(E_j)$. But given our definition for $P_\mathcal{E}$, we also have that $P_\mathcal{E}(A) = \sum_{E_j \in \mathcal{E}} P(A \mid E_j)P_\mathcal{E}(E_j)$. This implies that there is a partition $\mathcal{E}$ for which $P(A \mid E_j) = P_\mathcal{E}(A \mid E_j)$ is true for all $A \in \mathcal{F}$ and all $E_j \in \mathcal{E}$.

4. An Empirical Specification of $P_\mathcal{E}$

In this Section, we show how to compute DPK updating for $P_\mathcal{E}(A)$ via an empirically specified sequence of partitions, which in turn determines a sequence of empirical probability measures. Utilizing it eases the analyst of the burden of making a full subjective probabilistic assessment for the elements of $\mathcal{E}$.

Recall that $\mathcal{E}' = (E_i)_{i=1}^n = \{X^{-1}(x_i)\}_{i=1}^n$, and $\mathcal{E} = \{E_j\}_{j=1}^{m+1}$, where $E_1, \dots, E_m$ are the unique elements of $\mathcal{E}'$, and $E_{m+1} = (\cup_{j=1}^m E_j)^c$. Denote by $\Delta(\Omega, \mathcal{F})$ the set of all finitely additive probability measures on $(\Omega, \mathcal{F})$. Then, consider the empirical probability measure $P^{emp} \in \Delta(\Omega, \mathcal{F})$ such that, if $E_{m+1} \neq \emptyset$,

$$P^{emp}(E_j) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{I}(E_j = E_i), \tag{1}$$

for all $j \in \{1, \dots, m\}$, where $\mathbb{I}$ denotes the indicator function, and

$$P^{emp}(E_{m+1}) = 1 - \sum_{j=1}^m P^{emp}(E_j). \tag{2}$$

If instead $E_{m+1} = \emptyset$, $P^{emp}(E_j) = 1/n \sum_{i=1}^n \mathbb{I}(E_j = E_i)$, for all $j \in \{1, \dots, m\}$, and $P^{emp}(E_{m+1}) = 0$. We require that, for all $E_j \in \mathcal{E}$,

$$P_\mathcal{E}(E_j) = \beta(n)P(E_j) + [1 - \beta(n)]\, P^{emp}(E_j), \tag{3}$$

where $\beta(n)$ is a coefficient in $[0, 1]$ depending on n: the posterior probability $P_\mathcal{E}$ assigned to the elements E_j of partition $\mathcal{E}$ is a weighted average of the prior P and the empirical probability measure P^{emp}.[7] Performing the update in Section 3 then becomes a mechanical procedure. The coefficient $\beta(n)$ is specified by the agent and controls the extent of prior-data trade-off in the updated belief. The closer $\beta(n)$ is to 1, the "stickier" DPK is; that is, the less the collected observations influence the agent's (revised) beliefs, and vice versa the closer $\beta(n)$ is to 0. The facts that DPK is mechanical and that its stickiness is regulated by a parameter that is entirely under the agent's control makes our updating procedure mathematically and conceptually appealing.

In the remainder of this paper, we use the above procedure to assign updated probabilities to the elements of $\mathcal{E}$. We do so for two main reasons. The first one is that it makes studying the asymptotic behavior of subsequent updates of $P_\mathcal{E}$ relatively easy (see Theorem 4). The second one is that controlling the stickiness of the update via parameter $\beta(n)$ is desirable in many examples, and also from a computational point of view. An example of how to update subjective beliefs according to DPK is given in Section 7.1.[8]

5. Subsequent Updates

Let us denote the amount of data available at time $t = 1$ by n_1. Once at time $t = 2$ we observe new data points $x_{n_1+1}, \dots, x_{n_2}$, and update $P_\mathcal{E} \equiv P_{\mathcal{E}_1}$ to $P_{\mathcal{E}_1\mathcal{E}_2}$ via the empirical procedure presented in Section 4. That is, consider partition $\mathcal{E}_2 = \{E_j\}_{j=1}^{k+1}$ where $E_1, \dots, E_k$ are the unique elements in the collection $\mathcal{E}'' = (E_i)_{i=1}^{n_2} = \{X^{-1}(x_i)\}_{i=1}^{n_2}$, and $E_{k+1} = (\cup_{j=1}^k E_k)^c$. We equate $P_{\mathcal{E}_1\mathcal{E}_2}(E_j) = \beta(n_2)P_{\mathcal{E}_1}(E_j) + [1-\beta(n_2)]P_2^{emp}(E_j)$, for all $E_j \in \mathcal{E}_2$, where the $P_2^{emp}(E_j)$'s are computed similarly to Section 4, so we have $P_{\mathcal{E}_1\mathcal{E}_2}(A) = \sum_{E_j \in \mathcal{E}_2} P_{\mathcal{E}_1}(A \mid E_j)P_{\mathcal{E}_1\mathcal{E}_2}(E_j)$. Clearly, Proposition 1 is true also for $P_{\mathcal{E}_1\mathcal{E}_2}$.

Call $(P_{\mathcal{E}_1 \cdots \mathcal{E}_t})$, $t \in \mathbb{N}$, the sequence of successive updates of probability measure P representing the initial subjective beliefs of the agent around the elements of Ω, and $\mathbf{x}_t = (x_i)_{i=1}^{n_t}$ the collection of data points available at time t. Notice that $\#\mathcal{E}_t = \#\text{unique}(\mathbf{x}_t) + 1$, where # denotes the cardinality operator. That is, the number of elements of partition $\mathcal{E}_t$ is equal to the number of unique observations x_i collected up to time t plus 1, the complementary of the union of the other elements of $\mathcal{E}_t$. For convenience, from here on we write $P_{\mathcal{E}_t}$ in place of $P_{\mathcal{E}_1 \cdots \mathcal{E}_t}$ for all $t \in \mathbb{N}$.

Remark 2 *Notice that $n_t > n_{t-1}$ for all $t \in \mathbb{N}$, and $n_0 = 0$. That is, the amount of data points available at time t is always larger than that at time $t-1$; this implies that as $t \to \infty$, $n_t \to \infty$. In addition, we have that $P_{\mathcal{E}_t}$ depends on $n_1, \dots, n_t$ and $P_{\mathcal{E}_0}$; we denote this by $P_{\mathcal{E}_t} \equiv P_{\mathcal{E}_t}(n_1, \dots, n_t, P_0)$.*[9]

A consequence of how we build partitions is that, for any t, $\mathcal{E}_t$ is not coarser than $\mathcal{E}_{t-1}$. To see this, suppose $\mathcal{E}_{t-1}$ has $\ell+1$ many elements, that is, $\mathcal{E}_{t-1} = \{E_1^{\mathcal{E}_{t-1}}, \dots, E_\ell^{\mathcal{E}_{t-1}}, E_{\ell+1}^{\mathcal{E}_{t-1}}\}$. As we know, this means that $E_{\ell+1}^{\mathcal{E}_{t-1}} = (\cup_{j=1}^\ell E_j^{\mathcal{E}_{t-1}})^c$. Now suppose that in the next updating step we only observe one element x. If it is not a "novelty", then $\mathcal{E}_t = \mathcal{E}_{t-1}$. If

[7] In [7, Remark 9], the authors discuss an appealing choice of $\beta(n)$.

[8] A subtlety in moving from P to $P_\mathcal{E}$ is discussed in [7, Remark 3].

[9] To illustrate this, in [7, Remark 4] the authors write $P_{\mathcal{E}_2}$ in terms of n_1, n_2, and P_0; it is then easy to see how that can be generalized to any $t > 2$.

instead x is a new element, we have that $\mathcal{E}_t$ has $\ell+2$ many elements. In particular, $E_j^{\mathcal{E}_{t-1}} = E_j^{\mathcal{E}_t}$, for all $j \in \{1,\dots,\ell\}$, and $E_{\ell+1}^{\mathcal{E}_{t-1}} = E_{\ell+1}^{\mathcal{E}_t} \cup E_{\ell+2}^{\mathcal{E}_t}$. Of course, if we observe more elements, we further refine $E_{\ell+1}^{\mathcal{E}_{t-1}}$.

Proposition 3 *There exists a partition $\tilde{\mathcal{E}}$ that cannot be refined as a result of the updating process described in Sections 3 and 4.*

We now show how, under mild assumptions, the sequence of successive subjective beliefs updated according to the DPK procedure converges. Call Q the "objective" (finitely additive) probability measure on $(\Omega,\mathcal{F})$. In the language of Walley, it is the aleatory probability associated with our experiment [32, Sections 1.3.2, 2.11.2]. That is, Q is linked to a physical property of the analysis that does not depend on the observer. Formally, we have that for all $x \in \mathcal{X}$, $Q(X^{-1}(x)) = P_X(\{x\})$. On the other hand, an epistemic probability models logical or psychological degrees of belief of the agent. It is immediate to see, then, how the probability P specified by the agent at the beginning of the analysis is an epistemic probability. Recall that F_X denotes the cdf of P_X, and that the total variation distance d_{TV} is defined as $d_{TV}(\pi,\gamma) := \sup_{A\in\mathcal{F}} |\pi(A)-\gamma(A)|$, for all $\pi,\gamma \in \Delta(\Omega,\mathcal{F})$.

Theorem 4 *Suppose the following conditions hold,*

1. $\lim_{n_t\to\infty} \beta(n_t) = 0$;
2. *for every discontinuity point x of F_X, $F_X(x^+)-F(x^-) = P_X(\{x\})$;*
3. $F_X(-\infty) = 1 - F_X(\infty) = 0$.

Then, $P_{\mathcal{E}_t}$ converges to $\sum_{E\in\tilde{\mathcal{E}}} P_{\mathcal{E}_{t-1}}(\cdot \mid E)Q(E)$ almost surely as $n_t \to \infty$ in the total variation distance. [10]

Because as n_t grows to infinity the partition induced by collection $\{X^{-1}(x_i)\}_{i=1}^{n_t}$ approaches $\tilde{\mathcal{E}}$, we denote by $P_{\tilde{\mathcal{E}}}$ the limit we find in Theorem 4.

We tacitly assumed that for all nonempty $A \in \mathcal{F}$, the probability assigned to A by P (representing the agent's initial beliefs) is positive. In formulas, $P(A) > 0$, for all $\emptyset \neq A \in \mathcal{F}$. This assumption is not too stringent. For example, suppose the agent specifies P so that there is a collection of sets $\{A'_k\} \subset \mathcal{F}$ such that $A'_k \neq \emptyset$ and $P(A'_k) = 0$, for all k. Then, the agent should choose $\tilde{P} = (1-\epsilon)P + \epsilon R$ as probability encapsulating their initial beliefs, where R is a (finitely additive) probability measure belonging to the set $\mathcal{R} := \{R \in \Delta(\Omega,\mathcal{F}) : R(\{\omega\}) > 0, \forall \omega \in \Omega\}$, and ϵ is an arbitrarily small element of $(0,1)$. This procedure – a particular case of ϵ-contamination [4, 5, 15, 16] or *padding* [1] – keeps the initial beliefs essentially unaltered, and avoids complications coming from conditioning on zero probability events. In the future, we plan to deal with the delicate matter of conditioning on zero probability events in a more sophisticated way, possibly using techniques from the literature on lexicographic probabilities [6] or layers of zero probabilities [10].

Dynamic probability kinematics is not commutative. An in-depth study of this issue is given in [7, Remark 8]. Should the lack of commutativity worry the agent that intends to update their beliefs using DPK? The answer is no. Since successive partitions are induced by an increasing amount of collected data points, commutativity would mean that losing data yields no loss of information on the likelihood of the event $A \subset \Omega$ of interest. This is undesirable: the more we know about the composition of Ω, the better we want our assessment to be on the plausibility of event A. As Diaconis and Zabell point out in [11, Section 4.2, Remark 2], "noncommutativity is not a real problem for successive Jeffrey updating"; it is not a real problem for DPK either.

Despite DPK is not in general commutative, the limit probability $P_{\tilde{\mathcal{E}}}$ is the same regardless of the order in which data is collected. Suppose we collect observations in a different order in two different procedures. Call $(\mathcal{E}_t)$ and $(\mathcal{E}'_t)$ the sequences of successive partitions in the first and second procedures, respectively, and $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{E}}'$ the limit partitions for the first and second procedures, respectively.

Proposition 5 *Suppose that conditions 1-3 of Theorem 4 hold. Call $P_{\tilde{\mathcal{E}}}$ the almost sure limit of $(P_{\mathcal{E}_t})$ and $P_{\tilde{\mathcal{E}}'}$ the almost sure limit of $(P_{\mathcal{E}'_t})$ in the total variation metric as n_t goes to infinity. Then, $P_{\tilde{\mathcal{E}}} = P_{\tilde{\mathcal{E}}'}$.*

In PK, an agent's subjective probabilities over a fixed partition $\mathcal{E}$ undergo an exogenous change (a *Jeffrey shift*), which is then propagated across the rest of their probabilities in a natural manner. Crucially, PK does not specify what Jeffrey shift an agent's probabilities will undergo; it treats the choice of the Jeffrey shift as an input to the rule rather than part of the rule itself. Indeed, in the original interpretation of PK, the shift is usually a non-inferential change to the agent's degrees of belief that is not chosen consciously or freely, but rather e.g. the brute result of a perceptual process. Training the complicated network of neurons in our skull to translate perceptive and proprioceptive inputs into sensible Jeffrey shifts is something that Jeffrey did not believe should be formalized.

DPK is an updating technique that sits in between Bayes' and Jeffrey's rules. It can be seen, heuristically, as a map from specifications of statistical problems to choices of Jeffrey shift (which are then propagated in the usual way, via PK). While it is built as a particular case of PK, it uses the empirical distribution to assign probabilities to the elements of the updated partition $\mathcal{E}_t$. In order to mechanize

[10] In the countably additive case, it is enough to assume 1 to obtain the same result. This is also true for Propositions 5 and 6.

the procedure, it gives up the freedom of choosing whatever probability the agent feels correct to assign to the elements of $\mathcal{E}_t$.[11]

Notice for example that DPK can give significantly different answers about the appropriate Jeffrey shift to use on a given occasion than a trained medical expert. Though it may give different (potentially worse) answers than subject matter experts, it nevertheless enjoys the properties that we present in the next paragraph. The choice of adopting DPK instead of PK to update one's beliefs is based on such a trade-off.

If evidence is collected that does not belong to Ω, that is, if $\mathcal{X} \not\subset \Omega$, then using the inverse image X^{-1} of function X, DPK allows one to update their beliefs without first needing to enlarge Ω to $\Omega' = \Omega \times \mathcal{X}$. Notice also that, being a particular case of PK, DPK updates can be obtained by Bayesian updating in a larger space Ω' [11, Theorem 2.1].[12] There are two main reasons for not wanting to enlarge the state space: (i) reassessing our beliefs on a larger space requires us to extend our beliefs from the elements of 2^Ω to those of $2^{\Omega'}$ [12, 21];[13] (ii) updating probabilities on a larger sigma-algebra can be computationally costly. Besides simplifying the updating procedure by not requiring an enlarged state space, we also conjecture that DPK simplifies the treatment of nuisance parameters, a statement that will be verified in future work.

6. Working with Sets of Probabilities

In this Section, we generalize dynamic probability kinematics to dynamic imprecise probability kinematics (DIPK). To do so, we first introduce the concepts of lower probability, upper probability, and the core.

6.1. Concepts

Consider a generic set of probabilities Π on a measurable space $(\Omega, \mathcal{F})$. The *lower probability* of A associated with Π is defined as $\underline{P}(A) := \inf_{P \in \Pi} P(A)$, for all $A \in \mathcal{F}$. The *upper probability* of A associated with Π is defined as the conjugate to $\underline{P}(A)$, that is, $\overline{P}(A) := 1 - \underline{P}(A^c) = \sup_{P' \in \Pi} P'(A)$, for all $A \in \mathcal{F}$. Recall that $\Delta(\Omega, \mathcal{F})$ denotes the set of all finitely additive probability measures on $(\Omega, \mathcal{F})$. $\underline{P}$ completely characterizes its *core*[14]

$$\begin{aligned}\text{core}(\underline{P}) &= \{P \in \Delta(\Omega, \mathcal{F}) : P(A) \geq \underline{P}(A), \forall A \in \mathcal{F}\} \\ &= \{P \in \Delta(\Omega, \mathcal{F}) : \overline{P}(A) \geq P(A) \geq \underline{P}(A), \forall A \in \mathcal{F}\},\end{aligned}$$

where the second equality is a characterization. The core of $\underline{P}$ is the set of all probability measures on $(\Omega, \mathcal{F})$ that setwise dominate $\underline{P}$. Notice that the core is convex and weak*-compact [32, Theorem 3.6.2].[15]

To generalize DPK to DIPK, we first prescribe the agent to specify a set of probabilities $\mathcal{P}$, then to compute the lower probability associated with it.[16] The core of such lower probability represents the agent's initial beliefs. To update their beliefs, the agent computes the DPK update of the extrema of the core, that is, of the elements of the core that cannot be written as a convex combination of other elements. Their updated beliefs are represented by the convex hull of the updated extrema, which coincides with the core of the updated lower probability by [32, Theorem 3.6.2].

We assume that the number of extrema of $\mathcal{P}^{\text{co}}_{\mathcal{E}_0}$ is finite; we do so mainly for computational reasons [7, Remark 14].

We require the agent's beliefs to be represented by the core for two main reasons. The first, mathematical, one is to ensure that the belief set can be completely characterized by the lower probability, and that lower probability $\underline{P}$ is coherent [32, Section 3.3.3]. The second, philosophical, one is presented next.

At the beginning of the study, the *sensitivity analysis* approach to imprecise probabilities prescribes the agent to specify a set of possible (or plausible) candidates for the true or ideal probability measure P_T governing the events of interest [3]. As [32, Section 5.9] points out, this way of proceeding assumes the *axiom of ideal precision*: there exists a true probability measure P_T governing the random events, but it cannot be precisely known e.g. because we would need an infinitely long reflection to elicit it.

The philosophical motivation for the agent's beliefs being represented by the core of $\underline{P}$ is the following. A criticism brought forward by Walley in [32, Section 2.10.4.(c)] is that, given a lower probability $\underline{P}$, there is no cogent reason for which the agent should choose a specific P_T that dominates $\underline{P}$, or, for that matter, a collection of "plausible" probabilities. Because the core considers all finitely additive probability measures that dominate $\underline{P}$, it is the perfect instrument

[11] Such mechanical procedure is domain-invariant, that is, the same rule is used e.g. for medical data (see Section 7.1) and for soccer matches (see Section 7.2).

[12] In [11], the authors show that there exists a "duality" between Bayes' rule (BR) and PK. BR can be seen as a special case of PK, as we pointed out in Section 1.2, while at the same time we can obtain PK from BR if we enlarge the state space.

[13] For the imprecise version of DPK, that is, for DIPK, we can extend the agent's beliefs via Walley's extension [32, Chapter 3].

[14] By complete characterization, we mean that it is sufficient to know $\underline{P}$ to be able to completely specify core($\underline{P}$). To emphasize this aspect, some authors say that $\underline{P}$ is *compatible* with core($\underline{P}$) [14].

[15] Recall that in the weak* topology, a net $(P_\alpha)_{\alpha \in I}$ converges to P if and only if $P_\alpha(A) \to P(A)$, for all $A \in \mathcal{F}$.

[16] We work with the core instead of the convex hull of $\mathcal{P}$ because the lower probability completely characterizes the core, but does not completely characterize the convex hull [7, Remark 15].

to reconcile Walley's behavioral and sensitivity analysis interpretations.[17]

6.2. DIPK for Sets of Probabilities

The analysis begins with specifying a set $\mathcal{P} \subset \Delta(\Omega,\mathcal{F})$ of finitely additive probability measures on Ω. Consider $\underline{P} \equiv \underline{P}_{\mathcal{E}_0}$, the lower probability associated with $\mathcal{P}$. The set representing the agent's initial beliefs is given by $\mathcal{P}^{\text{co}}_{\mathcal{E}_0} = \text{core}(\underline{P}_{\mathcal{E}_0})$, where superscript co denotes the fact that $\mathcal{P}^{\text{co}}_{\mathcal{E}_0}$ is convex and compact. We also need to consider the set $\mathcal{P}_{\mathcal{E}_0} = \text{ex}\mathcal{P}^{\text{co}}_{\mathcal{E}_0}$ of extrema of $\mathcal{P}^{\text{co}}_{\mathcal{E}_0}$.

We then compute the DPK update of every element in $\mathcal{P}_{\mathcal{E}_0}$, and we obtain

$$\mathcal{P}_{\mathcal{E}_1} := \Bigg\{ P_{\mathcal{E}_1} : P_{\mathcal{E}_1}(A) = \sum_{E_j \in \mathcal{E}_1} P_{\mathcal{E}_0}(A \mid E_j) P_{\mathcal{E}_1}(E_j), \quad \forall A \in \mathcal{F}, P_{\mathcal{E}_0} \in \mathcal{P}_{\mathcal{E}_0} \Bigg\}.$$

After that, we compute $\mathcal{P}^{\text{co}}_{\mathcal{E}_1} = \text{Conv}(\mathcal{P}_{\mathcal{E}_1}) = \text{core}(\underline{P}_{\mathcal{E}_1})$, where $\underline{P}_{\mathcal{E}_1}$ is the updated lower probability, $\text{Conv}(\cdot)$ denotes the convex hull, and the last equality holds by [32, Theorem 3.6.2].

Repeating this procedure, we build two sequences, $(\mathcal{P}_{\mathcal{E}_t})$ and $(\mathcal{P}^{\text{co}}_{\mathcal{E}_t})$. Notice that for any $t \in \mathbb{N}$, the lower and upper probabilities associated with $\mathcal{P}_{\mathcal{E}_t}$ are equal to the lower and upper probabilities associated with $\mathcal{P}^{\text{co}}_{\mathcal{E}_t}$, respectively. An example of how to update subjective beliefs according to DIPK is given in Section 7.2.

Recall that d_{TV} denotes the total variation distance $d_{TV}(\pi,\gamma) := \sup_{A\in\mathcal{F}} |\pi(A) - \gamma(A)|$, for all $\pi,\gamma \in \Delta(\Omega,\mathcal{F})$. Suppose that assumptions 1-3 of Theorem 4 hold. Call $\mathcal{P}_{\tilde{\mathcal{E}}} := \{P_{\tilde{\mathcal{E}}} \in \Delta(\Omega,\mathcal{F}) : d_{TV}(P_{\mathcal{E}_t}, P_{\tilde{\mathcal{E}}}) \xrightarrow[n_t\to\infty]{a.s.} 0, P_{\mathcal{E}_t} \in \mathcal{P}_{\mathcal{E}_t}\}$. That is, $\mathcal{P}_{\tilde{\mathcal{E}}}$ is the set of limits (as n_t goes to infinity with probability 1 in the total variation metric) of the elements $P_{\mathcal{E}_t}$ of set $\mathcal{P}_{\mathcal{E}_t}$ representing the (extrema of the) agent's updated beliefs. We are sure $\mathcal{P}_{\tilde{\mathcal{E}}}$ is not empty by Proposition 3 and Theorem 4. Then, by construction, we have that $d_H(\mathcal{P}_{\mathcal{E}_t}, \mathcal{P}_{\tilde{\mathcal{E}}}) \to 0$ as n_t goes to infinity with probability one, where d_H denotes the Hausdorff metric $\max(\sup_{P\in\mathcal{P}_{\mathcal{E}_t}} d_{TV}(P, \mathcal{P}_{\tilde{\mathcal{E}}}), \sup_{P'\in\mathcal{P}_{\tilde{\mathcal{E}}}} d_{TV}(\mathcal{P}_{\mathcal{E}_t}, P'))$ and, in general, $d_{TV}(\pi,\Gamma) := \inf_{\gamma\in\Gamma} d_{TV}(\pi,\gamma)$, for all $\pi \in \Delta(\Omega,\mathcal{F})$ and all $\Gamma \subset \Delta(\Omega,\mathcal{F})$. Such a convergence is true also for $\mathcal{P}^{\text{co}}_{\mathcal{E}_t}$ and $\mathcal{P}^{\text{co}}_{\tilde{\mathcal{E}}}$, as shown in the next proposition.

Proposition 6 *If assumptions 1-3 of Theorem 4 hold, then $d_H(\mathcal{P}^{co}_{\mathcal{E}_t}, \mathcal{P}^{co}_{\tilde{\mathcal{E}}}) \to 0$ as n_t goes to infinity with probability one.*

[17] In the imprecise probabilities literature, agents are often required to specify coherent lower (and upper) probabilities [32, Section 2.5]. In [32, Section 3.3.3] the author shows that $\underline{P}$ is coherent if and only if it can be written as the infimum of a set $\mathcal{P}$ of finitely additive probability measures.

Let us discuss the importance of $\mathcal{P}^{\text{co}}_{\mathcal{E}_t}$ being convex and compact. Consider a generic set of probabilities Π on a measurable space $(\Omega,\mathcal{F})$. Suppose Π is finite, i.e. $\Pi = \{\pi_j\}_{j=1}^k$, for some $k \in \mathbb{N}$. Then, the lower probability associated with Π is equivalent to the one associated with its convex hull $\text{Conv}(\Pi)$. If instead Π is convex but open, then the lower probability associated with Π is equivalent to the one associated with its closure $\text{Cl}(\Pi)$. To this extent, lower probabilities are not able to detect "holes and dents" in their associated set of probabilities. This is why we need the sequence of convex and (weak*-)compact sets $(\mathcal{P}^{\text{co}}_{\mathcal{E}_t})$ to represent the agent's belief updating procedure.

Notice that assuming $P(A) > 0$, for all $\emptyset \neq A \in \mathcal{F}$, implies a *near-ignorance* assumption in the DIPK update. This means that every element in $\mathcal{P}_{\mathcal{E}_0} = \text{ex}\mathcal{P}^{\text{co}}_{\mathcal{E}_0}$ gives positive probability to all nonempty $A \in \mathcal{F}$. This is desirable because no finite sample is enough to annihilate a sufficiently extreme prior belief [7, Remark 17].

7. Examples of DPK and DIPK Updating

In this Section, we present two simple examples on how to update subjective beliefs according to DPK and DIPK procedures.

7.1. Trials of a New Surgical Procedure

We continue Example 1, and show how to frame it within the DPK paradigm. Recall that we wish to form a probabilistic opinion of a new surgical procedure to be performed three times at a new hospital. Upon one colleague's suggestion that another hospital performed this type of procedure with a success rate of 0.8, we update by considering random variable $X : \Omega \to \mathcal{X} = \{0,1,2,3\}$ whose distribution is unknown and such that $X(\omega)$ represents the number of 1's in ω. As we can see, $X^{-1}(3) = \{111\}$, $X^{-1}(2) = \{011, 101, 110\}$, $X^{-1}(1) = \{001, 010, 100\}$, $X^{-1}(0) = \{000\}$. The finest partition of Ω according to DPK, then, is given by $\tilde{\mathcal{E}} = \{E_0, E_1, E_2, E_3, E_4\}$, where $E_j = X^{-1}(j)$, $j \in \{0,1,2,3\}$, and $E_4 = \emptyset$. Recall that in DPK data points contribute information not through their sheer number, but rather the way they partition the space and assign relative frequencies. The information that our colleague provided us is equivalent to observing 1000 data points $x_1, \dots, x_{1000}$, out of which 512 are all 3's, 384 are all 2's, 96 are all 1's, and 8 are all 0's. This because the relative frequency Fr of the elements of $\mathcal{X}$ is $Fr(\{3\}) = 512/1000 = 1 \cdot 0.8^3$, $Fr(\{2\}) = 384/1000 = 3 \cdot 0.2 \cdot 0.8^2$, $Fr(\{1\}) = 96/1000 = 3 \cdot 0.2^2 \cdot 0.8$, and $Fr(\{0\}) = 8/1000 = 1 \cdot 0.2^3$. But why should they be derived in this way? We have that $Fr(\{3\}) = 1 \cdot 0.8^3$ because there is only one way of obtaining three successes, each of which has probability 0.8 in the procedures conducted at the hospital that our colleague informed us about. Instead, $Fr(\{2\}) = 3 \cdot 0.2 \cdot 0.8^2$ because there are three ways of obtaining two successes and one failure, where the probability of the latter is 0.2 according to our colleague. Finally, $Fr(\{1\}) = 3 \cdot 0.2^2 \cdot 0.8$ because there are three ways of obtaining one successes and two failures, and $Fr(\{0\}) = 1 \cdot 0.2^3$ because there is only one way of obtaining three failures.

Relative frequency Fr implies that $P_1^{emp}(E_0) = 0.008$, $P_1^{emp}(E_1) = 0.096$, $P_1^{emp}(E_2) = 0.384$, $P_1^{emp}(E_3) = 0.512$, and $P_1^{emp}(E_4) = 0$. This corresponds to collecting the following probabilistic evidence: three failures with probability 0.008, only one success with probability 0.096, two successes with probability 0.384, and three successes with probability 0.512. We are now ready to compute the DPK update of our initial P. Given the composition of the sample space $\mathcal{X}$, we have that $P(\{000\}) = p_0$, $P(\{001\}) = P(\{100\}) = P(\{010\}) = p_1$, $P(\{110\}) = P(\{101\}) = P(\{011\}) = p_2$, $P(\{111\}) = p_3$. Suppose $\beta(n_t) = 1/n_t$; in turn we have

$$\begin{aligned}
P_{\mathcal{E}_1}(\{000\}) &= \frac{p_0}{P_{\mathcal{E}_0}(E_0)}\left(\frac{p_0}{1000} + \frac{999}{1000}P_1^{emp}(E_0)\right)\\
&= 1 \cdot \left(\frac{p_0}{1000} + \frac{7.992}{1000}\right) = \frac{p_0 + 7.992}{1000},\\
P_{\mathcal{E}_1}(\{001\}) &= \frac{p_1}{P_{\mathcal{E}_0}(E_1)}\left(\frac{p_1}{1000} + \frac{999}{1000}P_1^{emp}(E_1)\right)\\
&= \frac{1}{3} \cdot \left(\frac{p_1}{1000} + \frac{95.904}{1000}\right) = \frac{p_1 + 95.904}{3000},\\
P_{\mathcal{E}_1}(\{011\}) &= \frac{p_2}{P_{\mathcal{E}_0}(E_2)}\left(\frac{p_2}{1000} + \frac{999}{1000}P_1^{emp}(E_2)\right)\\
&= \frac{1}{3} \cdot \left(\frac{p_2}{1000} + \frac{383.616}{1000}\right) = \frac{p_2 + 383.616}{3000},\\
P_{\mathcal{E}_1}(\{111\}) &= \frac{p_3}{P_{\mathcal{E}_0}(E_3)}\left(\frac{p_3}{1000} + \frac{999}{1000}P_1^{emp}(E_3)\right)\\
&= 1 \cdot \left(\frac{p_3}{1000} + \frac{511.488}{1000}\right) = \frac{p_3 + 511.488}{1000},
\end{aligned}$$

and $P_{\mathcal{E}_1}(\{001\}) = P_{\mathcal{E}_1}(\{010\}) = P_{\mathcal{E}_1}(\{100\})$, $P_{\mathcal{E}_1}(\{011\}) = P_{\mathcal{E}_1}(\{110\}) = P_{\mathcal{E}_1}(\{101\})$. We can see how, because of the composition of sample space $\mathcal{X}$, in the case of only one successful outcome the updated probability $P_{\mathcal{E}_1}$ assigned to $\{001\}$, $\{010\}$, and $\{100\}$ is exactly 1/3 of the mixture between the prior and the empirical probability of E_1. The same is true for the case of two successful outcomes.

To generalize the DPK updating presented here to a DIPK updating involving a set $\mathcal{P}$ of probability measures representing the initial beliefs of the agent one can follow the procedure explained in Section 7.2.

7.2. Soccer Match Results

This example is built on [32, Section 4.6.1]. Let $\Omega = \{W, D, L\}$ represent the result of soccer match Juventus Turin vs Inter Milan, where W denotes a win for Juventus Turin, D a draw, and L a loss for Juventus Turin. Let then $X : \Omega \to \mathcal{X} = \{0, 1\}$, where 1 denotes a useful result (a victory or a draw) and 0 denotes a defeat, so X can be thought of as a Bernoulli random variable with unknown parameter. It is immediate to see how the finest partition of Ω according to DPK is given by $\tilde{\mathcal{E}} = \{E_1, E_2, E_3\}$, where $E_1 = \{W, D\}$, $E_2 = \{L\}$, and $E_3 = \emptyset$. We call $P_{\mathcal{E}_t}$ the t-th update of $P \equiv P_{\mathcal{E}_0}$; $P_{\tilde{\mathcal{E}}}$ denotes the limit of sequence $(P_{\mathcal{E}_t})$.[18]

The data points $x_1, \ldots, x_n$ that we collect represent the outcomes of past matches. Because the two teams are well established and high-level, it is reasonable to assume that function X is fixed.

Let us describe how to perform a DIPK update of subjective beliefs in this context. Let the agent specify $\mathcal{P} \subset \Delta(\Omega, \mathcal{F})$, and suppose that the lower and upper probabilities $\underline{P} \equiv \underline{P}_{\mathcal{E}_0}$ and $\overline{P} \equiv \overline{P}_{\mathcal{E}_0}$ associated with $\mathcal{P}$ are such that $\underline{P}(W) = \underline{P}(D) = 0.27$, $\overline{P}(W) = \overline{P}(D) = 0.52$, $\underline{P}(L) = 0.21$, and $\overline{P}(L) = 0.31$.[19]

A simplex representation is given in Figure 1 where each assessment is represented by a line parallel to one side of the simplex.[20] The initial beliefs of the agent are encapsulated in $\mathcal{P}^{\text{co}}_{\mathcal{E}_0} = \text{core}(\underline{P})$. To update $\mathcal{P}^{\text{co}}_{\mathcal{E}_0}$ we need to find $\mathcal{P}_{\mathcal{E}_0} = \text{ex}\mathcal{P}^{\text{co}}_{\mathcal{E}_0}$. This is an easy job; it is sufficient to 1. Equate $P(\omega)$ to either $\underline{P}(\omega)$ or $\overline{P}(\omega)$ for two of the three events. The probability of the third is then determined; 2. Check which of the resulting P satisfies $\underline{P} \leq P \leq \overline{P}$. This procedure gives us four extreme points $\mathcal{P}_{\mathcal{E}_0} = \{P^{ex}_{1,\mathcal{E}_0}, P^{ex}_{2,\mathcal{E}_0}, P^{ex}_{3,\mathcal{E}_0}, P^{ex}_{4,\mathcal{E}_0}\}$ such that $(P^{ex}_{1,\mathcal{E}_0}(W), P^{ex}_{1,\mathcal{E}_0}(D), P^{ex}_{1,\mathcal{E}_0}(L)) = (0.52, 0.27, 0.21)$, $(P^{ex}_{2,\mathcal{E}_0}(W), P^{ex}_{2,\mathcal{E}_0}(D), P^{ex}_{2,\mathcal{E}_0}(L)) = (0.27, 0.42, 0.31)$, $(P^{ex}_{3,\mathcal{E}_0}(W), P^{ex}_{3,\mathcal{E}_0}(D), P^{ex}_{3,\mathcal{E}_0}(L)) = (0.42, 0.27, 0.31)$, and $(P^{ex}_{4,\mathcal{E}_0}(W), P^{ex}_{4,\mathcal{E}_0}(D), P^{ex}_{4,\mathcal{E}_0}(L)) = (0.27, 0.52, 0.21)$. The extrema $\mathcal{P}_{\mathcal{E}_0}$ of $\mathcal{P}^{\text{co}}_{\mathcal{E}_0}$ are the vertices of the grey trapezoid in Figure 1.

As of January 12, 2022, there have been 257 matches between the two teams, with 178 useful results for Juventus Turin and 79 wins for Inter Milan.[21] This is to say that we observe $x_1, \ldots, x_{257}$ such that 178 are 1's, and 79 are 0's. Then, to compute $\mathcal{P}^{\text{co}}_{\mathcal{E}_1}$ it is enough to update the extrema in $\mathcal{P}_{\mathcal{E}_0}$ so to obtain $\mathcal{P}_{\mathcal{E}_1}$, and then consider the convex hull of the latter. The partition induced by the collected data is $\mathcal{E}_1 = \{E_1, E_2, E_3\}$, and we have that $P_1^{emp}(E_1) = 178/257$,

[18] Notice that $\tilde{\mathcal{E}}$ is attained almost immediately: it is enough to observe $x_j \neq x_k$, for some $j \neq k$.

[19] We write $\underline{P}(\omega)$ in place of $\underline{P}(\{\omega\})$ and $\overline{P}(\omega)$ in place of $\overline{P}(\{\omega\})$, $\omega \in \{W, D, L\}$, for notational convenience.

[20] Notice that the higher the values assigned by P to $\{\omega\} \subset \Omega$, the closer the line representing $P(\{\omega\})$ is to vertex $\omega \in \{W, D, L\}$.

[21] Data available here.

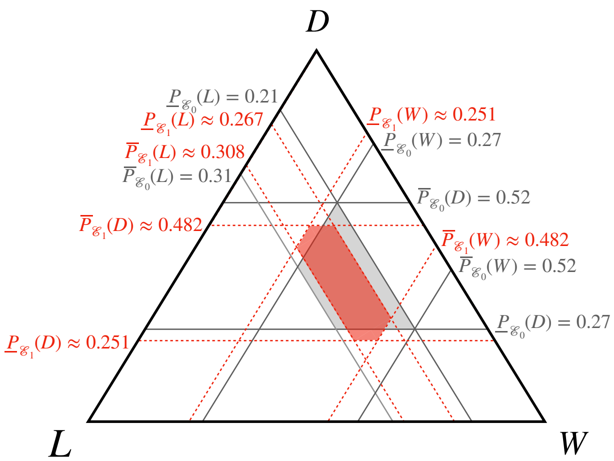

Figure 1: Visual representation of $\mathcal{P}^{\text{co}}_{\mathcal{E}_0}$ (the grey trapezoid) and of $\mathcal{P}^{\text{co}}_{\mathcal{E}_1}$ (the red hexagon) in our soccer example. $\underline{P}_{\mathcal{E}_0}$ is represented by the solid grey lines, while $\underline{P}_{\mathcal{E}_1}$ by the dashed red lines.

$P_1^{emp}(E_2) = 79/257$ and $P_1^{emp}(E_3) = 0$. This corresponds to collecting the following probabilistic evidence: Juventus Turin obtains a useful result with probability $178/257$, and it loses with probability $79/257$. Let us update $P^{ex}_{1,\mathcal{E}_0}$ to $P^{ex}_{1,\mathcal{E}_1}$. Suppose $\beta(n_t) = \frac{1}{\log(n_t+1)}$; we have

$$\begin{aligned}
P^{ex}_{1,\mathcal{E}_1}(W) &= \frac{P^{ex}_{1,\mathcal{E}_0}(W)}{P^{ex}_{1,\mathcal{E}_0}(E_1)} P^{ex}_{1,\mathcal{E}_1}(E_1) \\
&= \frac{0.52}{0.52+0.27}\left(\frac{0.52+0.27}{\log(258)} + \frac{\log(258)-1}{\log(258)}\cdot\frac{178}{257}\right) \\
&\approx 0.482, \\
P^{ex}_{1,\mathcal{E}_1}(D) &= \frac{P^{ex}_{1,\mathcal{E}_0}(D)}{P^{ex}_{1,\mathcal{E}_0}(E_1)} P^{ex}_{1,\mathcal{E}_1}(E_1) \\
&= \frac{0.27}{0.52+0.27}\left(\frac{0.52+0.27}{\log(258)} + \frac{\log(258)-1}{\log(258)}\cdot\frac{178}{257}\right) \\
&\approx 0.251, \\
P^{ex}_{1,\mathcal{E}_1}(L) &= \frac{P^{ex}_{1,\mathcal{E}_0}(L)}{P^{ex}_{1,\mathcal{E}_0}(E_2)} P^{ex}_{1,\mathcal{E}_1}(E_2) \\
&= 1\cdot\left(\frac{0.21}{\log(258)} + \frac{\log(258)-1}{\log(258)}\cdot\frac{79}{257}\right) \approx 0.267,
\end{aligned}$$

so $(P^{ex}_{1,\mathcal{E}_1}(W), P^{ex}_{1,\mathcal{E}_1}(D), P^{ex}_{1,\mathcal{E}_1}(L)) \approx (0.482, 0.251, 0.267)$. The other elements of $\mathcal{P}_{\mathcal{E}_0}$ are updated similarly. In particular, $(P^{ex}_{2,\mathcal{E}_1}(W), P^{ex}_{2,\mathcal{E}_1}(D), P^{ex}_{2,\mathcal{E}_1}(L)) \approx (0.271, 0.421, 0.308)$, $(P^{ex}_{3,\mathcal{E}_1}(W), P^{ex}_{3,\mathcal{E}_1}(D), P^{ex}_{3,\mathcal{E}_1}(L)) \approx (0.421, 0.271, 0.308)$, and $(P^{ex}_{4,\mathcal{E}_1}(W), P^{ex}_{4,\mathcal{E}_1}(D), P^{ex}_{4,\mathcal{E}_1}(L)) \approx (0.251, 0.482, 0.267)$. So we have that $\underline{P}_{\mathcal{E}_1}(W) \approx 0.251 \approx \underline{P}_{\mathcal{E}_1}(D)$, $\overline{P}_{\mathcal{E}_1}(W) \approx 0.482 \approx \overline{P}_{\mathcal{E}_1}(D)$, $\underline{P}_{\mathcal{E}_1}(L) \approx 0.267$, and $\overline{P}_{\mathcal{E}_1}(L) \approx 0.308$. As we can see from Figure 1, the graphical representation of $\mathcal{P}^{\text{co}}_{\mathcal{E}_1}$ is a hexagon (in red).

8. Conclusion

In this paper, we presented dynamic probability kinematics (DPK) and dynamic imprecise probability kinematics (DIPK). These methods dynamically update subjective beliefs stated in terms of precise and imprecise probabilities, in the presence of partial information (both DPK and DIPK) and of ambiguity (DIPK only). Two examples are provided to illustrate the procedures.

This work is just the first step towards a fully developed DIPK theory. In the future, we plan to relax the assumption that Ω needs to be at most countable and that $\mathcal{X}$ needs to be finite. Furthermore, we aim to generalize DIPK by allowing the agent to gather inconsistent evidence as in [26].

We also intend to let partial information be modeled via a set of probability distributions on $\mathcal{X}$, as empirical probabilities usually need a very large number of observations to estimate probabilities which are very close to zero or one to a good standard of relative accuracy. A similar idea is to use a set of empirical distributions to determine a set of partitions and Jeffrey shifts, and take the union of the posteriors recommended by DIPK.

After that, we plan to propose a way of performing statistical analysis based on DIPK updating. It is worth noting that lower probabilities are a special case of *lower previsions* [31, 32]. In the future, we will generalize DIPK to deal with these latter.

9. Proofs

Proof [Proof of Proposition 1] First, we have that $P_{\mathcal{E}}(A) \geq 0$, for all $A \in \mathcal{F}$. This comes by its definition, since it is defined as the summation of products of nonnegative quantities. Second, we have that $P_{\mathcal{E}}(\Omega) = 1$. This comes from the following

$$P_{\mathcal{E}}(\Omega) = \sum_{E_j \in \mathcal{E}} P(\Omega \mid E_j) P_{\mathcal{E}}(E_j) = \sum_{E_j \in \mathcal{E}} P_{\mathcal{E}}(E_j) = 1.$$

Finally, we have that if $\{A_i\}_{i\in I}$ is a finite, pairwise disjoint collection of events, then $P_{\mathcal{E}}(\cup_{i\in I} A_i) = \sum_{i\in I} P_{\mathcal{E}}(A_i)$. This because

$$\begin{aligned}
P_{\mathcal{E}}(\cup_{i\in I} A_i) &= \sum_{E_j \in \mathcal{E}} P\left(\cup_{i\in I} A_i \mid E_j\right) P_{\mathcal{E}}(E_j) \\
&= \sum_{E_j \in \mathcal{E}} \frac{P\left([\cup_{i\in I} A_i] \cap E_j\right)}{P(E_j)} P_{\mathcal{E}}(E_j)
\end{aligned}$$

$$= \sum_{E_j \in \mathcal{E}} \frac{P\left(\cup_{i \in I}\left[A_i \cap E_j\right]\right)}{P(E_j)} P_{\mathcal{E}}(E_j)$$
$$= \sum_{E_j \in \mathcal{E}} \frac{\sum_{i \in I} P\left(A_i \cap E_j\right)}{P(E_j)} P_{\mathcal{E}}(E_j)$$
$$= \sum_{i \in I} \sum_{E_j \in \mathcal{E}} \frac{P\left(A_i \cap E_j\right)}{P(E_j)} P_{\mathcal{E}}(E_j) = \sum_{i \in I} P_{\mathcal{E}}\left(A_i\right).$$

■

Proof [Proof of Proposition 3] Consider the limiting case where we have collected observations $(x_i)_{i \in \mathbb{N}}$. We have two cases. If $\cup_{i \in \mathbb{N}} x_i = \mathcal{X}$, then, since we observed all the elements of $\mathcal{X}$, and given the procedure in Sections 3 and 4 to refine the partition, it is immediate to see that the partition $\tilde{\mathcal{E}}$ induced by $(x_i)_{i \in \mathbb{N}}$ cannot be further refined. If instead $\cup_{i \in \mathbb{N}} x_i = \mathcal{X}_{reduced} \subsetneq \mathcal{X}$, then the elements of partition $\tilde{\mathcal{E}}$ will be the unique elements of the collection $\{X^{-1}(x_i)\}_{x_i \in \mathcal{X}_{reduced}}$, plus an extra one given by $(\cup_{x_i \in \mathcal{X}_{reduced}} X^{-1}(x_i))^c$. ■

Proof [Proof of Theorem 4] Let $t = 1$ and fix any $A \in \mathcal{F}$. Let $\#\mathcal{E}_1 = m + 1$ and assume without loss of generality that $E_{m+1} = \emptyset$. We have that $P_{\mathcal{E}_1}(A) = \sum_{E_j \in \mathcal{E}_1} P(A \mid E_j) P_{\mathcal{E}_1}(E_j)$, which is equal to

$$\sum_{E_j \in \mathcal{E}_1} P(A \mid E_j)\Bigg[\beta(n_1)P(E_j) + (1-\beta(n_1))\frac{1}{n_1}\sum_{i=1}^{n_1} \mathbb{I}(E_j = E_i)\Bigg]. \tag{4}$$

Let then $n_1 \to \infty$; following (4), we have that $\lim_{n_1 \to \infty} P_{\mathcal{E}_1}(A)$ is equal to

$$\sum_{E_j \in \tilde{\mathcal{E}}} \Bigg\{P(A \mid E_j)\Bigg[\lim_{n_1 \to \infty} \beta(n_1)P(E_j) + \lim_{n_1 \to \infty} \frac{1-\beta(n_1)}{n_1} \sum_{i=1}^{n_1} \mathbb{I}(E_j = E_i)\Bigg]\Bigg\}$$
$$= \sum_{E_j \in \tilde{\mathcal{E}}} \Bigg\{P(A \mid E_j)\Bigg[\lim_{n_1 \to \infty} \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbb{I}(E_j = E_i)\Bigg]\Bigg\}. \tag{5}$$

The equality in (5) is a consequence of assumption 1 in our statement. Assumptions 2 and 3 are needed to apply the finitely additive version of Glivenko-Cantelli theorem [9, Remark 2]. It ensures us that $F_X^{n_1}$ converges uniformly almost surely to F_X as $n_1 \to \infty$, where $F_X^{n_1}$ is the empirical cdf of $x_1, \ldots, x_{n_1}$. In turn, this implies that

$$\lim_{n_1 \to \infty} \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbb{I}(E_j = E_i) = Q(E_j)$$

with Q-probability 1, for all $E_j \in \tilde{\mathcal{E}}$. We considered $t = 1$ to highlight the dependence of the limiting distribution on the prior P. For a generic $t \in \mathbb{N}$, we have that

$$\lim_{n_t \to \infty} P_{\mathcal{E}_t}(A) = \sum_{E \in \tilde{\mathcal{E}}} P_{\mathcal{E}_{t-1}}(A \mid E) Q(E) \tag{6}$$

almost surely, for all $A \in \mathcal{F}$. Notice that $P_{\mathcal{E}_{t-1}}(A \mid E)$ does not depend on n_t, and $P_{\mathcal{E}_{t-1}}$ "contains" the prior as shown in Remark 2. We denote $P_{\tilde{\mathcal{E}}}(A) := \sum_{E \in \tilde{\mathcal{E}}} P_{\mathcal{E}_{t-1}}(A \mid E) Q(E)$, for all $A \in \mathcal{F}$. Finally, notice that (6) entails that $\lim_{n_t \to \infty} d_{TV}(P_{\mathcal{E}_t}, P_{\tilde{\mathcal{E}}}) = 0$ almost surely, concluding the proof. ■

Proof [Proof of Proposition 5] We first point out that $\tilde{\mathcal{E}} = \tilde{\mathcal{E}}'$. This because, no matter the order in which we collect data points $x_i \in \mathcal{X}$, in the limit we either end up observing all the elements of $\mathcal{X}$, or all the elements of $\mathcal{X}_{reduced}$ in the case $\cup_{i \in \mathbb{N}} x_i = \mathcal{X}_{reduced} \subsetneq \mathcal{X}$. So if $\tilde{\mathcal{E}}$ is finer than $\tilde{\mathcal{E}}'$, this means that there exists an ω that is mapped by X into two different values, a contradiction. If instead $\tilde{\mathcal{E}}$ is coarser than $\tilde{\mathcal{E}}'$, this means that $\tilde{\mathcal{E}}$ can be further refined, which contradicts Proposition 3. Then, the claim follows by the uniqueness of the limit of a sequence. ■

Proof [Proof of Proposition 6] Fix any $t \in \mathbb{N}$, and let $\mathcal{P}_{\mathcal{E}_t} = \{\check{P}_{k,\mathcal{E}_t}\}$. Pick any $P_{\mathcal{E}_t} \in \mathcal{P}^{\text{co}}_{\mathcal{E}_t}$. Then, by the convexity of $\mathcal{P}^{\text{co}}_{\mathcal{E}_t}$, there exists a collection $\{\alpha_k\} \subset [0,1]$ such that $\#\{\alpha_k\} = \#\mathcal{P}_{\mathcal{E}_t}$, $\sum_k \alpha_k = 1$, and $P_{\mathcal{E}_t}(A) = \sum_k \alpha_k \check{P}_{k,\mathcal{E}_t}(A)$, for all $A \in \mathcal{F}$. By construction and Theorem 4, given our assumptions we know that for all k, $d_{TV}(\check{P}_{k,\mathcal{E}_t}, \check{P}_{k,\tilde{\mathcal{E}}}) \to 0$ as n_t goes to infinity with probability 1, where $\check{P}_{k,\tilde{\mathcal{E}}}$ is an element of $\mathcal{P}_{\tilde{\mathcal{E}}} = \{\check{P}_{k,\tilde{\mathcal{E}}}\}$. So we can conclude that there is $P_{\tilde{\mathcal{E}}} \in \mathcal{P}^{\text{co}}_{\tilde{\mathcal{E}}}$ such that $P_{\tilde{\mathcal{E}}}(A) = \sum_k \alpha_k \check{P}_{k,\tilde{\mathcal{E}}}(A)$, for all $A \in \mathcal{F}$, and $d_{TV}(P_{\mathcal{E}_t}, P_{\tilde{\mathcal{E}}}) \to 0$ as n_t goes to infinity with Q-probability 1.

That is to say that for every element $P_{\mathcal{E}_t}$ of $\mathcal{P}^{\text{co}}_{\mathcal{E}_t}$, there is an element $P_{\tilde{\mathcal{E}}}$ of $\mathcal{P}^{\text{co}}_{\tilde{\mathcal{E}}}$ that $P_{\mathcal{E}_t}$ converges to (with probability 1 in the total variation metric). This immediately implies that the Hausdorff distance between $\mathcal{P}^{\text{co}}_{\mathcal{E}_t}$ and $\mathcal{P}^{\text{co}}_{\tilde{\mathcal{E}}}$ goes to 0 as n_t goes to infinity with probability 1. ■

Author Contributions

Both of the authors contributed equally to this paper.

Acknowledgments

The authors thank Sayan Murkherjee for inspiring this project and helpful discussions, Teddy Seidenfeld for a fruitful discussion about the commutativity of dynamic probability kinematics, Xiao-Li Meng for an insightful dialogue on information theory, Alessandro Zito for his help with Figure 1, and three anonymous referees for their constructive comments. Research of Ruobin Gong is supported in part by the National Science Foundation (DMS-1916002). Michele Caprio would like to acknowledge partial funding by the National Science Foundation (CCF-1934964) and the Army Research Office (ARO MURI W911NF2010080).

Appendix A. Further Related Literature

In Section 2, we inspected previous works extending PK to credal sets theory. Probability kinematics has been generalized to the contexts of Dempster-Shafer theory, evidence theory, neighborhood models theory, possibility theory, and maximum entropy theory as well.

Shafer [27] discusses Jeffrey's updating from a philosophical perspective, and is the first to consider its application to the context of Dempster-Shafer theory, for which belief functions – functions representing the degree of belief of the agent on a given event – and Dempster's updating rule play a central operational role. Ichihashi and Tanaka [17] and Smets [29] further study the generalization of Jeffrey's updating for belief functions defined on a finite state space. Ichihashi and Tanaka [17] point out how Shafer's approach is different from the normative Bayesian approach and is not a straight generalization of Jeffrey's rule, so they propose rules of conditioning for which Jeffrey's rule is a direct consequence of a special case. Smets [29] generalizes the results in [17] and shows that several forms of Jeffrey's updating rule can be defined so that they correspond to the geometrical rule of conditioning and to Dempster's rule of conditioning, respectively.

Ma et al. [25] provide a generalization of both Jeffrey's rule and Dempster's conditioning to propose an effective revision rule in the field of evidence theory. This is very interesting since when one source of evidence is less reliable than another, the idea is to let prior knowledge of an agent be altered only by some of the input information. The change problem is thus intrinsically asymmetric. To this extent, their model takes into account inconsistency between prior and input information. Other works that deal with a generalization of Jeffrey's rule within the framework of evidence theory are [30], in which the authors propose a generalization of probability kinematics where a priori knowledge and new evidence are all modeled by independent random sets, and [24] in which a priori knowledge and evidences are modelled by a probability distribution and a collection of multi-dimensional random sets, respectively.

Škulj [28] discusses the application of Jeffrey's rule to neighborhood models theory, in which ambiguity is captured by neighborhood of a classical probability measure P and presented in the form of interval probabilities $[L, U]$. This means that $P(A) \in [L(A), U(A)]$ for all $A \subset \Omega$, the state space of interest. The author shows that a neighborhood $[L, U]$ of a probability measure P whose lower envelope L is convex or bi-elastic with respect to the base probability measure [28, Definitions 3 and 4] is closed with respect to Jeffrey's rule of conditioning. This means that Jeffrey's posterior for $Q \in [L, U]$ still belongs to the interval.

Possibility theory [33] is a framework alternative to probability theory that is suitable for handling uncertain, imprecise and incomplete knowledge. In possibility theory, there are two different ways to define the conditioning depending on how possibility degrees are interpreted, one called quantitative possibility and the other called qualitative possibility. In [2], the authors investigate the existence and uniqueness of the posterior probabilities computed according to a possibilistic counterpart of Jeffrey's rule in both the quantitative and qualitative possibilistic frameworks.

References

[1] Mihir Bellare and Phillip Rogaway. Optimal asymmetric encryption. In Alfredo De Santis, editor, *Advances in Cryptology — EUROCRYPT'94*, pages 92–111. Springer Berlin Heidelberg, 1995.

[2] Salem Benferhat, Karim Tabia, and Karima Sedki. Jeffrey's rule of conditioning in a possibilistic framework: an analysis of the existence and uniqueness of the solution. *Annals of Mathematics and Artificial Intelligence*, 61:185–202, 2011.

[3] James O. Berger. The robust Bayesian viewpoint. In Joseph B. Kadane, editor, *Robustness of Bayesian Analyses*. Amsterdam : North-Holland, 1984.

[4] James O. Berger. *Statistical Decision Theory*. Springer Series in Statistics. Springer, New York, 2nd edition, 1985.

[5] James O. Berger and L. Mark Berliner. Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors. *The Annals of Statistics*, 14(2):461–486, 1986.

[6] Lawrence Blume, Adam Brandenburger, and Eddie Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica*, 59:61–79, 1991.

[7] Michele Caprio and Ruobin Gong. Dynamic imprecise probability kinematics. *Available at arXiv:2110.04382*, 2023.

[8] Michele Caprio and Sayan Mukherjee. Ergodic theorems for dynamic imprecise probability kinematics. *International Journal of Approximate Reasoning*, 152: 325–343, 2023.

[9] Robert Chen and S. Ramakrishnan. On the Glivenko-Cantelli theorem in a finitely additive setting. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 45(2):161–167, 1983.

[10] Giulianella Coletti and Romano Scozzafava. *Probabilistic Logic in a Coherent Setting*. Trends in Logic. Dordrecht : Springer, 2002.

[11] Persi Diaconis and Sandy L. Zabell. Updating subjective probability. *Journal of the American Statistical Association*, 77(380):822–830, 1982.

[12] Lester E. Dubins. On Lebesgue-like extensions of finitely additive measures. *The Annals of Probability*, 2(3):456–463, 1974.

[13] Daniel Ellsberg. Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 75(4): 643–669, 1961.

[14] Ruobin Gong and Xiao-Li Meng. Judicious judgment meets unsettling updating: dilation, sure loss, and Simpson's paradox. *Statistical Science*, 36(2):169–190, 2021.

[15] Peter J. Huber. The use of Choquet capacities in statistics. *Bulletin of the International Statistical Institute*, 4(45):181–191, 1973.

[16] Peter J. Huber and Elvezio M. Ronchetti. *Robust statistics*. Wiley Series in Probability and Statistics. Hoboken, New Jersey : Wiley, 2nd edition, 2009.

[17] Hidetomo Ichihashi and Hideo Tanaka. Jeffrey-like rules of conditioning for the Dempster-Shafer theory of evidence. *International Journal of Approximate Reasoning*, 3(2):143–156, 1989.

[18] Richard C. Jeffrey. *Contributions to the Theory of Inductive Probability*. PhD Thesis, Princeton University, Dept. of Philosophy, 1957.

[19] Richard C. Jeffrey. *The Logic of Decision*. Chicago : University of Chicago Press, 1965.

[20] Richard C. Jeffrey. Probable knowledge. In Imre Lakatos, editor, *The Problem of Inductive Logic*, volume 51 of *Studies in Logic and the Foundations of Mathematics*, pages 166 – 190. Elsevier, 1968.

[21] Anatolii G. Kusraev and Sergei A. Malyugin. Extensions of finitely additive measures. *Mathematical notes of the Academy of Sciences of the USSR*, 48: 658–661, 1990.

[22] Isaac Levi. *The Enterprise of Knowledge*. Cambridge, MA : MIT Press, 1980.

[23] David Lewis. Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85(3):297–315, 1976. ISSN 00318108, 15581470. URL http://www.jstor.org/stable/2184045.

[24] Hexin Lv, Ning Qiu, and Yongchuan Tang. Updating probabilistic knowledge using imprecise and uncertain evidence. In *Third International Conference on Natural Computation (ICNC 2007)*, volume 4, pages 624–628, 2007.

[25] Jianbing Ma, Weiru Lu, Didier Dubois, and Henri Prade. Bridging Jeffrey's rule, AGM revision and Dempster conditioning in the theory of evidence. *International Journal on Artificial Intelligence Tools*, 20(4):691–720, 2011.

[26] Sabina Marchetti and Alessandro Antonucci. Imaginary kinematics. In Amir Globerson and Ricardo Silva, editors, *UAI 2018: Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence*, pages 104–113, Monterey, California, USA, 2018. AUAI Press. URL http://auai.org/uai2018/proceedings/papers/42.pdf.

[27] Glenn Shafer. Jeffrey's rule of conditioning. *Philosophy of Science*, 48(3):337–362, 1981.

[28] Damjan Škulj. Jeffrey's conditioning rule in neighbourhood models. *International Journal of Approximate Reasoning*, 42(3):192–211, 2006.

[29] Philippe Smets. Jeffrey's rule of conditioning generalized to belief functions. In *Proceedings of the Ninth international conference on Uncertainty in Artificial Intelligence*, pages 500–505, 1993.

[30] Yongchuan Tang, Shouqian Sun, and Zhongyang Li. Conditional evidence theory and its application in knowledge discovery. *Lecture Notes in Computer Sciences*, 3007:500–505, 2004.

[31] Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley Series in Probability and Statistics. New York : Wiley, 2014.

[32] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*, volume 42 of *Monographs on Statistics and Applied Probability*. London : Chapman and Hall, 1991.

[33] Lotfi Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1:3–28, 1978.