Dynamic Precise and Imprecise Probability Kinematics

Michele Caprio & Ruobin Gong

PRECISE Center, Dept. of Computer and Information Science, University of Pennsylvania

Dept. of Statistics, Rutgers University

caprio@seas.upenn.edu, ruobin.gong@rutgers.edu

1. Probability kinematics

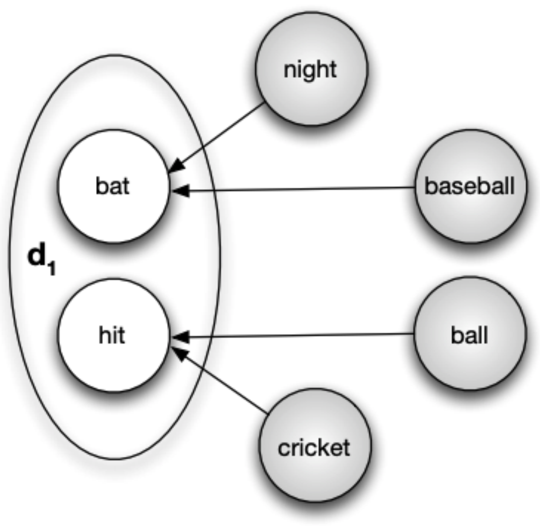
The goal of this paper is to introduce a method to update mechanically the subjective beliefs of an agent that faces ambiguity and who is only able to collect partial information.

PROBABILITY kinematics (PK), also known as Jeffrey's rule of updating, is the following

$$P^\star(A) = \sum_{E\in\mathcal{E}} P(A \mid E)P^\star(E).$$

It is valid when there is a partition $\mathcal{E}$ of the state space Ω such that $P^\star(A \mid E) = P(A \mid E)$, for all $A \subset \Omega$ and all $E \in \mathcal{E}$. It is useful when evidence is not propositional (i.e. it is be possible to represent it as a crisp subset), but instead is uncertain or partial.

PK 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.



2. Dynamic Probability Kinematics (DPK)

- Let Ω be the state space of interest, and assume it is at most countable.
- Suppose that P is a finitely additive probability measure on $(\Omega, \mathcal{F})$ representing an agent's initial beliefs around the elements of $\mathcal{F} = 2^\Omega$.
- 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.
- Call F_X its cdf, and assume $\mathcal{X}$ is finite.
- 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, $m \leq n$. Unique elements of $\mathcal{E}'$: $E_1, \dots, E_m$; complement of their union: $E_{m+1} = (\cup_{j=1}^m E_j)^c = \Omega \setminus \cup_{j=1}^m E_j$.

DPK is defined as

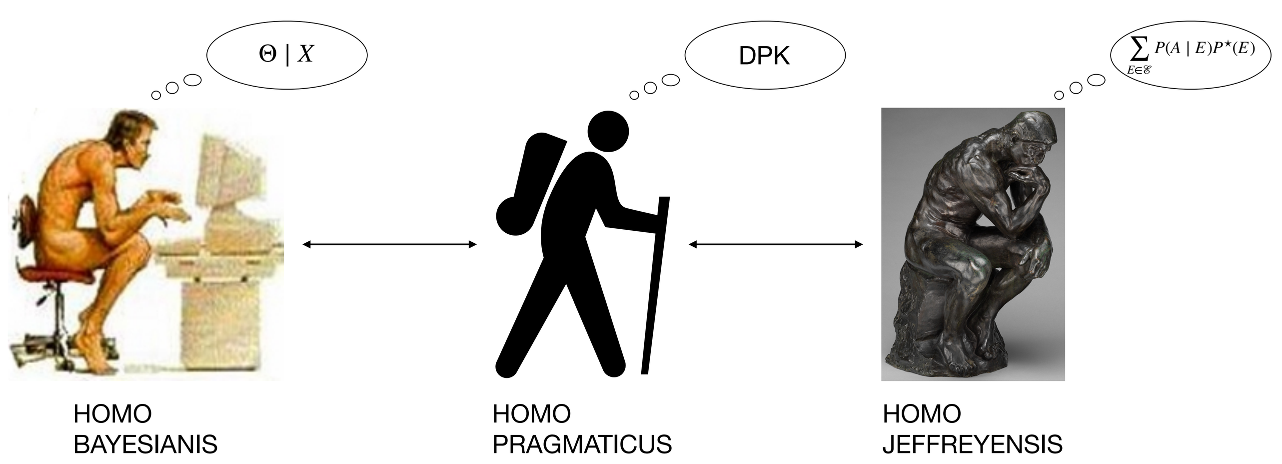
$$P_{\mathcal{E}}(A) := \sum_{E_j\in\mathcal{E}} P(A \mid E_j)P_{\mathcal{E}}(E_j),$$

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

where $\beta(n)$ is a coefficient in $[0, 1]$ depending on n, and, for all $j \in \{1, \dots, m\}$,

$$P^{emp}(E_j) := \begin{cases} \frac{1}{n+1}\sum_{i=1}^n \mathbb{I}(E_j = E_i), & \text{if } E_{m+1} \neq \emptyset \\ \frac{1}{n}\sum_{i=1}^n \mathbb{I}(E_j = E_i), & \text{if } E_{m+1} = \emptyset \end{cases}.$$

Of course, $P^{emp}(E_{m+1}) = 1 - \sum_{j=1}^m P^{emp}(E_j)$.



3. Jeffrey's rule vs DPK

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.

4. Subsequent DPK updates

Notice that as $t \to \infty$, $n_t \to \infty$. In addition, a consequence of how we build partitions is that, for any t, $\mathcal{E}_t$ is not coarser than $\mathcal{E}_{t-1}$.

Proposition 1 *There exists a partition $\tilde{\mathcal{E}}$ that cannot be refined as a result of the DPK updating process.*

Call Q the "objective" (finitely additive) probability measure on $(\Omega, \mathcal{F})$, $Q(X^{-1}(x)) = P_X(\{x\})$, for all $x \in \mathcal{X}$.

Theorem 2 *Suppose the following conditions hold,*

1. $\lim_{n_t\to\infty} \beta(n_t) = 0$;

2. for every discontinuity point x of F_X, $F_X(x^+) - F(x^-) = P_X(\{x\})$;

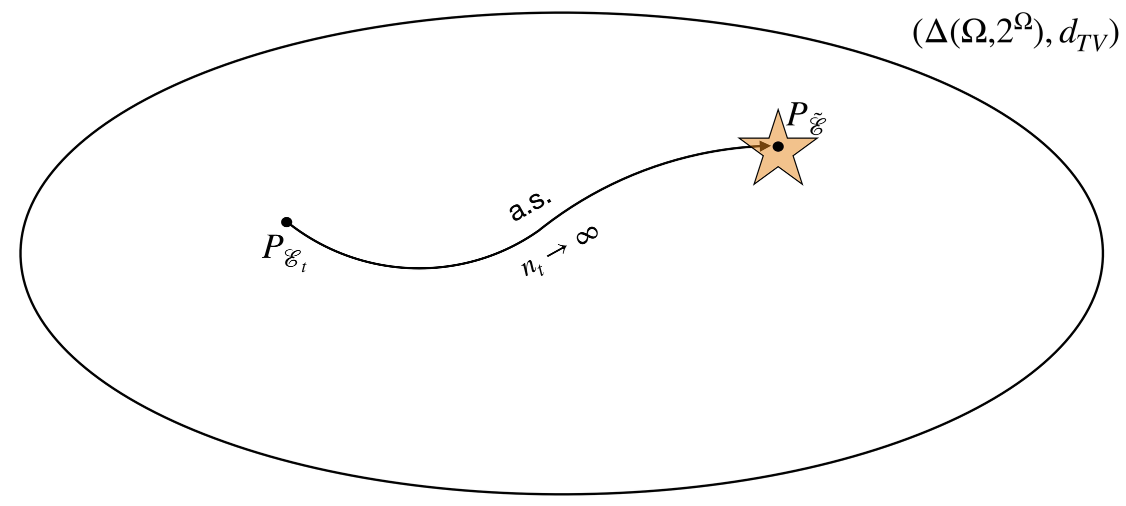
3. $F_X(-\infty) = 1 - F_X(\infty) = 0$.

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.

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 2.

Dynamic probability kinematics is not commutative. Despite that, the limit probability $P_{\tilde{\mathcal{E}}}$ is the same regardless of the order in which data is collected.

Proposition 3 *Suppose that conditions 1-3 of Theorem 2 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, $P_{\tilde{\mathcal{E}}} = P_{\tilde{\mathcal{E}}'}$.*



5. Working with sets of probabilities

Definition needed:

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

It is is convex and weak*-compact [1, Theorem 3.6.2]. We assume it has finite number of extrema. To generalize DPK to DIPK, we prescribe the agent to

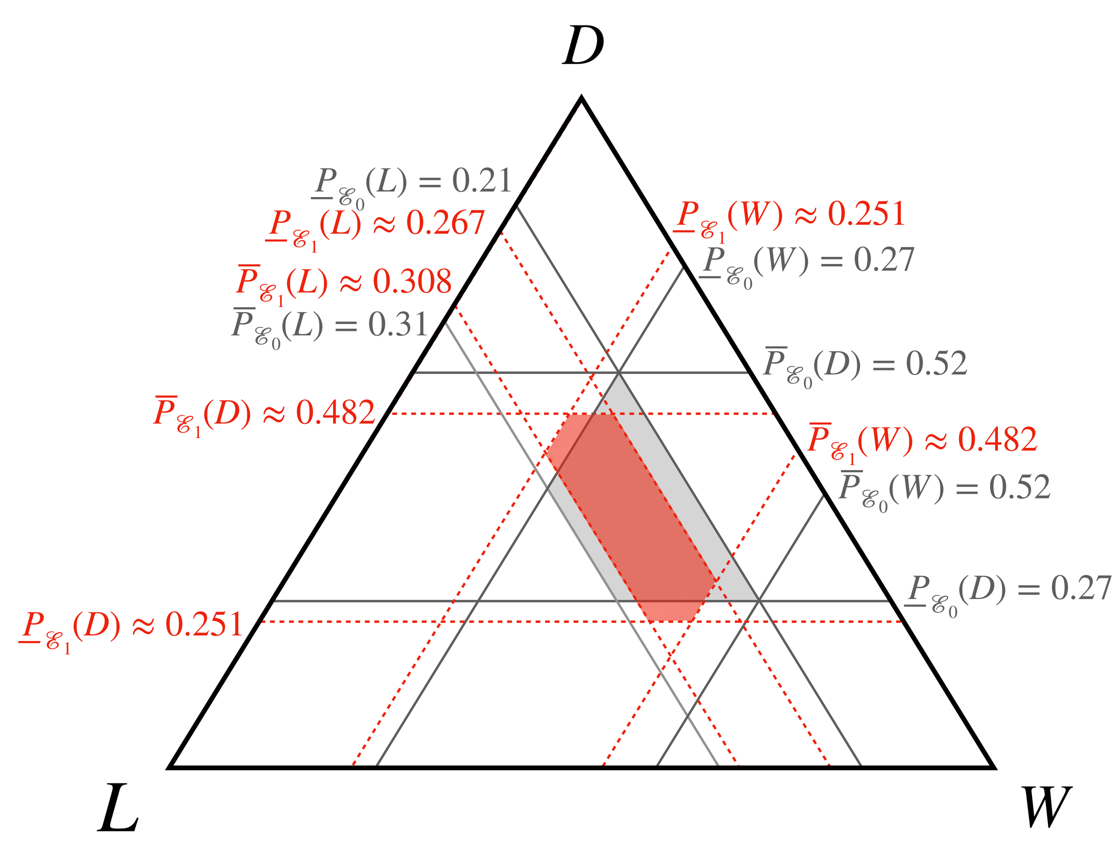
1. Specify a set of probabilities $\mathcal{P}$;
2. Compute the lower probability $\underline{P}$ associated with it. The core core$(\underline{P}) \equiv \mathcal{P}_0^{\text{co}}$ of such lower probability represents the agent's initial beliefs.
3. Consider the set $\mathcal{P}_{\mathcal{E}_0} = \text{ex}\mathcal{P}_{\mathcal{E}_0}^{\text{co}}$ of extrema of $\mathcal{P}_{\mathcal{E}_0}^{\text{co}}$
4. Compute the DPK update of every element in $\mathcal{P}_{\mathcal{E}_0}$.
5. Compute $\mathcal{P}_{\mathcal{E}_1}^{\text{co}} = \text{Conv}(\mathcal{P}_{\mathcal{E}_1}) = \text{core}(\underline{P}_{\mathcal{E}_1})$, where $\underline{P}_{\mathcal{E}_1}$ is the updated lower probability [1, Theorem 3.6.2].

Proposition 4 *If assumptions 1-3 of Theorem 2 hold, then $d_H(\mathcal{P}_{\mathcal{E}_t}^{\text{co}}, \mathcal{P}_{\tilde{\mathcal{E}}}^{\text{co}}) \to 0$ as n_t goes to infinity with probability one.*

6. Examples of DIPK updating: Soccer match results

Let $\Omega = \{W, D, L\}$ represent the result of soccer match Juventus Turin vs Inter Milan. Let then $X : \Omega \to \mathcal{X} = \{0, 1\}$, where 1 denotes a useful result and 0 denotes a defeat. The finest partition of Ω according to DPK is given by $\tilde{\mathcal{E}} = \{E_1, E_2, E_3\}$, where $E_1 = X^{-1}(1) = \{W, D\}$, $E_2 = X^{-1}(0) = \{L\}$, and $E_3 = \emptyset$. The data points $x_1, \dots, x_n$ that we collect represent the outcomes of past matches.

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 $\overline{P}(L) = 0.31$. As of January 12, 2022, there have been 257 matches between the two teams, 178 useful results for Juventus Turin and 79 wins for Inter Milan. We have that $P_1^{emp}(E_1) = 178/257$, $P_1^{emp}(E_2) = 79/257$ and $P_1^{emp}(E_3) = 0$.



References and whole paper can be found here:

ISIPTA 2023, 25º International Symposium on Imprecise Probabilities: Theories and Applications. 11-14 July 2023, Oviedo, Spain