Une courte introduction aux probabilités imprécises

S. Destercke [1,2]

[1] Institut de Radioprotection et de Sûreté Nucléaire (IRSN)
Cadarache, France

[2] Institut de Recherche en Informatique de Toulouse (IRIT)
Université Paul Sabatier

Séminaire IMPEC

Institut de Recherche en Informatique de Toulouse
CNRS
INPT
UPS
UT1

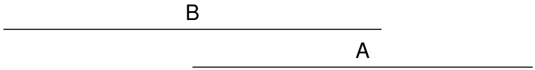
IRSN
INSTITUT DE RADIOPROTECTION ET DE SÛRETÉ NUCLÉAIRE

La modélisation d'incertitudes

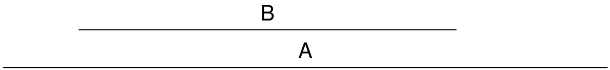
Intervalles

La seule information disponible est $X \in B$. Etant donné mon information, un ensemble A et deux mesures d'incertitudes $\underline{P}(A), \overline{P}(A)$ mesurant la potentialité que $X \in A$, trois situations peuvent arriver:

- $A \cap B \neq \{\emptyset, B\} \rightarrow \overline{P}(A) = 1, \underline{P}(A) = 0$ (ignorance)



- $A \supseteq B \rightarrow \overline{P}(A) = 1, \underline{P}(A) = 1$ (certitude)



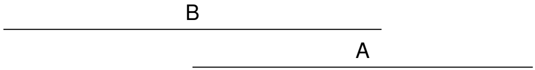
- $A \cap B = \emptyset \rightarrow \overline{P}(A) = 0, \underline{P}(A) = 0$ (certitude)

La modélisation d'incertitudes

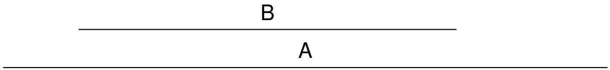
Probabilités

On considère que X prend ses valeurs sur un ensemble fini $\mathcal{X}$, et que l'incertitude est modélisée par une distribution de probabilité ayant pour support B.

1. $A \cap B \neq \{\emptyset, B\} \rightarrow \overline{P}(A) = \underline{P}(A) = P(A) = \sum_{x \in A} p(x)$ (certitude)



2. $A \supseteq B \rightarrow \overline{P}(A) = \underline{P}(A) = 1$ (certitude)



3. $A \cap B = \emptyset \rightarrow \overline{P}(A) = 0, \underline{P}(A) = 0$ (certitude)

Modèle général: Probabilités inférieures

Modèle général

L'incertitude sur une variable X prenant ses valeurs sur un ens. fini $\mathcal{X}$ est modélisée par une de probabilité inférieure $\underline{P}$ vérifiant

1. $\underline{P}(\emptyset) = 0, \underline{P}(\mathcal{X}) = 1$ (bornes)
2. pour tout sous-ens. $A \subseteq B \subseteq \mathcal{X}$, $\underline{P}(A) \leq \underline{P}(B)$ (monotonie)
3. pour tout sous-ens. $A, B \subseteq \mathcal{X}$, $A \cap B = \emptyset$, $\underline{P}(A) + \underline{P}(B) \leq \underline{P}(A \cup B)$ (sous-additivité)

A partir de cette mesure, on peut définir la mesure duale de probabilité supérieure $\overline{P}$ tel que, pour tout $A \subseteq \mathcal{X}$,

$$\overline{P}(A) = 1 - \underline{P}(A^c)$$

qui vérifient les propriétés 1 et 2 de $\underline{P}$ ainsi que

- pour tout sous-ens. $A, B \subseteq \mathcal{X}$, $A \cap B = \emptyset$, $\overline{P}(A) + \overline{P}(B) \geq \overline{P}(A \cup B)$ (sur-additivité)

Deux visions des probabilités imprécises

Robustification des probabilités classiques

Les probabilités imprécises sont vues comme des ensembles de probabilités, modélisant notre méconnaissance de la "vraie" probabilités. Elles offrent de nouveaux outils (plus pratiques) et de nouvelles approches par rapport à l'analyse Bayesienne robuste classique.

Approche similaire à celle de Huber, Berger, ...

Modèles d'un état de connaissance visant à faciliter sa manipulation

La probabilité inférieure modélise un état de connaissance à propos d'une variable, sans qu'on présuppose l'existence d'un modèle parfait et précis sous-jacent.

Approche similaire à celles de Walley, Williams, A.C.B. Smith (De Finetti) ou encore de Shafer et Vovk (Cournot, Kolmogorov) ...

Interprétation: en a-t-on besoin?

Réponse: Oui et non!

Mais Non!

- Analyse de sensibilité d'un modèle
- Calibrage d'un modèle
- Dessin industriel robuste

Mais Si!

- Manipulation de connaissances (Intelligence artificielle, logique sous incertitudes)
- Prise de décision basée sur notre connaissance (Analyse de risque, de fiabilité, décision multicritère)
- Justifications + théoriques/philosophiques

Modèle général: Probabilités inférieures

Ensemble crédal

Une probabilité inférieure $\underline{P}$ induit un ensemble de probabilités (souvent appelé ensemble crédal) $\mathcal{P}_{\underline{P}}$ tel que

$$\mathcal{P}_{\underline{P}} = \{P \in \mathbb{P}_{\mathcal{X}} | \forall A \subseteq \mathcal{X},\ P(A) \geq \underline{P}(A)\}$$

avec $\mathbb{P}_{\mathcal{X}}$ l'ensemble des mesures de probabilités sur $\mathcal{X}$. Une probabilité inférieure est dite

- **consistente** si $\mathcal{P}_{\underline{P}} \neq \emptyset$
- **cohérente** si $\underline{P}(A) = \inf_{P \in \mathcal{P}_{\underline{P}}} P(A)$

En pratique

Difficile de manipuler $2^{\mathcal{X}}$ valeurs. On se restreindra souvent à des modèles qui peuvent être décrits par un nombre limité de valeurs.

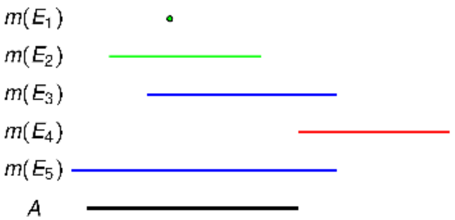
Un modèle utile: les fonctions de croyances

Modèle général

Incertitude sur X décrite par une distribution de masses de croyances $m : 2^{\mathcal{X}} \to [0,1]$ t.q. $\sum_{E \subseteq \mathcal{X}} m(E)$. A partir de cette distribution sont définies deux fonctions d'ensembles telles que, pour $A \subseteq \mathcal{X}$:

$$Bel(A) = \sum_{E \subseteq A} m(E) \quad \text{(Crédibilité)}; \quad Pl(A) = \sum_{E \cap A \neq \emptyset} m(E) = 1 - Bel^c(A) \quad \text{(Plausibilité)}$$

Bijection entre Bel et m (transformée de Möbius).



$Bel(A) = m(E_1) + m(E_2)$

$Pl(A) = m(E_1) + m(E_2) + m(E_3) + m(E_5)$

Ensemble de probabilité induit

Une fonction de croyance *Bel* interprétée comme une probabilité inférieure génère l'ensemble crédal

$$\mathcal{P}_{Bel} = \{P \in \mathbb{P}_{\mathcal{X}} | \forall A \subseteq \mathcal{X},\ Bel(A) \leq P(A)\}$$

Probabilités et intervalles

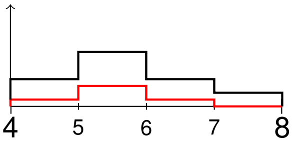
- On retrouve une probabilité P lorsque les masses ne portent que sur les singletons.
- On retrouve l'intervalle A lorsque $m(A) = 1$

Un modèle classique: contamination ϵ

Soit m_1 modélisant une probabilité et m_2 modélisant l'ignorance (i.e. $m_2(\mathcal{X}) = 1$), et $1 - \epsilon$ la "confiance" qu'on accorde au modèle précis m_1. Le modèle de contamination m_ϵ s'écrit alors:

$$m_\epsilon = (1 - \epsilon)m_1 + (\epsilon)m_2$$

Intervalles de probabilités (De campos, Huete, Moral)



Histogrammes imprécis, échantillons de petites tailles sur données multinomiales

Définition

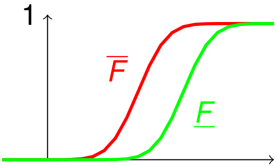
Ensemble $L = \{[l(x), u(x)] | x \in \mathcal{X}\}$ d'intervalles sur éléments de $\mathcal{X}$. Ensemble de probabilités induit

$$\mathcal{P}_L = \{P \in \mathbb{P}_{\mathcal{X}} | \forall x,\ l(x) \leq p(x) \leq u(x)\}$$

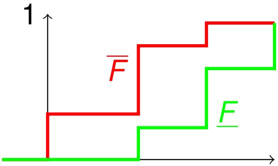
Probabilité inf. de $A \subseteq \mathcal{X}$:

$$\underline{P}(A) = \max(\sum_{x \in A} l(x), 1 - \sum_{x \in A^c} u(x))$$

P-boîtes : cumulées imprécises



P-boîte continue



P-boîte discrète

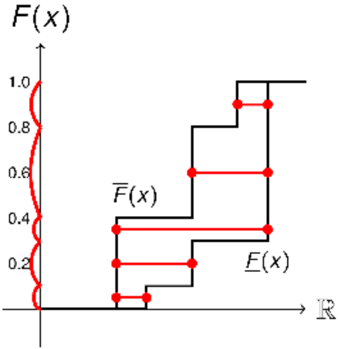
Définition

Paire de distributions cumulées $[\underline{F}, \overline{F}]$ sur $\mathbb{R}$.
Ensemble de probabilités induit

$$\mathcal{P}_{[\underline{F},\overline{F}]} = \{P | \forall x,\ \underline{F}(x) \leq P((-\infty, x) \leq \overline{F}(x)\}$$

Opinions d'experts, petits échantillons
+ distance Kolmogorov-Smirnov

Les p-boîtes comme fonctions de croyance



Une P-boîte (discrète) $[\underline{F}, \overline{F}]$ a une probabilité inférieure qui est ∞-monotone $\Rightarrow$ on peut construire une distribution de masse correspondante

Soit $0 = \alpha_0 < \alpha_1 < \ldots < \alpha_M = 1$ l'ensemble (fini) de valeurs prises par les distributions $\overline{F}, \underline{F}$.

Fct de Croy. Correspondante avec ens. focaux E_i tels que, pour $i = 1, \ldots, M$:

$$E_i = \{x \in \mathbb{R} | \overline{F}(x) \geq \alpha_i \wedge \sup\{\underline{F}(r') | r' \in \mathbb{R}; \underline{F}(r') < \underline{F}(r)\} < \alpha_i\}$$

$$m(E_i) = \alpha_i - \alpha_{i-1}$$

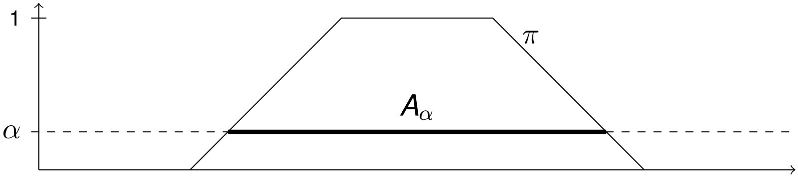
Distributions de possibilités

Definition

Une distribution de possibilité π est une fonction $\pi : \mathcal{X} \rightarrow [0, 1]$ et $\exists x \in \mathcal{X}$ t.q. $\pi(x) = 1$ à partir de laquelle on définit deux mesures (duales)

- Mesure de possibilité: $\Pi(A) = \sup_{x \in A} \pi(x)$ (maxitive)
- Mesure de Nécessité: $N(A) = 1 - \Pi(A^c)$

Une alpha-coupe se définit comme $A_\alpha = \{x \in \mathcal{X} | \pi(x) \geq \alpha\}$ (elle est dite stricte si l'inégalité est stricte)



Les possibilités comme fonctions de croyance

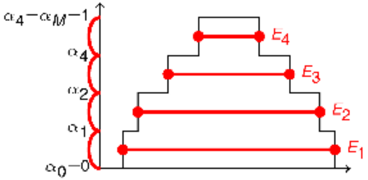
Une mesure de nécessité est ∞-monotone, et correspond donc à une fonction de croyance particulière, dont les ensembles focaux sont emboîtés.

Soit
$0 = \alpha_0 < \alpha_1 < \ldots < \alpha_M = 1$
l'ensemble (fini) de valeurs prises par la distribution π.

Fct de Croy. Correspondante avec ens. focaux E_i tels que, pour $i = 1, \ldots, M$:

$$E_i = \{x \in \mathcal{X} | \pi(x) \geq \alpha_i\}$$

$$m(E_i) = \alpha_i - \alpha_{i-1}$$



Les possibilités comm ens. de probabilités

L'ensemble crédal induit par une possibilité π est tel que

$$\mathcal{P}_\pi = \{P \in \mathbb{P}_\mathcal{X} | \forall A \subseteq \mathcal{X},\ P(A) \geq N(A)\}$$

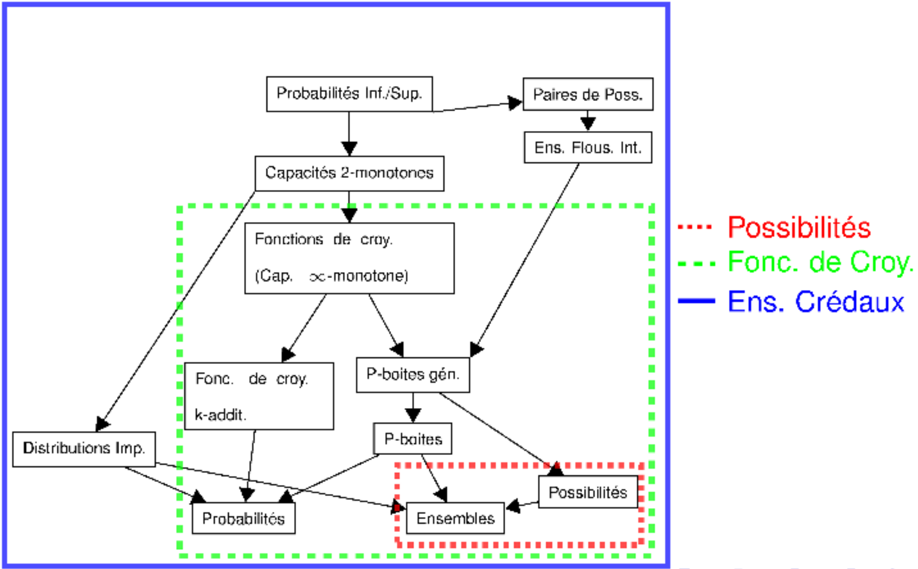
Autre formulation

Etant donné π, une mesure de probabilité P est dans $\mathcal{P}_\pi$ ssi pour tout α in $(0, 1]$:

$$1 - \alpha \leq P(\{x \in \mathcal{X} | \pi(x) > \alpha\})$$

$\rightarrow$ permet d'interpréter les possibilités comme des intervalles emboîtés avec niveaux de confiances inférieures.

Une cartographie générale



Quelques opérations impossibles avec le cadre classique

- Etendre l'intersection/union d'ensemble requiert une définition de la notion d'inclusion. Rien de tel en probabilités classiques, ce qui restreint la manière dont on peu agréger les informations.
- Peu/pas de places pour des utilités imprécises ou des aides à la décision qui proposent un ensemble de décisions optimales (plutôt qu'une)
- Modélisation de l'ignorance (pas d'informations a priori)
- Modélisation de l'imprécision (voir Walley, Chapitre 2)
- Techniques discutables pour traiter des données manquantes/imprécises (impossibilité de considérer l'ensemble des modèles cohérents)

Battons en brèche quelques idées pré-conçues (1)

Ha, vous faites de la logique floue?

Non, la logique floue ne parle **pas de modélisation d'incertitude**, il s'agit:

- d'étendre la logique classique
- de modéliser des notions graduelles, des termes linguistiques (but original)
- de fournir des approximateurs universels linéaires par partie (équivalence RN)

non-probabiliste classique $\neq$ logique floue

Battons en brèche quelques idées pré-conçues (2)

Vous utilisez des outils sans assises théoriques solides, au mieux se sont des usines à gaz, au pire des choses totalement inutiles.

Plus ou moins vrai il y a 20 ans... depuis, pas mal de choses ont été faites:

- Nombreux chercheurs ont justifié, théoriquement et philosophiquement, l'approche probabiliste imprécises (Smets, Walley, Shafer/Vovk, Seidenfeld, Williams, . . .) et le besoin de la prendre en compte
- Nombreux résultats classiques ont été étendus au cas imprécis (Lois des grands nombres, notions d'indépendances, chaînes de markov, réseaux Bayésiens et d-séparation)
- Les approches permettent de répondre à des problèmes théoriques rencontrés par les probabilités classiques (e.g. confrontation fréquentistes/Bayésiens, mutliples lois a priori dans les modèles de Dirichlet)

Mais de nombreux résultats restent à démontrer, et certains points à éclaircir!

Battons en brèche quelques idées pré-conçues (3)

C'est une belle théorie, mais elle est inapplicable/vous n'avez pas d'appli!

- De plus en plus d'applications. . . :
  - **Intelligence artificielle**: Ext. Réseaux Bayésiens app à reconnaissance militaire, risque de glissement de terrains (Marco Zaffalon, Alessandro Antonucci), fusion multi-capteurs (fonctions de croyances, Thierry Denoeux)
  - **Dessin industriel** en Aéronautique (Oberguggenberger et al.) et en aérospatial (Neumaier & Fuchs)
  - **Classification**: Amélioration des performances des arbres de décisions (Thomas Augustin), robustification de méthodes classiques (Modèle imprécis de Dirichlet, J-M Bernard)
  - **Etudes de risques**: environnementaux (Ferson), climatologie (Kriegler/Held), Radioprotection/sureté nucléaire (nous)
- . . . et les outils qui vont avec (implémentations en R, C++, Matlab).
- Parfois plus (ou tout aussi) efficaces que probabilités (construction de régions de confiance mutli-dimensionelles convexes + optimisation, chaînes de Markov imprécises)

Cependant, il est vrai qu'il reste beaucoup de travail à ce niveau!

Conclusions

Au final, les probabilités imprécises

- sont très satisfaisantes d'un point de vue intellectuel
- permettent parfois d'apporter un nouvel éclairage sur certaines problématiques
- donnent des réponses parfois meilleures que les approches probabilistes classiques

Je dois me convertir tout de suite ?

- Non, si les probabilités/intervalles suffisent à résoudre votre problème ou répondent à vos besoins, ou si vous êtes convaincus de leur universalité.
- D'un point de vue ingénieur, pas forcément... est-ce que les utiliser améliore le processus?

Mais ces théories ont nettement plus que le mérite d'exister! On peut pê s'en passer, mais pas les rejeter en bloc sans au moins en avoir une idée.