Sets of Probability Distributions and Independence

Fabio G. Cozman
Escola Politécnica, Universidade de São Paulo,
São Paulo, SP - Brazil

Abstract

This paper discusses concepts of independence and their relationship with convexity assumptions in the theory of sets of probability distributions. The paper offers an organized review of the literature and some new ideas (on regular conditional independence and exchangeability/"strong independence"). Finally, the connection between recent developments on the axiomatization of non-binary preferences, and its impact on "strict" independence, are analyzed.

1 Introduction

The goal of this paper is to offer an updated analysis of independence concepts for sets of probability distributions. A review of the recent literature on independence concepts for sets of probability distributions is presented in Section 3, where many approaches and arguments are organized into a few related strands. The discussion about independence is so rich that it can shed light on many other aspects of the theory of sets of distributions. Section 4 reviews the pioneering arguments of Kyburg and Pittarelli on the behavioral justification of general sets of probability distributions, and connects these arguments with recent work on axiomatization of non-binary preferences. Connections between these recent results and independence concepts are presented.

2 Convex Bayesianism and E-admissibility

Strict Bayesianism employs a single probability distribution for decision making and deliberation, while *convex* Bayesianism employs a convex set of probability distributions for the same purposes [46]. A convex Bayesian should take the convex hull of any given set of distributions; the resulting convex set is the set of permissible resolutions for the conflict amongst probability distributions.

We refer to a set of probability distributions as a *credal set*; if a credal set contains distributions for variable X, then it is denoted by $K(X)$. The same terminology and notation is used to refer to sets of full conditional measures. (Convexity here means that if P_1 and P_2 are in a credal set, then $\alpha P_1 + (1 - \alpha)P_2$ is also in the credal set for $\alpha \in (0, 1)$.)

The use of credal sets is often justified through partially ordered preferences. Suppose one is to buy or sell random variables X_i; to simplify matters, suppose all X_i are real valued and bounded. One may postulate a binary relation $\succ$, indicating by $X \succ Y$ that "X is preferred to Y." If $\succ$ is a complete order, then relatively simple axioms [26] yield a representation for $\succ$ in terms of a single probability measure P:

$$X \succ Y \quad \text{iff} \quad E_P[X] > E_P[Y]\,.$$

If $\succ$ is a *partial* order, then similarly simple axioms yield a representation for $\succ$ in terms of a set of probability measures K [29, 64, 53]:

$$X \succ Y \quad \text{iff} \quad E_P[X] > E_P[Y] \text{ for all } P \in K.$$

Note that any two credal sets with identical vertices produce the same $\succ$. Interest has focused on the unique *maximal* credal set that represents $\succ$, as this credal set embodies a "least commitment" strategy: every distribution that is allowed by $\succ$ is included in K. Indeed there has been almost unanimous agreement that there is no behavioral reason to move beyond such maximal credal sets as representations for partial preferences. As we will see in Section 4, Kyburg and Pittarelli challenge such common wisdom in a clever way that has only recently been revived.

Another way to produce credal sets is through one-sided betting, a scheme that was explored by Smith [60] and Williams [68, 69], and later investigated in great detail by Walley [64]. In this case the *lower expectation* $\underline{E}[X]$ for X is interpreted as the supremum of prices one is willing to pay so as to purchase X (Walley refers to X as a *gamble*). Conditions usually imposed on the functional $\underline{E}$ are [64]:

$$\underline{E}[X] \geq \inf X; \quad \underline{E}[\alpha X] = \alpha \underline{E}[X] \text{ for } \alpha \geq 0; \quad \underline{E}[X+Y] \geq \underline{E}[X] + \underline{E}[Y]\,.$$

Any functional $\underline{E}[X]$ satisfying these conditions is the lower envelope of a set of expectation functionals or, equivalently, any functional $\underline{E}[X]$ can be represented by a credal set. The *maximal* credal set $K_{\underline{E}}$ that represents $\underline{E}[X]$ is convex. The literature has, for the most part, accepted that only the maximal (convex) credal set is of interest, and that smaller (nonconvex) credal sets have no apparent behavioral justification. Again, this issue is addressed in Section 4.

There are several criteria for decision making with credal sets. To simplify the discussion, suppose one has a set $\mathcal{A}$ of variables and one or more variables must be selected according to some criterion. Each variable is interpreted as an *act* with consequences expressed in utiles; that is, higher values of X are more desirable than lower values of X.

The Γ-*minimax* criterion selects any act with maximum lower expected value [3, 27],

$$\arg \max_{X \in \mathcal{A}} \min_{P \in K} E_P[X]\,.$$

The *maximality* criterion selects any act X such that no other act $Y \in \mathcal{A}$ is preferred to X in a binary comparison [58] [64, Sect. 3.9]. Maximality focuses on the maximal elements of the partial order $\succ$. That is, X is *maximal* if

$$\text{there is no } Y \in \mathcal{A} \text{ such that } E_P[Y - X] > 0 \text{ for all } P \in K.$$

Finally, the *E-admissibility* criterion selects any act X that is maximal under at least one probability measure $P \in K$ [46]. That is, X is *E-admissible* if

$$\text{there is } P \in K \text{ such that } E_P[X - Y] \geq 0 \text{ for all } Y \in \mathcal{A}.$$

Kyburg and Pittarelli explain E-admissibility as follows [42]. Let $K_{\mathcal{A}}(X)$ be the (convex) set of probability distributions relative to which act X maximizes expected value against acts in $\mathcal{A}$. Then X is E-admissible iff

$$K_{\mathcal{A}}(X) \cap K \neq \emptyset.$$

There are other criteria for decision-making with credal sets, such as Γ-maximax and interval dominance [61].

3 Independence concepts for credal sets

The starting point of our discussion of independence is the widely used concept of *stochastic independence*: X and Y are *stochastically independent* if

$$P(X \in A | Y \in B) = P(X \in A) \quad \text{whenever } P(Y \in B) > 0,$$

for all events A and B in appropriate algebras. This definition is equivalent to

$$P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$$

for all events A and B in appropriate algebras.

We devote this section to an analysis of recent research on independence concepts in the context of credal sets. Even defining independence and conditional independence appears to be challenging. The literature is rich in results, but many questions remain open. Section 3.1 reviews the main definitions in the literature, leaving concerns about conditioning, null events, and countable additivity to Section 3.2. Independence concepts are briefly compared in Section 3.3, and Sections 3.4 and 3.5 focus on *strong* independence.

3.1 Confirmational, strict, strong, epistemic independence... and others

A direct generalization of stochastic independence to the context of credal sets is *strict* independence[1]: X and Y are *strictly independent* if, for all $P \in K$,

$$P(X \in A | Y \in B) = P(X \in A) \quad \text{whenever } P(Y \in B) > 0.$$

This straightforward definition violates convexity. Consider an example of Jeffrey's [42, Sect. IV.B]. Take binary variables X and Y. Suppose $K(X, Y)$ is the convex hull of two distributions P_1 and P_2 such that $P_1(X = 0) = P_1(Y = 0) = 1/3$ and $P_2(X = 0) = P_2(Y = 0) = 2/3$. Suppose X and Y are strictly independent; hence X and Y are stochastically independent with respect to

[1]This term is due to Teddy Seidenfeld.

P_1 and P_2. Now take the distribution $P_{1/2} = P_1/2 + P_2/2$; we have $P_{1/2} \in K(X, Y)$ by convexity. However, X and Y are not stochastically independent with respect to $P_{1/2}$, as

$$
\begin{aligned}
P_{1/2}(X = 0, Y = 0) &= P_1(X = 0)P_1(Y = 0)/2 + P_2(X = 0)P_1(Y = 0)/2 \\
&= 5/18 \neq 1/4 = P_{1/2}(X = 0)P_{1/2}(Y = 0).
\end{aligned}
$$

The clash between strict independence and convexity is already explicit in Levi's pioneering work on convex Bayesianism [46, Chap. 10]. Levi defines Y to be *confirmationally irrelevant* to X if

$$
K(X|Y \in B) = K(X) \quad \text{for nonempty } \{Y \in B\}, \tag{1}
$$

and notes that confirmational irrelevance is not the same as strict independence. Levi also states that expectations are not affected by failure of convexity in strict independence; his implicit message is that, as expectations are only affected by the convex hull of a credal set, one is allowed to take the convex hull whenever necessary.

This suggests the following somewhat convoluted definition: X and Y are *strongly independent* when $K(X, Y)$ is the convex hull of a credal set that satisfies strict independence. A more concise description of strong independence is possible, if all credal sets are assumed closed. Using lower/upper expectations, we have: X and Y are strongly independent iff for any bounded function $f(X, Y)$,

$$
\underline{E}[f(X, Y)] = \min\left(E_P[f(X, Y)] \;:\; P = P_X P_Y\right). \tag{2}
$$

In 1982 Walley and Fine called this expression, with $f(X, Y)$ restricted to indicator functions only, an *independent product* [65]. Independent products (this restricted form of strong independence) are adopted by Weichselberger in his theory of interval probability (under the name *mutual independence*) [66, 67].

Walley later used the term *type-1 product* to refer to Expression (2) in his highly influential book [64]. Walley reserves the term *type-2 product* to the situation where all marginals are equal.

In his book Walley also proposes a different concept, *epistemic irrelevance*: Y is epistemically irrelevant to X if for any bounded function $f(X)$,

$$
\underline{E}[f(X)|Y \in B] = \underline{E}[f(X)] \quad \text{for nonempty } \{Y \in B\}.
$$

If credal sets are closed and convex, then epistemic irrelevance is identical to Levi's confirmational irrelevance.[2] Epistemic irrelevance is what Smith refers to as independence in his pioneering work on medial odds [60]. And when restricted to events epistemic irrelevance is similar to what Weichselberger calls *canonical* independence of B to A ($\overline{E}[A|B] = \overline{E}[A]$ and $\underline{E}[A|B] = \underline{E}[A]$) [66, 67].

Epistemic irrelevance is not symmetric: Y may be epistemically irrelevant to X while X is not epistemically irrelevant to Y [13, 64]. Walley's clever proposal is to create a symmetric concept out of epistemic irrelevance, following Keynes' approach to independence [37]: X and Y are

[2] Definitions involving equalities among lower expectations tend to produce closed convex sets. Walley's theory takes lower expectations as primitives; it does not matter whether one builds a lower expectation from a closed convex credal set or not, in the end the definitions and calculations work as if closure and convex hulls were taken at each step (for instance, see examples in two papers co-authored by Walley [7, 13]).

epistemically independent if Y is epistemically irrelevant to X and X is epistemically irrelevant to Y.

Walley's book is contemporary with Kuznetsov's book on interval probabilities and interval expectations [40]. Kuznetsov proposes yet another concept of independence with a nice interpretation in terms of interval arithmetic.[3] Denote by $EI[X]$ the interval $[\underline{E}[X], \overline{E}[X]]$; then: X and Y are *Kuznetsov independent* if, for any bounded functions $f(X)$ and $g(Y)$,

$$EI[f(X)g(Y)] = EI[f(X)] \times EI[g(Y)],$$

where $\times$ is interval multiplication.[4] An equivalent formulation is: for any bounded functions $f(X)$ and $g(Y)$,

$$\underline{E}[f(X)g(Y)] = \inf \left(E_P[f(X)g(Y)] \quad : \quad P = P_X P_Y \right). \tag{3}$$

Expressions (2) and (3) are quite similar, particularly if the credal sets are closed. The expressions are however not equivalent: this can be seem easily when $K(X)$ and $K(Y)$ are vacuous [41], and with more difficulty when these marginal credal sets are not vacuous [11].

Many variations on the previous definitions are possible. For instance, one might say that Y is irrelevant to X if for any bounded function $f(X)$,

$$\underline{E}[f(X)|Y \in B'] = \underline{E}[f(X)|Y = B''] \qquad \text{for nonempt } \{Y \in B'\}, \{Y \in B''\}.$$

Such a definition is *not* equivalent to epistemic irrelevance, and it seems too weak. For instance we can have vacuous credal sets $K(X|Y = y)$ for every y, and still $K(X)$ can be a non-vacuous credal set (even a singleton[5]). It seems bizarre to say that Y is then irrelevant to X.

Several other variations appeared in the literature between 1990 and 2000, and terminology became somewhat confusing. Part of this activity was inherited from the research in Dempster-Shafer and possibility theories, where concepts of conditioning and independence were intensely debated during that decade.[6]

In 1995 de Campos and Moral tried to organized the field into a number of distinct concepts of independence [19]. Their *type-2* independence is strong independence as defined previously (that is, each vertex of a convex $K(X, Y)$ satisfies stochastic independence). Their *type-3* independence obtains when $K(X, Y)$ is the convex hull of *all* product distributions $P_X P_Y$, where $P_X \in K(X)$ and $P_Y \in K(Y)$. To minimize the number of different concepts, it makes sense to consider this type-3 independence as a special case of strong independence; indeed, often one sees the term *strong extension* to refer to the largest credal set that satisfies strong independence with given marginal sets $K(X)$ and $K(Y)$. Finally, de Campos and Moral have a variation on confirmational irrelevance: Y is *type-5* irrelevant to X if

$$R(X|Y \in B) = K(X) \qquad \text{whenever } \overline{P}(Y \in B) > 0,$$

[3] Actually, it seems that Kuznetsov adopts strong independence as the main concept of independence and proposes his new concept as a secondary idea [41].

[4] If $a = [\underline{a}, \overline{a}]$ and $b = [\underline{b}, \overline{b}]$, $a \times b = [\underline{ab}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{ab}]$.

[5] That is, we may have an extreme case of *dilation* [54], where $K(X)$ contains a single distribution but $K(X|Y = y)$ is vacuous for each y. Dilation often obtains when epistemic irrelevance is adopted [13].

[6] Concepts of independence were created for many different definitions of conditioning and of product distributions. For instance, one may take Dempster conditioning (indicated by a subscript D in the conditioning bar) and require $\overline{P}(X|_D Y) = \overline{P}(X, Y) / \overline{P}(Y) = \overline{P}(X)$ whenever $\overline{P}(X) > 0$; that is, $\overline{P}(X, Y) = \overline{P}(X)\, \overline{P}(Y)$. This is related (mathematically at least) to Shafer's concept of *cognitive independence* [59, 70, 71].

where $R(X|Y \in B)$ denotes the set

$$\{P(\cdot|Y \in B) : P \in K(X,Y)\,; P(Y \in B) > 0\}. \tag{4}$$

The set R defined in Expression (4) is related to what Walley calls *regular extension* [64],[7] and is often used as a definition of conditioning [66]. The following interesting example is given by de Campos and Moral [19]: suppose X and Y are binary, and $K(X,Y)$ is the convex hull of two distributions P_1 and P_2 such that $P_1(X=0, Y=0) = P_2(X=1, Y=1) = 1$. Then strong independence obtains, but neither Y is type-5 irrelevant to X, nor X is type-5 irrelevant to Y.

In 1999 Couso et al presented an influential review of independence concepts [6, 7]. They used yet another terminology: their *independence in the selection* is strong independence as defined previously, and their *strong independence* is de Campos and Moral's type-3 independence. Finally, their *repetition independence* refers to Walley's *type-2 product*.[8]

3.2 Conditional independence, null events, and full conditional measures

In this section we discuss conditional independence and null events, two issues that have grown in importance through the years, and have become a laboratory for all sorts of foundational problems regarding credal sets.

3.2.1 Conditional independence

Any concept of independence can be modified to express *conditional independence*, simply by conditioning on every value of some variable. For instance, *conditional epistemic irrelevance* of Y to X given Z obtains when, for all bounded functions $f(X)$,

$$\underline{E}[f(X)|Y \in B, Z = z] = \underline{E}[f(X)|Z = z] \qquad \text{for nonempty } \{Y \in B, Z = z\}.$$

Conditional epistemic independence is then the "symmetrized" concept. Likewise, *conditional Kuznetsov independence* of X and Y given Z obtains when for all bounded functions $f(X)$, $g(Y)$,

$$EI[f(X)g(Y)|Z = z] = EI[f(X)|Z = z] \times EI[g(Y)|Z = z] \qquad \text{for nonempty } \{Z = z\}.$$

And we define *conditional strict independence* by imposing elementwise conditional stochastic independence[9]: every probability distribution must satisfy

$$P(X \in A, Y \in B|Z = z) = P(X \in A|Z = z)\, P(Y \in B|Z = z) \qquad \text{whenever } P(Z = z) > 0.$$

[7] Regular extension is however different as it defines conditioning even if $\overline{P}(Y \in B) = 0$. In regular conditioning, one may produce different credal sets $R(X|A \cap B)$ by conditioning first on A and then on B, and by conditioning directly on $A \cap B$ (thanks to Teddy Seidenfeld for pointing this out to me).

[8] Couso et al also discuss two other situations that are not directly relevant to the concerns of the present paper: (1) the set $K(X,Y)$ is the largest set with given marginals $K(X)$ and $K(Y)$ and no further constraints; (2) the set $K(X,Y)$ is specified through a belief function where the joint mass assignment satisfies stochastic independence. Recall that a belief function can always be expressed as a probability measure (the mass assignment) over the subsets of the possibility space, plus a multi-valued mapping. Couso et al call this latter concept *random set independence* (a similar concept had been called a *belief function product* by Walley and Fine [65]).

[9] Moral and Cano describe three variants on conditional strict independence [47], basically by considering ways to extend given marginal and conditional credal sets on X and Y given Z; these alternative concepts are perhaps better understood as forms of *extension* given marginal and conditional credal sets.

We adopt the same scheme for *conditional strong independence*; that is, strong independence conditional on every value z of Z.

3.2.2 Full conditional measures

This review has so far presented concepts of irrelevance/independence with little care concerning *null* events; that is, events of zero probability. The usual attitude in probability theory is to ignore null events as they almost surely do not obtain. But in the context of credal sets one cannot ignore an event with zero *lower* probability but nonzero *upper* probability. For instance, take confirmational irrelevance. Suppose we refuse to define a conditional credal set $K(X|Y=y)$ whenever $\underline{P}(Y=y)=0$, as for instance proposed by Giron and Rios [29]. Then it seems advisable to modify the clause in Expression (1): Y is "irrelevant" to X if

$$K(X|Y \in B) = K(X) \qquad \text{whenever } \underline{P}(Y \in B) > 0.$$

Now suppose $K(Y)$ is such that for each event $\{Y \in B\}$ there is a distribution P in $K(Y)$ where $P(Y \in B) = 0$. Then Y is irrelevant to any other variable!

So, we must somehow allow conditioning on events that *may* be null. In fact, there exists a perfectly reasonable way to condition on null events. The solution is to resort to *full conditional measures*, where one takes conditional probability as a primitive concept [24, 21]. A full conditional measure is a set function $P(\cdot|\cdot)$ on $\mathcal{E} \times \mathcal{E}\backslash\emptyset$, where $\mathcal{E}$ is an algebra of events, such that for events A, B in $\mathcal{E}$ and C in $\mathcal{E}\backslash\emptyset$ we have

$$P(A|C) \geq 0; \quad P(\Omega|C) = 1; \quad P(A \cup B|C) = P(A|C) + P(B|C) \text{ if } A \cap B = \emptyset; \text{ and}$$

$$P(A \cap B|C) = P(A|B \cap C)\, P(B|C) \text{ if } B \cap C \neq \emptyset.$$

Unsurprisingly, both Levi's and Walley's theories adopt full conditional measures: confirmational/epistemic irrelevance/independence are defined without any clause concerning lower or upper probability zero.

As a digression, note that once one adopts full conditional measures, it seems advisable to base any concept of independence on conditioning instead of on product measures. For if one requires only the product $P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$ for independence, then it may happen that X and Y are declared independent while $P(X \in A|Y \in B) \neq P(X \in A)$ for some A, B.[10]

Full conditional measures are extremely elegant and have been advocated for a variety of reasons [5, 24, 30, 55], but they do have an associated cost. The usual assumption of countable additivity ties conditioning (in general spaces) to Radon-Nikodym derivatives; but these derivatives may fail to be full conditional measures [57]. Characterizing the situations where full conditional measures exist under countable additivity seems to be a hard (and mostly open) problem [1, 39]. Even though some authors have preferred to ignore these existence problems [8], it seems that in general one is forced into finite additivity when full conditional measures are

[10] A theory with full conditional measures and a "product" definition of independence has been used, for instance, by Cowell et al [8].

adopted [55]. This is indeed the path taken by Levi and Walley (the latter imposes additional conditions of conglomerability on lower expectations).

In short, confirmational/epistemic irrelevance/independence seem to require a combination of full conditional measures and finite addivity, perhaps with some assumption of conglomerability.

Strict and strong independence can also be adapted to the the specificities of full conditional measures. Thus we are led to some new ramifications in our tree of concepts:

- *Full* strict irrelevance of Y to X is elementwise epistemic irrelevance of Y to X; that is, epistemic irrelevance of Y to X for each $P \in K$. *Full* strict independence of X and Y now becomes epistemic irrelevance of Y to X and epistemic irrelevance of X to Y.
- And likewise for *full* strong irrelevance and *full* strong independence.

At this point the reader may despair as it seems we are quickly exahusting the possible names for concepts of irrelevance/independence. However, difficulties with null events have not been exhausted yet. Several authors have noted that epistemic independence is a relatively weak concept *even when applied to a single full conditional measure*, and for this reason various modifications to epistemic irrelevance (for a single measure) have been proposed [12, 31, 32, 63]. It does not seem that such modified concepts of irrelevance/independence have been applied to credal sets yet, but they offer important paths to follow.

3.2.3 Avoiding full conditional measures

One might try to avoid the complexities of full conditional measures by adopting conditional strict/strong/Kuznetsov independence, as these concepts can be expressed without resort to conditioning. Another possibility is to take de Campos and Moral's type-5 irrelevance as the basis for conditional irrelevance: Y is *conditionally regularly irrelevant* to X given Z if

$$R(X|Y \in B, Z = z) = R(X|Z = z) \quad \text{whenever } \overline{P}(Y \in B, Z = z) > 0.$$

Conditional regular independence is then the "symmetrized" concept. It does not seem that conditional regular irrelevance/independence have been explored in the literature.

3.3 Comparing concepts

The previous sections listed about a dozen concepts of (conditional) irrelevance/independence. They are not identical: for instance, epistemic independence implies Kuznetsov independence which in turn implies strong independence, but the reverse implications are false in general [11]. We note:

1. Confirmational/epistemic irrelevance/independence seem to be the most intuitive concepts once convexity is required (an alternative is to use de Campos and Moral's type-5 irrelevance so as to stay away from full conditional measures).

2. Kuznetsov independence appears as a natural generalization of the product form of stochastic independence, but it is hard to imagine a behavioral justification for it.

3. Strict independence is easy to state and obviously close to stochastic independence; however it violates convexity and the usual behavioral interpretations through partial preferences and one-sided betting.

4. Finally, strong independence has some of the appeal of strict independence and satisfies convexity; however it is the most difficult to justify: Why should one start with a nonconvex concept and take the convex hull of the resulting set of measures? We examine this question in Sections 3.4 and 3.5.

It is instructive to compare some of the properties displayed by these concepts. For instance, we might wonder which concepts lead to laws of large numbers. This particular question has been mostly settled by de Cooman et al [20], who have produced quite general laws of large numbers for assumptions of epistemic irrelevance. As their results are derived from very weak assumptions, they are valid for most, if not all, concepts of irrelevance/independence one might contemplate. So we must look for other potential differences among concepts. One possibility then is to investigate the *graphoid* properties, as these have been advocated as axioms for independence concepts [14, 15, 49]. Graphoid properties are stated for a ternary relation $(X \perp\!\!\!\perp Y \,|\, Z)$, which we should read as "independence of X and Y given Z". We have [28]:[11]

Symmetry: $(X \perp\!\!\!\perp Y \,|\, Z) \Rightarrow (Y \perp\!\!\!\perp X \,|\, Z)$

Decomposition: $(X \perp\!\!\!\perp (W, Y) \,|\, Z) \Rightarrow (X \perp\!\!\!\perp Y \,|\, Z)$

Weak union: $(X \perp\!\!\!\perp (W, Y) \,|\, Z) \Rightarrow (X \perp\!\!\!\perp W \,|\, (Y, Z))$

Contraction: $(X \perp\!\!\!\perp Y \,|\, Z)$ & $(X \perp\!\!\!\perp W \,|\, (Y, Z)) \Rightarrow (X \perp\!\!\!\perp (W, Y) \,|\, Z)$

Several results on the relationship between concepts of independence and graphoid properties can be found in the literature. Conditional strict and strong independence satisfy the graphoid properties [9]. Conditional confirmational/epistemic independence and conditional Kuznetsov independence fail contraction [13] even when all probabilities are positive.[12]

It seems that no study of graphoid properties of conditional regular independence is available in the literature. Under the assumption that all credal sets are closed and convex, and additionally that conglomerability holds, conditional regular independence satisfies symmetry, decomposition and weak union.[13] Clearly conditional regular independence fails contraction, as epistemic independence fails contraction even in the absence of zero probabilities.

[11]These properties are often called *semi-graphoid* properties [49].

[12]Moral has investigated a version of epistemic irrelevance for sets of desirable gambles (related but not equivalent to credal sets), satisfying a different set of graphoid properties [48]. Also the extensions to strict/strong independence proposed by by Moral and Cano [47] fail different sets of graphoid properties.

[13]Symmetry holds by definition; decomposition and weak union are proved as for epistemic independence using conglomerability [13], because the lower expectations completely characterize the credal sets that are by assumption closed and convex. Conglomerability is necessary: Take (W, Y) as the product measure of a fair coin and a finitely additive measure that imposes uniform probability over the integers [35]; then each (W, Y) has probability zero and (W, Y) is conditionally regularly irrelevant to any X; but W may be conditionally regularly relevant to X.

As a digression, note that failure of the contraction property greatly affects the theory of statistical models such as Markov chains and Bayesian networks [49]. Consider a simple Markov chain $W \to X \to Y \to Z$. The usual theory prescribes that any variable is stochastically independent of its predecessors given its immediate predecessor. From this assumption, other independences can be derived; for instance, that (W, X) and Z are conditionally stochastically independent given Y. But if we replace stochastic independence by epistemic independence, then it is possible to construct a Markov chain where (W, X) and Z are *not* conditionally epistemically independent given Y [10, Example 1].

When events of zero probability are present, the adoption of full conditional measures leads to additional complications. Then confirmational/epistemic independence may even fail decomposition and weak union [13]. It should be noted that when zero probabilities are present, confirmational/epistemic may fail weak union even for a single full conditional measure [12, 63]. Thus, failure of weak union can be observed in full strict/strong irrelevance/independence in the presence of events of zero probability. Besides, the known proofs of several graphoid properties of epistemic irrelevance/independence rely on conglomerability properties that may fail when countable additivity is discarded [12]; which graphoid properties hold under finite additivity is an open question. There are also modified concepts of independence that save weak union but fail contraction [12, 32]), but the study of such concepts in the context of credal sets is an open question.

There might be other valuable ways to compare concepts of irrelevance/independence. For instance, we might look at computational properties: what is the complexity of inference under each one of the concepts. However there are very few results in the literature: only strong independence has received attention [18], and some algorithms have been produced for epistemic independence [16]. The verdict on this matter is yet to be decided.

3.4 Justifying strong independence

The previous subsections attempted to present, in a somewhat organized form, the current landscape concerning concepts of independence for credal sets. An important question deserves further study: how can we keep convexity and stay close to stochastic independence? One solution is to find ways to justify strong independence so that it can be adopted whenever necessary.

One might try to justify strong independence using what Walley calls the *sensitivity* interpretation of credal sets [64]. Suppose several experts agree on stochastic independence among variables but disagree on specific probability values. The experts then adopt a credal set containing distributions that factorize appropriately, plus the convex combinations of these distributions, arguing that such convex combinations do not affect their collective preferences. However this argument fails because as shown in Section 4, these convex combinations do affect decision making for non-binary preferences.

Besides, one might like to state independence directly, using just partial preferences, lower expectations, credal sets; that is, without resorting to stochastic independence. A proposal to this effect was independently derived around 2000 by Cozman [9] and by Moral and Cano [47]. To understand the proposal, consider the following example. Take two binary variables X and Y so that $P(X = 0) \in [2/5, 1/2]$ and $P(Y = 0) \in [2/5, 1/2]$. The largest credal set $K(X, Y)$ satisfying epistemic independence has six vertices [64, Sect. 9.3.4]:

$$[1/4, 1/4, 1/4, 1/4], [4/25, 6/25, 6/25, 9/25], [1/5, 1/5, 3/10, 3/10], [1/5, 3/10, 1/5, 3/10],$$

$$[2/9, 2/9, 2/9, 1/3], [2/11, 3/11, 3/11, 3/11],$$

with vectors denoting $[P(X=0, Y=0), P(X=0, Y=1), P(X=1, Y=0), P(X=1, Y=1)]$. Suppose we learn that $P(Y=0) = 4/9$. One option is to "intersect" $K(X, Y)$ with the constraint $P(Y=0) = 4/9$; that is, to form a new credal set

$$K'(X, Y) = K(X, Y) \cap \{P : P(Y=0) = 4/9\}. \quad (5)$$

However, X and Y are not epistemically independent with respect to $K'(X, Y)$: the distribution $[2/9, 2/9, 2/9, 1/3]$ belongs to $K'(X, Y)$, and for this distribution $P(Y=0|X=1) = 2/5$, so $\underline{P}(Y=0|X=1) \leq 2/5 < 4/9 = \underline{P}(Y=0)$ with respect to $K'(X, Y)$. Epistemic independence is not preserved through Expression (5).

The situation just outlined reminds one of Jeffrey's rule [34]. In Jeffrey's rule we start with a distribution $P_{X,Y}$, we change P_Y but we keep the conditional distribution $P_X(\cdot|Y)$ intact. Moreover, if we have a single distribution P, then X and Y are stochastically independent iff P_X does not change through Jeffrey's rule for any change in P_Y and vice-versa [22, Theorem 3.3]. A similar result holds for credal sets: suppose a new $K(Y)$ (or $K(X)$) is given and a modification to $K(X, Y)$ is made by pointwise application Jeffrey's rule; then if X and Y are still epistemically independent after any such change, X and Y are fully strongly independent [10].

Moral and Cano follow the same idea but use a better strategy that avoids any potential controversies on how to apply Jeffrey's rule on credal sets [47]. They note that to produce strong independence it is not necessary to employ arbitrary modifications to the marginal credal sets. Their approach uses two definitions:

- Assessment $E[f(X)] \geq \alpha$ [resp. $E[g(Y)] \geq \beta$] for bounded $f(X)$ [resp. $g(Y)$] is *compatible* with marginal credal set $K(X)$ [resp. $K(Y)$] if there is a distribution P in $K(X)$ [resp. $K(Y)$] that satisfies $E_P[f(X)] \geq \alpha$ [resp. $E_P[g(Y)] \geq \beta$].
- A credal set $K(X, Y)$ is *combined* with assessment $\underline{E}[f(X)] \geq \alpha$ [resp. $\underline{E}[g(Y)] \geq \beta$] for bounded function $f(X)$ [resp. $g(Y)$] by eliminating all distributions in $K(X, Y)$ that do not satisfy the assessment.

The following theorem generalizes somewhat the basic result by Moral and Cano:

Theorem 1 *Variables X and Y are conditionally fully strongly independent given Z iff they are conditionally epistemically independent given Z after $K(X, Y|Z=z)$ is combined with an arbitrary collection of assessments that are compatible either with $K(X)$ or with $K(Y)$ for any value z of Z.*

The proof of this theorem is obtained by following all steps in Moral and Cano's proof of their Theorem 2. Changes are formally small but worthy mentioning. First, note that Moral and Cano formulate their theorem so as to generate the *largest* credal sets satisfying strong independence *and* constraints only in marginal credal sets $K(X)$ and $K(Y)$ (more precisely: they obtain "strong independence" in their sense; that is, they obtain strong extension). There is no reason to restrict attention to this situation. Second, it should be noted that conditional *full* strong independence is obtained from epistemic independence; to really produce conditional strong independence the theorem must replace epistemic independence by a concept of independence that ignores conditioning on events of zero probability. That is:

Theorem 2 *Variables X and Y are conditionally strongly independent given Z iff they are conditionally regularly independent given Z after $K(X, Y|Z = z)$ is combined with an arbitrary collection of assessments that are compatible either with $K(X)$ or with $K(Y)$ for any value z of Z.*

While Theorem 2 is a viable strategy to justify strong independence, this approach does have a weakness. For suppose we have a credal set $K(X, Y)$ that satisfies several assessments and epistemic independence of X and Y. Now we learn about a collection of new assessments on X and on Y. Shouldn't we just form a new credal set satisfying *all* assessments, new and old, *and* epistemic independence of X and Y? Theorems 1 and 2 assume that assessments are combined without concern to the very judgement of independence we wish to preserve.

3.5 Strong independence through partial exchangeability

An alternative way to justify (at least some varieties of) strong independence is to employ *exchangeability*. In the remainder of this section we investigate this idea; it does not seem that it has been explored yet in the literature. To simplify the discussion, countable additivity is assumed. but similar conclusions hold if countable additivity is dropped.

To start, consider a vector of m binary variables $\mathbf{X} = [X_1, \ldots, X_m]$. Denote by π_m a permutation of integers $\{1, \ldots, m\}$, and by $\pi_m(i)$ the ith integer in the permutation. Denote by $\{\mathbf{X} = \mathbf{x}\}$ the event $\cap_{i=1}^m \{X_i = x_i\}$, and by $\{\pi_m \mathbf{X} = \mathbf{x}\}$ the event $\cap_{i=1}^m \{X_{\pi_m(i)} = x_i\}$.

Variables $X_1, \ldots, X_m$ are *exchangeable* when [64, Chap. 9]:[14]

$$\underline{E}[\{\mathbf{X} = \mathbf{x}\} - \{\pi_m \mathbf{X} = \mathbf{x}\}] = 0 \quad \text{for any permutation } \pi_m. \tag{6}$$

That is, the order of variables does not matter: trading $\{\mathbf{X} = \mathbf{x}\}$ for $\{\pi_m \mathbf{X} = \mathbf{x}\}$ does not seem advantageous in the one-sided betting interpretation of $\underline{E}$.

As noted by Walley, we have

$$\begin{aligned} 0 = \underline{E}[\{\mathbf{X} = \mathbf{x}\} - \{\pi_m \mathbf{X} = \mathbf{x}\}] &\leq \overline{E}[\{\mathbf{X} = \mathbf{x}\} - \{\pi_m \mathbf{X} = \mathbf{x}\}] \\ &= -\underline{E}[\{\pi_m \mathbf{X} = \mathbf{x}\} - \{\mathbf{X} = \mathbf{x}\}] = 0. \end{aligned}$$

Hence every distribution in the credal set $K(X_1, \ldots, X_n)$ satisfies

$$P(\mathbf{X} = \mathbf{x}) = P(\pi_m \mathbf{X} = \mathbf{x}) \quad \text{for any permutation } \pi_m.$$

In words: Expression (6) implies *elementwise* exchangeability in the usual de Finetti's sense [21].

[14]Following de Finetti, we do not differentiate between an event and its indicator function.

Focus on a distribution P satisfying exchangeability for a moment. Suppose we have m exchangeable variables $X_1, \ldots, X_m$ and we examine the first n variables $X_1, \ldots, X_n$ for $n \leq m$. These n variables are also exchangeable, and [33]

$$P(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0) = \sum_{r=k}^{m-n+k} \frac{\binom{m-n}{r-k}}{\binom{m}{r}} P\left(\sum_{i=1}^{m} X_i = r\right).$$

Now if an infinitely long sequence of exchangeable variables is contemplated ($m \to \infty$), de Finetti's representation theorem yields [33, 52]:

$$P(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0) = \int_0^1 \theta^k (1-\theta)^{n-k} dF(\theta).$$

Here θ is the probability of $\{X_1 = 1\}$, and the distribution function $F(\theta)$ acts as a "prior" over θ.

In our setting, we have a set $K(\theta)$ of distributions $F(\theta)$, one for each distribution $P \in K(X_1, \ldots, X_n)$. Moreover, exchangeability is a "convex" concept in the sense that if two distributions satisfy exchangeability of $X_1, \ldots, X_n$, then any convex combination of these distributions also satisfy exchangeability. Consequently, $K(\theta)$ is a convex credal set.

Thus strong independence of $X_1, \ldots, X_n$ obtains when each vertex of $K(\theta)$ assigns probability 1 to a particular value of θ. As every vertex of $K(X_1, \ldots, X_n)$ is a product measure with identical marginals, we have obtained what Walley calls a type-2 product (Section 3.1). In short, we have easily reduced a judgement of strong independence to a judgement of exchangeability plus a condition on $K(\theta)$.[15] By using suitable versions of de Finetti's representation theorem [52], one can extend the previous argument to non-binary variables. One can also modify the argument to obtain "convexified" sets of Markov chains, using judgements of *Markov exchangeability* [23, 72]. The overall idea leads to forms of strong independence that are difficult to justify otherwise. Besides, exchangeability leads directly to the *product form* of strong independence, thus avoiding conditional distributions and issues of conditioning on events of zero probability.

With some additional imagination, the previous argument can be modified to obtain general strong independence. Consider two binary variables X_1 and Y_1 such that $K(X_1)$ and $K(Y_1)$ are different. Imagine that we observe X_1 and Y_1 repeatedly, creating a sequence of exchangeable variables $\mathbf{X} = [X_1, \ldots, X_m]$ with marginals $K(X_1)$, and a sequence of exchangeable variables $\mathbf{Y} = [Y_1, \ldots, Y_m]$ with marginals $K(Y_1)$. What else could we impose on these sequences? Consider the following judgement of *partial exchangeability* [4, 21, 45]:

$$\underline{E}[\{\mathbf{X} = \mathbf{x}\}\{\mathbf{Y} = \mathbf{y}\} - \{\pi'_m \mathbf{X} = \mathbf{x}\}\{\pi''_m \mathbf{Y} = \mathbf{y}\}] = 0 \quad \text{for any permutations } \pi'_m, \pi''_m.$$

Then, as $m \to \infty$, we have [4, Theorem 4.13] a convex set of distribution functions $F(\theta, \vartheta)$ such that for every P,

$$P(X_1 = x_1, Y_1 = y_1) = \int_{[0,1]^2} \theta^{x_1}(1-\theta)^{1-x_1} \vartheta^{x_2} (1-\vartheta)^{1-x_2} dF(\theta, \vartheta),$$

[15] One may even produce subsets of the strong extension by suitably choosing $K(\theta)$.

If each vertex of this set assigns probability one to a pair (θ, ϑ), we obtain strong independence of X_1 and Y_1.

This argument for strong independence of X_1 and Y_1 may not be as appealing as the previous one for type-2 products, as it requires additional exchangeable sequences of observations. However the central idea is rather simple: X_1 and Y_1 are strongly independence if, whatever we do to an exchangeable sequence of observations of X_1, probabilities for exchangeable observations of Y_1 are not affected.

4 Set-based Bayesianism and non-binary preferences

In the previous section we examined recent work on the connection between credal sets, independence concepts, and convexity. It is time to examine a suggestion made by Kyburg and Pittarelli in 1992 [44]: that we should drop convexity altogether and adopt strict independence. The difficulty with this prescription is that general credal sets do not seem to enjoy a foundation on partial preferences/one-sided betting.[16] Even though Kyburg and Pittarelli did not solve this problem completely, they did touch on a few critical elements of the solution.

The main insight here lies on the computation of E-admissible acts *amongst several acts.* Kyburg and Pittarelli discuss the following example [42, Sect. IVD]. Consider a possibility space with three states $\{s_1, s_2, s_3\}$. Suppose a credal set contains two distributions P_1 and P_2 such that

$$P_1(s_1) = 1/8, \quad P_1(s_2) = 3/4, \quad P_1(s_3) = 1/8,$$

$$P_2(s_1) = 3/4, \quad P_2(s_2) = 1/8, \quad P_2(s_3) = 1/8,$$

and consider the selection of an E-admissible decision amongst acts $\{a_1, a_2, a_3\}$, with decision matrix

	s_1	s_2	s_3
a_1	3	3	4
a_2	2.5	3.5	5
a_3	1	5	4.

Now with respect to P_1 and P_2, a_1 and a_3 are E-admissible but a_2 is not; with respect to the convex hull of $\{P_1, P_2\}$, all acts are E-admissible. That is, *there is a difference between a set of distributions and its convex hull when one chooses amongst several acts.*

The amusing irony here is that E-admissibility is a concept advanced by Isaac Levi, the main proponent of convex Bayesianism; Kyburg and Pittarelli basically take one of Levi's proposals against the other.

One might then ask: Can we axiomatize preferences amongst sets of acts, so as to obtain general credal sets? This is in fact the path followed by Seidenfeld et al in important recent

[16] Another potential difficulty with general credal sets is the computational cost of dealing with nonconvex sets. However, the computational experience of the last decade has shown that once independence relations are used, the computational benefits of convexity are rather diminished (for instance, when all vertices of a credal set factorize, the computational cost is dominated by factorization and convexity is not important [2, 25, 17]). Again, independence appears as a central issue, albeit indirectly.

work [56]. In a sequence of papers [36, 51, 56], Seidenfeld and colleagues have axiomatized general credal sets using non-binary preferences.

A quick summary of Seidenfeld et al's theory is as follows [56]. Consider a closed set $\mathcal{A}$ of acts. For any subset D of $\mathcal{A}$, a *rejection function* R identifies the subset of D containing all acts that are *not* admissible within D. This subset is denoted by $R(D)$. Now a rejection function is *coherent* if $R(D)$ is always the set of non-E-admissible acts within D, with respect to a set of pairs of utility function/probability distributions. Seidenfeld et al impose a set of axioms on rejection functions and prove that any rejection function that satisfies their axioms iff it is coherent [56, Theorems 3 and 4]. To understand the kinds of axioms that are proposed, consider the first axiom: An inadmissible act cannot become admissible (a) when new acts are added to the set of acts; and (b) when inadmissible acts are deleted from the set of acts.[17] Another axiom states that if $d \in R(\text{convexhull}(D))$, then $d \in R(D)$. That is, an inadmissible act amongst a set of mixed acts cannot become admissible just by removing the mixtures. Three additional axioms are imposed, paralleling the sure-thing, Archimedean and dominance axioms typically adopted in standard decision theory [26]. The axioms are not simple to state, and perhaps an exercise for the future is to trim down Seidenfeld et al's theory to a small set of intuitive axioms. In any case, their approach is entirely successful, as their axioms do yield general sets of probability distributions.

It is convenient to explore these ideas again in the context of independence. Suppose we wish to determine whether events A and B are strictly independent.[18] Construct five acts $a_0, \ldots, a_4$:

	AB	A^cB	AB^c	A^cB^c
a_0	0	0	0	0
a_1	$1-\alpha$	$-\alpha$	0	0
a_2	$-(1-\alpha)$	α	0	0

	AB	A^cB	AB^c	A^cB^c
a_3	0	0	$1-\beta$	$-\beta$
a_4	0	0	$-(1-\beta)$	β

Now these five acts serve as a test for strict independence: if we observe that for every $\alpha, \beta \in (0, 1)$ such that $\alpha \neq \beta$ we have some act rejected, we can conclude that A and B are strictly independent.

Proof. Suppose otherwise; that is, suppose $\forall \alpha, \beta \in (0, 1) : \alpha \neq \beta \rightarrow R(a_0, \ldots, a_4) \neq \emptyset$ but A and B are not strictly independent. Then there is P such that $P(A|B) \neq P(A|B^c)$. Take $\alpha = P(A|B)$ and $\beta = P(A|B^c)$ and note that $E_P[a_i] = 0$ for $a_0, \ldots, a_4$; so all acts are E-admissible and $R(a_0, \ldots, a_4) = \emptyset$, a contradiction. Note that when A and B are strictly independent, then for each P we have $E_P[a_0] = 0$, $E_P[a_1] = (P(A) - \alpha)P(B)$, $E_P[a_2] = (\alpha - P(A))P(B)$, $E_P[a_3] = (P(A) - \beta)P(B)$, $E_P[a_4] = (\beta - P(A))P(B)$, so if $\alpha \neq \beta$ we indeed have $R(a_0, \ldots, a_4) \neq \emptyset$. □

At this point one might consider revisiting confirmational irrelevance/independence in the context of general credal sets. However the resulting concepts do not seem very interesting: conditional confirmational independence then fails *all* graphoid properties except symmetry even when probabilities are all positive (contraction already fails for convex credal sets; failure of decomposition and weak union is depicted in Table 1).

[17] More precisely: If $D_2 \subseteq R(D_1)$, then: (a) if $D_1 \subseteq D_3$, then $D_2 \subseteq R(D_3)$; and (b) if $D_3 \subseteq D_2$, then $D_2 \backslash D_3 \subseteq R(\text{closure}(D_1 \backslash D_3))$.

[18] This example is based on a very compact, but perhaps a little more difficult to grasp, example produced by Teddy Seidenfeld. In his example strict independence is generated without any assumption on the marginal probabilities of A and B with only four acts that have a very intuitive meaning. His derivation seems not to be published at this point.

	1	2	3	4
$P(X=0\|W=0,Y=0)$, $P(W=0,Y=0)$	α, 1/4	α, 1/4	α, 1/4	β, $\frac{\beta/2}{\alpha+\beta}$
$P(X=0\|W=0,Y=1)$, $P(W=0,Y=1)$	α, 1/4	α, 1/4	α, 1/4	β, $\frac{\alpha/2}{\alpha+\beta}$
$P(X=0\|W=1,Y=0)$, $P(W=1,Y=0)$	α, $\frac{\alpha/2}{\alpha+\beta}$	α, $\frac{(1-\alpha)/2}{2-(\alpha+\beta)}$	α, 1/4	β, 1/4
$P(X=0\|W=1,Y=1)$, $P(W=1,Y=1)$	α, $\frac{\beta/2}{\alpha+\beta}$	α, $\frac{(1-\beta)/2}{2-(\alpha+\beta)}$	α, 1/4	β, 1/4

	5	6	7
$P(X=0\|W=0,Y=0)$, $P(W=0,Y=0)$	β, $\frac{(1-\beta)/2}{2-(\alpha+\beta)}$	$\frac{\alpha+\beta}{2}$, 1/4	β, 1/4
$P(X=0\|W=0,Y=1)$, $P(W=0,Y=1)$	β, $\frac{(1-\alpha)/2}{2-(\alpha+\beta)}$	$\frac{\alpha+\beta}{2}$, 1/4	α, 1/4
$P(X=0\|W=1,Y=0)$, $P(W=1,Y=0)$	β, 1/4	α, 1/4	$\frac{\alpha+\beta}{2}$, 1/4
$P(X=0\|W=1,Y=1)$, $P(W=1,Y=1)$	β, 1/4	β, 1/4	$\frac{\alpha+\beta}{2}$, 1/4

Table 1: Failure of decomposition and weak union for conditional confirmational independence in the context of general credal sets. Variables W, X and Y are binary, and $\alpha, \beta \in (0, 1)$. The credal set $K(W, X, Y)$ contains only the seven distributions in the tables (one per column). Then (W, Y) and X are confirmationally independent, but $K(X|W = w, Y = y) \neq K(X|Y = y)$ (failure of weak union) and $K(X|Y = y) \neq K(X)$ (failure of decomposition).

5 Conclusion

The present paper focused on independence concepts for credal sets, with particular attention to issues related to convexity. Section 3 reviewed and tried to organize the existing literature on the issue; that section also contributed with an analysis of conditional regular independence and a proposal for connecting exchangeability with strong independence. Section 4 focused on the use of E-admissibility to generate general sets of distributions, an idea hinted at by Kyburg and Pittarelli, and taken to fruition by Seidenfeld et al [56].

To conclude, a few words on some notable concepts of independence for credal sets:

- Epistemic irrelevance/independence is quite intuitive and simple to state for convex credal sets. However, it is difficult to handle computationally, and it fails the contraction property even when all probabilities are positive. Moreover, epistemic irrelevance/independence requires full conditional measures and their associated challenges (then failing other graphoid properties when zero probabilities are present). Someone disinclined to use full conditional measures might adopt conditional regular independence.
- Strict independence is simple to state and inherits all the familiar properties of stochastic independence. Due to the now available axiomatization of general credal sets, strict independence can be given behavioral substance. Someone inclined to full conditional measures might adopt full strict independence; however, currently there is no axiomatization of general credal sets that produces a set of full conditional measures, and future work on this issue would be welcome.

As for strong independence, it stays uncomfortably between epistemic independence (which is more intuitive for convex credal sets) and strict independence (which is more general and easier to justify). The popularity of strong independence seems to be due solely to a desire to keep at once stochastic independence and convexity. However, we need not be so negative, as there are situations where strong independence does make sense: for instance, when exchangeability leads to mixtures of product measures (Section 3.5).

Acknowledgements

Thanks first to the organizers of the SIPTA school.

Thanks to Michael Pittarelli for clarifying some of the history behind his collaboration with Henry Kyburg Jr.; to Luis Gustavo Esteves for helpful discussions concerning Markov exchangeability; to Thomas Augustin for help with Weichselberger's theory; and many thanks to Teddy Seidenfeld for patiently explaining to me the theory of choice functions and its relationship with strict independence.

The author is partially supported by CNPq, and acknowledges significant support by FAPESP through grant 04/09568-0.

References

[1] T. E. Armstrong. Countably additive full conditional probabilities. *Proceedings of the American Mathematical Society*, 107(4):977–987, 1989.

[2] J. Berger and E. Moreno. Bayesian robustness in bidimensional models: Prior independence. *Journal of Statistical Planning and Inference*, 40:161–176, 1994.

[3] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, 1985.

[4] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory.* John Wiley and Sons, 1994.

[5] G. Coletti and R. Scozzafava. *Probabilistic Logic in a Coherent Setting.* Trends in logic, 15. Kluwer, Dordrecht, 2002.

[6] I. Couso, S. Moral, and P. Walley. Examples of independence for imprecise probabilities. In *Proc. of the First Int. Symp. on Imprecise Probabilities and Their Applications*, pages 121–130, Ghent, Belgium, 1999. Imprecise Probabilities Project, Universiteit Gent.

[7] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, 5:165–181, 2000.

[8] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems.* Springer-Verlag, New York, 1999.

[9] F. G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

[10] F. G. Cozman. Separation properties of sets of probabilities. In *Conf. on Uncertainty in Artificial Intelligence*, pages 107–115, San Francisco, 2000.

[11] F. G. Cozman. Constructing sets of probability measures through Kuznetsov's independence condition. In *Proc. of the Second Int. Symp. on Imprecise Probabilities and Their Applications*, pages 104–111, Ithaca, New York, 2001.

[12] F. G. Cozman and T. Seidenfeld. Independence for full conditional measures, graphoids and Bayesian networks. Technical report BT/PMR/0711, Escola Politecnica, Universidade de Sao Paulo, São Paulo, Brazil, 2007.

[13] F. G. Cozman and P. Walley. Graphoid properties of epistemic irrelevance and independence. *Annals of Mathematics and Artificial Intelligence*, 45:173–195, 2005.

[14] A. P. Dawid. Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society B*, 41:1–31, 1979.

[15] A. P. Dawid. Separoids: A mathematical framework for conditional independence and irrelevance. *Annals of Mathematics and Artificial Intelligence*, 32(1-4):335–372, 2001.

[16] C. Polpo de Campos and F. G. Cozman. Computing lower and upper expectations under epistemic independence. *Int. Journal of Approximate Reasoning*, 44(3):244–260, 2007.

[17] Cassio Polpo de Campos and Fabio Gagliardi Cozman. Inference in credal networs using multilinear programming. In *Proc. of the Second Starting AI Researchers' Symposium (STAIRS)*, pages 50–61, Amsterdam, The Netherlands, 2004. IOS Press.

[18] C. Polpo de Campos and F. Gagliardi Cozman. The inferential complexity of Bayesian and credal networks. In *Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 1313–1318, Edinburgh, United Kingdom, 2005.

[19] L. de Campos and S. Moral. Independence concepts for convex sets of probabilities. In *Conf. on Uncertainty in Artificial Intelligence*, pages 108–115, San Francisco, California, 1995. Morgan Kaufmann.

[20] G. de Cooman and E. Miranda. Weak and strong laws of large numbers for coherent lower previsions. *Journal of Statistical Planning and Inference*, 138(8):2409–2432, 2008.

[21] B. de Finetti. *Theory of probability, vol. 1-2.* Wiley, New York, 1974.

[22] P. Diaconis and S. L. Zabell. Updating subjective probability. *Journal of the American Statistical Association*, 77(380):822–829, 1982.

[23] P. Diaconis and D. Freedman. de Finetti's theorem for Markov chains. *The Annals of Probability*, 8(1):115–130, 1980.

[24] L. E. Dubins. Finitely additive conditional probability, conglomerability and disintegrations. *Annals of Statistics*, 3(1):89–99, 1975.

[25] E. Fagiuoli and M. Zaffalon. 2U: An exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106(1):77–107, 1998.

[26] P. C. Fishburn. *Utility Theory for Decision Making.* John Wiley and Sons, Inc., New York, 1970.

[27] P. Gardenfors and N. E. Sahlin. Unreliable probabilities, risk taking and decision making. *Synthese*, 53:361–386, 1982.

[28] D. Geiger, T. Verma, and J. Pearl. d-separation: from theorems to algorithms. *Uncertainty in Artificial Intelligence 5*, pages 139–148, 1990.

[29] F. J. Giron and S. Rios. Quasi-Bayesian behaviour: A more realistic approach to decision making? In *Bayesian Statistics*, pages 17–38. University Press, Valencia, Spain, 1980.

[30] A. Hajek. What conditional probability could not be. *Synthese*, 137:273–323, 2003.

[31] J. Y. Halpern. Lexicographic probability, conditional probability, and nonstandard probability. In *Proc. of the Eighth Conf. on Theoretical Aspects of Rationality and Knowledge*, pages 17–30, 2001.

[32] P. J. Hammond. Elementary non-Archimedean representations of probability for decision theory and games. In P. Humphreys, editor, *Patrick Suppes: Scientific Philosopher; Volume 1*, pages 25–59. Kluwer, Dordrecht, The Netherlands, 1994.

[33] D. Heath and W. Sudderth. de Finetti's theorem on exchangeable variables. *The American Statistician*, 30(4):188–189, 1976.

[34] R. Jeffrey. *The Logic of Decision*. McGraw-Hill, 1965.

[35] J. B. Kadane and A. O'Hagan. Using finitely additive probability: Uniform distributions on the natural numbers. *Journal of the American Statistical Association*, 90(430):626–631, 1995.

[36] J. B. Kadane, M. J. Schervish, and T. Seidenfeld. A Rubinesque theory of decision. In *A Festschrift for Herman Rubin*, pages 45–55. Institute of Mathematical Statistics, 2004.

[37] John Maynard Keynes. *A Treatise on Probability*. Macmillan and Co., London, 1921.

[38] D. Kikuti, F. G. Cozman, and C. Polpo de Campos. Partially ordered preferences in decision trees: computing strategies with imprecision in probabilities. In *IJCAI Workshop on Advances in Preference Handling*, pages 118–123, Edinburgh, United Kingdom, 2005.

[39] P. Krauss. Representation of conditional probability measures on Boolean algebras. *Acta Mathematica Academiae Scientiarum Hungaricae*, 19(3-4):229–241, 1968.

[40] V. P. Kuznetsov. *Interval Statistical Methods*. Radio i Svyaz Publ., (in Russian), 1991.

[41] V. P. Kuznetsov. Auxiliary problems of statistical data processing: Interval approach. In *Extended Abstracts of APIC95: Int. Workshop on Applications of Interval Computations*, pages 123–129, 1995.

[42] H. E. Kyburg Jr. and M. Pittarelli. Set-based Bayesianism. *IEEE Transactions on Systems, Man and Cybernetics A*, 26(3):324–339, 1996.

[43] H. Kyburg Jr. and M. Pittarelli. Some problems for convex Bayesians. In *Conf. on Uncertainty in Artificial Intelligence*, pages 149–154, 1992.

[44] H. E. Kyburg Jr. and M. Pittarelli. Set-based Bayesianism. Technical Report UR CSD;TR 407 (http://hdl.handle.net/1802/765), University of Rochester, Computer Science Department, March 1992.

[45] F. Lad. *Operational Subjective Statistical Methods: a Mathematical, Philosophical, and Historical Introduction.* John Wiley, New York, 1996.

[46] I. Levi. *The Enterprise of Knowledge.* MIT Press, Cambridge, Massachusetts, 1980.

[47] S. Moral and A. Cano. Strong conditional independence for credal sets. *Annals of Mathematics and Artificial Intelligence*, 35:295–321, 2002.

[48] S. Moral. Epistemic irrelevance on sets of desirable gambles. *Annals of Mathematics and Artificial Intelligence*, 45(1-2):197–214, 2005.

[49] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, California, 1988.

[50] K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges.* Academic Press, 1982.

[51] M. Schervish, T. Seidenfeld, J. Kadane, and I. Levi. Extensions of expected utility theory and some limitations of pairwise comparisons. In *Proc. of the Third Int. Symp. on Imprecise Probabilities and their Applications*, pages 496–510. Carleton Scientific, 2003.

[52] M. J. Schervish. *Theory of Statistics.* Springer-Verlag, New York, 1995.

[53] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. Decisions without ordering. In *Acting and Reflecting*, pages 143–170. Kluwer Academic Publishers, 1990.

[54] T. Seidenfeld and L. Wasserman. Dilation for sets of probabilities. *Annals of Statistics*, 21(9):1139–1154, 1993.

[55] T. Seidenfeld. Remarks on the theory of conditional probability: some issues of finite versus countable additivity. In *Statistics – Philosophy, Recent History, and Relations to Science*, pages 167–177. Kluwer Academic, 2001.

[56] T. Seidenfeld, M. Schervish, and J. Kadane. Coherent choice functions under uncertainty. In *Int. Symp. on Imprecise Probability: Theories and Applications.* Prague, Czech Republic, 2007.

[57] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. Improper regular conditional distributions. *The Annals of Probability*, 29(4):1612–1624, 2001.

[58] A. Sen. Social choice theory: a reexamination. *Econometrica*, 45:53–89, 1977.

[59] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[60] C. A. B. Smith. Consistency in statistical inference and decision. *Journal Royal Statistical Society B*, 23:1–25, 1961.

[61] M. Troffaes. Decision making with imprecise probabilities: A short review. *SIPTA Newsletter*, pages 4–7, 2004.

[62] L. V. Utkin and T. Augustin. Powerful algorithms for decision making under prior information and general ambiguity attitudes. In *Proc. of the Fourth Int. Symp. on Imprecise Probabilities and Their Applications*, pages 349–358, 2005.

[63] B. Vantaggi. Graphical models for conditional independence structures. In *Second Int. Symp. on Imprecise Probabilities and Their Applications*, pages 332–341. Shaker, 2001.

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

[65] P. Walley and T. L. Fine. Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, 10(3):741–761, 1982.

[66] K. Weichselberger. The theory of interval-probability as a unifying concept for uncertainty. *Int. Journal of Approximate Reasoning*, 24(2-3):149–170, 2000.

[67] K. Weichselberger, T. Augustin (assistant), and A. Wallner (assistant). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept.* Physica-Verlag Heidelberg, 2001.

[68] P. M. Williams. Coherence, strict coherence and zero probabilities. *Fifth Int. Congress of Logic, Methodology and Philos. Sci.*, VI:29–30, 1975.

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

[70] B. Ben Yaghlane, P. Smets, and K. Mellouli. Belief function independence: I. the marginal case. *Int. Journal of Approximate Reasoning*, 29(1):47–70, 2002.

[71] B. Ben Yaghlane, P. Smets, and K. Mellouli. Belief function independence: II. the conditional case. *Int. Journal of Approximate Reasoning*, 31(1-2):31–75, 2002.

[72] A. Zaman. A finite form of de Finetti's theorem for stationary Markov exchangeability. *The Annals of Probability*, 14(4):1418–1427, 1986.