Synthese (2012) 186:577–600
DOI 10.1007/s11229-011-9999-0

Sets of probability distributions, independence, and convexity

Fabio G. Cozman

Received: 8 August 2011 / Accepted: 8 August 2011 / Published online: 9 September 2011

Abstract This paper analyzes concepts of independence and assumptions of convexity in the theory of sets of probability distributions. The starting point is Kyburg and Pittarelli's discussion of "convex Bayesianism" (in particular their proposals concerning E-admissibility, independence, and convexity). The paper offers an organized review of the literature on independence for sets of probability distributions; new results on graphoid properties and on the justification of "strong independence" (using exchangeability) are presented. Finally, the connection between Kyburg and Pittarelli's results and recent developments on the axiomatization of non-binary preferences, and its impact on "complete" independence, are described.

Keywords Sets of probability distributions · Independence · Decision-making · Preferences · Convexity

1 Introduction

This paper analyzes concepts of independence in the theory of sets of probability distributions. Special effort is made to organize the various existing proposals and arguments into a few strands, and in particular to relate the quest for definitions of independence with the question of whether to require convexity of sets of probability distributions.

The starting point of this paper is an analysis of three papers published between 1992 and 1996 by Kyburg and Pittarelli, where they argue against convexity for sets of probability distributions, and advocate the use of general (not necessarily convex) sets

F. G. Cozman (✉)
Escola Politécnica, Universidade de São Paulo, São Paulo, SP, Brazil
e-mail: fgcozman@usp.br

of distributions to reason about uncertainty (Kyburg and Pittarelli 1992a,b, 1996).[1] In their arguments against convexity, Kyburg and Pittarelli analyze both E-admissibility and independence. Their contributions to these two topics are noteworthy and deserve recognition.

Section 2 presents needed background, in particular on the computation of E-admissible decisions. Sections 3 and 4 move to the main focus of this paper; namely, concepts of independence for sets of probability distributions. Section 3 organizes the literature on this topic and introduces new results on graphoid properties and on strong independence. Section 4 reviews work on axiomatization of non-binary preferences, including pioneering arguments by Kyburg and Pittarelli, and investigates the connection between such results and concepts of independence.

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 (Levi 1980). 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 later 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, a few additional axioms (Fishburn 1970, Chap. 13) 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 similar axioms yield a representation for $\succ$ in terms of a set of probability measures K (Giron and Rios 1980; Seidenfeld et al. 1990; Walley 1991):

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

Note that any two credal sets with identical convex hulls produce the same partial order $\succ$. Interest has focused on the unique *maximal* credal set that represents $\succ$,

[1] These three papers by Kyburg and Pittarelli were the result of collaboration between the authors at the University of Rochester, where Pittarelli was a visiting Professor. The joint research grew out of Pittarelli's PhD work and Kyburg's longtime concern about interval/set probabilities (and his friendly rivalry with Isaac Levi, a pioneer in "convex" Bayesianism).

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 only such maximal credal sets have behavioral meaning, and only them can work as representations for partial preferences. As we will see in Sect. 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 (1961) and Williams (1975, 2007), and later investigated in great detail by Walley (1991). In this case the *lower expectation* $\underline{E}[X]$ of X is interpreted as the supremum of prices one is willing to pay for X (Walley refers to X as a *gamble*). Conditions usually imposed on the functional $\underline{E}$ are (Walley 1991):

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

Any functional $\underline{E}[X]$ satisfying these conditions is the lower envelope of a set of expectation functionals; 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 (Walley 1991, Sect. 3.6.1). 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 Sect. 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 (Berger 1985; Gardenfors and Sahlin 1982),

$$\arg \sup_{X \in \mathcal{A}} \inf_{P \in K} E_P[X].$$

The *maximality* criterion selects any act X such that no other act $Y \in \mathcal{A}$ is preferred to X in a binary comparison (Sen 1977; Walley 1991, Sect. 3.9). Maximality focuses on the maximal elements of the partial order $\succ$. That is, X is *maximal* if

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

Hence maximality of X can be determined by considering all pairs (X, Y) for $Y \in \mathcal{A}$.

Finally, the *E-admissibility* criterion selects any act X that is maximal under at least one probability measure $P \in K$ (Levi 1980, Sect. 4.8). That is, X is *E-admissible* if

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

Kyburg and Pittarelli characterize E-admissibility as follows. 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 (Troffaes 2004). Kyburg and Pittarelli concentrate on E-admissibility, perhaps because their interest in convex Bayesianism led them to focus on Isaac Levi's work, where convexity and E-admissibility are deeply interwined.

At this point it is convenient to pause and mention a significant contribution in Kyburg and Pittarelli longer papers (Kyburg and Pittarelli 1996, 1992b). Suppose one has a finite set of acts $\mathcal{A}$ and a credal set K. It is not immediately clear how one would determine whether an act $X \in \mathcal{A}$ is E-admissible; it seems that it would be easier to determine whether X is a maximal act (just test all binary comparisons). Indeed one can find a clear statement of this perception in Troffaes excellent 2004 literature review: "from a computational viewpoint, maximality is apparently to be preferred over E-admissibility" (Troffaes 2004). As a response to Troffaes' review, in 2005 two papers reported on methods that can identify E-admissible acts: Utkin and Augustin (2005) examined one-step decision problems with pure and mixed acts, while Kikuti et al. (2005) discussed sequential decision making with independence relations and pure acts only. In both cases, the central insight was that, to verify whether a given act X is E-admissible, one must verify whether there is P such that:

1. $P \in K$;
2. $E_P[X - Y] \geq 0$ for all $Y \in \mathcal{A}$ such that $Y \neq X$.

So, the E-admissibility question is equivalent to feasibility of a set of inequalities.

It turns out that such an insight is already present in Kyburg and Pittarelli's work (Kyburg and Pittarelli 1996, Sect. III.B). The feasibility problem is derived from the condition $K_{\mathcal{A}}(X) \cap K \neq \emptyset$; their analysis focuses on pure acts without independence relations, but clearly all the elements of the solution are derived there.

3 Independence and convexity

Kyburg and Pittarelli suggest that convex Bayesianism is unable to cope with natural constraints one might like to impose on sets of probability distributions. For instance, they describe situations where one might be willing to adopt a set containing disjoint sets of distributions; an example is a biased coin where the probability of heads is *either* 1/2 *or* 1/3.

Other situations discussed by Kyburg and Pittarelli involve concepts of independence. Their discussion is based on generalizations of the widely used concept of stochastic independence. Recall that X and Y are *stochastically independent* if

$$P(X \in A | Y \in B) = P(X \in A) \quad \text{whenever } P(Y \in B) > 0, \tag{1}$$

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 the rich literature on independence concepts for credal sets. Section 3.1 reviews the main definitions in the literature, leaving concerns about conditioning, null events, and countable additivity to Sect. 3.2. Independence concepts are briefly compared in Sect. 3.3, and Sects. 3.4 and 3.5 focus on *strong* independence.

3.1 Confirmational, complete, strong, epistemic independence… and others

A direct generalization of stochastic independence to the context of credal sets is *complete* independence[2]: X and Y are *completely independent* if, for each $P \in K$,

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

This straightforward definition violates convexity, as emphasized by Kyburg and Pittarelli through an example of Jeffrey's (1965, 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 completely independent; hence X and Y are stochastically independent with respect to 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

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

Kyburg and Pittarelli emphasize the failure of convexity by presenting a clever argument based on Dutch Book (Kyburg and Pittarelli 1996, Sect. IVC): they show that an agent is sure to lose in the long run by betting using convex combinations of product measures when she knows that events are actually stochastically independent for every possible distribution.

The clash between complete independence and convexity is already explicit in Levi's pioneering work on convex Bayesianism (Levi 1980, Chap. 10). Levi defines Y to be *confirmationally irrelevant* to X if

$$K(X|Y \in B) = K(X) \quad \text{for nonempty } \{Y \in B\}, \tag{2}$$

and notes that confirmational irrelevance is not the same as complete independence. Levi argues that in the end expectations are not affected by the failure of convexity in complete 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 neces-

[2] This term is due to Seidenfeld (2007a).

sary. This suggests a slightly convoluted definition: X and Y are *strongly independent* when $K(X, Y)$ is the convex hull of a credal set that satisfies complete independence.

Levi refers to strong independence as *confirmational irrelevance in the strong sense* and as *strong confirmational irrelevance* (Levi 1980, Sect. 10.6). The term *strong independence* apparently appears first in a technical report written in 1982 by Walley (1982, Appendix), where the term refers to a concept that is almost identical to the one we just defined (the difference lies on the treatment of events with zero lower probability). Several other terms have been used since then to refer to strong independence; we try to organize the terminology in the remainder of this section.

A special case of strong independence is obtained when the credal set $K(X, Y)$ contains *every* product of a distribution from $K(X)$ and a distribution from $K(Y)$. In this case we have, for any bounded function $f(X, Y)$,

$$\underline{E}[f(X,Y)] = \inf\left(E_{P_X \times P_Y}[f(X,Y)] \ : P_X \in K(X)\,, P_Y \in K(Y)\right). \tag{3}$$

The joint credal set $K(X, Y)$ is then called a *strong extension* of $K(X)$ and $K(Y)$. Walley used the term *type-1 product* to refer to Expression (3) in his highly influential 1991 book (Walley 1991, Chap. 9).[3] Walley reserved the term *type-2 product* to the situation where all marginals are equal.

In a long technical report published in 1982, Walley (1982) proposed the concept of *epistemic independence*. In his 1991 book Walley prefers to build epistemic independence out of the concept of *epistemic irrelevance*, as we do now.[4] As defined in Walley's book,[5] 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. Indeed, definitions involving equalities among lower expectations tend to produce closed convex sets.

Epistemic irrelevance is not symmetric: Y may be epistemically irrelevant to X while X is not epistemically irrelevant to Y (Cozman and Walley 2005; Walley 1991). The clever idea in Walley's book is to create a symmetric concept out of epistemic irrelevance, following Keynes' approach to independence (Keynes 1921): X and Y are *epistemically independent* if Y is epistemically irrelevant to X and X is epistemically irrelevant to Y.

[3] In 1982 Walley and Fine considered a similar expression where constraints are restricted to events (Walley and Fine 1982, Sect. 3.1); that is, they require only that $K(X, Y)$ satisfies, for any event $A(X)$ defined by X and any event $B(Y)$ defined by Y, the constraint $\underline{P}(A(X)B(Y)) = \inf\left(P_X(A(X)) \times P_Y(B(Y)) : P_X \in K(X)\,, P_Y \in K(Y)\right)$. The credal set is then called an *independent product*; Weichselberger refers to independent products by the term *mutual independence* (Weichselberger 2000; Weichselberger et al. 2001).

[4] There is actually a small difference between the treatment of conditioning between the 1982 and the 1991 versions of epistemic independence, but for now we ignore this difference.

[5] Epistemic irrelevance is what Smith refers to just as *independence* in his pioneering work on medial odds (Smith 1961). The weaker pair of conditions $\overline{P}(A|B) = \overline{P}(A)\,, \underline{P}(A|B) = \underline{P}(A)$ is termed *canonical* independence of B to A by Weichselberger (2000) and Weichselberger et al. (2001).

Walley's book is contemporary with Kuznetsov's book on interval probabilities and interval expectations (Kuznetsov 1991). Kuznetsov proposed yet another concept of independence with a nice interpretation in terms of interval arithmetic [it seems Kuznetsov took strong independence as the main concept of independence and proposed his new concept as a secondary idea (Kuznetsov 1995)]. Denote by $EI[X]$ the interval $[\underline{E}[X], \overline{E}[X]]$; X and Y are *Kuznetsov independent* if, for any bounded functions $f(X)$ and $g(Y)$,

$$EI[f(X)g(Y)] = EI[f(X)] \times EI[g(Y)], \tag{4}$$

where $\times$ is interval multiplication (that is, if $a = [\underline{a}, \overline{a}]$ and $b = [\underline{b}, \overline{b}]$, $a \times b = [\min \gamma, \max \gamma]$ for $\gamma = \{\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}\}$).

To some extent, Expressions (3) and (4) are conceptually similar, particularly if the credal sets are closed. However, the expressions are not equivalent in general (Cozman 2001).

Many variations on the previous definitions are possible. Indeed, several variations appeared in the literature between 1990 and 2000, and terminology became somewhat confusing. Part of this research activity was motivated by results in Dempster–Shafer and possibility theories, where concepts of conditioning and independence were intensely debated during that decade. 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* (Shafer 1976; Yaghlane et al. 2002a,b). In the present paper we do not discuss theories that are not based on probability distributions.

In 1995 de Campos and Moral tried to organized the field into a small number of distinct concepts of independence (de Campos and Moral 1995). Their *type-2* independence is strong independence as defined previously (that is, $K(X, Y)$ is the convex hull of a set where each distribution 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 $P_Y \in K(Y)$. So in fact type-3 independence is just a name for the *strong extension* of 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

$$R(X|Y \in B) = K(X) \quad \text{whenever } \overline{P}(Y \in B) > 0, \tag{5}$$

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

$$\{P(\cdot|Y \in B) : P \in K(X, Y) \ \text{ and } \ P(Y \in B) > 0\}. \tag{6}$$

The set R defined in Expression (6) is often used as a definition of conditioning (Weichselberger 2000), and it is related to what Walley calls *regular extension* (Walley 1991) (however, regular extension is different in that it defines conditioning even if $\overline{P}(Y \in B) = 0$).

The following example is given by de Campos and Moral (1995), to show that strong and type-5 irrelevance differ in notable ways in the presence of zero probabilities. Suppose X and Y are binary, and $K(X, Y)$ is the convex hull of two distributions P_1 and P_2 such that

$$P_1(X = 0, Y = 0) = P_2(X = 1, Y = 1) = 1.$$

Even though strong independence obtains, neither Y is type-5 irrelevant to X nor X is type-5 irrelevant to Y.

It is natural to define type-5 independence by "symmetrizing" type-5 irrelevance; that is, type-5 independence of X and Y is type-5 irrelevance of X to Y and type-5 irrelevance of Y to X. Curiously, type-5 independence is very similar to Walley's 1982 definition of epistemic independence (Walley 1982).

In 1999, Couso et al. presented an influential review of independence concepts (Couso et al. 1999, 2000). 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 (that is, strong extension). Finally, their *repetition independence* refers to Walley's *type-2 product*.[6]

To conclude this section, the following summary may be useful.

- Complete independence is an intuitive generalization of stochastic independence, but it fails convexity.
- Strong independence is an ad hoc combination of stochastic independence and convexity; several special cases of strong independence have been investigated.
- Epistemic independence is quite elegant, and several related concepts been proposed (most notably confirmational irrelevance).
- Kuznetsov independence is a rather different concept that is inspired by interval arithmetic.

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. To avoid technical issues concerning the definition of conditional probability, assume all spaces are finite.

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*

[6] 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 such that the joint mass assignment satisfies stochastic independence. Couso et al call this latter concept *random set independence* [a similar concept had been called a *belief function product* by Walley and Fine (1982)].

complete independence imposes elementwise conditional stochastic independence[7]: every probability distribution must satisfy

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

We adopt the same scheme for *conditional strong independence*; that is, strong independence conditional on every value z of Z. Also, *conditional epistemic irrelevance* of Y to X given Z obtains when, for all bounded functions $f(X)$,

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

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

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

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. For instance, it may seem strange that the definition of confirmational irrelevance, given by Expression (2), does not contain a clause discarding conditioning events that have zero probability. The usual attitude in probability theory is to discard null events as these events almost surely do not obtain. But one cannot ignore an event with zero *lower* probability but nonzero *upper* probability. 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 (de Finetti 1974; Dubins 1975). 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

$$\begin{aligned} &P(A|C) \geq 0; \quad P(\Omega|C) = 1;\\ &P(A \cup B|C) = P(A|C) + P(B|C) \quad \text{if } A \cap B = \emptyset; \quad \text{and}\\ &P(A \cap B|C) = P(A|B \cap C)\, P(B|C) \quad \text{if } B \cap C \neq \emptyset. \end{aligned}$$

[7] Moral and Cano describe three variants on conditional complete independence (Moral and Cano 2002), 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.

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

Full conditional measures are extremely elegant and have been advocated for a variety of reasons (Coletti and Scozzafava 2002; Dubins 1975; Hajek 2003; Seidenfeld 2001), 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 (Seidenfeld et al. 2001). Characterizing the situations where full conditional measures exist under countable additivity seems to be a hard (and mostly open) problem (Armstrong 1989; Krauss 1968). Even though some authors have preferred to ignore these existence problems (Cowell et al. 1999), it seems that in general one is forced into finite additivity when full conditional measures are adopted (Seidenfeld 2001).

In short, confirmational and epistemic irrelevance/independence seem to require a combination of full conditional measures and finite additivity. This is indeed the path taken by Levi and Walley (the latter imposes additional conditions of conglomerability on lower expectations).

3.2.3 Full conditional measures and complete/strong independence

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 and B such that $P(Y \in B) = 0$.

Complete and strong independence can be adapted to the specificities of full conditional measures as follows.

- *Full* complete 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* complete independence of X and Y is full complete irrelevance of Y to X and full complete irrelevance of X to Y.
- And likewise for *full* strong irrelevance and *full* strong independence (that is, they are "convexified" versions of full complete irrelevance and full complete independence).

At this point the reader may despair as it seems we are quickly exhausting 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 (Cozman and Seidenfeld 2007; Halpern 2001; Hammond 1994; Vantaggi 2001). It does not seem that such modified concepts of irrelevance/independence have been applied to credal sets yet, but they offer possible paths to follow.

3.2.4 Conditional regular irrelevance and independence

One way to avoid the complexities of full conditional measures 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

$$R(X|Y \in B, Z = z) = R(X|Z = z) \quad \text{whenever } \overline{P}(Y \in B, Z = z) > 0. \tag{7}$$

Conditional regular independence is then the "symmetrized" concept. It does not seem that conditional regular independence has been explored in the literature; Sect. 3.3 presents an initial analysis of its properties.

3.3 Comparing concepts: laws of large numbers and graphoid properties

The previous sections listed about a dozen concepts of (conditional) irrelevance/independence, and a number of variants and special cases. These concepts are not identical; for instance, epistemic independence, Kuznetsov independence and strong independence generate distinct constraints (Cozman 2001). We revisit the summary at the end of Sect. 3.1, now with additional commentary:

1. Complete independence is easy to state and obviously close to stochastic independence; however it violates convexity (thus failing to have a behavioral interpretation through partial preferences and one-sided betting).
2. Strong independence has some of the appeal of complete 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 Sects. 3.4 and 3.5.
3. Confirmational/epistemic irrelevance/independence seem to be the most intuitive concepts once convexity is required. Conditional regular irrelevance/independence is a viable alternative if one wishes to stay away from full conditional measures.
4. Kuznetsov independence appears as an "interval" generalization of the product form of stochastic independence, but it is hard to imagine a behavioral justification for it.

It is instructive to compare these concepts. For instance, we might try to classify concepts on the basis of which laws of large numbers they imply, if any. This particular question has been mostly settled by de Cooman and Miranda (2008), who have produced quite general laws of large numbers for assumptions of epistemic irrelevance. Their results can be extended even to unbounded variables (Cozman 2009). As such laws of large numbers are derived from very weak assumptions of epistemic irrelevance, they are valid for most, if not all, concepts of irrelevance/independence one might contemplate. Consequently, we must look for other potential differences among concepts.

One possibility is to compare concepts of independence using the *graphoid* properties (Dawid 1979, 2001; Pearl 1988). Graphoid properties are stated for a ternary relation $(X \perp\!\!\!\perp Y \mid Z)$ that should be understood as "independence of X and Y given Z" (Geiger et al. 1990):

Symmetry	$(X \perp\!\!\!\perp Y \mid Z) \Rightarrow (Y \perp\!\!\!\perp X \mid Z)$
Decomposition	$(X \perp\!\!\!\perp (W, Y) \mid Z) \Rightarrow (X \perp\!\!\!\perp Y \mid Z)$
Weak union	$(X \perp\!\!\!\perp (W, Y) \mid Z) \Rightarrow (X \perp\!\!\!\perp W \mid (Y, Z))$
Contraction	$(X \perp\!\!\!\perp Y \mid Z)\ \&\ (X \perp\!\!\!\perp W \mid (Y, Z)) \Rightarrow (X \perp\!\!\!\perp (W, Y) \mid Z)$

(These properties are often called *semi-graphoid* properties (Pearl 1988), and the term *graphoid* is then reserved for a larger set of properties.)

Several results on the relationship between concepts of independence and graphoid properties can be found in the literature. Conditional complete and strong independence satisfy all graphoid properties (Cozman 2000a). Conditional confirmational/epistemic independence and conditional Kuznetsov independence fail contraction (Cozman and Walley 2005) even when all probabilities are positive.[8]

Full conditional measures introduce additional complications. Confirmational/epistemic independence may then fail even decomposition and weak union when zero probabilities are present (Cozman and Walley 2005). In fact confirmational/epistemic independence may fail weak union even for a *single* full conditional measure when zero probabilities are present (Cozman and Seidenfeld 2007; Vantaggi 2001). Consequently, failure of weak union can be also observed in full complete and full strong irrelevance/independence in the presence of events of zero probability. As noted at the end of Sect. 3.2.3, other concepts of independence have been proposed for a *single* full conditional measure (Cozman and Seidenfeld 2007; Hammond 1994), but no study of such concepts has been conducted in the context of *sets* of full conditional measures.

In Sect. 3.2.4 we introduced conditional regular independence, a concept of independence that might behave appropriately in the presence of zero probabilities. However, conditional regular independence does not fare well with respect to graphoid properties. Symmetry holds by definition and contraction fails as it can fail already for epistemic independence with positive lower probabilities. More troubling is the fact that conditional regular independence can fail decomposition and weak union when lower probabilities are zero [by adapting previous examples aimed at epistemic independence (Cozman and Walley 2005, Example 1)]. The frustrating fact is that, while stochastic independence satisfies decomposition and weak union by resorting to clauses forbidding conditioning events of zero probability (Expression 1), similar clauses used in conditional regular independence do not have a similar effect. While it is possible to add assumptions of positivity on lower probabilities and then to prove versions of decomposition and weak union [by adapting the proofs for epistemic independence (Cozman and Walley 2005, Theorem 1)], the value of conditional regular independence, if any, is yet to be determined.

As a digression, note that failure of the contraction property greatly affects the theory of statistical models such as Markov chains and Bayesian networks (Pearl 1988).

[8] 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 (Moral 2005). Also the variants of complete/strong independence proposed by Moral and Cano (2002) fail different sets of graphoid properties.

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, W and Z are conditionally stochastically independent given X. Failure of contraction destroys such implications. For instance, if we replace stochastic independence by epistemic independence, then it is possible to construct a Markov chain where W and Z are *not* conditionally epistemically independent given X (Cozman 2000b, Example 1).

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 (de Campos and Cozman 2005), and some algorithms have been produced for epistemic independence (de Campos and Cozman 2007). 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. Most of this landscape has been produced after the collaboration between Kyburg and Pittarelli. However, their questions are as sharp as ever; in particular, how can we stay close to stochastic independence while keeping convexity? A possible solution is to properly justify strong independence.

One might try to justify strong independence using what Walley calls the *sensitivity* interpretation of credal sets (Walley 1991). Suppose several experts agree that X and Y are stochastically independent; however they disagree on specific probability values. The experts then adopt a credal set containing distributions that factorize according to stochastic independence, plus the convex combinations of these distributions. They do so by accepting that such convex combinations do not affect their collective preferences. As far as preferences are represented by binary comparisons, as in Sect. 2, this argument for convex combinations is a powerful one. However the argument breaks down when one notes, as we do in Sect. 4, that convex combinations do affect non-binary preferences.

An alternative strategy is to obtain strong independence directly, without even mentioning stochastic independence. Proposals to this effect were independently concocted around 2000 by Moral and Cano (2002) and by Cozman (2000b). To understand the proposals, consider the following example. Take two binary variables X and Y so that

$$P(X = 0) \in [2/5, 1/2] \quad \text{and} \quad P(Y = 0) \in [2/5, 1/2].$$

The largest credal set $K(X, Y)$ satisfying these assessments and epistemic independence of X and Y has six vertices (Walley 1991, Sect. 9.3.4):

$$[1/4, 1/4, 1/4, 1/4], [4/25, 6/25, 6/25, 9/25],$$
$$[1/5, 1/5, 3/10, 3/10], [1/5, 3/10, 1/5, 3/10],$$
$$[2/9, 2/9, 2/9, 1/3], [2/11, 3/11, 3/11, 3/11],$$

where each vector denotes a distribution as

$$[P(X=0, Y=0)\,,\, P(X=0, Y=1)\,,\, P(X=1, Y=0)\,,\, P(X=1, Y=1)].$$

Suppose we learn that $P(Y = 0) = 4/9$. What should we do with this new assessment? One option is to "intersect" $K(X, Y)$ with the constraint $\{P : P(Y = 0) = 4/9\}$; that is, to form a new credal set

$$K'(X, Y) = K(X, Y) \cap \{P : P(Y = 0) = 4/9\}. \tag{8}$$

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, with respect to $K'(X, Y)$,

$$\underline{P}(Y = 0|X = 1) \le 2/5 < 4/9 = \underline{P}(Y = 0)\,.$$

Epistemic independence of X and Y, satisfied by $K(X, Y)$, is not preserved through Expression (8).

The situation just outlined reminds one of Jeffrey's rule (1965). 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. 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 (Diaconis and Zabell 1982, Theorem 3.3). A similar result holds for credal sets, as follows (Cozman 2000b). Suppose we change either $K(X)$ or $K(Y)$ into a new marginal credal set, and we modify $K(X, Y)$ by pointwise application Jeffrey's rule: if X and Y are still epistemically independent after any such change, then X and Y are fully strongly independent.

Moral and Cano follow the same idea (Moral and Cano 2002), but use a better strategy that avoids potential controversies on how to apply Jeffrey's rule on credal sets. Their approach uses two definitions, where $f(X)$ is a bounded function:

– Assessment $E[f(X)] \ge \alpha$ is *compatible* with marginal credal set $K(X)$ if there is $P \in K(X)$ that satisfies $E_P[f(X)] \ge \alpha$.
– Joint credal set $K(X, Y)$ is *combined* with assessment $\underline{E}[f(X)] \ge \alpha$ 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 (2002, Theorem 2):

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 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. Note that Moral and Cano formulate their theorem so as to generate the strong extension of $K(X)$ and $K(Y)$ (that is, the *largest* credal sets satisfying strong independence and the assessments defining $K(X)$ and $K(Y)$). There is no reason to restrict attention to this situation. Note also that conditional *full* strong independence is obtained from epistemic independence.

While Theorem 1 is a viable strategy to justify (full) strong independence, this approach does have a conceptual weakness. For suppose we have a credal set $K(X, Y)$ that is the largest set satisfying both: (1) a set of assessments (that is, a set of constraints on probabilities); and (2) epistemic independence of X and Y. Now we receive a new compatible assessment on Y. Should we:

- combine this new assessment with $K(X, Y)$ by intersection, as done in Expression (8); or
- construct the largest credal set that simultaneously satisfies the original assessments, the new assessment, and epistemic independence of X and Y?

These two approaches may lead to different credal sets. Cano and Moral assume that the first approach is the natural one; that is, a new assessment is always combined with the currently held credal set. However, we might choose to recompute the joint credal set with all available assessments and judgements of independence. This latter approach breaks the argument embedded in Theorem 1.

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.

3.5.1 Exchangeability for binary variables

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 $\cap_{i=1}^{m}\{X_{\pi_m(i)} = x_i\}$. Following de Finetti, we do not differentiate between an event and its indicator function.

Variables $X_1, \ldots, X_m$ are *exchangeable* when (Walley 1991, Chap. 9):

$$\underline{E}[\{\mathbf{X} = \mathbf{x}\} - \{\pi_m\mathbf{X} = \mathbf{x}\}] = 0 \quad \text{for any permutation } \pi_m. \tag{9}$$

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

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

Consequently, for every distribution $P \in K(X_1, \ldots, X_n)$,

$$E_P[\{\mathbf{X} = \mathbf{x}\} - \{\pi_m \mathbf{X} = \mathbf{x}\}] = 0;$$

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

In words: Expression (9) implies *elementwise* exchangeability in the usual de Finetti's sense (de Finetti 1974).

Fix a distribution P satisfying exchangeability for a moment. If we examine a subset $X_1, \ldots, X_n$ of variables, for $n \leq m$, these n variables are also exchangeable. Then (Heath and Sudderth 1976),

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

Now if an infinitely long sequence of exchangeable variables is contemplated ($m \to \infty$), de Finetti's representation theorem yields (Heath and Sudderth 1976; Schervish 1995):

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

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

3.5.2 Strong independence from exchangeability: binary variables

With this machinery in hand, we return to the problem of justifying strong independence. Suppose we have binary variables $X_1, \ldots, X_n$, and we judge these variables to be the initial fragment of an infinite sequence of exchangeable variables. By de Finetti's representation theorem, each distribution in the joint credal set $K(X_1, \ldots, X_n)$ is a mixture of factorizing distributions, and the mixture is characterized by a distribution function $F(\theta)$. It should be noted that exchangeability is a "convex" concept in the sense that if two distributions P_1 and P_2 satisfy exchangeability of $X_1, \ldots, X_n$, then any convex combination $\alpha P_1 + (1-\alpha)P_2$ of these distributions also satisfy exchangeability. Hence if $F_1(\theta)$ and $F_2(\theta)$ are distribution functions obtained through exchangeability, then so is the distribution $\alpha F_1(\theta) + (1-\alpha)F_2(\theta)$.

Strong independence of $X_1, \ldots, X_n$ obtains when we select a convex set $K(\theta)$ such that each vertex of $K(\theta)$ assigns probability 1 to a particular value of θ. By doing so, we produce a type-2 product in Walley's terminology (Sect. 3.1), as every vertex of $K(X_1, \ldots, X_n)$ becomes a product measure with identical marginals. In short, we have produced a judgement of strong independence from a judgement of exchangeability plus a condition on $K(\theta)$. One may even produce subsets of the strong extension by selecting a smaller $K(\theta)$.

3.5.3 Strong independence from exchangeability

The previous argument can be extended to non-binary variables, using suitable versions of de Finetti's representation theorem (Schervish 1995). Similarly, one can also modify the argument to obtain "convexified" sets of Markov chains, using judgements of *Markov exchangeability* (Diaconis and Freedman 1980; Zaman 1986).

With some additional imagination, the previous argument can also be modified to obtain strong independence of variables that do not necessarily have identical marginal credal sets. Consider two binary variables X_1 and Y_1 with values 0 and 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* (Bernardo and Smith 1994; de Finetti 1974; Lad 1996): for any permutations π'_m, π''_m,

$$\underline{E}\big[\{\mathbf{X} = \mathbf{x}\}\{\mathbf{Y} = \mathbf{y}\} - \{\pi'_m\mathbf{X} = \mathbf{x}\}\{\pi''_m\mathbf{Y} = \mathbf{y}\}\big] = 0.$$

Then, as $m \to \infty$, we have (Bernardo and Smith 1994, Theorem 4.13) a convex set of distribution functions $F(\theta, \vartheta)$ such that for every P we can write for some $F(\theta, \vartheta)$:

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

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, but the central idea is rather simple: X_1 and Y_1 are strongly independent 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 Kyburg and Pittarelli's suggestion: that we should drop convexity altogether and adopt complete independence. The difficulty with this prescription is that general credal sets do not seem to have a justification based on partial preferences/one-sided betting.[9] 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 (Kyburg and Pittarelli 1996, 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 act 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 using E-admissibility.

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 work (Seidenfeld et al. 2007) that greatly advances previous efforts (Kadane et al. 2004; Schervish et al. 2003).

A quick summary of Seidenfeld et al's theory is as follows (Seidenfeld et al. 2007). 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 $\mathsf{R}(D)$. Seidenfeld et al impose a set of axioms on rejection functions and prove that a rejection function satisfies their axioms if and only if it can be represented through non-E-admissible acts with respect to a set of pairs of utilities/probabilities (Seidenfeld et al. 2007, Theorems 3 and 4).

To understand the kinds of axioms that are proposed by Seidenfeld et al, consider their first axiom: An inadmissible act cannot become admissible (a) when new acts

[9] 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 whenever independence relations are used, for whatever concept of independence we have described, 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 (Berger and Moreno 1994; Fagiuoli and Zaffalon 1998; de Campos and Cozman 2004).

are added to the set of acts; and (b) when inadmissible acts are deleted from the set of acts.[10] Another axiom states that if $d \in \mathsf{R}(\text{convexhull}(D))$, then $d \in \mathsf{R}(D)$. 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 (Fishburn 1970). 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.

We now explore these ideas again in the context of independence. Suppose we wish to determine whether events A and B are completely independent, under the assumption that $0 < \underline{P}(A)\,, \underline{P}(B) \leq \overline{P}(A)\,, \overline{P}(B) < 1$. Construct five acts $a_0, \ldots, a_4$ as follows:

	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
a_3	0	0	$1-\beta$	$-\beta$
a_4	0	0	$-(1-\beta)$	β

These five acts serve as a test for complete independence[11]: 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 completely independent. For 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 completely 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. Conversely, note that when A and B are completely 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 $a_0 \in R(a_0, \ldots, a_4) \neq \emptyset$ (using the fact that lower and upper probabilities are neither zero nor one).

As a short digression, at this point one might consider revisiting confirmational irrelevance/independence (Expression (2)) in the context of general credal sets. That is, we might consider

$$K(X|Y \in B) = K(X)$$

as a condition that applies for arbitrary credal sets (not worrying, for the moment, about how to handle events of zero probability). Alas, the resulting irrelevance concept and

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

[11] This discussion is based on a very compact example produced by Teddy Seidenfeld. In his example complete 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.

its symmetrized independence version do not seem very interesting. Conditional confirmational independence fails *all* graphoid properties except symmetry even when probabilities are all positive. Contraction already fails for convex credal sets (Cozman and Walley 2005, Example 1); failure of decomposition and weak union is depicted in Table 1.

5 Conclusion

Kyburg and Pittarelli's joint papers touch on many central issues in uncertain reasoning. Their papers critically evaluate strict Bayesianism, qualitative and interval probability, maximum entropy methods, and, with special emphasis, convex Bayesianism. A notable byproduct of their discussion of convex Bayesianism is a method to compute E-admissible strategies, as discussed in Sect. 2.

Kyburg and Pittarelli present three drawbacks of convex Bayesianism. First, its inability to deal with common kinds of assessments such as (complete) independence. Second, its vulnerability to "long run" forms of Dutch Book when (complete) independence is assessed. Third, the sensitivity of E-admissibility (and consequently of decision making) to lack of convexity. Kyburg and Pittarelli argue that convexity should not be required, and that one should represent uncertainty through general credal sets; they call this prescription "Set-Based Bayesianism."

The present paper focused on the relationship between independence and 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 justify general sets of distributions, an idea hinted at by Kyburg and Pittarelli, and taken to fruition by Seidenfeld et al. (2007).

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 and independence require 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 (however most difficulties with graphoid properties persist).
- Complete independence is simple to state and inherits all the familiar properties of stochastic independence. Due to the now available axiomatization of general credal sets, complete independence can be given behavioral substance. Someone inclined to full conditional measures might adopt full complete 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 the more intuitive concept of epistemic independence and the easier to handle concept of complete independence. The popularity of strong independence seems to be due solely to a desire to keep at

Table 1 Failure of decomposition and weak union for conditional confirmational independence in the context of general credal sets

	P_1	P_2	P_3	P_4	P_5	P_6	P_7
$P(X=0 \mid W=0, Y=0)$, $P(W=0, Y=0)$	α, $1/4$	α, $1/4$	α, $1/4$	β, $\frac{\beta/2}{\alpha+\beta}$	β, $\frac{(1-\beta)/2}{2-(\alpha+\beta)}$	$\frac{\alpha+\beta}{2}$, $1/4$	β, $1/4$
$P(X=0 \mid W=0, Y=1)$, $P(W=0, Y=1)$	α, $1/4$	α, $1/4$	α, $1/4$	β, $\frac{\alpha/2}{\alpha+\beta}$	β, $\frac{(1-\alpha)/2}{2-(\alpha+\beta)}$	$\frac{\alpha+\beta}{2}$, $1/4$	α, $1/4$
$P(X=0 \mid W=1, Y=0)$, $P(W=1, Y=0)$	α, $\frac{\alpha/2}{\alpha+\beta}$	α, $\frac{(1-\alpha)/2}{2-(\alpha+\beta)}$	α, $1/4$	β, $1/4$	β, $1/4$	α, $1/4$	$\frac{\alpha+\beta}{2}$, $1/4$
$P(X=0 \mid W=1, Y=1)$, $P(W=1, Y=1)$	α, $\frac{\beta/2}{\alpha+\beta}$	α, $\frac{(1-\beta)/2}{2-(\alpha+\beta)}$	α, $1/4$	β, $1/4$	β, $1/4$	β, $1/4$	$\frac{\alpha+\beta}{2}$, $1/4$

Variables W, X and Y are binary, and $\alpha, \beta \in (0,1), \alpha \neq \beta$. The credal set $K(W, X, Y)$ contains *only* the seven distributions in the tables (one distribution per column, specified through $P(X|W, Y)$ and $P(W, Y)$). For this credal set, (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) once stochastic independence and convexity. Section 3.5 presented a perhaps more positive argument where strong independence is a consequence of exchangeability.

Acknowledgments 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 complete independence. The author is partially supported by CNPq, and acknowledges significant support by FAPESP through research grants 04/09568-0 and 08/03995-5.

References

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

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer.

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

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory*. New York: Wiley.

Coletti, G., & Scozzafava, R. (2002). *Probabilistic logic in a coherent setting*. Trends in logic, 15. Dordrecht: Kluwer.

Couso, I., Moral, S., & Walley, P. (1999). Examples of independence for imprecise probabilities. In G. de Cooman, F. Cozman, S. Moral, & P. Walley (Eds.), *Proceedings of the first international symposium on imprecise probabilities and their applications*, Ghent, Belgium (pp. 121–130).

Couso, I., Moral, S., & Walley, P. (2000). A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy, 5*, 165–181.

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., & Spiegelhalter, D. J. (1999). *Probabilistic networks and expert systems*. New York: Springer.

Cozman, F. G. (2000a). Credal networks. *Artificial Intelligence, 120*, 199–233.

Cozman, F. G. (2000b). Separation properties of sets of probabilities. In C. Boutilier & M. Goldszmidt (Eds.), *Conference on uncertainty in artificial intelligence*, San Francisco (pp. 107–115).

Cozman, F. G. (2001). Constructing sets of probability measures through Kuznetsov's independence condition. In *Proceedings of the second international symposium on imprecise probabilities and their applications*, Ithaca, NY, USA (pp. 104–111).

Cozman, F. G. (2009). Concentration inequalities and laws of large numbers under epistemic irrelevance. In *Proceedings of the sixth international symposium on imprecise probabilities and their applications*, Durham, UK.

Cozman, F. G., & Seidenfeld, T. (2007). Independence for full conditional measures, graphoids and Bayesian networks. Technical report, PMR, Escola Politecnica, Universidade de Sao Paulo, São Paulo, Brazil.

Cozman, F. G., & Walley, P. (2005). Graphoid properties of epistemic irrelevance and independence. *Annals of Mathematics and Artificial Intelligence, 45*, 173–195.

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

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

de Campos, C. P., & Cozman, F. G. (2004). Inference in credal networks using multilinear programming. In E. Onaindia & S. Staab (Eds.), *Proceedings of the second starting AI researchers' symposium (STAIRS)*, Amsterdam, The Netherlands (pp. 50–61).

de Campos, C. P., & Cozman, F. G. (2005). The inferential complexity of Bayesian and credal networks. In *International joint conference on artificial intelligence (IJCAI)*, Edinburgh, UK (pp. 1313–1318).

de Campos, C. P., & Cozman, F. G. (2007). Computing lower and upper expectations under epistemic independence. *International Journal of Approximate Reasoning, 44*(3), 244–260.

de Campos, L., & Moral, S. (1995). Independence concepts for convex sets of probabilities. In P. Besnard & S. Hanks (Eds.), *XI conference on uncertainty in artificial intelligence*, San Francisco, CA, USA (pp. 108–115).

de Cooman, G., & Miranda, E. (2008). Weak and strong laws of large numbers for coherent lower previsions. *Journal of Statistical Planning and Inference, 138*(8), 2409–2432.
de Finetti, B. (1974). *Theory of probability* (Vols. 1–2). New York: Wiley.
Diaconis, P., & Freedman, D. (1980). de Finetti's theorem for Markov chains. *The Annals of Probability, 8*(1), 115–130.
Diaconis, P., & Zabell, S. L. (1982). Updating subjective probability. *Journal of the American Statistical Association, 77*(380), 822–829.
Dubins, L. E. (1975). Finitely additive conditional probability, conglomerability and disintegrations. *Annals of Statistics, 3*(1), 89–99.
Fagiuoli, E., & Zaffalon, M. (1998). 2U: An exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence, 106*(1), 77–107.
Fishburn, P. C. (1970). *Utility theory for decision making*. New York: Wiley.
Gardenfors, P., & Sahlin, N. E. (1982). Unreliable probabilities, risk taking and decision making. *Synthese, 53*, 361–386.
Geiger, D., Verma, T., & Pearl, J. (1990). d-Separation: From theorems to algorithms. *Uncertainty in Artificial Intelligence, 5*, 139–148.
Giron, F. J., & Rios, S. (1980). Quasi-Bayesian behaviour: A more realistic approach to decision making? In J. M. Bernardo, J. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics* (pp. 17–38). Valencia: University Press.
Hajek, A. (2003). What conditional probability could not be. *Synthese, 137*, 273–323.
Halpern, J. Y. (2001). Lexicographic probability, conditional probability, and nonstandard probability. In *Proceedings of the eighth conference on theoretical aspects of rationality and knowledge* (pp. 17–30).
Hammond, P. J. (1994). Elementary non-Archimedean representations of probability for decision theory and games. In P. Humphreys (Ed.), *Patrick suppes: Scientific philosopher* (Vol. 1, pp. 25–59). Dordrecht: Kluwer.
Heath, D., & Sudderth, W. (1976). de Finetti's theorem on exchangeable variables. *The American Statistician, 30*(4), 188–189.
Jeffrey, R. (1965). *The logic of decision*. New York: McGraw-Hill.
Kadane, J. B., Schervish, M. J., & Seidenfeld, T. (2004). A Rubinesque theory of decision. In *A festschrift for Herman Rubin* (pp. 45–55). Beachwood, OH: Institute of Mathematical Statistics.
Keynes, J. M. (1921). *A treatise on probability*. London: Macmillan and Co.
Kikuti, D., Cozman, F. G., & de Campos, C. P. (2005). Partially ordered preferences in decision trees: Computing strategies with imprecision in probabilities. In *IJCAI workshop on advances in preference handling*, Edinburgh, UK (pp. 118–123).
Krauss, P. (1968). Representation of conditional probability measures on Boolean algebras. *Acta Mathematica Academiae Scientiarum Hungaricae, 19*(3–4), 229–241.
Kuznetsov, V. P. (1991). Interval statistical methods. Radio i Svyaz Publ (in Russian).
Kuznetsov, V. P. (1995). Auxiliary problems of statistical data processing: Interval approach. In *Extended abstracts of APIC95: International workshop on applications of interval computations* (pp. 123–129).
Kyburg, Jr., H., & Pittarelli, M. (1992a). Some problems for convex Bayesians. In *Conference on uncertainty in artificial intelligence* (pp. 149–154).
Kyburg, Jr., H. E., & Pittarelli, M. (1992b). Set-based Bayesianism. technical report UR CSD; TR 407. University of Rochester, Computer Science Department. http://hdl.handle.net/1802/765.
Kyburg, H. E., & Pittarelli, M. (1996). Set-based Bayesianism. *IEEE Transactions on Systems, Man and Cybernetics A, 26*(3), 324–339.
Lad, F. (1996). *Operational subjective statistical methods: A mathematical, philosophical, and historical, and introduction*. New York: Wiley.
Levi, I. (1980). *The enterprise of knowledge*. Cambridge, MA: MIT Press.
Moral, S. (2005). Epistemic irrelevance on sets of desirable gambles. *Annals of Mathematics and Artificial Intelligence, 45*(1–2), 197–214.
Moral, S., & Cano, A. (2002). Strong conditional independence for credal sets. *Annals of Mathematics and Artificial Intelligence, 35*, 295–321.
Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.
Schervish, M. J. (1995). *Theory of statistics*. New York: Springer.

Schervish, M., Seidenfeld, T., Kadane, J., & Levi, I. (2003). Extensions of expected utility theory and some limitations of pairwise comparisons. In J.-M. Bernard, T. Seidenfeld, & M. Zaffalon (Eds.), *Proceedings of the third international symposium on imprecise probabilities and their applications* (pp. 496–510).

Seidenfeld, T. (2001). Remarks on the theory of conditional probability: Some issues of finite versus countable additivity. In *Statistics–philosophy, recent history, and relations to science* (pp. 167–177). Dordrecht: Kluwer.

Seidenfeld, T. (2007). Conditional independence, imprecise probabilities, null-events and graph-theoretic models. Seimour Geisser Distinguished Lecture, School of Statistics, University of Minnesota, Minneapolis, MN.

Seidenfeld, T., Schervish, M. J., & Kadane, J. B. (1990). Decisions without ordering. In W. Sieg (Ed.), *Acting and reflecting* (pp. 143–170). Dordrecht: Kluwer.

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

Seidenfeld, T., Schervish, M., & Kadane, J. (2007). Coherent choice functions under uncertainty. In *International symposium on imprecise probability: Theories and applications*, Prague, Czech Republic.

Sen, A. (1977). Social choice theory: A reexamination. *Econometrica, 45*, 53–89.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.

Smith, C. A. B. (1961). Consistency in statistical inference and decision. *Journal Royal Statistical Society B, 23*, 1–25.

Troffaes, M. (2004). Decision making with imprecise probabilities: A short review. *SIPTA Newsletter, 2*, 4–7.

Utkin, L. V., & Augustin, T. (2005). Powerful algorithms for decision making under prior information and general ambiguity attitudes. In *Proceedings of the fourth international symposium on imprecise probabilities and their applications* (pp. 349–358).

Vantaggi, B. (2001). Graphical models for conditional independence structures. In Second international symposium on imprecise probabilities and their applications (pp. 332–341).

Walley, P. (1982). The elicitation and aggregation of beliefs. Technical report 23, Department of Statistics, University of Warwick, Warwick, UK

Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. London: Chapman and Hall.

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

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

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

Williams, P. M. (1975). Coherence, strict coherence and zero probabilities. In *Fifth international congress in logic, methodology and philosophy of science* (Vol. VI, pp. 29–30).

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

Yaghlane, B. B., Smets, P., & Mellouli, K. (2002a). Belief function independence: I. The marginal case. *International Journal of Approximate Reasoning, 29*(1), 47–70.

Yaghlane, B. B., Smets, P., & Mellouli, K. (2002b). Belief function independence: II. The conditional case. *International Journal of Approximate Reasoning, 31*(1–2), 31–75.

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