The Annals of Statistics
2017, Vol. 45, No. 6, 1–36
https://doi.org/10.1214/16-AOS1537

EXTENDED CONDITIONAL INDEPENDENCE AND APPLICATIONS IN CAUSAL INFERENCE

BY PANAYIOTA CONSTANTINOU AND A. PHILIP DAWID

University of Warwick and University of Cambridge

The goal of this paper is to integrate the notions of stochastic conditional independence and variation conditional independence under a more general notion of extended conditional independence. We show that under appropriate assumptions the calculus that applies for the two cases separately (axioms of a separoid) still applies for the extended case. These results provide a rigorous basis for a wide range of statistical concepts, including ancillarity and sufficiency, and, in particular, the Decision Theoretic framework for statistical causality, which uses the language and calculus of conditional independence in order to express causal properties and make causal inferences.

1. Introduction. Conditional independence is a concept that has been widely studied and used in Probability and Statistics. The idea of treating conditional independence as an abstract concept with its own calculus was introduced by Dawid [12], who showed that many results and theorems concerning statistical concepts such as ancillarity, sufficiency, causality, *etc.*, are just applications of general properties of conditional independence—extended to encompass stochastic and nonstochastic variables together. Properties of conditional independence have also been investigated by Spohn [50] in connection with causality, and Pearl and Paz [45], Pearl [44], Geiger et al. [34], Lauritzen et al. [42] in connection with graphical models. For further related theory, see [15–19, 27, 30].

In this paper, we consider two separate concepts of conditional independence: *stochastic conditional independence*, which involves solely stochastic variables, and *variation conditional independence*, which involves solely nonstochastic variables. We argue that, although these concepts are fundamentally different in terms of their mathematical definitions, they share a common intuitive understanding as "irrelevance" relations. This allows them to satisfy the same set of rules (axioms of a separoid [18]). Armed with this insight, we unify the two notions into the more general concept of *extended conditional independence*, and show that (under suitable technical conditions) extended conditional independence also satisfies the axioms of a separoid. This justifies the hitherto informal or implicit application of these axioms in a number of previous works [11, 14, 20, 21, 24–26, 28, 29, 31, 37].

Received December 2015; revised December 2016.
MSC2010 subject classifications. Primary 62A99; secondary 60A05.
Key words and phrases. Conditional independence, extended conditional independence, sufficiency, ancillarity, causality, separoid.

To motivate the need for such a theory, we recall some fundamental concepts of statistics. First, consider the concept of *ancillarity* [33]. Let $\boldsymbol{X} := (X_1, X_2, \ldots, X_n)$ be a random sample from a probability distribution with unknown parameter θ, and let $T = T(\boldsymbol{X})$ be a statistic. T is called an *ancillary statistic for* θ if its distribution does not depend on the value of θ [2]. For example, consider an independent and identically distributed sample $(X_1, X_2, \ldots, X_n)$ from the normal $\mathcal{N}(\theta, 1)$ distribution. Then the range $T := \max\{X_1, X_2, \ldots, X_n\} - \min\{X_1, X_2, \ldots, X_n\}$ is an ancillary statistic, because its distribution does not change as θ changes. Ancillary statistics can be used to recover information lost by reducing the data to the maximum likelihood estimate [36]. For our purposes, we remark that the definition of ancillarity can be understood intuitively as requiring the independence of the stochastic variable T from the nonstochastic variable θ. We can express this property using the now standard (conditional) independence notation introduced by Dawid [12]: $T \perp\!\!\!\perp \theta$.

Another example is the notion of *sufficiency* [33]. With notation as above, T is a *sufficient statistic for* θ if the conditional distribution of the full data $\boldsymbol{X}$, given the value of $T(\boldsymbol{X})$, does not depend on the value of the parameter θ ([6], page 272). For example, consider an independent and identically distributed sample $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ from the Poisson distribution with mean θ. Then the sample total $T = X_1 + X_2 + \cdots + X_n$ is a sufficient statistic for θ, since the distribution of $\boldsymbol{X}$, given $T = t$, is multinomial $\mathcal{M}(t; 1/n, \ldots, 1/n)$ for all θ. Here, we emphasize that sufficiency can be expressed intuitively as: "Given T, $\boldsymbol{X}$ is independent of θ", where $\boldsymbol{X}$ and T are stochastic variables and θ is a nonstochastic variable. Using conditional independence notation: $\boldsymbol{X} \perp\!\!\!\perp \theta \mid T$.

A further application of these ideas emerges from the area of causality: in particular, the Decision Theoretic framework of statistical causality [23]. In this framework, the language and calculus of conditional independence are fundamental for expressing and manipulating causal concepts. The Decision Theoretic framework differentiates between observational and interventional regimes, using a nonstochastic variable to index the regimes. Typically, we consider the regime under which data can be collected (the *observational* regime) and a number of *interventional* regimes that we wish to compare. Since we mostly have access to purely observational data, we focus on extracting information from the observational regime relevant to the interventional regimes. Then the conditions that would justify transfer of information across regimes can be expressed in the language of conditional independence. To illustrate this, suppose we are interested in assessing the effect of a binary treatment variable T on a disease outcome variable Y (*e.g.*, recovery). Denote

$$T = \begin{cases} 0 & \text{for control treatment,} \\ 1 & \text{for active treatment.} \end{cases}$$

We consider three regimes, indexed by a nonstochastic variable Σ:

$$\Sigma = \begin{cases} \varnothing & \text{denotes the observational regime,} \\ 0 & \text{denotes the interventional regime under control treatment,} \\ 1 & \text{denotes the interventional regime under active treatment.} \end{cases}$$

In the simplest case, we might entertain the following (typically unrealistic) property: for either treatment choice $T = 0, 1$, the conditional distribution of the disease variable Y, given the treatment variable T, is the same in the observational and the corresponding interventional regime. We can express this property, using conditional independence notation, as $Y \perp\!\!\!\perp \Sigma \mid T$. Such a property, when it can be taken as valid, would allow us to use the observational regime to make causal inference directly. However, in most cases this assumption will not be easy to defend. Consequently, we would like to explore alternative, more justifiable, conditions, which would allow us to make causal inference. For such exploration, a calculus of extended conditional independence becomes a necessity. Some abstract theory underlying such a calculus was presented in [15]. The current paper develops and extends that theory in a somewhat more concrete setting. Some further detail is presented in [7].

The layout of the paper is as follows. In Section 2, we give the definition of a *separoid*, an algebraic structure with five axioms, and show that stochastic conditional independence and variation conditional independence both satisfy these axioms. In Section 3, we rigorously define *extended conditional independence*, a combined form of stochastic and variation conditional independence, and explore conditions under which extended conditional independence satisfies the separoid axioms, for the most part restricting to cases where the left-most term in an extended conditional independence relation is purely stochastic. In Section 4, we take a Bayesian approach, which allows us to deduce the axioms when the regime space is discrete. Next, using a more direct measure-theoretic approach, we show in Section 5 that the axioms hold when all the stochastic variables are discrete, and likewise in the presence of a dominating regime. In Section 6, we introduce a slight weakening of extended conditional independence, for which the axioms apply without further conditions. Next, Section 7 attempts to extend the analysis to cases where nonstochastic variables appear in the left-most term. Our analysis is put to use in Section 8, which gives some examples of its applications in causal inference, illustrating how extended conditional independence, equipped with its separoid calculus, provides a powerful tool in the area. We conclude in Section 9 with some comments on the usefulness of combining the theory of extended conditional independence with the technology of graphical models.

2. Separoids.

In this section, we describe the algebraic structure called a *separoid* [18]: a three-place relation on a join semilattice, subject to five axioms.

Let V be a set with elements denoted by $x, y, \dots$, and $\leq$ a quasiorder (a reflexive and transitive binary relation) on V. For $x, y \in V$, if $x \leq y$ and $y \leq x$, we say that x and y are *equivalent* and write $x \approx y$. For a subset $A \subseteq V$, z is a *join* of A if $a \leq z$ for all $a \in A$, and it is a minimal element of V with that property; we write $z = \bigvee A$; similarly, z is a *meet* of A ($z = \bigwedge A$) if $z \leq a$ for all $a \in A$, and it is a maximal element of V with that property. We write $x \vee y$ for $\bigvee\{x, y\}$, and $x \wedge y$ for $\bigwedge\{x, y\}$.

Clearly, if z and w are both joins (resp., meets) of A, then $z \approx w$. We call $(V, \leq)$ (or, when $\leq$ is understood, just V) a *join semilattice* if there exists a join for any nonempty finite subset; similarly, $(V, \leq)$ is a *meet semilattice* if there exists a meet for any nonempty finite subset. When $(V, \leq)$ is both a meet and join semilattice, it is a *lattice*.

DEFINITION 2.1 (Separoid). Given a ternary relation $\cdot \perp\!\!\!\perp \cdot \mid \cdot$ on V, we call $\perp\!\!\!\perp$ a *separoid* (on $(V, \leq)$), or the triple $(V, \leq, \perp\!\!\!\perp)$ a *separoid*, if:

S1: $(V, \leq)$ is a join semilattice

and

P1: $x \perp\!\!\!\perp y \mid z \Rightarrow y \perp\!\!\!\perp x \mid z$
P2: $x \perp\!\!\!\perp y \mid y$
P3: $x \perp\!\!\!\perp y \mid z$ and $w \leq y \Rightarrow x \perp\!\!\!\perp w \mid z$
P4: $x \perp\!\!\!\perp y \mid z$ and $w \leq y \Rightarrow x \perp\!\!\!\perp y \mid (z \vee w)$
P5: $x \perp\!\!\!\perp y \mid z$ and $x \perp\!\!\!\perp w \mid (y \vee z) \Rightarrow x \perp\!\!\!\perp (y \vee w) \mid z$.

The following lemma shows that, in P4 and P5, the choice of join does not change the property.

LEMMA 2.1. *Let $(V, \leq, \perp\!\!\!\perp)$ be a separoid and $x_i, y_i, z_i \in V$ $(i = 1, 2)$ with $x_1 \approx x_2$, $y_1 \approx y_2$ and $z_1 \approx z_2$. If $x_1 \perp\!\!\!\perp y_1 \mid z_1$, then $x_2 \perp\!\!\!\perp y_2 \mid z_2$.*

PROOF. See Corollary 1.2 in [18]. □

DEFINITION 2.2 (Strong separoid). We say that the triple $(V, \leq, \perp\!\!\!\perp)$ is a *strong separoid* if we strengthen S1 in Definition 2.1 to

S1′: $(V, \leq)$ is a lattice

and in addition to P1–P5 we require

P6: if $z \leq y$ and $w \leq y$, then $x \perp\!\!\!\perp y \mid z$ and $x \perp\!\!\!\perp y \mid w \Rightarrow x \perp\!\!\!\perp y \mid (z \wedge w)$.

2.1. *Stochastic conditional independence as a separoid.* The concept of stochastic conditional independence is a familiar example of a separoid (though not, without further conditions, a strong separoid [13]).

Let $(\Omega, \mathcal{A})$, $(F, \mathcal{F})$ be measurable spaces, and $Y : \Omega \to F$ a random variable. We denote by $\sigma(Y)$ the σ-algebra generated by Y, that is, $\{Y^{-1}(C) : C \in \mathcal{F}\}$. We write $Y : (\Omega, \mathcal{A}) \to (F, \mathcal{F})$ to imply that Y is measurable with respect to the σ-algebras $\mathcal{A}$ and $\mathcal{F}$; equivalently, $\sigma(Y) \subseteq \mathcal{A}$.

LEMMA 2.2. *Let $Y : (\Omega, \sigma(Y)) \to (F_Y, \mathcal{F}_Y)$ and $Z : (\Omega, \sigma(Z)) \to (F_Z, \mathcal{F}_Z)$ be surjective random variables. Suppose that $\mathcal{F}_Y$ contains all singleton sets $\{y\}$. Then the following are equivalent*:

(i) $\sigma(Y) \subseteq \sigma(Z)$.
(ii) *There exists measurable $f : (F_Z, \mathcal{F}_Z) \to (F_Y, \mathcal{F}_Y)$ such that $Y = f(Z)$.*

PROOF. See the online supplementary material [8]. □

In the sequel, whenever we invoke Lemma 2.2 we shall implicitly assume that its conditions are satisfied. In most of our applications of Lemma 2.2, both $(F_Y, \mathcal{F}_Y)$ and $(F_Z, \mathcal{F}_Z)$ will be the real or extended real line equipped with its Borel σ-algebra $\mathcal{B}$.

We recall Kolmogorov's definition of conditional expectation ([4], page 445).

DEFINITION 2.3 (Conditional expectation). Let X be an integrable real-valued random variable defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let $\mathcal{G} \subseteq \mathcal{A}$ be a σ-algebra. A random variable Y is called (a version of) the *conditional expectation of X given $\mathcal{G}$*, and we write $Y = \mathbb{E}(X \mid \mathcal{G})$, if

(i) Y is $\mathcal{G}$-measurable; and
(ii) Y is integrable and $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(Y\mathbb{1}_A)$ for all $A \in \mathcal{G}$.

When $\mathcal{G} = \sigma(Z)$, we may write $\mathbb{E}(X \mid Z)$ for $\mathbb{E}(X \mid \mathcal{G})$.

It can be shown that $\mathbb{E}(X \mid \mathcal{G})$ exists, and any two versions of it are almost surely equal ([4], page 445). In particular, if X is $\mathcal{G}$-measurable then $\mathbb{E}(X \mid \mathcal{G}) = X$ a.s. Thus for any integrable function $f(X)$, $\mathbb{E}\{f(X) \mid X\} = f(X)$ a.s. Also by (ii) for $A = \Omega$, $\mathbb{E}(X) = \mathbb{E}\{\mathbb{E}(X \mid \mathcal{G})\}$. We will use this in the form $\mathbb{E}(X) = \mathbb{E}\{\mathbb{E}(X \mid Y)\}$.

DEFINITION 2.4 (Conditional independence). Let X, Y, Z be random variables on $(\Omega, \mathcal{A}, \mathbb{P})$. We say that *$X$ is (conditionally) independent of Y given Z (with respect to $\mathbb{P}$)* and write $X \perp\!\!\!\perp_s Y \mid Z$ $[\mathbb{P}]$, or just $X \perp\!\!\!\perp_s Y \mid Z$ when $\mathbb{P}$ is understood, if:

$$\text{For all } A_X \in \sigma(X), \qquad \mathbb{E}(\mathbb{1}_{A_X} \mid Y, Z) = \mathbb{E}(\mathbb{1}_{A_X} \mid Z) \qquad \text{a.s. } [\mathbb{P}].$$

We refer to the above property as *stochastic conditional independence*; we use the subscript s under $\perp\!\!\!\perp$ ($\perp\!\!\!\perp_s$) to emphasize that we refer to this stochastic definition.

To prove the separoid axioms, we need equivalent forms of the above definition.

PROPOSITION 2.3. *Let* X, Y, Z *be random variables on* $(\Omega, \mathcal{A}, \mathbb{P})$. *Then the following are equivalent*:

(i) $X \perp\!\!\!\perp_s Y \mid Z$.

(ii) *For all real, bounded and measurable functions* $f(X)$, $\mathbb{E}\{f(X) \mid Y, Z\} = \mathbb{E}\{f(X) \mid Z\}$ a.s.

(iii) *For all real, bounded and measurable functions* $f(X)$, $g(Y)$, $\mathbb{E}\{f(X)g(Y) \mid Z\} = \mathbb{E}\{f(X) \mid Z\}\mathbb{E}\{g(Y) \mid Z\}$ *a.s.*

(iv) *For all* $A_X \in \sigma(X)$ *and all* $A_Y \in \sigma(Y)$, $\mathbb{E}(\mathbb{1}_{A_X \cap A_Y} \mid Z) = \mathbb{E}(\mathbb{1}_{A_X} \mid Z)\mathbb{E}(\mathbb{1}_{A_Y} \mid Z)$ *a.s.*

PROOF. See [5], page 42. □

Henceforth, we write $X \preceq Y$ when $X = f(Y)$ for some measurable function f.

THEOREM 2.4 (Axioms of conditional independence). *Let* X, Y, Z, W *be random variables on* $(\Omega, \mathcal{A}, \mathbb{P})$. *Then the following properties hold* (*the descriptive terms are those assigned by Pearl* [43]):

P1^s (*Symmetry*): $X \perp\!\!\!\perp_s Y \mid Z \Rightarrow Y \perp\!\!\!\perp_s X \mid Z$.
P2^s: $X \perp\!\!\!\perp_s Y \mid Y$.
P3^s (*Decomposition*): $X \perp\!\!\!\perp_s Y \mid Z$ *and* $W \preceq Y \Rightarrow X \perp\!\!\!\perp_s W \mid Z$.
P4^s (*Weak Union*): $X \perp\!\!\!\perp_s Y \mid Z$ *and* $W \preceq Y \Rightarrow X \perp\!\!\!\perp_s Y \mid (W, Z)$.
P5^s (*Contraction*): $X \perp\!\!\!\perp_s Y \mid Z$ *and* $X \perp\!\!\!\perp_s W \mid (Y, Z) \Rightarrow X \perp\!\!\!\perp_s (Y, W) \mid Z$.

PROOF. See the online supplementary material [8]. □

Theorem 2.4 shows that stochastic conditional independence satisfies the axioms of a separoid. Denoting by V the set of all random variables defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and equipping V with the quasiorder $\preceq$, $(V, \preceq)$ becomes a join semilattice and the triple $(V, \preceq, \perp\!\!\!\perp)$ is then a separoid.

Using stochastic conditional independence in an axiomatic way, we can mechanically prove many useful conditional independence results.

EXAMPLE 2.1. Let X, Y, Z be random variables on $(\Omega, \mathcal{A}, \mathbb{P})$. Then $X \perp\!\!\!\perp_s Y \mid Z$ implies that $(X, Z) \perp\!\!\!\perp_s Y \mid Z$.

PROOF. Applying P1^s to $X \perp\!\!\!\perp_s Y \mid Z$, we obtain

$$Y \perp\!\!\!\perp_s X \mid Z. \tag{2.1}$$

By P2^s, we obtain

$$Y \perp\!\!\!\perp_s (X, Z) \mid (X, Z). \tag{2.2}$$

Applying P3^s to (2.2), we obtain

$$Y \perp\!\!\!\perp_s Z \mid (X, Z) \tag{2.3}$$

and applying P5^s to (2.1) and (2.3), we obtain

$$Y \perp\!\!\!\perp_s (X, Z) \mid Z. \tag{2.4}$$

The result follows by applying P1^s to (2.4). □

EXAMPLE 2.2 (Nearest Neighbour Property of a Markov Chain). Let X_1, X_2, X_3, X_4, X_5 be random variables on $(\Omega, \mathcal{A}, \mathbb{P})$ and suppose that:

(i) $X_3 \perp\!\!\!\perp_s X_1 \mid X_2$,
(ii) $X_4 \perp\!\!\!\perp_s (X_1, X_2) \mid X_3$,
(iii) $X_5 \perp\!\!\!\perp_s (X_1, X_2, X_3) \mid X_4$.

Then $X_3 \perp\!\!\!\perp_s (X_1, X_5) \mid (X_2, X_4)$.

PROOF. See [12]. □

2.2. *Variation conditional independence as a separoid.* Variation conditional independence, which concerns solely nonstochastic variables, is another, indeed much simpler, example of a separoid.

Let $\mathcal{S}$ be a set with elements denoted by, for example, σ, and let V be the set of all functions with domain $\mathcal{S}$ and arbitrary range space. The elements of V will be denoted by, for example, $X, Y, \ldots$. We do not require any additional properties or structure such as a probability measure, measurability, *etc.* We write $X \preceq Y$ to denote that X is a function of Y, that is, $Y(\sigma_1) = Y(\sigma_2) \Rightarrow X(\sigma_1) = X(\sigma_2)$. The equivalence classes for this quasiorder correspond to partitions of $\mathcal{S}$. Then $(V, \preceq)$ forms a join semilattice, with join $X \vee Y$ the function $(X, Y) \in V$.

The *(unconditional) image of* Y is $R(Y) := Y(\mathcal{S}) = \{Y(\sigma) : \sigma \in \mathcal{S}\}$. The *conditional image of* X*, given* $Y = y$ is $R(X \mid Y = y) := \{X(\sigma) : \sigma \in \mathcal{S}, Y(\sigma) = y\}$. For simplicity of notation, we will sometimes write $R(X \mid y)$ instead of $R(X \mid Y = y)$, and $R(X \mid Y)$ for the function $R(X \mid Y = \cdot)$.

DEFINITION 2.5. We say that X *is variation (conditionally) independent of* Y *given* Z (on Ω) and write $X \perp\!\!\!\perp_v Y \mid Z\ [\mathcal{S}]$ (or, if $\mathcal{S}$ is understood, just $X \perp\!\!\!\perp_v Y \mid Z$) if:

$$\text{for any } (y, z) \in R(Y, Z), \qquad R(X \mid y, z) = R(X \mid z).$$

We use the subscript v under $\perp\!\!\!\perp$ ($\perp\!\!\!\perp_v$) to emphasize that we refer to the above nonstochastic definition. In parallel with the stochastic case, we have equivalent forms of the above definition.

PROPOSITION 2.5. *The following are equivalent*:

(i) $X \perp\!\!\!\perp_v Y \mid Z$.
(ii) *The function* $R(X \mid Y, Z)$ *of* (Y, Z) *is a function of* Z *alone*.
(iii) *For any* $z \in R(Z)$, $R(X, Y \mid z) = R(X \mid z) \times R(Y \mid z)$.

PROOF. See the online supplementary material [8]. □

PROPOSITION 2.6. *The following are equivalent*:

(i) $W \preceq Y$.
(ii) *there exists* $f : R(Y) \to R(W)$ *such that* $W = f(Y)$.

PROOF. See the online supplementary material [8]. □

THEOREM 2.7 (Axioms of variation independence). *Let* X, Y, Z, W *be functions on* $\mathcal{S}$. *Then the following properties hold*:

P1^v: $X \perp\!\!\!\perp_v Y \mid Z \Rightarrow Y \perp\!\!\!\perp_v X \mid Z$.
P2^v: $X \perp\!\!\!\perp_v Y \mid Y$.
P3^v: $X \perp\!\!\!\perp_v Y \mid Z$ *and* $W \preceq Y \Rightarrow X \perp\!\!\!\perp_v W \mid Z$.
P4^v: $X \perp\!\!\!\perp_v Y \mid Z$ *and* $W \preceq Y \Rightarrow X \perp\!\!\!\perp_v Y \mid (W, Z)$.
P5^v: $X \perp\!\!\!\perp_v Y \mid Z$ *and* $X \perp\!\!\!\perp_v W \mid (Y, Z) \Rightarrow X \perp\!\!\!\perp_v Y \mid (W, Z)$.

PROOF. See the online supplementary material [8]. □

The above theorem shows that variation independence satisfies the axioms of a separoid. Indeed—and in contrast with stochastic conditional independence—variation independence also satisfies the axioms of a strong separoid [19].

3. Extended conditional independence. There are numerous contexts in which it would appear fruitful to merge the concepts of stochastic conditional independence and variation conditional independence, allowing both stochastic and nonstochastic variables to appear together.

EXAMPLE 3.1 (Inference in the presence of nuisance parameters [10, 14]). Let X have a distribution governed by a parameter Σ, and let T be a function of X. We call T a *cut* in this model [1] if there exists a pair of parameters Θ, Φ (functions of Σ), such that:

(i) The marginal distribution of T depends only on Θ

(ii) The conditional distributions of X, given T, depend only on Φ
(iii) Θ and Φ are variation independent.

In this case, T has been termed *S-sufficient* for Θ, and, it has been suggested, inference about Θ might reasonably be based on the marginal distribution of T, which by (i) does only involve Θ. The justification for ignoring any information obtainable from the neglected aspect of the model, namely the conditional distribution of X given T, is based on the argument that, by (ii), this is directly informative only about Φ, and so by (iii) supplies no information "logically relevant" to Θ.

The above definition could be rephrased, intuitively, in terms of conditional independence:

(i) $T \perp\!\!\!\perp \Phi \mid \Theta$
(ii) $X \perp\!\!\!\perp \Theta \mid (T, \Phi)$
(iii) $\Theta \perp\!\!\!\perp \Phi$.

Now a Bayesian, who regards (Θ, Φ) as random, could take $\perp\!\!\!\perp = \perp\!\!\!\perp_s$, so interpreting (i) and (ii) straightforwardly as stochastic independence properties. Such an interpretation for (iii) would however require *a priori* independence of Θ and Φ, which is a much stronger property than variation independence. When this can be assumed, we can apply the separoid properties of stochastic independence to derive

$$\Theta \perp\!\!\!\perp (X, \Phi) \mid T. \tag{3.1}$$

This in turn is equivalent to

$$\Theta \perp\!\!\!\perp X \mid T, \tag{3.2}$$

$$\Theta \perp\!\!\!\perp \Phi \mid X. \tag{3.3}$$

With this stochastic interpretation, (3.2) affirms that the Bayesian's posterior distribution of Θ, given the full data X, is the same as that based on T alone; while (3.3) shows that the prior independence of Θ and Φ is preserved in their joint posterior distribution.

What though of the non-Bayesian, who wishes to derive the consequences of (i)–(iii) without assigning stochastic status to the parameters? This requires an extended interpretation of conditional independence. We shall clarify and rigorise such an extended interpretation of the statements (i) and (ii) in Definition 3.2 below.

Statement (iii) is now to be interpreted as variation independence. Theory addressing the combination of this with the other statements, and their interpretation, will be introduced in Section 7: this will give meaning to, and justify, the conclusions (3.1) and (3.3). Conclusion (3.2), however, which might be interpreted, informally, as justifying the basing of any inference about Θ on T alone, is not justified by our analysis; indeed its very interpretation remains in need of further clarification.

There is a basic intuitive similarity between the notions of stochastic conditional independence and variation independence. A statement like $X \perp\!\!\!\perp_s Y \mid Z$ for stochastic variables, or $X \perp\!\!\!\perp_v Y \mid Z$ for nonstochastic variables, reflects our informal understanding that, having already obtained information about Z, further information about Y will not affect the uncertainty (suitably understood) about X. Building on this intuitive interpretation, one can extend $X \perp\!\!\!\perp_s Y \mid Z$ to the case that one or both of Y and Z involve nonstochastic variables, such as parameters or regime indicators. Such an extended version of conditional independence would embrace the notions of ancillarity, sufficiency, causality, *etc.*

The first authors to consider sufficiency in a general abstract setting were Halmos and Savage [39]. Removing any assumption such as the existence of a probability mass function or a density with respect to a common measure, sufficiency is defined as follows.

DEFINITION 3.1 (Sufficiency). Consider a a random variable X, and a family $\mathcal{P} = \{\mathbb{P}_\theta\}$ of probability distributions for X, indexed by $\theta \in \boldsymbol{\Theta}$. A statistic $T = T(X)$ is *sufficient* for $\mathcal{P}$, or for θ, if for any real, bounded and measurable function h, there exists a function $w(T)$ such that, for any $\theta \in \Theta$,

$$\mathbb{E}_\theta\{h(X) \mid T\} = w(T) \qquad \text{a.s. } [\mathbb{P}_\theta].$$

Interpreting the definition carefully, we require that, for any real, bounded and measurable $h(X)$, there exist a single function $w(T)$ that serves as a version of the conditional expectation $\mathbb{E}_\theta\{h(X) \mid T\}$ under $\mathbb{P}_\theta$, simultaneously for all $\theta \in \Theta$.

In the Decision Theoretic framework, we consider, instead of the parameter space $\boldsymbol{\Theta}$, a space $\mathcal{S}$ of different regimes, typically σ,[1] under which data can be observed. We thus consider a family $\mathcal{P} = \{\mathbb{P}_\sigma : \sigma \in \mathcal{S}\}$ of probability measures over a suitable space $(\Omega, \mathcal{A})$. A stochastic variable, such as $X : (\Omega, \sigma(X)) \to (\mathbb{R}, \mathcal{B})$, can have different distributions under the different regimes $\sigma \in \mathcal{S}$. We write $\mathbb{E}_\sigma(X \mid Y)$ to denote a version of the conditional expectation $\mathbb{E}(X \mid Y)$ under regime σ: this is defined a.s. $[\mathbb{P}_\sigma]$. We also consider nonstochastic variables, functions defined on $\mathcal{S}$, which we term *decision variables*. Decision variables give us full or partial information about which regime is operating. For the rest of the paper, we denote by Σ the identity function on $\mathcal{S}$.

We aim to extend Definition 3.1 to express a statement like $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, where X, Y, Z are stochastic variables and Θ, Φ decision variables. In order to formalise such a statement, we first describe what we would like a conditional independence statement like $X \perp\!\!\!\perp \Theta \mid \Phi$ to reflect intuitively: that the distribution of X, given the information carried by (Θ, Φ) jointly about which

[1]The regime indicator σ is not to be confused with the σ-algebra generated by X, denoted by $\sigma(X)$.

regime is operating, is in fact fully determined by the value of Φ alone. However, in order for this to make sense, we must assume that Φ and Θ together do fully determine the regime $\sigma \in S$ operating, and thus, the distribution of X in this regime. Formally, we require that the function (Θ, Φ) defined on $\mathcal{S}$ be an injection: if $\Theta(\sigma_1) = \Theta(\sigma_2)$ and $\Phi(\sigma_1) = \Phi(\sigma_2)$, then $\sigma_1 = \sigma_2$. (Note that, unless explicitly stated, we do not require that Θ and Φ be variation independent). In this case, we say that Φ and Θ are *complementary* (on $\mathcal{S}$), or that Θ *is complementary to* Φ (on $\mathcal{S}$). The property of complementarity extends in an obvious way to more than two decision variables.

This leads to the following definition.

DEFINITION 3.2. Let X, Y and Z be stochastic variables, and let Φ and Θ be complementary decision variables. We say that X *is (conditionally) independent of* (Y, Θ) *given* (Z, Φ) and write $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ if, for all $\phi \in \Phi(\mathcal{S})$ and all real, bounded and measurable functions h, there exists a function $w_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma\{h(X) \mid Y, Z\} = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{3.4}$$

We will refer to this definition of conditional independence as *extended conditional independence*. Note that the only important property of Θ in the above definition is that it be complementary to Φ; beyond this, the actual form of Θ becomes irrelevant (we could even take $\Theta = \Sigma$). Henceforth, we will write down a conditional independence statement involving decision variables only when those variables are complementary.

REMARK 3.1. Assume that $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ and consider $w_\phi(Z)$ as in Definition 3.2. Then

$$\begin{aligned}
\mathbb{E}_\sigma\{h(X) \mid Z\} &= \mathbb{E}_\sigma\big[\mathbb{E}_\sigma\{h(X) \mid Y, Z\} \mid Z\big] \qquad \text{a.s. } \mathbb{P}_\sigma \\
&= \mathbb{E}_\sigma\{w_\phi(Z) \mid Z\} \qquad \text{a.s. } \mathbb{P}_\sigma \\
&= w_\phi(Z) \qquad \text{a.s. } \mathbb{P}_\sigma.
\end{aligned}$$

Thus, $w_\phi(Z)$ also serves as a version of $\mathbb{E}_\sigma\{h(X) \mid Z\}$ for all $\sigma \in \Phi^{-1}(\phi)$.

The following example shows that, even when (3.4) holds, we cannot use just any version of the conditional expectation in one regime to serve as a version of the conditional expectation in another regime. This is because two versions of the conditional expectation can differ on a set of probability zero, but a set of probability zero in one regime could have positive probability in another.

EXAMPLE 3.2. Let the regime space be $S = \{\sigma_0, \sigma_1\}$, let binary variable T represent the treatment taken (where $T = 0$ denotes placebo and $T = 1$ denotes active treatment), and let X be an outcome of interest. Regime σ_t ($t = 0, 1$) represents the interventional regime under treatment t: in particular, $\mathbb{P}_{\sigma_j}(T = j) = 1$.

We consider the situation where the treatment is ineffective, so that X has the same distribution in both regimes. We then have $X \perp\!\!\!\perp \Sigma \mid T$—since, for any function $h(X)$, we can take as $\mathbb{E}_\sigma\{h(X) \mid T\}$, for both σ_0 and σ_1, the (constant) common expectation of $h(X)$ in both regimes.[2]

In particular, suppose X has expectation 1 in both regimes. Then the function $w(T) \equiv 1$ is a version both of $\mathbb{E}_{\sigma_0}\{h(X) \mid T\}$ and of $\mathbb{E}_{\sigma_1}\{h(X) \mid T\}$. That is, $\mathbb{E}_{\sigma_0}(X \mid T) = 1$ almost surely $[\mathbb{P}_{\sigma_0}]$, and $\mathbb{E}_{\sigma_1}(X \mid T) = 1$ almost surely $[\mathbb{P}_{\sigma_1}]$.

Now consider the functions:

$$k_0(t) = 1 - t \quad \text{and} \quad k_1(t) = t.$$

We can see that $k_0(T) = w(T)$ almost surely $[\mathbb{P}_{\sigma_0}]$, so that $k_0(T)$ is a version of $\mathbb{E}_{\sigma_0}\{h(X) \mid T\}$; similarly, $k_1(T)$ is a version of $\mathbb{E}_{\sigma_1}\{h(X) \mid T\}$. However, almost surely, under both $\mathbb{P}_{\sigma_0}$ and $\mathbb{P}_{\sigma_1}$, $k_0(T) \neq k_1(T)$. Hence, neither of these variables can replace $w(T)$ in supplying a version of $\mathbb{E}_\sigma(X \mid T)$ simultaneously in both regimes.

We now introduce some equivalent versions of Definition 3.2.

PROPOSITION 3.1. *Let X, Y, Z be stochastic variables and let Φ, Θ be complementary decision variables. Then the following are equivalent*:

(i) $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$.

(ii) *For all $\phi \in \Phi(\mathcal{S})$ and all real, bounded and measurable function h_1, there exists a function $w_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$ and all real, bounded and measurable functions h_2,*

$$\mathbb{E}_\sigma\{h_1(X)h_2(Y) \mid Z\} = w_\phi(Z)\mathbb{E}_\sigma\{h_2(Y) \mid Z\} \qquad a.s.\ [\mathbb{P}_\sigma].$$

(iii) *For all $\phi \in \Phi(\mathcal{S})$ and all $A_X \in \sigma(X)$, there exists a function $w_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$ and all $A_Y \in \sigma(Y)$,*

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X \cap A_Y} \mid Z) = w_\phi(Z)\mathbb{E}_\sigma(\mathbb{1}_{A_Y} \mid Z) \qquad a.s.\ [\mathbb{P}_\sigma].$$

(iv) *For all $\phi \in \Phi(\mathcal{S})$ and all $A_X \in \sigma(X)$, there exists a function $w_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,*

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, Z) = w_\phi(Z) \qquad a.s.\ [\mathbb{P}_\sigma]. \tag{3.5}$$

PROOF. See the online supplementary material [8]. $\square$

Using Proposition 3.1, we can obtain further properties of extended conditional independence. For example, we can show that Definition 3.2 can be equivalently

[2]Indeed, this encapsulates the still stronger property $X \perp\!\!\!\perp (\Sigma, T)$.

expressed in two simpler statements of extended conditional independence, or that when all the decision variables are confined to the right-most term symmetry does follow. In Section 3.1, we will show still more properties.

PROPOSITION 3.2. *Let X, Y, Z be stochastic variables and Φ, Θ complementary decision variables. Then the following are equivalent*:

(i) $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$
(ii) $X \perp\!\!\!\perp Y \mid (Z, \Phi, \Theta)$ *and* $X \perp\!\!\!\perp \Theta \mid (Z, \Phi)$.

PROOF. (i) $\Rightarrow$ (ii). Since $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, for all $\phi \in \Phi(\mathcal{S})$ and $A_X \in \sigma(X)$, there exists $w_\phi(Z)$ such that for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma(1_{A_X} \mid Y, Z) = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]$$

which proves that $X \perp\!\!\!\perp Y \mid (Z, \Phi, \Theta)$. Also, by Remark 3.1,

$$\mathbb{E}_\sigma(1_{A_X} \mid Z) = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]$$

which proves that $X \perp\!\!\!\perp \Theta \mid (Z, \Phi)$.

(ii) $\Rightarrow$ (i). Since $X \perp\!\!\!\perp Y \mid (Z, \Phi, \Theta)$, for all $\sigma \in \mathcal{S}$ and $A_X \in \sigma(X)$, there exists $w_\sigma(Z)$ such that

$$\mathbb{E}_\sigma(1_{A_X} \mid Y, Z) = w_\sigma(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{3.6}$$

By Remark 3.1,

$$\mathbb{E}_\sigma(1_{A_X} \mid Z) = w_\sigma(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{3.7}$$

Since $X \perp\!\!\!\perp \Theta \mid (Z, \Phi)$, for all $\phi \in \Phi(\mathcal{S})$ and $A_X \in \sigma(X)$ there exists $w_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma(1_{A_X} \mid Z) = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{3.8}$$

By (3.7) and (3.8),

$$w_\sigma(Z) = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma].$$

Thus, by (3.6),

$$\mathbb{E}_\sigma(1_{A_X} \mid Y, Z) = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma],$$

which proves that $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$. $\square$

PROPOSITION 3.3. *Let X, Y, Z be stochastic variables, and Σ the regime indicator. Then $X \perp\!\!\!\perp Y \mid (Z, \Sigma)$ if and only if $X \perp\!\!\!\perp_s Y \mid Z$ under $\mathbb{P}_\sigma$ for all $\sigma \in \mathcal{S}$.*

PROOF. Follows from Proposition 2.3 and Proposition 3.1. $\square$

COROLLARY 3.4. $X \perp\!\!\!\perp Y \mid (Z, \Sigma) \Rightarrow Y \perp\!\!\!\perp X \mid (Z, \Sigma)$.

3.1. *Some separoid properties*. Comparing Definition 3.2 for extended conditional independence with Definition 2.4 for stochastic conditional independence, we observe a close technical, as well as intuitive, similarity. This suggests that these two concepts should have similar properties, and motivates the conjecture that the separoid axioms of conditional independence will continue to hold for the extended concept. In this section, we show that this is indeed so, in complete generality, for a subset of the axioms. However, in order to extend this to other axioms we need to impose additional conditions; this we shall develop in later sections.

One important difference between extended conditional independence and stochastic conditional independence concerns the symmetry axiom P1. Whereas symmetry holds universally for stochastic conditional independence, its application to extended conditional independence is constrained by the fact that, for Definition 3.2 even to make sense, the first term x in an extended conditional independence relation of the form $x \perp\!\!\!\perp y \mid z$ must be fully stochastic, whereas the second term y can contain a mixture of stochastic and nonstochastic variables—in which case it would make no sense to interchange x and y. This restricted symmetry also means that each of the separoid axioms P2–P5 has a possibly nonequivalent "mirror image" version, obtained by interchanging the first and second terms in each relation. Yet another restriction is that the decision variables featuring in any extended conditional independence assertion must be complementary.

The following theorem demonstrates certain specific versions of the separoid axioms.

THEOREM 3.5. *Let X, Y, Z, W be stochastic variables, Φ, Θ complementary decision variables, and Σ the regime indicator. Then the following properties hold*:

P1′: $X \perp\!\!\!\perp Y \mid (Z, \Sigma) \Rightarrow Y \perp\!\!\!\perp X \mid (Z, \Sigma)$.
P2′: $X \perp\!\!\!\perp (Y, \Sigma) \mid (Y, \Sigma)$.
P3′: $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ *and* $W \preceq Y \Rightarrow X \perp\!\!\!\perp (W, \Theta) \mid (Z, \Phi)$.
P4′: $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ *and* $W \preceq Y \Rightarrow X \perp\!\!\!\perp (Y, \Theta) \mid (Z, W, \Phi)$.
P4*a*′: $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ *and* $\Lambda \preceq \Theta \Rightarrow X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi, \Lambda)$.
P5′: $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ *and* $X \perp\!\!\!\perp W \mid (Y, Z, \Theta, \Phi) \Rightarrow X \perp\!\!\!\perp (Y, W, \Theta) \mid (Z, \Phi)$.

PROOF.

P1′. Proved in Proposition 3.3.
P2′. Let $\sigma \in \mathcal{S}$ and $A_X \in \sigma(X)$. Then for all $A_Y \in \sigma(Y)$,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X \cap A_Y} \mid Y) = \mathbb{1}_{A_Y} \mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y) \qquad \text{a.s. } [\mathbb{P}_\sigma]$$

which completes the proof.
P3′. Let $\phi \in \Phi(\mathcal{S})$ and $A_X \in \sigma(X)$. Since $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, there exists $w_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, Z) = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma].$$

Since $W \preceq Y$, it follows from Lemma 2.2 that $\sigma(W) \subseteq \sigma(Y)$, and thus $\sigma(W, Z) \subseteq \sigma(Y, Z)$. Then

$$\begin{aligned}\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid W, Z) &= \mathbb{E}_\sigma\{\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, Z) \mid W, Z\} \qquad \text{a.s. } [\mathbb{P}_\sigma] \\ &= \mathbb{E}_\sigma\{w_\phi(Z) \mid W, Z\} \qquad \text{a.s. } [\mathbb{P}_\sigma] \\ &= w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]\end{aligned}$$

which completes the proof.

P4′. Let $\phi \in \Phi(\mathcal{S})$ and $A_X \in \sigma(X)$. Since $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, there exists $w_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, Z) = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma].$$

Since $W \preceq Y$, it follows from Lemma 2.2 that $\sigma(W) \subseteq \sigma(Y)$, and thus $\sigma(Y, Z, W) = \sigma(Y, Z)$. Then

$$\begin{aligned}\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, Z, W) &= \mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, Z) \qquad \text{a.s. } [\mathbb{P}_\sigma] \\ &= w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]\end{aligned}$$

which completes the proof.

P4*a*′. Follows readily from Definition 3.2.

P5′. Let $\phi \in \Phi(\mathcal{S})$ and $A_X \in \sigma(X)$. Since $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, there exists $w_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, Z) = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma].$$

Since $W \preceq Y$, it follows from Lemma 2.2 that $\sigma(W) \subseteq \sigma(Y)$, and thus $\sigma(Y, W, Z) = \sigma(Y, Z)$. Then

$$\begin{aligned}\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, W, Z) &= \mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, Z) \qquad \text{a.s. } [\mathbb{P}_\sigma] \\ &= w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]\end{aligned}$$

which completes the proof. □

Lack of symmetry introduces some complications as the mirror image variants of axioms P3′, P4′ and P5′ do not automatically follow.

Consider the following statements, which mirror P3′–P5′:

P3″: $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ and $W \preceq X \Rightarrow W \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$.

P4″: $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ and $W \preceq X \Rightarrow X \perp\!\!\!\perp (Y, \Theta) \mid (Z, W, \Phi)$.

P5″: $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ and $W \perp\!\!\!\perp (Y, \Theta) \mid (X, Z, \Phi) \Rightarrow (X, W) \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$.

P3″ follows straightforwardly, and P5″ will be proved to hold in Proposition 3.7 below. However, P4″ presents some difficulty.

LEMMA 3.6. *Let X, Y, Z, W be stochastic variables and Φ, Θ be complementary decision variables. Then*

$$X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi) \quad \textit{and} \quad W \preceq X \Rightarrow (W, Z) \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi).$$

PROOF. Since $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, for all $\phi \in \Phi(\mathcal{S})$ and all $A_X \in \sigma(X)$ there exists $w_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, Z) = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{3.9}$$

To prove that $(W, Z) \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, let $\phi \in \Phi(\mathcal{S})$ and $A_{W,Z} \in \sigma(W, Z)$. We will show that there exists $a_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_{W,Z}} \mid Y, Z) = a_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{3.10}$$

Consider

$$\mathcal{D} = \{A_{W,Z} \in \sigma(W, Z) : \text{there exists } a_\phi(Z) \text{ such that (3.10) holds}\}$$

and

$$\Pi = \{A_{W,Z} \in \sigma(W, Z) : A_{W,Z} = A_W \cap A_Z \text{ for } A_W \in \sigma(W) \text{ and } A_Z \in \sigma(Z)\}.$$

Then $\sigma(\Pi) = \sigma(W, Z)$ ([46], page 73). We will show that $\mathcal{D}$ is a d-system that contains Π. We can then apply Dynkin's lemma ([4], page 42), to conclude that $\mathcal{D}$ contains $\sigma(\Pi) = \sigma(W, Z)$.

To show that $\mathcal{D}$ contains Π, let $A_{W,Z} = A_W \cap A_Z$ with $A_W \in \sigma(W)$ and $A_Z \in \sigma(Z)$. Then

$$\begin{aligned}\mathbb{E}_\sigma(\mathbb{1}_{A_W}\mathbb{1}_{A_Z} \mid Y, Z) &= \mathbb{1}_{A_Z}\mathbb{E}_\sigma(\mathbb{1}_{A_W} \mid Y, Z) \qquad \text{a.s. } [\mathbb{P}_\sigma] \\ &= \mathbb{1}_{A_Z} w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma] \text{ by (3.9)}.\end{aligned}$$

Now define $a_\phi(Z) := \mathbb{1}_{A_Z} w_\phi(Z)$ and we are done.

To show that $\mathcal{D}$ is a d-system, first note that $\Omega \in \mathcal{D}$. Also, for $A_1, A_2 \in \mathcal{D}$ such that $A_1 \subseteq A_2$, we can readily see that $A_2 \setminus A_1 \in \mathcal{D}$. Now consider an increasing sequence $(A_n : n \in \mathbb{N})$ in $\mathcal{D}$ and denote by $a_\phi^{A_n}(Z)$ the corresponding function such that (3.10) holds. Then $A_n \uparrow \bigcup_n A_n$ and $\mathbb{1}_{A_n} \uparrow \mathbb{1}_{\bigcup_n A_n}$ pointwise. Thus, by conditional monotone convergence ([32], page 193),

$$\begin{aligned}\mathbb{E}_\sigma(\mathbb{1}_{\bigcup_n A_n} \mid Y, Z) &= \lim_{n\to\infty} \mathbb{E}_\sigma(\mathbb{1}_{A_n} \mid Y, Z) \qquad \text{a.s. } [\mathbb{P}_\sigma] \\ &= \lim_{n\to\infty} a_\phi^{A_n}(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma].\end{aligned}$$

Now define $a_\phi(Z) := \lim_{n\to\infty} a_\phi^{A_n}(Z)$ and we are done. $\square$

PROPOSITION 3.7. *Let X, Y, Z, W be stochastic variables and Φ, Θ complementary decision variables. Then*

$$\textbf{P5}'':\quad \left.\begin{array}{c} X \perp\!\!\!\perp (Y,\Theta) \mid (Z,\Phi) \\ \textit{and} \\ W \perp\!\!\!\perp (Y,\Theta) \mid (X,Z,\Phi) \end{array}\right\} \Rightarrow (X,W) \perp\!\!\!\perp (Y,\Theta) \mid (Z,\Phi).$$

PROOF. Following the same approach as in the proof of Lemma 3.6, to prove that $(X,W) \perp\!\!\!\perp (Y,\Theta) \mid (Z,\Phi)$ it is enough to show that, for all $\phi \in \Phi(\mathcal{S})$ and all $A_{X,W} = A_X \cap A_W$ where $A_X \in \sigma(X)$ and $A_W \in \sigma(W)$, there exists $w_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_{X,W}} \mid Y, Z) = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{3.11}$$

Since $W \perp\!\!\!\perp (Y,\Theta) \mid (X,Z,\Phi)$, for all $\phi \in \Phi(\mathcal{S})$ and all $A_W \in \sigma(W)$ there exists $w_\phi^1(X,Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_W} \mid X, Y, Z) = w_\phi^1(X,Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{3.12}$$

Also by Lemma 3.6,

$$X \perp\!\!\!\perp (Y,\Theta) \mid (Z,\Phi) \Rightarrow (X,Z) \perp\!\!\!\perp (Y,\Theta) \mid (Z,\Phi).$$

Thus, for all $\phi \in \Phi(\mathcal{S})$ and all $h(X,Z)$, there exists $w_\phi^2(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma\{h(X,Z) \mid Y, Z\} = w_\phi^2(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{3.13}$$

Let $\phi \in \Phi(\mathcal{S})$ and $A_{X,W} = A_X \cap A_W$, where $A_X \in \sigma(X)$ and $A_W \in \sigma(W)$. Then

$$\begin{aligned} \mathbb{E}_\sigma(\mathbb{1}_{A_X\cap A_W} \mid Y, Z) &= \mathbb{E}_\sigma\{\mathbb{E}_\sigma(\mathbb{1}_{A_X\cap A_W} \mid X, Y, Z) \mid Y, Z\} \qquad \text{a.s. } [\mathbb{P}_\sigma] \\ &= \mathbb{E}_\sigma\{\mathbb{1}_{A_X}\mathbb{E}_\sigma(\mathbb{1}_{A_W} \mid X, Y, Z) \mid Y, Z\} \qquad \text{a.s. } [\mathbb{P}_\sigma] \\ &= \mathbb{E}_\sigma\{\mathbb{1}_{A_X} w_\phi^1(X,Z) \mid Y, Z\} \qquad \text{a.s. } [\mathbb{P}_\sigma] \text{ by (3.12)} \\ &= w_\phi^2(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma] \text{ by (3.13)}, \end{aligned}$$

which proves (3.11). □

What we have shown in this section (without making any specific assumptions about the nature of the stochastic variables or the regime space) is that axioms P2′–P5′ as well as the mirror axioms P3″ and P5″ hold in full generality. However, validity of P4″ in full generality remains an open problem. In the subsequent sections, we will study P4″ under certain additional conditions. In Section 4, we take a Bayesian approach and develop a set-up that allows us to prove P4″ under the assumption of a discrete regime space (Corollary 4.6), and in Section 5 we take a measure-theoretic approach, which allows us to prove P4″ under the assumption of discreteness of the stochastic variables (Proposition 5.1) or the existence of a dominating regime (Proposition 5.2).

4. A Bayesian approach. In the present section, we introduce a Bayesian construction in an attempt to justify the remaining separoid axioms. We extend the original space in order to construe both stochastic and nonstochastic variables as measurable functions on the new space and create an analogy between extended conditional independence and stochastic conditional independence. Similar ideas can be found in a variety of contexts in probability theory and statistics. Examples include Poisson random processes [40], pages 82–84, or Bayesian approaches to statistics [41]. We will see that, under the assumption of a discrete regime space, extended conditional independence and stochastic conditional independence are equivalent. Thus, we can continue to apply all the properties P1^s–P5^s of Theorem 2.4.

Consider a measurable space $(\Omega, \mathcal{A})$ and a regime space $\mathcal{S}$ and let $\mathcal{F}$ be a σ-algebra of subsets of $\mathcal{S}$. We can expand the original space $(\Omega, \mathcal{A})$ and consider the product space $\Omega \times \mathcal{S}$ with its corresponding σ-algebra $\mathcal{A} \otimes \mathcal{F}$, where $\mathcal{A} \otimes \mathcal{F} := \sigma(\mathcal{A} \times \mathcal{F}) := \sigma(\{A \times B : A \in \mathcal{A}, B \in \mathcal{F}\})$. Thus, we can regard all stochastic variables $X, Y, Z, \dots$ defined on $(\Omega, \mathcal{A})$ also as defined on $(\Omega \times \mathcal{S}, \mathcal{A} \otimes \mathcal{F})$ and all $\mathcal{F}$-measurable decision variables $\Theta, \Phi, \dots$ defined on $\mathcal{S}$ also as defined on $(\Omega \times \mathcal{S}, \mathcal{A} \otimes \mathcal{F})$. To see this, consider any stochastic variable $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}_X)$. For any such X, we define $X^* : (\Omega \times \mathcal{S}, \mathcal{A} \otimes \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_X)$ by $X^*(\omega, \sigma) = X(\omega)$. It is readily seen that X^* is $\mathcal{A} \otimes \mathcal{F}$-measurable. Similarly, for a decision variable $\Theta : \mathcal{S} \to \Theta(\mathcal{S})$ we will denote by $\Theta^* : (\Omega \times \mathcal{S}, \mathcal{A} \otimes \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_X)$ the function defined by $\Theta^*(\omega, \sigma) = \Theta(\sigma)$. We will use similar conventions for all the variables we consider.

REMARK 4.1. Note that for any stochastic variable X as above, $\sigma(X^*) = \sigma(X) \times \{\mathcal{S}\}$. Similarly, for any decision variable Θ^*, $\sigma(\Theta^*) = \{\Omega\} \times \sigma(\Theta)$. Thus,

$$\begin{aligned}
\sigma(X^*, \Theta^*) &= \sigma\big(\{A_{X^*} \cap A_{\Theta^*} : A_{X^*} \in \sigma(X^*), A_{\Theta^*} \in \sigma(\Theta^*)\}\big) \\
&\qquad \text{(see [46], page 73)} \\
&= \sigma\big(\{A_{X^*} \cap A_{\Theta^*} : A_{X^*} \in \sigma(X) \times \{\mathcal{S}\}, A_{\Theta^*} \in \{\Omega\} \times \sigma(\Theta)\}\big) \\
&= \sigma\big(\{(A_X \times \mathcal{S}) \cap (\Omega \times A_\Theta) : A_X \in \sigma(X), A_\Theta \in \sigma(\Theta)\}\big) \\
&= \sigma\big(\{A_X \times A_\Theta : A_X \in \sigma(X), A_\Theta \in \sigma(\Theta)\}\big) \\
&=: \sigma(X) \otimes \sigma(\Theta).
\end{aligned}$$

Thus, for any $\sigma \in \mathcal{S}$ and $A_{X^*} \in \sigma(X^*)$, the function $\mathbb{1}^\sigma_{A_{X^*}} : \Omega \to \{0, 1\}$ defined by $\mathbb{1}^\sigma_{A_{X^*}}(\omega) := \mathbb{1}_{A_{X^*}}(\omega, \sigma)$ does not depend on σ. It is equal to $\mathbb{1}_{A_X}$, for $A_X \in \sigma(X)$ such that $A_{X^*} = A_X \times \{\mathcal{S}\}$. Also for $A_{X^*,\Theta^*} \in \sigma(X^*, \Theta^*)$, the function $\mathbb{1}_{A_{X^*,\Theta^*}}$ is $(\sigma(X) \otimes \sigma(\Theta))$-measurable, and by Lemma 4.1, for $\sigma \in \mathcal{S}$, the function $\mathbb{1}^\sigma_{A_{X^*,\Theta^*}} : \Omega \to \{0, 1\}$ defined by $\mathbb{1}^\sigma_{A_{X^*,\Theta^*}}(\omega) := \mathbb{1}_{A_{X^*,\Theta^*}}(\omega, \sigma)$ is $\sigma(X)$-measurable. $\mathbb{1}^\sigma_{A_{X^*,\Theta^*}}(\omega)$ is equal to $\mathbb{1}_{A^\sigma_X}$ for $A^\sigma_X \in \sigma(X)$ such that A^σ_X is the section of A_{X^*,Θ^*} at σ.

Now in the initial space $(\Omega, \mathcal{A})$, we can talk about extended conditional independence and in the product space $(\Omega \times \mathcal{S}, \mathcal{A} \otimes \mathcal{F})$, after we equip it with a probability measure, we can talk about stochastic conditional independence. To rigorously justify the equivalence of extended conditional independence and stochastic conditional independence, we will need the following results.

LEMMA 4.1. *Let $f : \Omega \times \mathcal{S} \to \mathbb{R}$ be $\mathcal{A} \otimes \mathcal{F}$-measurable. Define for all $\sigma \in \mathcal{S}$, $f_\sigma : \Omega \to \mathbb{R}$ by $f_\sigma(\omega) := f(\omega, \sigma)$. Then f_σ is $\mathcal{A}$-measurable. If further f is bounded, define for all $\sigma \in \mathcal{S}$, $\mathbb{E}_\sigma(f_\sigma) : \mathcal{S} \to \mathbb{R}$ by $\mathbb{E}_\sigma(f_\sigma) := \int_\Omega f_\sigma(\omega)\mathbb{P}_\sigma(d\omega)$. Then the function $\sigma \mapsto \mathbb{E}_\sigma(f_\sigma)$ is bounded and $\mathcal{F}$-measurable.*

PROOF. See [4], page 231, Theorem 18.1, and page 234, Theorem 18.3. □

Now let π be a probability measure on $(F, \mathcal{F})$. For $A^* \in \mathcal{A} \otimes \mathcal{F}$, define

$$\mathbb{P}^*(A^*) := \int_{\mathcal{S}} \int_\Omega \mathbb{1}_{A^*}(\omega, \sigma)\mathbb{P}_\sigma(d\omega)\pi(d\sigma). \tag{4.1}$$

THEOREM 4.2. *$\mathbb{P}^*$ is the unique probability measure on $\mathcal{A} \otimes \mathcal{F}$ such that*

$$\mathbb{P}^*(A \times B) = \int_B \mathbb{P}_\sigma(A)\pi(d\sigma) \tag{4.2}$$

for all $A \in \mathcal{A}$ and $B \in \mathcal{F}$.

PROOF. See the online supplementary material [8]. □

THEOREM 4.3. *Let $f : \Omega \times \mathcal{S} \to \mathbb{R}$ be an $\mathcal{A} \otimes \mathcal{F}$-measurable integrable function. Then*

$$\mathbb{E}^*(f) = \int_{\mathcal{S}} \int_\Omega f(\omega, \sigma)\mathbb{P}_\sigma(d\omega)\pi(d\sigma). \tag{4.3}$$

PROOF. See the online supplementary material [8]. □

In the above theorems, we have constructed a new probability measure $\mathbb{P}^*$ on the measurable space $(\Omega \times \mathcal{S}, \mathcal{A} \otimes \mathcal{F})$ and also obtained an expression for the integral of a $\mathcal{A} \otimes \mathcal{F}$-measurable function under $\mathbb{P}^*$. We now use this expression to justify the analogy between extended conditional independence and stochastic conditional independence in the case of a discrete regime space.

4.1. *Discrete regime space.* We now suppose that $\mathcal{S}$ is discrete, and take $\mathcal{F}$ to comprise all subsets of $\mathcal{S}$. In particular, every decision variable is $\mathcal{F}$-measurable. Moreover, in this case we can, and shall, require $\pi(\{\sigma\}) > 0$ for all $\sigma \in \mathcal{S}$.

Now (4.3) becomes

$$\mathbb{E}^*(f) = \sum_{\sigma \in \mathcal{S}} \int_\Omega f(\omega, \sigma)\mathbb{P}_\sigma(d\omega)\pi(\sigma)$$
$$= \sum_{\sigma \in \mathcal{S}} \mathbb{E}_\sigma(f_\sigma)\pi(\sigma).$$

THEOREM 4.4. *Let X, Y, Z be $\mathcal{A}$-measurable functions on Ω, and let Φ, Θ be decision variables on $\mathcal{S}$, where $\mathcal{S}$ is discrete. Suppose that Θ is complementary to Φ. Then $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ if and only if $X^* \perp\!\!\!\perp_s (Y^*, \Theta^*) \mid (Z^*, \Phi^*)$.*

PROOF. $\Rightarrow$: Since $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, by Proposition 3.1, for all $\phi \in \Phi(\mathcal{S})$ and all $A_X \in \sigma(X)$, there exists a function $w_\phi(Z)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, Z) = w_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma], \tag{4.4}$$

that is,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X}\mathbb{1}_{A_{Y,Z}}) = \mathbb{E}_\sigma\{w_\phi(Z)\mathbb{1}_{A_{Y,Z}}\} \qquad \text{whenever } A_{Y,Z} \in \sigma(Y, Z). \tag{4.5}$$

To show that $X^* \perp\!\!\!\perp_s (Y^*, \Theta^*) \mid (Z^*, \Phi^*)$, by Proposition 2.3 we need to show that, for all $A_{X^*} \in \sigma(X^*)$, there exists a function $w(Z^*, \Phi^*)$ such that

$$\mathbb{E}^*(\mathbb{1}_{A_{X^*}} \mid Y^*, \Theta^*, Z^*, \Phi^*) = w(Z^*, \Phi^*) \qquad \text{a.s. } [\mathbb{P}^*],$$

that is,

$$\mathbb{E}^*(\mathbb{1}_{A_{X^*}}\mathbb{1}_{A_{Y^*,\Theta^*,Z^*,\Phi^*}}) = \mathbb{E}^*\{w(Z^*, \Phi^*)\mathbb{1}_{A_{Y^*,\Theta^*,Z^*,\Phi^*}}\} \tag{4.6}$$

whenever $A_{Y^*,\Theta^*,Z^*,\Phi^*} \in \sigma(Y^*, \Theta^*, Z^*, \Phi^*)$.

Let $A_{X^*} \in \sigma(X^*)$ and define $w(z^*, \phi^*) = w_{\phi^*}(z^*)$ as in (4.5). Then for all $A_{Y^*,\Theta^*,Z^*,\Phi^*} \in \sigma(Y^*, \Theta^*, Z^*, \Phi^*)$,

$$\begin{aligned}\mathbb{E}^*(\mathbb{1}_{A_{X^*}}\mathbb{1}_{A_{Y^*,\Theta^*,Z^*,\Phi^*}}) &= \sum_{\sigma \in \mathcal{S}} \mathbb{E}_\sigma(\mathbb{1}_{A_X}\mathbb{1}_{A^\sigma_{Y,Z}})\pi(\sigma)\\ &= \sum_{\sigma \in \mathcal{S}} \mathbb{E}_\sigma\{w_\phi(Z)\mathbb{1}_{A^\sigma_{Y,Z}}\}\pi(\sigma) \qquad \text{by (4.5)}\\ &= \mathbb{E}^*\{w(Z^*, \Phi^*)\mathbb{1}_{A_{Y^*,\Theta^*,Z^*,\Phi^*}}\}\end{aligned}$$

which proves (4.6).

$\Leftarrow$: To show that $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, let $\phi \in \Phi(\mathcal{S})$ and $A_X \in \sigma(X)$. Then, for any $\sigma_0 \in \Phi^{-1}(\phi)$,

$$\begin{aligned}\mathbb{E}_{\sigma_0}(\mathbb{1}_{A_X}\mathbb{1}_{A_{Y,Z}})\pi(\sigma_0) &= \sum_{\sigma \in \mathcal{S}} \mathbb{E}_\sigma\{\mathbb{1}_{A_X}\mathbb{1}_{A_{Y,Z}}\mathbb{1}_{\sigma_0}(\sigma)\}\pi(\sigma)\\ &= \mathbb{E}^*(\mathbb{1}_{A_{X^*}}\mathbb{1}_{A_{Y,Z}\times\{\sigma_0\}})\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}^*\{w(Z^*, \Phi^*)\mathbb{1}_{A_{Y,Z}\times\{\sigma_0\}}\} \qquad \text{by (4.6)} \\
&= \sum_{\sigma\in\mathcal{S}} \mathbb{E}_\sigma\{w(Z; \Phi(\sigma))\mathbb{1}_{A_{Y,Z}}\mathbb{1}_{\sigma_0}(\sigma)\}\pi(\sigma) \\
&= \mathbb{E}_{\sigma_0}\{w(Z; \Phi(\sigma_0))\mathbb{1}_{A_{Y,Z}}\}\pi(\sigma_0).
\end{aligned}
$$

Since $\pi(\sigma_0) > 0$, we have proved (4.5) with $w_\phi(z) = w(z, \phi)$. $\square$

COROLLARY 4.5. *Suppose we are given a collection of extended conditional independence properties as in the form of Definition 3.2. If the regime space $\mathcal{S}$ is discrete, any deduction made using the separoid axioms of stochastic conditional independence will be valid, so long as, in both premisses and conclusions, each conditional independence statement involves complementary decision variables, and no decision variable appears in the left-most term. We are however allowed to violate these conditions in intermediate steps of an argument.*

COROLLARY 4.6. *In the case of a discrete regime space, we have*:

P4″: $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ *and* $W \preceq X \Rightarrow X \perp\!\!\!\perp (Y, \Theta) \mid (Z, W, \Phi)$.

5. Other approaches. Inspecting the proof of Theorem 4.4, we see that the assumption of discreteness of the regime space $\mathcal{S}$ is crucial. If we have an uncountable regime space $\mathcal{S}$ and assign a distribution π over it, the arguments for the forward direction will still apply but the arguments for the reverse direction will not. Intuitively, this is because (4.1) holds almost everywhere but not necessarily everywhere. Thus, we cannot immediately extend it to hold for all $\sigma \in \mathcal{S}$ as in (4.4). However, using another, more direct, approach we can still deduce P4″ if we impose appropriate conditions. In particular, this will hold if the stochastic variables are discrete. Alternatively, we can use a domination condition on the set of regimes.

5.1. *Discrete variables.*

PROPOSITION 5.1. *Let X, Y, Z, W be discrete stochastic variables, and Φ, Θ complementary decision variables. Then*

P4″: $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ *and* $W \preceq X \Rightarrow X \perp\!\!\!\perp (Y, \Theta) \mid (Z, W, \Phi)$.

PROOF. To show that $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, W, \Phi)$ we need to show that, for all $\phi \in \Phi(\mathcal{S})$ and all $A_X \in \sigma(X)$, there exists $w_\phi(Z, W)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$
\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Y, Z, W) = w_\phi(Z, W) \qquad \text{a.s. } [\mathbb{P}_\sigma],
$$

that is,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X}\mathbb{1}_{A_{Y,Z,W}}) = \mathbb{E}_\sigma\{w_\phi(Z,W)\mathbb{1}_{A_{Y,Z,W}}\} \quad \text{whenever } A_{Y,Z,W} \in \sigma(Y,Z,W). \tag{5.1}$$

Observe that it is enough to show (5.1) for $A_{Y,Z,W} \in \sigma(Y,Z,W)$ such that $\mathbb{P}_\sigma(A_{Y,Z,W}) > 0$. Also since X, Y, Z and W are discrete, we need to show (5.1) only for sets of the form $\{X = x\}$ and $\{Y = y, Z = z, W = w\}$. Thus, it is enough to show that, for all $\phi \in \Phi(\mathcal{S})$ and all x, there exists $w_\phi(Z, W)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$, and all y, z, w such that $\mathbb{P}_\sigma(Y = y, Z = z, W = w) > 0$,

$$\mathbb{E}_\sigma(\mathbb{1}_{\{X=x\}}\mathbb{1}_{\{Y=y,Z=z,W=w\}}) = \mathbb{E}_\sigma\{w_\phi(Z,W)\mathbb{1}_{\{Y=y,Z=z,W=w\}}\}.$$

Let $\phi \in \Phi(\mathcal{S})$. For $\sigma \in \Phi^{-1}(\phi)$ and all x, y, z, w such that $\mathbb{P}_\sigma(Y = y, Z = z, W = w) > 0$,

$$\begin{aligned}
&\mathbb{E}_\sigma(\mathbb{1}_{\{X=x\}}\mathbb{1}_{\{Y=y,Z=z,W=w\}}) \\
&\quad = \mathbb{P}_\sigma(X = x, Y = y, Z = z, W = w) \\
&\quad = \mathbb{P}_\sigma(X = x \mid Y = y, Z = z, W = w)\mathbb{P}_\sigma(Y = y, Z = z, W = w) \\
&\quad = \frac{\mathbb{P}_\sigma(X = x, W = w \mid Y = y, Z = z)}{\mathbb{P}_\sigma(W = w \mid Y = y, Z = z)}\mathbb{P}_\sigma(Y = y, Z = z, W = w).
\end{aligned} \tag{5.2}$$

Since $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ and $W \preceq X$, there exist $w_\phi^1(Z)$ and $w_\phi^2(Z)$ such that

$$\mathbb{E}_\sigma(\mathbb{1}_{\{X=x,W=w\}} \mid Y, Z) = w_\phi^1(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]$$

and

$$\mathbb{E}_\sigma(\mathbb{1}_{\{W=w\}} \mid Y, Z) = w_\phi^2(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]$$

where $w_\phi^1(Z) = 0$ unless $w = W(x)$.

Define

$$w_\phi(z) = \begin{cases} \dfrac{w_\phi^1(z)}{w_\phi^2(z)} & \text{if } w_\phi^2(z) \neq 0, \\ 0 & \text{if } w_\phi^2(z) = 0 \end{cases}$$

and note that $w_\phi^2(z) \neq 0$ when $\mathbb{P}_\sigma(Y = y, Z = z, W = w) \neq 0$. Also note that since $w_\phi(Z)$ is $\sigma(Z)$-measurable it is also $\sigma(W, Z)$-measurable. Returning to (5.2), we get

$$\begin{aligned}
\mathbb{E}_\sigma(\mathbb{1}_{\{X=x\}}\mathbb{1}_{\{Y=y,Z=z,W=w\}}) &= w_\phi(z)\mathbb{P}_\sigma(Y = y, Z = z, W = w) \\
&= \mathbb{E}_\sigma\{w_\phi(Z)\mathbb{1}_{\{Y=y,Z=z,W=w\}}\}
\end{aligned}$$

which completes the proof. $\square$

5.2. *Dominating regime*. Another approach is based on the existence of a dominating regime in a set of distributions, that is, one such that every other distribution in the set is absolutely continuous with respect to it. This will automatically be the case if all the distributions under consideration have the same support. Such a domination condition is the foundation of the abstract version of the Fisher–Neyman factorisation criterion for sufficiency [39], using the Radon–Nikodym theorem. However, our analysis does not use that construction.

DEFINITION 5.1 (Dominating regime). Let $\mathcal{S}$ index a set of probability measures on $(\Omega, \mathcal{A})$. For $\mathcal{S}_0 \subseteq \mathcal{S}$, we say that $\sigma^* \in \mathcal{S}_0$ is a *dominating regime in* $\mathcal{S}_0$, if, for all $\sigma \in \mathcal{S}_0$, $\mathbb{P}_\sigma \ll \mathbb{P}_{\sigma^*}$; that is,

$$\mathbb{P}_{\sigma^*}(A) = 0 \quad \Rightarrow \quad \mathbb{P}_\sigma(A) = 0 \qquad \text{for all } A \in \mathcal{A} \text{ and all } \sigma \in \mathcal{S}_0.$$

PROPOSITION 5.2. *Let X, Y, Z, W be stochastic variables, and Φ, Θ complementary decision variables. Suppose that, for all $\phi \in \Phi(\mathcal{S})$, there exists a dominating regime $\sigma_\phi \in \Phi^{-1}(\phi)$. Then*

P4″: $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ *and* $W \preceq X \Rightarrow X \perp\!\!\!\perp (Y, \Theta) \mid (Z, W, \Phi)$.

PROOF. By Proposition 3.2, it suffices to prove the following two statements:

$$X \perp\!\!\!\perp Y \mid (Z, W, \Phi, \Theta) \tag{5.3}$$

and

$$X \perp\!\!\!\perp \Theta \mid (Z, W, \Phi). \tag{5.4}$$

To prove (5.3), we will use Proposition 3.3 and prove equivalently that $Y \perp\!\!\!\perp X \mid (Z, W, \Phi, \Theta)$. Note first that since $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, by Proposition 3.2 it follows that $X \perp\!\!\!\perp Y \mid (Z, \Phi, \Theta)$, and by Proposition 3.3 it follows that $Y \perp\!\!\!\perp X \mid (Z, \Phi, \Theta)$. Also, since $W \preceq X$, by Lemma 2.2 it follows that $\sigma(W) \subseteq \sigma(X)$. Let $(\phi, \theta) \in (\Phi, \Theta)(\mathcal{S})$, $\sigma = (\Phi, \Theta)^{-1}(\phi, \theta)$ and $A_Y \in \sigma(Y)$. Then

$$\begin{aligned} \mathbb{E}_\sigma(\mathbb{1}_{A_Y} \mid X, Z, W) &= \mathbb{E}_\sigma(\mathbb{1}_{A_Y} \mid X, Z) \qquad \text{a.s. } [\mathbb{P}_\sigma] \text{ since } \sigma(W) \subseteq \sigma(X) \\ &= \mathbb{E}_\sigma(\mathbb{1}_{A_Y} \mid Z) \qquad \text{a.s. } [\mathbb{P}_\sigma] \text{ since } Y \perp\!\!\!\perp X \mid (Z, \Phi, \Theta), \end{aligned}$$

which proves that $Y \perp\!\!\!\perp X \mid (Z, W, \Phi, \Theta)$.

To prove (5.4), let $\phi \in \Phi(\mathcal{S})$ and $A_X \in \sigma(X)$. We will show that there exists $w_\phi(Z, W)$ such that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mid Z, W) = w_\phi(Z, W) \qquad \text{a.s. } [\mathbb{P}_\sigma],$$

that is,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X} \mathbb{1}_{A_{Z,W}}) = \mathbb{E}_\sigma\{w_\phi(Z, W)\mathbb{1}_{A_{Z,W}}\} \qquad \text{whenever } A_{Z,W} \in \sigma(Z, W).$$

Let $A_{Z,W} \in \sigma(Z, W)$ and note that

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X}\mathbb{1}_{A_{Z,W}}) = \mathbb{E}_\sigma\{\mathbb{E}_\sigma(\mathbb{1}_{A_X}\mathbb{1}_{A_{Z,W}} \mid Z)\}. \tag{5.5}$$

Since $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, by Lemma 3.6 it follows that $(X, Z) \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, and by Proposition 3.2 that $(X, Z) \perp\!\!\!\perp \Theta \mid (Z, \Phi)$. Also, since $W \preceq X$ there exists $a_\phi(Z)$ such that

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X}\mathbb{1}_{A_{Z,W}} \mid Z) = a_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{5.6}$$

In particular, for the dominating regime $\sigma_\phi \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_{\sigma_\phi}(\mathbb{1}_{A_X}\mathbb{1}_{A_{Z,W}} \mid Z) = a_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_{\sigma_\phi}]$$

and thus

$$\mathbb{E}_{\sigma_\phi}\{\mathbb{E}_{\sigma_\phi}(\mathbb{1}_{A_X}\mathbb{1}_{A_{Z,W}} \mid Z, W) \mid Z\} = a_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_{\sigma_\phi}].$$

Since $\mathbb{P}_\sigma \ll \mathbb{P}_{\sigma_\phi}$ for all $\sigma \in \Phi^{-1}(\phi)$, it follows that, for all $\sigma \in \Phi^{-1}(\phi)$,

$$\mathbb{E}_{\sigma_\phi}\{\mathbb{E}_{\sigma_\phi}(\mathbb{1}_{A_X}\mathbb{1}_{A_{Z,W}} \mid Z, W) \mid Z\} = a_\phi(Z) \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{5.7}$$

Thus, by (5.6) and (5.7), we get that

$$\mathbb{E}_\sigma(\mathbb{1}_{A_X}\mathbb{1}_{A_{Z,W}} \mid Z) = \mathbb{E}_{\sigma_\phi}\{\mathbb{E}_{\sigma_\phi}(\mathbb{1}_{A_X}\mathbb{1}_{A_{Z,W}} \mid Z, W) \mid Z\} \qquad \text{a.s. } [\mathbb{P}_\sigma]. \tag{5.8}$$

Similarly,

$$\begin{aligned} &\mathbb{E}_{\sigma_\phi}\{\mathbb{E}_{\sigma_\phi}(\mathbb{1}_{A_X}\mathbb{1}_{A_{Z,W}} \mid Z, W) \mid Z\} \\ &\qquad = \mathbb{E}_\sigma\{\mathbb{E}_{\sigma_\phi}(\mathbb{1}_{A_X}\mathbb{1}_{A_{Z,W}} \mid Z, W) \mid Z\} \qquad \text{a.s. } [\mathbb{P}_\sigma]. \end{aligned} \tag{5.9}$$

Returning to (5.5), it follows that

$$\begin{aligned} \mathbb{E}_\sigma(\mathbb{1}_{A_X}\mathbb{1}_{A_{Z,W}}) &= \mathbb{E}_\sigma[\mathbb{E}_{\sigma_\phi}\{\mathbb{1}_{A_{Z,W}}\mathbb{E}_{\sigma_\phi}(\mathbb{1}_{A_X} \mid Z, W) \mid Z\}] \qquad \text{by (5.8)} \\ &= \mathbb{E}_\sigma[\mathbb{E}_\sigma\{\mathbb{1}_{A_{Z,W}}\mathbb{E}_{\sigma_\phi}(\mathbb{1}_{A_X} \mid Z, W) \mid Z\}] \qquad \text{by (5.9)} \\ &= \mathbb{E}_\sigma\{\mathbb{1}_{A_{Z,W}}\mathbb{E}_{\sigma_\phi}(\mathbb{1}_{A_X} \mid Z, W)\}. \end{aligned}$$

□

6. Pairwise conditional independence. Yet another path is to relax the notion of extended conditional independence. Here, we introduce a weaker version that we term *pairwise extended conditional independence*.

DEFINITION 6.1. Let X, Y and Z be stochastic variables and let Θ and Φ be complementary decision variables. We say that X *is pairwise (conditionally) independent of* (Y, Θ) *given* (Z, Φ), and write $X \perp\!\!\!\perp_p (Y, \Theta) \mid (Z, \Phi)$, if for all $\phi \in \Phi(\mathcal{S})$, all real, bounded and measurable functions h, and all pairs $\{\sigma_1, \sigma_2\} \in \Phi^{-1}(\phi)$, there exists a function $w_\phi^{\sigma_1,\sigma_2}(Z)$ such that

$$\mathbb{E}_{\sigma_1}\{h(X) \mid Y, Z\} = w_\phi^{\sigma_1,\sigma_2}(Z) \qquad \text{a.s. } [\mathbb{P}_{\sigma_1}]$$

and

$$\mathbb{E}_{\sigma_2}\{h(X) \mid Y, Z\} = w_\phi^{\sigma_1,\sigma_2}(Z) \qquad \text{a.s. } [\mathbb{P}_{\sigma_2}].$$

It is readily seen that extended conditional independence implies pairwise extended conditional independence, but the converse is false. In Definition 6.1, for all $\phi \in \Phi(\mathcal{S})$, we only require a common version for the corresponding conditional expectation for every pair of regimes $\{\sigma_1, \sigma_2\} \in \Phi^{-1}(\phi)$, but we do not require that these versions agree on one function that can serve as a version for the corresponding conditional expectation simultaneously in all regimes $\sigma \in \Phi^{-1}(\phi)$.

Under this weaker definition, the analogues of P1$'$ to P5$'$, and of P3$''$ and P5$''$, can be seen to hold just as in Section 3.1. Also, by confining attention to two regimes at a time and applying Corollary 4.6, the analogue of P4$''$ will hold without further conditions.

It can be shown that, when there exists a dominating regime, pairwise extended conditional independence is equivalent to extended conditional independence. The argument parallels that of [39], who show that, under domination, pairwise sufficiency implies sufficiency. This property can be used to supply an alternative proof of Proposition 5.2.

7. Further extensions. So far we have studied extended conditional independence relations of the form $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, where the left-most term is fully stochastic. We now wish to extend this to the most general expression, of the form $(X, K) \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, where X, Y, Z are stochastic variables and K, Θ, Φ are complementary decision variables, and investigate the validity of the separoid axioms. An example of such an extended language appears in Example 3.1, where, by formal application of the separoid properties to (i), (ii) and (iii), we can derive $(T, \Theta) \perp\!\!\!\perp \Phi$ and $(X, \Phi) \perp\!\!\!\perp \Theta \mid T$.

Consider first the expression $K \perp\!\!\!\perp \Theta \mid Z$. Our desired intuitive interpretation of this is that conditioning on the stochastic variable Z renders the decision variables K and Θ variation independent. We need to turn this intuition into a rigorous definition, taking account of the fact that Z may have different distributions in the different regimes $\sigma \in \mathcal{S}$, whereas K and Θ are functions defined on $\mathcal{S}$.

One way to interpret this intuition is to consider, for each value z of Z, the set $\mathcal{S}_z$ of regimes that for which z is a “possible outcome”, and ask that the decision variables be variation independent on this restricted set. In order to make this rigorous, we shall require that $Z : (\Omega, \mathcal{A}) \to (F_Z, \mathcal{F}_Z)$ where $(F_Z, \mathcal{F}_Z)$ is a topological space with its Borel σ-algebra, and introduce

$$\mathcal{S}_z := \{\sigma \in \mathcal{S} : \mathbb{P}_\sigma(Z \in U) > 0 \text{ for every open set } U \subseteq \mathcal{F}_Z \text{ containing } z\}.$$

In particular, when Z is discrete, with the discrete topology,

$$\mathcal{S}_z := \{\sigma \in \mathcal{S} : \mathbb{P}_\sigma(Z = z) > 0\}.$$

We now formalise the slightly more general expression $K \perp\!\!\!\perp \Theta \mid (Z, \Phi)$ in the following definition.

DEFINITION 7.1. Let Z be a stochastic variable and K, Θ, Φ complementary decision variables. We say that Θ *is (conditionally) independent of K given* (Z, Φ), and write $\Theta \perp\!\!\!\perp K \mid (Z, \Phi)$ if, for all $z \in Z(\Omega)$, $\Theta \perp\!\!\!\perp_v K \mid \Phi\ [\mathcal{S}_z]$.

We now wish to introduce further definitions, to allow stochastic and decision variables to appear together in the left-most term of a conditional independence statement. Recall that in Definition 3.2, we defined $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ only when Θ and Φ are complementary on $\mathcal{S}$. Similarly, we will define $(X, K) \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ only when K, Θ and Φ are complementary. Our interpretation of $(X, K) \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ will now be a combination of Definitions 3.2 and 7.1. We start with a special case.

DEFINITION 7.2. Let Y, Z be stochastic variables, and K, Θ, Φ complementary decision variables. We say that (Y, Θ) *is (conditionally) independent of K given* (Z, Φ), and write $(Y, \Theta) \perp\!\!\!\perp K \mid (Z, \Phi)$, if $Y \perp\!\!\!\perp K \mid (Z, \Phi, \Theta)$ and $\Theta \perp\!\!\!\perp K \mid (Z, \Phi)$. In this case, we may also say that K *is (conditionally) independent of* (Y, Θ) *given* (Z, Φ), and write $K \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$.

Finally, we have the general definition.

DEFINITION 7.3. Let X, Y, Z be stochastic variables and K, Θ, Φ complementary decision variables. We say that (X, K) *is (conditionally) independent of* (Y, Θ) *given* (Z, Φ), and write $(X, K) \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, if

$$X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi, K) \tag{7.1}$$

and

$$K \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi). \tag{7.2}$$

REMARK 7.1. From Definition 7.3 and Proposition 3.2, $(X, K) \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ is equivalent to:

$$X \perp\!\!\!\perp Y \mid (Z, \Phi, K, \Theta), \tag{7.3}$$

$$X \perp\!\!\!\perp \Theta \mid (Z, \Phi, K), \tag{7.4}$$

$$Y \perp\!\!\!\perp K \mid (Z, \Phi, \Theta), \tag{7.5}$$

$$\Theta \perp\!\!\!\perp K \mid (Z, \Phi). \tag{7.6}$$

7.1. *Separoid properties*. We now wish to investigate the extent to which versions of the separoid axioms apply to the above general definition. In this context, the relevant set V is the set of pairs of the form (Y, Θ), where Y is a stochastic variable defined on Ω and Θ is a decision variable defined on $\mathcal{S}$.

For a full separoid treatment, we also need to introduce a quasiorder $\preceq$ on V. A natural definition would be: $(W,\Lambda)\preceq(Y,\Theta)$ if $W=f(Y)$ for some measurable function f (also denoted by $W\preceq Y$) and $\Lambda=h(\Theta)$ for some function h. Then $(V,\preceq)$ becomes a join semilattice, with join $(Y,\Theta)\vee(W,\Lambda)\approx((Y,W),(\Theta,\Lambda))$. Again, whenever we consider a relation $(X,K)\perp\!\!\!\perp(Y,\Theta)\mid(Z,\Phi)$ we require that K,Θ,Φ be complementary.

THEOREM 7.1 (Separoid-type properties). *Let X,Y,Z,W be stochastic variables and K,Θ,Φ complementary decision variables. Then the following properties hold*:

P1^g: $(X,K)\perp\!\!\!\perp(Y,\Theta)\mid(Z,\Phi)\Rightarrow(Y,\Theta)\perp\!\!\!\perp(X,K)\mid(Z,\Phi)$.
P2^g: *If Θ and K are complementary*, $(X,K)\perp\!\!\!\perp(Y,\Theta)\mid(Y,\Theta)$.
P3^g: $(X,K)\perp\!\!\!\perp(Y,\Theta)\mid(Z,\Phi),\ W\preceq Y\Rightarrow(X,K)\perp\!\!\!\perp(W,\Theta)\mid(Z,\Phi)$.
P4^g: *Under the conditions of Corollary* 4.6, *Proposition* 5.1 *or Proposition* 5.2, $(X,K)\perp\!\!\!\perp(Y,\Theta)\mid(Z,\Phi),\ W\preceq Y\Rightarrow(X,K)\perp\!\!\!\perp(Y,\Theta)\mid(Z,W,\Phi)$.
P4a^g: $(X,K)\perp\!\!\!\perp(Y,\Theta)\mid(Z,\Phi),\ \Lambda\preceq\Theta\Rightarrow(X,K)\perp\!\!\!\perp(Y,\Theta)\mid(Z,\Phi,\Lambda)$.
P5^g: $\left.\begin{array}{c}(X,K)\perp\!\!\!\perp(Y,\Theta)\mid(Z,\Phi)\\ \textit{and}\\ (X,K)\perp\!\!\!\perp W\mid(Y,Z,\Theta,\Phi)\end{array}\right\}\Rightarrow(X,K)\perp\!\!\!\perp(Y,W,\Theta)\mid(Z,\Phi)$.

PROOF.

P1^g. We need to show:

$$Y\perp\!\!\!\perp X\mid(Z,\Phi,\Theta,K),\tag{7.7}$$

$$Y\perp\!\!\!\perp K\mid(Z,\Phi,\Theta),\tag{7.8}$$

$$X\perp\!\!\!\perp\Theta\mid(Z,\Phi,K),\tag{7.9}$$

$$K\perp\!\!\!\perp\Theta\mid(Z,\Phi).\tag{7.10}$$

(7.8) and (7.9) hold automatically by (7.5) and (7.4). Also applying P1$'$ to (7.3) we deduce (7.7). Rephrasing (7.6) in terms of variation independence, we have that, for all $z\in Z(\Omega)$, $\Theta\perp\!\!\!\perp_v K\mid\Phi\ [\mathcal{S}_z]$. Thus, applying P1^v to (7.6) we deduce that, for all $z\in Z(\Omega)$, $K\perp\!\!\!\perp_v\Theta\mid\Phi\ [\mathcal{S}_z]$, that is, (7.10).

P2^g. We need to show:

$$X\perp\!\!\!\perp(Y,\Theta)\mid(Y,\Theta,K),\tag{7.11}$$

$$Y\perp\!\!\!\perp K\mid(Y,\Theta),\tag{7.12}$$

$$\Theta\perp\!\!\!\perp K\mid(Y,\Theta).\tag{7.13}$$

By P2$'$, we have that $X\perp\!\!\!\perp(Y,\Theta,K)\mid(Y,\Theta,K)$ which is identical to (7.11). To show (7.12), let $\theta\in\Theta(\mathcal{S})$ and $A_Y\in\sigma(Y)$. We seek $w_\theta(Y)$ such that, for all $\sigma\in\Theta^{-1}(\theta)$,

$$\mathbb{E}_\sigma(\mathbb{1}_{A_Y}\mid Y)=w_\theta(Y)\qquad\text{a.s. }[\mathbb{P}_\sigma].$$

But note that

$$\mathbb{E}_\sigma(\mathbb{1}_{A_Y} \mid Y) = \mathbb{1}_{A_Y} \qquad \text{a.s. } [\mathbb{P}_\sigma].$$

To show (7.13), let $y \in Y(\Omega)$. By P2^v, we have that

$$K \perp\!\!\!\perp_v \Theta \mid \Theta \, [\mathcal{S}_y]. \tag{7.14}$$

Applying P1^v to (7.12), we deduce that $\Theta \perp\!\!\!\perp_v K \mid \Theta \, [\mathcal{S}_y]$, that is, (7.13).

P3^g. We need to show:

$$X \perp\!\!\!\perp (W, \Theta) \mid (Z, \Phi, K), \tag{7.15}$$

$$W \perp\!\!\!\perp K \mid (Z, \Phi, \Theta), \tag{7.16}$$

$$\Theta \perp\!\!\!\perp K \mid (Z, \Phi). \tag{7.17}$$

Since $W \preceq Y$, applying P3$'$ to (7.1) we deduce (7.15), and applying P3$''$ to (7.4) we deduce (7.16); while (7.17) is identical to (7.6).

P4^g. We need to show:

$$X \perp\!\!\!\perp (Y, \Theta) \mid (W, Z, \Phi, K), \tag{7.18}$$

$$Y \perp\!\!\!\perp K \mid (W, Z, \Phi, \Theta), \tag{7.19}$$

$$\Theta \perp\!\!\!\perp K \mid (W, Z, \Phi). \tag{7.20}$$

(7.18) follows from (7.1) by P4$'$. Also (under the given conditions), (7.19) follows from (7.5) by P4$''$.

For simplicity, we restrict the proof of (7.20) to the case that all stochastic variables are discrete.

We first note that (7.6) is equivalent to: for all z,

$$\Theta \perp\!\!\!\perp_v K \mid \Phi \, [S_z], \tag{7.21}$$

where $S_z = \{(\theta, \phi, k) : \mathbb{P}_{\theta,\phi,k}(Z = z) > 0\}$; while (7.20) is equivalent to: for all (w, z),

$$\Theta \perp\!\!\!\perp_v K \mid \Phi \, [S_{w,z}], \tag{7.22}$$

where $S_{w,z} = \{(\theta, \phi, k : \mathbb{P}_{\theta,\phi,k}(W = w, Z = z) > 0\}$. Now $\mathbb{P}_{\theta,\phi,k}(W = w, Z = z) = \mathbb{P}_{\theta,\phi,k}(Z = z) \times \mathbb{P}_{\theta,\phi,k}(W = w \mid Z = z)$. Hence, the additional constraint, over and above that of S_z, is $\mathbb{P}_{\theta,\phi,k}(W = w \mid Z = z) > 0$. But by (7.5) and P3$''$, $W \perp\!\!\!\perp K \mid (Z, \Phi, \Theta)$, so that $\mathbb{P}_{\theta,\phi,k}(W = w \mid Z = z)$ is a function only of (θ, ϕ). Thus, (7.22) imposes, on (7.21), the condition $I(\Theta, \Phi) = 1$, where $I(\theta, \phi)$ is the indicator function of the property $\mathbb{P}_{\theta,\phi}(W = w \mid Z = z) > 0$. That is to say, (7.22) would follow from

$$\Theta \perp\!\!\!\perp_v K \mid \big(\Phi, I(\Theta, \Phi)\big)[S_z].$$

But this follows from (7.21) and the separoid properties of $\perp\!\!\!\perp_v$.

P4a^g. We need to show:

$$X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi, K, \Lambda), \tag{7.23}$$

$$Y \perp\!\!\!\perp K \mid (Z, \Phi, \Theta, \Lambda), \tag{7.24}$$

$$\Theta \perp\!\!\!\perp K \mid (Z, \Phi, \Lambda). \tag{7.25}$$

Now (7.23) follows from (7.1) and P4a'; while (7.24) is a trivial consequence of (7.5). As for (7.25), we note that (7.6) can be expressed as: for all z, $\Theta \perp\!\!\!\perp_v K \mid \Phi\ [S_z]$. By the separoid properties of $\perp\!\!\!\perp_v$, this implies $\Theta \perp\!\!\!\perp_v K \mid (\Phi, \Lambda)\ [S_z]$, which is (7.25).

P5^g. We need to show:

$$X \perp\!\!\!\perp (Y, W, \Theta) \mid (Z, \Phi, K), \tag{7.26}$$

$$(Y, W) \perp\!\!\!\perp K \mid (Z, \Phi, \Theta), \tag{7.27}$$

$$\Theta \perp\!\!\!\perp K \mid (Z, \Phi). \tag{7.28}$$

Since $(X, K) \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$, (7.1), (7.5) and (7.6) hold; while since $(X, K) \perp\!\!\!\perp W \mid (Y, Z, \Theta, \Phi)$, we have

$$X \perp\!\!\!\perp W \mid (Y, Z, \Theta, \Phi, K), \tag{7.29}$$

$$W \perp\!\!\!\perp K \mid (Y, Z, \Theta, \Phi). \tag{7.30}$$

Then (7.28) holds by (7.6). Also applying P5$'$ to (7.1) and (7.29), we deduce (7.26) and applying P5$''$ to (7.5) and (7.30) we deduce (7.27). □

We illustrate the above theory by applying it to Example 3.1.

THEOREM 7.2. *Suppose*

(i) $T \perp\!\!\!\perp \Phi \mid \Theta$
(ii) $X \perp\!\!\!\perp \Theta \mid (T, \Phi)$
(iii) $\Theta \perp\!\!\!\perp \Phi$.

Then $\Theta \perp\!\!\!\perp (X, \Phi) \mid T$.

PROOF. Apply P5^g to (i) and (iii), taking X, Y, Z and Φ trivial, and replacing K by Φ and W by T. We deduce $\Phi \perp\!\!\!\perp (T, \Theta)$. Then by P4^g (which in this special case can be shown to apply without further conditions), we derive $\Phi \perp\!\!\!\perp \Theta \mid T$. Combining this with (ii) yields $\Theta \perp\!\!\!\perp (X, \Phi) \mid T$. □

COROLLARY 7.3. *Under conditions supporting P4^g, $\Theta \perp\!\!\!\perp \Phi \mid (X, T)$. In particular, if $T \preceq X$ then $\Theta \perp\!\!\!\perp \Phi \mid X$.*

PROOF. From P4^g and P3^g. □

8. Applications of extended conditional independence to causality. The driving force behind this work was the need to establish a rigorous basis for a wide range of statistical concepts—in particular the Decision Theoretic framework for statistical causality. In the Decision Theoretic framework, we have stochastic variables whose outcomes are determined by nature, and decision variables that are functions of a regime indicator Σ that governs the probabilistic regime generating the stochastic variables. Using the language of extended conditional independence, we are able to express and manipulate conditions that allow us to transfer probabilistic information between regimes, and thus use information gleaned from one regime to understand a different, unobserved regime of interest.

Here, we illustrate, with two examples, how the language and the calculus of extended conditional independence can be applied to identify causal quantities. For numerous further applications and illustrations, see [3, 20–24, 26, 31, 35, 37, 38].

EXAMPLE 8.1 (Average causal effect). Suppose we are concerned with the effect of a binary treatment T (with value 1 denoting active treatment, and 0 denoting placebo) on a disease variable Y. There are 3 regimes of interest, indicated by a regime indicator Σ: $\Sigma = 1$ (resp., $\Sigma = 0$) denotes the situation where the patient is assigned treatment $T = 1$ (resp., $T = 0$) by external intervention; whereas $\Sigma = \varnothing$ indicates an observational regime, in which T is chosen, in some random way beyond the analyst's control,"by nature". For example, the data may have been gathered by doctors or in hospitals, and the criteria on which the treatment decisions were based not recorded.

A typical focus of interest is the *Average Causal Effect* (ACE) [35, 37],

$$\mathrm{ACE} := \mathbb{E}_1(Y) - \mathbb{E}_0(Y),$$

where $\mathbb{E}_\sigma(\cdot) = \mathbb{E}(\cdot \mid \sigma)$ denotes expectation under regime $\Sigma = \sigma$. This is a direct comparison of the average effects of giving treatment versus placebo for a given patient. However, in practice, for various reasons (ethical, financial, pragmatic, *etc.*), we may not be able to observe Y under these interventional regimes, and then cannot compare them directly. Instead, we might have access to data generated under the observational regime, where other variables might affect both the treatment choice and the variable of interest. In such a case, the distribution of the outcome of interest, for a patient receiving treatment $T = t$, cannot necessarily be assumed to be the same as in the corresponding interventional regime $\Sigma = t$.

However, if the observational data have been generated and collected from a randomised control trial (*i.e.*, the sample is randomly chosen and the treatment is randomly allocated), we could reasonably impose the following extended conditional independence condition:

$$Y \perp\!\!\!\perp \Sigma \mid T. \tag{8.1}$$

This condition expresses the property that, given information on the treatment T, the distribution of Y is independent of the regime—in particular, the same under interventional and observational conditions. When it holds, we are, intuitively, justified in identifying $\mathbb{E}_t(Y)$ with $\mathbb{E}_\varnothing(Y \mid T = t)$ $(t = 0, 1)$, so allowing estimation of ACE from the available data.

To make this intuition precise, note that, according to Definition 3.2, property (8.1) implies that there exists $w(T)$ such that, for all $\sigma \in \{\varnothing, 0, 1\}$,

$$\mathbb{E}_\sigma(Y \mid T) = w(T) \qquad \text{a.s. } [\mathbb{P}_\sigma].$$

Now in the interventional regimes, for $t = 0, 1$, $\mathbb{P}_t(T = t) = 1$. Thus, for $t = 0, 1$,

$$w(t) = \mathbb{E}_t(Y \mid T = t) = \mathbb{E}_t(Y) \qquad \text{a.s. } [\mathbb{P}_t].$$

Since both $w(t)$ and $\mathbb{E}_t(Y)$ are nonrandom real numbers, we thus must have

$$w(t) = \mathbb{E}_t(Y). \tag{8.2}$$

Also, in the observational regime,

$$\mathbb{E}_\varnothing(Y \mid T) = w(T) \qquad \text{a.s. } [\mathbb{P}_\varnothing].$$

Thus (so long as in the observational regime both treatments are allocated with positive probability), we obtain, for $t = 0, 1$,

$$\begin{aligned} \mathbb{E}_\varnothing(Y \mid T = t) &= w(t) \\ &= \mathbb{E}_t(Y) \qquad \text{by (8.2)}. \end{aligned}$$

Then

$$\begin{aligned} \text{ACE} &= \mathbb{E}_1(Y) - \mathbb{E}_0(Y) \\ &= \mathbb{E}_\varnothing(Y \mid T = 1) - \mathbb{E}_\varnothing(Y \mid T = 0) \end{aligned}$$

and so ACE can be estimated from the observational data.

EXAMPLE 8.2 (Dynamic treatment strategies). Suppose we wish to control some variable of interest through a sequence of consecutive actions [47–49]. An example in a medical context is maintaining a critical variable, such as blood pressure, within an appropriate risk-free range. To achieve such control, the doctor will administer treatments over a number of stages, taking into account, at each stage, a record of the patient's history, which provides information on the level of the critical variable, and possibly other related measurements.

We consider two sets of stochastic variables: $\mathcal{L}$, a set of *observable* variables, and $\mathcal{A}$, a set of *action* variables. The variables in $\mathcal{L}$ represent initial or intermediate symptoms, reactions, personal information, *etc.*, observable between consecutive treatments, and over which we have no direct control; they are perceived as generated and revealed by nature. The action variables $\mathcal{A}$ represent the treatments, which we could either control by external intervention, or else leave to nature (or the doctor) to determine.

An alternating ordered sequence $\mathcal{I} := (L_1, A_1, \ldots, L_n, A_n, L_{n+1} \equiv Y)$ with $L_i \subseteq \mathcal{L}$ and $A_i \in \mathcal{A}$ defines an *information base*, the interpretation being that the specified variables are observed in this time order. Thus, at each stage $i = 1, \ldots, n$ we will have a realisation of the random variable (or set of random variables) $L_i \subseteq \mathcal{L}$, followed by a value for the variable $A_i \in \mathcal{A}$. After the realisation of the final $A_n \in \mathcal{A}$, we will observe the outcome variable $L_{n+1} \in \mathcal{L}$, which we also denote by Y.

In such problems, we might be interested to evaluate and compare different strategies, that is, well-specified algorithms that take as input the recorded history of a patient at each stage and give as output the choice (possibly randomised) of the next treatment to be allocated. These strategies constitute interventional regimes, for which we would like to make inference. However, it may not be possible to implement all (or any) of these strategies to gather data, so we may need to rely on observational data and hope that it will be possible to use these data to estimate the interventional effects of interest.

We thus take the regime space to be $\mathcal{S} = \{\varnothing\} \cup \mathcal{S}^*$, where $\varnothing$ labels the observational regime under which data have been gathered, and $\mathcal{S}^*$ is the collection of contemplated interventional strategies. We denote the regime indicator, taking values in $\mathcal{S}$, by Σ. In order to identify the effect of some strategy $s \in \mathcal{S}^*$ on the outcome variable Y, we aim to estimate, from the observational data gathered under regime $\Sigma = \varnothing$, the expectation $\mathbb{E}_s\{k(Y)\}$, for some appropriate function $k(\cdot)$ of Y, that would result from application of strategy s.

One way to compute $\mathbb{E}_s\{k(Y)\}$ is by identifying the overall joint density of $(L_1, A_1, \ldots, L_n, A_n, Y)$ in the interventional regime of interest s. Factorising this joint density, we have

$$p_s(y, \bar{l}, \bar{a}) = \left\{\prod_{i=1}^{n+1} p_s(l_i \mid \bar{l}_{i-1}, \bar{a}_{i-1})\right\} \times \left\{\prod_{i=1}^{n} p_s(a_i \mid \bar{l}_i, \bar{a}_{i-1})\right\}$$

with $l_{n+1} \equiv y$. Here, $\bar{l}_i$ denotes $(l_1, \ldots, l_i)$, *etc.*

In order to compute $\mathbb{E}_s\{k(Y)\}$, we thus need the following terms:

(i) $p_s(a_i \mid \bar{l}_i, \bar{a}_{i-1})$ for $i = 1, \ldots, n$.
(ii) $p_s(l_i \mid \bar{l}_{i-1}, \bar{a}_{i-1})$ for $i = 1, \ldots, n+1$.

Since s is an interventional regime, corresponding to a well-defined treatment strategy, the terms in (i) are fully specified by the treatment protocol. So we only need to get a handle on the terms in (ii).

One assumption that would allow this is *simple stability*, expressed as

$$L_i \perp\!\!\!\perp \Sigma \mid (\overline{L}_{i-1}, \overline{A}_{i-1}) \qquad (i = 1, \ldots, n+1).$$

This says, intuitively, that the distribution of L_i, given all past observations, is the same in both the interventional and the observational regimes. When it holds (and assuming that the conditioning event occurs with positive probability in the observational regime) we can replace $p_s(l_i \mid \bar{l}_{i-1}, \overline{a}_{i-1})$ in (ii) with its observationally estimable counterpart, $p_\varnothing(l_i \mid \bar{l}_{i-1}, \overline{a}_{i-1})$. We then have all the ingredients needed to estimate the interventional effect $\mathbb{E}_s\{k(Y)\}$.[3]

However, in many cases the presence of unmeasured variables, both influencing the actions taken under the observational regime and affecting their outcomes, would not support a direct assumption of simple stability. Denote these additional variables by U_i $(i = 1, \dots, n)$. A condition that might be more justifiable in this context is *extended stability*, expressed as

$$(L_i, U_i) \perp\!\!\!\perp \Sigma \mid (\overline{L}_{i-1}, \overline{U}_{i-1}\overline{A}_{i-1}) \qquad (i = 1, \dots, n+1).$$

This is like simple stability, but taking the unmeasured variables also into account.

Now extended stability does not, in general, imply simple stability. But using the machinery of extended conditional independence, we can explore when, in combination with further conditions that might also be justifiable—for example, sequential randomisation or sequential irrelevance [24, 26]—simple stability can still be deduced, and hence $\mathbb{E}_s\{k(Y)\}$ estimated.

9. Discussion. We have presented a rigorous account of the hitherto informal concept of extended conditional independence, and indicated its fruitfulness in numerous statistical contexts, such as ancillarity, sufficiency, causal inference, *etc.*

Graphical models, in the form of Directed Acyclic Graphs (DAGs), are often used to represent collections of conditional independence properties amongst stochastic variables [9, 42], and we can then use graphical techniques (in particular, the *d-separation*, or the equivalent *moralisation*, criterion) to derive, in a visual and transparent way, implied conditional independence properties that follow from the assumptions and the separoid axioms. When such graphical models are extended to *Influence Diagrams*, incorporating both stochastic and nonstochastic variables, the identical methods support causal inference [20]. Numerous applications may be found in [23]. The theory developed in this paper formally justifies this extended methodology.

SUPPLEMENTARY MATERIAL

Some Proofs (DOI: 10.1214/16-AOS1537SUPP; .pdf). Supplementary material, comprising proofs of Lemma 2.2, Theorem 2.4, Proposition 2.5, Proposition 2.6, Theorem 2.7, Proposition 3.1, Theorem 4.2 and Theorem 4.3, is available online.

[3]The actual computation can be streamlined using *G-recursion* [26].

REFERENCES

[1] BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. MR0489333

[2] BASU, D. (1964). Recovery of ancillary information. *Sankhya A* **26** 3–16. MR0172366

[3] BERZUINI, C. R., DAWID, A. P. and DIDELEZ, V. (2012). Assessing dynamic treatment strategies. In *Causality*: *Statistical Perspectives and Applications* (C. R. Berzuini, A. P. Dawid and L. Bernardinelli, eds.) Chapter 8 85–100. Wiley, Chichester.

[4] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York. MR1324786

[5] BORKAR, V. S. (1995). *Probability Theory*: *An Advanced Course*, 1st ed. Springer, New York. MR1367959

[6] CASELLA, G. and BERGER, R. L. (2001). *Statistical Inference*, 2nd ed. Duxbury Press. MR1051420

[7] CONSTANTINOU, P. (2013). Conditional independence and applications in statistical causality. PhD Dissertation, Univ. Cambridge.

[8] CONSTANTINOU, P. and DAWID, A. P. (2017). Supplement to "Extended conditional independence and applications in causal inference." DOI:10.1214/16-AOS1537SUPP.

[9] COWELL, R. G., DAWID, A. P., LAURITZEN, S. L. and SPIEGELHALTER, D. J. (2007). *Probabilistic Networks and Expert Systems*: *Exact Computational Methods for Bayesian Networks*, 2nd ed. Springer, Berlin. MR1697175

[10] DAWID, A. P. (1975). On the concepts of sufficiency and ancillarity in the presence of nuisance parameters. *J. R. Stat. Soc., B* **37** 248–258.

[11] DAWID, A. P. (1976). Properties of diagnostic data distributions. *Biometrics* **32** 647–658.

[12] DAWID, A. P. (1979). Conditional independence in statistical theory. *J. R. Stat. Soc., B* **41** 1–31.

[13] DAWID, A. P. (1979). Some misleading arguments involving conditional independence. *J. R. Stat. Soc., B* **41** 249–252. MR0547250

[14] DAWID, A. P. (1980). A Bayesian look at nuisance parameters. In *Bayesian Statistics* (*Valencia*, 1979) (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 167–203. Univ. Press, Valencia. With discussion. MR0638877

[15] DAWID, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.* **8** 598–617. MR0568723

[16] DAWID, A. P. (1985). Invariance and independence in multivariate distribution theory. *J. Multivariate Anal.* **17** 304–315. MR0813238

[17] DAWID, A. P. (1998). Conditional independence. In *Encyclopedia of Statistical Sciences*, *Vol.* 2 (S. Kotz, C. B. Read and D. L. Banks, eds.) 146–155. Wiley-Interscience.

[18] DAWID, A. P. (2001). Separoids: A mathematical framework for conditional independence and irrelevance. *Ann. Math. Artif. Intell.* **32** 335–372.

[19] DAWID, A. P. (2001). Some variations on variation independence. In *Proceedings of Artificial Intelligence and Statistics* 2001 (T. Jaakkola and T. Richardson, eds.) 187–191.

[20] DAWID, A. P. (2002). Influence diagrams for causal modelling and inference. *Int. Stat. Rev.* **70** 161–189.

[21] DAWID, A. P. (2003). Causal inference using influence diagrams: The problem of partial compliance. In *Highly Structured Stochastic Systems* (P. J. Green, N. L. Hjort and S. Richardson, eds.). *Oxford Statist. Sci. Ser.* **27** 45–81. Oxford Univ. Press, Oxford. With part A by Elja Arjas and part B by James M. Robins. MR2082406

[22] DAWID, A. P. (2012). The decision-theoretic approach to causal inference. In *Causality*: *Statistical Perspectives and Applications* (C. R. Berzuini, A. P. Dawid and L. Bernardinelli, eds.) Chapter 4 25–42. Wiley, Chichester.

[23] DAWID, A. P. (2015). Statistical causality from a decision-theoretic perspective. *Annual Review of Statistics and Its Application* **2** 273–303.

[24] DAWID, A. P. and CONSTANTINOU, P. (2014). A formal treatment of sequential ignorability. *Stat. Biosci.* **6** 166–188.

[25] DAWID, A. P. and DICKEY, J. M. (1977). Likelihood and Bayesian inference from selectively reported data. *J. Amer. Statist. Assoc.* **72** 845–850.

[26] DAWID, A. P. and DIDELEZ, V. (2010). Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview. *Stat. Surv.* **4** 184–231. MR2740837

[27] DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272–1317. [Correction: *Ann. Statist.* **23** (1995) 1864.]

[28] DAWID, A. P. and MORTERA, J. (1996). Coherent analysis of forensic identification evidence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 425–443. MR1377842

[29] DAWID, A. P. and MORTERA, J. (1998). Forensic identification with imperfect evidence. *Biometrika* **85** 835–849. [Correction: *Biometrika* **86** (1999) 974.] MR1666687

[30] DAWID, A. P. and STUDENÝ, M. (1999). Conditional products: An alternative approach to conditional independence. In *Artificial Intelligence and Statistics* 99 (D. Heckerman and J. Whittaker, eds.) 32–40. Morgan Kaufmann.

[31] DIDELEZ, V., DAWID, A. P. and GENELETTI, S. (2006). Direct and indirect effects of sequential treatments. In *Proceedings of the* 22*nd Annual Conference on Uncertainty in Artificial Intelligence* (R. Dechter and T. S. Richardson, eds.) 138–146. AUAI Press, Arlington, VA.

[32] DURRETT, R. (2013). *Probability: Theory and Examples*, 4th ed. Cambridge Univ. Press. MR2722836

[33] FISHER, R. A. (1925). Theory of statistical estimation. *Math. Proc. Cambridge Philos. Soc.* **22** 700–725.

[34] GEIGER, D., VERMA, T. and PEARL, J. (1990). Identifying independence in Bayesian networks. *Networks* **20** 507–534. MR1064736

[35] GENELETTI, S. and DAWID, A. P. (2011). Defining and identifying the effect of treatment on the treated. In *Causality in the Sciences* (P. M. Illari, F. Russo and J. Williamson, eds.) 728–749.

[36] GHOSH, M., REID, N. and FRASER, D. A. S. (2010). Ancillary statistics: A review. *Statist. Sinica* **20** 1309–1332. MR2777327

[37] GUO, H. and DAWID, A. P. (2010). Sufficient covariates and linear propensity analysis. In *Proceedings of the Thirteenth International Workshop on Artificial Intelligence and Statistics* (*AISTATS*) 2010 (Y. W. Teh and D. M. Titterington, eds.). *Journal of Machine Learning Research Workshop and Conference Proceedings* **9** 281–288.

[38] GUO, H., DAWID, A. P. and BERZUINI, G. M. (2016). Sufficient covariate, propensity variable and doubly robust estimation. In *Statistical Causal Inferences and Their Applications in Public Health Research*. (H. He, P. Wu and D.-G. Chen, eds.). Springer. To appear.

[39] HALMOS, P. R. and SAVAGE, L. J. (1949). Application of the Radon–Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Stat.* **1** 225–241. MR0030730

[40] KINGMAN, J. F. C. (1993). *Poisson Processes*, 2nd ed. *Oxford Studies in Probability* **3**. Oxford Univ. Press, New York. MR1207584

[41] KOLMOGOROFF, A. (1942). Sur l'estimation statistique des paramètres de la loi de Gauss. *Bull. Acad. Sci. URSS. Sér. Math.* [*Izvestia Akad. Nauk SSSR*] **6** 3–32. MR0007959

[42] LAURITZEN, S. L., DAWID, A. P., LARSEN, B. N. and LEIMER, H.-G. (1990). Independence properties of directed Markov fields. *Networks* **20** 491–505. MR1064735

[43] PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, 1st ed. Morgan Kaufmann, San Mateo, CA. MR0965765

[44] PEARL, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge Univ. Press, Cambridge. MR1744773

[45] PEARL, J. and PAZ, A. (1986). Graphoids: Graph-based logic for reasoning about relevance relations or when would x tell you more about y if you already know z? In *Advances in Artificial Intelligence II*, *Seventh European Conference on Artificial Intelligence*, *ECAI* 86 (J. B. H. du Boulay, D. Hogg and L. Steels, eds.) 357–363.

[46] RESNICK, S. I. (2014). *A Probability Path*, 1st ed. Birkhäuser. MR3135152

[47] ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math*. *Modelling* **7** 1393–1512. MR0877758

[48] ROBINS, J. (1987). Addendum to "A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect". *Comput*. *Math*. *Appl*. **14** 923–945. MR0922792

[49] ROBINS, J. (1989). The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology*: *A Focus on AIDS* (L. Sechrest, H. Freeman and A. Mulley, eds.) 113–159.

[50] SPOHN, W. (1994). On the properties of conditional independence. In *Patrick Suppes*: *Scientific Philosopher* (P. Humphreys, ed.). *Synthese Lib.* **233** 173–196. Kluwer Acad. Publ., Dordrecht. MR1349350

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
UNITED KINGDOM
E-MAIL: p.constantinou.1@warwick.ac.uk

STATISTICAL LABORATORY
CENTRE FOR MATHEMATICAL SCIENCES
WILBERFORCE ROAD
CAMBRIDGE CB3 0WB
UNITED KINGDOM
E-MAIL: apd@statslab.cam.ac.uk