Dependence and Conditional Independence in Economic Models∗

(PRELIMINARY VERSION)

Luciano I. de Castro†

March 17, 2010

Abstract

De Finetti’s Theorem asserts that a sequence of exchangeable variables are conditionally independent, that is, dependence under exchangeability is nothing but conditionally independence. However, de Finetti’s theorem is valid only with a infinite number of random variables, which makes this result unsuitable for application in many economic models. In this paper, we prove that dependence is conditionally independence in a completely general setting (even without exchangeability). Moreover, we prove the existence of a minimally informative random variable that makes types conditionally independent. If this variable is known by all privately informed individuals, then all results that are valid under independence are also valid for such a model.

JEL Classification Numbers: C62, C72, D44, D82.

Keywords: affiliation, positive dependence, statistical dependence of types, conditional independence, de Finetti’s theorem, minimally informative random variable, auctions, pure strategy equilibrium, revenue ranking.


†Department of Economics, University of Illinois at Urbana-Champaign. 1206 South Sixth St., Champaign, IL, 61822. E-mail: decastro.luciano@gmail.com.
1 Introduction

Asymmetric information is a central theme in modern economics, not only in game theory, but also in industrial organization, general equilibrium, group decision, finance and many other subdisciplines. Most models assume that each agent privately knows a random variable, and these random variables are statistically independent. Although independence is convenient for theoretical manipulations, it is considered a restrictive and unrealistic assumption. Independence is regarded as restrictive because it is satisfied by a “knife-edge” set of distributions, and unrealistic because there are many potential sources of correlation in the real world: media, education, culture or even evolution. Perceiving these limitations early on, economists tried to surpass the mathematical difficulties and include statistical dependence in their models.

The introduction of affiliation was a milestone in the study of dependence in economics.¹ This remarkable contribution was made by Milgrom and Weber (1982a), who borrowed a statistical concept (multivariate total positivity of order 2, MTP₂) and applied it to a general model of symmetric auctions.² But the success of affiliation is not restricted to auction theory. Whenever information is important, affiliation may potentially be applied. In fact, researchers in many different areas of economics and finance used the concept to obtain useful results.³

However, as with any scientific achievement, affiliation has limitations, which are discussed by de Castro (2007). That paper serves as a motivation for the search of alternative approaches to the study of dependence in economics. Responding to this motivation, the present paper establishes new facts that may lead to different

---

¹Wilson (1969, 1977) was a precursor in the use of dependence and conditional independence in auction theory.

²In two previous papers, Milgrom (1981b) and Milgrom (1981a) presented results that used a particular version of the concept, under the name “monotone likelihood ratio property” (MLRP). It is also clear that Wilson (1969) and Wilson (1977) influenced the development of the affiliation idea. Nevertheless, the concept was fully developed and the term affiliation first appeared in Milgrom and Weber (1982a). See also Milgrom and Weber (1982b). When there is a density function, the property had been previously studied by statisticians under different names. Lehmann (1966) calls it Positive Likelihood Ratio Dependence (PLRD), Karlin (1968) calls it Total Positivity of order 2 (TP₂) for the case of two variables or Multivariate Total Positivity of Order 2 (MTP₂) for the multivariate case.

³For instance, Bergin (2001) used affiliation to obtain a generalization of a theorem by Aumann (1976) for the aggregation of information by a set of individuals; Persico (2000) proved a theorem about the usefulness of information for a decision maker under affiliation; and Sobel (2006) also used affiliation to study aggregation of information by groups. This list represents just a very small sample of papers; it would be almost impossible to cite all applications.
approaches in the study of dependence.

The main result is Theorem 4.2, which shows that all economic models fall into three cases: (1) types are independent; (2) types are dependent, but after receiving their private information, players’ conditional beliefs about other player’s types are independent; (3) types are truly statistically dependent, but there always exists a variable that makes types conditionally independent. In this case, we say that such variable _conditionally splits_ the types. This last part of Theorem 4.2 _does not_ come from de Finetti’s theorem. Theorem 4.2 also states that there is a minimally informative random variable that makes the signals of all individuals (conditionally) independent, that is, there is a minimally informative conditional splitter. In general there is no “least” informative conditional splitter in the sense of inclusion by $\sigma$-fields, but we introduce a new definition that establishes such an existence. Section ?? then discusses how this result can illuminate the study of dependence in economic models. We show that any result that holds under independence also holds true if the conditional splitter is known, even if the types are statistically dependent. Since a conditional splitter always exits and dependence is not an issue if the players know such conditional splitter, any approach to dependence has to rule out first the possibility that this conditional splitter is known by the players. Interestingly, this task may be very difficult to do in empirical works, but it can be done in laboratory settings with a control of players’ knowledge. This empirical/experimental question is fundamental in order to determine if (and in which circumstances) dependence is relevant in economic models.

We adopt a standard model of information, but it seems convenient to define our notation and setup in an explicit form. This is the purpose of section 2. Section 3 revises de Finetti’s theorem and shows its main limitation, that Theorem 4.2 covers. Section 4 contains the main results. Section 5 briefly reviews the literature and section 6 discusses potential directions for future work, not only in theory but also in econometrics and experimental economics.

## 2 Notation

For our purposes, it is sufficient to describe the information structure of an economic situation. We do not specify the payoffs or the actions, with the meaning that any such specifications are allowed, provided they are consistent with the private information structure.

There are $n$ individuals, $i = 1, \ldots, n$. Individual $i$ receives private information $t_i \in T_i$, where $T_i$ is a Polish space (complete metrizable topological space). $T_i$
is not assumed to be countable, finite or compact, but it can be. The usual notations $T \equiv \prod_{i=1}^{n} T_i$ and $t = (t_i, t_{-i}) = (t_1, ..., t_n) \in T$ are adopted. Assume that $(T, \mathcal{T}, \tau)$ is the corresponding Borel probabilistic space.

We may also consider sequences of random variables, $X_1, ..., X_n, ...$, with general distribution denoted by $Pr(\cdot)$. They are exchangeable if for any $n$ and permutation $\pi$, $(X_1, ..., X_n)$ has the same distribution as $(X_{\pi(1)}, ..., X_{\pi(n)})$.

3 De Finetti’s Theorem

Conditional independence models assume that the signals of bidders are conditionally independent, given a variable $v$ (the intrinsic value of the object, for instance). Since symmetry is the same as exchangeability, which is the main assumption of de Finetti’s Theorem, some specialists seem to believe that de Finetti’s Theorem implies that conditional independence holds in symmetric models without loss of generality. De Finetti’s theorem states the following:

**De Finetti’s Theorem.** Consider a sequence of random variables $X_1, X_2, ..., X_n$, and assume that they are exchangeable, that is, assume that the distribution of $(X_1, ..., X_n)$ is equal to the distribution of $(X_{\pi(1)}, ..., X_{\pi(n)})$, for any $n$ and any permutation $\pi : \mathbb{N} \rightarrow \mathbb{N}$. Then, there is a random variable $Q$ such that all $X_1, X_2, ...$, are conditionally independent (and identically distributed) given $Q$.

Unfortunately, however, de Finetti’s theorem is not valid for standard models of games, even assuming symmetry. The reason is that standard games (auctions, in particular) consider only a finite number of players and, hence, a finite number of random variables. De Finetti’s theorem is valid only for an (infinite) sequence of random variables. The following example illustrates the problem:

---

4By Borel space, we mean that $\mathcal{T}$ is the $\sigma$-field generated by the open sets of the metric space $T$. Also, types are withdraw with respect to $\tau$.

5De Finetti proved this theorem for the case where the $X_i$ are Bernoulli variables. Hewitt and Savage (1955) extended it to the general setting. The statement above is somewhat vague. A precise statement is as follows: Let $X_1, X_2, ...$, be an exchangeable sequence of random variables with values in a set $S$. Then there exists a probability measure $\mu$ on the set of probability measures $\Delta(S)$ such that for all measurable sets $A_1, ..., A_n$,

$$Pr(X_1 \in A_1, ..., X_n \in A_n) = \int_{\Delta(S)} Q(A_1) \cdots Q(A_n) \mu(dQ).$$

6One can assume that there are an infinite number of potential players in the game, but for
Example 3.1 Consider two random variables, $X_1$ and $X_2$, taking values in $\{0, 1\}$, with joint distribution given by: $P(X_1 = 0, X_2 = 1) = P(X_1 = 1, X_2 = 0) = \frac{1}{2} - \varepsilon$ and $P(X_1 = 0, X_2 = 0) = P(X_1 = 1, X_2 = 1) = \varepsilon$. It is easy to see that $X_1$ and $X_2$ are symmetric (exchangeable). In the appendix, we show that the conclusion of de Finetti’s Theorem cannot hold if $\varepsilon < 1/4$.

Thus, de Finetti’s Theorem does not imply that conditional independence is a generic condition in economic models. The use of de Finetti’s theorem is granted only if our model has an infinite number of individuals and they are symmetric (exchangeable). Otherwise, the theorem does not apply. Although those assumptions can be considered reasonable in some settings, they are excessively restrictive in other models. The purpose of this paper is exactly to offer an alternative statement for finite number of not necessarily symmetric individuals.

4 Conditional independence for finite non-symmetric variables

This section shows that there always exists a minimally informative random variable that makes any set of random variables conditionally independent. Theorem 4.2 also provides a classification of the cases of statistical dependence where the dependence matters and where it does not matter (all the results that hold under independence can be used). This result suggests that independence may be actually less restrictive that appears at first sight. However, of course there are cases where dependence is truly important.

Consider the following definition:

**Definition 4.1 (Conditional splitter)** Let $(\Omega, \Sigma, \Pr)$ be a probabilistic space, such that $\Omega$ is a Polish (complete separable metrizable) space. Given $\sigma$-fields $\mathcal{F}_i \subset \Sigma$, $i = 1, \ldots, n$, we say that $\mathcal{F}$ conditionally splits (or $\mathcal{F}$ is a conditional splitter of) $\mathcal{F}_1, \ldots, \mathcal{F}_n$ if $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are conditionally independent given $\mathcal{F}$. We say that a some reason only a finite number of them actually participate. Then, one can apply de Finetti’s theorem. However, this will be of course with a loss of generality.

7See also Proposition 4.7 below. Example 3.1 generalizes an example given by Diaconis and Freedman (1980). They prove an approximation version of de Finetti’s theorem for a finite set of random variables. See a discussion of their paper after Proposition 4.7.
variable $Z$ conditionally splits variables $X^1, \ldots, X^n$ if the $\sigma$-field generated by $Z$, denoted $\sigma(Z)$, conditionally splits $\sigma(X^1), \ldots, \sigma(X^n)$.

It is useful to observe that the information content of a variable $Z$ must refer to the $\sigma$-field $\sigma(Z)$ associated with it, and not with the variable’s value. For instance, it is clear that the variable $Y = 2Z$ contains exactly the same information as $Z$ does and, as natural, $\sigma(Y) = \sigma(Z)$ but $Y \neq Z$. The following theorem, whose statement and proof are slightly informal (a more formal statement and proof are given in appendix ??), is the main result of this section.

**Theorem 4.2** Consider a game of asymmetric information with $n$ players, such that each player $i = 1, \ldots, n$ receives a random variable (type) $t_i \in T_i$ and there is a joint distribution on all types. Then one of the following three alternatives happens:

1. the types are statistically independent;

2. the types are statistically dependent, but the common knowledge information is a conditional splitter;\footnote{This means that when each player receives his type, it becomes common knowledge that the (conditional) beliefs are independent.}

3. the types are statistically dependent, the common knowledge information is not a conditional splitter, but a conditional splitter exists.\footnote{Thus, if players become aware of the outcome of this conditional splitter, case 3 is converted into case 2.}

Moreover, in this case:

(a) any conditional splitter contains strictly more information than it is common knowledge for the players;\footnote{Note that this statement requires proof, as it could in principle be the case that a random variable with information completely different from the common knowledge information is a conditional splitter. Note also a subtle aspect of this statement: if $Z$ is a conditional splitter and $C$ is a variable representing the common knowledge information, statement 3(a) says only that $\sigma(Z) \supset \sigma(C)$. It may happen that there exists a variable $Y$, with $\sigma(Y) \not\subset \sigma(C)$ such that if the players are informed of $Y$, the types will be conditionally independent. This only means that $\sigma(Y, C)$ is a conditional splitter and this does not contradict statement 3(a), since obviously $\sigma(Y, C) \supset \sigma(C)$.}
(b) there exists a conditional splitter that is minimally informative (in the sense of inclusion by $\sigma$-fields);\textsuperscript{12}

(c) if the support of types is a finite set, there is an algorithm to find all the minimally informative conditional splitters;\textsuperscript{13}

(d) in general, there is not a least informative conditional splitter in the sense of inclusion by $\sigma$-fields;\textsuperscript{14}

(e) nevertheless, in the finite case there is a least informative conditional splitter in the sense of proximity of conditional beliefs.\textsuperscript{15}

Case 1 in Theorem 4.2 is familiar and requires no comments. Cases 2 and 3 seem to require less and deliver more than de Finetti’s theorem. While de Finetti’s theorem requires exchangeability and an infinite number of random variables (as we showed in example 3.1), case 3 in Theorem 4.2 covers the case of a finite number of random variables that are not necessarily exchangeable and states the existence of a variable that makes all types conditionally independent. More important, at first glance, this existence seems to be in contradiction with example 3.1. The contradiction is, of course, only apparent. The difference between the two settings is that Theorem 4.2 does not assume nor does it deliver symmetric distributions, while de Finetti’s theorem requires the conditional distribution to be symmetric (exchangeable) but also delivers identical (i.i.d.) distributions.

\textsuperscript{12}By “minimally informative” we mean the following: if $Y$ is another variable that makes the types conditionally independent, and $Y$ contains less information than $Z$ (i.e., $\sigma(Y) \subset \sigma(Z)$), then $Y$ contains as much information as $Z$ ($\sigma(Y) = \sigma(Z)$). Note that the existence of minimally informative variables is not trivial: conditional independence is not preserved under the intersection of $\sigma$-fields.

\textsuperscript{13}Naturally, our algorithm produces $\sigma$-fields, not exactly specific variables.

\textsuperscript{14}By least informative conditional splitter we mean the following: if $Y$ is another conditional splitter, then $\sigma(Y) \supset \sigma(Z)$. This least informative conditional splitter would contain strictly more information that the common knowledge, unless we are in case 2 instead of case 3 of this Theorem.

\textsuperscript{15}This statement is inaccurate. We represent conditional expectations by the associated Markov transitions and observe that they can be seen as functions in $L^2$. Therefore, we can show that there is a unique conditional splitter that is the closest one (in the standard $L^2$ norm) to the conditional expectation given the common knowledge information. A formal description of our notion requires a number of technical definitions, which we prefer to postpone to the appendix. There are good reasons for using this definition, but, of course, it can be disputed. Our objective here is just to show that it is possible to give a reasonable definition of “least informative” conditional splitter that allows to obtain existence. Since this is not a central point for this paper’s objective, we will not discuss this further.
Therefore, Theorem 4.2 is incomparable to de Finetti’s theorem. The following example illustrates this matter.

**Example 4.3 (Example 3.1 continued.)**

Consider the same distribution described in Example 3.1 and fix $\varepsilon = \frac{1}{6} < \frac{1}{4}$.

As stated in Example 3.1, there is no variable $Z$ such that $X|Z$ and $Y|Z$ are conditionally independent and symmetric. However, consider a variable $Z \in \{0, 1\}$ such that the joint distribution with $X$ and $Y$ is given by:

\[
\begin{bmatrix}
a_{00} & a_{01} & a_{001} & a_{011} \\
a_{10} & a_{11} & a_{101} & a_{111}
\end{bmatrix} =
\begin{bmatrix}
\frac{3}{48} & \frac{1}{48} & \frac{5}{48} & \frac{15}{48} \\
\frac{15}{48} & \frac{5}{48} & \frac{1}{48} & \frac{3}{48}
\end{bmatrix},
\]

where $a_{ijk} = \Pr(X = i, Y = j, Z = k)$, for $(i, j, k) \in \{0, 1\}^3$. If we write $a_{ijk}$ for $\Pr(X = i, Y = j|Z = k)$, for $(i, j, k) \in \{0, 1\}^3$, we obtain:

\[
\begin{align*}
\begin{bmatrix}
a_{00} & a_{01} \\
a_{10} & a_{11}
\end{bmatrix} &=
\begin{bmatrix}
\frac{3}{24} & \frac{1}{24} \\
\frac{15}{24} & \frac{5}{24}
\end{bmatrix};
\begin{bmatrix}
a_{001} & a_{011} \\
a_{101} & a_{111}
\end{bmatrix} &=
\begin{bmatrix}
\frac{5}{24} & \frac{15}{24} \\
\frac{1}{24} & \frac{3}{24}
\end{bmatrix}.
\end{align*}
\]

Note that $\Pr(X = 0|Z = 0) = \frac{1}{6}$ and that $\Pr(Y = 0|Z = 0) = \frac{3}{4}$. Therefore, $\Pr(X = 0, Y = 0|Z = 0) = \frac{3}{4} = \Pr(X = 0|Z = 0) \Pr(Y = 0|Z = 0)$. A similar verification can be done to all other cases, showing that $X|Z$ is independent of $Y|Z$ and, consequently, $Z$ makes $X$ and $Y$ conditionally independent (is a conditional splitter). Note, however, that the claim in example 3.1 is not violated, since the conditional distributions in (1) are not symmetric.

Figure 2 below illustrates cases 2 and 3 in Theorem 4.2. There are two players with continuous types $t_1$ and $t_2$ whose support is the union of the rectangles showed in Figure 2. The random variable $Z$ indicates the rectangle that contains the realization of $t_1$ and $t_2$; the types are conditionally independent given the rectangle (which occurs, for instance, if the distribution is uniform in each rectangle). The darker the rectangle, the bigger the probability of that rectangle. Note that the information contained in the variable $Z$ is common knowledge in case 2, and it is not completely informative (does not imply the knowledge of the other player’s type) in case 3. Note also that even in case 3, it is possible that some realization of $Z$ is common knowledge, as shown in Figure 2 (d).

Theorem 4.7 below provides necessary and sufficient condition for the conclusion of de Finetti’s theorem with a finite number of variables in a setting that covers examples 3.1 and 4.3.
Although case 2 seems special, it is not possible to say, from a theoretical point of view, whether it is typical or not in economic applications. An example will illustrate this claim. Suppose that an econometrician observes data on wine auctions. Analyzing the data, the econometrician observes that the bids (and therefore the values, assuming symmetry and affiliation) are extremely positively correlated. However, the vintage, the producer (and in some cases, the previous prices) of that wine are common knowledge to market participants. If the econometrician also has this common knowledge information and controls (conditions) on it, this large correlation will reduce and maybe even disappear. If it disappears, this corresponds exactly to case 2. In sum, it may be very difficult for an econometrician to separate cases 2 and 3. In this sense, the additional information provided for case 3 in Theorem 4.2 may be useful to understand the kind of dependence that the econometrician finds.

The proof of Theorem 4.2 also gives conditions that characterize the three cases in its statement (see lemma A.16). Although the proof is too long to describe here, it is useful to state separately a step in the proof that may be of interest by itself.

**Proposition 4.4** Let $F_1, ..., F_n$ be the sub-$\sigma$-fields of $\Sigma$. There exists a $\sigma$-field $Z \subset \Sigma$ such that $F_1, ..., F_n$ are conditionally independent given $Z$. More specifically, given random variables $X^1, ..., X^n$, then there exists a random variable $Z$ that makes them conditionally independent.

While Proposition 4.4 only states the existence of a random variable that makes the types conditionally independent, Theorem 4.2 deals also with minimally informative variables. Thus, Proposition 4.4 can be proven with a complete

---

17The results presented in Theorem 4.2 can also shed some light in the problem of unobserved heterogeneity, which is a topic not completely understood in the econometric literature. See Athey and Haile (2005) and Krasnokutskaya (2009).
informative variable $Z$, i.e., a variable that contains all the information about $X^1$, $X^n$. That this variable exists may not seem completely obvious. In particular, since $X = (X^1, \ldots, X^n)$ is an $n$-dimensional, one may be confused by the fact that all information in a model with multidimensional information can be summarized by a one-dimensional variable ($Z$). As it turns out, the information contained in any vector of random variable (with values in $\mathbb{R}^n$) can be summarized by a single-dimensional random variable.\footnote{This observation is interesting by itself, because it suggests that the gap between the results obtained for one-dimensional and multidimensional information models is not due to information complexity, but rather to techniques employed to obtain those results, such as techniques based on order, monotonicity, calculus, etc.}

The following (coding) argument can be instructive.

Assume, without loss of generality, that $X^i$ is in $[0, 1]$. For each $i = 1, 2, \ldots$, write $X^i$ as $0.X_{i1}X_{i2}X_{i3}$. From this, define $Z$ as the r.v. in $[0, 1]$ whose realization is given by:

$$Z = 0.X_{11}X_{21} \ldots X_{n1}X_{12}X_{22} \ldots X_{n2}X_{13}X_{23} \ldots$$

Recall that two random variables $X$ and $Y$ are conditional independent given $Z$ if and only if the additional knowledge of $Y$ does not improve the assessment of $X$ once one knows $V$, that is, $\Pr(X|Z) = \Pr(X|Y, Z)$. It is easy to see that $Z$ contains all the information that $(X^1, X^2, \ldots, X^n)$ contains, that is, $\Pr(X^i|Z) = \Pr(X^i|X^j, Z)$.

Unfortunately, however, this “coding argument” is not a formal proof because it considers conditioning with respect to null events ($Z = z$). As it is well-known, a number of paradoxes may arise from this kind of procedure. See, for example, Billingsley (1995, Exercise 33.1, p. 441) and the following quote: “There are pathological examples showing that the interpretation of conditional probabilities in terms of an observer with partial information breaks down in certain cases.” (Billingsley (1995, p. 437)) This “proof” is useful, however, from an intuitive point of view, because it appeals directly to the notion of information. Consider the following quote from Billingsley (1995, p. 58-9): “The heuristic equating of $\sigma$-field and information is helpful even though it sometimes breaks down, and of course proofs are indifferent to whatever illusions and vagaries brought them into existence.”

The confusion about the use of de Finetti’s theorem to deliver conditional independence (discussed after example 3.1) shows that the statement and proof of Proposition 4.4 is already helpful.\footnote{For finite variables, the result contained in Proposition 4.4 is the main result of Suppes and} The result itself, however, is not totally useful
from a practical point of view, because it is based on a fully informative random variable. This fact highlights the minimal informativeness established by Theorem 4.2. This result is of economic importance, because the less information is sufficient to make bidders conditionally independent, the easier it is for bidders to acquire it. And if conditional independence turns out to be common knowledge, as in case 2, then dependence does not matter: we can apply any result that is valid for independent variables.

**Lemma 4.5** Any result that holds with independent types also holds with the statistically dependent variables in case 2. \(^{20}\)

**Proof.** Although the types in case 2 are statistically dependent, they are independent given the common knowledge information observed in the interim stage. Following the arguments in Harsanyi (1967-8), this interim stage is exactly the realistic stage from where we construct the ex ante stage (and impose the common prior assumption). For the given realization of types in the interim stage, it is not necessary that the ex ante stage be the original one: it can just be the event whose occurrence is common knowledge. In that event, the variables are independent. The result follows. □

An immediate corollary of Lemma 4.5 is:

**Corollary 4.6** The Revenue Equivalence Theorem (RET) holds for dependent types in case 2 (provided the other assumptions of this theorem also hold).

In particular, the variables illustrated in Figure 2 (b) are truly affiliated, but the RET holds for this case. Thus, it is important for the study of dependence in economics to be able to distinguish when we are in case 2. For econometric applications, this seems a rather difficult task, because one has to know what is common knowledge among the participants, and this piece of information may be unobservable to the econometrician. In the wine auctions example given above, it is possible that more than just the vintage is common knowledge; for example, some brochure could have been distributed or some explanation (information) Zanotti (1981). The generalization for continuous variables is not straightforward because, as we discussed above, it is not correct to condition on zero measure events. See remark ?? in appendix ?? for a review of results related to Theorem 4.2 and Proposition 4.4.

\(^{20}\)This statement is slightly informal, but its content should be clear: the conclusions hold conditionally to each piece of common knowledge information. For instance, we can have monotone pure strategy equilibria in each common knowledge part, but this does not imply that the equilibria will be overall monotonic.
about the object may have been given at the the time of the auction, which the econometrician is not aware and/or cannot use as a control.

Even if the situation corresponds to case 3 instead of case 2, it is possible that \( Z \) can be learned with the repetition of the game. Alternatively, the information contained in \( Z \) may be available for acquisition (from a consultant or a spy, for instance). If \( Z \) becomes common knowledge, case 3 will reduce to case 2. Thus statistical dependence is only relevant in case 3 with a \( Z \) that cannot be (is not) learned. Even if a conditional splitter cannot be learned, it may happen that after conditioning to the common knowledge, the remaining correlation is so small that independence already gives a good approximation, since revenue and equilibria vary continuously with the distribution.\(^{21}\) In sum, independence can already give a good approximation in many cases of interest. However, if this is true or not is an empirical/experimental question, not a theoretical one. The following summarizes the discussion in this section:

1. There is always a random variable (r.v.) \( Z \) that makes the types of the bidders conditionally independent.

2. If this r.v. is common knowledge, then the appropriate framework is that of independent types models.

3. If this r.v. can be learned with time, then the game will converge to the previous situation.

4. Only if the r.v. cannot become common knowledge, then a model with dependence is truly necessary.

   Therefore:

5. It is important to test whether or not independence is a good approximation in economically relevant environments.

### 4.1 De Finetti’s Theorem for a finite number of random variables

It was noted earlier that Theorem 4.2 ensures the existence of conditional independence but it does not deliver de Finetti’s Theorem implication that the conditional

\(^{21}\)This statement can be formalized by using the sup-norm in the space of continuous densities. This observation is sufficiently simple and intuitive, so that we do not pursue its formalization. See, however, items 3 and 4 in Theorem ?? below for related results.
probabilities are identical, that is, symmetric. Although the implication given in Theorem 4.2 is sufficient for our purposes, in some applications one may be interested in the stronger implication, which we will call symmetric conditional independence. This terminology means that the variables are not only conditionally independent given $Z$, but also symmetrically (identically) distributed. The full exploration of this question is beyond the scope of this paper, but we present here a theorem in a simple setting in which the question can be totally clarified.

Theorem 4.7 Let $X$ and $Y$ be symmetric (exchangeable) binary random variables that are statistically dependent. Then there is a binary random variable $Z$ that makes $X$ and $Y$ symmetrically conditionally independent if and only if $X$ and $Y$ are positively correlated.

Theorem 4.7 generalizes the claim made in Example 3.1. It should be noted that Diaconis (1977) and Diaconis and Freedman (1980) have previously studied versions of de Finetti’s theorem for a finite set of variables, but these papers do not present any result comparable to Theorem 4.7. They are interested in how the conclusion of de Finetti’s theorem is asymptotically true for a finite set of exchangeable variables when the size of the set increases. In contrast, we are interested in a fixed (and small) set of random variables and look for a (necessary and sufficient) condition for de Finetti’s conclusion to hold exactly.

5 Related literature

Although Proposition 4.4 is known by specialists, we were unable to find a good reference for it. The closest reference is Suppes and Zanotti (1981), who state the existence of a (fully informative) $Z$ when $t_i$ are binary variables. It is interesting to note that Holland and Rosenbaum (1986, p.1525) quote Suppes and Zanotti’s theorem and say that their proof “is easily generalized to the discrete case” and that “any distribution on $\mathbb{R}^J$ may be approximated arbitrarily well by a discrete distribution of $\mathbb{R}^J$, and Theorem 1 [Suppes-Zanotti] applies to such discrete approximation.” They do not state, discuss or give any references for the result established in Proposition 4.4. We also found no reference for Proposition 4.4 in Probability text-books. Mouchart and Rolin (1984) and van Putten and van Schuppen (1985) prove the first part of Proposition 4.4 (that is, Lemma A.7 in the appendix), for the case of $n = 2$. It is clear that those authors could have stated and proved Lemma A.7, in which case we would just quote them. A similar comment
is valid for the existence of random variables (the second part of Proposition 4.4), which does not follow immediately from Lemma A.7. The closest references for Theorem 4.2 that we were able to find (unfortunately after obtaining our results) was section 4 of Mouchart and Rolin (1984) (see also section 5 of van Putten and van Schuppen (1985) and Mouchart and Rolin (1985)). They discuss what they call “σ-algebraic realization problem”, which is the problem of finding a “minimal conditional independence relation $\text{CI}_{\text{min}}$” and proved item 3(b) of Theorem 4.2 for the particular case of $n = 2$ players. The classification and the algorithm provided by Theorem 4.2 are completely new. The contribution given in item 3(e) of Theorem 4.2 (definition and existence of least informative conditional splitter) is also new.

6 Final remarks

Theorem 4.2 shows that it is not possible to rule out independence without a control of what is common knowledge for the participants. In other words, it is virtually impossible to assert what kind of dependence is typical from a purely theoretical point of view. This impossibility raises the question of testing dependence in more general terms. Again, Theorem 4.2 implies that this test requires special care with characteristics that are unobservable but may be (commonly) known by market participants.

Experimental studies could shed light on the actual distribution of values across individuals, controlling for the common knowledge. It would be very helpful to develop methods to determine the values that people attribute to objects in an auction and whether those values are correlated or not. With respect to econometrics, an obvious need is to develop methods to test the affiliation of bidders’ values, controlling for the common knowledge (if this is possible). It would also be useful to develop techniques to describe the kind of dependence of the bids in real auctions. It would be very helpful to learn whether the kind of dependence is different across different markets and how these differences can be characterized. For instance, is there less correlation in Internet auctions, where the participants are consumers with almost no interaction, than in auctions where the participants are firms or professionals acting in the same industry? Yet another direction of research would be the development of econometric techniques to deal with dependence out of affiliation (see de Castro and Paarsch (2008)).

In sum, there is much yet to be done to fully understand dependence in economics.
A Proofs

A.1 Proof of Example 3.1.

Although this is implied by Theorem 4.7, we give here a direct proof for this example. Let \( p \) denote the probability of Heads and let \( \mu \) be a distribution over coins. Then:

\[
\text{Pr}(\text{Heads,Heads}) = \varepsilon = \int (p)^2 \mu(dp), \quad \text{and} \quad \text{Pr}(\text{Tails,Tails}) = \varepsilon = \int (1-p)^2 \mu(dp) = \int (1) \mu(dp) + \int (-2p) \mu(dp) + \int (p)^2 \mu(dp) = 1 - 2E[p] + \varepsilon.
\]

Then, \( 1 - 2E[p] = 0 \), or \( E[p] = 1/2 \). This implies: \( \text{Var}[p] = \int (p - E[p])^2 \mu(dp) = \int (p^2 - p + \frac{1}{4}) \mu(dp) = \int (p)^2 \mu(dp) - \frac{1}{4} = \varepsilon - \frac{1}{4} \). Since \( \text{Var}[p] \) is non-negative, \( \varepsilon \geq \frac{1}{4} \).

A.2 Preliminary probabilistic facts

Let \((\Omega, \Sigma, \text{Pr})\) be a probability space, where \( \Omega \) is a Polish (separable complete metrizable) space and \( \Sigma \), its Borel \( \sigma \)-field, unless otherwise specified. Let \( \mathcal{B} \) denote the Borel \( \sigma \)-field in \( \mathbb{R} \) (the set of real numbers). Given a \( \sigma \)-field \( \mathcal{F} \subset \Sigma \), a random variable \( Y \) is \( \mathcal{F} \)-measurable if \( Y^{-1}(B) \equiv \{ \omega : Y(\omega) \in B \} \in \mathcal{F} \) for every \( B \in \mathcal{B} \). Let \( \sigma(Y) \) denote the smallest \( \sigma \)-field with respect to which \( Y \) is measurable. If \( \mathcal{C} \) is a class of sets, \( \sigma(\mathcal{C}) \) denotes the \( \sigma \)-field generated by \( \mathcal{C} \), that is, the smallest \( \sigma \)-field containing \( \mathcal{C} \). Let \( \Sigma^o \) denote the set of \( \Sigma \)-measurable null sets, that is, \( \Sigma^o \equiv \{ A \in \Sigma : \Pr(A) = 0 \} \). The completion of \( \mathcal{F} \subset \Sigma \) is \( \bar{\mathcal{F}} \) given by \( \sigma(\mathcal{F} \cup \Sigma^o) \). As usual, \( F \Delta G \) denotes \( (F \cap G^c) \cup (F^c \cap G) \). We have the following:

Lemma A.1 \( F \in \bar{\mathcal{F}} \) if and only if there is \( G \in \mathcal{F} \) such that \( \Pr(F \Delta G) = 0 \).

**Proof.** Define \( \bar{\mathcal{F}} \equiv \{ F \in \Sigma : \exists G \in \mathcal{F} \text{ such that } \Pr(F \Delta G) = 0 \} \). It is clear that \( \bar{\mathcal{F}} \supseteq \mathcal{F} \cup \Sigma^o \). The fact that \( F^c \Delta G^c = F \Delta G \) implies that \( \bar{\mathcal{F}} \) is closed to complementation. If \( F_1, ..., F_n, ... \in \bar{\mathcal{F}} \), with corresponding \( G_1, ..., G_n, ... \in \mathcal{F} \), such that \( \Pr(F_n \Delta G_n) = 0 \), let \( F = \bigcup_{n \in \mathbb{N}} F_n \) and \( G = \bigcup_{n \in \mathbb{N}} G_n \). Then, \( F^c \cap G = (\cap_n F_n^c) \cap \cup_n G_n \subset \cup_{n \in \mathbb{N}}(F_n^c \Delta G_n) \) and similarly for \( F \cap G^c \). Thus, \( F \Delta G \subset \bigcup_{n \in \mathbb{N}}(F_n \Delta G_n) \). Therefore, countable additivity implies that \( \bar{\mathcal{F}} \) is closed to countable unions, which shows that \( \bar{\mathcal{F}} \) is a \( \sigma \)-field. Since it contains \( \mathcal{F} \cup \Sigma^o \), \( \bar{\mathcal{F}} \supseteq \mathcal{F} \). On the other hand, assume that \( F \in \bar{\mathcal{F}} \) and let \( G \in \mathcal{F} \) be such that \( F \Delta G \in \Sigma^o \). Then \( G \cap F^c \in \Sigma^o \subset \bar{\mathcal{F}} \) and \( F \cup G = G \cup (F \Delta G) \in \bar{\mathcal{F}} \). Therefore, \( F = (F \cup G) \setminus (G \cap F^c) \in \bar{\mathcal{F}} \), which shows that \( \bar{\mathcal{F}} \supseteq \mathcal{F} \) and concludes the proof. 

Definition A.2 We say that \( \mathcal{F} \) and \( G \) are equivalent \( \sigma \)-fields if \( \bar{\mathcal{F}} = \bar{G} \).
We will consider random variables \(X^i: \Omega \rightarrow \mathbb{R}\), for \(i = 1, \ldots, n\) and \(Z\). The vector \((X^1, \ldots, X^n)\) will be denoted by \(X\). We will denote the \(\sigma\)-fields \(\sigma(X^i)\), \(\sigma(X_1, \ldots, X_n)\) and \(\sigma(Z)\) by \(X^i\), \(X\) and \(Z\), respectively.

**Definition A.3** Given a \(\sigma\)-field \(\mathcal{F} \subseteq \Sigma\), a regular conditional probability given \(\mathcal{F}\) is a function \(Q: \Omega \times \Sigma \rightarrow \mathbb{R}_+\) satisfying:

(a) \(\omega \mapsto Q(\omega, A)\) is \(\mathcal{F}\)-measurable, for any \(A \in \Sigma\);

(b) for every \(B \in \mathcal{F}\) and \(A \in \Sigma\),

\[
\int_B Q(\omega, A) \, d\Pr(\omega) = \Pr(A \cap B);
\]

(c) for each \(\omega\), \(Q(\omega, \cdot)\) is a (countably additive) probability measure on \((\Omega, \Sigma)\).

In this case, the (regular) conditional probability \(Q(\omega, A)\) will be denoted by \(\Pr(A | \mathcal{F})_\omega\). Following the usual practice, \(\omega\) will be sometimes omitted in \(\Pr(\cdot | \mathcal{F})_\omega\).

Although conditional probabilities always exist, sometimes there does not exist a regular conditional probability. See Billingsley (1995, Exercise 33.11, p. 443). However, regular conditional probabilities always exist if \(\Omega\) is a Polish (complete, separable, metrizable) space and \(\Sigma\) is its Borel field, as we assume here. (See Billingsley (1995, Theorem 33.3, p. 439)). Thus, we will consider only regular conditional probabilities in what follows, and refer to them simply as conditional probabilities.

**Definition A.4** We say that the sub-\(\sigma\)-fields \(\mathcal{F}^1, \ldots, \mathcal{F}^n\) are conditionally independent given \(\mathcal{F} \subseteq \Sigma\) if for any \(A^i \in \mathcal{F}^i \subseteq \Sigma\), for \(i = 1, \ldots, n\), we have:

\[
\Pr(\cap_i A^i | \mathcal{F}) = \prod_i \Pr(A^i | \mathcal{F}).
\]

We denote this by \(\perp \perp (\mathcal{F}^1, \ldots, \mathcal{F}^n) | \mathcal{F}\) or, if \(n = 2\), by \(\mathcal{F}^1 \perp \perp \mathcal{F}^2 | \mathcal{F}\) and we say that \(\mathcal{F}\) conditionally splits (or is a conditional splitter of) \((\mathcal{F}^1, \ldots, \mathcal{F}^n)\). We say that the random variables (r.v.) \(X^1, \ldots, X^n\) are conditionally independent given a r.v. \(Z\) if \(\perp \perp (X^1, \ldots, X^n) | Z\) and we also denote this by \(\perp \perp (X^1, \ldots, X^n) | Z\). In this case, we also say that \(Z\) conditionally splits (or is a conditional splitter of) \((X^1, \ldots, X^n)\).

---

\(^{22}\)Our theory can be generalized for random variables taking values in more general spaces, but this seems sufficient for our purposes and avoid unnecessary complications.

\(^{23}\)If only the first two conditions are satisfied, we have just a conditional probability (not regular).
The following non-trivial results will be needed:

**Lemma A.5 (Mutual Conditional Independence)** The \( \sigma \)-fields \( F_t, t \in T \) are conditionally independent given \( F \) (denoted \( \perp \perp F_t | F \)) if and only if \( F_S \perp \perp F_{T \setminus S} | F \) for all sets \( S \subset T \), where \( F_S \equiv \bigvee_{t \in S} F_t \) is the \( \sigma \)-field generated by \( \bigcup_{t \in S} F_t \).

**Proof.** It is sufficient to adapt the proof of Kallenberg (2002, Lemma 3.8 (ii), p. 51) to conditional independence. 

**Lemma A.6 (Doob)** For any \( \sigma \)-fields \( F, G \) and \( H \), \( F \perp \perp G | H \) if and only if
\[
\Pr[H | F, G] = \Pr[H | G] \text{ a.s., } \forall H \in \mathcal{H}.
\]

**Proof.** See Kallenberg (2002, Proposition 6.6, p. 110).

### A.3 Proof of Proposition 4.4

To prove Proposition 4.4, we need the following lemma, which is in fact the first half of that proposition.

**Lemma A.7** Let \( F_1, \ldots, F^n \) be the sub-\( \sigma \)-fields of \( \Sigma \). There exist a \( \sigma \)-field \( Z \subseteq \Sigma \) such that \( F_1, \ldots, F^n \) are conditionally independent given \( Z \).

**Proof.** For each set \( S \subset N \equiv \{1, \ldots, n\} \), let \( Z^S \equiv \bigvee_{i \in S} F_i \), that is, \( Z^S \) denotes the \( \sigma \)-field generated by \( \bigcup_{i \in S} F_i \). For simplicity, we write \( Z \) instead of \( Z^N \). By lemma A.5, it is sufficient to prove that \( Z^S \) and \( Z^{S'} \) are conditionally independent given \( Z \) for all \( S \subset N \). By lemma A.6, this follows by establishing \( \Pr[H | Z^S, Z] = \Pr[H | Z] \), for all \( H \in Z^{S'} \). But since \( \sigma(Z^S, Z) = Z \), the conditional probability is the same in both sides of this equation. 

Observe that the existence of \( Z \) given in Lemma A.7 above is yet not sufficient for the last statement in Proposition 4.4 because it is not always true that given a \( \sigma \)-field \( Z \), there exists a r.v. \( Z \) such that \( Z = \sigma(Z) \). For this to hold, it is necessary that \( Z \) is countably generated. A \( \sigma \)-field \( F \) is countably generated if there is a countable class of sets \( A_1, A_2, \ldots \) such that \( F = \sigma(A_1, A_2, \ldots) \). Indeed, we have the following:
Claim A.8  There exists r.v. $Z$ such that $\mathcal{F} = \sigma(Z)$ if and only if $\mathcal{F}$ is countably generated.\footnote{See Billingsley (1995, Exercise 20.1, p. 270). We include the proof for completeness.}

Proof. If $\mathcal{F} = \sigma(A_1, A_2, \ldots)$, define: $Z(\omega) = \sum_{k=1}^{\infty} 3^{-k} 1_{A_k}(\omega)$. It is easy to see that $\sigma(Z) = \sigma(A_1, A_2, \ldots)$. Conversely, assume that $\mathcal{F} = \sigma(Z)$ for some r.v. $Z$. Since $Z$ is $\mathcal{F}$-measurable, for each $B \in \mathcal{B}$, $Z^{-1}(B) \in \mathcal{F}$. Take a countable class of sets $A_1, A_2, \ldots$ which generate $\mathcal{B}$ (for instance, the class of intervals $(a, b)$, for $a, b \in \mathbb{Q}$). Since $A_k \in \mathcal{B}$, $Z^{-1}(A_k) \in \mathcal{F}$, which implies $Z^{-1}(\sigma(A_1, A_2, \ldots)) = Z^{-1}(B) \subset \mathcal{F}$. But since $\mathcal{F} = \sigma(Z)$ is the smallest $\sigma$-field with respect to which $Z$ is measurable, $\mathcal{F} \subset \sigma(Z^{-1}(B)) = Z^{-1}(B)$, concluding the proof of the claim.

Given a countably generated $\sigma$-field $\mathcal{F}$ and a sub-$\sigma$-field $\mathcal{G} \subset \mathcal{F}$, it is not necessarily true that $\mathcal{G}$ is countably generated. To see a counterexample, let $\mathcal{F}$ be the borelianos in $[0, 1]$ and $\mathcal{G}$ the class of all countable or cocountable subsets of $\mathbb{Q}$ (a set is said to be cocountable if its complement is countable). It is not difficult to verify that $\mathcal{G}$ is a sub-$\sigma$-field of $\mathcal{F}$, which is not countably generated. Therefore, the rest of the proof of Proposition 4.4 requires the use of equivalent equivalent fields (see definition A.2). From Lemma A.1, we know that two $\sigma$-fields $\mathcal{G}, \mathcal{G}' \subset \Sigma$ are equivalent if for every $B \in \mathcal{G}$, there is a $B' \in \mathcal{G}'$ such that $\Pr(B \Delta B') = 0$ and for every $B' \in \mathcal{G}'$, there is a $B \in \mathcal{G}$, such that $\Pr(B \Delta B') = 0$. Consider the following two facts:

Lemma A.9  Every sub-$\sigma$-field $\mathcal{G}$ of $\Sigma$ is equivalent to a countably generated sub-$\sigma$-field $\mathcal{G}'$.

Proof. Given $E, F \in \Sigma$, let $d(E, F) \equiv \Pr(E \Delta F)$. This defines a pseudo-metric in $\Sigma$. Since $\Sigma$ is the Borel $\sigma$-field in $\Omega$, there is a countable set of sets $\{E_n\}$ such that $\Sigma = \sigma(\{E_n\})$. Since the ring generated by $\{E_n\}$ is also countable, then we may assume that $\{E_n\}$ is a ring. By Halmos (1974, Theorem 13.D), for every $m \in \mathbb{N}$ and set $A \in \Sigma$ (in particular, for every $A \in \mathcal{G}$), there is a integer $n$ such that $\Pr(A \Delta E_n) < 1/m$. Let $B_{m,n}$ denote this set. Then, the collection of sets $\{B_{m,n}\} \subset \{E_n\}$ is countable and it is dense in $\mathcal{G}$. Therefore, $\mathcal{G}' \equiv \sigma(\{B_{m,n}\})$ is countably generated and is equivalent to $\mathcal{G}$.}

Lemma A.10  If $\mathcal{F}$ and $\mathcal{G}$ are equivalent $\sigma$-fields then $\Pr(\cdot | \mathcal{F}) = \Pr(\cdot | \mathcal{G})$ (a.e.).
Proof. It can be shown that a function \( f : \Omega \to \mathbb{R} \) is \( \mathcal{F} \)-measurable if and only if there is a \( \tilde{f} : \Omega \to \mathbb{R} \) which is equal to \( f \) a.e. and is \( \bar{\mathcal{F}} \)-measurable. Thus, if \( \omega \mapsto Q(\omega, A) \) is a probability kernel which is \( \mathcal{F} \)-measurable, it is \( \bar{\mathcal{F}} \)-measurable (up to a null set). Since the second condition in the definition of conditional expectation is also satisfied for \( \bar{\mathcal{F}} \), then \( \Pr(\cdot|\mathcal{F}) = \Pr(\cdot|\bar{\mathcal{F}}) \). This implies the result.

Conclusion of the proof of proposition 4.4: Lemmas A.7, A.9 and A.10 imply that there exist a random variable \( Z \) such that \( \perp \perp (X^1, ..., X^n)|Z \), which concludes the proof.

A.4 Theorem 4.2 for the finite support case

In this section, we will assume that each \( X^i \) have a finite support \( S^i = \{x_1^i, ..., x_k^i\} \).\(^{25}\) remain true We will assume the following trivial condition:\(^{26}\)

Assumption A.11 For all \( x \in S^1 \times \cdots \times S^n \), there is \( \omega \in \Omega \) such that \( X(\omega) = x \).

Lemma A.12 Assume that \( \mathcal{F} \) is a \( \sigma \)-field formed by a finite partition \( \Pi = \{C_k\}_{k \in K} \) of \( \Omega \). For any \( A \in \Sigma \),

\[
\Pr(A|\mathcal{F})_{\omega} = \sum_{k \in K} \Pr(A|C_k)_{\omega} 1_{C_k}(\omega). \quad (a.e.)
\]

Proof. It is easy to see that the expression on the right above satisfies the two conditions for being a conditional probability. On the other hand, let \( J \subset K \) be the set of indices \( j \) such that \( \Pr(C_j) > 0 \). Then \( \Pr(\cup_{j \in J} C_j) = \sum_{j \in J} \Pr(C_j) = 0 \) and the sets \( C_j \) for \( j \notin J \) can be ignored. Since \( \Pr(A|\mathcal{F})_{\omega} \) is \( \mathcal{F} \)-measurable, it must be constant in each \( C_j \) for \( j \in J \) and, therefore, it must be \( \Pr(A|\mathcal{F})_{\omega} = \Pr(A|C_k) \) for almost all \( \omega \in C_k \). This concludes the proof. \( \blacksquare \)

The following definition will be useful in the sequel.

\(^{25}\)Most results and proofs are exactly the same for the case of countable support, although with some potential complication in the notation. Naturally, the algorithm described at the end of this subsection is restricted to the finite support case.

\(^{26}\)It is easy to construct examples of spaces that do not satisfy this. For example, let \( \Omega = \{a, b\} \), \( n = 2 \), \( X^1(a) = X^2(b) = 0 \) and \( X^1(b) = X^2(a) = 1 \). Then, \((0, 0) \in S^1 \times S^2 \) but there is no \( \omega \in \Omega \) such that \( X(\omega) = (0, 0) \). However, in this case, we can easily add new zero-probability states to the space and change it to a space that satisfy it.
**Definition A.13** A set \( C \in \Sigma \) is admissible if \( \forall x = (x^1, ..., x^n) \in S^1 \times \cdots \times S^n \),

\[
\Pr(\{X^i = x^i\} \cap C) > 0, \forall i \Rightarrow \Pr(\{X = x\} \cap C) > 0.
\]

We say that a partition \( \Pi \) is admissible if for all \( C \in \Pi \), \( C \) is admissible.

To understand the concept of admissible set, consider the following example.

**Example A.14** \( \Omega = \{a, b, c, d\} \); \( \Pi = \{\{a\}, \{b, c, d\}\} \) and \( \Pr(\omega) > 0, \forall \omega \in \Omega \). Let the values of \( X(\omega) \) be given as in Figure 4. While the set \( \{a\} \in \Pi \) is admissible, the set \( C = \{b, c, d\} \) is not, because \( \Pr(X_1 = 0, X_2 = 1 | C) = 0 \) while \( \Pr(X_1 = 0 | C) \Pr(X_2 = 1 | C) > 0 \). Note also that \( \Pi' = \{\{a, b\}, \{c, d\}\} \) is an admissible partition.

![Figure 4: Admissible and non-admissible partitions.](image)

We have the following:

**Proposition A.15** \( X^1, ..., X^n \) are conditionally independent given \( F \) if and only if \( \Pi \) is admissible and

\[
(\Pr(C))^{n-1} = \frac{\Pr(\{X^1 = x^1\} \cap C) \cdots \Pr(\{X^n = x^n\} \cap C)}{\Pr(X = x \cap C)}\), \quad (3)
\]

for all \( C \in \Pi \) and \( x = (x^1, ..., x^n) \in S^1 \times \cdots \times S^n \) s.t. \( \Pr(\{X = x\} \cap C) > 0 \).

**Proof.** Necessity. Observe that Lemma A.12 gives

\[
\Pr(A|\Sigma) = \sum_{C \in \Pi} \Pr(A|C) \omega C(\omega) \quad (a.e.),
\]

which implies that the conditional independence must hold for all \( C \in \Pi \) for which \( \Pr(A \cap C) > 0 \). If \( \Pi \) is not admissible, then there exists \( x = (x^1, ..., x^n) \in \)
We can assume that \( x \in \omega \) there exists \( X \) that this implies that \( \Pr(\tilde{\prod}_{i=1}^{n} \Pr(\{X^i = x^i\} | C) > 0 \). Then \( \mathcal{A}^i, \cdots, \mathcal{A}^n \) is not conditionally independent given \( \Sigma \), a contradiction.

Since \( \{x^i\} \in \mathcal{A}^i \), for \( i = 1, \cdots, n \), necessity of the second condition comes directly from the conditional independence requirement:

\[
\Pr(X^1 = x^1, \cdots, X^n = x^n | C) = \frac{\Pr(\{X^1 = x^1\} | C) \cdots \Pr(\{X^n = x^n\} | C)}{\Pr(C)} = \frac{\Pr(\{X^1 = x^1\} \cap \cdots \cap \{X^n = x^n\})}{\Pr(C)^n}.
\]

**Sufficiency.** From Lemma A.12, it is sufficient to check the conditional independence condition for each \( C \in \Pi \) satisfying \( \Pr(C) > 0 \). So, fix an element \( C \) of an admissible partition \( \Pi \). For \( i = 1, \cdots, n \), fix \( A^i \in \mathcal{A}^i \) and let \( A = A^1 \cap \cdots \cap A^n \). Let \( \tilde{S}^i \subset S^i \) be the set of the values of \( X^i \) implied by \( A^i \), that is, \( \tilde{S}^i \equiv X^i(A^i) \). If for some \( i \) and \( x^i \in \tilde{S}^i \), \( \Pr(\{X^i = x^i\} \cap C) = 0 \), we can redefine \( \tilde{A}^i = A^i \setminus \{X^i = x^i\} \) and \( \tilde{A} = A \setminus \{X^i = x^i\} \) so that \( \Pr(\tilde{A} | C) = \Pr(A | C) \) and \( \Pr(\tilde{A}^i | C) = \Pr(A^i | C) \). In other words, we can assume without loss of generality that \( \Pr(\{X^i = x^i\} \cap C) > 0 \) for all \( i \) and \( x^i \in \tilde{S}^i \). Since \( \Pi \) is admissible, this implies that \( \Pr(\{X = x\} \cap C) > 0 \) for all \( x \in \tilde{S} \equiv \tilde{S}^1 \times \cdots \times \tilde{S}^n \). It is clear that \( X(A) \subset \tilde{S} \). Assume that there exists \( x \in \tilde{S} \setminus X(A) \). By assumption A.11, there exists \( \omega \in \Omega \) such that \( X(\omega) = x = (x^1, \cdots, x^n) \). Without loss of generality, we can assume that \( \omega \in A^i \) for all \( i \). Thus, \( \omega \in A^1 \cap \cdots \cap A^n = A \). This implies that \( x = X(\omega) \in X(A) \), which is a contradiction. This shows that \( X(A) = \tilde{S} \), that is, \( A = \cup_{x \in \tilde{S}} X^{-1}(x) \). It is also clear that \( A^i = \cup_{x^i \in \tilde{S}^i} (X^i)^{-1}(x^i) \).

Since \( \Pr(\{X = x\} \cap C) > 0 \) for all \( x \in \tilde{S} \), the assumption implies that \( \Pr(X = x | C) = \prod_{i=1}^{n} \Pr(X^i = x^i | C) \) for all \( x \in \tilde{S} \). Therefore,

\[
\Pr(A | C) = \sum_{x \in \tilde{S}} \Pr(X = x | C) = \sum_{x \in \tilde{S}} \prod_{i=1}^{n} \Pr(X^i = x^i | C)
\]

\[
= \sum_{x^i \in \tilde{S}^i} \cdots \sum_{x^n \in \tilde{S}^n} \prod_{i=1}^{n} \Pr(X^i = x^i | C)
\]

\[
= \Pr(A^1 | C) \cdots \Pr(A^n | C),
\]

\[\text{Since } X^i(\omega) = x^i \in \tilde{S}^i, \text{ if } \Pr(\{\omega\}) > 0, \text{ then } \omega \in A^i \in \mathcal{A}^i; \text{ otherwise, we can put } A^i \equiv A^i \cup \{\omega\} \in \mathcal{A}^i. \text{ Note that this change does not affect the previous assumption that } \Pr(\{X^i = x^i\} \cap C) > 0.\]
as we wanted to show.

The above result gives an algorithm to find (all) partitions $\Pi$ that make the variables $X^1, \ldots, X^n$ conditionally independent. The variable $Z$ will be just the indication of the element of the partition $\Pi$ that contains the true realization of types. The algorithm can be roughly described as follows:

**Input:** the finite probabilistic space $(\Omega, \Sigma, \Pr)$ and the partitions $\Pi^1, \ldots, \Pi^n$ generating $\mathcal{X}^1, \ldots, \mathcal{X}^n$.

1. Find the common knowledge partition: $\Pi^0 \equiv \bigwedge_{i=1}^n \Pi^i$, that is, the finest common coarsening partition.

2. Test whether $\Pi^0$ is admissible. If $\Pi^0$ is not admissible, find a coarser refinement of $\Pi^0$ that is admissible and call it $\Pi^0$.

3. Put $\Pi := \emptyset$. Let $\Pi^0 = \{C_1, \ldots, C_K\}$. For $k = 1, \ldots, K$, do:
   
   (a) Test whether $C_k$ satisfies (3) for all $x = (x^1, \ldots, x^n) \in S^1 \times \cdots \times S^n$ such that $\Pr(\{X = x\} \cap C) > 0$. If it satisfies, do $\Pi := \Pi \cup \{C_k\}$. If $C_k$ does not satisfy (3) for some $x$, do the following:
   
   (b) Calculate $P(\cdot) = \Pr(\cdot | C_k)$, that is the conditional probability given $C_k$;
   
   (c) Obtain $\Sigma_{C_k} = \Sigma \cap C_k$;
   
   (d) Derive and $\Pi^{i}_{C_k} \equiv \Pi^i \cap C_k$;
   
   (e) Call this program for the input $(C_k, \Sigma_{C_k}, P)$ and partitions $\Pi^{1}_{C_k}, \ldots, \Pi^{n}_{C_k}$. The outcome will be a partition $\Pi_k = \{C^1_k, \ldots, C^m_k\}$ of $C_k$. Put $\Pi := \Pi \cup \{C^1_k, \ldots, C^m_k\}$.

**Output:** the partition $\Pi$.

This algorithm will produce all partitions that make $X^1, \ldots, X^n$ conditionally independent if step 2 considers all partitions that are coarser refinements of $\Pi^0$ and admissible. It will stop at some point because the fully informative partition makes $X^1, \ldots, X^n$ conditionally independent (see Proposition 4.4).

---

28 The algorithm will run in exponential time, but we conjecture that this cannot be significantly improved, unless $P = NP$. 

22
Gossner, Kalai, and Weber (2009) also have the existence of minimal conditional structure for finite or countable setting treated in this subsection.\textsuperscript{29} It should be noted however that the important contribution of this subsection is not existence, since this is known at least since McKean (1963, p. 343, property e) (see also Mouchart and Rolin (1984, Theorem 4.3)). The main contribution of this subsection is the algorithm that it provides. The development of such an algorithm was an open question: see van Putten and van Schuppen (1985) and van Schuppen (1982). Gossner, Kalai, and Weber (2009)’s existence proof is based in the Zorn’s Lemma, which is obviously non-constructive. They also provide a characterization of the case where the common knowledge partition makes the variables conditionally independent. However, their definition puts together cases 1 and 2 of of Theorem 4.2. We offer a different characterization of all cases below.

\textbf{Lemma A.16} Let $S^i$ denote the finite or countable the support of the $X^i$. Consider the following two conditions: (i) the common knowledge partition is the (unique) outcome of the above algorithm; (ii) there exist $x^i \in \prod_{i=1}^{n} S^i$ such that $\Pr(X = x) = 0$\footnote{I am grateful to Olivier Gossner for pointing out this paper to me. This happened after I had obtained my results.}.\footnote{This lemma is also true without the restriction to finite or countable support. In this case, condition (i) should be that the common knowledge partition makes the variables conditionally independent; and condition (ii) should be that there exist $B^i \subset S^i$ such that $\Pr(B^i) > 0, \forall i$, but $\Pr(B^1 \times \cdots \times B^n) = 0$.}

If only condition (i) holds, but not (ii), we are in case 1 of Theorem 4.2. If both conditions hold, we are in case 2. If condition (i) does not hold, we are in case 3.

\textbf{Proof.} It is clear that if condition (i) does not hold, we are in case 3. Under condition (i), the common knowledge partition makes the types conditionally independent. Now, condition (ii) holds, that is, if there exists $x = (x^1, ..., x^n) \in \prod_{i=1}^{n} S^i$ such that $\Pr(X = x) = 0$, this means that there is $i$ and $j$ such that $x^i$ and $x^j$ cannot be in the same element of common knowledge partition. Therefore, this partition is not trivial, which means that the variables are not independent, that is, they are not in case 1. On the other hand, if condition (ii) does not hold, we have independence (case 1). $\blacksquare$

This section establishes the existence of minimally informative conditional splitters in the finite support case, provides an algorithm to find all of them and establishes the classification of cases 1, 2 and 3 given in Theorem 4.2. The other statements of Theorem 4.2 for the case of finite support are included in the discussion for the general case, which is done in the next subsection.
A.5 Theorem 4.2 for the general case.

The minimally informative conditional splitters are defined as follows:

**Definition A.17 (Minimally informative conditional splitters)** A variable $Z$ that makes $X^1, \ldots, X^n$ conditional independent is minimally informative, denoted $\perp \perp (X^1, \ldots, X^n)_{\min} Z$, if there is no $\sigma$-field $F$ such that: (i) $\perp \perp (X^1, \ldots, X^n)|F$, and (ii) $F$ is equivalent to $Z$.

A variable $Z$ that makes $X^1, \ldots, X^n$ conditional independent is the least informative if for every $\sigma$-field $F$ such that: $\perp \perp (X^1, \ldots, X^n)|F$, we have $F \subset \bar{Z}$.

Lemma A.16 gives the classification of cases in Theorem 4.2. The following lemma establishes statement 3(a) in Theorem 4.2:

**Lemma A.18** If $\perp \perp (X^1, \ldots, X^n)|F$, the common knowledge $\sigma$-field $K$ is included in $F$.

**Proof.** It is easy to see that the common knowledge $\sigma$-field is given by $K = \bigcap_{i=1}^n \mathcal{X}^i$. Let $H \in K$ and $\perp \perp (X^1, \ldots, X^n)|F$. We want to prove that $H \in F$. For each set $S \subset N \equiv \{1, \ldots, n\}$, let $\mathcal{X}^S \equiv \bigvee_{i \in S} \mathcal{X}^i$, that is, $\mathcal{X}^S$ denotes the $\sigma$-field generated by $\bigcup_{i \in S} \mathcal{X}^i$. By Lemmas A.5 and A.6 $\perp \perp (X^1, \ldots, X^n)|F$ is equivalent to $P[H|F] = P[H|\mathcal{X}^S, F]$, for all $H \in \mathcal{X}^S$ and $S \subset \{1, \ldots, n\}$. Since $H \in K = \bigcap_{i=1}^n \mathcal{X}^i \subset \mathcal{X}^S$, $P[H|F] = P[H|\mathcal{X}^S, F]$. Also, $K \subset \sigma(\mathcal{X}^S, F)$ implies that $H$ is $\sigma(\mathcal{X}^S, F)$-measurable and $P[H|\mathcal{X}^S, F] = 1_H$. But $P[H|F] = 1_H$ implies that $H \in F$, as we wanted to show.

For a proof of statement 3(b), see Mouchart and Rolin (1984, Theorem 4.3) (see also McKean (1963, p. 343, property e)). The algorithm mentioned in item 3(c) for the finite support case is presented in section A.4 above. The following example establishes statement 3(d), that is, it shows that it may not exist a least informative conditionally splitter.

**Example A.19** Let $\Omega = \{a, b, c\}$ and $\Pr(\{\omega\}) = \frac{1}{3}$, for all $\omega \in \Omega$. Consider the following $\sigma$-fields:

\[
\mathcal{F}_1 = \sigma(\{a, b\}, \{c\}), \\
\mathcal{F}_2 = \sigma(\{a\}, \{b, c\}).
\]

Note that $\mathcal{F}_1$ and $\mathcal{F}_2$ are not independent because $\Pr(\{b\}) \neq \Pr(\{a, b\}) \Pr(\{b, c\})$. However, it is not difficult to verify that $\mathcal{F}_1 \perp \perp \mathcal{F}_2|\mathcal{F}_1$ and $\mathcal{F}_1 \perp \perp \mathcal{F}_2|\mathcal{F}_2$ and these two different $\sigma$-fields are minimally informative. Thus, it is not possible that there is a least informative conditional splitter.
The proof of item 3(e) is given in the next subsection.

A.6 Proof of item 3(e) of Theorem 4.2

Although we will prove item 3(e) for the finite case, we will introduce definitions for the general case below, since these are more common.

**Definition A.20** A Markov transition from \((S, \mathcal{S})\) to \((X, \mathcal{X})\) is a Borel measurable function \(T : S \rightarrow \Delta(X)\), where \(\Delta(X)\) is the set of all measures in the measurable space \((X, \mathcal{X})\), and the topology in \(\Delta(X)\) (for giving its Borel sets) is its weak* topology, i.e., the \(\sigma(\Delta(X), C_b)\)-topology.

**Definition A.21** A Markov kernel from \((S, \mathcal{S})\) to \((X, \mathcal{X})\) is a function \(k : S \times X \rightarrow [0, 1]\) satisfying the following:

1. For each \(s \in S\), the set function \(k(s, \cdot) : X \rightarrow [0, 1]\) is a probability measure.
2. For each \(A \in \mathcal{X}\), the mapping \(k(\cdot, A) : S \rightarrow [0, 1]\) is \(\mathcal{S}\)-measurable.

The reader will note that a conditional probability as defined in definition A.3 is just a Markov kernel as defined by definition A.21. The following two results are informative. They are respectively theorems 19.12 and 19.13 of Aliprantis and Border (2006, p.630).

**Lemma A.22** Let \(S\) and \(X\) be separable metrizable spaces. Then for a mapping \(T : S \rightarrow \Delta(X)\) the following statements are equivalent.

1. \(T\) is a Markov transition, that is, \(T\) is Borel measurable.
2. The function \(k : S \times B_X \rightarrow [0, 1]\), defined by \(k(s, A) = T_s(A)\), is a Markov kernel.

**Lemma A.23** Let \(S\) and \(X\) be separable metrizable spaces. Then for a mapping \(k : S \times B_X \rightarrow [0, 1]\) the following statements are equivalent.

1. The function \(k\) is a Markov kernel.
2. The function \(T : S \rightarrow \Delta(X)\), defined by \(s \mapsto T_s(\cdot) = k(s, \cdot)\), is a Markov transition.
For now, let $Y$ denote the set $\Delta(\Omega)$, and let $\mathcal{Y}$ denote the weak* topology mentioned in definition A.20. Let $M \subset Y$ denote the set of all independent (product) measures, that is,

$$M = \{ \mu \in \Delta(\Omega, \Sigma) : \mu = \mu^1 \times \cdots \times \mu^n, \mu^i \in \Delta(\Omega, X^i) \}$$

It is easily seen that $M$ is $\mathcal{Y}$-measurable. The following result clarifies the relationship between Markov transitions and conditional probabilities.

**Lemma A.24** A Markov transition $T : \Omega \to \Delta(\Omega) = Y$ from $(\Omega, \mathcal{F})$ to $(\Omega, \Sigma)$ represents the conditional probability given $\mathcal{F}$ if and only if for every $B \in \mathcal{F}$ and $A \in \Sigma$,

$$\int_B T_\omega(A) \, d\Pr(\omega) = \Pr(A \cap B)$$

where $T_\omega(A)$ represents the probability of the set $A \in \Sigma$ under the measure $T(\omega) \in \Delta(\Omega)$.

In this case, $\mathcal{F}$ conditionally splits $(X^1, \ldots, X^n)$ if and only if $T(\Omega) \subset M$.

**Proof.** The first part is immediate from definitions A.3, A.20 and A.21 and Lemmas A.22, A.23. The second part comes from the definition of $M$ above and definition A.4. ■

**Corollary A.25** The set $C$ of Markov transitions $T : \Omega \to \Delta(\Omega)$ that conditionally splits $(X^1, \ldots, X^n)$ is a closed convex subset of all Markov transitions.\(^{31}\)

**Proof.** This comes directly from the two characterizing conditions given in Lemma A.24 above. ■

The following result establishes the fact 3(e) of Theorem 4.2 for $\Omega$ finite. It is useful to introduce the notation: $L^p(\mathcal{F}, X)$ for the space $L^p((\Omega, \mathcal{F}, \Pr), X)$, $1 \leq p \leq \infty$, where $X$ is a Banach space (see Diestel and Uhl (1977)). If $X = \mathbb{R}$, then we will write just $L^p(\mathcal{F})$ instead of $L^p(\mathcal{F}, \mathbb{R})$. Also, in the references below, DU stands for Diestel and Uhl (1977), while DS abbreviates Dunford and Schwartz (1958).

**Proposition A.26** Let $T^0$ denote the Markov transition representing the conditional probability given the common knowledge $\sigma$-field $\mathcal{K}$. If $\Omega$ is finite, then there is a unique $T \in C$ which realizes the minimal distance from $T^0$ to $C$.

\(^{31}\)We will use this result for $\Omega$ finite, so that the topology of $Y = \Delta(\Omega)$ is not important.
Proof. Let \( n \) be the number of points in \( \Omega \). Then a Markov transition is a function \( T : \Omega \rightarrow \Delta(\Omega) \subset \mathbb{R}^n \). Then, \( C \) can be seen as a subset of \( L^2(\Sigma, \mathbb{R}^n) \), and \( T^0 \not\in C \) (because we are considering case 3 of Theorem 4.2), there exists a unique point \( \bar{T} \in C \) that realizes \( \inf_{T \in C} ||T - T^0|| \), where \( || \cdot || \) denotes the \( L^2(\Sigma, \mathbb{R}^n) \)-norm (see DS, IV.4.2, p. 248). 

Remark A.27 Note that Proposition A.26 only holds because \( C \) is a subset of a Hilbert space and we used its norm. If we have just a Banach space instead of a Hilbert space, it is not always true that there is a unique point minimizing the distance from \( C \) to \( P^0 \). For example, consider the set \( D = \{(x_1, x_2) \in [0, 1]^2 : x_1 = x_2\} \) in \( \mathbb{R}^2 \) with the sum norm: \( ||(x_1, x_2)|| = |x_1| + |x_2| \). The distance of the point \( x^0 = (1, 0) \) to \( D \) is 1 = \( ||(x_1, x_2) - (1, 0)|| = |x_1 - 1| + |x_2| \) for all \( (x_1, x_2) \in D \). However, every Hilbert space is reflexive (DS, IV.4.6, p. 250) but \( L^2(\Sigma, X) \) is reflexive if and only if \( X \) is reflexive (see DU, Corollary IV.1.2, p. 100). Unfortunately, \( \Delta(\Omega) \) is reflexive if and only if it is finite dimensional (for instance, if \( \Omega \) is finite) (DS, IV.13.21, p. 341). These observations show that the ideas in Proposition A.26 do not extend directly to general \( \Omega \).

A.7 Proof of Theorem 4.7.

Let the distribution of \( X \) and \( Y \) be given by the table below, that is, \( \Pr(X = 1, Y = 0) = b \).

<table>
<thead>
<tr>
<th></th>
<th>( Y = 0 )</th>
<th>( Y = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X = 0 )</td>
<td>( a )</td>
<td>( b )</td>
</tr>
<tr>
<td>( X = 1 )</td>
<td>( b )</td>
<td>( d )</td>
</tr>
</tbody>
</table>

Our purpose is to find a binary variable \( Z \), with joint distribution with \( X \) and \( Y \) described by the tables below,

<table>
<thead>
<tr>
<th>( Z = 0 )</th>
<th>( Y = 0 )</th>
<th>( Y = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X = 0 )</td>
<td>( u a )</td>
<td>( v b )</td>
</tr>
<tr>
<td>( X = 1 )</td>
<td>( v b )</td>
<td>( w d )</td>
</tr>
</tbody>
</table>

\[ \text{\(32\)Since all } T \in C \text{ take value in } \Delta(\mathbb{R}^n) \text{, which is compact, then } C \text{ is actually a subset of } L^\infty(\Sigma, \mathbb{R}^n) \subset L^2(\Sigma, \mathbb{R}^n).}\]

\[ \text{\(33\)The norm of Hilbert space is given by an inner product. Naturally, this example fails with the euclidian norm, which comes from an inner product.}\]
such that the conditional independence conditions are satisfied:

\[
\begin{align*}
&adw = b^2v^2 \\
&ad(1-u)(1-w) = b^2(1-v)^2 \\
&u, v, w \in (0, 1)
\end{align*}
\]

Let us define \( r \equiv \frac{b^2}{ad} \). From the first equation above, we obtain: \( w = \frac{v^2}{u} \).

Thus, \( w < 1 \) is equivalent to \( rv^2 < u \). The second equation simplifies to \( (1-u-w+uw) = r(1-2v+v^2) \) and, using the first equation, to \( 1-u-w = r(1-2v) \).

Substituting \( w = \frac{v^2}{u} \) we obtain:

\[
1-u-r \frac{v^2}{u} = r(1-2v) \iff u^2 - u[1-r(1-2v)] + rv^2 = 0.
\]

Observe that the case \( r = 1 \) corresponds to independence of \( X \) and \( Y \), in which case \( u = v = w \) can be anything. So, we assume \( r \neq 1 \) in what follows.

The solution is:

\[
u = \frac{1-r(1-2v) \pm \sqrt{(1-r(1-2v))^2 - 4rv^2}}{2} \tag{4}
\]

The conditions \( u, v, w \in (0, 1) \) will be satisfied if \( u, v \in (0, 1) \) and \( rv^2 < u \). Let us define the polynomial \( P(u) \equiv u^2 - u[1-r(1-2v)] + rv^2 \) and observe that:

\[
\begin{align*}
P(0) &= 0^2 - 0[1-r(1-y-z)] + rv^2 = rv^2 > 0; \\
P(rv^2) &= (rv^2)^2 - (rv^2)[1-r(1-2v)] + rv^2 \\
&= rv^2(rv^2 - 1 + r - 2rv + 1) = r^2v^2(1-v)^2 > 0; \\
P(1) &= 1^2 - 1[1-r(1-2v)] + rv^2 = r(1-2v+2v) = r(1-v)^2 > 0.
\end{align*}
\]

Since \( u \) must be a root of the polynomial \( P(u) \), the conditions \( u, v \in (0, 1) \) and \( rv^2 < u \) are equivalent to \( v \in (0, 1) \) and:

\[
rv^2 < \frac{1-r(1-2v)}{2} < 1 \iff 1 - \sqrt{\frac{2-r}{r}} < v < \min\left\{ \frac{1+r}{2r}, \frac{1+r}{2} \right\};
\]

\[
[1-r(1-2v)]^2 \geq 4rv^2 \iff \begin{cases} v \geq \frac{r-1}{2(r-\sqrt{r})} & \text{if } r > 1 \smallskip \\
v \leq \frac{1-r}{2(\sqrt{r}-1)} & \text{if } r < 1
\end{cases}
\]

28
Note that the inequality \( rv^2 < u \iff rv^2 < \frac{1-r(1-2v)}{2} \) cannot be satisfied for \( r > 2 \). For \( r \in (1, 2] \),

\[
\frac{r - 1}{2(r - \sqrt{r})} > \frac{1 + \sqrt{2-r}}{2} \\
\iff (\sqrt{r} - 1)(\sqrt{r} + 1) > \frac{1}{\sqrt{r}}(\sqrt{r} + \sqrt{2-r})(r - \sqrt{r}) \\
\iff \sqrt{r} + 1 > \sqrt{r} + \sqrt{2-r},
\]

which is obviously true for \( r \in (1, 2] \). Therefore, the conditions cannot be met if \( r > 1 \). On the other hand, if \( r < 1 \), then \( \sqrt{\frac{2-r}{r}} > 1 \) and \( \frac{1+r}{2r} > 1 \), in which case the inequalities \( \frac{1-\sqrt{2-r}}{2} < v < \min\{\frac{1+\sqrt{2-r}}{2}, \frac{1+r}{2r}\} \) are trivially satisfied for any \( v \in (0, 1) \). We simplify all conditions to: \( v \in (0, 1) \) and \( v \leq \frac{1-r}{2(\sqrt{r} - r)} \), but this last condition is also trivially satisfied for any \( v \in (0, 1) \), since:

\[
\frac{1-r}{2(\sqrt{r} - r)} \geq 1 \\
\iff 1 - r \geq 2\sqrt{r} - 2r \\
\iff (\sqrt{r} - 1)^2 \geq 0.
\]

Therefore, if \( r < 1 \), we can choose any \( v \in (0, 1) \), obtain \( u \) from (4) and put \( w = rv^2 \). This will give the decomposition that we wanted. Note that the condition \( r < 1 \) is equivalent to \( ad > b^2 \), which is just positive correlation.

References


---

34 This gives another proof for the statement in Example 3.1.


