Extended (Pre-)Version of a paper accepted at FLAIR 32, 2019

Influence-based Independence

Özgür L. Özçep and **Felix Kuhr** and **Ralf Möller**
Institute of Information Systems (IFIS)
University of Lübeck
Lübeck, Germany
{oezcep,kuhr,moeller}@ifis.uni-luebeck.de

Abstract

Conditional independence structures describe independencies of one set of variables from another set of variables conditioned upon a third set of variables. These structures are invaluable means for compact representations of knowledge because independencies can be exploited for useful factorizations. Conditional independence structures appear in different disguise in various areas of knowledge representation, be it the conditional independence of sets of random variables in probabilistic graphical models such as Bayesian networks or as conditional functions related to belief revision, or as independencies induced by (embedded) multivalued dependencies in data bases. This paper investigates conditional independencies for Boolean functions using Fourier analysis. We define three notions of independence based on the notion of influence of a variable on a function and draw connections to multivalued dependencies.

1 Introduction

Conditional independence (CI) structures (Studený 1993; Studeny 2005; Wang 2010; Dawid 2001) describe independencies of one set of variables from another set of variables conditioned upon a third set of variables. They are invaluable means for compact representations of knowledge as independencies can be exploited for useful factorizations.

Conditional independence structures appear in different disguise in the area of knowledge representation (Studený 1993): In statistics and probabilistic graphical modelling CIs occur as independencies of conditional probabilities for sets of random variables. Here they are used to represent full joint distributions compactly with up to exponential savings by Bayesian Networks or more specifically by causal networks (Pearl 2009); CIs were investigated from an epistemological perspective with conditional functions (Spohn 1994); CIs appear implicitly as independencies induced by multivalued dependencies in databases (Fagin 1977), where the factorizations resulting from multivalued dependencies are used to define and design normalized DBs (= databases); and last but not least, under the term *(ir)relevance*, CIs also play a role for rational belief revision (Parikh 1999; Özçep 2016), where one is interested in eliminating at most those sentences from a knowledge base that are relevant for the possibly inconsistent trigger information.

In its most abstract form, CIs can be investigated axiomatically by stating closure rules on triples $(A \perp\!\!\!\perp B | C)$ as initially done in the graphoid framework (Pearl and Paz 1985). Such an axiomatization is useful for the comparison of different classes of CIs by comparing the axioms modelled by each class. As for the models of such triples, various mathematical structures have been exploited in order to capture the essence of CIs, be it separoids (Dawid 2001), imsets (Studeny 2005), cain algebras (Wang 2010), or—as an efficient data structure for processing Boolean functions—sum-product networks (Poon and Domingos 2012).

The focus of this paper is on CIs for Boolean functions. Treating Boolean functions as the limits of full joint distributions over Boolean random variables, we develop a first notion of conditional independence based on probabilistic (conditional) independence: Random variables (RVs) become propositional variables, the full joint distribution on RVs $X_1, \ldots, X_n$ becomes an n-ary Boolean function $f : \{0,1\}^n \longrightarrow \{0,1\}$, and the marginal of a probability distribution becomes the forgetting operator of a Boolean function.

As a first result of this paper, we show that the induced notion of conditional independence corresponds exactly to embedded multivalued dependencies from database theory when treating the assignments making the Boolean function true as a table. A useful consequence of this theorem is that insights, techniques, and results from one area (probabilistic graphical models) can be transferred easily to another area (databases), and vice versa.

Many interesting results regarding Boolean functions were achieved with the framework of Fourier analysis (O'Donnell 2014) and were applied to various areas of computer science such as property testing, circuit analysis, social choice, machine learning, cryptography etc. Fourier analysis rests on a unique canonical representation of functions as a multilinear polynomial where the monomials are parity functions. A fundamental measure used in Fourier analysis is the notion of the influence of a variable on the outcome of a function. We use (adapted versions of) influence to describe intuitive measures for the dependence between two (sets of) variables. These measures induce, as special cases, three notions of independence.

As a result of our Fourier analysis we present decomposition theorems that shed light onto the different dimensions of the influence measures and their inter-relations. With the last result of the Fourier analysis for influences we give the closing bracket for the narrative that was opened with probabilistic conditional independencies and embedded multivalued dependencies: We show that multivalued dependencies (of some sufficiently general form) can be decomposed into summands that are described by our influence-based dependency measures.

An extended version of this paper with proofs is provided under the URL `https://tinyurl.com/yah9mkyc`.

2 Preliminaries

Let $[n] = \{1, 2, \ldots, n-1, n\}$ be the set of all natural numbers (different from zero) up to $n \in \mathbb{N} \setminus \{0\}$. Similarly, we define $[X_n] = \{X_1, \ldots, X_n\}$ to be the set of propositional variables X_i up to index n from a fixed family of propositional variables $(X_i)_{i \in \mathbb{N} \setminus \{0\}}$. We are going to deal mainly with Boolean functions with finite arity operating on tuples over the 2-element domain of Boolean values $\mathbb{B} = \{0, 1\}$, where, intuitively 0 stands for the truth value FALSE and 1 for the truth value TRUE. Due to this intuition we call the tuples x for which $f(x) = 1$ verifiers or models of f, and the tuples y such that $f(y) = 0$ falsifiers or anti-models of f. Tuples of Boolean values will be denoted by x, y, z and indexed variants. x_i for $i \in [n]$ and n-tuple x is the element in the i-th position of x. The arity/length of the tuples will be clear from the context. The n-ary Boolean functions are of the form $f : \mathbb{B}^n \longrightarrow \mathbb{B}$, for some $n \in \mathbb{N}$. When referring to the positions of an n-ary Boolean function we usually use the variables $X_1, \ldots, X_n$. We are going to use the following unary functions $f : \mathbb{B} \longrightarrow \mathbb{B}$: negation or "not" function defined by $\neg(x) = 1 - x$; the unary 1-function or "always true" function defined by $1(x) = 1$, and the 0-function or "always false" function defined by $0(x) = 0$. The binary Boolean functions $f : \mathbb{B} \times \mathbb{B} \longrightarrow \mathbb{B}$ that we consider are: conjunction alias "and" function defined by $\wedge(x, y) = x \wedge y = x \cdot y$; disjunction alias "or" function defined by $\vee(x, y) = x \vee y = \neg(\wedge(\neg x, \neg y)) = x + y - x \cdot y$; implication alias "if-then" function defined by $\rightarrow (x, y) = x \rightarrow y = \neg(x) \vee y$; inverse implication defined by $\leftarrow (x, y) = \rightarrow (y, x)$; biimplication alias "if-and only-if" function defined by $\leftrightarrow (x, y) = \wedge(\rightarrow (x, y), \rightarrow (y, x)) =$ and the xor function alias "exclusive or" function defined by $x \oplus y = \leftrightarrow (x, \neg(y))$. As "and","or" and "xor" functions are associative we define these for arbitrary arities so that it makes sense to use them with running indices. For example, for a sequence of Boolean values $(x^1, \ldots, x^n)$ the expression $\bigvee_{1 \leq i \leq n} x^i$ can be read as $(\ldots((x^1 \vee x^2) \vee x^3) \vee \ldots x^n$.

Let x be an n-tuple of Boolean values, $i \in [n]$ and $b \in \mathbb{B}$. Then $x[x_i/b]$ is an *i-variant* and is defined to be the same as x except for the position i which is set to b. This can be extended to variants on a whole set of positions in an intuitive way. Formally: For a set of positions $A = \{i_1, \ldots, i_k\} \subseteq [n]$ with cardinality k and y a k-tuple $y = (y_1, \ldots, y_k)$ of Boolean values and an n-ary tuple x, $x[A/y]$ denotes an A variant of x: $x[A/y] = x[x_{i_1}/y_1, \ldots, x_{i_k}/y_k]$.

For $J \subseteq [n]$ we use $f_{J\restriction z}$ or even shorter $f_{\restriction z}$ as a shorthand for the function resulting from f by fixing the non-J positions to the values of the vector z.

Let $A = \{i_1, \ldots, i_k\} \subseteq [n]$ be a set of positions with cardinality k, and B the set of all other positions, $B = [n] \setminus A$; further let f be an n-ary Boolean function. Then the *A-marginal* f_A^m ("m" for marginal) is the n-ary function defined by $f_A^m(x_1, \ldots, x_n) = \bigvee_{y \in \mathbb{B}^{n-k}} f(x[B/y])$, i.e., f_A^m is built by considering all position variants of the input vector x on positions different from those in A and then cumulating the result w.r.t. disjunction. The naming convention chosen here is drawn from the notion of a max-marginal known in probability theory. As the A-marginal depends actually only on the values at the A-positions one can consider it also as a k-ary function: In this case we will talk about the A-projection of f and denote it by f_A^p. If the arity does not matter or is clear from the context, then we do not distinguish them, i.e., drop the superscript.

Even more loosely (but conveniently), we will use the notation $f(X_1, \ldots, X_n)$, with propositional variables X_i, in order to denote an n-ary Boolean function and, e.g., $f(X_1, X_3)$ to denote its $\{1, 3\}$-marginal. In this case we also talk about the $\{X_1, X_3\}$-marginal.

An immediately verifiable lemma (due to the monotonicity of disjunction) says that the less one projects out the stronger the function (in a logical sense):

Lemma 1. *For all $A, B \subseteq [n]$: If $A \subseteq B$, then for all $x \in \mathbb{B}^n$: $f_B^m(x) \leq f_A^m(x)$.*

3 Conditional Independence and Marginals

This section is devoted to developing a notion of independence on Boolean functions motivated from the notion of conditional independence as used in probability theory. The main step towards this is the definition of a notion corresponding to (not exactly being) conditional probabilities.

Definition 1. *Let $A, B \subseteq [X_n]$. The A, B-*conditional *of an n-ary function f is defined as the n-ary Boolean function $f_{A|B}$ (also written as $f(A|B)$) defined for all $x \in \mathbb{B}^n$ by $f_{A|B}(x) = f_B(x) \rightarrow f_{A \cup B}(x)$.*

Note the analogy to the conditional probability $P(A|B)$ which stands for the probability that A is the case given that B is known and which is defined for $P(B) \neq 0$ by $P(A|B) = P(A, B)/P(B)$. Because in the Boolean case the normalization factor $P(B)$ is always 0 or 1, one has to account only for the the case that $P(B) = 0$. And this is handled above via implication $\rightarrow$.

With the notion of a conditional of a Boolean function we can define conditional independence.

Definition 2. *For an n-ary Boolean function f and sets of propositional variables $A, B, C \subseteq [X_n]$ we say that A* is independent of B conditioned on C w.r.t. f, *for short $A \perp\!\!\!\perp B|C(f)$, iff for all $x \in \mathbb{B}^n$:*

$$\neg(f_C(x)) \vee \neg(f_{A|C}(x)) \vee \neg(f_{B|C}(x)) \vee (f_{A|B \cup C}(x) \leftrightarrow f_{A|C}(x)) \quad = \quad 1(x)$$

If $C = \emptyset$, then A is said to be (un-conditionally) independent of B w.r.t. f, for short: $A \perp\!\!\!\perp B(f)$.

This corresponds one-to-one to conditional independence in probability theory (Neopolitan 2003, p. 30). This notion of independence has an alternative representation which is the content of the following observation:

Observation 1. *For an n-ary Boolean function f and sets of propositional variables $A, B, C \subseteq [X_n]$ A is independent of B conditioned on C iff the following holds for all $x \in \mathbb{B}^n$:*

$$\neg(f_C(x)) \vee (f_{A \cup B|C}(x) \quad \leftrightarrow \quad (f_{A|C}(x) \wedge f_{B|C}(x))) = 1(x) \qquad (1)$$

As the notion of conditional independence developed here is an adaptation of conditional independence from probability theory, one might ask oneself whether it captures really logical dependencies between propositional variables. As Example 1 below shows, the notion of (un-)conditional independence according to Definition 2 captures indeed the usual logical dependencies between two variables. And this is no coincidence: The condition expressed in Equation (6) corresponds exactly to the notion of logical independence defined for sets of columns in the table (relation) of a relational data base (Fagin 1977): If $X_1, \ldots, X_n$ are the columns in a table/relation R, then for $A \subseteq [X_n]$ of cardinality k the k-ary relation R_A is the set of k-tuples x such that there is some $n-k$ tuple y such that $(x, y) \in R$. Then for sets $A, B, C \subseteq [X_n]$ one says that A is independent of B conditioned on C iff: For all $|A|$-ary tuples x, $|B|$-ary tuples y and $|C|$-ary tuples z: If $(x, z) \in R_{A \cup C}$ and $(y, z) \in R_{B \cup C}$, then $(x, y, z) \in R_{A \cup B \cup C}$. This is denoted $A \perp\!\!\!\perp B | C(R)$. (Note the twist in terminology: an embedded multivalued *dependency* induces a conditional *independency*.)

The strengthening of embedded multivalued dependencies are called multivalued dependencies: For these one additionally has the constraint that $A \cup B \cup C = [X_n] =$ the whole set of positions. The usual notation in this case is $C \twoheadrightarrow A$ and $C \twoheadrightarrow B$, or shorter $C \twoheadrightarrow A|B$.

For n-ary Boolean functions there are two natural sets associated with it that can play the role of an n-ary relation: The set $f^{-1}(1)$ of its models (also called the *onset*) and the set $f^{-1}(0)$ of non-models, which is exactly the set of models of its negation (sometimes called the *offset*). Considering the former, one sees immediately that conditional independence as defined by (6) is nothing else than embedded multivalued dependency for $f^{-1}(1)$ treated as an n-ary relation.

Proposition 1. *For Boolean functions f conditional independence, where marginals are defined as maximum-marginals, and conditional independence induced by embedded multivalued dependencies, where the models of f are considered as a DB-relation, are the same.*

This result is in so far remarkable, as in general conditional independence induced by embedded multivalued dependency over a database does not lead to the same CI triples as probabilistic conditional independence (see (Studený 1993)). But considering only propositional logic and max-marginals eliminates the difference.

Example 1. *The simplest nontrivial case of independence in this setting is for binary Boolean functions f. Consider $A = \{X_1\}, B = \{X_2\}$ and $C = \emptyset$. For illustration purposes we show that for $f = X_1 \wedge X_2$, X_1 is independent of X_2: $f_A = X_1$, $f_B = X_2$, hence if $f_A(x_1, x_2) = x_1 = 1$ and $f_B(x_1, x_2) = x_2 = 1$, then also $f(x_1, x_2) = x_1 \wedge x_2$. On the other hand, e.g., for $f = X_1 \rightarrow X_2$, X_1 and X_2 are not independent. Because $f_A(x) = f_B(x) = 1(x)$ for all x, but for $x = (1, 0)$ $f(x) = 0$. Similar calculations show that the binary Boolean functions f that fulfill the (unconditional) independence constraint $A \perp\!\!\!\perp B(f)$ are all 16 functions except the following six ones:*

1. $X_1 \rightarrow X_2$ (subsumption), 2. $X_2 \rightarrow X_1$ (supersumption), 3. $\neg X_1 \rightarrow X_2$ (covering), 4. $X_1 \rightarrow \neg X_2$ (disjointness), 5. $X_1 \leftrightarrow X_2$ (equivalence), 6. $X_1 \leftrightarrow \neg X_2$ (complement). The functions that do not satisfy the conditional independence property are the ones that describe the usual dependencies (given with their usual names in brackets) one encounters, say, in the area of semantic integration. The remaining functions are the ones with independent positions: 1 (constant 1 function), 0 (constant 0 function), the unary functions X_1, $\neg X_1$, X_2, $\neg X_2$ and the binary functions $X_1 \wedge X_2$, $\neg X_1 \wedge X_2$, $X_1 \wedge \neg X_2$, $\neg X_1 \wedge \neg X_2$.

4 Fourier Analysis of Boolean Functions

We are going to investigate independency of variables in a Boolean function using Fourier analysis. In this section we provide the necessary terminology, based mainly on (O'Donnell 2014).

For the following Fourier analytic discussions we are going to follow the convention (see (O'Donnell 2014)) of representing Boolean functions as functions $f : \mathbb{B}_\pm^n \longrightarrow \mathbb{B}_\pm$ where both, the domain and the range, are (cartesian products) of the two-element set $\mathbb{B}_\pm = \{1, -1\} \subseteq \mathbb{R}$. The translation between this and the previous representation is handled by $\chi : \mathbb{B} \longrightarrow \mathbb{B}_\pm$ mapping $0 \in \mathbb{B}$ to $1 \in \mathbb{B}_\pm$ and $1 \in \mathbb{B}$ to $-1 \in \mathbb{B}_\pm$: $\chi(0) = 1$, $\chi(1) = -1$. The definitions of the Boolean operators change accordingly. For example, the negation operator now becomes $\neg(x) = -x$ (where $-$ is the minus sign from the field of real numbers $\mathbb{R}$.)

Position $i \in [n]$ in a Boolean function f is said to be *pivotal for f* iff flipping the bit in i-position changes the value of f, formally, iff $f(x) \neq f(x^{\oplus i})$. Here we use $x^{\oplus i}$ to denote the tuple identical to x except that the bit at position i is changed to its complement. The *influence Inf_i of position i on f* is given as the relative frequency of those n-bit tuples $x \in \mathbb{B}_\pm^n$ which are pivotal for f:

$$\begin{aligned} \mathrm{Inf}_i(f) &= Pr_{x \sim \{-1,1\}}(f(x) \neq f(x^{\oplus i})) \\ &= E_{x \sim \{-1,1\}}(1_{f(x) \neq f(x^{\oplus i})}) \\ &= \frac{1}{2^n} \cdot \#\{x \in \mathbb{B}_\pm^n \mid f(x) \neq f(x^{\oplus i})\} \end{aligned}$$

Here and below we follow (O'Donnell 2014) in defining the relevant notions using the terminology of probability theory. $x \sim \mathbb{B}_\pm^n$ means that x is chosen from the uniform distribution over $\mathbb{B}_\pm^n$, $Pr(\cdot)$ denotes the corresponding probability distribution and $E(\cdot)$ the expectation value. $1_P(\cdot)$ is the characteristic function of a predicate P for elements $x \in \mathbb{B}_\pm^n$. The influence measure can be described analytically with the notion of a (partial) derivative using Fourier analysis. The i^{th} *discrete derivative operator* D_i maps any (n-ary) Boolean function f to the n-ary Boolean function

$D_i(f) : \mathbb{B}^n_\pm \longrightarrow \{-1,1\}$ defined for all $x \in \mathbb{B}^n_\pm$ by $D_i f(x) = \frac{f(x[x_i/1]) - f(x[x_i/-1])}{2}$. The square of the derivative is the indicator function for the property of i being pivotal for f. Hence the influence can also be written as

$$\begin{aligned} \text{Inf}_i(f) &= Pr_{x\sim\{-1,1\}}(f(x) \neq f(x^{\oplus i})) \\ &= E_{x\sim\{-1,1\}} D_i f(x)^2 \end{aligned}$$

Influences can be calculated conveniently using the fact that every Boolean function has a so-called unique *Fourier expansion*. The expansion is based on the fact that the parity functions $x^S : \mathbb{B}^n_\pm \longrightarrow \mathbb{B}_\pm$ defined by $x^S = \prod_{i\in S} x_i$ (and setting $x^\emptyset = 1$) for $S \subseteq [n]$ make up an orthonormal basis in the vector space of all Boolean functions of an arbitrary but fixed arity. The scalar product of two functions f, g is defined as $\langle f, g\rangle = E_{x\sim\mathbb{B}^n_\pm}(f(x) \cdot g(x)) = \frac{1}{2^n} \cdot \sum_{x\in\mathbb{B}_\pm} f(x) \cdot g(x)$. The *Fourier expansion* of f is $f(x) = \sum_{S\subseteq[n]} \hat{f}(S)x^S$ where $\hat{f}(S) \in \mathbb{R}$ denotes the *Fourier coefficients*. The set of Fourier coefficients is also called the *Fourier spectrum*. A Fourier expansion for functions $f : \mathbb{B}^n \longrightarrow \mathbb{R}$ over the domain $\mathbb{B}^n$ can derived by extending the definition of the χ-function to arbitrary sets $S \subseteq S$: $\chi_S(x) = \Pi_{i\in S}\chi(x_i) = -1^{\sum_{i\in S} X_i}$.

$$f(x) = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S \tag{2}$$

In fact, if $f : \mathbb{B}^n_\pm \longrightarrow \mathbb{B}_\pm$ is a function having domain $\mathbb{B}^n_\pm$ and range $\mathbb{B}_\pm$, where its Fourier expansion is given by the polynomial $p(x)$, then the Fourier expansion $q(x)$ of f represented over $\mathbb{B}$ is given as $q(x) = \frac{1}{2} - \frac{1}{2}p(1 - 2x_1, \ldots, 1-2x_n)$. In the other direction it holds that $p(x) = 1 - 2p(\frac{1-x_1}{2}, \ldots, \frac{1+x_n}{2})$. In particular, the Fourier coefficients of f w.r.t. these two domains are related by

$$\hat{f}^{\mathbb{B}}(S) = \begin{cases} \frac{1}{2}(1 - \hat{f}^{\mathbb{B}_\pm}(S)) & \text{if } S = \emptyset \\ -\frac{1}{2}\hat{f}^{\mathbb{B}_\pm}(S) & \text{else} \end{cases} \tag{3}$$

Plancherels Theorem states that the scalar product of two functions f and g is the sum of pointwise multiplication of the Fourier coefficients of f and g, i.e., $\langle f, g\rangle = \sum_{x\in\mathbb{B}_\pm} \hat{f}(x) \cdot \hat{g}(x)$. It's special case is *Parseval's theorem* stating that $\langle f, f\rangle = \sum_{x\in\mathbb{B}_\pm} \hat{f}(x)^2$. Using the simple structure of the Fourier spectrum of the derivatives, Equation (2) and Parseval's theorem, the following representation of the influence can be derived:

$$\text{Inf}_i(f) = \sum_{S\subseteq[n], i\in S} \hat{f}(x)^2 \tag{4}$$

The convolution $f * g$ of two functions $f, g : \mathbb{B}^n_\pm \longrightarrow \mathbb{R}$ is defined by $f * g(x) = \sum_{y\sim\mathbb{B}^n_\pm} f(y) \cdot g(y \cdot x)$ where $y \cdot x$ denotes bitwise multiplication of the bits in y and x. In case f, g are probability distributions, the convolution describes the probability for the sum of the random variables corresponding to f and g. The Fourier coefficients of a convolution are calculated as the product of the Fourier coeefficients of the components: $\widehat{f * g}(S) = \hat{f} \cdot \hat{g}$. Dually, the Fourier coefficients of the product of two functions f, g is the convolution of the Fourier coefficients of the components:

$$\widehat{f \cdot g}(S) = \hat{f} * \hat{g}(S) \tag{5}$$

5 Influence-Based Dependencies

The (in)dependency notions described in the first section lead to crisp criterions according to which two sets of propositional variables are dependent—conditioned on other propositional variables. In this section we are going to describe a measure that quantifies the dependency between sets of variables using a notion similar to that of influence known in the area of Boolean analysis. This influence-based notion of (in)dependency is quite different from the (in)dependence notions mentioned above, even so that the limit case of dependency measure = 0, i.e., no influence at all, does not coincide with the notion of conditional independence induced by embedded multivalued dependency.

We consider in this section two influence-based notions of dependency. For the second notion we are able to give an analytical expression using the notion of influence Inf_i of a variable on a function f.

Our first notion of dependence uses the notion of influence of position j on i w.r.t. Boolean function f as given in the following definition.

Definition 3. *For an n-ary Boolean function f and positions $i, j \in [n]$ the influence of j on i of type I, for short $Inf^I_{j,i}(f)$, is the relative frequency of those x for which i is not pivotal, but becomes pivotal with flipped j:*

$$\begin{aligned} Inf^I_{j,i}(f) &= Pr_{x\sim\mathbb{B}^n_\pm}(f(x) = f(x^{\oplus i}) \textit{ and} \\ &\quad f(x^{\oplus j}) \neq f(x^{\oplus j, \oplus i})) \end{aligned}$$

Here we used the notation $x^{\oplus j,\oplus i}$ denoting the outcome of first flipping the value at position j and then flipping the value at position i in x. As long as $i \neq j$, this is the same as flipping both in parallel.

This notion of the influence of a position j on a position i induces a notion of independence that is different from embedded multivalued independence. This new notion of independence is defined to hold between i and j iff the type-I-influence of j on i is zero.

Definition 4. *Position i (variable X_i) is* type-I-influence-independent *of position j (variable X_j) (for $i \neq j$) w.r.t. n-ary Boolean function f iff $Inf^I_{j,i}(f) = 0$.*

Remark 1. *Here and in the following we are going to consider mainly the special case of (unconditioned) independence of a variable from another variable.*

Using a different notion of derivative than the one introduced in the preliminary section, we can give a description of $\text{Inf}^{\text{I}}_{j,i}(f)$ based on Fourier coefficients. The new notion of a derivative, denoted $d_i f$, for functions $f : \mathbb{B}^n_\pm \longrightarrow \mathbb{B}_\pm$ is the exact adaptation of the derivative notion for functions $f : \mathbb{B}^n \longrightarrow \mathbb{B}$ over the Galois field as used, e.g., in (Vichniac 1990), and is defined as $d_i f(x) = f(x) \cdot f(x^{\oplus i})$. The indicator function for an i-induced flip f can be defined with this new notion of derivative as follows: $\theta_i(x) = \frac{1-d_i f(x)}{2}$. Then the opposite condition, i.e. that i is not pivotal for f on x, is expressed by $1 - \theta_i(x) = \frac{1+d_i f(x)}{2} = \frac{1+f(x)\cdot f(x^{\oplus i})}{2}$. A generalization to two variables describes the derivative w.r.t. both positions i, j: $d_{ij} f(x) = f(x) \cdot f(x^{\oplus i, \oplus j})$. An indicator function θ_{ij} indicating whether a flip in both positions

i, j leads to a flip in the f-value is accordingly given as $\theta_{ij}(x) = \frac{1-d_{ij}(x)}{2}$.

Theorem 1. *For any n-ary Boolean function $f : \mathbb{B}_\pm^n \longrightarrow \mathbb{B}_\pm$ and positions $i \neq j \in [n]$ it holds: i is type-I-influence-independent of j w.r.t. f iff for all x:*

$$f_{\restriction 1,1}(x) f_{\restriction -1,1}(x) = f_{\restriction 1,-1}(x) f_{\restriction -1,-1}(x))$$

The expression found in Theorem 1 shows that the induced notion of independence is much weaker than the notion of embedded multivalued dependence: Consider the special case of $[X_n] \setminus \{X_i, X_j\} \twoheadrightarrow X_i | X_j$. In this case one requires that if $f_{\restriction 1,1}(x) = -1 = f_{\restriction -1,1}(x)$, then also $f_{\restriction -1,-1}(x) = -1 = f_{\restriction 1,-1}(x)$, which is much stronger than to say $f_{\restriction 1,1}(x) f_{\restriction -1,1}(x) = f_{\restriction 1,-1}(x) f_{\restriction -1,-1}(x))$.

The following example illustrates the simple case of binary Boolean functions $f(X_1, X_2)$ and lists all functions with X_1 being independent of X_2 of type-I.

Example 2. *The set of binary functions f such that its positions $1, 2$ are type I influence-independent are the following ones: X_1, X_2, $\neg X_1$, $\neg X_2$, 1,-1,$X_1 \wedge X_2$,$\neg X_1 \wedge X_2$, $X_1 \wedge \neg X_2$, $\neg X_1 \wedge \neg X_2$, $X_1 \leftrightarrow X_2$, $X_1 \oplus X_2$.*

The expression $\frac{1}{4} - \frac{1}{4} E(f(x) f(x^{\oplus i}) f(x^{\oplus j}) f(x^{\oplus j \oplus i})$ derived in the proof of the Theorem 1 (see extended version) can be further broken down to the Fourier coefficients of f using the convolution of functions and Equation (5):

$$\begin{aligned} \text{Inf}^{\text{I}}_{j,i}(f) &= \frac{1}{4}(1 - \sum_{S \subseteq [n]} [\hat{f} * \widehat{f(()^{\oplus i})}](S) \cdot \\ &\quad [\widehat{f(()^{\oplus j})} * \widehat{f(()^{\oplus i \oplus j})}](S)) \end{aligned}$$

(In the above formula we used $f(()^{\oplus i})$ to denote the composition of f and the i-flip function $()^{\oplus i}$.)

Whereas the first notion of influence-based dependence measures the possible j-induced flips of f that hold independently of whether $x_i = +1$ or $x_i = -1$, the second related notion considers the number of f-flips which are not induced by i alone but by flipping both i and j.

Definition 5. *For an n-ary Boolean function f and positions $i, j \in [n]$ the influence of j on i, of type II, for short $\mathit{Inf}^{II}_{j,i}(f)$, is the relative frequency of those x for which i is not pivotal for f, but flipping both i, j leads to a change of the f-value:*

$$\mathit{Inf}^{II}_{j,i}(f) = Pr_{x \sim \mathbb{B}^n_\pm}(f(x) = f(x^{\oplus i}) \textit{ and } f(x) \neq f(x^{\oplus j, \oplus i}))$$

We give an analytic description of the type-II-influence via the Fourier spectrum. The condition used in the definition of the influence of type II holds if there is no flip induced by i and there is an $\{i, j\}$-induced flip. Using indicators for derivatives d_i and d_{ij} leads to the indicator function $\zeta(x) = \frac{1+d_i(f)}{2} \cdot \frac{1-f(x) \cdot f(x^{\oplus i, \oplus j})}{2}$. These considerations with some algebraic calculations results in the following representation of influence of type II via Fourier coefficients:

Theorem 2. *If $i \neq j$, then $\mathit{Inf}^{II}_{j,i}(f) = \sum_{\substack{S \subseteq [n] \\ i \notin S, j \in S}} \hat{f}(S)^2$.*

A simple corollary is the following connection between the influence $\text{Inf}_j(f)$ of j on the function f and the influence $\text{Inf}^{\text{II}}_{j,i}(f)$ of j on i w.r.t. f of type II: $\text{Inf}_j(f) = \text{Inf}^{\text{II}}_{j,i}(f) + \sum_{S \subseteq [n], i \in S, j \in S} \hat{f}(S)^2$.

6 Relation to (Embedded) MVDs

Independence of type II describes in essence that part of the common influence of $\{i, j\}$ on j that does not rest on the influence of i on f alone. This notion of common influence is known as *coalition influence*. In general, the coalition influence of a set of positions $J \subseteq [n]$ on an n-ary Boolean function f (O'Donnell 2014, p.274) is defined as the probability of finding an assignment z over non-J positions such that flipping one of the J-positions is pivotal for f with fixed z. Formally: $\text{Inf}_J(f) = Pr_{z \sim \mathbb{B}^{[n] \setminus J}_\pm}(f_{J \restriction z} \text{ is not constant})$. A refinement of this notion is that of *coalition influence towards* $b \in \mathbb{B}_\pm$ which describes the probability that flips in J lead to f becoming evaluated to b, formally: $\text{Inf}^b_J(f) = Pr_{z \sim \mathbb{B}^{[n] \setminus J}_\pm}(f_{J \restriction z} \text{can be made } b) - Pr[f = b]$. Coalition influence does not permit a neat analytical representation that holds for all Boolean functions f. But coalition influence is related to multivalued dependency as shown below.

The independence of position A from position B (with $C = \emptyset$) induced by multivalued dependency (Sect. 3) reads within the words of influential coalition as follows. Using the abbreviation $J = [n] \setminus \{A, B\}$ we get:

$$\begin{aligned} \forall a, b, \in \mathbb{B}_\pm : \quad & \text{Inf}^{-1}_J(f_{\restriction a}) = 0 \text{ or } \text{Inf}^{-1}_J(f_{\restriction b}) = 0 \text{ or} \\ & \text{Inf}^{-1}_J(f_{\restriction a}) = 1 = \text{Inf}^{-1}_J(f_{\restriction b}) \end{aligned}$$

As the coalition influence is combinatorial, this rephrasing of logical independence does not lead immediately to a characterization via the squares of the Fourier coefficients $\hat{f}(S)^2$. In fact, considering the Fourier expansions for the binary Boolean functions f reveals that a potential characterization cannot rest only on the squares of the Fourier coefficients, i.e., the signs of the Fourier coefficients are relevant for the decision whether independence holds.

A simple formula on the Fourier coefficients that characterizes the independence of $A = \{X_1\}$ and $B = \{X_2\}$ for the special case of Boolean binary functions is given in the following observation.

Observation 2. *$A = \{X_1\}$ and $B = \{X_2\}$ are independent w.r.t. $f : \mathbb{B}^2 \rightarrow \mathbb{B}$ (resp. $f : \mathbb{B}^2_\pm \rightarrow \mathbb{B}_\pm$) iff $\hat{f}(\emptyset) \cdot \hat{f}(\{X_1, X_2\}) = \hat{f}(\{X_1\}) \cdot \hat{f}(\{X_2\})$ (resp. $(\hat{f}(\emptyset) - 1) \cdot \hat{f}(\{X_1, X_2\}) = \hat{f}(\{X_1\}) \cdot \hat{f}(\{X_2\})$).*

Ordinary MVDs can be broken down to the sum of influences as developed in the preceding sections. We get a decomposition of the probability that a Boolean vector x falsifies the MVD conditions into four summands: the influence of j on f, the influence of i on f, the dependence of i of j of type I, and an additional term $E(\frac{1+f(x)f(x^{\oplus i})f(x^{\oplus j})}{2})$. This term defined below describes the probability that there is an i-flip if $f(x^{\oplus j}) = -1$ or no flip if $f(x^{\oplus j}) = 1$.

Definition 6. *For an n-ary Boolean function f and positions $i, j \in [n]$, the influence of j on i of type III, for short $\mathit{Inf}^{III}_{j,i}(f)$, is the relative frequency of those x for which i is pivotal for f and flipping j gives f-value -1 or i is not pivotal and flipping j gives f-value 1.*

$$\begin{aligned} \mathit{Inf}^{III}_{j,i}(f) &= Pr_{x \sim \mathbb{B}^n_\pm}(f(x^{\oplus j}) = -1 \ \&\ f(x) \neq f(x^{\oplus i}) or \\ &\quad f(x^{\oplus j}) = 1 \ \&\ f(x) = f(x^{\oplus i})) \end{aligned}$$

The following theorem formalizes the decomposition.

Theorem 3. *The probability X of an x falsifying the MVD $[X_n] \setminus \{X_i, X_j\} \twoheadrightarrow X_i \mid X_j$ is described by* $X = \frac{1}{4}(\frac{1}{2}\mathit{Inf}_i(f) + \frac{1}{2}\mathit{Inf}_j(f) + \frac{1}{2}\mathit{Inf}^{I}_{j,i} + \mathit{Inf}^{III}_{j,i} - 1)$.

7 Related Work

We gave criteria for identifying independencies between variables of a Boolean function within Fourier coefficients. Fourier expansions are flat structures: They consider a fixed basis of parity functions. Sum-product networks introduced in (Poon and Domingos 2012) describe a deep architecture for Boolean functions in which conditional independencies are implemented directly as structural constraints.

The approaches of (Dawid 2001; Wang 2010; Studeny 2005) are mathematical theories meant to capture the essence of CI structures. With our approach we did not add another structure but considered the well-established Fourier expansion of Boolean functions and the notion of influence as a means for analyzing independence.

Influence plays a pivotal role in the measures of dependence and the induced notions of independence. The notion of coalition influence as used in Sect. 5 and Sect. 6 is relevant for many different properties of Boolean functions but is not easily describable analytically. The considerations on coalition influences of (Ben-Or and Linial 1985) and (Filmus 2016) may help in finding an analytical expression.

The notions of influence used here bears some resemblance to the notion of influence in causal networks (Pearl 2009) and, in particular, to instrumental variables as used in linear causal networks (Brito and Pearl 2002).

Conditional independencies can be used for the decomposition of different data structures. In (Parikh 1999) decompositions of propositional knowledge bases using so-called splittings are considered. The notion of irrelevance induced by finest splitting bears strong similarities to CIs.

8 Conclusion and Outlook

The Fourier analysis framework is an invaluable means for the analysis of Boolean functions. In this paper, we relied on this framework in order to identify criteria on conditional independencies using the notion of flips and influences of variables. As far as we know, the results of this paper are the first to draw connections between influence and independence, in particular independence induced by multivalued dependencies. So, with this paper we provide the foundation for further interesting investigations of CIs with Fourier analysis.

One open problem for future investigations is a full characterization of (embedded) multivalued dependencies with Fourier coefficients. The simple Fourier characterization of binary functions having independent arguments given in Section 6 works because one does not have to account for $\vee$-marginals in this sample case. The combinatorics of coalition influences can be tamed by constraining the class of Boolean functions. A possible strategy is to consider functions defined not on the whole Boolean cube but on some subset, e.g., a Boolean slice, i.e., a subset of $\mathbb{F}^n$ where all vectors have the same weight (Filmus 2016).

References

Ben-Or, M., and Linial, N. 1985. Collective coin flipping, robust voting schemes and minima of Banzhaf values. In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, 408–416.

Brito, C., and Pearl, J. 2002. Generalized instrumental variables. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, UAI'02, 85–93. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

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

Fagin, R. 1977. Multivalued dependencies and a new normal form for relational databases. *ACM Trans. Database Syst.* 2(3):262–278.

Filmus, Y. 2016. An orthogonal basis for functions over the boolean hypercube. *The Electronic Journal of Combinatorics* 23(1).

Neopolitan, R. E. 2003. *Learning Bayesian Networks*. Pearson.

O'Donnell, R. 2014. *Analysis of Boolean Fuctions*. Cambridge University Press.

Özçep, Ö. L. 2016. Minimality postulates for ontology revision. In Baral, C.; Delgrande, J. P.; and Wolter, F., eds., *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, 2016*, 589–592. AAAI Press.

Parikh, R. 1999. Beliefs, belief revision, and splitting languages. In Moss, L.; Ginzburg, J.; and de Rijke, M., eds., *Logic, Language and Computation*, volume 2. CSLI Publications. 266–278.

Pearl, J., and Paz, A. 1985. Graphoids: a graph-based logic for reasoning about relevance relations. Technical Report R53 CSD-850038, University of California.

Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. New York, NY, USA: CUP, 2nd edition.

Poon, H., and Domingos, P. 2012. Sum-Product Networks: A New Deep Architecture. *ArXiv e-prints*.

Spohn, W. 1994. On the properties of conditional independence. In Humphreys, P., ed., *Patrick Suppes: Scientific Philosopher: Volume 1. Probability and Probabilistic Causality*, 173–196. Dordrecht: Springer Netherlands.

Studený, M. 1993. Formal properties of conditional independence in different calculi of AI. In Clarke, M.; Kruse, R.; and Moral, S., eds., *Symbolic and Quantitative Approaches to Reasoning and Uncertainty: ECSQARU '93 Granada, Spain, November 8–10, 1993 Proceedings*, 341–348. Springer.

Studeny, M. 2005. *Probabilistic Conditional Independence Structures*. Santa Clara, CA, USA: Springer-Verlag.

Vichniac, G. 1990. Boolean derivatives on cellular automata. *Phys. D* 45(1-3):63–74.

Wang, J. 2010. A universal algebraic approach for conditional independence. *Annals of the Institute of Statistical Mathematics* 62(4):747–773.

A Proofs

A.1 Proof of Lemma 1

If $f_B^m(x) = 1$, this means that there is a vector $y \in \mathbb{B}^n$ that is identical with x for the variables in B and $f(y) = 1$. But this y is also identical with x for the variables in $A \subseteq B$ and so $f_A^m(x) = 1$.

A.2 Proof of Proposition 1

Due to Lemma 1 the direction from left to right in the equation of Observation 1 is always true. So the equation reduces to

$$1 \quad = \quad \neg(f_C(x)) \vee (f_{A\cup B|C}(x) \leftarrow (f_{A|C}(x) \wedge f_{B|C}(x)))$$

Moreover, instantiating the definitions of the conditionals and equivalence constructions one gets

$$1 \quad = \quad f_C(x) \rightarrow ([f_C(x) \rightarrow (f_{A\cup C}(x) \wedge f_{B\cup C}(x))] \rightarrow [f_C(x) \rightarrow f_{A\cup B\cup C}(x)])$$

This again is equivalent to

$$1 \quad = \quad (f_C(x) \wedge f_{A\cup C}(x) \wedge f_{B\cup C}(x)) \rightarrow f_{A\cup B\cup C}(x)])$$

Due to Lemma 1 the antecedens of the implication can even be reduced such that we get

$$1(x) \quad = \quad (f_{A\cup C}(x) \wedge f_{B\cup C}(x)) \rightarrow f_{A\cup B\cup C}(x)]) \quad (6)$$

The condition expressed in Equation (6) is nothing else than that of embedded multivalued dependencies.

A.3 Proof of Theorem 1

$$\begin{aligned}
\mathrm{Inf}^{\mathrm{I}}_{j,i}(f) &= E(\frac{1+f(x)f(x^{\oplus i})}{2} \cdot \frac{1-f(x^{\oplus j})f(x^{\oplus j\oplus i})}{2}) \\
&= \frac{1}{4}E(1+f(x)f(x^{\oplus i}) - f(x^{\oplus j})f(x^{\oplus j\oplus i}) - \\
&\quad f(x)f(x^{\oplus i})f(x^{\oplus j})f(x^{\oplus j\oplus i})) \\
&= \frac{1}{4}E(1 - f(x)f(x^{\oplus i})f(x^{\oplus j})f(x^{\oplus j\oplus i})) \\
&= \frac{1}{4} - \frac{1}{4}E(f(x)f(x^{\oplus i})f(x^{\oplus j})f(x^{\oplus j\oplus i})) \\
&= \frac{1}{4} - \frac{1}{4}\cdot\frac{1}{2^n}\sum_{x\in\mathbb{B}_\pm^{n-2}} \cdot(f_{\upharpoonright 1,1}(x)f_{\upharpoonright -1,1}(x)\cdot \\
&\quad f_{\upharpoonright 1,-1}(x)f_{\upharpoonright -1,-1}(x))
\end{aligned}$$

Hence, $\mathrm{Inf}^{\mathrm{I}}_{j,i}(f) = 0$ iff $2^{n-2} = \sum_{x\in\mathbb{B}_\pm^{n-2}} 4 \cdot (f_{\upharpoonright 1,1}(x)f_{\upharpoonright -1,1}(x)f_{\upharpoonright 1,-1}(x)f_{\upharpoonright -1,-1}(x)$, i.e., iff for all $x \in \mathbb{B}_\pm^{n-2}$ it holds that $f_{\upharpoonright 1,1}(x)f_{\upharpoonright -1,1}(x)f_{\upharpoonright 1,-1}(x)f_{\upharpoonright -1,-1}(x)) = 1$, which means that for all x: $f_{\upharpoonright 1,1}(x)f_{\upharpoonright -1,1}(x) = f_{\upharpoonright 1,-1}(x)f_{\upharpoonright -1,-1}(x))$.

A.4 Proof of Theorem 2

Let $E(\cdot) = E_{x\sim\mathbb{B}_\pm^n}(\cdot)$ for the following.

$$\begin{aligned}
&\mathrm{Inf}^{\mathrm{II}}_{j,i}(f) = E(\zeta(x)) \\
&= E(\tfrac{1+f(x)\cdot f(x^{\oplus i})}{2} \cdot \tfrac{1-f(x)\cdot f(x^{\oplus i,\oplus j})}{2}) \\
&= E(\tfrac{1+f(x)\cdot f(x^{\oplus i}) - f(x)^2 f(x^{\oplus i}) f(x^{\oplus i,\oplus j}) - f(x)f(x^{\oplus i,\oplus j})}{4}) \\
&= \tfrac{1}{4}\big(E(1) + E(f(x)\cdot f(x^{\oplus i})) - E(f(x^{\oplus i})f(x^{\oplus i,\oplus j})) - \\
&\quad E(f(x)f(x^{\oplus i,\oplus j}))\big) \\
&= \tfrac{1}{4}\big(1 + \textstyle\sum_{S\subseteq[n]} \hat{f}(S)\cdot \widehat{f\circ(\cdot)^{\oplus i}}(S) \\
&\quad - \textstyle\sum_{S\subseteq[n]} \widehat{f\circ(\cdot)^{\oplus i}}(S)\cdot \widehat{f\circ(\cdot)^{\oplus i,\oplus j}}(S) \\
&\quad - \textstyle\sum_{S\subseteq[n]} \hat{f}(S)\cdot \widehat{f\circ(\cdot)^{\oplus i,\oplus j}}(S)\big)
\end{aligned}$$

Now one has to calculate the Fourier coefficients. For example, in order to find $\widehat{f\circ(\cdot)^{\oplus i}}(S)$ one considers the Fourier expansion of f for the argument $x^{\oplus i}$: $f(x^{\oplus i}) = \sum_{S\subseteq[n]} \hat{f}(S)\cdot(x^{\oplus i})^S$. As $(x^{\oplus i})^S = x^S$ if $i \notin S$ and $(x^{\oplus i})^S = -x^S$ if $i \in S$ one gets $f\circ(\cdot)^{\oplus i}(x) = f(x^{\oplus i}) = \sum_{S\subseteq[n], i\notin S} \hat{f}(S)\cdot x^S - \sum_{S\subseteq[n], i\in S} \hat{f}(S)\cdot x^S$. So $\widehat{f\circ(\cdot)^{\oplus i}}(S) = \hat{f}(S)$ for $i \notin S$, else $\widehat{f\circ(\cdot)^{\oplus i}}(S) = -\hat{f}(S)$. Similarly one calculates that $\widehat{f\circ(\cdot)^{\oplus i,\oplus j}}(S) = \hat{f}(S)$ if $i,j \notin S$ or $i \in S$ & $j \notin S$ and $\widehat{f\circ(\cdot)^{\oplus i,\oplus j}}(S) = -\hat{f}(S)$ else, i.e., if $i \in S \& j \notin S$ or $i \notin S \& j \in S$. Substituting these in the equation above and using the fact that $1 = \sum_{S\subseteq[n]} \hat{f}(S)^2$ leads to the following chain of equations:

$$\begin{aligned}
&\mathrm{Inf}^{\mathrm{II}}_{j,i}(f) \\
&= \tfrac{1}{4}\big(\textstyle\sum_{S\subseteq[n]} \hat{f}(S)^2 \\
&\quad + \textstyle\sum_{S\subseteq[n], i\notin S} \hat{f}^2(S) - \sum_{S\subseteq[n], i\in S} \hat{f}^2(S) \\
&\quad - \textstyle\sum_{S\subseteq[n], i,j\notin S} \hat{f}^2(S) - \sum_{S\subseteq[n], i\in S, j\notin S} \hat{f}^2(S) \\
&\quad + \textstyle\sum_{S\subseteq[n], i\notin S, j\in S} \hat{f}^2(S) + \sum_{S\subseteq[n], i,j\in S} \hat{f}^2(S) \\
&\quad - \textstyle\sum_{S\subseteq[n], i,j\notin S} \hat{f}^2(S) - \sum_{S\subseteq[n], i,j\in S} \hat{f}^2(S) \\
&\quad + \textstyle\sum_{S\subseteq[n], i\notin S, j\in S} \hat{f}^2(S) + \sum_{S\subseteq[n], i\in S, j\notin S} \hat{f}^2(S)\big) \\
&= \tfrac{1}{4}\big(2\textstyle\sum_{S\subseteq[n], i\notin S} \hat{f}(S)^2 - 2\sum_{S\subseteq[n], i,j\notin S} \hat{f}^2(S) \\
&\quad + 2\textstyle\sum_{S\subseteq[n], i\notin S, j\in S} \hat{f}^2(S)\big) \\
&= \tfrac{1}{4}\big(4\textstyle\sum_{S\subseteq[n], i\notin S, j\in S} \hat{f}^2(S)\big) = \sum_{S\subseteq[n], i\notin S, j\in S} \hat{f}^2(S)
\end{aligned}$$

A.5 Proof of Observation 2

For the proof we work with the representation of binary Boolean functions $f : \mathbb{B}^2 \rightarrow \mathbb{B}$ w.r.t. $\mathbb{B}$, and then derive the result for functions f represented w.r.t. $\mathbb{B}_\pm$. Consider the independence condition expressed in Equation (6). The other direction $\leftarrow$ holds due to the definition of the marginal. Using this definition of the marginal and considering the case where $C = \emptyset$, the condition can be written as

For all $a, b \in \mathbb{B}$: $f(a,b)\cdot f(\neg a, \neg b) - f(a, \neg b)\cdot f(a, \neg b) = 0$

Instantiating the pairs (a,b) with all pairs in $\mathbb{B}$ leads four times to the same equation

$$f(0,0)\cdot f(1,1) - f(0,1)\cdot f(1,0) = 0 \quad (7)$$

Consider the Fourier expansion of $f(x)$ according to Equation (2):

$\hat{f}(\emptyset) + \hat{f}(\{X_1\})(-1)^{X_1} + \hat{f}(\{X_2\})(-1)^{X_2} + \hat{f}(\{X_2\})(-1)^{X_1+X_2}$. Instantiating all pairs in this expansion gives

$$\begin{aligned}
f(0,0) &= \hat{f}(\emptyset) + \hat{f}(\{X_1\}) + \hat{f}(\{X_2\}) + \hat{f}(\{X_1, X_2\}) \\
f(1,1) &= \hat{f}(\emptyset) - \hat{f}(\{X_1\}) - \hat{f}(\{X_2\}) + \hat{f}(\{X_1, X_2\}) \\
f(0,1) &= \hat{f}(\emptyset) + \hat{f}(\{X_1\}) - \hat{f}(\{X_2\}) + \hat{f}(\{X_1, X_2\}) \\
f(1,0) &= \hat{f}(\emptyset) - \hat{f}(\{X_1\}) + \hat{f}(\{X_2\}) + \hat{f}(\{X_1, X_2\})
\end{aligned}$$

This results in $f(0,0) \cdot f(1,1) = (\hat{f}(\emptyset) + \hat{f}(\{X_1, X_2\})^2 - (\hat{f}(\{X_1\}) + \hat{f}(\{X_2\}))^2$ and $f(0,1) \cdot f(1,0) = (\hat{f}(\emptyset) - \hat{f}(\{X_1, X_2\})^2 - (\hat{f}(\{X_1\}) - \hat{f}(\{X_2\}))^2$. Putting these into the independency constraint in Eq. (7) leads to the simple constraint

$$\hat{f}(\emptyset) \cdot \hat{f}(\{X_1, X_2\}) = \hat{f}(\{X_1\}) \cdot \hat{f}(\{X_2\})$$

So X_1 is independent of X_2 iff the Fourier coefficient for $\{X_1, X_2\}$ is the product of the Fourier coefficients of X_1 and X_2 normalized by the inverse of the Fourier coefficient of the empty set $\frac{1}{\hat{f}(\emptyset)}$. In case one represents f w.r.t. $\mathbb{B}_\pm$ this constraint translates to (using Eq. 3): $(\hat{f}(\emptyset) - 1) \cdot \hat{f}(\{X_1, X_2\}) = \hat{f}(\{X_1\}) \cdot \hat{f}(\{X_2\})$.

A.6 Proof of Theorem 3

Consider the multivalued dependency $[X_n] \setminus \{X_i, X_j\} \twoheadrightarrow X_i \mid X_j$. For ease of description let us assume that $i = 1, j = 2$. This condition says that for a function $f : \mathbb{B}_\pm^n \longrightarrow \mathbb{B}_\pm$ any $x \in \mathbb{B}_\pm^{n-2}$ fulfills the predicate *mvd*(x, f) defined as follows:

$$\begin{aligned}
&(f(1,1,x) = -1 \text{ and } -1 = f(-1,-1,x) \quad \text{iff} \\
&f(1,-1,x) = -1 \text{ and } -1 = f(-1,1,x))
\end{aligned} \tag{8}$$

This in turn is exactly the case iff for all $x \in \mathbb{B}_\pm^{n-2}$ the following equation holds:

$$\begin{aligned}
&\overbrace{\frac{1}{2}f(1,1,x) + \frac{1}{2}f(-1,-1,x) - \frac{1}{2}f(1,1,x)f(-1,-1,x)}^{\alpha} = \\
&\quad \underbrace{\frac{1}{2}f(1,-1,x) + \frac{1}{2}f(-1,1,x) - \frac{1}{2}f(1,-1,x)f(-1,1,x)}_{\beta}
\end{aligned}$$

Using the fact that two expressions α, β evaluating to a value in $\mathbb{B}_\pm$ are different exactly iff the indicator function $1 - \frac{\alpha \cdot \beta}{2}$ evaluates to 1 leads to the condition that $0 = P(1 - \frac{\alpha \cdot \beta}{2})$. Resubstituting the expressions for α, β and rearranging the terms gives:

$$\begin{aligned}
0 &= P(1 - \frac{\alpha \cdot \beta}{2}) \\
&= \frac{1}{8}E(4 - 1 - f(1,1,x)f(1,-1,x) - \\
&\quad f(1,1,x)f(-1,1,x) + f(1,1,x)f(1,-1,x)f(-1,1,x) \\
&\quad -f(-1,-1,x)f(1,-1,x) - f(-1,-1,x)f(-1,1,x) \\
&\quad +f(-1,-1,x)f(1,-1,x)f(-1,1,x) \\
&\quad +f(1,-1,x)f(1,1,x)f(-1,-1,x) \\
&\quad +f(-1,1,x)f(1,1,x)f(-1,-1,x) \\
&\quad -f(1,1,x)f(-1,-1,x)f(1,-1,x)f(-1,1,x))
\end{aligned}$$

The different summands can be described using i and j flips, so that the following equation results:

$$\begin{aligned}
0 &= P(1 - \frac{\alpha \cdot \beta}{2}) \\
&= \frac{1}{4}\Big(\frac{1}{2}E(\frac{1 - f(x)f(x^{\oplus j})}{2}) + \frac{1}{2}E(\frac{1 - f(x)f(x^{\oplus i})}{2}) \\
&\quad +E(\frac{1 + f(x)f(x^{\oplus i})f(x^{\oplus j})}{2}) \\
&\quad +\frac{1}{4}E(\frac{1 - f(x)f(x^{\oplus i})f(x^{\oplus j})f(x^{\oplus i \oplus j})}{2}) - E(1)\Big)
\end{aligned}$$