NO ELEVENTH CONDITIONAL INGLETON INEQUALITY

TOBIAS BOEGE

ABSTRACT. A rational probability distribution on four binary random variables $X,Y,Z,U$ is constructed which satisfies the conditional independence relations $[X \perp Y]$, $[X \perp Z \mid U]$, $[Y \perp U \mid Z]$ and $[Z \perp U \mid XY]$ but whose entropy vector violates the Ingleton inequality. This settles a recent question of Studený (IEEE Trans. Inf. Theory vol. 67, no. 11) and shows that there are, up to symmetry, precisely ten inclusion-minimal sets of conditional independence assumptions which make the Ingleton inequality hold. The last case in the classification of which of these inequalities are essentially conditional is also settled.

1. Summary

This note answers Open Question 1 and one half of Open Question 2 raised by Milan Studený in his recent article [Stu21] on conditional Ingleton information inequalities on four discrete random variables $X,Y,Z,U$. The first result is the following rational binary distribution represented by its atomic probabilities $p_{ijkt} = P(X = i, Y = j, Z = k, U = t)$:

\[
\begin{align*}
    p_{0000} &= 20/77, \quad p_{0001} = 0, \quad p_{0010} = 0, \quad p_{0011} = 0, \\
    p_{0100} &= 40/693, \quad p_{0101} = 4/99, \quad p_{0110} = 10/693, \quad p_{0111} = 2/99, \\
    p_{1000} &= 20/693, \quad p_{1001} = 40/693, \quad p_{1010} = 1/693, \quad p_{1011} = 2/99, \\
    p_{1100} &= 0, \quad p_{1101} = 0, \quad p_{1110} = 0, \quad p_{1111} = 2/11,
\end{align*}
\]

which satisfies solely the four conditional independence statements $[X \perp Y]$, $[X \perp Z \mid U]$, $[Y \perp U \mid Z]$ and $[Z \perp U \mid XY]$ and on which the Ingleton expression evaluates to a negative number close to $-0.00757$. This example settles simultaneously the last three open cases in the classification of CI-type conditional Ingleton inequalities on four discrete random variables and shows that all ten of them were already described in [Stu21].

With the knowledge of all conditional Ingleton inequalities, we also settle the last remaining case in the classification of their essential conditionality. The results are summarized in:

**Theorem.** On four discrete random variables $X,Y,Z,U$ there are precisely ten inclusion-minimal conditional independence assumptions which make Ingleton’s inequality $\square(XY \mid ZU) \geq 0$ hold for entropy vectors (up to the symmetries $X \leftrightarrow Y$ and $Z \leftrightarrow U$ in the Ingleton expression), namely:

\[
\begin{align*}
    (1.1) & \quad [Z \perp U] & \quad (2.1) & \quad [X \perp Y] \wedge [X \perp Y \mid Z] \\
    (1.2) & \quad [X \perp Z] & \quad (2.2) & \quad [X \perp Y \mid Z] \wedge [Y \perp U \mid Z] \\
    (1.3) & \quad [X \perp Z \mid Y] & \quad (2.3) & \quad [X \perp Z \mid U] \wedge [X \perp U \mid Z] \\
    (1.4) & \quad [X \perp Y \mid ZU] & \quad (2.4) & \quad [X \perp Z \mid U] \wedge [Z \perp U \mid X] \\
    (1.5) & \quad [X \perp Z \mid YU] & \quad (2.5) & \quad [X \perp Z \mid U] \wedge [Y \perp Z \mid U]
\end{align*}
\]

The conditional Ingleton inequalities given by (1.1)--(1.5) are not essentially conditional but the ones given by (2.1)--(2.5) are essentially conditional.

Date: March 18, 2024.

2020 Mathematics Subject Classification. 94A17, 62R01, 94-08.

Key words and phrases. entropy region, conditional information inequality, conditional independence, essential conditionality, Ingleton inequality, rationality.
These results are derived computationally. Section 2 gives an introduction to the topic of conditional Ingleton inequalities and recalls the previous results leading to the question answered here. For basics on polymatroids and their role in conditional independence and information theory we refer to the excellent exposition in [Stu21]. The computational methodologies used to find the above distribution and to prove essential conditionality of inequality (2.5) are explained in Sections 3 and 4, respectively. Section 5 collects further remarks and observations. The source code in Macaulay2 [M2] and Mathematica [WM] behind various steps in the computations and auxiliary data produced using 4ti2 [4ti2] and normaliz [Nor] are available at
https://mathrepo.mis.mpg.de/ConditionalIngleton/.

2. On conditional Ingleton inequalities

2.1. Ingleton inequality and entropy region. Suppose that \( X, Y, Z, U \) are subspaces in a finite-dimensional (left or right) vector space over a field (or division ring). For this data, the Ingleton inequality asserts that

\[
0 \leq \dim\langle X, Z \rangle + \dim\langle X, U \rangle + \dim\langle Y, Z \rangle + \dim\langle Y, U \rangle + \dim\langle Z, U \rangle - \\
\dim\langle X, Y \rangle - \dim\langle Z \rangle - \dim\langle U \rangle - \dim\langle X, Z, U \rangle - \dim\langle Y, Z, U \rangle,
\]

where \( \dim(\cdot) \) is the dimension of the subspace spanned by its arguments. This rank function \( h : 2^N \to \mathbb{R}, i \mapsto \dim(I) \) on subsets of \( N = \{X, Y, Z, U\} \) is a polymatroid, i.e., it is normalized: \( h(\emptyset) = 0 \), non-decreasing: \( h(I) \leq h(J) \) for \( I \subseteq J \), and submodular: \( h(I) + h(J) \geq h(I \cup J) + h(I \cap J) \) for all \( I, J \subseteq N \). The set of polymatroids over an \( n \)-element set forms a rational polyhedral cone in \( \mathbb{R}^{2^n} \) denoted by \( H_n \). The Ingleton expression \( \square(XY|ZU) \) is the linear functional in \( h \) which appears on the right-hand side of (□). Hence, non-negativity of the inner product \( \square(XY|ZU) : h \) is a necessary condition for a polymatroid \( h \) to be linearly representable over some division ring. The necessity was found by Ingleton [Ing71] through an analysis of the Vámos matroid, the prototypical example of a non-linear matroid.

Now let \( X, Y, Z, U \) denote jointly distributed random variables which take only finitely many states. These random variables are referred to as discrete with finiteness being implicit. If \( X \) attains \( q \) states, without loss of generality from the set \( \{q\} := \{1, \ldots, q\} \), with positive probabilities \( p(X = i) \), then its Shannon entropy is the expression

\[
H(X) := E_X[\log 1/p] = \sum_{i=1}^{q} p(X = i) \log 1/p(X = i).
\]

The entropy vector of jointly distributed discrete random variables \( X_1, \ldots, X_n \) assigns to each subset \( I \subseteq [n] \) the entropy of the vector-valued discrete random variable \( X_I := (X_i : i \in I) \). We denote the entropy region, the set of all points in \( \mathbb{R}^{2^n} \) which occur as entropy vectors of \( n \) discrete random variables, by \( H_n \). The choice of basis for the logarithm changes the scale of all entropy vectors and does not change any of the considerations in this paper.

Fujishige [Fuj8] observed that the non-negativity of Shannon’s information measures implies that entropy vectors are polymatroids and thus entropy vectors are sometimes called entropic polymatroids. A result of Matúš [Mat97, Lemma 10] implies that every integer polymatroid which is linearly representable by a subspace arrangement over a field is a scalar multiple of an entropic one. Hence, it makes sense to reinterpret Ingleton’s functional \( \square(XY|ZU) \) by replacing \( \dim(\cdot) \) with \( H(\cdot) \). But whereas the inequality \( \square \geq 0 \) is valid for linear polymatroids, it fails for the more general entropic ones. This paper is concerned with special types of assumptions on entropy vectors which guarantee that the Ingleton inequality holds.

2.2. Discrete representability of CI structures. Even though the Ingleton inequality does not hold universally for entropy vectors, it was a key tool in the characterization of conditional independence (CI) structures which are representable by four discrete random variables. This classification was achieved in the series of papers [MS95, Mat95, Mat99] by Matúš and Studený and we use this section to outline the role of the Ingleton inequality in this work.
Let $I, J, K \subseteq [n]$. The common shorthand notation \(IJ := I \cup J\) applies to these subsets. For a polymatroid \(h\) and \(I, J, K \subseteq [n]\), we employ the difference expression

\[
\triangle(I, J|K) \cdot h := h(IK) + h(JK) - h(JJK) - h(K),
\]

that is, \(\triangle(I, J|K)\) is a linear functional on \(\mathbb{R}^{2^n}\). The non-negativity of this functional on \(\mathbf{H}_n\) is guaranteed by the submodular inequalities. Its vanishing makes \(IK\) and \(JK\) a modular pair.

If \(h\) is the entropy vector of random variables \((X_i : i \in [n])\), then \(\triangle(I, J|K)\) is known as the conditional mutual information of subvectors \(X_I\) and \(X_J\) given \(X_K\) and its vanishing is equivalent to the conditional independence \([X_I \perp \!\!\!\perp X_J | X_K]\). Recall from [Stu21, Section II.D] that the study of conditional independence (excluding functional dependence) can be reduced to the elementary CI statements, i.e., the equalities \(\triangle(i, j|K) = 0\) where \(i\) and \(j\) are distinct singletons and \(K\) is a subset of \(N\) not containing \(i\) or \(j\). These functionals define facets of \(\mathbf{H}_n\) and even supporting hyperplanes of \(\mathbf{H}_n^*\) with non-empty intersection. A set \(\mathcal{L}\) of elementary CI statements on \(n\) random variables, also called a CI structure, is representable if and only if there exists \(h \in \mathbf{H}_n^*\) such that \(\triangle(i, j|K) \cdot h = 0 \iff [i \perp \!\!\!\perp j | K] \in \mathcal{L}\). The CI structure defined by any polymatroid \(h\) in this way is denoted by \([h]\).

Let \(\mathbf{H}_4^1\) denote the subcone of \(\mathbf{H}_4\) (whose ground set elements are labeled \(X, Y, Z, U\)) which consists of polymatroids satisfying the Ingleton inequalities \(\square(IJ|KL) \geq 0\) for every possible permutation \(I, J, K, L\) of \(X, Y, Z, U\). There are \(\binom{4}{2} = 6\) unique such inequalities because the Ingleton expression is invariant under exchanging \(I \leftrightarrow J\) and \(K \leftrightarrow L\). One key insight of [MS95] is that the extreme rays of \(\mathbf{H}_4^1\) are a subset of those of \(\mathbf{H}_4\) and that they are all probabilistically representable. This implies that every CI structure \([h]\), for \(h \in \mathbf{H}_4^1\), is representable; this condition is of polyhedral nature and can easily be checked using linear programming. Miraculously, even in the non-Ingleton regime, the Ingleton inequality is the main obstruction to entropicness: namely, in [Mat99] sets of CI statements \(\mathcal{L}\) is described such that whenever a polymatroid \(h\) is entropic and satisfies \(\mathcal{L} \subseteq [h]\), then \(\square(XY|ZU) \cdot h \geq 0\) holds. This is a conditional information inequality in the sense of [KR13], formally written as \(\mathcal{L} \Rightarrow \square(XY|ZU) \geq 0\) and called a conditional Ingleton inequality. It is important to emphasize that a conditional information inequality is not required to hold for general polymatroids (in which case it would be a consequence of the polyhedral geometry of \(\mathbf{H}_4\) and not very informative) but only for entropic polymatroids. An inequality such as \(\mathcal{L} \Rightarrow \square(XY|ZU) \geq 0\) allows one to conclude that a CI structure containing \(\mathcal{L}\) cannot be representable if the cone of its realizing polymatroids does not intersect the cone given by \(\square(XY|ZU) \geq 0\); which is again a polyhedral condition that can be computed easily.

**Convention.** The definition of conditional information inequality in [KR13] allows arbitrary linear assumptions \(p_1 \cdot h \geq 0 \land \cdots \land p_s \cdot h \geq 0\) to imply a linear conclusion \(q \cdot h \geq 0\). Conditional independence assumptions are a special case of this using \(\triangle\) functionals. In this work, “conditional information inequality” will always refer to the special case of CI-type inequality.

While the precise shape of \(\mathbf{H}_4^*\) or even its closure \(\overline{\mathbf{H}}_4^1\) in the euclidean topology (which is known to be a convex cone [ZY97]) remains unknown to date (cf. [Mat07] and [GMM17] for a challenging open problem), conditional information inequalities help to delimit it in ways that go beyond linear inequalities and hence make it possible to describe differences between the entropy region and its closure. This becomes significant, for example, when information-theoretic optimization problems such as channel capacity computations are solved not in terms of their original parameters and non-linear objective functions but in terms of linear programs over the entropy region; this is done in Shannon’s original paper [Sha48, Theorem 10] and has since then become a standard technique. In this case, the optimum is attained on the boundary of \(\overline{\mathbf{H}}_n^*\). Even if it can be located, it is not clear whether the optimizer is entropic and hence corresponds to a real probability distribution or if it can only be approximated arbitrarily well by distributions.

The knowledge of which CI structures are representable can be viewed as combinatorial information about the intricate boundary structure of \(\mathbf{H}_4^*\). Namely, given a set of CI assumptions \(\mathcal{L}\) which define a subspace \(U = \{ h \in \mathbb{R}^{16} : \triangle(i, j|K) \cdot h = 0 \text{ for all } [i \perp \!\!\!\perp j | K] \in \mathcal{L} \}\), the question

\[
\triangle(I, J|K) \cdot h := h(IK) + h(JK) - h(JJK) - h(K),
\]
is which other inequalities $\triangle \geq 0$ are tight at every point in $H_n^* \cap U$? Calling the set of implied statements $M$, this proves a conditional independence inference rule $L \Rightarrow M$ for representable CI structures. Unlike the geometric shape of $H_1^*$, this combinatorial, CI-theoretic information about its boundary is completely available due to the series of papers by Matúš and Studený. Studený’s recent paper [Stu21] revisits this series and shows that all inference properties for four discrete random variables can be deduced from conditional Ingleton inequalities in addition to the common Shannon information inequalities. Each of the ten conditional Ingleton inequalities presented in [Stu21] is necessary to obtain all the CI inference rules. In this paper we prove that there are no further, in the CI-theoretic sense “extraneous”, conditional Ingleton inequalities.

2.3. Masks and conditional Ingleton inequalities. One way to obtain conditional Ingleton inequalities is to rewrite the functional $\Box(X|Y|ZU)$ as a linear combination of difference expressions $\triangle(i,j|K)$ in the dual space $(\mathbb{R}^{16n})^*$. Some of these masks of the Ingleton expression were found in [MS95] and are also discussed in [Stu21, Section II.G]:

\begin{align*}
(M.1) & \quad \Box(X|Y|ZU) = \triangle(Z, U|X) + \triangle(Z, U|Y) + \triangle(X, Y) - \triangle(Z, U) \\
(M.2) & \quad = \triangle(Z, U|Y) + \triangle(X, Z|U) + \triangle(X, Y) - \triangle(X, Z) \\
(M.3) & \quad = \triangle(X, Y|Z) + \triangle(X, Z|U) + \triangle(Z, U|Y) - \triangle(X, Z|Y) \\
(M.4) & \quad = \triangle(X, Y|Z) + \triangle(X, Y|U) + \triangle(Z, U|XY) - \triangle(X, Y|ZU) \\
(M.5) & \quad = \triangle(X, Y|Z) + \triangle(X, Z|U) + \triangle(Z, U|Y) - \triangle(X, Z|YU).
\end{align*}

These masks prove (1.1)–(1.5); indeed mask (M.1), for example, implies (1.1): $[Z \perp U] \Rightarrow \Box(X|Y|ZU) \geq 0$ due to the non-negativity of all difference expressions. Under the symmetries $X \leftrightarrow Y$ and $Z \leftrightarrow U$ which fix $\Box(X|Y|ZU)$, these five masks generate fourteen distinct conditional Ingleton inequalities, displayed below in groups by symmetry class:

\begin{align*}
\big([X \perp Z], \ [Y \perp Z], \ [X \perp U], \ [Y \perp U]\big), \\
\big([X \perp Z \mid Y], \ [Y \perp Z \mid X], \ [X \perp U \mid Y], \ [Y \perp U \mid X]\big), \\
\big([X \perp Y \mid ZU]\big), \\
\big([X \perp Z \mid YU], \ [Y \perp Z \mid XU], \ [X \perp U \mid YZ], \ [Y \perp U \mid XZ]\big).
\end{align*}

In [Stu21, Section IV] five further conditional Ingleton inequalities are proved which require two CI assumptions. They expand to fourteen conditional inequalities under symmetry as well. Studený has ruled out five other sets of CI assumptions by counterexamples and reduced the possibilities for an eleventh conditional Ingleton inequality to three CI structures, namely the sets strictly above $L_0 = [X \perp Z \mid U] \land [Y \perp U \mid Z]$ and below $L = [X \perp Z \mid U] \land [Y \perp U \mid Z] \land [X \perp Y] \land [Z \perp U \mid XY]$.

The verification of this claim by hand is tedious. The process can be delegated to a SAT solver such as CaDiCaL [CDCL] as follows. There are 24 elementary CI statements $[i \perp j \mid K]$ on four random variables; introduce one boolean variable for each of them. If a CI structure implies the Ingleton inequality, then so does every superset. If a counterexample exists for a set of CI assumptions, then every subset is ruled out by the same counterexample. Using the ten known conditional Ingleton inequalities, Studený’s five counterexamples and the conjectured minimal and maximal unsolved cases $L_0$ and $L$ — and all their symmetric variants —, a boolean formula can be constructed whose satisfying assignments are all CI structures which are not covered and are potential assumptions for an eleventh conditional Ingleton inequality. The solver quickly decides that the formula is unsatisfiable and hence proves that all unsolved cases are between $L_0$ and $L$. More details and source code for this computation are available on our MathRepo page.

The objective of the next section is to construct a probability distribution satisfying $L$ and violating the Ingleton inequality. Known examples of this kind are usually hand-crafted, rational distributions with small denominators derived by careful exploitation of zero patterns and symmetries; cf. [KR13, Stu21]. We present a different, computer-assisted and heuristic methodology to find counterexamples in information theory rooted in algebra and relying on symbolic computations as well as numerical non-linear optimization.
3. Construction of the distribution

3.1. Circuits, masks and scores. The 24 difference expressions $\Delta(i, j|K)$ and the Ingleton expression $\square(XY|ZU)$ are elements in the dual space $\mathbb{R}^{16^*}$. Choosing the standard basis there, they can be identified with vectors which make up the columns of a $16 \times 25$ matrix. The circuits of this matrix, i.e., the non-zero integer vectors in its kernel with inclusion-minimal support and coprime non-zero entries, can be computed using the software dt12 [dt12]; cf. [Stu96, Chapter 4]. There are 10,814 such circuits and among them 6,814 which give a non-zero coefficient to $\square(XY|ZU)$. These circuits are the shortest possible ways of writing $\square$ as a linear combination of $\triangle$. The 14 shortest circuits require only four $\triangle$ terms and among them they are precisely the 14 symmetric images of (M.1)–(M.5). All masks are available on our website.

Based on the circuits, we obtain short masks which are closely related to the two subcases $\mathcal{L}_1 = \{Z \perp U \mid XY\} \land \mathcal{L}_0$ and $\mathcal{L}_2 = \{X \perp Y\} \land \mathcal{L}_0$ of the model $\mathcal{L} = \mathcal{L}_1 \land \mathcal{L}_2$. All three cases remained open in Studený’s analysis, but $\mathcal{L}_0$ was settled in [Stu21, Example 5]. The mask

\[
\square(XY|ZU) = \Delta(X, Y|ZU) + \Delta(X, Z|UY) + \Delta(Y, U|Z) - \Delta(Y, U|XZ) + \Delta(Z, U|XY),
\]

(\textit{f.1})

\textit{can be confirmed by plugging in the definitions of $\triangle$ and $\square$. It was selected to simplify as much as possible under the CI assumptions $\mathcal{L}_1$ which would otherwise contribute positive quantities to the Ingleton expression. Given that $\mathcal{L}_1$ holds, the mask (\textit{f.1}) yields}

\[
-\square(XY|ZU) = \Delta(X, Z|YU) + \Delta(Y, U|XZ) - \Delta(X, Y|ZU)
\]

\[
= H(Y|XZ) + H(Y|XU) - H(X|ZU) =: \varrho_1(X, Y, Z, U).
\]

Analogously one proves

\[
\square(XY|ZU) = \Delta(X, Y) - \Delta(X, Z) + \Delta(X, Z|U) - \Delta(Y, U) + \Delta(Y, U|Z) + \Delta(Z, U),
\]

(\textit{f.2})

\textit{which under $\mathcal{L}_2$ yields}

\[
-\square(XY|ZU) = \Delta(X, Z) + \Delta(Y, U) - \Delta(Z, U)
\]

\[
= H(ZU) - H(Z|X) - H(U|Y) =: \varrho_2(X, Y, Z, U).
\]

The functions $\varrho_1$ and $\varrho_2$ are referred to as the non-Ingleton scores on $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively. On the distributions satisfying the respective CI statements, they equal the value of $-\square(XY|ZU)$ but they involve fewer terms and are thus easier to evaluate and to differentiate. Both scores coincide on the intersection $\mathcal{L}$ of the models $\mathcal{L}_1$ and $\mathcal{L}_2$.

We continue with a geometric analysis of the space of binary distributions in the model $\mathcal{L}_1$ and extend these findings to derive a binary distribution for $\mathcal{L}$ with positive non-Ingleton score.

3.2. Parametrization of $\mathcal{L}_1$. A joint distribution of four binary random variables is given by a $2 \times 2 \times 2 \times 2$ tensor with real, non-negative entries $p_{ijkl}$ which sum to one. With all four indices ranging in $\{0, 1\}$, these represent the atomic probabilities of the sixteen joint events. The CI statements of $\mathcal{L}_1$ prescribe quadratic equations on these probabilities:

\[
[Z \perp U \mid XY] \iff \begin{cases}
p_{0000} \cdot p_{0011} = p_{0001} \cdot p_{0010}, \\
p_{0100} \cdot p_{0111} = p_{0101} \cdot p_{0110}, \\
p_{1000} \cdot p_{1011} = p_{1001} \cdot p_{1010}, \\
p_{1100} \cdot p_{1111} = p_{1101} \cdot p_{1110}.
\end{cases}
\]

\[
[X \perp Z \mid U] \iff \begin{cases}
(p_{0000} + p_{0100}) \cdot (p_{0101} + p_{1101}) = (p_{0001} + p_{0101}) \cdot (p_{0100} + p_{1100}), \\
(p_{0001} + p_{0101}) \cdot (p_{0111} + p_{1111}) = (p_{0011} + p_{0111}) \cdot (p_{0101} + p_{1101}),
\end{cases}
\]

\[
[Y \perp U \mid Z] \iff \begin{cases}
(p_{0000} + p_{0100}) \cdot (p_{0101} + p_{1101}) = (p_{0001} + p_{0101}) \cdot (p_{0100} + p_{1100}), \\
(p_{0010} + p_{0110}) \cdot (p_{0111} + p_{1111}) = (p_{0011} + p_{0111}) \cdot (p_{0100} + p_{1110}).
\end{cases}
\]
These equations are studied in algebraic statistics; see [Sul18, Proposition 4.1.6] for their derivation. It is in general difficult to derive a rational parametrization of a given CI model. To simplify this task, we impose the support pattern which already appears in [Stu21, Example 5]: suppose that $p_{0001} = p_{0010} = p_{0101} = p_{1001} = p_{1101} = 0$ and all other variables are positive. From now on, we regard only this linear slice of the CI models for $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}$.

Under these additional constraints, the above eight equations together with the condition that all probabilities sum to one can be resolved to yield the rational parametrization

$$p_{0100} = \frac{p_{0101}p_{0110}}{p_{0111}}, \quad p_{1000} = \frac{p_{1001}p_{1010}}{p_{1011}},$$

$$p_{0111} = \frac{p_{0101}(p_{0111} + p_{1111})}{p_{1011}(p_{1011} + p_{1111})}, \quad p_{0000} = \frac{p_{0101}(p_{0111} + p_{1111})}{p_{1011}(p_{1011} + p_{1111})}, \quad p_{1010} = \frac{p_{0110}p_{1011}}{p_{1011} + p_{1111}}, \quad p_{0101} = \frac{p_{0110}p_{1011}}{p_{1011} + p_{1111}},$$

$$p_{0001} = \frac{p^2_{1011}(1 - 2p_{0110} - 2p_{1011}) + p_{1011}p_{1111}(1 - p_{0110} - 3p_{1011} - p_{1111})}{(p_{0110} + p_{1011})(2p_{1011} + p_{1111})}.$$  

With six zero conditions and seven equations (two of the CI equations trivialize under the zero constraints), this leaves the three parameters $p_{0110}$, $p_{1011}$ and $p_{1111}$. The positivity conditions on the ten non-zero probabilities turn into non-linear inequalities and these are the only remaining constraints on the parameters. Thus, this defines a three-dimensional basic semialgebraic set $\mathcal{T}_1$.

3.3. Numerical optimization and a rational point. The Ingleton inequality is not an algebraic function of the parameters but a transcendental one. Hence, algebraic techniques like Gröbner bases or cylindrical algebraic decomposition cannot be directly applied to decide if there exist parameters on which $\square(XY|ZU)$ is negative. This question can be reformulated as whether a system of integer polynomial equations and inequalities in variables and exponentials of variables has a real solution. Thus, it is a question in the existential theory of the real numbers with exponentiation. The decidability of this theory is an open problem known as Tarski’s Exponential Function Problem and hence no general symbolic algorithms are available today to solve it; see [MW96] for a starting point on this topic.

Instead of symbolic techniques, we employ optimization. Mathematica’s FindMaximum function, when started on the values $(\frac{1}{16}, \frac{1}{16}, \frac{1}{16})$, numerically finds a local maximum of $q_1$ on $\mathcal{T}_1$ with value $0.0198$ at the parameters $p_{0110} = 0.36179$, $p_{1011} = 0.01463$ and $p_{1111} = 0.27455$. By continuity, $q_1$ remains positive in a small neighborhood of this point. Searching for a local minimum of $q_1$ in the range

$$(\square) \quad \frac{1}{6} \leq p_{0110} \leq \frac{3}{6}, \quad \frac{1}{160} \leq p_{1011} \leq \frac{3}{160}, \quad \frac{1}{8} \leq p_{1111} \leq \frac{3}{8}$$

yields a positive value, indicating that this region is likely to contain many points violating the Ingleton inequality. Based on this heuristic, we want to find a distribution in this range which satisfies the system $\mathcal{T}$ consisting of the inequalities of $\mathcal{T}_1$ and the additional CI equation for $[X \perp Y]$ which rewrites under the parametrization $(\ast)$ to

$$p_{1011}^2(1 + p_{1111})^3 + p_{1010}p_{1111}(2p_{1011} + p_{1111}^2 + p_{1011}p_{1111}^2 + 3p_{1111} + 4p_{1111}) + p_{0110}(p_{1011} + 3p_{1111} + p_{1111}^2 + 2p_{0111}p_{1111}^2 + 2p_{1011}p_{1111} + p_{0111}^2 + 2p_{1111}^2 + 8p_{1111}^3)$$

$$= p_{1111}(2p_{0111} + 3p_{1111}p_{1111} + p_{1111}^2)(p_{1011}^2 + p_{1011}p_{1111} + p_{1011}p_{1111}^2).$$
This equation can be resolved for \( p_{0110} = f(p_{0111}, p_{1111}) \) where \( f \) is a (lengthy) algebraic function involving rational functions of its arguments and a single square root. The system \( T \) together with the bounds (□) define a semialgebraic set and Mathematica’s \texttt{FindInstance} function quickly returns a solution typically with large denominators and an algebraic number of extension degree 2 over \( \mathbb{Q} \). This distribution proves \( L \neq □(XY | ZU) \geq 0 \). A rough map of where such counterexamples lie in the space \( T \) is given in Figure 1.

However, to confirm the Ingleton violation without numerical approximations, we seek a distribution with rational probabilities. The distribution is rational if \( p_{0110}, p_{0111}, p_{1111} \) can be chosen rational, which hinges on the square root in the algebraic function \( f \) determining \( p_{0110} \). The term under the square root, expressed in \( p_{1111} = a/b \) and \( p_{1111} = c/d \) with \( a, b, c, d \in \mathbb{N} \), reads

\[
\frac{1}{b^8d^{12}} \left( b^6c^{12} + 10ab^7c^{11}d - 2b^8c^{11}d + 41a^2b^6c^{10}d^2 - 16ab^7c^{10}d^2 + b^8c^{10}d^2 + 88a^3b^5c^9d^3 - 46a^2b^6c^9d^3 + 6ab^7c^9d^3 + 104a^4b^6c^8d^4 - 44a^3b^5c^7d^4 + 112a^2b^6c^7d^4 + 64a^5b^3c^7d^5 + 44a^4b^4c^7d^5 + 2a^3b^5c^7d^5 - 2a^2b^6c^7d^5 + 16ab^2c^6d^6 + 136a^5b^3c^5d^6 - 6a^4b^4c^5d^6 - 14a^3b^5c^5d^6 + 112a^2b^6c^5d^6 + 26a^5b^3c^5d^7 - 42a^4b^4c^5d^7 + 32a^7bc^4d^8 + 68a^6b^2c^4d^8 - 70a^5b^3c^4d^8 + 4b^4c^4d^8 + 56a^7bc^3d^9 - 68a^6b^2c^3d^9 + 4a^5b^3c^3d^9 + 16a^8c^2d^{10} - 36a bc^2d^{10} + 6a^6b^2c^2d^{10} - 8a^8 cd^{11} + 4a^7bcd^{11} + a^8d^{12} \right).
\]

The denominator is always a square, so it suffices to find, in accordance with (□), four positive integers \( b \leq 160a \leq 3b \) and \( d \leq 8c \leq 3d \) which make the parenthesized numerator into a square. An exhaustive search through small denominators \( b, d \) turns up \( p_{1111} = 2/99 \) and \( p_{1111} = 2/11 \) satisfying this criterion, because their value

\[
937 129 691 803 487 846 400 = 30 612 574 080
\]

is a perfect square. The resulting rational value \( p_{0110} = f(2/99, 2/11) = 10/693 \) does not satisfy (□) but it still yields a positive non-Ingleton score. To see this, consider the score \( \varrho_2 \) of the distribution with the given parameters, write all fractions with their common denominator 693 and assemble all terms under one

\[
(\exp \varrho_2)^{693} = \frac{24^{24} \cdot 30^{30} \cdot 141^{141} \cdot 168^{168} \cdot 201^{201} \cdot 228^{228} \cdot 294^{294} \cdot 300^{300} \cdot 693^{693}}{11^{11} \cdot 154^{154} \cdot 198^{198} \cdot 220^{220} \cdot 252^{252} \cdot 308^{308} \cdot 441^{441} \cdot 495^{495}}
\]

the violation of the Ingleton inequality is just a matter of comparing the integers in the numerator and denominator — a standard task which every computer algebra system with exact arithmetic on big integers will perform. The former is approximately \( 219.148 \times 10^{5190} \) and the latter \( 1.14751 \times 10^{5190} \). Thus, the fraction is greater than one and the non-Ingleton score is positive. Numerically, the score and hence the negative of the Ingleton expression \( □(XY | ZU) \) is approximately 0.00757. The distribution in its entirety is given in the beginning of this note.

4. Classification of essentially conditional Ingleton inequalities

4.1. Essential conditionality. The second part of our theorem concerns essential conditionality, a notion introduced in [KR13]. Given a conditional information inequality \( L \Rightarrow □(XY | ZU) \geq 0 \) one may ask if it arises from a valid unconditional information inequality of the form

\[
□(XY | ZU) + \sum_{|i \perp j| \in \mathcal{L}} \lambda_{|i \perp j|}[K] \Delta(i, j|K) \geq 0,
\]

with Lagrange multipliers \( \lambda_{|i \perp j|}[K] \geq 0 \). The existence of multipliers which make (□) a valid information inequality constitutes an “unconditional” proof of the conditional inequality \( L \Rightarrow □(XY | ZU) \geq 0 \); otherwise this inequality is essentially conditional. The masks (M.1)–(M.5) show that the conditional Ingleton inequalities (1.1)–(1.5) are in fact not essentially conditional. Among the first examples of essentially conditional inequalities due to Kaced and Romashchenko [KR13] are the conditional Ingleton inequalities (2.1)–(2.4). Hence, the only remaining case in the classification of essential conditionality for conditional Ingleton inequalities is the inequality (2.5) which was recently discovered by Studený [Stu21].
A curve of this type converges to a distribution which is uniform on its support. It is well-known that every invalid information inequality can be refuted by such a distribution — however, this result requires unbounded state spaces. The typical argument in [KR13] expands the terms in \((\square_\lambda)\) as power series in \(\varepsilon\) around zero and compares convergence orders to conclude that a small enough value of \(\varepsilon\) leads to a violation of the inequality.

4.2. Sampling for a counterexample. If \(\lambda\) is a tuple of Lagrange multipliers that makes \((\square_\lambda)\) true and \(\mu \geq \lambda\) componentwise, then \(\mu\) also makes \((\square_\lambda)\) true since the \(\triangle\) functionals are non-negative on the entropy region. Hence there is no loss of generality in assuming that all multipliers are equal and arbitrarily large but fixed. To prove essential conditionality we construct counterexamples to \((\square_\lambda)\) depending continuously on \(\lambda \to \infty\), i.e., a curve of counterexamples. The curves proving essential conditionalities in [KR13] all follow a simple combinatorial recipe:

1. Commit to state space sizes for all four random variables; usually they are all assumed to be binary. This gives rise to 16 real parameters \(P = \{p_{\lambda} \mid \lambda \in \{0000, \ldots, 1111\}\}\).

2. Choose a partition of \(P\) into four subsets \(A, B, C, D\) and assign the probabilities

\[
\begin{array}{cccc}
 p \in A & p \in B & p \in C & p \in D \\
 \frac{1}{|A| + |B|} & \frac{1}{|A| + |B|} - \varepsilon & \frac{|B|}{|C|} \varepsilon & 0
\end{array}
\]

with a real, positive parameter \(\varepsilon \to 0\). To ensure that the result is a probability distribution we require \(|A \cup B| > 0\) and \(|C| > 0\).

A curve of this type converges to a distribution which is uniform on its support. It is well-known [Cha01] that every invalid information inequality can be refuted by such a distribution — however, this result requires unbounded state spaces. The typical argument in [KR13] expands the terms in \((\square_\lambda)\) as power series in \(\varepsilon\) around zero and compares convergence orders to conclude that a small enough value of \(\varepsilon\) leads to a violation of the inequality.

Sampling distributions according to the above algorithm and using criteria based on the limit behavior of the power series coefficients obtained via Mathematica’s Series function eventually turns up the following sparse proof of essential conditionality for (2.5):

\[
p_{0000} = 0, \quad p_{0001} = 0, \quad p_{0010} = \frac{1}{5} - \varepsilon, \quad p_{0011} = 0,
\]

\[
p_{0100} = 0, \quad p_{0101} = 0, \quad p_{0110} = \frac{1}{5}, \quad p_{0111} = 0,
\]

\[
p_{1000} = 0, \quad p_{1001} = 0, \quad p_{1010} = \frac{1}{5}, \quad p_{1011} = 0,
\]

\[
p_{1100} = \varepsilon, \quad p_{1101} = \frac{1}{5}, \quad p_{1110} = 0, \quad p_{1111} = \frac{1}{5}.
\]

The CI assumptions of (2.5) are only satisfied in the limit \(\varepsilon = 0\) since

\[
\triangle(X, Z|U) = \triangle(Y, Z|U) = \frac{1}{5} \log \left( \frac{27}{(3 - 5\varepsilon)^5 \cdot (1 + 5\varepsilon)} \right) = \log(3) \varepsilon - \frac{10}{3} \varepsilon^2 + \frac{100}{27} \varepsilon^3 + \mathcal{O}(\varepsilon^4).
\]

This makes it possible to violate the Ingleton inequality, and indeed:

\[
\square(XY|ZU) = \log \left( \frac{\sqrt{27}}{8000} \cdot \frac{(1/5 - \varepsilon)^{1/5 - \varepsilon} \cdot (4/5 - \varepsilon)^{4/5 - \varepsilon} \cdot (2/5 + \varepsilon)^{2/5 + \varepsilon} \cdot \varepsilon^\varepsilon}{(2/5 - \varepsilon)^{2(2/5 - \varepsilon)} \cdot (3/5 - \varepsilon)^{3/5 - \varepsilon} \cdot (1/5 + \varepsilon)^{3(1/5 + \varepsilon)}} \right)
\]

\[= (\log(30\varepsilon) - 1) \varepsilon - \frac{155}{25} \varepsilon^2 + \frac{11525}{864} \varepsilon^3 + \mathcal{O}(\varepsilon^4).\]

The expression \((\square_\lambda)\) in our case is

\[
\square(XY|ZU) + \lambda(\triangle(X, Z|U) + \triangle(Y, Z|U)) = (-1 + 2\lambda \log(3) + \log(30\varepsilon)) \varepsilon + \mathcal{O}(\varepsilon^2)
\]

whose \(\varepsilon\)-order coefficient tends to \(-\infty\) as \(\varepsilon \to 0\) for any fixed \(\lambda\). Hence, every unconditional version of (2.5) can be violated on our curve of distributions, which proves essential conditionality.
5. Remarks

(1) The distribution constructed in Section 3 satisfies the four CI statements in \( \mathcal{L} \) and none other. This can be checked computationally but it also follows from Section 2.3 since every superset of \( \mathcal{L} \) implies the Ingleton inequality.

(2) The entropy vector of that distribution is a conic combination of twelve extreme rays of \( H_4 \) (corresponding to the twelve coatoms in the lattice of semimatroids above \( \mathcal{L} \); cf. [MS95]). The only ray which violates the Ingleton inequality is not entropic. Thus, our construction gives an *entropic* conic combination of these not necessarily entropic polymatroids where the non-Ingleton component has sufficiently high weight.

(3) All counterexamples to potential conditional Ingleton inequalities with inclusion-minimal assumptions [Stu21, Section IV.B] as well as all proofs of essential conditionality [KR13, Section IV.A] require only rational binary distributions. This is remarkable insofar as there exist CI inference rules which are valid for binary random vectors but not in general; see [Mat18]. Whether every wrong CI inference rule can be refuted by a rational distribution is equivalent to [Mat99, Conjecture] and still open.

(4) The method of [Mat18] to construct binary distributions with prescribed CI structure using the Fourier–Stieltjes transform even produces distributions close to the uniform distribution. This allows one to concentrate on satisfying the CI equations only, because every binary tensor close to the uniform distribution has strictly positive entries and thus yields a positive probability distribution after multiplying all entries by a normalizing constant. Matúš’s parametrization of the model \( \mathcal{L}_2 \) depends on a solution to the associated *solvability system* whose components appear as exponents of the parameters. The smallest integral solution to the solvability system is \((x_{12}, x_{13}, x_{14}, x_{23}, x_{24}, x_{34}) = (1, 2, 1, 1, 2, 1)\); see [Mat18, Theorem 1] for details. In the nomenclature of this theorem (and its proof), the non-Ingleton score is then given by

\[
(\gamma^2 + 1) \log(\gamma^2 + 1) + \frac{1}{2} (\gamma - 1) \log(\gamma - 1) - (\gamma^2 - 1) \log(\gamma^2 - 1) - \frac{1}{2} (\gamma + 1) \log(\gamma + 1)
\]

for \( \gamma \) small but positive. This function in \( \gamma \) has one root in the interval \((0, 1)\) where it passes from negative on the left to positive values on the right. The root has the approximate value of 0.72766. Using cylindrical algebraic decomposition in *Mathematica*, it can be verified that Matúš’s construction does not produce tensors with non-negative entries (ergo probability distributions) if \( \gamma > 0.727 \) is imposed. It remains open whether there exist counterexamples to the validity of the Ingleton inequality subject to \( \mathcal{L} \) and arbitrarily close to uniform or even just without zero entries.

(5) The same method applies to the search for a proof of essential conditionality in Section 4 because the CI assumptions \([X \perp \perp Z \mid U] \land [Y \perp \perp Z \mid U]\) have conditioning sets of size one. Moreover, this statistical model has a rational parametrization: its conditionals with respect to \( U \) belong to the *marginal independence model* \([X \perp \perp Z] \land [Y \perp \perp Z]\) which has a monomial parametrization in Möbius coordinates by [BPS22]. Lastly, the entropy vectors arising from those distributions in the marginal independence model which have no private information have been completely characterized by [Mat06]. The random search carried out in Section 4 found a counterexample more quickly than any of these approaches.

(6) Combinatorial and group-theoretic constructions of distributions with large violations of the Ingleton inequality have been investigated in [BN12] in the context of the four-atom conjecture, which was then refuted in [MC16].

(7) The last part of Open Question 2 in [Stu21] concerns validity of (2.1)–(2.5) for almost-entropic points. As mentioned in Remark 4.1 some cases are settled in [KR13] with different answers. The status of (2.2) and of (2.5) is open.

Acknowledgements. I thank the anonymous referee for hints for improving the presentation. I would also like to thank Mima Stanojkovski and Rosa Winter for their immediate interest, code samples and an inspiring discussion about finding rational points on varieties — even though the brute force approach turned out to succeed more quickly this time.
References

[BN12] Nigel Boston and Ting-Ting Nan: Large violations of the Ingleton inequality. In 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 1588–1593 (2012). DOI: 10.1109/Allerton.2012.6483410.

[BPS22] Tobias Boege, Sonja Petrović and Bernd Sturmfels: Marginal independence models. In Proceedings of the 2022 International Symposium on Symbolic and Algebraic Computation, ISSAC ’22, pp. 263–271. Association for Computing Machinery (ACM) (2022). DOI: 10.1145/3476446.3536193.

[Cha01] Terence H. Chan: A combinatorial approach to information inequalities. Commun. Inf. Syst., 1(3):241–253 (2001). DOI: 10.4310/CIS.2001.v1.n3.a1.

[Fuj78] Satoru Fujishige: Polymatroidal dependence structure of a set of random variables. Inf. Control, 39(1):55–72 (1978). DOI: 10.1016/0019-9958(78)91063-X.

[GMM17] Arley Gómez, Carolina Mejía and Juan Andrés Montoya: Defining the almost-entropic regions by algebraic inequalities. Int. J. Inf. Coding Theory, 4(1):1–18 (2017). DOI: 10.1504/IJICT.2017.081456.

[Ing71] Aubrey W. Ingleton: Representation of matroids. In Dominic J. A. Welsh, ed.: Combinatorial Mathematics and its Applications. Proceedings of a Conference held at the Mathematical Institute, Oxford, from 7–10 July, 1969, pp. 149–167 (1971).

[KR13] Tarik Kaced and Andrei Romashchenko: Conditional information inequalities for entropic and almost entropic points. IEEE Trans. Inf. Theory, 59(11):7149–7167 (2013). DOI: 10.1109/TIT.2013.2274614.

[Mat95] František Matúš: Conditional independences among four random variables. II. Combin. Probab. Comput., 4(4):407–417 (1995). DOI: 10.1017/S0963548300001747.

[Mat97] František Matúš: Conditional independence structures examined via minors. Ann. Math. Artif. Intell., 21(1):99–30 (1997). DOI: 10.1023/A:1018957117081.

[Mat99] František Matúš: Conditional independences among four random variables. III. Final conclusion. Combin. Probab. Comput., 8(3):269–276 (1999). DOI: 10.1017/S0963548399003740.

[Mat06] František Matúš: Piecewise linear conditional information inequality. IEEE Trans. Inf. Theory, 52(1):236–238 (2006). DOI: 10.1109/TIT.2005.860438.

[Mat07] František Matúš: Infinitely many information inequalities. In Proceedings of the IEEE ISIT 2007, pp. 41–44 (2007).

[Mat18] František Matúš: On patterns of conditional independences and covariance signs among binary variables. Acta Math. Hung., 154(2):511–524 (2018). DOI: 10.1007/s10474-018-0799-6.

[MC16] František Matúš and László Csirmaz: Entropy region and convolution. IEEE Trans. Inf. Theory, 62(11):6007–6018 (2016). DOI: 10.1109/TIT.2016.2601598.

[MS95] František Matúš and Milan Studený: Conditional independences among four random variables. I. Combin. Probab. Comput., 4(3):209–278 (1995). DOI: 10.1017/S0963548300001644.

[MW96] Angus J. Macintyre and Alex J. Wilkie: On the decidability of the real exponential field. In Piergiorgio Odifreddi, ed.: Kreiseliana: About and around Georg Kreisel, pp. 441–467. A. K. Peters (1996).

[Sha48] Claude E. Shannon: A mathematical theory of communication. Bell Syst. Tech. J., 27:379–423, 623–656 (1948). DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[Stu96] Bernd Sturmfels: Gröbner bases and convex polytopes, vol. 8. American Mathematical Society (AMS) (1996). DOI: 10.1090/ulect/008.

[Stu21] Milan Studený: Conditional independence structures over four discrete random variables revisited: conditional Ingleton inequalities. IEEE Trans. Inf. Theory, 67(11):7030–7049 (2021). DOI: 10.1109/TIT.2021.3104250.

[Sul18] Seth Sullivant: Algebraic Statistics, vol. 194 of Graduate Studies in Mathematics. American Mathematical Society (AMS) (2018). DOI: 10.1090/gsm/194.

[ZY97] Zhen Zhang and Raymond W. Yeung: A non-shannon-type conditional inequality of information quantities. IEEE Trans. Inf. Theory, 43(6):1982–1986 (1997). DOI: 10.1109/18.641561.

Mathematical software

[4ti2] 4ti2 team: 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at https://4ti2.github.io. Version 1.6.9.

[CDCL] Armin Biere: CaDiCaL at the SAT Race 2019. In Marijn Heule, Matti Järvisalo and Martin Suda, eds.: Proc. of SAT Race 2019 – Solver and Benchmark Descriptions, vol. B-2019-1 of Department of Computer Science Series of Publications B, pp. 8–9. University of Helsinki (2019). Available at https://github.com/arminbie/icalc. Version 1.3.1.

[M2] Daniel R. Grayson and Michael E. Stillman: Macaulay2, a software system for research in algebraic geometry. Available at https://www.math.uiuc.edu/Macaulay2/. Version 1.16.

[Nor] Winfried Bruns, Bogdan Ichim, Christof Söger and Ulrich von der Ohe: Normaliz. Algorithms for rational cones and affine monoids. Available at https://www.normaliz.uni-osnabrueck.de. Version 3.8.4.

[WM] Wolfram Research, Inc.: Mathematica. Champaign, IL, 2018. Version 11.3.

Tobias Boege, Department of Mathematics, KTH Royal Institute of Technology, Sweden

Email address: post@taboege.de