Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Properties of Unique Information

by

Johannes Rauh and Maik Schünemann

Preprint no.: 4 2020

PROPERTIES OF UNIQUE INFORMATION

JOHANNES RAUH AND MAIK SCHÜNEMANN

Abstract. We study the measure of unique information $UI(T : X \setminus Y)$ defined by Bertschinger et al. [2014] within the framework of information decompositions. We study uniqueness and support of the solutions to the optimization problem underlying the definition of UI. We give necessary conditions for non-uniqueness of solutions with full support in terms of the cardinalities of T, X and Y and in terms of conditional independence constraints. Our results help to speed up the computation of $UI(T : X \setminus Y)$, most notably in the case where T is binary. In the case that all variables are binary, we obtain a complete picture where the optimizing probability distributions lie.

Contents

1. Introduction 1
Summary of results and outline 2
2. The optimization domain Δ_P 3
3. Support and uniqueness of the optimum 5
4. The case of binary T 8
4.1. Independence properties for optimizers in the interior 8
4.2. The case of restricted support 10
4.3. Statistics for uniqueness and support of optimizers for binary T 11
5. The all binary case 12
6. Examples 14
References 15

1. Introduction

Bertschinger et al. [2014] introduced an information measure $UI(T : X \setminus Y)$ which they called *unique information*. The function UI is proposed within the framework of information decompositions [Williams and Beer, 2010] to quantify the amount of information about T that is contained in X but not in Y. Similar quantities within this framework have been proposed by Harder et al. [2013], Ince [2017], James et al. [2018] and Niu and Quinn [2019]. Among them, the quantity UI is characterized that it is the only one with a full axiomatic characterization. Although it has received a lot of attention by theorists [see e.g. Rauh et al., 2014], so far, applications have focused on other measures, because UI is difficult to compute, although there has been recent progress [Banerjee et al., 2018].

Date: January 1, 2020.
2010 *Mathematics Subject Classification.* 94A15; 94A17.
Key words and phrases. Information decomposition, unique information.

The function UI is defined by means of an optimization problem. Let T, X, Y be random variables with finite state spaces $\mathcal{T}, \mathcal{X}, \mathcal{Y}$ and with a joint distribution P. Let $\Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$ be the set of all joint distributions of such random variables, and let

$$\Delta_P = \Big\{ Q \in \Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}} : Q(X = x, T = t) = P(X = x, T = t), \\ Q(Y = y, T = t) = P(Y = y, T = t) \text{ for all } x \in \mathcal{X}, y \in \mathcal{Y}, t \in \mathcal{T} \Big\}$$

be the set of all joint distributions that have the same pair marginals as P for the pairs (X, T) and (Y, T). Then

$$UI(T : X \setminus Y) = \min_{Q \in \Delta_P} I_Q(T : X|Y), \tag{1}$$

where $I_Q(T : X|Y)$ denotes the conditional mutual information of T and X given Y, computed with respect to Q. Due to the invariances in Δ_P, the optimization problem in (1) can be reformulated as follows:

$$\min_{Q \in \Delta_P} I_Q(T : X|Y) = H(T|Y) - \max_{Q \in \Delta_P} H(T|X, Y). \tag{2}$$

This paper studies UI, focusing on the following two questions:

(1) When is there a unique solution to the optimization problems in (2)?
(2) When is there a solution in the relative interior of Δ_P?

In the framework of information decomposition, the solutions to the optimization problems (2) are distributions with "zero synergy about T." Thus, understanding these solutions sheds light on the concept of synergy. If the solution is unique, there is a unique way to combine the random variables X and Y without synergy about T that preserves the (X, T)- and (Y, T)-marginals. If the solution is not unique, there are many different such possibilities. Moreover, a unique solution Q^* might be used to "localize" the information decomposition, in the sense of Finn and Lizier [2018] (although there might be conceptual problems, because the support of Q^* might not satisfy $\operatorname{supp}(Q^*) \supseteq \operatorname{supp}(P)$).

A better understanding of the optimization problems also helps in the computation of UI. For example, our results allow to compute UI in constant time in the case that all random variables are binary, by explicitly solving the optimization problem. In the case where $\mathcal{T}$ is binary, an optimum in the interior of Δ_P can be found by solving a linear programming problem.

Summary of results and outline. Section 2 describes the optimization domain Δ_P and its support in dependence of P.

Section 3 summarizes general facts about the optimization problem that hold for arbitrary $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{T}$. In particular, it is shown that under some conditions on the support of Δ_P, if both conditional independence statements $T \perp\!\!\!\perp_P X$ and $T \perp\!\!\!\perp_P Y$ hold, it follows that the optimum is not unique.

Section 4 specializes to the case where T is binary. In this case, if there is an optimizer in the interior, then this optimizer satisfies a conditional independence constraint. In general, the optimizer is not unique. We analyze how often the optimum lies in the interior or at the boundary of Δ_P and how often an optimum in the interior is unique as a function of the cardinalities of $\mathcal{X}, \mathcal{Y}$ when sampling P uniformely from $\Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$.

Section 5 gives a complete picture for the case where all variables are binary. A closed form expression is given for optimizers that lie in the interior of Δ_P. If the optimizer does not lie in the interior, the optimum is attained at an extremal point of Δ_P.

Finally, Section 6 collects examples that demonstrates that the conditions of some of our results are indeed necessary.

2. The optimization domain Δ_P

Fix a joint distribution $P \in \Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$. Since the marginal of T is constant on Δ_P, the support of T, which we denote by $\mathcal{T}' := \{t \in \mathcal{T} : P(T = t) > 0\}$, is also constant on Δ_P.

Any distribution $Q \in \Delta_P$ is characterized uniquely by the conditional probabilities $Q(X, Y|T = t)$ for $t \in \mathcal{T}'$. The map

$$P \in \Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}} \mapsto \big(P(X, Y|T = t)\big)_{t\in\mathcal{T}'}$$

(where $\mathcal{T}'$ depends on P) induces a linear bijection

$$\Delta_P = \bigtimes_{t\in\mathcal{T}'} \Delta_{P,t},$$

where

$$\Delta_{P,t} = \Big\{Q \in \Delta_{\mathcal{X},\mathcal{Y}} : Q(X = x) = P(X = x|T = t), \\ Q(Y = y) = P(Y = y|T = t)\Big\},$$

and $\Delta_{\mathcal{X},\mathcal{Y}}$ is the set of all probability distributions of random variables X, Y with finite state spaces $\mathcal{X}, \mathcal{Y}$. For example, when X and Y are binary, $\Delta_{P,t}$ is a line segment (which may degenerate to a point) for all $t \in \mathcal{T}'$. Thus, Δ_P is a product of line segments; that is, a hypercube (up to a scaling). If T is also binary, then Δ_P is a rectangle (a product of two line segments), which may degenerate to a line segment or even a point depending on the support of P.

In the following, for $Q \in \Delta_P$ and $t \in \mathcal{T}'$, we write $Q_t := Q(X, Y|T = t)$ for the conditional distribution of X, Y given that $T = t$. The product structure of Δ_P implies: if $Q \in \Delta_P$ lies on the boundary of Δ_P, then at least one of the Q_t lies on the boundary of $\Delta_{P,t}$. Moreover, Q lies on the boundary of $\Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$. Hence, the boundaries of the polytopes Δ_P or $\Delta_{P,t}$ are characterized by the vanishing of coordinates.

Remark 2.1. In the following, the expression *boundary of* Δ_P refers to the relative boundary. If P lies on the boundary of $\Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$, then Δ_P may be a subset of the boundary of $\Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$. This happens if and only if one coordinate vanishes throughout $\Delta_{P,t}$ (and thus one coordinate vanishes throughout Δ_P). In this case, Δ_P is part of the boundary of $\Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$. However, the (relative) boundary of Δ_P is a strict subset of Δ_P, and the same holds for $\Delta_{P,t}$.

Let A be the linear map that maps a joint distribution $P \in \Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$ to the pair $(P(X, T), P(Y, T))$ of marginal distributions. Then

$$\Delta_P = (P + \ker(A)) \cap \Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}.$$

The difference of any two elements of Δ_P belongs to $\ker(A)$. Conversely, the elements of $\ker(A)$ can be used to move within each Δ_P. A generating set of $\ker(A)$ is given by the vectors

$$\gamma_{t;x,x';y,y'} = \delta_{t,x,y} + \delta_{t,x',y'} - \delta_{t,x,y'} - \delta_{t,x',y}, \qquad x, x' \in \mathcal{X}, y, y' \in \mathcal{Y}, \tag{3}$$

where $\delta_{t,x,y}$ denotes the dirac measure supported at $T = t, X = x, Y = y$. These vectors are not linearly dependent. One way to choose a linearly independent subset is to fix $x_0 \in \mathcal{X}$, $y_0 \in \mathcal{Y}$. Then the set

$$\Gamma := \big\{\gamma_{t;x,x_0;y,y_0} : x \in \mathcal{X} \setminus \{x_0\}, y \in \mathcal{Y} \setminus \{y_0\}\big\}$$

is a basis of $\ker(A)$.

Remark 2.2. Apart from being symmetric, the larger dependent set has the following advantage, which is reminiscent of the Markov basis property [Diaconis and Sturmfels, 1998]: Any two points $Q, Q' \in \Delta_P$ can be connected by a path in Δ_P by applying a sequence of multiples of the elements $\gamma_{t;x,x';y,y'}$. The same is not true if we restrict x', y' to x_0, y_0: if $Q(X = x_0) = 0$, then adding a multiple of $\gamma_{t;x,x_0;y,y_0}$ for any $x \in \mathcal{X}$, $y \in \mathcal{Y}$ leads to a negative entry.

Let V be the set of distributions $Q_0 \in \Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$ that have a factorization of the form

$$Q_0(t, x, y) = Q_0(t)Q_0(x|t)Q_0(y|t).$$

Thus, V consists of all joint distributions that satisfy the Markov chain $X - T - Y$. For each $P \in \Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$, the intersection $\Delta_P \cap V$ contains precisely one element $Q_0 = Q_0(P)$; namely

$$Q_0(t, x, y) = P(t)P(x|t)P(y|t). \tag{4}$$

A general distribution $Q \in \Delta$ can thus be expressed uniquely in the form

$$Q = Q_0 + \sum_{t,x',y'} P(t)\gamma_{t,x',y'}\gamma_{t;x^0;x;y^0,y} \tag{5}$$

with $Q_0 = Q_0(Q) \in V$ and $\gamma = (\gamma_{t,x',y'})_{t,x'\neq x^0,y'\neq y^0}$ denoting the coefficients with respect to Γ.

Let $\operatorname{supp}(\Delta_P) := \bigcup_{Q\in\Delta_P} \operatorname{supp}(Q)$ be the largest support of an element of Δ_P. Generic elements of Δ_P have support $\operatorname{supp}(\Delta_P)$. We also let

$$\begin{aligned}\operatorname{supp}(\Delta_{P,t}) &:= \bigcup_{Q\in\Delta_P} \operatorname{supp}(Q_t) \\ &= \big\{(x, y) \in \mathcal{X} \times \mathcal{Y} : (t, x, y) \in \operatorname{supp}(\Delta_P)\big\} \text{ for } t \in \mathcal{T}'.\end{aligned}$$

If Δ_P is a singleton, then $P = Q_0$. In this case, $\operatorname{supp}(\Delta_P) = \operatorname{supp}(P)$, and $\operatorname{supp}(\Delta_{P,t}) = \operatorname{supp}(P_t)$. For $t \in \mathcal{T}'$ let $\mathcal{X}_t = \big\{x \in \mathcal{X} : P(X = x|T = t) > 0\big\}$ and $\mathcal{Y}_t = \big\{y \in \mathcal{Y} : P(Y = y|T = t) > 0\big\}$. It follows from the definitions:

Lemma 2.3. *Let $t \in \mathcal{T}'$. Then $\operatorname{supp}(\Delta_{P,t}) = \operatorname{supp}(Q_{0,t}) = \mathcal{X}_t \times \mathcal{Y}_t$. Moreover, $\operatorname{supp}(\Delta_P) = \operatorname{supp}(Q_0)$. Thus, Q_0 has maximal support in Δ_P.*

The next lemma follows from Lemma 2.3 and the definitions:

Lemma 2.4. *Let $t \in \mathcal{T}$, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. The following statements are equivalent:*

1. *Δ_P lies in the face of $\Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$ defined by $Q(t, x, y) = 0$.*
2. *$(t, x, y) \notin \operatorname{supp}(\Delta_P)$.*
3. *Every $Q \in \Delta_P$ satisfies $Q(t, x, y) = 0$.*
4. *$Q_0 := Q_0(P)$ satisfies $Q_0(t, x, y) = 0$.*
5. *$P(T = t, Y = y)P(T = t, X = x) = 0$.*

Lemma 2.5. *Let $t \in \mathcal{T}'$. The following are equivalent:*

(1) $\Delta_{P,t}$ *is a singleton.*
(2) *At least one of* $\mathcal{X}_t$, $\mathcal{Y}_t$ *is a singleton.*

Proof. Condition 2. in the lemma captures precisely when it is not possible to add a multiple of some $\gamma_{t;x,x';y,y'}$ to P or, in fact, to any $Q \in \Delta_P$ (cf. Remark 2.2). $\square$

3. Support and uniqueness of the optimum

This section studies the uniqueness of the optimizer and the question, when it lies on the boundary of Δ_P. There are many relations between uniqueness and support of the optimizers: Lemma 3.1 states that, if the optimizer is not unique, then there are optimizers with restricted support. Theorems 3.6, 3.7, 3.8 and 3.10 prove that either the optimizer lies at the boundary or it is not unique under a variety of different assumptions that involve the cardinalities of $|\mathcal{X}|$, $|\mathcal{Y}|$ and $|\mathcal{T}|$ or conditional independence conditions.

Lemma 3.1. *If the optimizer is not unique, then there exists an optimizer on the boundary of* Δ_P.

Proof. Suppose that there are two distinct optimizers $Q_1, Q_2 \in \Delta_P$. By convexity of the target function on Δ_P, the convex hull of Q_1 and Q_2 consists of optimizers. Since the target function is analytic, at least in the interior of Δ_P, any measure on the line through Q_1 and Q_2 is an optimizer. This line intersects the boundary of Δ_P. $\square$

The derivative of $I_Q(T : X|Y)$ in the direction of $\gamma_{t;x,x';y,y'}$ at Q equals

$$\log\left(\frac{Q(t,x,y)Q(t,x',y')}{Q(t,x,y')Q(t,x',y)}\cdot\frac{Q(x,y')Q(x',y)}{Q(x,y)Q(x',y')}\right) = \log\left(\frac{Q(t|x,y)Q(t|x',y')}{Q(t|x,y')Q(t|x',y)}\right), \tag{6}$$

assuming that the probabilities in the logarithm are positive. Otherwise, the partial derivative has to be computed as a limit.

Remark 3.2. The vanishing of the directional derivative of $I_Q(T : X|Y)$ can be seen as a determinantal condition: all derivatives (6) vanish if and only if for all $t \in \mathcal{T}'$ the determinants of all 2×2-submatrices of the matrix $(Q(t|x,y))_{x,y} \in \mathbb{R}^{\mathcal{X}\times\mathcal{Y}}$ vanish; that is, if and only if these matrices have rank one. As $\sum_{t\in\mathcal{T}'} Q(t|x,y) = 1$ for all x, y, the sum of these rank-one matrices is again of rank one.

Conversely, let $\tilde{Q}_1, \dots, \tilde{Q}_k$ be non-negative rank-one matrices such that the sum $\tilde{Q} = \tilde{Q}_1 + \dots + \tilde{Q}_k$ is non-zero and again of rank one; say $q = v^t w$ with v, w non-negative. Let $V = \mathrm{diag}(v)$, $W = \mathrm{diag}(w)$, and let $q_t = V^{-1} q_t W^{-1}$ for $t = 1, \dots, k$. Then $q_1 + \dots + q_k = V^{-1}\tilde{Q}W^{-1}$ is the matrix with all entries equal to one. Thus, the matrices q_t for $t = 1, \dots, k$ can be interpreted as matrices of conditional probabilities $q(t|X,Y)$. Together with any distribution of the pair (X, Y), one obtains a distribution $q(T, X, Y)$ at which all directional derivatives of $I_Q(T : X|Y)$ vanish.

Lemma 3.3. *Let* Q^* *be a minimizer of* $I_Q(T : X|Y)$ *for* $Q \in \Delta_P$, *and let* $(t, x, y) \in \mathrm{supp}(\Delta_P)$. *If* $Q^*(t, x, y) = 0$, *then* $Q^*(x, y) = 0$. *Thus,* $Q^*(t', x, y) = 0$ *for all* $t' \in \mathcal{T}$.

Proof. Suppose that $Q^*(t, x, y) = 0$, but that $Q^*(x, y) > 0$. Then there exist x', y' such that $Q_\epsilon := Q^* + \epsilon\gamma_{t;x,x';y,y'}$ is non-negative for $\epsilon > 0$ small enough (and thus $Q_\epsilon \in \Delta_P$). In particular, $Q^*(t, x', y), Q^*(t, x, y') > 0$.

Since Q^* is a minimizer, the partial derivative (6) at Q^* must be non-negative. Note that, by assumption, $Q^*(t,x,y) = 0$. If all four probabilities in the denominator of the fraction in the logarithm were non-zero, then the partial derivative would be equal to minus infinity. Thus, either $Q^*(x,y)$ or $Q^*(x',y')$ must vanish.

Suppose that $Q^*(x,y) > 0$. Then $Q^*(x',y') = 0$. Hence, $Q^*(t,x',y') = 0$, and so

$$\frac{Q_\epsilon(t,x,y)Q_\epsilon(t,x',y')}{Q_\epsilon(t,x,y')Q_\epsilon(t,x',y)} \cdot \frac{Q_\epsilon(x,y')Q\epsilon(x',y)}{Q_\epsilon(x,y)Q_\epsilon(x',y')} \\ = \frac{\epsilon^2 Q_\epsilon(x,y')Q_\epsilon(x',y)}{Q_\epsilon(t,x,y')Q_\epsilon(t,x',y)Q_\epsilon(x,y)\epsilon} = O(\epsilon).$$

Thus, the partial derivative diverges as $\log(\epsilon)$ to $-\infty$ as $\epsilon \to 0$, contradicting the fact that Q^* is a local minimizer. Therefore, $Q^*(x,y) = 0$. □

If $Q^*(t,x,y) = 0$ and $Q^*(t,x',y) > 0$, $Q^*(t,x,y') > 0$ for some $t \in \mathcal{T}'$, $x,x' \in \mathcal{X}$, $y,y' \in \mathcal{Y}$, then the partial derivative at Q^* in the direction of $\gamma_{t;x,x';y,y'}$ is

$$\log\left(\frac{Q^*(t,x',y')Q^*(x,y')Q^*(x',y)}{Q^*(t,x,y')Q^*(t,x',y)Q^*(x',y')}\right).$$

Therefore,

$$Q^*(t,x',y')Q^*(x,y')Q^*(x',y) \geq Q^*(t,x,y')Q^*(t,x',y)Q^*(x',y'),$$

or

$$\frac{Q^*(t,x',y')}{Q^*(x',y')} \geq \frac{Q^*(t,x,y')}{Q^*(x,y')}\frac{Q^*(t,x',y)}{Q^*(x',y)}.$$

It is wellknown that entropy is strictly concave and that conditional entropy is concave. From the proof of this fact, it is easy to analyze where conditional entropy is strictly concave.

Lemma 3.4. *The conditional entropy $H(A|B)$ is concave in the joint distribution of A,B. It is strictly concave, with the exception of those directions where $P(A|B)$ is constant. That is:*

$$\lambda H_{P_1}(A|B) + (1-\lambda)H_{P_2}(A|B) \leq H_{\lambda P_1 + (1-\lambda)P_2}(A|B)$$

with equality if and only if $P_1(A|B) = P_2(A|B)$ a.e.

Proof. Let θ be a Bernoulli random variable with parameter λ, and consider the joint distribution P of θ, A and B given by

$$P(A,B,\theta) = \begin{cases} \lambda P_1(A,B), & \text{if } \theta = 0, \\ (1-\lambda)P_2(A,B), & \text{if } \theta = 1. \end{cases}$$

Then

$$H_{\lambda P_1+(1-\lambda)P_2}(A|B) = H_P(A|B) \geq H_P(A|B,\theta) \\ = \lambda H_{P_1}(A|B) + (1-\lambda)H_{P_2}(A|B).$$

Equality holds if and only if A is independent of θ given B; that is:

$$P_1(A|B) = P(A|B,\theta=0) = P(A|B,\theta=1) = P_2(A|B).$$ □

Lemma 3.5. *Let $Q_1, Q_2 \in \Delta_P$ be two maximizers of $\max_{Q\in\Delta_P} H_Q(T|XY)$. Then $Q_1(T|XY) = Q_2(T|XY)$.*

Proof. We may assume that $Q_1 \neq Q_2$. By assumption, $H_Q(T|XY)$ is constant on the line segment between Q_1 and Q_2. Thus, on this line segment $H_Q(T|XY)$ is not strictly concave. By Lemma 3.4, $Q_1(T|XY) = Q_2(T|XY)$. □

Theorem 3.6. *Suppose that $|\mathcal{T}| < \max\{|\mathcal{X}|, |\mathcal{Y}|\}$. If there exists an optimizer of $\max_{Q\in\Delta_P} H_Q(T|XY)$ with full support, then the optimizer is not unique.*

Proof. Suppose that $Q^* \in \arg\max_{Q\in\Delta_P} H_Q(T|XY)$ has full support. The proof proceeds by finding a direction within Δ_P in which $H_Q(T|XY)$ is not strictly concave. Consider the linear equation

$$Q(t, x, y) = Q^*(t|x, y)Q(x, y) \quad \text{for } Q \in \Delta_P. \tag{7}$$

If $Q' \in \Delta_P$ solves this equation, then, by Lemma 3.4, the function $H_Q(T|X, Y)$ is affine on the line connecting Q^* and Q'. Since Q^* is a maximizer, $H_Q(T|X, Y)$ is constant on this line, whence any point on this line is a maximizer. Thus, to prove the theorem, it suffices to show that there exists a solution $Q' \neq Q*$ in Δ_P to (7).

By Remark 3.2, for every $t \in \mathcal{T}'$, there exists a pair of non-negative vectors v_t, w_t such that $Q^*(t|x, y) = v_t^t w_t$. The assumption $|\mathcal{T}| < \max\{|\mathcal{X}|, |\mathcal{Y}|\}$ implies that there exist vectors $v_0, w_0 \neq 0$ with $v_0^t v_t = 0 = w_0^t w_t$ for all $t \in \mathcal{T}'$. For $\epsilon \in \mathbb{R}$ let

$$Q_\epsilon(x, y) := Q^*(x, y) + \epsilon v_{0,x}^t w_{0,y}.$$

Then

$$\sum_{x\in\mathcal{X},y\in\mathcal{Y}} Q_\epsilon(x, y) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} Q^*(x, y) + \epsilon \sum_{x\in\mathcal{X},y\in\mathcal{Y}} v_{0,x}w_{0,y} = 1,$$

because

$$\begin{aligned}\sum_{x\in\mathcal{X},y\in\mathcal{Y}} v_{0,x}w_{0,y} = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} v_{0,x}w_{0,y} \sum_{t\in\mathcal{T}'} Q^*(t|x, y)&\\ = \sum_{t\in\mathcal{T}'} \sum_{x\in\mathcal{X}} v_{0,x}v_{t,x} \sum_{y\in\mathcal{Y}} w_{0,y}w_{t,y} = 0.&\end{aligned}$$

Therefore, if ϵ is sufficiently close to zero, then Q_ϵ defines a probability distribution for X and Y.

Extend Q_ϵ to a joint distribution of T, X, Y by $Q_\epsilon(t, x, y) = Q^*(t|x, y)Q_\epsilon(x, y)$. Then Q_ϵ satisfies (7). It remains to show that $Q_\epsilon \in \Delta_{Q^*}$. From

$$\begin{aligned}Q_\epsilon(t, x) - Q^*(t, x) &= \sum_{y\in\mathcal{Y}} \big(Q_\epsilon(t, x, y) - Q^*(t, x, y)\big)\\ &= \sum_{y\in\mathcal{Y}} Q^*(t|x, y)\big(Q_\epsilon(x, y) - Q^*(x, y)\big) = \epsilon v_{t,x}v_{0,x} \sum_{y\in\mathcal{Y}} w_{t,y}w_{0,y} = 0\end{aligned}$$

follows $Q_\epsilon(T, X) = Q^*(T, X)$. The equality $Q_\epsilon(T, Y) = Q^*(T, Y)$ follows similarly. □

Theorem 3.7. *Let $|\mathcal{T}| < |\mathcal{Y}|$, and suppose that $UI(T : X \setminus Y) = 0$. If there is an optimizer of $\max_{Q\in\Delta_P} H_Q(T|XY)$ with full support, then the optimizer is not unique.*

Proof. The proof of Theorem 3.6 can be adapted. Under the assumptions of the theorem, if Q^* is an optimizer, then $Q^*(t|x, y) = Q^*(t|x)$ does not depend on x. Therefore, one may choose $w_{t,y} = 1$ for all $y \in \mathcal{Y}, t \in \mathcal{T}$ and $v_{t,x} = Q^*(t|x)$.

To construct w_0, it now suffices that $|\mathcal{X}| \geq 2$, since all vectors w_t, $t \in \mathcal{T}$, are identical. $\square$

Theorem 3.8. *Suppose that $H(X), H(Y) > 0$. If $T \perp\!\!\!\perp_P X$ and $T \perp\!\!\!\perp_P Y$, then $\arg\max_{Q\in\Delta_P} H_Q(T|X,Y)$ is not unique.*

Proof. Let $Q_0 = Q_0(P) = P_T P_{X|T} P_{Y|T} = P_T P_X P_Y \in \Delta_P$. By construction, $T \perp\!\!\!\perp_{Q_0} (X,Y)$. Since $H_Q(T|X,Y) \leq H(T)$ for $Q \in \Delta_P$ and since Q_0 achieves equality, Q_0 maximizes $H(T|X,Y)$ on Δ_P.

Due to the assumption of positive entropy, there exist $x_0, x_1 \in \mathcal{X}$, $y_0, y_1 \in \mathcal{Y}$ with $P_X(x_0) > 0$, $P_X(x_1) > 0$, $P_Y(y_0) > 0$ and $P_Y(y_1) > 0$. For $\delta \in \mathbb{R}$ let

$$Q_\delta(t,x,y) := Q_0(t,x,y) + \delta p_T(t)\gamma_{t;x_0,x_1;y_0,y_1}.$$

If $|\delta|$ is small enough, then Q_δ is non-negative and hence belongs to Δ_P. For such δ, the conditional $Q_\delta(x,y|t)$ does not depend on t, whence $T \perp\!\!\!\perp_{Q_\delta} (X,Y)$. Thus, all such Q_δ are maximizers of $H_Q(T|X,Y)$ for $Q \in \Delta_P$. $\square$

Example 3.9. Let P be the distribution of three independent uniform binary random variables T, X, Y, and let P' be the joint distribution where X, T are uniform independent binary random variables and where $X = Y$. Then $\Delta_P = \Delta_{P'}$, and both P and P' maximize $H_Q(T|X,Y)$ for $Q \in \Delta_P$.

This example is the same as Example 31 by Bertschinger et al. [2014]. Ironically, Bertschinger et al. [2014] remarked that the optimization problem is ill-conditioned, but they failed to observe the non-uniqueness of the optimum in this case.

The following technical result generalizes Theorem 3.8. It is illustrated by Example 6.2.

Theorem 3.10. *Suppose that $T \perp\!\!\!\perp_P X \,\big|\, Y$ and $T \perp\!\!\!\perp_P Y \,\big|\, X$. If there exist $x_0 \in \mathcal{X}$, $y_0 \in \mathcal{Y}$ with $P(X = x_0, Y = y_0) > 0$ and $H(X|Y = y_0) \neq 0 \neq H(Y|X = x_0)$, then $\max_{Q\in\Delta_P} H_Q(T|X,Y)$ is not unique.*

Proof. If $T \perp\!\!\!\perp_P X \,\big|\, Y$, then $I_P(T : X|Y) = 0$. From this it follows that $P \in \arg\min_{Q\in\Delta_P} I_Q(T : X|Y)$. The probability distributions that satisfy $T \perp\!\!\!\perp_P X \,\big|\, Y$ and $T \perp\!\!\!\perp_P Y \,\big|\, X$ have first been characterized by Fink [2011]; see also the reformulation by Rauh and Ay [2014]. This characterization implies that there are partitions $\mathcal{X} = \mathcal{X}'_1 \cup \cdots \cup \mathcal{X}'_b$ and $\mathcal{Y} = \mathcal{Y}'_1 \cup \cdots \cup \mathcal{Y}'_b$ such that $\mathrm{supp}(P) \subseteq \mathcal{X}'_1 \times \mathcal{Y}'_1 \cup \cdots \cup \mathcal{X}'_b \times \mathcal{Y}'_b$ and such that $T \perp\!\!\!\perp_P \{X, Y\} \,\big|\, X \in \mathcal{X}'_i, Y \in \mathcal{Y}'_i$ for $i = 1, \ldots, b$. There exists $i_0 \in \{1, \ldots, b\}$ such that $x_0 \in \mathcal{X}'_{i_0}$ and $y_0 \in \mathcal{Y}'_{i_0}$. Since $H(X|Y = y_0) \neq 0 \neq H(Y|X = x_0)$, there exist $x_1 \in \mathcal{X}'_{i_0} \setminus \{x_0\}$ and $y_1 \in \mathcal{Y}'_{i_0} \setminus \{y_0\}$ with $P(x_1, y_0) > 0$ and $P(x_0, y_1) > 0$. For $\delta > 0$ let

$$P_\delta = P + \delta \cdot P(T|X,Y)\gamma_{t;x_0,x_1;y_0;y_1}.$$

If δ is positive and small enough, then P_δ is a probability distribution in Δ_P with $\mathrm{supp}(P) = \mathrm{supp}(P_\delta)$. Moreover, $T \perp\!\!\!\perp_{P_\delta} \{X, Y\} \,\big|\, X \in \mathcal{X}'_i, Y \in \mathcal{Y}'_i$ for $i = 1, \ldots, b$. Hence, $T \perp\!\!\!\perp_{P_\delta} X \,\big|\, Y$ and $T \perp\!\!\!\perp_{P_\delta} Y \,\big|\, X$, and so $P_\delta \in \arg\min_{Q\in\Delta_P} I_Q(T : X|Y)$. $\square$

4. The case of binary T

4.1. **Independence properties for optimizers in the interior.**

If $T \perp\!\!\!\perp_P X \,\big|\, Y$ or $T \perp\!\!\!\perp_P Y \,\big|\, X$, then P solves the PID optimization problem (2). The next theorem is a partial converse in the case of binary T:

Theorem 4.1. *Let T be binary. Assume that Δ_P has full support and that $\tilde{Q} \in \arg\max_{Q\in\Delta_P} H_Q(T|X,Y) \cap \mathring{\Delta}_P$. Then, $T \perp\!\!\!\perp_{\tilde{Q}} X \,\big|\, Y$ or $T \perp\!\!\!\perp_{\tilde{Q}} Y \,\big|\, X$. Thus, either $UI(T : X \setminus Y) = 0$ or $UI(T : Y \setminus X) = 0$.*

Remark 4.2. The proof of the theorem relies on the vanishing condition of the directional derivatives. Thus, the conclusion still holds when $\tilde{Q}$ does not have full support, as long as all directional derivatives of the target function $H_Q(T|X,Y)$ exist and vanish at $\tilde{Q}$. By Remark 3.2, this happens if and only if for any $t \in \mathcal{T}'$ the matrix $(\tilde{Q}(t|x,y))_{x,y} \in \mathbb{R}^{\mathcal{X}\times\mathcal{Y}}$ has rank one.

Remark 4.3. When $\mathcal{T}$ has cardinality three or more, the statement of the theorem becomes false; see Example 6.1. This is related to the fact that there exist three positive rank-one-matrices the sum of which has again rank one, cf. Remark 3.2. When the support of Δ_P is not full, the statement of the theorem becomes false, even when all variables are binary; see Example 6.4

Remark 4.4. Theorem 4.1 can be used to efficiently compute UI (and the corresponding bivariate information decomposition) when the optimum lies in the interior of Δ_P, as searching for conditional independences in Δ_P constitutes solving a linear programming problem (see the proof of Theorem 4.5). If no solution in the interior is found, $\max_{Q\in\partial\Delta_P}(H_Q(T|X,Y))$ has to be solved.

Proof. Under the assumption that the optimum is attained in the interior of Δ_P, it is characterized by $\frac{\partial H_q(T|X,Y)}{\partial_{\gamma_{t,x,y}}} = 0$. This leads to the set of equations

$$\log \frac{\tilde{Q}(t|x,y_0)\tilde{Q}(t|x_0,y)}{\tilde{Q}(t|x_0,y_0)\tilde{Q}(t|x,y)} = 0 \ ,$$

for $t \in \{0,1\}$, $x \in \mathcal{X} \setminus \{x_0\}$ and $y \in \mathcal{Y} \setminus \{y_0\}$. Since T is binary, for fixed x, y, this leads to the conditions

$$\begin{aligned} \tilde{Q}(0|x,y_0)\tilde{Q}(0|x_0,y) &= \tilde{Q}(0|x_0,y_0)\tilde{Q}(0|x,y) \\ \tilde{Q}(1|x,y_0)\tilde{Q}(1|x_0,y) &= \tilde{Q}(1|x_0,y_0)\tilde{Q}(1|x,y) \ . \end{aligned}$$

Using $\tilde{Q}(0|x,y) = 1 - \tilde{Q}(1|x,y)$, these equations rewrite to

$$\begin{aligned} \tilde{Q}(0|x,y_0)\tilde{Q}(0|x_0,y) &= \tilde{Q}(0|x_0,y_0)\tilde{Q}(0|x,y) \\ \tilde{Q}(0|x,y_0) + \tilde{Q}(0|x_0,y) &= \tilde{Q}(0|x_0,y_0) + \tilde{Q}(0|x,y) \ . \end{aligned}$$

These equations imply

$$\begin{aligned} &(\tilde{Q}(0|x,y_0) - \tilde{Q}(0|x_0,y_0))(\tilde{Q}(0|x_0,y) - \tilde{Q}(0|x_0,y_0)) \\ &\qquad = \tilde{Q}(0|x,y_0)\tilde{Q}(0|x_0,y) - \tilde{Q}(0|x,y_0)\tilde{Q}(0|x_0,y_0) \\ &\qquad\qquad - \tilde{Q}(0|x_0,y_0))\tilde{Q}(0|x_0,y) + \tilde{Q}(0|x_0,y_0))^2 \\ &\qquad = \tilde{Q}(0|x_0,y_0)\big(\tilde{Q}(0|x,y) - \tilde{Q}(0|x,y_0) - \tilde{Q}(0|x_0,y) + \tilde{Q}(0|x_0,y_0)\big) = 0. \end{aligned}$$

Therefore, for fixed values of x and y, there are only two possible solutions:

$$\begin{aligned} I(x,y) &: \tilde{Q}(t|x_0,y_0) = \tilde{Q}(t|x,y_0) \text{ and } \tilde{Q}(t|x,y) = \tilde{Q}(t|x_0,y) \text{ for all } t, \\ II(x,y) &: \tilde{Q}(t|x_0,y_0) = \tilde{Q}(t|x_0,y) \text{ and } \tilde{Q}(t|x,y) = \tilde{Q}(t|x,y_0) \text{ for all } t. \end{aligned}$$

Let $\mathcal{X}' = \mathcal{X}\setminus\{x_0\}$ and $\mathcal{Y}' = \mathcal{Y}\setminus\{y_0\}$. By what has been shown so far, $A_I \cup A_{II} = \mathcal{X}'\times\mathcal{Y}'$, where

$$A_I = \big\{(x,y) : x\in\mathcal{X}', y\in\mathcal{Y}', I(x,y) \text{ holds}\big\},$$
$$A_{II} = \big\{(x,y) : x\in\mathcal{X}', y\in\mathcal{Y}', II(x,y) \text{ holds}\big\}.$$

We next show that either $A_I = \mathcal{X}'\times\mathcal{Y}'$ or $A_{II} = \mathcal{X}'\times\mathcal{Y}'$ (or both).

Suppose that A_I is not empty. Let $(x,y)\in A_I$, and let $y'\in\mathcal{Y}'\setminus\{y\}$. If $II(x,y')$ holds, then $\tilde{Q}(t|x,y') = \tilde{Q}(t|x,y_0) = \tilde{Q}(t|x_0,y_0) = \tilde{Q}(t|x_0,y')$. Thus, $I(x,y')$ also holds, which implies $(x,y')\in A_I$. Thus, $A_I\subset\mathcal{X}'\times\mathcal{Y}'$ is of the form $A_I = \mathcal{X}_I'\times\mathcal{Y}'$, where $\mathcal{X}_I'\subseteq\mathcal{X}'$.

Similarly, $A_{II} = \mathcal{X}'\times\mathcal{Y}_{II}'$, where $\mathcal{Y}_{II}'\subseteq\mathcal{Y}'$. If $A_I\neq\emptyset$ and $A_{II}\neq\emptyset$, then $A_I\cap A_{II}\neq\emptyset$; say $(x',y')\in A_I\cap A_{II}$. Let $(x,y)\in A_I$. Then $\tilde{Q}(t|x,y) = \tilde{Q}(t|x',y) = \tilde{Q}(t|x',y')$ for all t. Similarly, if $(x,y)\in A_{II}$. Then $\tilde{Q}(t|x,y) = \tilde{Q}(t|x,y') = \tilde{Q}(t|x',y')$ for all t. Thus, all conditional distributions of t given any $(x,y)\in\mathcal{X}\times\mathcal{Y}$ are identical, and so $A_I = A_{II} = \mathcal{X}'\times\mathcal{Y}'$.

The theorem follows from the following observation: if $A_I = \mathcal{X}'\times\mathcal{Y}'$, then $T \perp\!\!\!\perp_{\tilde{Q}} X \,\big|\, Y$, and if $A_{II} = \mathcal{X}'\times\mathcal{Y}'$, then $T \perp\!\!\!\perp_{\tilde{Q}} Y \,\big|\, X$. □

As a corollary to Theorem 3.6:

Theorem 4.5. *Under the assumptions of Theorem 4.1, the optimizer is not unique for $|\mathcal{X}|>2, |\mathcal{Y}|>2$.*

More precisely, under the assumptions of Theorem 4.1, Theorem 3.7 implies that the optimizer is not unique

- when $UI(T : X\setminus Y) = 0$ and $|\mathcal{Y}|>2$, or
- when $UI(T : Y\setminus X) = 0$ and $|\mathcal{X}|>2$.

4.2. **The case of restricted support.** With a little more effort, the analysis of Theorem 4.1 extends to the case where Δ_P has restricted support. For any $t\in\mathcal{T}' = \{0,1\}$ let $\mathcal{X}_t = \operatorname{supp}(P(X|T=t))$ and $\mathcal{Y}_t = \operatorname{supp}(P(Y|T=t))$. Lemma 2.3 says that $\operatorname{supp}(\Delta_{P,t}) = \mathcal{X}_t\times\mathcal{Y}_t$.

For any $t\in\mathcal{T}$ let $\bar{t} = 1-t$. If $x\notin\mathcal{X}_t$, then $P(T=\bar{t}|X=x)=1$. Therefore, $T\perp\!\!\!\perp Y \,\big|\, \{X=x\}$ for all $x\in\mathcal{X}\setminus\mathcal{X}_t$. Similarly, $T\perp\!\!\!\perp X \,\big|\, \{Y=y\}$ for all $y\in\mathcal{Y}\setminus\mathcal{Y}_t$. Thus, to prove that $T\perp\!\!\!\perp Y \,\big|\, X$, say, it suffices to look at $\mathcal{X}_0\cap\mathcal{X}_1$.

Lemma 4.6.
1. *If $\mathcal{X}_0\cap\mathcal{X}_1=\emptyset$, then $T\perp\!\!\!\perp_Q Y \,\big|\, X$ for any $Q\in\Delta_P$.*
2. *If $\mathcal{Y}_0\cap\mathcal{Y}_1=\emptyset$, then $T\perp\!\!\!\perp_Q X \,\big|\, Y$ for any $Q\in\Delta_P$.*
3. *Suppose that $\mathcal{X}_0\cap\mathcal{X}_1\neq\emptyset\neq\mathcal{Y}_0\cap\mathcal{Y}_1$.*
   (a) *If $\mathcal{X}_t\setminus\mathcal{X}_{\bar{t}}\neq\emptyset$ and $\mathcal{Y}_t\setminus\mathcal{Y}_{\bar{t}}\neq\emptyset$ for some $t\in\mathcal{T}'$, then there is no maximizer of $\max_{Q\in\Delta_P} H(T|X,Y)$ in $\mathring{\Delta}_P$.*
   (b) *If $\mathcal{X}_t\setminus\mathcal{X}_{\bar{t}}\neq\emptyset$ and if there exists $Q^*\in\mathring{\Delta}_P\cap\arg\max_{Q\in\Delta_P} H(T|X,Y)$, then $T\perp\!\!\!\perp_{Q^*} Y \,\big|\, \{X, Y\in\mathcal{Y}_t\}$ (i.e., with respect to Q^*, T is independent of Y given X, given that $Y\in\mathcal{Y}_t$).*
   (c) *If $\mathcal{Y}_t\setminus\mathcal{Y}_{\bar{t}}\neq\emptyset$ and if there exists $Q^*\in\mathring{\Delta}_P\cap\arg\max_{Q\in\Delta_P} H(T|X,Y)$, then $T\perp\!\!\!\perp_{Q^*} X \,\big|\, \{Y, X\in\mathcal{X}_t\}$.*

Proof. Statements (1) and (2): If $\mathcal{X}_0\cap\mathcal{X}_1=\emptyset$, then T is a function of X for any $Q\in\Delta_P$, whence $T\perp\!\!\!\perp_Q Y \,\big|\, X$. Statement (2) follows similarly.

Statement (3a): Let $x_0 \in \mathcal{X}_0 \cap \mathcal{X}_1$, $y_0 \in \mathcal{Y}_0 \cap \mathcal{Y}_1$, $x_1 \in \mathcal{X}_t \setminus \mathcal{X}_{\bar{t}} \neq \emptyset$ and $y_1 \in \mathcal{Y}_t \setminus \mathcal{Y}_{\bar{t}} \neq \emptyset$. Suppose that $q \in \mathring{\Delta}_P$. Then $Q(t, x_0, y_0) > 0$ and $Q(\bar{t}, x_0, y_0) > 0$, whence $Q(t|x_0, y_0) \neq 1$. Then the derivative of $H(T|X, Y)$ in the direction of $\gamma_{t;x_0,x_1;y_0,y_1}$ is

$$\log \frac{Q(t|x_0, y_0)Q(t|x_1, y_1)}{Q(t|x_0, y_1)Q(t|x_1, y_0)} = \log Q(t|x_0, y_0) \neq 0.$$

Statement (3b): If $|\mathcal{Y}_t| = 1$, then Y is constant when conditioning on $Y \in \mathcal{Y}_t$, whence the conclusion holds trivially. Let $y_0, y_1 \in \mathcal{Y}_t$ with $y_0 \neq y_1$, let $x_0 \in \mathcal{X}_0 \cap \mathcal{X}_1$, and let $x_1 \in \mathcal{X}_t \setminus \mathcal{X}_{\bar{t}} \neq \emptyset$. The derivative of $H(T|X, Y)$ at Q^* in the direction of $\gamma_{t;x_0,x_1;y_0,y_1}$ is

$$\log \frac{Q^*(t|x_0, y_0)Q^*(t|x_1, y_1)}{Q^*(t|x_0, y_1)Q^*(t|x_1, y_0)} = \log \frac{Q^*(t|x_0, y_0)}{Q^*(t|x_0, y_1)}.$$

By assumption, this derivative vanishes at Q^*, whence $Q^*(t|x_0, y_0) = Q^*(t|x_0, y_1)$, which proves the statement. ☐

Theorem 4.7. *Let T be binary, and suppose that $Q^* \in \arg\max_{Q \in \Delta_P} H(T|X, Y)$ lies in $\mathring{\Delta}_P$.*

- *If $\mathcal{X}_0 = \mathcal{X}_1$ and $\mathcal{Y}_0 \neq \mathcal{Y}_1$, then $T \perp\!\!\!\perp_{Q^*} X \,\big|\, Y$.*
- *If $\mathcal{Y}_0 = \mathcal{Y}_1$ and $\mathcal{X}_0 \neq \mathcal{X}_1$, then $T \perp\!\!\!\perp_{Q^*} Y \,\big|\, X$.*

Proof. The theorem follows from Lemma 4.6 (3b) and (3c). ☐

4.3. **Statistics for uniqueness and support of optimizers for binary T.** To better understand whether the optimizer typically lies in the interior of Δ_P and whether it is typically unique, we uniformly sampled joint distributions $P \in \Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$ for binary $\mathcal{T}$ and different cardinalities of $|\mathcal{X}|, |\mathcal{Y}|$. Uniform sampling from $\Delta_{T,X,Y}$ was performed with the Kraemers' method [Smith and Tromble, 2004]. Based on 1000 samples, the following percentage of optima were found in the interior of Δ_P:

$\lvert\mathcal{X}\rvert/\lvert\mathcal{Y}\rvert$	2	3	4	5
2	76.6	51.4	76.0	75.2
3	-	51.4	55.2	59.8
4	-	-	53.8	50.4
5	-	-	-	47.1

The percentage of solutions found in the interior of Δ_P decreases with increasing cardinality of $|\mathcal{X}|$ and $|\mathcal{Y}|$. The following table lists the same percentages for $|\mathcal{X}| = |\mathcal{Y}| = k$ for different values of k.

k:	5	6	7	8	9	10	11	12
optimizer in interior [%]:	47.1	47.8	47.2	43.9	40.6	41.0	42.4	37.4
k:	13	14	15	16	17	18	19	20
optimizer in interior [%]:	35.1	37.3	38.0	37.5	34.9	32.5	33.2	29.1

Again, each entry corresponds to 1000 samples.

When sampling uniformly, we only find distributions with full support. In accordance with Theorem 4.5, we only find unique optima in the interior of Δ_P for cardinalities $2 \times 2 \times k$:

T	X	Y	$P(t,x,y)$
0	0	0	$a(bd+\gamma_1)$
0	0	1	$a(b(1-d)-\gamma_1)$
0	1	0	$a((1-b)d-\gamma_1)$
0	1	1	$a((1-b)(1-d)+\gamma_1)$
1	0	0	$(1-a)(ce+\gamma_2)$
1	0	1	$(1-a)(c(1-e)-\gamma_2)$
1	1	0	$(1-a)((1-c)e-\gamma_2)$
1	1	1	$(1-a)((1-c)(1-e)+\gamma_2)$

TABLE 1. Parameterization of $2\times 2\times 2$ distributions

k:	2	3	4	5
optimizer unique [%]:	100	27.6	8.3	2.9

5. The all binary case

If X, Y and T are all binary, $\Delta_{\mathcal{T},\mathcal{X},\mathcal{Y}}$ has 7 dimensions, which split in 5 dimensions for V and 2 dimensions for Δ_P.

Throughought this section we assume that $\mathcal{T}' = \{0,1\} = \mathcal{X} = \mathcal{Y}$. In the following, V is parameterized by the variables

$$a = P_T(0), \qquad \begin{aligned} b &= P_{X|T}(0|0), & d &= P_{Y|T}(0|0),\\ c &= P_{X|T}(0|1), & e &= P_{Y|T}(0|1), \end{aligned} \tag{8}$$

and by the coefficients γ_1, γ_2 of $a\gamma_{0;0;1;0;1}, (1-a)\gamma_{1;0;1;0;1}$. Table 1 makes the parametrization 5 explicit.

Δ_P is a rectangle. The allowed parameter domain is

$$\begin{aligned} -\min\{bd,(1-b)(1-d)\} &\le \gamma_1 \le \min\{b(1-d),(1-b)d\}\\ -\min\{ce,(1-c)(1-e)\} &\le \gamma_2 \le \min\{c(1-e),(1-c)e\}\ . \end{aligned}$$

The lower and upper bounds on γ_i will be denoted by $\gamma_{i_{\min}}$ and $\gamma_{i_{\max}}$ respectively.

The following holds:

1. $\Delta_{p,0}$ is a singleton iff $b\in\{0,1\}$ or $d\in\{0,1\}$.
2. $\Delta_{p,1}$ is a singleton iff $c\in\{0,1\}$ or $e\in\{0,1\}$.
3. Δ_P is a singleton iff both conditions are met. Thus, Δ_P degenerates to a single point precisely in the following four cases:
   (a) $H(X|T)=0$;
   (b) $H(Y|T)=0$;
   (c) $H(X|T=0)=0$ and $H(Y|T=1)=0$;
   (d) $H(X|T=1)=0$ and $H(Y|T=0)=0$.

In the all-binary case, Theorem 4.1 slightly generalizes:

Theorem 5.1. *Let X,Y,T be binary. Suppose that Δ_p is not a singleton in case (c) or (d). If $\tilde{Q} = \arg\max_{Q\in\Delta_P} H_Q(T|X,Y) \in \mathring{\Delta}_P$, then $T \perp\!\!\!\perp_{\tilde{Q}} X \,\big|\, Y$ or $T \perp\!\!\!\perp_{\tilde{Q}} Y \,\big|\, X$.*

Remark 5.2. Example 6.4 shows that the conclusion does not in general hold in the singleton cases (c) and (d).

Proof. The singleton cases (a) and (b) are trivial, and the remaining cases follow from Theorem 4.7. □

In the all-binary case, uniqueness can be completely characterized:

Theorem 5.3. $\arg\max_{Q\in\Delta_P} H_Q(T|X,Y)$ *is unique, unless* $b=c$ *and* $d=e$.

Proof. If $\arg\max_{Q\in\Delta_P} H_Q(T|X,Y)$ is not unique, then $\arg\max_{Q\in\mathring{\Delta}_P} H_Q(T|X,Y)$ is not unique either (by Lemma 3.3), so we may restrict attention to maximizers in the interior of Δ_P.

First assume that Δ_P has full support. As shown in Theorem 4.1 and its proof, there are two cases *I* and *II* to consider. Inserting the parameterization from above and using the injectivity of $\frac{1}{1+x}$ leads for case *I* to the equations [1]

$$\begin{aligned}\frac{ce+\gamma_2}{bd+\gamma_1} &= \frac{(1-c)e-\gamma_2}{(1-b)d-\gamma_1}\\ \frac{c(1-e)-\gamma_2}{b(1-d)-\gamma_1} &= \frac{(1-c)(1-e)+\gamma_2}{(1-b)(1-d)+\gamma_1}\ ,\end{aligned}$$

which simplify to

$$\begin{aligned}\gamma_2 d-\gamma_1 e &= de(b-c)\\ \gamma_1(1-e)-\gamma_2(1-d) &= (1-d)(1-e)(b-c)\ .\end{aligned}$$

Rearranging for γ_1,γ_2 leads to

$$\begin{aligned}\gamma_1(d-e) &= d(b-c)(1-d)\\ \gamma_2(d-e) &= e(b-c)(1-e)\ .\end{aligned}\tag{9}$$

For $d\neq e$, there exists a unique solution. For $b=c$, the optimum is Q_0 itself.

Similarly, case *II* reduces to

$$\begin{aligned}\gamma_2 b-\gamma_1 c &= bc(d-e)\\ \gamma_1(1-c)-\gamma_2(1-b) &= (1-b)(1-c)(d-e)\end{aligned}$$

and rearranging for γ_1,γ_2 gives

$$\begin{aligned}\gamma_1(b-c) &= b(d-e)(1-b)\\ \gamma_2(b-c) &= c(d-e)(1-c)\ .\end{aligned}\tag{10}$$

Again, there exists a unique solution for $b\neq c$ and Q_0 is the optimum for $d=e$.

Now assume that Δ_P is a line. Following the proof of Theorem 5.1, assume that $b=0$. Plugging the parametrization from above into the equality $Q(1|10)=Q(1|11)$ gives

$$\frac{(1-a)\big((1-c)e-\gamma_2\big)}{(1-a)\big((1-c)e-\gamma_2\big)+P(010)} = \frac{(1-a)\big((1-c)(1-e)+\gamma_2\big)}{(1-a)\big((1-c)(1-e)+\gamma_2\big)+P(011)}.$$

If $P(010)=0$, then $P(011)=0$, and conversely; otherwise, this equation has no solution. In this case $P(010)=P(011)=0$, the sum $P(01)=P(010)+P(011)=a$ vanishes, which contradicts $\mathcal{T}'=\{0,1\}$. Thus, $P(010)\neq 0$ and $P(011)\neq 0$. Using injectivity of $x\mapsto\frac{1}{1+x}$ and cancelling $(1-a)$, this is equivalent to

$$\frac{(1-c)e-\gamma_2}{P(010)} = \frac{(1-c)(1-e)+\gamma_2}{P(011)}.\tag{11}$$

[1]No solutions exist for which one denominator equals 0. The same applies for case *II*.

This equation is linear in γ_2 and has a single unique solution, since the coefficient $\frac{1}{P(010)} + \frac{1}{P(011)}$ in front of γ_2 is positive. □

Only the case where the maximizer lies on the boundary of Δ_P remains to be analyzed.

Theorem 5.4. *Assume that* $\tilde{Q} = \arg\max_{Q\in\Delta_P} H_Q(T|X,Y)$ *lies at the boundary of* Δ_P*. Then, it is attained either at* $(\gamma_{1_{\min}}, \gamma_{2_{\min}})$ *or* $(\gamma_{1_{\max}}, \gamma_{2_{\max}})$*.*

Proof. If Δ_P is degenerate, then either $\gamma_{1_{\min}} = \gamma_{1_{\max}}$ or $\gamma_{2_{\min}} = \gamma_{2_{\max}}$, and the theorem becomes trivial. Otherwise, the statement follows from Lemma 3.3. □

To sum up, assuming $H(X) > 0$, $H(Y) > 0$ and $H(T) > 0$, there are five cases:

(1) $b = c$ and $d = e$. In this case, $X \perp\!\!\!\perp Y \,\big|\, T$, and $\arg\max_{Q\in\Delta_P} H_Q(T|X,Y)$ is not unique, but consists of the diagonal of Δ_P.
(2) $T \perp\!\!\!\perp_{\tilde{Q}} X \,\big|\, Y$ for the unique $\tilde{Q} = \arg\max_{Q\in\Delta_P} H_Q(T|X,Y)$.
(3) $T \perp\!\!\!\perp_{\tilde{Q}} Y \,\big|\, X$ for the unique $\tilde{Q} = \arg\max_{Q\in\Delta_P} H_Q(T|X,Y)$.
(4) The unique maximizer lies at $(\gamma_{1_{\min}}, \gamma_{2_{\min}})$.
(5) The unique maximizer lies at $(\gamma_{1_{\max}}, \gamma_{2_{\max}})$.

The last four cases intersect. For example, the intersection of the last four cases contains the distribution $\frac{1}{2}\delta_{000} + \frac{1}{2}\delta_{111}$ (see Fink [2011], Rauh and Ay [2014] for a discussion of the intersection of cases (2) and (3)).

The five cases can be identified by checking certain polynomial equalities among the parameters a, b, c, d, e. Therefore, the five cases correspond to five semi-algebraic sets of probability distributions. For example, case (2) holds if and only if the unique solution (γ_1, γ_2) to (9) satisfies $\gamma_{i_{\min}} \leq \gamma_i \leq \gamma_{i_{\max}}$ for $i = 1, 2$, which can be formulated as eight polynomial inequalities.

These results make it possible to exactly solve $\arg\max_{Q\in\Delta_P}(H_Q(T|X,Y)$ by checking whether the solutions of (9), (10) or (11) lie in Δ_P and otherwise using the maximum of $H(T|X,Y)$ at $(\gamma_{1_{\min}}, \gamma_{2_{\min}})$ and $(\gamma_{1_{\max}}, \gamma_{2_{\max}})$.

6. Examples

Example 6.1 (For ternary T, maximizers with full support need not satisfy CI statements)**.** Let X, Y be binary random variables with $P(X,Y)$ arbitrary (of full support), and let T be ternary with

$$(P(T=1|X=x,Y=y))_{x,y} = \begin{pmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{12} & \frac{1}{8} \end{pmatrix},$$
$$(P(T=2|X=x,Y=y))_{x,y} = \begin{pmatrix} \frac{1}{3} & \frac{1}{8} \\ \frac{5}{24} & \frac{5}{64} \end{pmatrix},$$
$$(P(T=3|X=x,Y=y))_{x,y} = \begin{pmatrix} \frac{1}{3} & \frac{3}{8} \\ \frac{17}{24} & \frac{51}{64} \end{pmatrix}$$

Then P minimizes $I_Q(T:X|Y)$ on Δ_P (cf. Remark 3.2), and one can check that P is the unique minimizer on Δ_P (it is impossible to find a line in Δ_P such that the two points at which this line hits the boundary satisfy the conclusion of Lemma 3.3). P has full support, but there is no conditional independence statement.

Example 6.2. Consider the distributions

x	y	t	$P(x,y,t)$
0	0	0	$\frac{1}{6}$
0	0	1	$\frac{1}{6}$
1	1	0	$\frac{1}{6}$
1	1	1	$\frac{1}{6}$
2	2	0	$\frac{1}{9}$
2	2	1	$\frac{2}{9}$

x	y	t	$P'(x,y,t)$
0	1	0	$\frac{1}{6}$
0	1	1	$\frac{1}{6}$
1	0	0	$\frac{1}{6}$
1	0	1	$\frac{1}{6}$
2	2	0	$\frac{1}{9}$
2	2	1	$\frac{2}{9}$

Then $T \perp\!\!\!\perp_P Y \,\big|\, X$ and $T \perp\!\!\!\perp_{P'} Y \,\big|\, X$, and $P' \in \Delta_P$. It follows that $I_P(T:Y|X) = I_{P'}(T:Y|X) = 0$, whence P and P' are both minimizers. The same holds true for any convex combination of P and P'. Note that P and P' (more generally: any convex combination of P and P') have restricted support: the probability of $\big\{X = 2, Y \neq 2\big\}$ vanishes.

Example 6.3 (The all-binary case where Δ_P is a line)**.** Consider the $2 \times 2 \times 2$ distribution given by $e = 0$ and $a, b, c, d = \frac{1}{2}$

T	X	Y	$P(t,x,y)$
0	0	0	$\frac{1}{8}$
0	0	1	$\frac{1}{8}$
0	1	0	$\frac{1}{8}$
0	1	1	$\frac{1}{8}$
1	0	1	$\frac{1}{4}$
1	1	1	$\frac{1}{4}$

Δ_P degenerates to a line $P + \gamma_1 \gamma_{0;0,1;0,1}$ with support $-\frac{1}{8} \leq \gamma_1 \leq \frac{1}{8}$. The conditional entropy is

$$\begin{aligned} H_{\gamma_1}(T|X,Y) &= (\tfrac{3}{8} - \gamma_1) H_{\gamma_1}(T|0,1) + (\tfrac{3}{8} + \gamma_1) H_{\gamma_1}(T|1,1) \\ &= (\tfrac{3}{8} - \gamma_1) h\left(\frac{\frac{1}{8} - \gamma_1}{\frac{3}{8} - \gamma_1}, \frac{\frac{1}{4}}{\frac{3}{8} - \gamma_1}\right) + (\tfrac{3}{8} + \gamma_1) h\left(\frac{\frac{1}{8} + \gamma_1}{\frac{3}{8} + \gamma_1}, \frac{\frac{1}{4}}{\frac{3}{8} + \gamma_1}\right). \end{aligned}$$

By symmetry and Lemma 3.5, the unique maximizer of $H_{\gamma_1}(T|X,Y)$ lies at $\gamma_1 = 0$, that is, P is the unique solution to the optimization problem. In this case, P equals Q_0; that is, $X \perp\!\!\!\perp_P Y \,\big|\, T$ holds. Moreover, $T \perp\!\!\!\perp_P X \,\big|\, Y$ holds.

Example 6.4 (The all-binary case where Δ_P is a singleton)**.** Consider the $2 \times 2 \times 2$ distribution given by $b = e = 1$ and $a, c, d = \frac{1}{2}$:

T	X	Y	$P(t,x,y)$
0	0	0	$\frac{1}{4}$
0	0	1	$\frac{1}{4}$
1	0	0	$\frac{1}{4}$
1	1	0	$\frac{1}{4}$

Here, Δ_P is a singleton. Neither $T \perp\!\!\!\perp_P Y \,\big|\, X$ nor $T \perp\!\!\!\perp_P X \,\big|\, Y$ holds.

References

Pradeep Kr. Banerjee, Johannes Rauh, and Guido Montúfar. Computing the unique information. In *Proc. IEEE ISIT*, pages 141–145. IEEE, 2018.

Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014.

Persi Diaconis and Bernd Sturmfels. Algebraic algorithms for sampling from conditional distributions. *Annals of Statistics*, 26:363–397, 1998.

Alex Fink. The binomial ideal of the intersection axiom for conditional probabilities. *Journal of Algebraic Combinatorics*, 33(3):455–463, 2011.

Conor Finn and Joseph T. Lizier. Pointwise partial information decomposition using the specificity and ambiguity lattices. *Entropy*, 20(4):297, 2018.

Malte Harder, Christoph Salge, and Daniel Polani. A bivariate measure of redundant information. *Phys. Rev. E*, 87:012130, Jan 2013.

Robin Ince. Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy*, 19(7):318, 2017.

Ryan James, Jeffrey Emenheiser, and James Crutchfield. Unique information via dependency constraints. *Journal of Physics A*, 52(1):014002, 2018.

Xueyan Niu and Christopher Quinn. A measure of synergy, redundancy, and unique information using information geometry. In *Proc. IEEE ISIT*, 2019.

Johannes Rauh and Nihat Ay. Robustness, canalyzing functions and systems design. *Theory in Biosciences*, 133(2):63–78, 2014.

Johannes Rauh, Nils Bertschinger, Eckehard Olbrich, and Jürgen Jost. Reconsidering unique information: Towards a multivariate information decomposition. In *Proc. IEEE ISIT*, pages 2232–2236, 2014.

Noah A Smith and Roy W Tromble. Sampling uniformly from the unit simplex. *Johns Hopkins University, Tech. Rep*, 29, 2004.

Paul Williams and Randall Beer. Nonnegative decomposition of multivariate information. *arXiv:1004.2515v1*, 2010.

MPI für Mathematik in den Naturwissenschaften, Leipzig
E-mail address: jrauh@mis.mpg.de

Institute for Theoretical Physics, University of Bremen
E-mail address: maik@neuro.uni-bremen.de