About the mutual (conditional) information

Renato Renner Ueli Maurer

Department of Computer Science
Swiss Federal Institute of Technology (ETH Zurich)
CH-8092 Zurich, Switzerland
E-mail: {renner,maurer}@inf.ethz.ch

Abstract

In general, the mutual information between two random variables X and Y, $I(X;Y)$, might be larger or smaller than their mutual information conditioned on some additional information Z, $I(X;Y|Z)$. Such additional information Z can be seen as output of a channel C taking as input X and Y. It is thus a natural question, with applications in fields such as information theoretic cryptography, whether conditioning on the output Z of a fixed channel C can potentially increase the mutual information between the inputs X and Y.

In this paper, we give a necessary, sufficient, and easily verifiable criterion for the channel C (i.e., the conditional probability distribution $P_{Z|XY}$), such that $I(X;Y) \geq I(X;Y|Z)$ holds for every joint distribution of the random variables X and Y. Furthermore, the result is generalized to channels with n inputs (for $n \in \mathbb{N}$), that is, to conditional probability distributions of the form $P_{Z|X_1 \cdots X_n}$.

1 Introduction

The mutual information $I(X;Y)$ between two random variables X and Y is one of the basic measures in information theory. It can be interpreted as the amount of information that X gives on Y (or vice versa). In general, additional information, i.e., conditioning on an additional random variable Z, can either increase or decrease this mutual information.[1] Without loss of generality[2], this additional information Z can be seen as output of a channel C with input (X,Y), which is fully specified by the conditional probability distribution $P_{Z|XY}$.

In the following, we investigate the question whether for a fixed conditional probability distribution $P_{Z|XY}$ (i.e., a fixed channel C with input (X,Y) and output Z), conditioning on Z can increase the mutual information between X and Y. We give a sufficient criterion, depending only on $P_{Z|XY}$, such that this is not the case, i.e., $I(X;Y) \geq I(X;Y|Z)$ for all distributions P_{XY}. The criterion is also necessary in the sense that, if it is not satisfied, there exists a probability distribution P_{XY} such that $I(X;Y) < I(X;Y|Z)$. Moreover, since our criterion is basically a simple information theoretic expression, it can easily be handled, and the verification of whether it is satisfied by a given conditional probability distribution $P_{Z|XY}$ is efficient.

One possible application of this result is in the field of information theoretic cryptography, where it is used for the analysis of secret-key agreement protocols[3], but this application is not discussed in this extended abstract.

This paper is organized as follows. In Section 2, the notation and some definitions are introduced. The main theorem is stated and proved in Section 3. A generalization of the result to channels with more than two inputs, i.e., to probability distributions of the form $P_{Z|X_1 \cdots X_n}$ for some $n \in \mathbb{N}$, is described in Section 4.

[1] Let for example $X = Y = Z$ be three (uniformly distributed) binary random variables. Then, conditioning on Z decreases the mutual information between X and Y. On the other hand, for two independent binary random variables X and Y, conditioning on $Z = X \oplus Y$ increases their mutual information.

[2] at least in the context where only the three random variables X, Y and Z are considered

[3] See [3] for an example of information-theoretically secure secret-key agreement.

2 Definitions

Let in the following p be a conditional probability distribution of the form

$$p : (x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \longmapsto p(z|x, y),$$

i.e., for each pair $(x, y) \in \mathcal{X} \times \mathcal{Y}$, $p(\cdot|x, y)$ is a probability distribution of a random variable with range $\mathcal{Z}$.

The conditional probability distribution p uniquely defines a channel[4] C taking as input two random variables X and Y with ranges $\mathcal{X}$ and $\mathcal{Y}$, respectively, and giving an output Z in the range $\mathcal{Z}$. The main goal of this paper is to investigate the question whether for such a fixed channel C conditioning on the channel output Z can increase the mutual information (i.e., the correlation) between the two inputs X and Y with arbitrary joint distribution P_{XY}. This motivates the following definition.

Definition 2.1. *The conditional probability distribution p is called* correlation free *if*

$$I(X; Y) \geq I(X; Y|Z)$$

for all P_{XY}, where X, Y and Z are random variables[5] *distributed according to P_{XY} and $P_{Z|XY} := p$.*

Similar to the joint probability distribution P_{UV} of two random variables U and V, which is the product of P_U and P_V if and only if U and V are statistically independent, we will see in Section 3 that the conditional probability distribution p can be written as a product if and only if it is correlation free.

Definition 2.2. *The conditional probability distribution p is called* multiplicative *if it is the product of two functions r and s depending only on (z, x) and (z, y), respectively, i.e.,*

$$p(z|x, y) = r(z, x) \cdot s(z, y)$$

for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$.

It is easy to decide whether a given conditional probability distribution p is multiplicative. The following lemma shows that one only has to check the conditional independence of a certain pair of random variables.

Lemma 2.3. *The conditional probability distribution p is multiplicative if and only if*

$$I(X; Y|Z) = 0$$

for two independent random variables X and Y with uniform distribution on their ranges $\mathcal{X}$ and $\mathcal{Y}$, respectively (i.e., $P_{XY}(x, y) = c$ for some constant c), and Z with $P_{Z|XY} := p$.

Proof. From

$$p(z|x, y) = P_{Z|XY}(z|x, y) = \frac{P_Z(z) \cdot P_{XY|Z}(x, y|z)}{P_{XY}(x, y)} = \frac{P_Z(z)}{c} \cdot P_{XY|Z}(x, y|z)$$

it is obvious that p is multiplicative if and only if (for each fixed $z \in \mathcal{Z}$ with $P_Z(z) > 0$) the probability distribution $P_{XY|Z=z}$ can be written as a product of a function depending only on x and a function depending only on y, which is equivalent to the independence of X and Y conditioned on Z, i.e., $I(X; Y|Z) = 0$. $\square$

3 Main Result

Theorem 3.1. *A conditional probability distribution*

$$p : (x, y, z) \longmapsto p(z|x, y)$$

is correlation free if and only if it is multiplicative.

[4] such that p is its probability transition matrix

[5] In this work, we restrict to random variables with finite entropy. However, their range might be infinite.

Proof. If p is not multiplicative, then it follows from Lemma 2.3 that $I(X;Y|Z) > 0$ for random variables X, Y and Z with $P_{XY} = c$ (where c is some constant) and $P_{Z|XY} = p$, while, obviously, $I(X;Y) = 0$. Consequently, conditioning on Z increases the mutual information between X and Y, i.e., p is not correlation free.

It thus remains to be shown that if for any random variables X, Y and Z, the conditional probability distribution $P_{Z|XY}$ is multiplicative, then $I(X;Y) \geq I(X;Y|Z)$. The argument will be subdivided into two parts, where in the first part, the implication is proven for a special case, called *deterministic case*. In the second part, we will make use of this result to prove the general case.

Let X, Y and Z be random variables with ranges $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, respectively, such that $P_{Z|XY}$ is multiplicative, i.e., $P_{Z|XY}(z|x,y) = r(z,x) \cdot s(z,y)$. For the deterministic case, we additionally assume that all values $r(z,x)$ and $s(z,y)$ are either 0 or 1, which obviously implies that $P_{Z|XY}(z|x,y)$ is also 0 or 1 (for all x,y,z). The value of Z is thus uniquely determined by X and Y, i.e.,

$$H(Z|XY) = 0. \tag{1}$$

The main idea for the proof of the deterministic case is to introduce an additional random variable Y' with range $\mathcal{Y}$ and $P_{Y'|ZX}(y|z,x) := P_{Y|Z}(y|z)$ (for all x,y,z). Hence we have

$$\begin{aligned} I(Y;Z) &= I(Y';Z) & (2)\\ I(X;Y'|Z) &= 0, & (3) \end{aligned}$$

i.e., $X \to Z \to Y'$ is a Markov chain. Moreover, the value of Z is uniquely determined by the values of X and Y', i.e.,

$$H(Z|XY') = 0. \tag{4}$$

This can be seen as follows. Assume by contradiction that $H(Z|XY') > 0$. Then there exist (at least) two different values $z_1 \neq z_2$ and x, y such that $P_{XY'Z}(x,y,z_1) > 0$ and $P_{XY'Z}(x,y,z_2) > 0$. Since for $i = 1$ and $i = 2$

$$0 < \sum_y P_{XY'Z}(x,y,z_i) = \sum_y P_{XYZ}(x,y,z_i) = r(z_i,x) \cdot \sum_y s(z_i,y) \cdot P_{XY}(x,y),$$

and similarly

$$0 < s(z_i,y) \cdot \sum_x r(z_i,x) \cdot P_{XY}(x,y),$$

the factors $r(z_i,x)$ and $s(z_i,y)$ must be nonzero and thus by assumption be equal to 1. Consequently, the probabilities $P_{Z|XY}(z_1|x,y)$ and $P_{Z|XY}(z_2|x,y)$ are both equal to 1, which is a contradiction.

From (1) and (4) the expressions $I(XY;Z)$ and $I(XY';Z)$ are both equal to $H(Z)$ and thus, using (2),

$$I(X;Z) + I(Y;Z) = I(X;Z) + I(Y';Z) \geq I(X;Z) + I(Y';Z|X) = I(XY';Z) = I(XY;Z) \tag{5}$$

where the inequality follows from the fact that $X \to Z \to Y'$ is a Markov chain (see (3)). Making use of some basic information theoretic equalities shows that

$$I(X;Z) + I(Y;Z) \geq I(XY;Z) \iff I(X;Y) \geq I(X;Y|Z), \tag{6}$$

which concludes the proof for the deterministic case.

To prove the general case, we again assume that for given random variables X, Y and Z the conditional probability distribution $P_{Z|XY}$ is multiplicative, but this time, the factors r and s of $P_{Z|XY}$ might take on any value in the interval $[0,1]$.[6]

The main idea is to reduce this case to the deterministic case by constructing new random variables $\bar{X}$, $\bar{Y}$ and $\bar{Z}$ with

$$I(X;Y) = I(\bar{X};\bar{Y}) \tag{7}$$

and for which the conditional probability distribution $P_{\bar{Z}|\bar{X}\bar{Y}}$ is again multiplicative, i.e., for all x, y and z

$$P_{\bar{Z}|\bar{X}\bar{Y}}(z|x,y) = \bar{r}(z,x) \cdot \bar{s}(z,y) \tag{8}$$

[6] This is the most general case, since any two factors r and s can be replaced by $\tilde{r} := r \cdot c$ and $\tilde{s} := s/c$ where c is a function only depending on z, such that the function values of $\tilde{r}$ and $\tilde{s}$ are in the interval $[0,1]$.

where $\bar{r}(z,x), \bar{s}(z.y) \in \{0,1\}$. Additionally, the range of $\bar{Z}$ should consist of two disjoint sets $\mathcal{A}$ and $\mathcal{B}$, where

$$
\begin{aligned}
I(\bar{X};\bar{Y}|\bar{Z},\bar{Z}\in\mathcal{A}) &\geq I(X;Y|Z) &(9)\\
I(\bar{X};\bar{Y}|\bar{Z},\bar{Z}\in\mathcal{B}) &\geq I(X;Y). &(10)
\end{aligned}
$$

Then, from the result in the deterministic case, $I(\bar{X};\bar{Y}) \geq I(\bar{X};\bar{Y}|\bar{Z})$, and thus

$$
\begin{aligned}
I(X;Y) = I(\bar{X};\bar{Y}) \geq I(\bar{X};\bar{Y}|\bar{Z}) &= P[\bar{Z}\in\mathcal{A}]\cdot I(\bar{X};\bar{Y}|\bar{Z},\bar{Z}\in\mathcal{A}) + P[\bar{Z}\in\mathcal{B}]\cdot I(\bar{X};\bar{Y}|\bar{Z},\bar{Z}\in\mathcal{B})\\
&\geq P[\bar{Z}\in\mathcal{A}]\cdot I(X;Y|Z) + (1-P[\bar{Z}\in\mathcal{A}])\cdot I(X;Y)
\end{aligned}
$$

which implies $I(X;Y) \geq I(X;Y|Z)$.

The main task is thus to find such a construction of random variables $\bar{X}$, $\bar{Y}$ and $\bar{Z}$ satisfying (7), (8), (9) and (10), which will be sketched in the remaining part of this section. However, in this extended abstract, we skip the proof that the following construction fulfills the above conditions.

Without loss of generality, assume that the ranges $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ of the random variables X, Y and Z, respectively, are finite and that the function values of r and s are rational numbers.[7] We thus can write

$$
r(z,x) = \frac{\phi(z,x)}{\rho} \qquad\qquad s(z,y) = \frac{\psi(z,y)}{\sigma}
$$

with appropriate constants $\rho,\sigma\in\mathbb{N}$ and functions ϕ and ψ with ranges $\{0,1,\ldots,\rho\}$ and $\{0,1,\ldots,\sigma\}$, respectively. Moreover, to simplify the notation, set $\mathcal{Z} = \{0,1,\ldots,\gamma-1\}$ for an appropriate $\gamma\in\mathbb{N}_0$.

Let U and V be independent and uniformly distributed random variables with ranges $\mathcal{U} := \{0,1,\ldots,\alpha-1\}$ and $\mathcal{V} := \{0,1,\ldots,\beta-1\}$, respectively, where $\alpha := \rho\cdot\gamma$ and $\beta := \sigma\cdot\gamma$. Moreover, for all $x\in\mathcal{X}$ and $y\in\mathcal{Y}$ let P_x and Q_y be independent and uniformly distributed random variables with ranges $\mathcal{U}$ and $\mathcal{V}$, respectively. In the following, the $|\mathcal{X}|$-tuple $(P_{x\in\mathcal{X}})$ will be denoted as $\mathbf{P}$, and the $|\mathcal{Y}|$-tuple $(Q_{y\in\mathcal{Y}})$ as $\mathbf{Q}$. The random variables $\bar{X}$ and $\bar{Y}$ should then be defined as triples $(X,U,\mathbf{Q})$ and $(Y,V,\mathbf{P})$, respectively. Condition (7) is thus an immediate consequence of the independence of U, V, $\mathbf{P}$ and $\mathbf{Q}$.

For the construction of $\bar{Z}$ we need some additional notation. Let for all $x\in\mathcal{X}$, $y\in\mathcal{Y}$ and $z\in\mathcal{Z}$ the sets $A_{x,z}$ and $B_{y,z}$ be defined as

$$
A_{x,z} := \{u\in\mathcal{U}\,|\,z\leq\frac{u}{\rho}<z+r(z,x)\} \qquad\qquad B_{y,z} := \{v\in\mathcal{V}\,|\,z\leq\frac{v}{\sigma}<z+s(z,y)\}.
$$

Note that for any given x, the sets $A_{x,z}$ (for all z), and for any given y, the sets $B_{y,z}$ are disjoint. Set

$$
C_{x,y} := \{(u,v)\in\mathcal{U}\times\mathcal{V}\,|\,\exists\, z\in\mathcal{Z}\,:\, u\in A_{x,z}\,\wedge\, v\in B_{y,z}\}.
$$

It is easy to verify that, for fixed x,y,u,v with $(u,v)\in C_{x,y}$, there is exactly one element $z\in\mathcal{Z}$ such that $u\in A_{x,z}$ and $v\in B_{y,z}$, which we will denote as $z(x,y,u,v)$. Furthermore, the cardinality of the set $C_{x,y}$ is equal to $\rho\cdot\sigma$, and thus the cardinalities of both $C_{x,y}$ and its complement $\bar{C}_{x,y} := \mathcal{U}\times\mathcal{V}\backslash C_{x,y}$ are independent of x and y. This allows us to define a family of functions

$$
\Gamma_k = (\Gamma_k^{(1)},\Gamma_k^{(2)}) : \mathcal{X}\times\mathcal{Y}\longrightarrow\mathcal{U}\times\mathcal{V}
$$

parameterized by $k\in\{1,\ldots\kappa\}$ where $\kappa := |\bar{C}_{x,y}|$, such that for each fixed pair $(x,y)\in\mathcal{X}\times\mathcal{Y}$ the function

$$
k\longmapsto\Gamma_k(x,y) = (\Gamma_k^{(1)}(x,y),\Gamma_k^{(2)}(x,y)) \tag{11}
$$

is a bijection between $\{1,\ldots,\kappa\}$ and $\bar{C}_{x,y}$.

The random variable $\bar{Z}$ will be defined such that the value of $\bar{Z}$ is uniquely defined by $\bar{X} = (X,U,\mathbf{Q})$ and $\bar{Y} = (Y,V,\mathbf{P})$. We will distinguish two cases: $(\bar{U},\bar{V})\in C_{X,Y}$ and $(\bar{U},\bar{V})\in\bar{C}_{X,Y}$, where $\bar{U} := U - P_X \pmod{\alpha}$ and $\bar{V} := V - Q_Y \pmod{\beta}$. In the first case (if $(\bar{U},\bar{V})\in C_{X,Y}$), let

[7] It is easy to verify that any triple of random variables X, Y and Z with multiplicative conditional probability distribution $P_{Z|XY}$ can be approximated by a triple of random variables X', Y' and Z' having finite range and for which $P_{Z'|X'Y'}$ is multiplicative with rational factors, such that the mutual information between X and Y (given Z) is arbitrarily close to the mutual information between X' and Y' (given Z').

the random variable $\bar{Z}$ be defined as the triple $(Z', \mathbf{P}, \mathbf{Q})$ with $Z' := z(X, Y, \bar{U}, \bar{V})$ and let $\mathcal{A}$ be the set of all these values of $\bar{Z}$.

If $(\bar{U}, \bar{V}) \in \bar{C}_{X,Y}$, then let $\bar{Z}$ be a triple $(K, \mathbf{U}, \mathbf{V})$ where K satisfies

$$\Gamma_K(X, Y) = (\bar{U}, \bar{V}), \tag{12}$$

and where $\mathbf{U} := (U_{x \in \mathcal{X}})$ and $\mathbf{V} := (V_{y \in \mathcal{Y}})$ are $|\mathcal{X}|$- and $|\mathcal{Y}|$-tuples with

$$U_x := \Gamma_K^{(1)}(x, Y) + P_x \pmod{\alpha} \qquad \forall x \in \mathcal{X} \tag{13}$$

$$V_y := \Gamma_K^{(2)}(X, y) + Q_y \pmod{\beta} \qquad \forall y \in \mathcal{Y}. \tag{14}$$

Note that, since the function (11) is a bijection, the value of K is uniquely determined by (12). $\mathcal{B}$ is then defined as the set of all possible values of $\bar{Z}$ in this case. □

4 Generalization

The conditional probability distribution $P_{Z|XY}$ studied in the previous sections corresponds to a channel taking a pair of random variables as input. However, it is a natural question whether our considerations can be extended to channels with more than two inputs.

Let therefore p be a conditional probability distribution of the form

$$p: (x_1, \ldots, x_n, z) \longmapsto p(z|x_1, \ldots, x_n)$$

for some $n \in \mathbb{N}$. Then, there is a canonical extension of Definition 2.2 including this more general type of conditional probability distributions.

Definition 4.1. *The conditional probability distribution p is called* multiplicative *if it can be written as a product*

$$p(z|x_1, \ldots, x_n) = r_1(z, x_1) \cdots r_n(z, x_n)$$

for appropriate functions $r_1, \ldots, r_n$.

On the other hand, the generalization of the definition of correlation freeness is motivated by the expression

$$I(X;Z) + I(Y;Z) \geq I(XY;Z)$$

which is equivalent to $I(X;Y) \geq I(X;Y|Z)$ (see (6)).

Definition 4.2. *The conditional probability distribution p is called* correlation free *if*

$$\sum_{i=1}^{n} I(X_i;Z) \geq I(X_1 \cdots X_n;Z)$$

for any choice of random variables $X_1, \ldots, X_n$ and Z with $P_{Z|X_1 \cdots X_n} := p$.

It turns out that, for these extended definitions, the equivalence between correlation freeness and the multiplicative property of conditional probability distributions still holds.

Theorem 4.3. *A conditional probability distribution*

$$p: (x_1, \ldots, x_n, z) \longmapsto p(z|x_1, \ldots, x_n)$$

is correlation free if and only if it is multiplicative.

Proof. We first show by induction that any multiplicative conditional probability distribution p is correlation free. Let therefore $X_1, \ldots, X_n$ and Z be random variables such that $P_{Z|X_1 \cdots X_n}$ is multiplicative and assume that the implication is proven for probability distributions conditioned on $n-1$ random variables, i.e.,

$$\sum_{i=1}^{n-1} I(X_i;Z) \geq I(X_1 \cdots X_{n-1};Z). \tag{15}$$

Hence,

$$\sum_{i=1}^{n} I(X_i;Z) \geq I(X_1 \cdots X_{n-1};Z) + I(X_n;Z) \geq I(X_1 \cdots X_n;Z)$$

where the last inequality is equivalent to $I(X_1 \cdots X_{n-1}; X_n) \geq I(X_1 \cdots X_{n-1}; X_n|Z)$ (see (6)) and therefore a direct consequence of Theorem 3.1. (Note that if $P_{Z|X_1 \cdots X_n}$ is multiplicative, then $P_{Z|XY}$ for $X = (X_1, \ldots, X_{n-1})$ and $Y = X_n$ is multiplicative as well.) Since (15) is trivially satisfied for $n = 2$, the assertion follows by induction on n.

It remains to be proven that correlation freeness of a probability distribution p implies that p is multiplicative. First, note that for random variables $X_1, \ldots, X_n$ and Z

$$I(X_1 \cdots X_n;Z) \geq \sum_{i=1}^{n} I(X_i;Z). \tag{16}$$

if $X_1, \ldots, X_n$ are mutually independent. Second, recall that $0 = I(X;Y) < I(X;Y|Z)$ for independent and uniformly distributed random variables X and Y if $P_{Z|XY}$ is not multiplicative (see first section of the proof of Theorem 3.1). Again (see (6)), this is equivalent to the inequality

$$I(XY;Z) > I(X;Z) + I(Y;Z). \tag{17}$$

Assume by contradiction that p is not multiplicative, and let $X_1, \ldots, X_n$ be uniformly distributed independent random variables and Z be distributed according to $P_{Z|X_1 \cdots X_n} := p$. Then, there is an index k such that $P_{Z|XY}$ is not multiplicative for $X := X_k$ and $Y := X_1 \cdots X_{k-1} X_{k+1} \cdots X_n$. Hence, from (17) and (16)

$$I(X_1 \cdots X_n;Z) > I(X_k;Z) + I(X_1 \cdots X_{k-1}X_{k+1} \cdots X_n;Z) \geq \sum_{i=1}^{n} I(X_i;Z),$$

i.e., $P_{Z|X_1 \cdots X_n}$ is not correlation free. □

5 Concluding Remarks

We have investigated for which cases the mutual information between arbitrarily distributed random variables X and Y can not be increased when conditioning on additional information Z about X and Y, which is determined by a fixed conditional probability distribution $P_{Z|XY}$. Clearly, Z can be considered as the output of a channel C with two inputs, X and Y, and probability transition matrix $P_{Z|XY}$. Our main theorem gives a necessary and sufficient criterion for the channel C, i.e., for $P_{Z|XY}$, such that $I(X;Y) \geq I(X;Y|Z)$ for all distributions P_{XY}. Furthermore, we have shown that this result can be generalized to channels with more than two inputs.

Combining the main Theorem 3.1 and Lemma 2.3, our criterion and its consequence can be formulated as follows: If for a fixed channel C specified by $P_{Z|XY}$, conditioning on the output Z does not increase the mutual information between independent and uniformly distributed inputs X and Y,[8] then conditioning on the output of C can not increase the mutual information between inputs X and Y having any arbitrary joint distribution P_{XY}.

References

[1] R. E. Blahut. *Principles and Practice of Information Theory.* Addison-Wesley Series in Electrical and Computer Engineering. Addison-Wesley, Reading, Massachusetts, 1990.

[2] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley Series in Telecommunications. Wiley, New York, 1991.

[3] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, 39(3):733–742, 1993.

[4] S. Roman. *Coding and Information Theory.* Graduate Texts in Mathematics. Springer, New York, 1992.

[5] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, 1948.

[8] Since the mutual information between two independent inputs X and Y is zero, this means that $I(X;Y|Z) = 0$.