On the Discontinuity of the Shannon Information Measures

Siu-Wai Ho
Dept. of Information Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
Email: swho4@ie.cuhk.edu.hk

Raymond W. Yeung
Dept. of Information Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
Email: whyeung@ie.cuhk.edu.hk

Abstract—It is well known that the Shannon information measures are continuous functions of the probability distribution when the support is finite. This, however, does not hold when the support is countably infinite. In this paper, we investigate the continuity of the Shannon information measures for countably infinite support. With respect to a distance based on the Kullback-Liebler divergence, we use two different approaches to show that all the Shannon information measures are in fact discontinuous at all probability distributions with countably infinite support.

I. INTRODUCTION

The Shannon information measures are functions mapping a probability distribution into a real value. If the input probability distribution is restricted to a finite support, then it is well known that the Shannon information measures are continuous functions. Harremoës [1] has shown that the Shannon entropy of a probability distribution with countably infinite support can be discontinuous, and he has given conditions for it to be continuous. For a sequence of probability distributions P_n and a fixed probability distribution Q, [1] uses the Kullback-Leibler divergence $D(P_n||Q)$ to define the convergence of P_n to Q. However, if $D(Q||P_n)$ is used instead, results different from [1] can be obtained, and they are discussed in this paper. We use two different approaches to show that all the Shannon information measures are discontinuous at all probability distributions with countably infinite support.

The rest of this paper is organized as follows. In Section II, we will first define the convergence of a sequence of probability distributions P_n to a fixed probability distribution Q with respect to the distance $D(Q||P_n)$. Based on this definition of convergence, a continuous function of a probability distribution is defined. In Section III, we will discuss a class of sequences of probability distributions such that each of them converges to the distribution $\{1, 0, 0, ...\}$, but the entropy of the sequence does not converge to 0. Then entropy will be shown to be discontinuous at all probability distributions with countably infinite support. In Section IV, mutual information will be shown to be discontinuous following two different approaches. By extending these two approaches in Section V, the discontinuity of conditional mutual information will be discussed. Finally, all the Shannon information measures are shown to be discontinuous at all probability distributions with countably infinite support.

II. CONVERGENCE AND CONTINUITY

Definition 1 (Left Convergence): A sequence of probability distributions $\{P_n\}$ is said to **left converge** to a probability distribution Q if for every $\varepsilon > 0$ there exists an integer $K(\varepsilon)$ such that $D(Q||P_n) < \varepsilon$ for all $n \geq K(\varepsilon)$. If $\{P_n\}$ left converges to Q, we write

$$P_n \xrightarrow{D_L} Q,$$

and call Q the **limit** of $\{P_n\}$. When there is no ambiguity, we also write

$$\lim_{n \to \infty} P_n = Q.$$

Two characteristics of the above definition of convergence are noteworthy: a) The Kullback-Leibler distance measure is used instead of the L_1 distance measure. Due to Pinsker's inequality (cf. [2], p.22), if a sequence of probability distributions left converges with respect to the Kullback-Leibler distance measure, then the sequence converges to the same distribution with respect to the L_1 distance measure. Therefore, convergence with respect to the Kullback-Leibler distance measure is stronger than convergence with respect to the L_1 distance measure. b) We use $D(Q||P_n)$ instead of $D(P_n||Q)$ in the definition. Our results are, therefore, different from the results obtained in [1].

We now define the continuity of a function based on the notion of left convergence.

Definition 2: Let $\mathcal{A}$ be a subset of all probability distributions and let $Q \in \mathcal{A}$. Then a function $f : \mathcal{A} \to \mathcal{R}$ is **continuous** at Q if, given any $\varepsilon > 0$, there exists a $\delta > 0$ such that if P is any distribution in $\mathcal{A}$ satisfying $D(Q||P) < \delta$, then $|f(P) - f(Q)| < \varepsilon$.

If f fails to be continuous at Q, then we say that f is **discontinuous** at Q. The following proposition, which is an alternative form of Definition 2, will be used to test the continuity of a function.

Proposition 1: Let $\mathcal{A}$ be a subset of all probability distributions and let $Q \in \mathcal{A}$. Then a function $f : \mathcal{A} \to \mathcal{R}$ is

discontinuous at Q if there exists a sequence $\{P_n\}$ in $\mathcal{A}$ such that

$$P_n \xrightarrow{D_L} Q,$$

but $f(P_n)$ does not converge to $f(Q)$, i.e.

$$\lim_{n \to \infty} f(P_n) \neq f(Q).$$

III. THE DISCONTINUITY OF ENTROPY

In this section, we first discuss a probability distribution $\mathcal{D}_n^{(\alpha,\beta)}$ which will be used to show that entropy is discontinuous at the probability distribution $\nu = \{1, 0, 0, ...\}$. This distribution will subsequently be used to prove that the entropy is discontinuous at all probability distributions with countably infinite support.

For fixed real numbers α and β and an integer n, where $\alpha > 1$, $\beta > 0$, and $n > \alpha$, let $\mathcal{D}_n^{(\alpha,\beta)}$ be a probability distribution such that one of the elements is $1 - (\frac{\log \alpha}{\log n})^\beta$, n of them are $\frac{1}{n}(\frac{\log \alpha}{\log n})^\beta$ and the rest are all 0, i.e.,

$$\mathcal{D}_n^{(\alpha,\beta)} = \left\{ 1 - \left(\frac{\log \alpha}{\log n}\right)^\beta, \frac{1}{n}\left(\frac{\log \alpha}{\log n}\right)^\beta, \frac{1}{n}\left(\frac{\log \alpha}{\log n}\right)^\beta, ..., 0, 0, ... \right\}.$$

We write $\frac{\log \alpha}{\log n}$ instead of $\frac{1}{\log_\alpha n}$ so that all the logarithms in this paper are in the same base. For any given α and β, we will show that the sequence $\{\mathcal{D}_n^{(\alpha,\beta)}\}$ converges to $\nu = \{1, 0, 0, ...\}$.

Consider

$$D(\nu \| \mathcal{D}_n^{(\alpha,\beta)}) = -\log\left(1 - \left(\frac{\log \alpha}{\log n}\right)^\beta\right),$$

which tends to 0 when n tends to infinity. Therefore, for any α and β,

$$\mathcal{D}_n^{(\alpha,\beta)} \xrightarrow{D_L} \nu.$$

The entropy of $\mathcal{D}_n^{(\alpha,\beta)}$ is given by

$$H(\mathcal{D}_n^{(\alpha,\beta)}) = -\left[1 - \left(\frac{\log \alpha}{\log n}\right)^\beta\right] \log\left[1 - \left(\frac{\log \alpha}{\log n}\right)^\beta\right]$$
$$-n\left[\frac{1}{n}\left(\frac{\log \alpha}{\log n}\right)^\beta\right] \log\left[\frac{1}{n}\left(\frac{\log \alpha}{\log n}\right)^\beta\right]. \quad (1)$$

The first term on the R.H.S. in (1) tends to $-1 \cdot \log 1 = 0$ as $n \to \infty$. Rewrite the second term as

$$\left(\frac{\log \alpha}{\log n}\right)^\beta \log n + \left(\frac{\log \alpha}{\log n}\right)^\beta \log\left[\left(\frac{\log n}{\log \alpha}\right)^\beta\right]. \quad (2)$$

As $n \to \infty$, the second term in (2) tends to 0 since $\beta > 0$. For the first term, we have

$$\lim_{n \to \infty} (\log \alpha)^\beta \cdot \frac{\log n}{(\log n)^\beta} = \lim_{n \to \infty} (\log \alpha)^\beta (\log n)^{1-\beta}.$$

Therefore,

$$\lim_{n \to \infty} H(\mathcal{D}_n^{(\alpha,\beta)}) = \begin{cases} 0 & \beta > 1, \\ \log \alpha & \beta = 1, \\ \infty & 0 < \beta < 1. \end{cases} \quad (3)$$

Thus we have the following proposition.

Proposition 2: The Shannon entropy is discontinuous at the probability distribution $\nu = \{1, 0, 0, ...\}$.

Proof: By taking $\alpha > 1$ and $0 < \beta \leq 1$ in (3), we have

$$\lim_{n \to \infty} H(\mathcal{D}_n^{(\alpha,\beta)}) = \infty \neq H(\nu),$$

but

$$\mathcal{D}_n^{(\alpha,\beta)} \xrightarrow{D_L} \nu.$$

By Proposition 1, H is discontinuous at the probability distribution ν. $\blacksquare$

Now, we are going to prove that entropy is discontinuous not only at ν but also at any probability distribution. Consider $\mathcal{D}_n = \mathcal{D}_n^{(\alpha,0.5)}$ so that

$$\mathcal{D}_n = \left\{ 1 - \frac{1}{\sqrt{\log_\alpha n}}, \frac{1}{n\sqrt{\log_\alpha n}}, \frac{1}{n\sqrt{\log_\alpha n}}, ..., 0, ... \right\} (4)$$

and $\lim_{n \to \infty} H(\mathcal{D}_n) = \infty$ from (3). In (4), since α, the base of the logarithms, can be any real number strictly larger than 1, we can simply write

$$\mathcal{D}_n = \left\{ 1 - \frac{1}{\sqrt{\log n}}, \frac{1}{n\sqrt{\log n}}, \frac{1}{n\sqrt{\log n}}, ..., 0, 0, ... \right\}. \quad (5)$$

Theorem 3: For any probability distribution $\mathcal{P}^0 = \{p_0, p_1, ...\}$ with countably infinite support such that $H(\mathcal{P}^0) < \infty$, there exists a sequence of probability distributions $\mathcal{P}_n$ such that $H(\mathcal{P}_n)$ is bounded for integers $n \geq 2$ and $\mathcal{P}_n \xrightarrow{D_L} \mathcal{P}^0$ but $\lim_{n \to \infty} H(\mathcal{P}_n) = \infty$ and $H(\mathcal{P})$ is discontinuous at $\mathcal{P}^0$.

Proof: In Proposition 2, we have shown the claim in the theorem at the probability distribution ν and so we assume $\mathcal{P}^0 \neq \nu$. Then, without loss of generality, we can assume that $0 < p_0 < 1$. Let $q = 1 - p_0$ and

$$\mathcal{P}^1 = \{0, q^{-1}p_1, q^{-1}p_2, q^{-1}p_3, ...\},$$

which is seen to be a probability distribution.

Recall the definition of $\mathcal{D}_n$ in (5). Let the probability distributions of random variables V and W_n be $\mathcal{P}^1$ and $\mathcal{D}_n$, respectively. Let the probability distribution of an independent binary random variable Z be such that $p(Z = 0) = q$ and $p(Z = 1) = 1 - q$. By letting

$$X_n = \begin{cases} V & \text{if } Z = 0 \\ W_n & \text{if } Z = 1, \end{cases}$$

the probability distribution of X_n for $n \geq 2$ is given by

$$\mathcal{P}_n^2 = \left\{ (1-q) - \frac{1-q}{\sqrt{\log n}}, p_1 + \frac{1-q}{n\sqrt{\log n}}, \right.$$
$$\left. p_2 + \frac{1-q}{n\sqrt{\log n}}, ..., p_n + \frac{1-q}{n\sqrt{\log n}}, p_{n+1}, p_{n+2}, ... \right\}$$
$$= \left\{ p_0 - \frac{1-q}{\sqrt{\log n}}, p_1 + \frac{1-q}{n\sqrt{\log n}}, p_2 + \frac{1-q}{n\sqrt{\log n}}, \right.$$
$$\left. ..., p_n + \frac{1-q}{n\sqrt{\log n}}, p_{n+1}, p_{n+2}, ... \right\}. \quad (6)$$

Consider

$$D(\mathcal{P}^0||\mathcal{P}_n^2) = p_0 \log \frac{p_0}{p_0 - \frac{1-q}{\sqrt{\log n}}} + \sum_{i=1}^{n} p_i \log \frac{p_i}{p_i + \frac{1-q}{n\sqrt{\log n}}}$$
$$+ \sum_{i=n+1}^{\infty} p_i \log \frac{p_i}{p_i}$$
$$= p_0 \log \frac{p_0}{p_0 - \frac{1-q}{\sqrt{\log n}}} + \sum_{i=1}^{n} p_i \log \frac{p_i}{p_i + \frac{1-q}{n\sqrt{\log n}}}.$$

The first term tends to zero when $n \to \infty$. Without loss of generality, we assume that $p_i > 0$ for all i (otherwise we simply omit all i such that $p_i = 0$ in a summation). Let

$$J_n = \sum_{i=1}^{n} p_i \log \frac{p_i}{p_i + \frac{1-q}{n\sqrt{\log n}}},$$

which is less than 0 for all n. By the log-sum inequality (see, for example, [2]), we have

$$J_n \geq \left(\sum_{i=1}^{n} p_i\right) \log \frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} \left(p_i + \frac{1-q}{n\sqrt{\log n}}\right)}$$
$$= \left(\sum_{i=1}^{n} p_i\right) \log \frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} p_i + \frac{1-q}{\sqrt{\log n}}}.$$

When $n \to \infty$, we have

$$\lim_{n \to \infty} J_n \geq \lim_{n \to \infty} (\sum_{i=1}^{n} p_i) \log \frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} p_i + \frac{1-q}{\sqrt{\log n}}}$$
$$= (1 - p_0) \cdot \log \frac{(1 - p_0)}{(1 - p_0) + 0}$$
$$= 0.$$

Therefore, $\lim_{n \to \infty} J_n = 0$. Thus we have proved that

$$\lim_{n \to \infty} D(\mathcal{P}^0||\mathcal{P}_n^2) = 0,$$

which means that

$$\mathcal{P}_n^2 \xrightarrow{D_L} \mathcal{P}^0.$$

On the other hand,

$$\lim_{n \to \infty} H(\mathcal{P}_n^2) = \lim_{n \to \infty} H(X_n)$$
$$\geq \lim_{n \to \infty} H(X_n|Z)$$
$$= \lim_{n \to \infty} (qH(V) + (1-q)H(W_n))$$
$$= qH(V) + (1-q) \lim_{n \to \infty} H(\mathcal{D}_n)$$
$$= \infty \qquad (7)$$
$$> H(\mathcal{P}^0).$$

By Proposition 1, we conclude that H is discontinuous at $\mathcal{P}^0$.

By comparing $\mathcal{P}^0$ and $\mathcal{P}_n^2$, the upper bound for $H(\mathcal{P}_n^2)$ is given by

$$H(\mathcal{P}_n^2) \leq \log(n+1) + H(\mathcal{P}^0).$$

Therefore $H(\mathcal{P}_n^2)$ is bounded for integers $n \geq 2$. ■

IV. THE DISCONTINUITY OF MUTUAL INFORMATION

In this section, the discontinuity of mutual information is discussed. Let $\tilde{\mathcal{P}}^0 = \{p(xy)\}$ be a joint probability distribution for random variables X and Y where $p(xy)$ is the probability that X equals x and Y equals y. We can assume without loss of generality that $p(00) > 0$, because if $p(00) = 0$, we can always make $p(00) > 0$ with an appropriate reindexing of the alphabets of X and Y.

Let $q = 1 - p(00)$ and

$$\tilde{\mathcal{P}}^1 = \begin{bmatrix} 0 & q^{-1}p(10) & \cdots \\ q^{-1}p(01) & q^{-1}p(11) & \cdots \\ \vdots & \vdots & \end{bmatrix}.$$

By letting $\tilde{\mathcal{D}}_n$ be a diagonal matrix with diagonal equal to the distribution $\mathcal{D}_n$ in (5), we have

$$\tilde{\mathcal{D}}_n = \begin{bmatrix} 1 - \frac{1}{\sqrt{\log n}} & 0 & 0 & \cdots \\ 0 & \frac{1}{n\sqrt{\log n}} & 0 & \cdots \\ 0 & 0 & \frac{1}{n\sqrt{\log n}} & \\ \vdots & \vdots & & \ddots \end{bmatrix}.$$

Let the probability distribution of an independent binary random variable Z be such that $p(Z = 0) = q$ and $p(Z = 1) = 1 - q$. By letting the joint probability distribution of (X, Y) be $\tilde{\mathcal{P}}_n^2$ for $n \geq 2$ and

$$\tilde{\mathcal{P}}_n^2 = \begin{cases} \tilde{\mathcal{P}}^1 & \text{if } Z = 0 \\ \tilde{\mathcal{D}}_n & \text{if } Z = 1 \end{cases}$$
$$= \begin{bmatrix} 1 - q - \frac{1-q}{\sqrt{\log n}} & p(10) & \cdots \\ p(01) & p(11) + \frac{1-q}{n\sqrt{\log n}} & \\ \vdots & & \ddots \end{bmatrix} \quad (8)$$

Theorem 4: Let $\tilde{\mathcal{P}}^0$ be a joint probability distribution for random variables X and Y with countably infinite support for both of the marginal probability distributions such that $I_{X;Y}(\tilde{\mathcal{P}}^0) < \infty$. There exists a sequence of probability distributions $\tilde{\mathcal{P}}_n$ such that $I_{X;Y}(\tilde{\mathcal{P}}_n)$ is finite and $\tilde{\mathcal{P}}_n \xrightarrow{D_L} \tilde{\mathcal{P}}^0$ but $\lim_{n \to \infty} I_{X;Y}(\tilde{\mathcal{P}}_n) = \infty$. Thus, $I_{X;Y}(\cdot)$ is discontinuous at $\tilde{\mathcal{P}}^0$.

Proof: By the same argument used in the proof of

$$\mathcal{P}_n^2 \xrightarrow{D_L} \mathcal{P}^0,$$

we can see that

$$\tilde{\mathcal{P}}_n^2 \xrightarrow{D_L} \tilde{\mathcal{P}}^0.$$

On the other hand, it can be shown that for a joint probability distribution with finite support,

$$I(X;Y) = qI(X;Y|Z=0) + (1-q)I(X;Y|Z=1)$$
$$+ I(X;Z) - I(X;Z|Y).$$

For joint probability distribution with countably infinite support, the summations in $I(X;Z|Y)$, $I(X;Z)$ and $I(X;Y|Z=0)$ are bounded by

$$I_{X;Z|Y}(\tilde{\mathcal{P}}_n^2) \leq \log 2,$$
$$I_{X;Z}(\tilde{\mathcal{P}}_n^2) \geq 0,$$
$$qI_{X;Y|Z=0}(\tilde{\mathcal{P}}_n^2) = qI_{X;Y}(\tilde{\mathcal{P}}^1) \geq 0,$$

and

$$\begin{aligned}(1-q)I_{X;Y|Z=1}(\tilde{\mathcal{P}}_n^2) &= (1-q)I_{X;Y}(\tilde{\mathcal{D}}_n) \\ &= (1-q)H(\mathcal{D}_n).\end{aligned}$$

Thus

$$\begin{aligned}I_{X;Y}(\tilde{\mathcal{P}}_n^2) &\geq 0 + (1-q)H(\mathcal{D}_n) + 0 - \log 2 \\ &= (1-q)H(\mathcal{D}_n) - \log 2.\end{aligned}$$

Therefore

$$\begin{aligned}\lim_{n\to\infty} I_{X;Y}(\tilde{\mathcal{P}}_n^2) &\geq \lim_{n\to\infty}(1-q)H(\mathcal{D}_n) - \log 2 \\ &= \infty.\end{aligned}$$

By Proposition 1, the function $I_{X;Y}(\tilde{\mathcal{P}})$ is discontinuous at the distribution $\tilde{\mathcal{P}}^0$.

By comparing $\tilde{\mathcal{P}}_n^2$ in (8) with $\tilde{\mathcal{P}}^0$, we can obtain an upper bound

$$I_{X;Y}(\tilde{\mathcal{P}}_n^2) \leq (4n+6)e^{-1}\log e + I_{X;Y}(\tilde{\mathcal{P}}^0)$$

for all integers n. The bound is not the tightest but it is enough to show that $I_{X;Y}(\tilde{\mathcal{P}}_n^2)$ is bounded for $n \geq 2$. $\blacksquare$

In the above theorem, we have constructed a sequence of probability distributions whose mutual information tends to infinity when $n \to \infty$ but the mutual information of each distribution inside the sequence is finite. In the next theorem, we will resort to a different method to show that mutual information is discontinuous.

Theorem 5: Let $\tilde{\mathcal{P}}^0$ be a joint probability distribution for random variables X and Y with countably infinite support for both of the marginal probability distributions such that $I_{X;Y}(\tilde{\mathcal{P}}^0) < \infty$. There exists a sequence of probability distributions $\tilde{\mathcal{P}}_n$ such that $\tilde{\mathcal{P}}_n \xrightarrow{D_L} \tilde{\mathcal{P}}^0$ but $I_{X;Y}(\tilde{\mathcal{P}}_n) = \infty$ for all integers n. Thus $I_{X;Y}(\cdot)$ is discontinuous at $\tilde{\mathcal{P}}^0$.

Proof: Let $\tilde{\Phi}$ be a joint probability distribution with $I(\tilde{\Phi}) = \infty$. An example is a diagonal matrix with elements equal to the distribution given in Example 2.46 in [2]. Let $\tilde{\phi} = 0.5\tilde{\mathcal{P}}^0 + 0.5\tilde{\Phi}$ and note that $D(\tilde{\mathcal{P}}^0||\tilde{\phi}) \leq \log 2$. Let the distribution of a binary random variable S_n be $(1-\frac{1}{n}, \frac{1}{n})$. Then a sequence of probability distributions $\tilde{\mathcal{P}}^n$ is constructed by letting (X,Y) be distributed according to $\tilde{\mathcal{P}}^0$ if $S_n = 0$ and according to $\tilde{\phi}$ if $S_n = 1$. We have

$$\tilde{\mathcal{P}}^n = (1-\frac{1}{n})\tilde{\mathcal{P}}^0 + \frac{1}{n}\tilde{\phi}.$$

By the convexity of the Kullback-Leibler distance, we have

$$\begin{aligned}&\lim_{n\to\infty} D(\tilde{\mathcal{P}}^0||\tilde{\mathcal{P}}^n) \\ &\leq \lim_{n\to\infty}(1-\frac{1}{n})D(\tilde{\mathcal{P}}^0||\tilde{\mathcal{P}}^0) + \frac{1}{n}D(\tilde{\mathcal{P}}^0||\tilde{\phi}) \\ &= 0.\end{aligned}$$

Therefore,

$$\tilde{\mathcal{P}}^n \xrightarrow{D_L} \tilde{\mathcal{P}}^0.$$

On the other hand, it can be shown that for a joint probability distribution with finite support,

$$I(X;Y) = I(X;Y|S_n) + I(X;S_n) - I(X;S_n|Y).$$

For a joint probability distribution with countably infinite support, the summations in $I(X;S_n|Y)$ and $I(X;S_n)$ are bounded by

$$I_{X;S_n|Y}(\tilde{\mathcal{P}}^n) \leq \log 2,$$

and

$$I_{X;S_n}(\tilde{\mathcal{P}}^n) \geq 0.$$

Since

$$Pr\{S_n = 0\}I_{X;Y|S_n=0}(\tilde{\mathcal{P}}^n) = Pr\{S_n = 0\}I_{X;Y}(\tilde{\mathcal{P}}^0)$$

and

$$Pr\{S_n = 1\}I_{X;Y|S_n=1}(\tilde{\mathcal{P}}^n) = Pr\{S_n = 1\}I_{X;Y}(\tilde{\phi}) = \infty,$$

we have

$$I_{X;Y|S_n}(\tilde{\mathcal{P}}^n) = \infty.$$

Thus

$$I_{X;Y}(\tilde{\mathcal{P}}^n) \geq \infty + 0 - \log 2 = \infty, \qquad (9)$$

for all integers n. Therefore,

$$\lim_{n\to\infty} I_{X;Y}(\tilde{\mathcal{P}}^n) = \infty \neq I_{X;Y}(\tilde{\mathcal{P}}^0).$$

By Proposition 1, the function $I_{X;Y}(\tilde{\mathcal{P}})$ is discontinuous at the distribution $\tilde{\mathcal{P}}^0$. $\blacksquare$

V. THE DISCONTINUITY OF CONDITIONAL MUTUAL INFORMATION AND THE SHANNON INFORMATION MEASURES

In this section, we extend the results obtained in the previous sections to conclude that all Shannon information measures are, indeed, discontinuous for any probability distribution with countably infinite support.

Theorem 6: For any joint probability distribution $\hat{\mathcal{P}}^0 = \{p(xyz)\}$ for random variables X, Y and Z with countably infinite support for the marginal probability distributions of X and Y such that $I_{X;Y|Z}(\hat{\mathcal{P}}^0) < \infty$, there exists a sequence of probability distributions $\hat{\mathcal{P}}_n$ such that $I_{X;Y|Z}(\hat{\mathcal{P}}_n)$ is finite and $\hat{\mathcal{P}}_n \xrightarrow{D_L} \hat{\mathcal{P}}^0$ but $\lim_{n\to\infty} I_{X;Y|Z}(\hat{\mathcal{P}}_n) = \infty$. Thus, the conditional mutual information $I_{X;Y|Z}(\cdot)$ is discontinuous at $\hat{\mathcal{P}}^0$.

Proof: Let the marginal distribution of Z be $\{p_Z(z)\}$. Without loss of generality, we assume that $p_Z(0) > 0$ and let $\gamma = p_Z(0)$. Since

$$I_{X;Y|Z}(\hat{\mathcal{P}}^0) = \sum_z p(z) I_{X;Y|Z=z}(\hat{\mathcal{P}}^0),$$

which is bounded and $\gamma > 0$ so that $I_{X;Y|Z=0}(\hat{\mathcal{P}}^0)$ is also bounded. In other words, the mutual information of the distribution

$$\tilde{\mathcal{P}}^0 = \begin{bmatrix} \gamma^{-1}p(000) & \gamma^{-1}p(100) & \cdots \\ \gamma^{-1}p(010) & \gamma^{-1}p(110) & \cdots \\ \vdots & \vdots & \end{bmatrix},$$

is bounded. Then by Theorem 4, we can construct a probability distribution $\tilde{\mathcal{P}}_n^2$ such that

$$\tilde{\mathcal{P}}_n^2 \xrightarrow{D_L} \tilde{\mathcal{P}}^0,$$

but

$$\lim_{n \to \infty} I_{X;Y}(\tilde{\mathcal{P}}_n^2) = \infty.$$

Replace $\{p(xy0)\}$ in $\hat{\mathcal{P}}^0$ by $\gamma \cdot \tilde{\mathcal{P}}_n^2$ and let the resulting joint probability distribution be $\hat{\mathcal{P}}_n^2$. Then

$$\hat{\mathcal{P}}_n^2 \xrightarrow{D_L} \hat{\mathcal{P}}^0,$$

but

$$\lim_{n \to \infty} I_{X;Y|Z}(\hat{\mathcal{P}}_n^2) = \infty$$
$$\neq I_{X;Y|Z}(\hat{\mathcal{P}}^0).$$

Thus the conditional mutual information $I_{X;Y|Z}(\cdot)$ is discontinuous at $\hat{\mathcal{P}}^0$. ∎

Note that conditional entropy is a special case of conditional mutual information. The discontinuity of entropy, mutual information and conditional mutual information has been carried out in Theorem 3, Theorem 4 and Theorem 6 respectively. Thus, for any Shannon information measure $\mathcal{H}(\mathcal{P})$ and for any probability distribution $\mathcal{P}^0$ with countably infinite support such that $\mathcal{H}(\mathcal{P}^0) < \infty$, there exists a sequence of probability distributions $\mathcal{P}_n$ such that $\mathcal{H}(\mathcal{P}_n)$ is finite and $\mathcal{P}_n \xrightarrow{D_L} \mathcal{P}^0$ but $\lim_{n \to \infty} \mathcal{H}(\mathcal{P}_n) = \infty$ and $\mathcal{H}(\mathcal{P})$ is discontinuous at $\mathcal{P}^0$.

By a method similar to the one used in the proof of Theorem 6, the results in Theorem 5 can be extended to conditional mutual information.

Theorem 7: For any joint probability distribution $\hat{\mathcal{P}}^0 = \{p(xyz)\}$ for random variables X, Y and Z with countably infinite support for the marginal probability distributions of X and Y such that $I_{X;Y|Z}(\hat{\mathcal{P}}^0) < \infty$, there exists a sequence of probability distributions $\hat{\mathcal{P}}_n$ such that $\hat{\mathcal{P}}_n \xrightarrow{D_L} \hat{\mathcal{P}}^0$ but $I_{X;Y|Z}(\hat{\mathcal{P}}_n) = \infty$ for all n. Thus, the conditional mutual information $I_{X;Y|Z}(\cdot)$ is discontinuous at $\hat{\mathcal{P}}^0$.

Proof: By the same setup of γ and $\tilde{\mathcal{P}}^0$ in the proof of Theorem 6 and by Theorem 5, we can construct a probability distribution $\tilde{\mathcal{P}}_n^2$ such that

$$\tilde{\mathcal{P}}_n^2 \xrightarrow{D_L} \tilde{\mathcal{P}}^0,$$

but for all integers n,

$$I_{X;Y}(\tilde{\mathcal{P}}_n^2) = \infty.$$

Again replace $\{p(xy0)\}$ in $\hat{\mathcal{P}}^0$ by $\gamma \cdot \tilde{\mathcal{P}}_n^2$ and let the resulting joint probability distribution be $\hat{\mathcal{P}}_n^2$. Then

$$\hat{\mathcal{P}}_n^2 \xrightarrow{D_L} \hat{\mathcal{P}}^0,$$

but for all integers n,

$$I_{X;Y|Z}(\hat{\mathcal{P}}_n^2) = \infty$$
$$\lim_{n \to \infty} I_{X;Y|Z}(\hat{\mathcal{P}}_n^2) = \infty$$
$$\neq I_{X;Y|Z}(\hat{\mathcal{P}}^0).$$

Thus the conditional mutual information $I_{X;Y|Z}(\cdot)$ is discontinuous at $\hat{\mathcal{P}}^0$. ∎

Note that entropy and conditional entropy are special cases of mutual information and conditional mutual information. The discontinuity of mutual information and conditional mutual information has been presented in Theorem 5 and Theorem 7 respectively. Thus, for any Shannon information measure $\mathcal{H}(\mathcal{P})$ and for any probability distribution $\mathcal{P}^0$ with countably infinite support such that $\mathcal{H}(\mathcal{P}^0) < \infty$, there exists a sequence of probability distributions $\mathcal{P}_n$ such that $\mathcal{P}_n \xrightarrow{D_L} \mathcal{P}^0$ but $\mathcal{H}(\mathcal{P}_n) = \infty$ for all integers n and $\mathcal{H}(\mathcal{P})$ is discontinuous at $\mathcal{P}^0$.

VI. Conclusion

Using "left convergence" of a sequence of discrete probability distributions, two different approaches have demonstrated that all the Shannon information measures are discontinuous at every distribution with countably infinite support.

References

[1] Peter Harremoës, "Information Topologies with Applications," vol. Special volume of Bolyi Series. Springer. To appear.
[2] Raymond W. Yeung, *A First Course in Information Theory*, Kluwer Academic/Plenum Publishers, 2002.