Generalized Source Coding Theorems and Hypothesis Testing: Part I – Information Measures

Po-Ning Chen
Dept. of Communications Engineering
National Chiao Tung University
1001, Ta-Hsueh Road, Hsin Chu
Taiwan 30050, R.O.C.

Fady Alajaji
Dept. of Mathematics and Statistics
Queen's University
Kingston, Ontario K7L 3N6
Canada

Key Words: entropy, mutual information, divergence, ε-capacity

Abstract

Expressions for ε-entropy rate, ε-mutual information rate and ε-divergence rate are introduced. These quantities, which consist of the quantiles of the asymptotic information spectra, generalize the inf/sup-entropy/information/divergence rates of Han and Verdú. The algebraic properties of these information measures are rigorously analyzed, and examples illustrating their use in the computation of the ε-capacity are presented. In Part II of this work, these measures are employed to prove general source coding theorems for block codes and the general formula of the Neyman-Pearson hypothesis testing type-II error exponent subject to upper bounds on the type-I error probability.

I. Introduction and Motivation

Entropy, divergence and mutual information are without a doubt the most important information theoretic quantities. They constitute the fundamental measures upon which information theory is founded. Given a discrete random variable X with distribution P_X, its entropy is defined by [3]

$$H(X) \stackrel{\triangle}{=} -\sum_x P_X(x) \log_2 P_X(x) = E_{P_X}\left[-\log P_X(X)\right].$$

$H(X)$ is a measure of the average amount of uncertainty in X. The divergence, on the other hand, measures the relative distance between the distributions of two random variables X and $\hat{X}$ that are defined on the same alphabet :

$$D(X\|\hat{X}) \stackrel{\triangle}{=} E_{P_X}\left[\log_2 \frac{P_X(X)}{P_{\hat{X}}(X)}\right].$$

As for the mutual information $I(X;Y)$ between random variables X and Y, it represents the average amount of information that Y contains about X. It is defined as the divergence between the joint distribution P_{XY} and the product distribution $P_X P_Y$:

$$I(X;Y) \stackrel{\triangle}{=} D(P_{XY}\|P_X P_Y) = E_{P_{XY}}\left[\log_2 \frac{P_{XY}(X,Y)}{P_X(X)P_Y(Y)}\right].$$

More generally, consider an input process $\boldsymbol{X}$ defined by a sequence of finite dimensional distributions [11]: $\boldsymbol{X} \stackrel{\triangle}{=} \{X^n = (X_1^{(n)}, \ldots, X_n^{(n)})\}_{n=1}^{\infty}$. Let $\boldsymbol{Y} \stackrel{\triangle}{=} \{Y^n = (Y_1^{(n)}, \ldots, Y_n^{(n)})\}_{n=1}^{\infty}$ be the corresponding output process induced by $\boldsymbol{X}$ via the channel $\boldsymbol{W} \stackrel{\triangle}{=} \{W^n = P_{Y^n|X^n} : \mathcal{X}^n \to \mathcal{Y}^n\}_{n=1}^{\infty}$, which is an arbitrary sequence of n-dimensional conditional distributions from $\mathcal{X}^n$ to $\mathcal{Y}^n$, where $\mathcal{X}$ and $\mathcal{Y}$ are the input and output alphabets respectively. The entropy rate for the source $\boldsymbol{X}$ is defined by [2], [3]

$$H(\boldsymbol{X}) \stackrel{\triangle}{=} \lim_{n\to\infty} \frac{1}{n} E\left[-\log P_{X^n}(X^n)\right],$$

assuming the limit exists. Similarly the expressions for the divergence and mutual information rates are given by

$$D(\boldsymbol{X}\|\hat{\boldsymbol{X}}) \stackrel{\triangle}{=} \lim_{n\to\infty} \frac{1}{n} E\left[\log \frac{P_{X^n}(X^n)}{P_{\hat{X}^n}(X^n)}\right],$$

and

$$I(\boldsymbol{X};\boldsymbol{Y}) \triangleq \lim_{n\to\infty} \frac{1}{n} E\left[\log \frac{P_{X^nY^n}(X^n,Y^n)}{(P_{X^n}(X^n)P_{Y^n})(Y^n)}\right],$$

respectively.

The above quantities have an operational significance established via Shannon's coding theorems when the stochastic systems under consideration satisfy certain *regularity* conditions (such as stationarity and ergodicity, or information stability) [9], [11]. However, in more complicated situations such as when the systems are non-stationary (with time-varying statistics), these information rates are no longer valid and lose their operational significance. This results in the need to establish new information measures which appropriately characterize the operational limits of arbitrary stochastic systems.

This is achieved in [10] and [11] where Han and Verdú introduce the notions of *inf/sup-entropy/information rates* and illustrate the key role these information measures play in proving a general lossless (block) source coding theorem and a general channel coding theorem. More specifically, they demonstrate that for an arbitrary finite-alphabet source $\boldsymbol{X}$, the expression for the minimum achievable (block) source coding rate is given by the *sup-entropy rate* $\bar{H}(\boldsymbol{X})$, defined as the limsup in probability of $(1/n)\log 1/P_{X^n}(X^n)$ [10]. They also establish in [11] the formulas of the ε-capacity C_ε and capacity[1] C of arbitrary single-user channels without feedback (not necessarily information stable, stationary, ergodic, etc.). More specifically, they show that

$$\sup_{\boldsymbol{X}} \sup\{R : F_{\boldsymbol{X}}(R) < \varepsilon\} \le C_\varepsilon \le \sup_{\boldsymbol{X}} \sup\{R : F_{\boldsymbol{X}}(R) \le \varepsilon\},$$

and

$$C = \sup_{\boldsymbol{X}} \underline{I}(\boldsymbol{X};\boldsymbol{Y}),$$

[1] *Definition* ([8],[11]): Given $0 < \varepsilon < 1$, an (n, M, ϵ) code for the channel $\boldsymbol{W}$ has blocklength n, M codewords and average (decoding) error probability not larger than ε. A non-negative number R is an ε-*achievable rate* if for every $\delta > 0$, there exist, for all n sufficiently large, (n, M, ϵ) codes with rate $\frac{1}{n}\log M > R - \delta$. The *supremum* of all ε-achievable rates is called the ε-*capacity*, C_ε. The *capacity* C is the supremum of rates that are ε-achievable for all $0 < \varepsilon < 1$ and hence $C = \lim_{\varepsilon\downarrow 0} C_\varepsilon$.

In other words, C_ε is the largest rate at which information can be conveyed over the channel such that the probability of decoding error is below a fixed threshold ε, for sufficiently large blocklengths. Furthermore, C represents the largest rate at which information can be transmitted over the channel with asymptotically *vanishing* error probability.

where

$$F_{\boldsymbol{X}}(R) \stackrel{\triangle}{=} \limsup_{n\to\infty} Pr[(1/n)\ i_{X^nY^n}(X^n;Y^n) \leq R],$$

$(1/n)\ i_{X^nY^n}(X^n;Y^n)$ is the sequence of normalized information densities defined by

$$i_{X^nY^n}(x^n;y^n) = \log\ \frac{P_{Y^n|X^n}(y^n|x^n)}{P_{Y^n}(y^n)},$$

and $\underline{I}(\boldsymbol{X};\boldsymbol{Y})$ is *inf-information rate* between $\boldsymbol{X}$ and $\boldsymbol{Y}$, which is defined as the *liminf in probability* of $(1/n)\ i_{X^nY^n}(X^n;Y^n)$.

By adopting the same technique as in [10] (also in [11]), general expressions for the capacity of single-user channels with feedback and for Neyman-Pearson type-II error exponents are derived in [5] and [4], respectively. Furthermore, an application of the type-II error exponent formula to the non-feedback and feedback channel reliability functions is demonstrated in [4] and [6].

The above inf/sup-entropy/information rates are expressed in terms of the *liminf/limsup in probability* of the normalized entropy/information densities. The *liminf in probability* of a sequence of random variables is defined as follows [10]: if A_n is a sequence of random variables, then its *liminf in probability* is the largest extended real number $\underline{U}$ such that for all $\xi > 0$,

$$\lim_{n\to\infty} Pr[A_n \leq \underline{U} - \xi] = 0. \tag{1.1}$$

Similarly, its *limsup in probability* is the smallest extended real number $\bar{U}$ such that for all $\xi > 0$,

$$\lim_{n\to\infty} Pr[A_n \geq \bar{U} + \xi] = 0. \tag{1.2}$$

Note that these two quantities are always defined; if they are equal, then the sequence of random variables converges in probability to a constant.

It is straightforward to deduce that equations (1.1) and (1.2) are respectively equivalent to

$$\liminf_{n\to\infty} Pr[A_n \leq \underline{U} - \xi] = \limsup_{n\to\infty} Pr[A_n \leq \underline{U} - \xi] = 0, \tag{1.3}$$

and

$$\liminf_{n\to\infty} Pr[A_n \geq \bar{U} + \xi] = \limsup_{n\to\infty} Pr[A_n \geq \bar{U} + \xi] = 0. \tag{1.4}$$

We can observe however that there might exist cases of interest where *only* the liminfs of the probabilities in (1.3) and (1.4) are equal to zero; while the limsups do *not* vanish. There are also other cases where *both* the liminfs and limsups in (1.3)-(1.4) do not vanish; but they are upper bounded by a prescribed threshold. Furthermore, there are situations where the interval $[\underline{U}, \bar{U}]$ does not contain only one point; for e.g., when A_n converges in distribution to another random variable. Hence, those points within the interval $[\underline{U}, \bar{U}]$ might possess a Shannon-theoretic operational meaning when for example A_n consists of the normalized entropy density of a given source.

The above remarks constitute the motivation for this work in which we generalize Han and Verdu's information rates and prove general data compression and hypothesis testing theorems that are the *counterparts* of their ε-capacity channel coding theorem [11].

In Part I, we propose generalized versions of the inf/sup-entropy/information/divergence rates. We analyze in detail the algebraic properties of these information measures, and we illustrate their use in the computation of the ε-capacity of arbitrary additive-noise channels. In Part II of this paper [7], we utilize these quantities to establish general source coding theorems for arbitrary finite-alphabet sources, and the general expression of the Neyman-Pearson type-II error exponent.

II. Generalized Information Measures

Definition 2.1 (Inf/sup-spectrum) If $\{A_n\}_{n=1}^{\infty}$ is a sequence of random variables, then its *inf-spectrum* $\underline{u}(\cdot)$ and its *sup-spectrum* $\bar{u}(\cdot)$ are defined by

$$\underline{u}(\theta) \triangleq \liminf_{n\to\infty} Pr\{A_n \leq \theta\},$$

and

$$\bar{u}(\theta) \triangleq \limsup_{n\to\infty} Pr\{A_n \leq \theta\}.$$

In other words, $\underline{u}(\cdot)$ and $\bar{u}(\cdot)$ are respectively the liminf and the limsup of the cumulative distribution function (CDF) of A_n. Note that by definition, the CDF of A_n – $Pr\{A_n \leq \theta\}$ – is non-decreasing and right-continuous. However, for $\underline{u}(\cdot)$ and $\bar{u}(\cdot)$, only the *non-decreasing* property remains [2].

Definition 2.2 (Quantile of inf/sup-spectrum) For any $0 \leq \delta \leq 1$, the *quantiles* $\underline{U}_\delta$ and $\bar{U}_\delta$ of the sup-pectrum and the inf-spectrum are defined by[3]

$$\underline{U}_\delta \triangleq \begin{cases} -\infty, & \text{if } \{\theta : \bar{u}(\theta) \leq \delta\} = \emptyset, \\ \sup\{\theta : \bar{u}(\theta) \leq \delta\}, & \text{otherwise,} \end{cases}$$

and

$$\bar{U}_\delta \triangleq \begin{cases} -\infty, & \text{if } \{\theta : \underline{u}(\theta) \leq \delta\} = \emptyset, \\ \sup\{\theta : \underline{u}(\theta) \leq \delta\}, & \text{otherwise,} \end{cases}$$

respectively. If follows from the above definitions that $\underline{U}_\delta$ and $\bar{U}_\delta$ are right-continuous and non-decreasing in δ.

Note that the liminf in probability $\underline{U}$ and the limsup in probability $\bar{U}$ of A_n satisfy

$$\underline{U} = \underline{U}_0,$$

and

$$\bar{U} = \bar{U}_{1^-},$$

respectively, where the superscript "-" denotes a strict inequality in the definition of $\bar{U}_{1^-}$; i.e.,

$$\bar{U}_{\delta^-} \triangleq \sup\{\theta : \underline{u}(\theta) < \delta\}.$$

Note also that

$$\underline{U} \leq \underline{U}_\delta \leq \bar{U}_\delta \leq \bar{U}.$$

Remark that $\underline{U}_\delta$ and $\bar{U}_\delta$ always exist. Furthermore, if $\underline{U}_\delta = \bar{U}_\delta \ \forall \ \delta \in [0, 1]$, then the sequence of random variables A_n converges in distribution to a random variable A, provided the distribution sequence of A_n is tight.

[2] It is pertinent to also point out that even if we do not require right-continuity as a fundamental property of a CDF, the spectrums $\underline{u}(\cdot)$ and $\bar{u}(\cdot)$ are *not* necessarily legitimate CDFs of (conventional real-valued) random variables since there might exist cases where the "probability mass escapes to infinity" (cf. [1, page 346]). A necessary and sufficient condition for $\underline{u}(\cdot)$ and $\bar{u}(\cdot)$ to be conventional CDFs (without requiring right-continuity) is that the sequence of distribution functions of A_n be *tight* [1, page 346]. Tightness is actually guaranteed if the alphabet of A_n is finite.

[3] Note that the usual definition of the quantile function $\phi(\delta)$ of a non-decreasing function $F(\cdot)$ is slightly different from our definition [1, page 190]: $\phi(\delta) = \sup\{\theta : F(\theta) < \delta\}$. Remark that if $F(\cdot)$ is strictly increasing, then the quantile is nothing but the inverse of $F(\cdot)$: $\phi(\delta) = F^{-1}(\delta)$.

For a better understanding of the quantities defined above, we depict them in Figure 1.

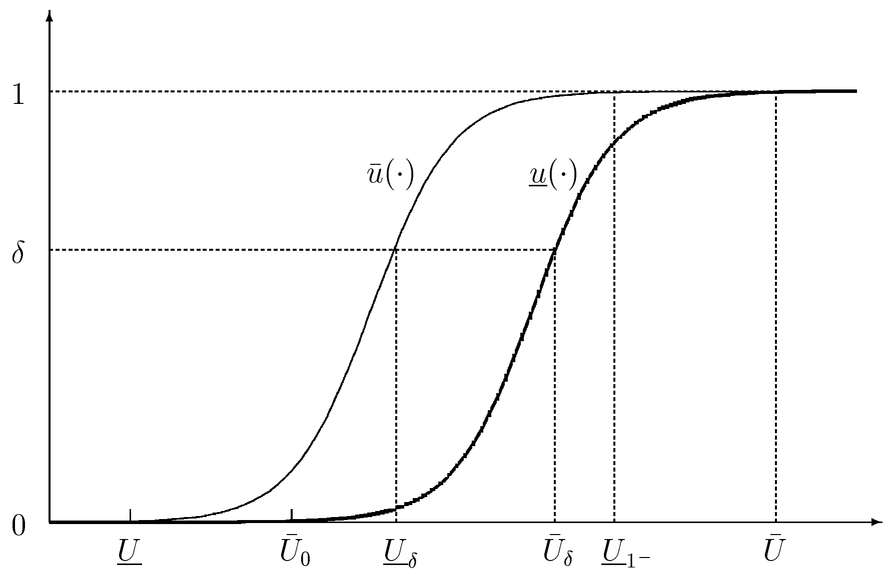

Figure 1: The asymptotic CDFs of a sequence of random variables $\{A_n\}_{n=1}^{\infty}$. $\bar{u}(\cdot)$ = sup-spectrum of A_n; $\underline{u}(\cdot)$ = inf-spectrum of A_n.

In the above definitions, if we let the random variable A_n equal the normalized entropy density of an arbitrary source $\boldsymbol{X}$, we obtain two generalized entropy measures for $\boldsymbol{X}$: the *δ-inf-entropy rate* $\underline{H}_\delta(\boldsymbol{X})$ and the *δ-sup-entropy rate* $\bar{H}_\delta(\boldsymbol{X})$ as described in Table 1. Note that the *inf-entropy-rate* $\underline{H}(\boldsymbol{X})$ and the *sup-entropy-rate* $\bar{H}(\boldsymbol{X})$ introduced in [10] are special cases of the δ-inf/sup-entropy rate measures:

$$\underline{H}(\boldsymbol{X}) = \underline{H}_0(\boldsymbol{X}), \quad \text{and} \quad \bar{H}(\boldsymbol{X}) = \bar{H}_{1^-}(\boldsymbol{X}).$$

Analogously, for an arbitrary channel $\boldsymbol{W} \triangleq P_{\boldsymbol{Y}|\boldsymbol{X}}$ with input $\boldsymbol{X}$ and output $\boldsymbol{Y}$ (or respectively for two observations $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$), if we replace A_n by the normalized information density (resp. by the normalized log-likelihood ratio), we get the δ-inf/sup-information rates (resp. δ-inf/sup-divergences rates) as shown in Table 1.

The algebraic properties of these newly defined information measures are investigated in the next section.

III. Properties of the Generalized Information Measures

Lemma 3.1 Consider two arbitrary random sequences, $\{A_n\}_{n=1}^{\infty}$ and $\{B_n\}_{n=1}^{\infty}$. Let $\bar{u}(\cdot)$ and $\underline{u}(\cdot)$ denote respectively the sup-spectrum and inf-spectrum of $\{A_n\}_{n=1}^{\infty}$. Similarly, let $\bar{v}(\cdot)$ and $\underline{v}(\cdot)$ denote respectively the sup-spectrum and inf-spectrum of $\{B_n\}_{n=1}^{\infty}$. Define $\underline{U}_{\delta} \triangleq \sup\{\theta : \bar{u}(\theta) \leq \delta\}$, $\bar{U}_{\delta} \triangleq \sup\{\theta : \underline{u}(\theta) \leq \delta\}$, $\underline{V}_{\delta} \triangleq \sup\{\theta : \bar{v}(\theta) \leq \delta\}$, $\bar{V}_{\delta} \triangleq \sup\{\theta : \underline{v}(\theta) \leq \delta\}$,

$$(\underline{U+V})_{\delta+\gamma} \triangleq \sup\{\theta : (\overline{u+v})(\theta) \leq \delta+\gamma\},$$

$$(\overline{U+V})_{\delta+\gamma} \triangleq \sup\{\theta : (\underline{u+v})(\theta) \leq \delta+\gamma\},$$

$$(\overline{u+v})(\theta) \triangleq \limsup_{n\to\infty} Pr\{A_n + B_n \leq \theta\},$$

and

$$(\underline{u+v})(\theta) \triangleq \liminf_{n\to\infty} Pr\{A_n + B_n \leq \theta\}.$$

Then the following statements hold.

1. $\underline{U}_{\delta}$ and $\bar{U}_{\delta}$ are both non-decreasing functions of $\delta \in [0,1]$.

2. For $\delta \geq 0$, $\gamma \geq 0$, and $1 \geq \delta+\gamma$,

$$(\underline{U+V})_{\delta+\gamma} \geq \underline{U}_{\delta} + \underline{V}_{\gamma}, \tag{3.5}$$

and

$$(\overline{U+V})_{\delta+\gamma} \geq \underline{U}_{\delta} + \bar{V}_{\gamma}. \tag{3.6}$$

3. For $\delta \geq 0$, $\gamma \geq 0$, and $1 > \delta+\gamma$,

$$(\underline{U+V})_{\delta} \leq \underline{U}_{\delta+\gamma} + \bar{V}_{(1-\gamma)^-}, \tag{3.7}$$

and

$$(\overline{U+V})_{\delta} \leq \bar{U}_{\delta+\gamma} + \bar{V}_{(1-\gamma)^-}. \tag{3.8}$$

***Proof*:** The proof of property 1 follows directly from the definitions of $\underline{U}_\delta$ and $\bar{U}_\delta$ and the fact that the inf-spectrum and the sup-spectrum are non-decreasing in δ.

To show (3.5), we first observe that

$$Pr\{A_n + B_n \le \underline{U}_\delta + \underline{V}_\gamma\} \le Pr\{A_n \le \underline{U}_\delta\} + Pr\{B_n \le \underline{V}_\gamma\}.$$

Then

$$\begin{aligned}
& \limsup_{n\to\infty} Pr\{A_n + B_n \le \underline{U}_\delta + \underline{V}_\gamma\} \\
\le\ & \limsup_{n\to\infty} \Big(Pr\{A_n \le \underline{U}_\delta\} + Pr\{B_n \le \underline{V}_\gamma\} \Big) \\
\le\ & \limsup_{n\to\infty} Pr\{A_n \le \underline{U}_\delta\} + \limsup_{n\to\infty} Pr\{B_n \le \underline{V}_\gamma\} \\
\le\ & \delta + \gamma,
\end{aligned}$$

which, by definition of $(\underline{U+V})_{\delta+\gamma}$, yields (3.5).

Similarly, we have

$$Pr\{A_n + B_n \le \underline{U}_\delta + \bar{V}_\gamma\} \le Pr\{A_n \le \underline{U}_\delta\} + Pr\{B_n \le \bar{V}_\gamma\}.$$

Then

$$\begin{aligned}
& \liminf_{n\to\infty} Pr\{A_n + B_n \le \underline{U}_\delta + \bar{V}_\gamma\} \\
\le\ & \liminf_{n\to\infty} \Big(Pr\{A_n \le \underline{U}_\delta\} + Pr\{B_n \le \bar{V}_\gamma\} \Big) \\
\le\ & \limsup_{n\to\infty} Pr\{A_n \le \underline{U}_\delta\} + \liminf_{n\to\infty} Pr\{B_n \le \bar{V}_\gamma\} \\
\le\ & \delta + \gamma,
\end{aligned}$$

which, by definition of $(\overline{U+V})_{\delta+\gamma}$, proves (3.6).

To show (3.7), we remark from (3.5) that $(\underline{U+V})_\delta + (\underline{-V})_\gamma \le (\underline{U+V-V})_{\delta+\gamma} = \underline{U}_{\delta+\gamma}$. Hence,

$$(\underline{U+V})_\delta \le \underline{U}_{\delta+\gamma} - (\underline{-V})_\gamma.$$

(Note that the cases $\varepsilon+\gamma = 1$ or $\gamma = 1$ are not allowed here because they result in $\underline{U}_1 = \underline{-V}_1 = \infty$, and the subtraction of two infinite terms is undefined. That is why the condition for property 2, $1 \ge \delta + \gamma$, is replaced by $1 > \delta + \gamma$ in property 3.)

The proof is completed by showing that

$$-(\underline{-V})_{\gamma} \leq \bar{V}_{(1-\gamma)^-}. \tag{3.9}$$

By definition,

$$\begin{aligned}
(\overline{-v})(\theta) &\triangleq \limsup_{n\to\infty} Pr\{-B_n \leq \theta\} \\
&= 1 - \liminf_{n\to\infty} Pr\{B_n < -\theta\} \\
&= 1 - \underline{v}(-\theta^+).
\end{aligned}$$

So $\underline{v}(-\theta^+) = 1 - (\overline{-v})(\theta)$. Then

$$\begin{aligned}
\bar{V}_{(1-\gamma)^-} &\triangleq \sup\{\theta : \underline{v}(\theta) < 1-\gamma\} \\
&\geq \sup\{\theta : \underline{v}(\theta^-) < 1-\gamma\} \\
&= \sup\{-\hat{\theta} : \underline{v}(-\hat{\theta}^+) < 1-\gamma\} \\
&= \sup\{-\hat{\theta} : 1 - (\overline{-v})(\theta) < 1-\gamma\} \\
&= -\inf\{\hat{\theta} : (\overline{-v})(\theta) > \gamma\} \\
&= -\sup\{\hat{\theta} : (\overline{-v})(\theta) \leq \gamma\} \\
&= -(\overline{-V})_{\gamma},
\end{aligned}$$

where the inequality follows from $\underline{v}(\theta) \geq \underline{v}(\theta^-)$. Finally, to show (3.8), we observe from (3.6) that $(\overline{U+V})_{\delta} + (\underline{-V})_{\gamma} \leq (\overline{U+V-V})_{\delta+\gamma} = \bar{U}_{\delta+\gamma}$. Hence,

$$(\overline{U+V})_{\delta} \leq \bar{U}_{\delta+\gamma} - (\underline{-V})_{\gamma}.$$

Using (3.9), we have the desired result. $\square$

If we take $\delta = \gamma = 0$ in (3.5) and (3.7), we obtain

$$(\underline{U+V}) \geq \underline{U} + \underline{V} \quad \text{and} \quad (\underline{U+V}) \leq \underline{U} + \bar{V},$$

which mean that the liminf in probability of a sequence of random variables $A_n + B_n$ is upper [resp. lower] bounded by the liminf in probability of A_n plus the limsup [resp. liminf] in probability of B_n. This fact is used in [11] to show that

$$\underline{H}(\boldsymbol{Y}) - \bar{H}(\boldsymbol{Y}|\boldsymbol{X}) \leq \underline{I}(\boldsymbol{X};\boldsymbol{Y}) \leq \underline{H}(\boldsymbol{Y}) - \underline{H}(\boldsymbol{Y}|\boldsymbol{X}),$$

which is a special case of property 3 in Lemma 3.2 .

The next lemmas will show some of the analogous properties of the generalized information measures.

Lemma 3.2 For δ, γ, $\delta+\gamma \in [0,1)$, the following statements hold.

1. $\bar{H}_\delta(\boldsymbol{X}) \geq 0$. $\bar{H}_\delta(\boldsymbol{X}) = 0$ if and only if the sequence $\{X^n = (X_1^{(n)}, \ldots, X_n^{(n)})\}_{n=1}^{\infty}$ is ultimately deterministic (in probability).

   (This property also applies to $\underline{H}_\delta(\boldsymbol{X})$, $\bar{I}_\delta(\boldsymbol{X};\boldsymbol{Y})$, $\underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y})$, $\bar{D}_\delta(\boldsymbol{X}\|\hat{\boldsymbol{X}})$, and $\underline{D}_\delta(\boldsymbol{X}\|\hat{\boldsymbol{X}})$.)

2. $\underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) = \underline{I}_\delta(\boldsymbol{Y};\boldsymbol{X})$ and $\bar{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) = \bar{I}_\delta(\boldsymbol{Y};\boldsymbol{X})$.

3.
$$\underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) \leq \underline{H}_{\delta+\gamma}(\boldsymbol{Y}) - \underline{H}_\gamma(\boldsymbol{Y}|\boldsymbol{X}), \tag{3.10}$$
$$\underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) \leq \bar{H}_{\delta+\gamma}(\boldsymbol{Y}) - \bar{H}_\gamma(\boldsymbol{Y}|\boldsymbol{X}), \tag{3.11}$$
$$\bar{I}_\gamma(\boldsymbol{X};\boldsymbol{Y}) \leq \bar{H}_{\delta+\gamma}(\boldsymbol{Y}) - \underline{H}_\delta(\boldsymbol{Y}|\boldsymbol{X}), \tag{3.12}$$
$$\underline{I}_{\delta+\gamma}(\boldsymbol{X};\boldsymbol{Y}) \geq \underline{H}_\delta(\boldsymbol{Y}) - \bar{H}_{(1-\gamma)^-}(\boldsymbol{Y}|\boldsymbol{X}), \tag{3.13}$$
   and
$$\bar{I}_{\delta+\gamma}(\boldsymbol{X};\boldsymbol{Y}) \geq \bar{H}_\delta(\boldsymbol{Y}) - \bar{H}_{(1-\gamma)^-}(\boldsymbol{Y}|\boldsymbol{X}). \tag{3.14}$$

4. $0 \leq \underline{H}_\delta(\boldsymbol{X}) \leq \bar{H}_\delta(\boldsymbol{X}) \leq \log|\mathcal{X}|$, where each $X_i^{(n)} \in \mathcal{X}$, $i = 1, \ldots, n$ and $n = 1, 2, \ldots$, and $\mathcal{X}$ is finite.

5. $\underline{I}_\delta(\boldsymbol{X},\boldsymbol{Y};\boldsymbol{Z}) \geq \underline{I}_\delta(\boldsymbol{X};\boldsymbol{Z})$.

***Proof*:** Property 1 holds because

$$Pr\left\{-\frac{1}{n}\log P_{X^n}(X^n) < 0\right\} = 0,$$
$$Pr\left\{\frac{1}{n}\log\frac{dP_{X^n}}{dP_{\hat{X}^n}}(X^n) < -\delta\right\} \leq \exp\{-\delta n\},$$

and

$$Pr\left\{\frac{1}{n}\log\frac{dP_{X^nY^n}}{d(P_{X^n}\times P_{Y^n})}(X^n,Y^n)<-\delta\right\}\leq\exp\{-\delta n\}.$$

Property 2 is an immediate consequence of the definition.

To show the inequalities in property 3 we first remark that

$$\frac{1}{n}h_{Y^n}(Y^n)=\frac{1}{n}i_{(X^n,Y^n)}(X^n;Y^n)+\frac{1}{n}h_{(X^n,Y^n)}(Y^n|X^n),$$

where $\frac{1}{n}h_{(X^n,Y^n)}(Y^n|X^n)\triangleq-\frac{1}{n}\log P_{Y^n|X^n}(Y^n|X^n)$. With this fact, (3.10) follows directly from (3.5), (3.11) and (3.12) follow from (3.6), (3.13) follows from (3.7), and (3.14) follows from (3.8).

Property 4 follows from the fact that $\bar{H}_\delta(\cdot)$ is non-decreasing in δ: $\bar{H}_\delta(\boldsymbol{X})\leq\bar{H}_{1^-}=\bar{H}(\boldsymbol{X})$, and that $\bar{H}(\boldsymbol{X})$ is the minimum achievable (i.e., with asymptotically negligible probability of decoding error) fixed-length coding rate for $\boldsymbol{X}$ as seen in [7, Theorem 3.2] and [10].

Property 5 can be proved using the fact that

$$\frac{1}{n}i_{(X^n,Y^n,Z^n)}(X^n,Y^n;Z^n)=\frac{1}{n}i_{(X^n,Z^n)}(X^n;Z^n)+\frac{1}{n}i_{(X^n,Y^n,Z^n)}(Y^n;Z^n|X^n).$$

By applying (3.5), and letting $\gamma=0$, we obtain the desired result. □

Lemma 3.3 (Data processing lemma) Fix $\delta\in[0,1)$. Suppose that for every n, X_1^n and X_3^n are conditionally independent given X_2^n. Then

$$\underline{I}_\delta(\boldsymbol{X}_1;\boldsymbol{X}_3)\leq\underline{I}_\delta(\boldsymbol{X}_1;\boldsymbol{X}_2).$$

***Proof*:** By property 5, we get

$$\underline{I}_\delta(\boldsymbol{X}_1;\boldsymbol{X}_3)\leq\underline{I}_\delta(\boldsymbol{X}_1;\boldsymbol{X}_2,\boldsymbol{X}_3)=\underline{I}_\delta(\boldsymbol{X}_1;\boldsymbol{X}_2),$$

where the equality holds because

$$\frac{1}{n}\log\frac{dP_{X_1^nX_2^nX_3^n}}{d(P_{X_1^n}\times P_{X_2^nX_3^n})}(x_1^n,x_2^n,x_3^n)=\frac{1}{n}\log\frac{dP_{X_1^nX_2^n}}{d(P_{X_1^n}\times P_{X_2^n})}(x_1^n,x_2^n).$$

□

Lemma 3.4 (Optimality of independent inputs) Fix $\delta \in [0,1)$. Consider a finite alphabet, discrete memoryless channel – i.e., $P_{Y^n|X^n} = \prod_{i=1}^n P_{Y_i|X_i}$, for all n. For any input $\boldsymbol{X}$ and its corresponding output $\boldsymbol{Y}$,

$$\underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) \le \underline{I}_\delta(\bar{\boldsymbol{X}};\bar{\boldsymbol{Y}}) = \underline{I}(\bar{\boldsymbol{X}};\bar{\boldsymbol{Y}}),$$

where $\bar{\boldsymbol{Y}}$ is the output due to $\bar{\boldsymbol{X}}$, which is an independent process with the same first order statistics as $\boldsymbol{X}$, i.e., $P_{\overline{X}^n} = \prod_{i=1}^n P_{X_i}$.

***Proof*:** First, we observe that

$$\frac{1}{n}\log\frac{dP_{Y^n|X^n}}{dP_{Y^n}}(X^n,Y^n) + \frac{1}{n}\log\frac{dP_{Y^n}}{dP_{\overline{Y}^n}}(X^n,Y^n) = \frac{1}{n}\log\frac{dP_{Y^n|X^n}}{dP_{\overline{Y}^n}}(X^n,Y^n).$$

In other words,

$$\frac{1}{n}\log\frac{dP_{X^nY^n}}{d(P_{X^n}\times P_{Y^n})}(X^n,Y^n) + \frac{1}{n}\log\frac{dP_{Y^n}}{dP_{\overline{Y}^n}}(X^n,Y^n) = \frac{1}{n}\log\frac{dP_{\overline{X}^n\overline{Y}^n}}{d(P_{\overline{X}^n}\times P_{\overline{Y}^n})}(X^n,Y^n).$$

By evaluating the above terms under $P_{X^nY^n}$ and letting

$$\bar{z}(\theta) \triangleq \limsup_{n\to\infty} P_{X^nY^n}\left\{\frac{1}{n}\log\frac{dP_{\overline{X}^n\overline{Y}^n}}{d(P_{\overline{X}^n}\times P_{\overline{Y}^n})}(X^n,Y^n) \le \theta\right\}$$

and

$$\underline{Z}_\delta(\bar{\boldsymbol{X}};\bar{\boldsymbol{Y}}) \triangleq \sup\{\theta : \bar{z}(\theta) \le \delta\},$$

we obtain from (3.5) (with $\gamma = 0$) that

$$\underline{Z}_\delta(\bar{\boldsymbol{X}};\bar{\boldsymbol{Y}}) \ge \underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) + \underline{D}(\boldsymbol{Y}\|\bar{\boldsymbol{Y}}) \ge \underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}),$$

since $\underline{D}(\boldsymbol{Y}\|\bar{\boldsymbol{Y}}) \ge 0$ by property 1 of Lemma 3.2 .

Note that the summable property of $(1/n)\log[dP_{\overline{X}^n\overline{Y}^n}/d(P_{\overline{X}^n}\times P_{\overline{Y}^n})](X^n,Y^n)$ (i.e., it is equal to $(1/n)\sum_{i=1}^n \log[dP_{\overline{X}_i\overline{Y}_i}/d(P_{\overline{X}_i}\times P_{\overline{Y}_i})](X_i,Y_i)$), the Chebyshev inequality and the finiteness of the channel alphabets imply

$$\underline{I}(\bar{\boldsymbol{X}};\bar{\boldsymbol{Y}}) = \underline{I}_\delta(\bar{\boldsymbol{X}};\bar{\boldsymbol{Y}}) \quad \text{and} \quad \underline{Z}(\bar{\boldsymbol{X}};\bar{\boldsymbol{Y}}) = \underline{Z}_\delta(\bar{\boldsymbol{X}};\bar{\boldsymbol{Y}}).$$

It finally remains to show that

$$\underline{I}(\bar{\boldsymbol{X}};\bar{\boldsymbol{Y}}) \ge \underline{Z}(\bar{\boldsymbol{X}};\bar{\boldsymbol{Y}}),$$

which is proved in [11, Theorem 10]. $\square$

IV. Examples for the Computation of ε-Capacity

In [11], Verdú and Han establish the general formulas for channel capacity and ε-capacity. In terms of the ε-inf-information rate, the expression of the ε-capacity becomes

$$\sup_{\boldsymbol{X}} \underline{I}_{\varepsilon^-}(\boldsymbol{X};\boldsymbol{Y}) \le C_\varepsilon \le \sup_{\boldsymbol{X}} \underline{I}_{\varepsilon}(\boldsymbol{X};\boldsymbol{Y}),$$

where $\varepsilon \in (0,1)$.

We now provide examples for the computation of C_ε. They are basically an extension of some of the examples provided in [11] for the computation of channel capacity.

Let the alphabet be binary $\mathcal{X} = \mathcal{Y} = \{0,1\}$, and let every output be given by

$$Y_i = X_i \oplus Z_i$$

where $\oplus$ represents the addition operation modulo-2 and $\boldsymbol{Z}$ is an *arbitrary* binary random process independent of $\boldsymbol{X}$.

To compute the ε-capacity we use the results of property 3 in Lemma 3.2 :

$$\underline{I}_{\varepsilon^-}(\boldsymbol{X};\boldsymbol{Y}) \ge \underline{H}_0(\boldsymbol{Y}) - \bar{H}_{(1-\varepsilon^-)^-}(\boldsymbol{Y}|\boldsymbol{X}) = \underline{H}_0(\boldsymbol{Y}) - \bar{H}_{(1-\varepsilon)}(\boldsymbol{Y}|\boldsymbol{X}), \tag{4.15}$$

and

$$\underline{I}_{\varepsilon}(\boldsymbol{X};\boldsymbol{Y}) \le \min\{\underline{H}_{\varepsilon+\gamma}(\boldsymbol{Y}) - \underline{H}_{\gamma}(\boldsymbol{Y}|\boldsymbol{X}), \bar{H}_{\varepsilon+\gamma}(\boldsymbol{Y}) - \bar{H}_{\gamma}(\boldsymbol{Y}|\boldsymbol{X})\}. \tag{4.16}$$

where $\varepsilon \ge 0$, $\gamma \ge 0$ and $1 > \varepsilon + \gamma$. The lower bound in (4.15) follows directly from (3.13) (by taking $\delta = 0$ and $\gamma = \varepsilon^-$). The upper bounds in (4.16) follow from (3.10) and (3.11) respectively.

$$\begin{aligned} C_\varepsilon &\le \sup_{\boldsymbol{X}} \underline{I}_{\varepsilon}(\boldsymbol{X};\boldsymbol{Y}) \\ &\le \sup_{\boldsymbol{X}} \left\{ \bar{H}_{\varepsilon+\gamma}(\boldsymbol{Y}) - \bar{H}_{\gamma}(\boldsymbol{Y}|\boldsymbol{X}) \right\}. \end{aligned}$$

Since the above inequality holds for all $0 \le \gamma < 1-\varepsilon$, we have:

$$\begin{aligned} C_\varepsilon &\le \inf_{0\le\gamma<1-\varepsilon} \sup_{\boldsymbol{X}} \left\{ \bar{H}_{\varepsilon+\gamma}(\boldsymbol{Y}) - \bar{H}_{\gamma}(\boldsymbol{Y}|\boldsymbol{X}) \right\}. \qquad (4.17) \\ &\le \inf_{0\le\gamma<1-\varepsilon} \left\{ \sup_{\boldsymbol{X}} \bar{H}_{\varepsilon+\gamma}(\boldsymbol{Y}) - \inf_{\boldsymbol{X}} \bar{H}_{\gamma}(\boldsymbol{Y}|\boldsymbol{X}) \right\}. \end{aligned}$$

By the symmetry of the channel, $\bar{H}_\gamma(\boldsymbol{Y}|\boldsymbol{X}) = \bar{H}_\gamma(\boldsymbol{Z})$ which is independent of $\boldsymbol{X}$. Hence,

$$\begin{aligned} C_\varepsilon &\leq \inf_{0\leq\gamma<1-\varepsilon}\left\{\sup_{\boldsymbol{X}} \bar{H}_{\varepsilon+\gamma}(\boldsymbol{Y}) - \bar{H}_\gamma(\boldsymbol{Z})\right\} \\ &\leq \inf_{0\leq\gamma<1-\varepsilon}\left\{\log 2 - \bar{H}_\gamma(\boldsymbol{Z})\right\} = \inf_{0\leq\gamma<1-\varepsilon}\left\{1 - \bar{H}_\gamma(\boldsymbol{Z})\right\} \end{aligned}$$

where the last step follows by taking a Bernoulli uniform input. Since $1-\bar{H}_\gamma(\boldsymbol{Z})$ is non-increasing in γ,

$$C_\varepsilon \leq 1 - \bar{H}_{(1-\varepsilon)^-}(\boldsymbol{Z}).$$

(Note that the superscript "-" indicates a strict inequality in the definition of $\bar{H}_\gamma(\cdot)$; this is consistent with the condition $\gamma + \varepsilon < 1$.)

On the other hand, we can derive the lower bound to C_ε by choosing a Bernoulli uniform input in (4.15). We thus obtain

$$1 - \bar{H}_{(1-\varepsilon)}(\boldsymbol{Z}) \leq C_\varepsilon \leq 1 - \bar{H}_{(1-\varepsilon)^-}(\boldsymbol{Z}).$$

Note that there are actually two upper bounds in (4.16). In this example, the first upper bound $1 - \underline{H}_{(1-\varepsilon)^-}(\boldsymbol{Z})$ (which is no less than $1 - \bar{H}_{(1-\varepsilon)^-}(\boldsymbol{Z})$) is a looser upper bound, and hence, can be omitted. In addition, we demonstrate in the above derivation that the computation of the upper bound to C_ε involves in general the infimum operation over the parameter γ. Therefore, if the optimizing input distribution does not have a "nice" property (such as independence and uniformity), then the computation of (4.17) may be complicated in general.

Remark: An *alternative* method to compute C_ε is to derive the channel sup-spectrum in terms of the inf-spectrum of the noise process. Under the optimizing equally likely Bernoulli input $\boldsymbol{X}^*$ we can write

$$\begin{aligned} \bar{i}_{(\boldsymbol{X}^*;\boldsymbol{Y})}(\theta) &\triangleq \limsup_{n\to\infty} Pr\left\{\frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n/X^n)}{P_{Y^n}(Y^n)} \leq \theta\right\} \\ &= \limsup_{n\to\infty} Pr\left\{\frac{1}{n}\log P_{Z^n}(Z^n) - \frac{1}{n}\log P_{Y^n}(Y^n) \leq \theta\right\} \\ &= \limsup_{n\to\infty} Pr\left\{\frac{1}{n}\log P_{Z^n}(Z^n) \leq \theta - 1\right\} \\ &= \limsup_{n\to\infty} Pr\left\{-\frac{1}{n}\log P_{Z^n}(Z^n) \geq 1 - \theta\right\} \\ &= 1 - \underline{h}_{\boldsymbol{Z}}((1-\theta)^-). \end{aligned}$$

Hence,

$$\begin{aligned}\underline{I}_{\varepsilon}(\boldsymbol{X}^*;\boldsymbol{Y}) &= \sup\left\{\theta : 1-\underline{h}_{\boldsymbol{Z}}((1-\theta)^-) \le \varepsilon\right\}\\ &= \sup\left\{\theta : \underline{h}_{\boldsymbol{Z}}((1-\theta)^-) \ge 1-\varepsilon\right\}\\ &= \sup\left\{(1-\beta) : \underline{h}_{\boldsymbol{Z}}(\beta^-) \ge 1-\varepsilon\right\}\\ &= 1+\sup\left\{(-\beta) : \underline{h}_{\boldsymbol{Z}}(\beta^-) \ge 1-\varepsilon\right\}\\ &= 1-\inf\left\{\beta : \underline{h}_{\boldsymbol{Z}}(\beta^-) \ge 1-\varepsilon\right\}\\ &= 1-\sup\left\{\beta : \underline{h}_{\boldsymbol{Z}}(\beta^-) < 1-\varepsilon\right\}\\ &= 1-\bar{H}_{(1-\varepsilon)^-}(\boldsymbol{Z}).\end{aligned}$$

Similarly,

$$\underline{I}_{\varepsilon^-}(\boldsymbol{X}^*;\boldsymbol{Y}) = 1-\bar{H}_{(1-\varepsilon)}(\boldsymbol{Z}).$$

Therefore,

$$1-\bar{H}_{(1-\varepsilon)}(\boldsymbol{Z}) = \underline{I}_{\varepsilon^-}(\boldsymbol{X}^*;\boldsymbol{Y}) \le C_\varepsilon \le \underline{I}_{\varepsilon}(\boldsymbol{X}^*;\boldsymbol{Y}) = 1-\bar{H}_{(1-\varepsilon)^-}(\boldsymbol{Z}).$$

Example 4.1 Let $\boldsymbol{Z}$ be an all-zero sequence with probability β and Bernoulli (with parameter p) with probability $1-\beta$. Then the sequence of random variables $(1/n)h_{Z^n}(Z^n)$ converges to atoms 0 and $h_b(p) \stackrel{\triangle}{=} -p\log p-(1-p)\log(1-p)$ with respective masses β and $1-\beta$. The resulting $\underline{h}_{\boldsymbol{Z}}(\theta)$ is depicted in Figure 2. From (4.18), we obtain $\bar{\imath}_{(\boldsymbol{X};\boldsymbol{Y})}(\theta)$ as shown in Figure 3.

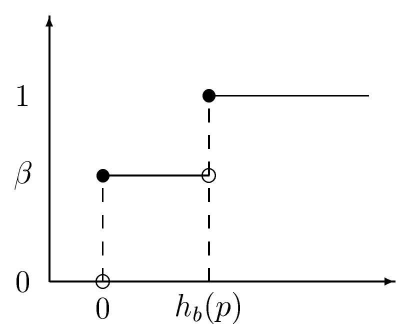

Figure 2: The spectrum of $(1/n)h_{Z^n}(Z^n)$ for Example 4.1.

Therefore,

$$C_\varepsilon = \begin{cases} 1-h_b(p), & \text{if } 0<\varepsilon<1-\beta; \\ 1, & \text{if } 1-\beta<\varepsilon<1 \end{cases}.$$

When $\varepsilon = 1-\beta$, C_ε lies somewhere between $1-h_b(p)$ and 1.

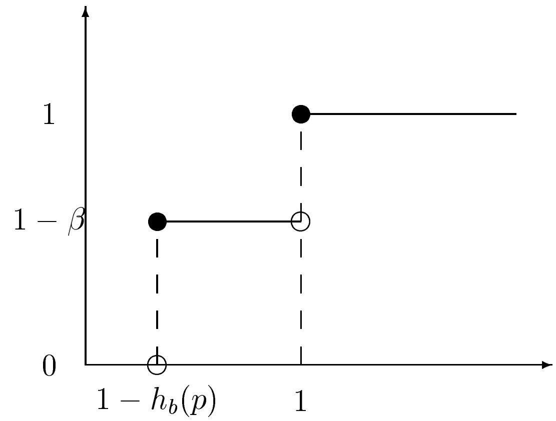

Figure 3: The spectrum of $(1/n)i_{(X^n;Y^n)}(X^n;Y^n)$ for Example 4.1.

Example 4.2 If $\boldsymbol{Z}$ is a non-stationary binary independent sequence with $Pr\{Z_1 = 1\} = p_i$, then by the uniform boundedness (in i) of the variance of random variable $-\log P_{Z_i}(Z_i)$, namely,

$$\begin{aligned}\mathrm{Var}[-\log P_{Z_i}(Z_i)] &\leq E[(\log P_{Z_i}(Z_i))^2] \\ &\leq \sup_{0<p_i<1} p_i(\log p_i)^2 + (1-p_i)(\log(1-p_i))^2 \\ &\leq 1,\end{aligned}$$

we have (by Chebyshev's inequality)

$$Pr\left\{\left|-\frac{1}{n}\log P_{Z^n}(Z^n) - \frac{1}{n}\sum_{i=1}^{n} H(Z_i)\right| < \gamma\right\} \to 0,$$

for any $\gamma > 0$. Therefore, $\bar{H}_{(1-\varepsilon)^-}(\boldsymbol{Z})$ is independent of ε, and C_ε is equal to 1 minus the largest cluster point of $(1/n)\sum_{i=1}^{n} H(Z_i)$, i.e.,

$$\bar{H}_{(1-\varepsilon)^-}(\boldsymbol{Z}) = \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} H(Z_i),$$

and

$$C_\varepsilon = 1 - \bar{H}(\boldsymbol{Z}) = 1 - \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} H(Z_i),$$

where $H(Z_i) = h_b(p_i)$. This result is illustrated in Figures 4 and 5.

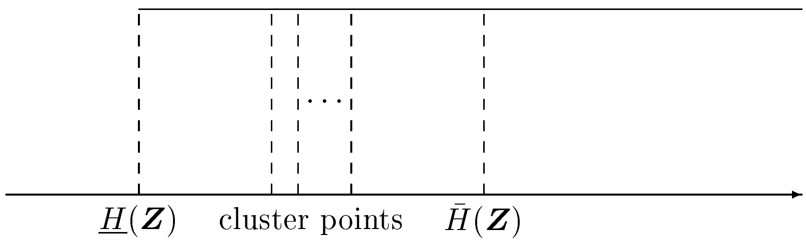

Figure 4: The spectrum of $(1/n)h_{Z^n}(Z^n)$ for Example 4.2.

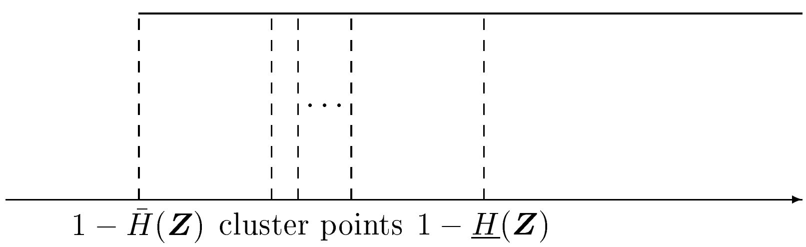

Figure 5: The spectrum of $(1/n)i_{(X^n;Y^n)}(X^n;Y^n)$ for Example 4.2.

V. Conclusions

In light of the work of Han and Verdú in [10] and [11], generalized entropy, mutual-information, and divergence rates are proposed. The properties of each of these information quantities are analyzed, and examples illustrating the computation of the ε-capacity of channels with arbitrary additive noise are presented.

In [7], we use these information measures to prove a generalized version of the Asymptotic Equipartition Property (AEP) and general source coding and hypothesis testing theorems.

Acknowledgment

The authors would like to thank Prof. S. Verdú for his valuable advice and constructive criticism which helped improve the paper.

References

1. P. Billingsley, *Probability and Measure*, Wiley, New York (1986).

2. R. E. Blahut, *Principles and Practice of Information Theory*, Addison Wesley, Massachusetts (1988).

3. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York (1991).

4. P.-N. Chen, "General formulas for the Neyman-Pearson type-II error exponent subject to fixed and exponential type-I error bounds," *IEEE Trans. Info. Theory*, IT–42(1):316–323, January (1996).

5. P.-N. Chen and F. Alajaji, "Strong converse, feedback channel capacity and hypothesis testing," *Journal of the Chinese Institute of Engineers,* vol. 18, pp. 777-785, November 1995; also in *Proceedings of CISS,* John Hopkins Univ., MD, USA, March (1995).

6. P.-N. Chen and F. Alajaji, "The reliability function of arbitrary channels with and without feedback," *Proceedings of the 18'th Biennial Symposium on Communications*, Queen's University, Kingston, Ontario, June (1996).

7. P.-N. Chen and F. Alajaji, "Generalized source coding theorems and hypothesis testing: Part II – Operational limits," *Journal of the Chinese Institute of Engineers,* to appear, May (1998).

8. I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems,* Academic, New York, 1981.

9. R. M. Gray, *Entropy and Information Theory.* Springer-Verlag, New York (1990).

10. T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Info. Theory*, IT–39(3):752–772, May (1993).

11. S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Info. Theory*, IT–40(4):1147–1157, Jul. (1994).

Nomenclature

$(1-\varepsilon)$-achievable data compaction rate	$T_{1-\varepsilon}(\boldsymbol{X})$
$(1-\varepsilon)$-achievable data compression rate at distortion D	$T_{1-\varepsilon}(D, \boldsymbol{X})$
δ-inf-divergence rate	$\underline{D}_{\delta}(\boldsymbol{X}\|\hat{\boldsymbol{X}})$
δ-inf-entropy rate	$\underline{H}_{\delta}(\boldsymbol{X})$
δ-inf-information rate	$\underline{I}_{\delta}(\boldsymbol{X};\boldsymbol{Y})$
δ-sup-divergence rate	$\bar{D}_{\delta}(\boldsymbol{X}\|\hat{\boldsymbol{X}})$
δ-sup-entropy rate	$\bar{H}_{\delta}(\boldsymbol{X})$
δ-sup-information rate	$\bar{I}_{\delta}(\boldsymbol{X};\boldsymbol{Y})$
ε-sup-distortion rate	$\bar{\Lambda}_{\varepsilon}(\boldsymbol{X},\boldsymbol{Y})$
ϵ-capacity	C_{ϵ}
channel capacity	C
channel transition distribution	$P_{W^n} = P_{Y^n\|X^n}$
distortion inf-spectrum	$\underline{\lambda}_{(\boldsymbol{X},f(\boldsymbol{X}))}(\theta)$
divergence inf-spectrum	$\underline{d}_{\boldsymbol{X}\|\hat{\boldsymbol{X}}}(\theta)$
divergence sup-spectrum	$\bar{d}_{\boldsymbol{X}\|\hat{\boldsymbol{X}}}(\theta)$
entropy density	$h_{X^n}(X^n)$
entropy inf-Spectrum	$\underline{h}_{\boldsymbol{X}}(\theta)$
entropy sup-Spectrum	$\bar{h}_{\boldsymbol{X}}(\theta)$
inf-divergence rate	$\underline{D}(\boldsymbol{X}\|\hat{\boldsymbol{X}})$
inf-entropy rate	$\underline{H}(\boldsymbol{X})$
inf-information rate	$\underline{I}(\boldsymbol{X};\boldsymbol{Y})$
information density	$i_{X^nW^n}(x^n;y^n)$
information inf-spectrum	$\underline{i}_{(\boldsymbol{X},\boldsymbol{Y})}(\theta)$
information sup-spectrum	$\bar{i}_{(\boldsymbol{X},\boldsymbol{Y})}(\theta)$
input alphabet	$\mathcal{A}$
input distributions	P_{X^n}
log-likelihood ratio	$d_{X^n}(X^n\|\hat{X}^n)$
output alphabet	$\mathcal{B}$
sup-divergence rate	$\bar{D}(\boldsymbol{X}\|\hat{\boldsymbol{X}})$
sup-entropy rate	$\bar{H}(\boldsymbol{X})$
sup-information rate	$\bar{I}(\boldsymbol{X};\boldsymbol{Y})$

Entropy Measures	
System	Arbitrary source $\boldsymbol{X}$
A_n : Norm. Entropy Density	$\frac{1}{n}h_{X^n}(X^n) \triangleq -\frac{1}{n}\log P_{X^n}(X^n)$
Entropy Sup-Spectrum	$\bar{h}_{\boldsymbol{X}}(\theta) \triangleq \limsup_{n\to\infty} Pr\left\{\frac{1}{n}h_{X^n}(X^n) \le \theta\right\}$
Entropy Inf-Spectrum	$\underline{h}_{\boldsymbol{X}}(\theta) \triangleq \liminf_{n\to\infty} Pr\left\{\frac{1}{n}h_{X^n}(X^n) \le \theta\right\}$
δ-Inf-Entropy Rate	$\underline{H}_\delta(\boldsymbol{X}) \triangleq \sup\{\theta : \bar{h}_{\boldsymbol{X}}(\theta) \le \delta\}$
δ-Sup-Entropy Rate	$\bar{H}_\delta(\boldsymbol{X}) \triangleq \sup\{\theta : \underline{h}_{\boldsymbol{X}}(\theta) \le \delta\}$
Sup-Entropy Rate	$\bar{H}(\boldsymbol{X}) \triangleq \bar{H}_{1^-}(\boldsymbol{X})$
Inf-Entropy Rate	$\underline{H}(\boldsymbol{X}) \triangleq \underline{H}_0(\boldsymbol{X})$

Mutual Information Measures	
System	Arbitrary channel $\boldsymbol{W} \triangleq P_{\boldsymbol{Y}\mid\boldsymbol{X}}$ with input $\boldsymbol{X}$ and output $\boldsymbol{Y}$
A_n : Norm. Information Density	$\frac{1}{n}i_{(X^n,Y^n)}(X^n;Y^n) \triangleq \frac{1}{n}\log\frac{dP_{X^nY^n}}{d(P_{X^n}\times P_{Y^n})}(X^n,Y^n)$
Information Sup-Spectrum	$\bar{i}_{(\boldsymbol{X},\boldsymbol{Y})}(\theta) \triangleq \limsup_{n\to\infty} Pr\left\{\frac{1}{n}i_{(X^n,Y^n)}(X^n;Y^n) \le \theta\right\}$
Information Inf-Spectrum	$\underline{i}_{(\boldsymbol{X},\boldsymbol{Y})}(\theta) \triangleq \liminf_{n\to\infty} Pr\left\{\frac{1}{n}i_{(X^n,Y^n)}(X^n;Y^n) \le \theta\right\}$
δ-Inf-Information Rate	$\underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) \triangleq \sup\{\theta : \bar{i}_{(\boldsymbol{X},\boldsymbol{Y})}(\theta) \le \delta\}$
δ-Sup-Information Rate	$\bar{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) \triangleq \sup\{\theta : \underline{i}_{(\boldsymbol{X},\boldsymbol{Y})}(\theta) \le \delta\}$
Sup-Information Rate	$\bar{I}(\boldsymbol{X};\boldsymbol{Y}) \triangleq \bar{I}_{1^-}(\boldsymbol{X};\boldsymbol{Y})$
Inf-Information Rate	$\underline{I}(\boldsymbol{X};\boldsymbol{Y}) \triangleq \underline{I}_0(\boldsymbol{X};\boldsymbol{Y})$

Divergence Measures	
System	Arbitrary sources $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$
A_n : Norm. Log-Likelihood Ratio	$\frac{1}{n}d_{X^n}(X^n\\|\hat{X}^n) \triangleq \frac{1}{n}\log\frac{dP_{X^n}}{dP_{\hat{X}^n}}(X^n)$
Divergence Sup-Spectrum	$\bar{d}_{\boldsymbol{X}\\|\hat{\boldsymbol{X}}}(\theta) \triangleq \limsup_{n\to\infty} Pr\left\{\frac{1}{n}d_{X^n}(X^n\\|\hat{X}^n) \le \theta\right\}$
Divergence Inf-Spectrum	$\underline{d}_{\boldsymbol{X}\\|\hat{\boldsymbol{X}}}(\theta) \triangleq \liminf_{n\to\infty} Pr\left\{\frac{1}{n}d_{X^n}(X^n\\|\hat{X}^n) \le \theta\right\}$
δ-Inf-Divergence Rate	$\underline{D}_\delta(\boldsymbol{X}\\|\hat{\boldsymbol{X}}) \triangleq \sup\{\theta : \bar{d}_{\boldsymbol{X}\\|\hat{\boldsymbol{X}}}(\theta) \le \delta\}$
δ-Sup-Divergence Rate	$\bar{D}_\delta(\boldsymbol{X}\\|\hat{\boldsymbol{X}}) \triangleq \sup\{\theta : \underline{d}_{\boldsymbol{X}\\|\hat{\boldsymbol{X}}}(\theta) \le \delta\}$
Sup-Divergence Rate	$\bar{D}(\boldsymbol{X}\\|\hat{\boldsymbol{X}}) \triangleq \bar{D}_{1^-}(\boldsymbol{X}\\|\hat{\boldsymbol{X}})$
Inf-Divergence Rate	$\underline{D}(\boldsymbol{X}\\|\hat{\boldsymbol{X}}) \triangleq \underline{D}_0(\boldsymbol{X}\\|\hat{\boldsymbol{X}})$

Table 1: Generalized information measures where $\delta \in [0,1]$.