Optimistic Shannon Coding Theorems for Arbitrary Single-User Systems*

Po-Ning Chen† and *Fady Alajaji*‡

†Department of Communication Engineering
National Chiao-Tung University
HsinChu, Taiwan, R.O.C.

‡Department of Mathematics and Statistics
Queen's University
Kingston, Ontario K7L 3N6, Canada

Abstract

The conventional definitions of the source coding rate and of channel capacity require the existence of reliable codes for *all sufficiently large blocklengths*. Alternatively, if it is required that good codes exist for *infinitely many blocklengths*, then *optimistic* definitions of source coding rate and channel capacity are obtained.

In this work, formulas for the optimistic minimum achievable fixed-length source coding rate and the minimum ε-achievable source coding rate for arbitrary finite-alphabet sources are established. The expressions for the optimistic capacity and the optimistic ε-capacity of arbitrary single-user channels are also provided. The expressions of the optimistic source coding rate and capacity are examined for the class of information stable sources and channels, respectively. Finally, examples for the computation of optimistic capacity are presented.

Index terms – Shannon theory, optimistic channel capacity, optimistic source coding rate, error probability, source-channel separation theorem.

* *IEEE Transactions on Information Theory*, to appear. This work was supported in part by an ARC grant from Queen's University, the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant OGP0183645, and the National Science Council of Taiwan, R.O.C., under Grant NSC 87-2213-E-009-139. Parts of this paper were presented at the IEEE International Symposium on Information Theory, MIT, Boston, MA, August 1998.

I Introduction

The conventional definition of the minimum achievable fixed-length source coding rate T for a source $\boldsymbol{Z}$ [13, Definition 4] requires the existence of reliable source codes for *all sufficiently large blocklengths*. Alternatively, if it is required that reliable codes exist for *infinitely many blocklengths*, a new, more *optimistic* definition of source coding rate (denoted by $\underline{T}$) is obtained [13]. Similarly, the *optimistic capacity* $\bar{C}$ is defined by requiring the existence of reliable channel codes for infinitely many blocklengths, as opposed to the definition of the conventional channel capacity C [14, Definition 1].

This concept of optimistic source coding rate and capacity has recently been investigated by Verdú *et.al* for *arbitrary* (not necessarily stationary, ergodic, information stable, etc.) sources and single-user channels [13, 14]. More specifically, they establish an additional *operational* characterization for the optimistic minimum achievable source coding rate ($\underline{T}$) by demonstrating that for a given channel, the classical statement of the source-channel separation theorem[1] holds for every channel if $\underline{T} = T$ [13]. In a dual fashion, they also show that for channels with $\bar{C} = C$, the classical separation theorem holds for every source. They also conjecture that $\underline{T}$ and $\bar{C}$ do not seem to admit a simple expression.

In this work, we demonstrate that $\underline{T}$ and $\bar{C}$ do indeed have a general formula. The key to these results is the application of the generalized sup-information rate introduced in [3, 4] to the existing proofs by Verdú and Han [14, 7] of the direct and converse parts of the conventional coding theorems. We also provide a general expression for the optimistic minimum ε-achievable source coding rate and the optimistic ε-capacity.

In Section II, we briefly introduce the generalized sup/inf-information/entropy rates which will play a key role in proving our optimistic coding theorems. In Section III, we provide the optimistic source coding theorems. They are shown based on *two recent bounds* due to Han [7] on the error probability of a source code as a function of its size. Interestingly, these bounds constitute the natural counterparts of the upper bound provided by Feinstein's Lemma and the Verdú-Han lower bound to the error probability of a channel code.

[1]By the "classical statement of the source-channel separation theorem," we mean the following. Given a source $\boldsymbol{Z}$ with (conventional) source coding rate $T(\boldsymbol{Z})$ and channel $\boldsymbol{W}$ with capacity C, then $\boldsymbol{Z}$ can be reliably transmitted over $\boldsymbol{W}$ if $T(\boldsymbol{Z}) < C$. Conversely, if $T(\boldsymbol{Z}) > C$, then $\boldsymbol{Z}$ cannot be reliably transmitted over $\boldsymbol{W}$. By reliable transmissibility of the source over the channel, we mean that there exits a sequence of source-channel codes such that the decoding error probability vanishes as the blocklength $n \to \infty$ (cf [13]).

Furthermore, we show that for information stable sources, the formula for $\underline{T}$ reduces to

$$\underline{T} = \liminf_{n\to\infty} \frac{1}{n} H(X^n).$$

This is in contrast to the expression for T, which is known to be

$$T = \limsup_{n\to\infty} \frac{1}{n} H(X^n).$$

The above result leads us to observe that for sources that are both stationary and information stable, the classical separation theorem is valid for every channel.

In Section IV, we present (without proving) the general optimistic channel coding theorems, and prove that for the class of information stable channels the expression of $\bar{C}$ becomes

$$\bar{C} = \limsup_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n),$$

while the expression of C is

$$C = \liminf_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n).$$

Finally, in Section V, we present examples for the computation of C and $\bar{C}$ for information stable as well as information unstable channels.

II ε-Inf/Sup-Information/Entropy Rates

Consider an input process $\boldsymbol{X}$ defined by a sequence of finite dimensional distributions [14]: $\boldsymbol{X} \triangleq \left\{X^n = \left(X_1^{(n)}, \ldots, X_n^{(n)}\right)\right\}_{n=1}^{\infty}$. Denote by $\boldsymbol{Y} \triangleq \left\{Y^n = \left(Y_1^{(n)}, \ldots, Y_n^{(n)}\right)\right\}_{n=1}^{\infty}$ the corresponding output process induced by $\boldsymbol{X}$ via the channel $\boldsymbol{W} \triangleq \left\{W^n = P_{Y^n|X^n} : \mathcal{X}^n \to \mathcal{Y}^n\right\}_{n=1}^{\infty}$, 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. We assume throughout this paper that $\mathcal{X}$ and $\mathcal{Y}$ are finite.

In [8, 14], Han and Verdú introduce the notions of inf/sup-information/entropy 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.

The *inf-information rate* $\underline{I}(\boldsymbol{X};\boldsymbol{Y})$ (resp. *sup-information rate* $\bar{I}(\boldsymbol{X};\boldsymbol{Y})$) between processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined in [8] as the *liminf in probability* (resp. *limsup in probability*) of the sequence of normalized information densities $(1/n)\ i_{X^nW^n}(X^n;Y^n)$, where

$$\frac{1}{n} i_{X^nW^n}(a^n; b^n) \triangleq \frac{1}{n} \log \frac{P_{Y^n|X^n}(b^n|a^n)}{P_{Y^n}(b^n)}.$$

When $\boldsymbol{X}$ is equal to $\boldsymbol{Y}$, $\bar{I}(\boldsymbol{X};\boldsymbol{X})$ (respectively, $\underline{I}(\boldsymbol{X};\boldsymbol{X})$) is referred to as the *sup* (respectively, *inf*) *entropy rate* of $\boldsymbol{X}$ and is denoted by $\bar{H}(\boldsymbol{X})$ (respectively, $\underline{H}(\boldsymbol{X})$).

The *liminf in probability* of a sequence of random variables is defined as follows [8]: if A_n is a sequence of random variables, then its *liminf in probability* is the largest extended real number $\underline{U}$ such that,

$$\lim_{n\to\infty} Pr[A_n < \underline{U}] = 0. \tag{1}$$

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

$$\lim_{n\to\infty} Pr[A_n > \bar{U}] = 0. \tag{2}$$

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) and (2) are respectively equivalent to

$$\liminf_{n\to\infty} Pr[A_n < \underline{U}] = \limsup_{n\to\infty} Pr[A_n < \underline{U}] = 0, \tag{3}$$

and

$$\liminf_{n\to\infty} Pr[A_n < \bar{U}] = \limsup_{n\to\infty} Pr[A_n > \bar{U}] = 0. \tag{4}$$

We can observe however that there might exist cases of interest where *only* the liminfs of the probabilities in (3) and (4) are equal to zero, while the limsups do *not* vanish. There are also other cases where *both* the liminfs and limsups in (3)-(4) do not vanish, but they are upper bounded by a prescribed threshold ε. 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. This remark constitutes the motivation to the recent work in [3, 4], where generalized versions of the inf/sub-information/entropy rates are established.

Definition 2.1 (Inf/sup spectrums [3, 4]) If $\{A_n\}_{n=1}^{\infty}$ is a sequence of random variables taking values in a finite set $\mathcal{A}$, then its *inf-spectrum* $\underline{u}(\cdot)$ and its *sup-spectrum* $\bar{u}(\cdot)$ are defined by

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

and

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

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 – $Pr\{A_n \le \theta\}$ – is non-decreasing and right-continuous. However, for $\underline{u}(\cdot)$ and $\bar{u}(\cdot)$, only the *non-decreasing* property remains.

Definition 2.2 (Quantile of inf/sup-spectrum [3, 4]) For any $0 \le \varepsilon \le 1$, the *quantiles* $\underline{U}_\varepsilon$ and $\bar{U}_\varepsilon$ of the sup-spectrum and the inf-spectrum are defined by

$$\underline{U}_\varepsilon \stackrel{\triangle}{=} \sup\{\theta : \bar{u}(\theta) \le \varepsilon\},$$

and

$$\bar{U}_\varepsilon \stackrel{\triangle}{=} \sup\{\theta : \underline{u}(\theta) \le \varepsilon\},$$

respectively. It follows from the above definitions that $\underline{U}_\varepsilon$ and $\bar{U}_\varepsilon$ are right-continuous and non-decreasing in ε. Note that Han and Verdú's liminf/limsup in probability of A_n are *special cases* of $\underline{U}_\varepsilon$ and $\bar{U}_\varepsilon$. More specifically, the following hold

$$\underline{U} = \underline{U}_0 \quad \text{and} \quad \bar{U} = \bar{U}_{1^-},$$

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

$$\bar{U}_{\varepsilon^-} \stackrel{\triangle}{=} \sup\{\theta : \underline{u}(\theta) < \varepsilon\}.$$

Note also that $\underline{U} \le \underline{U}_\varepsilon \le \bar{U}_\varepsilon \le \bar{U}$. Remark that $\underline{U}_\varepsilon$ and $\bar{U}_\varepsilon$ always exist. For a better understanding of the quantities defined above, we depict them in Figure 1. If we replace A_n by the normalized information (resp. entropy) density, we get the following definitions.

Definition 2.3 (ε-inf/sup-information rates [3, 4])

The *ε-inf-information rate* $\underline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y})$ (resp. *ε-sup-information rate* $\bar{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y})$) between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined as the quantile of the sup-spectrum (resp. inf-spectrum) of the normalized information density. More specifically,

$$\underline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y}) \stackrel{\triangle}{=} \sup\{\delta : \bar{i}_{\boldsymbol{X}\boldsymbol{W}}(\delta) \le \varepsilon\},$$

where $\bar{i}_{\boldsymbol{X}\boldsymbol{W}}(\delta) \stackrel{\triangle}{=} \limsup_{n\to\infty} Pr\left\{\frac{1}{n} i_{X^nW^n}(X^n;Y^n) \le \delta\right\}$, and

$$\bar{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y}) \stackrel{\triangle}{=} \sup\{\delta : \underline{i}_{\boldsymbol{X}\boldsymbol{W}}(\delta) \le \varepsilon\},$$

where $\underline{i}_{\boldsymbol{X}\boldsymbol{W}}(\delta) \stackrel{\triangle}{=} \liminf_{n\to\infty} Pr\left\{\frac{1}{n} i_{X^nW^n}(X^n;Y^n) \le \delta\right\}$.

Definition 2.4 (ε-inf/sup-entropy rates [3, 4]) The *ε-inf-entropy rate* $\underline{H}_\varepsilon(\boldsymbol{X})$ (resp. *ε-sup-entropy rate* $\bar{H}_\varepsilon(\boldsymbol{X})$) for a source $\boldsymbol{X}$ is defined as the quantile of the sup-spectrum (resp. inf-spectrum) of the normalized entropy density. More specifically,

$$\underline{H}_\varepsilon(\boldsymbol{X}) \stackrel{\triangle}{=} \sup\{\delta : \bar{h}_{\boldsymbol{X}}(\delta) \le \varepsilon\},$$

where $\bar{h}_{\boldsymbol{X}}(\delta) \triangleq \limsup_{n\to\infty} Pr\left\{\frac{1}{n}h_{X^n}(X^n) \le \delta\right\}$, and

$$\bar{H}_\varepsilon(\boldsymbol{X}) \triangleq \sup\{\delta : \underline{h}_{\boldsymbol{X}}(\delta) \le \varepsilon\},$$

where $\underline{h}_{\boldsymbol{X}}(\delta) \triangleq \liminf_{n\to\infty} Pr\left\{\frac{1}{n}h_{X^n}(X^n) \le \delta\right\}$, and $\frac{1}{n}h_{X^n}(X^n) \triangleq \frac{1}{n}\log\frac{1}{P_{X^n}(X^n)}$.

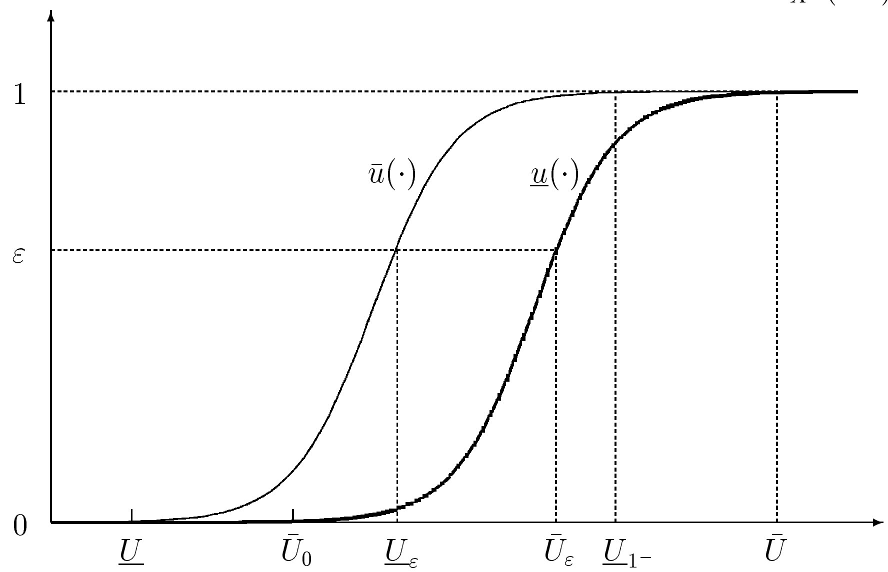

Figure 1: The asymptotic CDFs of a sequence of random variables $\{A_n\}_{n=1}^{\infty}$: $\bar{u}(\cdot)$ = sup-spectrum and $\underline{u}(\cdot)$ = inf-spectrum.

III Optimistic Source Coding Theorems

In [13], Vembu *et.al* characterize the sources for which the classical separation theorem holds for *every channel*. They demonstrate that for a given source $\boldsymbol{X}$, the separation theorem holds for every channel if its optimistic minimum achievable source coding rate ($\underline{T}(\boldsymbol{X})$) coincides with its conventional (or pessimistic) minimum achievable source coding rate ($T(\boldsymbol{X})$); i.e., if $\underline{T}(\boldsymbol{X}) = T(\boldsymbol{X})$.

We herein establish a general formula for $\underline{T}(\boldsymbol{X})$. We prove that for any source $\boldsymbol{X}$,

$$\underline{T}(\boldsymbol{X}) = \underline{H}_{1^-}(\boldsymbol{X}).$$

We also provide the general expression for the optimistic minimum ε-achievable source coding rate. We show these results based on *two new bounds* due to Han (one upper bound and one lower bound) on the error probability of a source code [7, Chapter 1]. The upper bound (Lemma 3.1) consists of the *counterpart* of Feinstein's Lemma for channel codes (cf for example [14, Theorem 1]), while the lower bound (Lemma 3.2) consists of the *counterpart* of the Verdú-Han lower bound on the error probability of a channel code ([14, Theorem 4]). As in the case of the channel coding bounds, both source coding bounds (Lemmas 3.1 and 3.2) hold for arbitrary sources and for arbitrary fixed blocklength.

Definition 3.5 An (n, M) fixed-length source code for X^n is a collection of M n-tuples $\mathscr{C}_n = \{c_1^n, \ldots, c_M^n\}$. The error probability of the code is $P_e^{(n)} \triangleq Pr\left[X^n \notin \mathscr{C}_n\right]$.

Definition 3.6 (Optimistic ε-achievable source coding rate) Fix $0 < \varepsilon < 1$. $R \geq 0$ is an optimistic ε-achievable rate if, for every $\gamma > 0$, there exists a sequence of (n, M) fixed-length source codes $\mathscr{C}_n$ such that

$$\frac{1}{n} \log M < R + \gamma \quad \text{and} \quad P_e^{(n)} \leq \varepsilon \quad \text{for infinitely many } n.$$

The infimum of all ε-achievable source coding rates for source $\boldsymbol{X}$ is denoted by $\underline{T}_\varepsilon(\boldsymbol{X})$. Also define $\underline{T}(\boldsymbol{X}) \triangleq \sup_{0<\varepsilon<1} \underline{T}_\varepsilon(\boldsymbol{X}) = \lim_{\varepsilon \downarrow 0} \underline{T}_\varepsilon(\boldsymbol{X})$ as the optimistic source coding rate.

Lemma 3.1 (Lemma 1.5 in [7]) Fix a positive integer n. There exists an (n, M) source block code $\mathscr{C}_n$ for P_{X^n} such that its error probability satisfies

$$P_e^{(n)} \leq Pr\left[\frac{1}{n} h_{X^n}(X^n) > \frac{1}{n} \log M\right].$$

Lemma 3.2 (Lemma 1.6 in [7]) Every (n, M) source block code $\mathscr{C}_n$ for P_{X^n} satisfies

$$P_e^{(n)} \geq Pr\left[\frac{1}{n} h_{X^n}(X^n) > \frac{1}{n} \log M + \gamma\right] - \exp\{-n\gamma\},$$

for every $\gamma > 0$.

We next use Lemmas 3.1 and 3.2 to prove *general* optimistic (fixed-length) source coding theorems.

Theorem 3.1 (Optimistic minimum ε-achievable source coding rate formula)
Fix $0 < \varepsilon < 1$. For any source $\boldsymbol{X}$,

$$\underline{H}_{\varepsilon^-}(\boldsymbol{X}) \leq \underline{T}_{1-\varepsilon}(\boldsymbol{X}) \leq \underline{H}_\varepsilon(\boldsymbol{X}).$$

Note that actually $\underline{T}_{1-\varepsilon}(\boldsymbol{X}) = \underline{H}_\varepsilon(\boldsymbol{X})$, except possibly at the points of discontinuities of $\underline{H}_\varepsilon(\boldsymbol{X})$ (which are countable).

Proof:

1. *Forward part (achievability):* $\underline{T}_{1-\varepsilon}(\boldsymbol{X}) \leq \underline{H}_{\varepsilon}(\boldsymbol{X})$

We need to prove the existence of a sequence of block codes $\{\mathscr{C}_n\}_{n\geq 0}$ such that, for every $\gamma > 0$, $(1/n)\log|\mathscr{C}_n| < \underline{H}_{\varepsilon}(\boldsymbol{X})+\gamma$ and $P_e^{(n)} \leq 1-\varepsilon$ for infinitely many n. Lemma 3.1 ensures the existence (for any $\gamma > 0$) of a source block code $\mathscr{C}_n = (n, \exp\{n(\underline{H}_{\varepsilon} + \gamma/2)\})$ with error probability

$$P_e^{(n)} \leq Pr\left\{\frac{1}{n}h_{X^n}(X^n) > \underline{H}_{\varepsilon} + \frac{\gamma}{2}\right\}.$$

Therefore,

$$\begin{aligned}
\liminf_{n\to\infty} P_e^{(n)} &\leq \liminf_{n\to\infty} Pr\left\{\frac{1}{n}h_{X^n}(X^n) > \underline{H}_{\varepsilon} + \frac{\gamma}{2}\right\} \\
&= 1 - \limsup_{n\to\infty} Pr\left\{\frac{1}{n}h_{X^n}(X^n) \leq \underline{H}_{\varepsilon}(\boldsymbol{X}) + \frac{\gamma}{2}\right\} \\
&< 1-\varepsilon, \qquad (5)
\end{aligned}$$

where (5) follows from the definition of $\underline{H}_{\varepsilon}(\boldsymbol{X})$. Hence, $P_e^{(n)} \leq 1-\varepsilon$ for infinitely many n.

2. *Converse part:* $\underline{T}_{1-\varepsilon}(\boldsymbol{X}) \geq \underline{H}_{\varepsilon^-}(\boldsymbol{X})$

Assume without loss of generality that $\underline{H}_{\varepsilon^-}(\boldsymbol{X}) > 0$. We will prove the converse by contradiction. Suppose that $\underline{T}_{1-\varepsilon}(\boldsymbol{X}) < \underline{H}_{\varepsilon^-}(\boldsymbol{X})$. Then $(\exists\ \gamma > 0)$ $\underline{T}_{1-\varepsilon}(\boldsymbol{X}) < \underline{H}_{\varepsilon^-}(\boldsymbol{X}) - 3\gamma$. By definition of $\underline{T}_{1-\varepsilon}(\boldsymbol{X})$, there exists a sequence of codes $\mathscr{C}_n$ such that

$$\frac{1}{n}\log|\mathscr{C}_n| < [\underline{H}_{\varepsilon^-}(\boldsymbol{X}) - 3\gamma] + \gamma$$

and

$$\liminf_{n\to\infty} P_e^{(n)} \leq 1-\varepsilon. \qquad (6)$$

By Lemma 3.2,

$$\begin{aligned}
P_e^{(n)} &\geq Pr\left[\frac{1}{n}h_{X^n}(X^n) > \frac{1}{n}\log|\mathscr{C}_n| + \gamma\right] - e^{-n\gamma} \\
&\geq Pr\left[\frac{1}{n}h_{X^n}(X^n) > (\underline{H}_{\varepsilon^-}(\boldsymbol{X}) - 2\gamma) + \gamma\right] - e^{-n\gamma}.
\end{aligned}$$

Therefore,

$$\liminf_{n\to\infty} P_e^{(n)} \geq 1 - \limsup_{n\to\infty} Pr\left[\frac{1}{n}h_{X^n}(X^n) \leq \underline{H}_{\varepsilon^-}(\boldsymbol{X}) - \gamma\right] > 1-\varepsilon,$$

where the last inequality follows from the definition of $\underline{H}_{\varepsilon^-}(\boldsymbol{X})$. Thus, a contradiction to (6) is obtained.

3. *Equality:* $\underline{H}_\varepsilon(\boldsymbol{X})$ is a non-decreasing function of ε; hence the number of discontinuous points is countable. For any continuous point ε, we have that $\underline{H}_\varepsilon(\boldsymbol{X}) = \underline{H}_{\varepsilon^-}(\boldsymbol{X})$, and thus $\underline{T}_\varepsilon(\boldsymbol{X}) = \underline{H}_\varepsilon(\boldsymbol{X})$. □

Theorem 3.2 (Optimistic minimum achievable source coding rate formula)
For any source $\boldsymbol{X}$,

$$\underline{T}(\boldsymbol{X}) = \underline{H}_{1^-}(\boldsymbol{X}).$$

***Proof*:**
By definition,

$$\underline{T}(\boldsymbol{X}) \triangleq \sup_{0<\varepsilon<1} \underline{T}_\varepsilon(\boldsymbol{X}) \geq \sup_{0<\varepsilon<1} \underline{H}_{\varepsilon^-}(\boldsymbol{X}) \geq \underline{H}_{1^-}(\boldsymbol{X}).$$

On the other hand, suppose that $\underline{H}_{1^-}(\boldsymbol{X}) < \underline{T}(\boldsymbol{X})$. Then $\exists\ \gamma > 0$ such that

$$\underline{H}_{1^-}(\boldsymbol{X}) < \underline{T}(\boldsymbol{X}) - \gamma.$$

But by definition of $\underline{T}(\boldsymbol{X})$, there exists $0 < \varepsilon = \varepsilon(\gamma) < 1$ such that

$$\underline{T}(\boldsymbol{X}) - \gamma < \underline{T}_\varepsilon(\boldsymbol{X}).$$

Therefore, $\underline{H}_{1^-}(\boldsymbol{X}) < \underline{T}(\boldsymbol{X}) - \gamma < \underline{T}_\varepsilon(\boldsymbol{X}) \leq \underline{H}_{1-\varepsilon}(\boldsymbol{X}) \leq \underline{H}_{1^-}(\boldsymbol{X})$, and a contradiction is obtained. □

We conclude this section by examining the expression of $\underline{T}(\boldsymbol{X})$ for information stable sources. It is already known (cf for example [13]) that for an information stable source $\boldsymbol{X}$,

$$T(\boldsymbol{X}) = \limsup_{n\to\infty} \frac{1}{n} H(X^n).$$

We herein prove a parallel expression for $\underline{T}(\boldsymbol{X})$.

Definition 3.7 (Information stable sources [13]) A source $\boldsymbol{X}$ is said to be information stable if $H(X^n) > 0$ for n sufficiently large, and $h_{X^n}(X^n)/H(X^n)$ converges in probability to one as $n \to \infty$, i.e.,

$$\limsup_{n\to\infty} Pr\left[\left|\frac{h_{X^n}(X^n)}{H(X^n)} - 1\right| > \gamma\right] = 0 \qquad \forall\gamma > 0,$$

where $H(X^n) = E[h_{X^n}(X^n)]$ is the entropy of X^n.

Lemma 3.3 Every information source $\boldsymbol{X}$ satisfies

$$\underline{T}(\boldsymbol{X}) = \liminf_{n\to\infty} \frac{1}{n} H(X^n).$$

***Proof*:**

1. $[\underline{T}(\boldsymbol{X}) \geq \liminf_{n\to\infty}(1/n)H(X^n)]$

Fix $\varepsilon > 0$ arbitrarily small. Using the fact that $h_{X^n}(X^n)$ is a (finite-alphabet) nonnegative bounded random variable, we can write the normalized block entropy as

$$\begin{aligned}\frac{1}{n}H(X^n) = E\left[\frac{1}{n}h_{X^n}(X^n)\right] &= E\left[\frac{1}{n}h_{X^n}(X^n)\ 1\left\{0 \leq \frac{1}{n}h_{X^n}(X^n) \leq \underline{H}_{1^-}(\boldsymbol{X}) + \varepsilon\right\}\right] \\ &+ E\left[\frac{1}{n}h_{X^n}(X^n)\ 1\left\{\frac{1}{n}h_{X^n}(X^n) > \underline{H}_{1^-}(\boldsymbol{X}) + \varepsilon\right\}\right]. \quad (7)\end{aligned}$$

From the definition of $\underline{H}_{1^-}(\boldsymbol{X})$, it directly follows that the first term in the right hand side of (7) is upper bounded by $\underline{H}_{1^-}(\boldsymbol{X}) + \varepsilon$, and that the liminf of the second term is zero. Thus

$$\underline{T}(\boldsymbol{X}) = \underline{H}_{1^-}(\boldsymbol{X}) \geq \liminf_{n\to\infty}\frac{1}{n}H(X^n).$$

2. $[\underline{T}(\boldsymbol{X}) \leq \liminf_{n\to\infty}(1/n)H(X^n)]$

Fix $\varepsilon > 0$. Then for infinitely many n,

$$\begin{aligned}Pr\left\{\frac{h_{X^n}(X^n)}{H(X^n)} - 1 > \varepsilon\right\} &= Pr\left\{\frac{1}{n}h_{X^n}(X^n) > (1+\varepsilon)\left(\frac{1}{n}H(X^n)\right)\right\} \\ &\geq Pr\left\{\frac{1}{n}h_{X^n}(X^n) > (1+\varepsilon)\left(\liminf_{n\to\infty}\frac{1}{n}H(X^n) + \varepsilon\right)\right\}.\end{aligned}$$

Since $\boldsymbol{X}$ is information stable, we obtain that

$$\liminf_{n\to\infty} Pr\left\{\frac{1}{n}h_{X^n}(X^n) > (1+\varepsilon)\left(\liminf_{n\to\infty}\frac{1}{n}H(X^n) + \varepsilon\right)\right\} = 0.$$

By the definition of $\underline{H}_{1^-}(\boldsymbol{X})$, the above implies that

$$\underline{T}(\boldsymbol{X}) = \underline{H}_{1^-}(\boldsymbol{X}) \leq (1+\varepsilon)\left(\liminf_{n\to\infty}\frac{1}{n}H(X^n) + \varepsilon\right).$$

The proof is completed by noting that ε can be made arbitrarily small. □

Observations:

- If the source $\boldsymbol{X}$ is *both* information stable *and* stationary, the above Lemma yields

$$T(\boldsymbol{X}) = \underline{T}(\boldsymbol{X}) = \lim_{n\to\infty}\frac{1}{n}H(X^n).$$

  This implies that given a stationary and information stable source $\boldsymbol{X}$, the classical separation theorem holds for every channel.

- Recall that both Lemmas 3.1 and 3.2 hold not only for arbitrary sources $\boldsymbol{X}$, but also for *arbitrary* fixed blocklength n. This leads us to conclude that they can analogously be employed to provide a simple proof to the conventional source coding theorems [8]:

$$T(\boldsymbol{X}) = \bar{H}(\boldsymbol{X}),$$

and

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

IV Optimistic Channel Coding Theorems

In this section, we state without proving the general expressions for the optimistic ε-capacity[2] ($\bar{C}_\varepsilon$) and for the optimistic capacity ($\bar{C}$) of arbitrary single-user channels. The proofs of these expressions are straightforward once the right definition (of $\bar{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y})$) is made. They employ Feinstein's Lemma and the Verdú-Han lower bound ([14, Theorem 4]), and follow the *same* arguments used in [14] to show the general expressions of the conventional channel capacity

$$C = \sup_{\boldsymbol{X}} \underline{I}_0(\boldsymbol{X};\boldsymbol{Y}) = \sup_{\boldsymbol{X}} \underline{I}(\boldsymbol{X};\boldsymbol{Y}),$$

and the conventional ε-capacity

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

We close this section by proving the formula of $\bar{C}$ for information stable channels.

Definition 4.8 (Channel block code) An (n, M) code for channel W^n with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is a pair of mappings

$$f : \{1, 2, \ldots, M\} \to \mathcal{X}^n$$

and

$$g : \mathcal{Y}^n \to \{1, 2, \ldots, M\}.$$

Its average error probability is given by

$$P_e^{(n)} \triangleq \frac{1}{M} \sum_{m=1}^{M} \sum_{\{y^n : g(y^n) \neq m\}} W^n(y^n | f(m)).$$

[2]The authors would like to point out that the expression of $\bar{C}_\varepsilon$ was also separately obtained in [11, Theorem 7].

Definition 4.9 (Optimistic ε-achievable rate) Fix $0 < \varepsilon < 1$. $R \geq 0$ is an optimistic ε-achievable rate if, for every $\gamma > 0$, there exists a sequence of (n, M) channel block codes such that

$$\frac{\log M}{n} > R - \gamma \quad \text{and} \quad P_e^{(n)} \leq \varepsilon \quad \text{for infinitely many } n.$$

Definition 4.10 (Optimistic ε-capacity $\bar{C}_\varepsilon$) Fix $0 < \varepsilon < 1$. The supremum of optimistic ε-achievable rates is called the optimistic ε-capacity, $\bar{C}_\varepsilon$.

Definition 4.11 (Optimistic capacity $\bar{C}$) The optimistic channel capacity $\bar{C}$ is defined as the supremum of the rates that are optimistic ε-achievable for all $0 < \varepsilon < 1$. It follows immediately from the definition that $\bar{C} = \inf_{0<\varepsilon<1} \bar{C}_\varepsilon = \lim_{\varepsilon \downarrow 0} \bar{C}_\varepsilon$ and that $\bar{C}$ is the supremum of all the rates R for which, for every $\gamma > 0$, there exists a sequence of (n, M) channel block codes such that

$$\frac{1}{n} \log M > R - \gamma \quad \text{and} \quad \liminf_{n \to \infty} P_e^{(n)} = 0.$$

Theorem 4.3 (Optimistic ε-capacity formula)

Fix $0 < \varepsilon < 1$. The optimistic ε-capacity $\bar{C}_\varepsilon$ satisfies

$$\sup_{\boldsymbol{X}} \bar{I}_{\varepsilon^-}(\boldsymbol{X}; \boldsymbol{Y}) \leq \bar{C}_\varepsilon \leq \sup_{\boldsymbol{X}} \bar{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y}). \tag{8}$$

Note that actually $\bar{C}_\varepsilon = \sup_{\boldsymbol{X}} \bar{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y})$, except possibly at the points of discontinuities of $\sup_{\boldsymbol{X}} \bar{I}_\varepsilon(\boldsymbol{X}; \boldsymbol{Y})$ (which are countable).

Theorem 4.4 (Optimistic capacity formula)

The optimistic capacity $\bar{C}$ satisfies

$$\bar{C} = \sup_{\boldsymbol{X}} \bar{I}_0(\boldsymbol{X}; \boldsymbol{Y}).$$

We next investigate the expression of $\bar{C}$ for information stable channels. The expression for the capacity of information stable channels is already known (cf for example [13])

$$C = \liminf_{n \to \infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n),$$

where

$$C_n \triangleq \sup_{X^n} \frac{1}{n} I(X^n; Y^n).$$

We prove a dual formula for $\bar{C}$.

Definition 4.12 (Information stable channels [6, 9]) A channel $\boldsymbol{W}$ is said to be information stable if there exists an input process $\boldsymbol{X}$ such that $0 < C_n < \infty$ for n sufficiently large, and

$$\limsup_{n\to\infty} Pr\left[\left|\frac{i_{X^nW^n}(X^n;Y^n)}{nC_n} - 1\right| > \gamma\right] = 0 \qquad \forall\gamma > 0.$$

Lemma 4.4 Every information stable channel $\boldsymbol{W}$ satisfies

$$\bar{C} = \limsup_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n;Y^n).$$

***Proof*:**

1. $[\bar{C} \le \limsup_{n\to\infty} \sup_{X^n} (1/n) I(X^n;Y^n)]$

By using a similar argument as in the proof of [14, Theorem 8, property h)], we have

$$\bar{I}_0(\boldsymbol{X};\boldsymbol{Y}) \le \limsup_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n;Y^n).$$

Hence,

$$\bar{C} = \sup_{\boldsymbol{X}} \bar{I}_0(\boldsymbol{X};\boldsymbol{Y}) \le \limsup_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n;Y^n).$$

2. $[\bar{C} \ge \limsup_{n\to\infty} \sup_{X^n} (1/n) I(X^n;Y^n)]$

Suppose $\tilde{\boldsymbol{X}}$ is the input process that makes the channel information stable. Fix $\varepsilon > 0$. Then for infinitely many n,

$$\begin{aligned}
& P_{\tilde{X}^nW^n}\left[\frac{1}{n} i_{\tilde{X}^nW^n}(\tilde{X}^n;Y^n) \le (1-\varepsilon)(\limsup_{n\to\infty} C_n - \varepsilon)\right] \\
\le\ & P_{\tilde{X}^nW^n}\left[\frac{i_{\tilde{X}^nW^n}(\tilde{X}^n;Y^n)}{n} < (1-\varepsilon)C_n\right] \\
=\ & P_{\tilde{X}^nW^n}\left[\frac{i_{\tilde{X}^nW^n}(\tilde{X}^n;Y^n)}{nC_n} - 1 < -\varepsilon\right].
\end{aligned}$$

Since the channel is information stable, we get that

$$\liminf_{n\to\infty} P_{\tilde{X}^nW^n}\left[\frac{1}{n} i_{\tilde{X}^nW^n}(\tilde{X}^n;Y^n) \le (1-\varepsilon)(\limsup_{n\to\infty} C_n - \varepsilon)\right] = 0.$$

By the definition of $\bar{C}$, the above immediately implies that

$$\bar{C} = \sup_{\boldsymbol{X}} \bar{I}_0(\boldsymbol{X};\boldsymbol{Y}) \ge \bar{I}_0(\tilde{\boldsymbol{X}};\boldsymbol{Y}) \ge (1-\varepsilon)(\limsup_{n\to\infty} C_n - \varepsilon).$$

Finally, the proof is completed by noting that ε can be made arbitrarily small. □

Observations:

- It is know that for discrete memoryless channels, the optimistic capacity $\bar{C}$ is equal to the (conventional) capacity C [14, 5]. The same result holds for $modulo - q$ additive noise channels with stationary ergodic noise. However, in general, $\bar{C} \geq C$ since $\bar{I}_0(\boldsymbol{X};\boldsymbol{Y}) \geq \underline{I}(\boldsymbol{X};\boldsymbol{Y})$ [3, 4].

- Remark that Theorem 11 in [13] holds if and only if

$$\sup_{\boldsymbol{X}} \underline{I}(\boldsymbol{X};\boldsymbol{Y}) = \sup_{\boldsymbol{X}} \bar{I}_0(\boldsymbol{X};\boldsymbol{Y}).$$

  Furthermore, note that, if $\bar{C} = C$ and there exists an input distribution $P_{\hat{\boldsymbol{X}}}$ that achieves C, then $P_{\hat{\boldsymbol{X}}}$ also achieves $\bar{C}$.

V Examples

We provide four examples to illustrate the computation of C and $\bar{C}$. The first two examples present information stable channels for which $\bar{C} > C$. The third example shows an information unstable channel for which $\bar{C} = C$. These examples indicate that information stability is neither necessary nor sufficient to ensure that $\bar{C} = C$ or thereby the validity of the classical source-channel separation theorem. The last example illustrates the situation where $0 < C < \bar{C} < C_{SC} < \log_2 |\mathcal{Y}|$, where C_{SC} is the channel strong capacity[3]. We assume in this section that all logarithms are in base 2 so that C and $\bar{C}$ are measured in bits.

A. Information Stable Channels

Example 5.1 Consider a nonstationary channel $\boldsymbol{W}$ such that at odd time instances $n = 1, 3, \cdots$, W^n is the product of the transition distribution of a binary symmetric channel with crossover probability 1/8 (BSC(1/8)), and at even time instances $n = 2, 4, 6, \cdots$, W^n is the product of the distribution of a BSC(1/4). It can be easily verified that this channel is information stable. Since the channel is symmetric, a Bernoulli(1/2) input achieves $C_n = \sup_{X^n}(1/n)I(X^n;Y^n)$; thus

$$C_n = \begin{cases} 1 - h_b(1/8), & \text{for } n \text{ odd;} \\ 1 - h_b(1/4), & \text{for } n \text{ even,} \end{cases}$$

[3] The strong (or strong converse) capacity C_{SC} is defined [2] as the infimum of the numbers R for which there exits $\gamma > 0$ such that for all (n, M) codes with $(1/n)\log M > R-\gamma$, $\liminf_{n\to\infty} P_e^{(n)} = 1$. This definition of C_{SC} implies that for any sequence of (n, M) codes with $\liminf_{n\to\infty}(1/n)\log M > C_{SC}$, $P_e^{(n)} > 1 - \varepsilon$ for every $\varepsilon > 0$ and for n sufficiently large. It is shown in [2] that $C_{SC} = \lim_{\varepsilon\uparrow 1} \bar{C}_\varepsilon = \sup_{\boldsymbol{X}} \bar{I}(\boldsymbol{X};\boldsymbol{Y})$.

where $h_b(a) \stackrel{\triangle}{=} -a\log_2 a - (1-a)\log_2(1-a)$ is the binary entropy function. Therefore, $\underline{C} = \liminf_{n\to\infty} C_n = 1 - h_b(1/4)$ and $\bar{C} = \limsup_{n\to\infty} C_n = 1 - h_b(1/8) > \underline{C}$.

Example 5.2 Here we use the information stable channel provided in [13, Section III] to show that $\bar{C} > \underline{C}$. Let $\mathcal{N}$ be the set of all positive integers. Define the set $\mathcal{J}$ as

$$\begin{aligned} \mathcal{J} &\stackrel{\triangle}{=} \{n \in \mathcal{N} : 2^{2i+1} \le n < 2^{2i+2}, i = 0, 1, 2, \ldots\} \\ &= \{2, 3, 8, 9, 10, 11, 12, 13, 14, 15, 32, 33, \cdots, 63, 128, 129, \cdots, 255, \cdots\}. \end{aligned}$$

Consider the following nonstationary symmetric channel $\boldsymbol{W}$. At times $n \in \mathcal{J}$, W_n is a BSC(0), whereas at times $n \notin \mathcal{J}$, W_n is a BSC(1/2). Put $W^n = W_1 \times W_2 \times \cdots \times W_n$. Here again C_n is achieved by a Bernoulli(1/2) input $\hat{X}^n$. We then obtain

$$C_n = \frac{1}{n}\sum_{i=1}^{n} I(\hat{X}_i; Y_i) = \frac{1}{n}\left[J(n)\cdot(1) + (n - J(n))\cdot(0)\right] = \frac{J(n)}{n},$$

where $J(n) \stackrel{\triangle}{=} |\mathcal{J} \cap \{1, 2, \cdots, n\}|$. It can be shown that

$$\frac{J(n)}{n} = \begin{cases} 1 - \dfrac{2}{3} \times \dfrac{2^{\lfloor \log_2 n \rfloor}}{n} + \dfrac{1}{3n}, & \text{for } \lfloor \log_2 n \rfloor \text{ odd;} \\ \dfrac{2}{3} \times \dfrac{2^{\lfloor \log_2 n \rfloor}}{n} - \dfrac{2}{3n}, & \text{for } \lfloor \log_2 n \rfloor \text{ even.} \end{cases}$$

Consequently, $\underline{C} = \liminf_{n\to\infty} C_n = 1/3$ and $\bar{C} = \limsup_{n\to\infty} C_n = 2/3$.

B. Information Unstable Channels

Example 5.3 *The Polya-contagion channel:* Consider a discrete additive channel with binary input and output alphabet $\{0, 1\}$ described by

$$Y_i = X_i \oplus Z_i, \quad i = 1, 2, \cdots,$$

where X_i, Y_i and Z_i are respectively the i-th input, i-th output and i-th noise, and $\oplus$ represents modulo-2 addition. Suppose that the input process is independent of the noise process. Also assume that the noise sequence $\{Z_n\}_{n\ge1}$is drawn according to the Polya contagion urn scheme [1, 10], as follows: an urn originally contains R red balls and B black balls with $R < B$; the *noise* just make successive draws from the urn; after each draw, it returns to the urn $1 + \Delta$ balls of the same color as was just drawn ($\Delta > 0$). The noise sequence $\{Z_i\}$ corresponds to the outcomes of the draws from the Polya urn: $Z_i = 1$ if ith ball drawn is red and $Z_i = 0$, otherwise. Let $\rho \triangleq R/(R+B)$ and $\delta \triangleq \Delta/(R+B)$. It is shown in [1] that the noise process $\{Z_i\}$ is stationary and nonergodic; thus the channel is information unstable.

From Lemma 2 and Section IV in [4, Part I], we obtain

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

and

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

It has been shown [1] that $-(1/n)\log P_{Z^n}(Z^n)$ converges in distribution to the continuous random variable $V \triangleq h_b(U)$, where U is beta-distributed $(\rho/\delta, (1-\rho)/\delta)$, and $h_b(\cdot)$ is the binary entropy function. Thus

$$\bar{H}_{1-\varepsilon}(\boldsymbol{Z}) = \bar{H}_{(1-\varepsilon)^-}(\boldsymbol{Z}) = \underline{H}_{1-\varepsilon}(\boldsymbol{Z}) = \underline{H}_{(1-\varepsilon)^-}(\boldsymbol{Z}) = F_V^{-1}(1-\varepsilon),$$

where $F_V(a) \triangleq Pr\{V \leq a\}$ is the cumulative distribution function of V, and $F_V^{-1}(\cdot)$ is its inverse [1]. Consequently, $C_\varepsilon = \bar{C}_\varepsilon = 1 - F_V^{-1}(1-\varepsilon)$, and $C = \bar{C} = \lim_{\varepsilon \downarrow 0} 1 - F_V^{-1}(1-\varepsilon) = 0$.

Example 5.4 Let $\tilde{W}_1, \tilde{W}_2, \ldots$ consist of the channel in Example 5.2, and let $\hat{W}_1, \hat{W}_2, \ldots$ consist of the channel in Example 5.3. Define a new channel $\boldsymbol{W}$ as follows:

$$W_{2i} = \tilde{W}_i \quad \text{and} \quad W_{2i-1} = \hat{W}_i \quad \text{for } i = 1, 2, \cdots.$$

As in the previous examples, the channel is symmetric, and a Bernoulli(1/2) input maximizes the inf/sup information rates. Therefore for a Bernoulli(1/2) input $\boldsymbol{X}$, we have

$$\begin{aligned}
& Pr\left\{\frac{1}{n}\log\frac{P_{W^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \leq \theta\right\} \\
= & \left\{\begin{array}{ll} Pr\left\{\dfrac{1}{2i}\left[\log\dfrac{P_{\tilde{W}^i}(Y^i|X^i)}{P_{Y^i}(Y^i)} + \log\dfrac{P_{\hat{W}^i}(Y^i|X^i)}{P_{Y^i}(Y^i)}\right] \leq \theta\right\}, & \text{if } n = 2i; \\ Pr\left\{\dfrac{1}{2i+1}\left[\log\dfrac{P_{\tilde{W}^i}(Y^i|X^i)}{P_{Y^i}(Y^i)} + \log\dfrac{P_{\hat{W}^{i+1}}(Y^{i+1}|X^{i+1})}{P_{Y^{i+1}}(Y^{i+1})}\right] \leq \theta\right\}, & \text{if } n = 2i+1; \end{array}\right. \\
= & \left\{\begin{array}{ll} 1 - Pr\left\{-\dfrac{1}{i}\log P_{Z^i}(Z^i) < 1 - 2\theta + \dfrac{1}{i}J(i)\right\}, & \text{if } n = 2i; \\ 1 - Pr\left\{-\dfrac{1}{i+1}\log P_{Z^{i+1}}(Z^{i+1}) < 1 - \left(2 - \dfrac{1}{i+1}\right)\theta + \dfrac{1}{i+1}J(i)\right\}, & \text{if } n = 2i+1. \end{array}\right.
\end{aligned}$$

The fact that $-(1/i)\log[P_{Z^i}(Z^i)]$ converges in distribution to the continuous random variable $V \triangleq h_b(U)$, where U is beta-distributed $(\rho/\delta, (1-\rho)/\delta)$, and the fact that

$$\liminf_{n\to\infty}(1/n)J(n) = 1/3 \quad \text{and} \quad \limsup_{n\to\infty}(1/n)J(n) = 2/3$$

imply that

$$\underline{i}_{\boldsymbol{XW}}(\theta) \triangleq \liminf_{n\to\infty} Pr\left\{\frac{1}{n}\log\frac{P_{W^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \theta\right\} = 1 - F_V\left(\frac{5}{3} - 2\theta\right),$$

and

$$\bar{i}_{\boldsymbol{XW}}(\theta) \triangleq \limsup_{n\to\infty} Pr\left\{\frac{1}{n}\log\frac{P_{W^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \theta\right\} = 1 - F_V\left(\frac{4}{3} - 2\theta\right).$$

Consequently,

$$\bar{C}_\varepsilon = \frac{5}{6} - \frac{1}{2}F_V^{-1}(1-\varepsilon) \qquad \text{and} \qquad C_\varepsilon = \frac{2}{3} - \frac{1}{2}F_V^{-1}(1-\varepsilon).$$

Thus

$$0 < C = \frac{1}{6} < \bar{C} = \frac{1}{3} < C_{SC} = \frac{5}{6} < \log_2|\mathcal{Y}| = 1.$$

References

[1] F. Alajaji and T. Fuja, "A communication channel modeled on contagion," *IEEE Trans. Inform. Theory*, Vol. 40, No. 6, November 1994.

[2] P.-N. Chen and F. Alajaji, "Strong converse, feedback capacity and hypothesis testing," *Proc. of CISS*, John Hopkins Univ., Baltimore, March 1995.

[3] P.-N. Chen and F. Alajaji, "Generalization of information measures," *Proc. Int. Symp. Inform. Theory & Applications*, Victoria, Canada, September 1996.

[4] P.-N. Chen and F. Alajaji, "Generalized source coding theorems and hypothesis testing," *Journal of the Chinese Institute of Engineering*, Vol. 21, No. 3, pp. 283-303, May 1998.

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

[6] R. L. Dobrushin, "General formulation of Shannon's basic theorems of information theory," *AMS Translations,* Vol. 33, pp. 323-438, AMS, Providence, RI, 1963.

[7] T. S. Han, *Information-Spectrum Methods in Information Theory*, (in Japanese), Baifukan Press, Tokyo, 1998.

[8] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, Vol. 39, No. 3, pp. 752–772, May 1993.

[9] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes,* Holden-Day, 1964.

[10] G. Polya, "Sur quelques points de la théorie des probabilités." *Ann. Inst. H. Poincarré*, Vol. 1, pp. 117-161, 1931.

[11] Y. Steinberg, "New converses in the theory of identification via channels," *IEEE Trans. Inform. Theory*, Vol. 44, No. 3, pp. 984–998, May 1998.

[12] Y. Steinberg and S. Verdú, "Simulation of random processes and rate-distortion theory," *IEEE Trans. Inform. Theory*, Vol. 42, No. 1, pp. 63-86, January 1996.

[13] S. Vembu, S. Verdú and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inform. Theory*, Vol. 41, No. 1, pp. 44–54, January 1995.

[14] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, Vol. 40, No. 4, pp. 1147–1157, July 1994.