LARGE DEVIATIONS AND DETERMINISTIC MEASURES OF INFORMATION

[†]Charalambos D. Charalambous,[*] [‡]Seddik M. Djouadi

[†]School of Information Technology and Engineering, University of Ottawa
Ottawa, Ontario, K1N 6N5, Canada
E-mail: chadcha@site.uottawa.ca

[‡]System Engineering Department, University of Arkansas at Little rock
Little Rock, AR 72204
E-mail: msdjoudi@ualr.edu

Keywords. Large Deviations, Max-Plus Algebra, Information Measures.

Abstract

In recent years, Large Deviations theory has found important applications in many areas of engineering and science including communication and control systems. The objective of this note is to introduce and explore connections between certain fundamental concepts of Large Deviations theory and Information Theory, by introducing deterministic measures of information. The connections are established through the so-called rate functional associated with the Large Deviations principle, which lead to a natural definition of (max,plus) deterministic measure of information.

1 Preliminary Mathematical Constructs

In this section we shall introduce the measures of interest, which in the next section, are related to Large Deviations Theory. Some of the definitions can be found in [2, 6, 1, 7, 8]. Let "+" and "." denote the usual addition and multiplication operations, defined over the reals $\Re$. Let $\bar{\Re} \stackrel{\triangle}{=} \{-\infty\} \cup \Re = [-\infty, \infty]$ denote the extended real line, which has all the properties of a compact interval, so that every subset $R \subset \bar{\Re}$ has a supremum and an infimum (although they can be $-\infty$ and ∞, respectively). The algebras of interest are the (log,plus) and the (max,plus), which are defined over the extended real line $\bar{\Re}$.

With respect to the (log,plus) algebra the operations of addition and multiplication are defined by

$$\begin{aligned} &a \oplus^{\epsilon} b \stackrel{\triangle}{=} \epsilon \log \Big\{ \exp(\frac{a}{\epsilon}) + \exp(\frac{b}{\epsilon}) \Big\}; \\ &a \otimes^{\epsilon} b \stackrel{\triangle}{=} \epsilon \log \Big\{ \exp(\frac{a}{\epsilon}) . \exp(\frac{b}{\epsilon}) \Big\}, \\ &a, b \in \bar{\Re}, \quad \epsilon \in (0, \infty) \end{aligned}$$

With respect to the (max,plus) algebra the operations of addition and multiplication are defined by

$$a \oplus b \stackrel{\triangle}{=} \max\{a, b\}; \quad a \otimes b \stackrel{\triangle}{=} a + b, \quad a, b \in \bar{\Re} \quad (1)$$

Note that $-\infty$ is the additive identity, and 0 is the multiplicative identity of the (log,plus) and the (max,plus) algebras, while there is no additive inverse. Moreover, for any sequence $\{a_i\}_{i=1}^n \subset \bar{\Re}$ which is independent of the parameter $\epsilon \in (0, \infty)$ we have

$$\lim_{\epsilon \rightarrow 0} \bigoplus_{i=1}^{n}{}^{\epsilon} a_i \stackrel{\triangle}{=} \lim_{\epsilon \rightarrow 0} \epsilon \log \Big\{ \sum_{i=1}^{n} \exp\Big(\frac{a_i}{\epsilon}\Big) \Big\}$$

[*]Adjunct Professor with the Department of Electrical & Computer Engineering, McGill University, Montréal, P.Q., CANADA H3A 2A7. This work was supported by the Natural Science and Engineering Research Council of Canada through an operating grant.

$$= \bigoplus_{i=1}^{n} a_i = \max_i a_i, \quad \forall \{a_i\}_{i=1}^n \subset \bar{\Re} \tag{2}$$

$$\lim_{\epsilon \to 0} \bigotimes_{i=1}^{n}{}^{\epsilon} a_i \stackrel{\triangle}{=} \lim_{\epsilon \to 0} \epsilon \log \Big\{ \prod_{i=1}^{n} \exp\Big\{\frac{a_i}{\epsilon}\Big\}\Big\}$$

$$= \bigotimes_{i=1}^{n} a_i = \sum_{i=1}^{n} a_i, \quad \forall \{a_i\}_{i=1}^n \subset \bar{\Re} \tag{3}$$

Consequently, the (log,plus) algebra converges to the (max,plus) algebra as $\epsilon \to 0$. This convergence is a property of Large Deviations theoty, and it applies to general complete separable metric spaces on which probability measures are defined (we shall say more about this in the next section).

Next, introduce the definitions measures with respect to the (log, plus) and the (max, plus) algebras. For $\epsilon \in [0, \infty)$, let I_ϵ^a and I_ϵ^m denote the additive and multiplicative identities of the (log,plus) operations $\oplus^\epsilon, \otimes^\epsilon$, and I_0^a and I_0^m those of the (max,plus) operations $\oplus, \otimes$; then $I_\epsilon^a = -\infty$ and $I_\epsilon^m = 0$ for any $\epsilon \in [0, \infty)$.

Definition 1.1 *Let $(\Omega, \mathcal{F})$ be a measurable space.*

1. *For $\epsilon \in [0, \infty)$, μ^ϵ is a measure with respect to the (log,plus) algebra if the following conditions hold.*

$$\begin{aligned} &i) \;\; \mu^\epsilon(A) \in [I_\epsilon^a, I_\epsilon^m], \quad \forall A \in \mathcal{F}; \\ &ii) \;\; \mu^\epsilon(\Omega) = I_\epsilon^m; \\ &iii) \;\; \mu^\epsilon\Big(\bigcup_i A_i\Big) = \bigoplus_i^{\epsilon} \mu^\epsilon(A_i), \\ &\qquad A_i \bigcap A_j = \emptyset, \forall i \neq j, \quad \forall \{A_i\} \in \mathcal{F}. \end{aligned}$$

2. *μ is a measure with respect to the (max,plus) algebra if the following conditions hold.*

$$\begin{aligned} &i) \;\; \mu(A) \in [I_0^a, I_0^m], \quad \forall A \in \mathcal{F}; \\ &ii) \;\; \mu(\Omega) = I_0^m; \\ &iii) \;\; \mu\Big(\bigcup_i A_i\Big) = \bigoplus_i \mu(A_i), \\ &\qquad A_i \bigcap A_j = \emptyset, \forall i \neq j, \quad \forall \{A_i\} \in \mathcal{F}. \end{aligned}$$

This coincides with the standard definition for the usual algebra. Next, we illustrate how the (log,plus) and the (max,plus) measures arize when one considers the probability of rare events, studied in the Theory of Large Deviations [4, 5, 3].

Let $\Big\{\Big(\Omega, \mathcal{F}, P^\epsilon\Big)\Big\}_{\epsilon > 0}$ be a family of probability measures indexed by $\epsilon > 0$.

For any $A \in \mathcal{F}$, define

$$\mu^\epsilon(A) \stackrel{\triangle}{=} \epsilon \log P^\epsilon(A), \quad A \in \mathcal{F}, \tag{4}$$

(with $\log(0) \stackrel{\triangle}{=} -\infty$). Then, for $\{A_i\}$ a countable collection of disjoint sets in $\mathcal{F}$ we have

$$\begin{aligned} &\mu^\epsilon(\bigcup_{i=1}^{\infty} A_i) = \bigoplus_{i=1}^{\infty}{}^{\epsilon} \mu^\epsilon(A_i), \\ &A_i \bigcap A_j = \emptyset, \forall i \neq j, \quad \{A_i\} \in \mathcal{F}. \end{aligned} \tag{5}$$

Therefore, (5) implies that P^ϵ is a (log,plus) probability measure indexed by ϵ.

Next, suppose the following limit exists

$$\begin{aligned} \mu(A) &\stackrel{\triangle}{=} \lim_{\epsilon \to 0} \mu^\epsilon(A) = \lim_{\epsilon \to 0} \epsilon \log P^\epsilon(A) \\ &= ess \sup \Big\{ I(x); x \in A \Big\}, \quad A \in \mathcal{F} \end{aligned} \tag{6}$$

where $I : \Omega \to [-\infty, 0]$ is an upper semicontinuous (u.s.c) function. In general, the above limit will exists whenever P^ϵ is absolutely continous with respect to the Lebesgue measure [4, 5, 3]. Then μ is a (max,plus) finite-additive measure which satisfies conditions i)-iii) of Definition 1.1 with $I_\epsilon^a = -\infty, I_\epsilon^m = 0$. Moreover, for all $I : \Omega \to [-\infty, 0]$ which are measurable with respect to $\mathcal{F}$, and $\{A_i\} \in \mathcal{F}$ countable and disjoint,

$$\begin{aligned} &\mu(\bigcup_{i=1}^{\infty} A_i) = ess \sup \Big\{ I(\omega) : \omega \in \bigcup_{i=1}^{\infty} A_i \Big\} \\ &= \sup_i \Big\{ ess \sup \{ I(\omega) : \omega \in A_i \} \Big\} = \sup_i \mu(A_i), \\ &A_i \cap A_j = \emptyset, \forall i \neq j, \quad \{A_i\}_{i=1}^n \in \mathcal{F}, \end{aligned}$$

thus μ is also countable additive. Unfortunately, μ does not have a density because $\mu(\{\omega\}) = -\infty, \forall \omega \in \Omega$. However, if we replace (6) by $\mu(A) \stackrel{\triangle}{=} \sup \Big\{ I(\omega); \omega \in A \Big\}, A \in \mathcal{F}$, then μ has I as its density. Moreover, the measure of the even $A \in \mathcal{F}$ can be expressed in terms of the indicator function of A with respect to the (max,plus) algebra as follows. Define the indicator function χ_A of $A \in \mathcal{F}$, by $\chi_A(\omega) = 0 = I_\epsilon^m$ if $\omega \in A$, and

$\chi_A(\omega) = -\infty = I^a_\epsilon$ if $\omega \notin A$. Then $I^m_\epsilon = 0$ is the indicator of $A = \Omega$ and $I^a_\epsilon = -\infty$ is the indicator of $A = \emptyset$. Therefore, $\mu(A) = \sup\Big\{\chi_A(\omega) + I(\omega); \omega \in \Omega\Big\}, \quad A \in \mathcal{F}$. Let $\mathcal{I}_A$ is the indicator function of the event $A \in \mathcal{F}$ associated with the family of probability spaces $\Big\{\Big(\Omega, \mathcal{F}, P^\epsilon\Big)\Big\}_{\epsilon>0}$, defined by $\mathcal{I}_A(\omega) \stackrel{\triangle}{=} \begin{cases} 1 & \text{if}\, \omega \in A \\ 0 & \text{if}\, \omega \notin A \end{cases}$. The above construction implies that the expectation of the indicator function $\chi_A, A \in \mathcal{F}$ with respect to the (max,plus) measure μ is $E_\mu[\chi_A] = 0 \otimes \mu(A) = \mu(A) = \lim_{\epsilon\to 0} \mu^\epsilon(A)$ $= \lim_{\epsilon \to 0} \epsilon E_{\mu^\epsilon}[\chi_A] = \lim_{\epsilon\to 0} \epsilon \log E_{P^\epsilon}[\mathcal{I}_A]$. Next, we may considerer (max,plus) simple functions of the form $f(\omega) \stackrel{\triangle}{=} \sum_{i=1}^n a_i \otimes \chi_{A_i}(\omega), \quad \{A_i\} \in \mathcal{F}, A_i \bigcap A_j = \emptyset, \forall i \neq j$. Then the expectation of f with respect to the (max,plus) measure μ is defined by $E_\mu[f] = \bigoplus_{i=1}^n \alpha_i \otimes \mu(A_i), \{A_i\} \in \mathcal{F}, A_i \bigcap A_j = \emptyset, \forall i \neq j$. Next, we may extend the definition of expectation to nonnegative measurable functions, and then measurable functions.

We shall show in the next section that for a family of probability measures which satisfies the Large Deviations Principle, the above construction of measures is directly obtained by invoking two fundamental results of Large Deviations Theory, namely, the Contraction Principle and the Laplace-Varadhan Lemma, provided every element of $\mathcal{F}$ is a continuity set of the action functional.

2 Large Deviations Theory and Related Deterministic Measures

Throughout we let $\mathcal{X}$ be a Polish space (e.g., complete separable metric space), $\mathcal{B}_\mathcal{X}$ the Borel algebra of $\mathcal{X}$, and $\{P^\epsilon\}_{\epsilon>0}$ a family of probability measures on $\mathcal{B}_\mathcal{X}$. Next we introduce the precise conditions for an underlying family of probability spaces to satisfy the LDP (see [4, 5, 3]).

Definition 2.1 *(Large Deviations Principle) [4, 5, 3]. Let $\Big\{\Big(\mathcal{X}, \mathcal{B}_\mathcal{X}, P^\epsilon\Big)\Big\}_{\epsilon>0}$ be a family of complete probability spaces indexed by ϵ and let*

$$\mu^\epsilon_\mathcal{X}(\mathcal{A}) = \epsilon \log P^\epsilon(\mathcal{A}), \quad \mu_\mathcal{X}(\mathcal{A}) \stackrel{\triangle}{=} \lim_{\epsilon\to 0} \mu^\epsilon_\mathcal{X}(\mathcal{A}), \quad \mathcal{A} \in \mathcal{B}_\mathcal{X}$$

provided the limit exists.
We say that this probability space satisfies the Large Deviations Principle (LDP) with real-valued rate function $I_\mathcal{X}(\cdot)$, denoted by $\Big\{\Big(\mathcal{X}, \mathcal{B}_\mathcal{X}\ , P^\epsilon\Big)\Big\}_{\epsilon>0} \sim I_\mathcal{X}(x)$ if there exists a function $I_\mathcal{X} : \mathcal{X} \to [-\infty, 0]$ called the action functional which satisfies the following properties.

1. *$-\infty \leq I_\mathcal{X}(x) \leq 0, \forall x \in \mathcal{X}$*
2. *$I_\mathcal{X}(\cdot)$ is Upper Semicontinuous (u.s.c))*
3. *For each $m > -\infty$ the set $\{x; m \leq I_\mathcal{X}(x)\}$ is a compact set in $\mathcal{X}$.*
4. *For each $\mathcal{C} \in \mathcal{B}_\mathcal{X}$*
$$\lim_{\epsilon\to 0} \sup \epsilon \log P^\epsilon(\mathcal{C}) \leq \sup_{x\in\bar{\mathcal{C}}} I_\mathcal{X}(x) \qquad (7)$$
where $\bar{\mathcal{C}}$ is the closure of the set $\mathcal{C} \in \mathcal{B}_\mathcal{X}$.
5. *For each $\mathcal{O} \in \mathcal{B}_\mathcal{X}$*
$$\lim_{\epsilon\to 0} \inf \epsilon \log P^\epsilon(\mathcal{O}) \geq \sup_{x\in\mathcal{O}^0} I_\mathcal{X}(x) \qquad (8)$$
where $\mathcal{C}^0$ is the interior of the set $\mathcal{O} \in \mathcal{B}_\mathcal{X}$.
6. *If $\mathcal{C} \in \mathcal{B}_\mathcal{X}$ is such that*
$$\sup_{x\in\mathcal{C}^0} I_\mathcal{X}(x) = \sup_{x\in\mathcal{C}} I_\mathcal{X}(x) = \sup_{x\in\bar{\mathcal{C}}} I_\mathcal{X}(x) \qquad (9)$$
then
$$\mu_\mathcal{X}(\mathcal{C}) = \sup_{x\in\mathcal{X}} I_\mathcal{C}(x) \qquad (10)$$
and $\mathcal{C} \in \mathcal{X}$ is called a continuity set of $I_\mathcal{X}(\cdot)$. If (9) holds for all elements $\mathcal{C} \in \mathcal{B}_\mathcal{X}$ then $\mathcal{B}_\mathcal{X}$ is called a continuity $\sigma-$algebra of $I_\mathcal{X}(\cdot)$.
In 4, 5, 6 the supremum over an empty set is defined to be $-\infty$.

Next, we shall introduce two fundamental Theorems associated with the LDP, which will lead to the conclusion that if $\mathcal{B}_\mathcal{X}$ is a family of continuity sets of $I_\mathcal{X}(\cdot)$ then $\mu_\mathcal{X}(\cdot)$ define above is a deterministic (max,plus) measure.

Theorem 2.2 *(Contraction Principle) [4, 5, 3] Let $\Big\{\Big(\mathcal{X}, \mathcal{B}_\mathcal{X}, P^\epsilon\Big)\Big\}_{\epsilon>0} \sim I_\mathcal{X}(x)$. Let $F^\epsilon : \mathcal{X} \to \mathcal{Y}$*

be a continuous map where $\mathcal{Y}$ is another complete separable metric space satisfying

$$\lim_{\epsilon\to 0} F^\epsilon = F \quad \text{uniformly on compact subsets of } \mathcal{X} \qquad (11)$$

Then the induced measures $\Big\{Q^\epsilon\Big\}_{\epsilon>0}$ on $\mathcal{Y}$, namely, $Q^\epsilon = P^\epsilon \circ F^{\epsilon,-1}$ satisfy the LDP with rate function $I_\mathcal{Y}(\cdot)$ given by

$$I_\mathcal{Y}(y) = \sup\Big\{I_\mathcal{X}(x); y = F(x), x\in\mathcal{X}\Big\} \qquad (12)$$

Lemma 2.3 *(Laplace-Varadhan Lemma) [4, 5, 3] Suppose $\Big\{\Big(\mathcal{X},\mathcal{B}_\mathcal{X},P^\epsilon\Big)\Big\}_{\epsilon>0} \sim I_\mathcal{X}(x)$. Then for any bounded continuous function $F(\cdot)$ on $\mathcal{X}$*

$$\begin{aligned}&\lim_{\epsilon\to 0}\epsilon\log\int_\mathcal{X}\exp\Big(\frac{F(x)}{\epsilon}\Big)dP^\epsilon(x)\\ &= \sup_{x\in\mathcal{X}}\Big\{F(x)+I_\mathcal{X}(x)\Big\}\end{aligned} \qquad (13)$$

The Laplace-Varadhan Lemma gives rise to the definition of the expectation with respect to the (max,plus) measure, (at least when every element of $\mathcal{B}_\mathcal{X}$ is a continuity element of $I_\mathcal{X}$, and thus satisfies (9)), among many other essential properties.

Assumption 2.4 *In subsequent discussions, and unless otherwise state, we assume that all LD statements are with respect to continuity sets of $I_\mathcal{X}(\cdot)$, so that (9) is satisfied and (10) is well defined.*

2.1 Deterministic Measures

Armed with the above statements we shall show that if the family of probability spaces $\Big\{\Big(\mathcal{X},\mathcal{B}_\mathcal{X},P^\epsilon\Big)\Big\}_{\epsilon>0}$ satisfies the LDP with rate $I_\mathcal{X}(\cdot)$, then the rate induces a (max,plus) measure defined by

$$\mu_\mathcal{X}(\mathcal{O}) \triangleq \lim_{\epsilon\to 0}\epsilon\log P^\epsilon(\mathcal{O}), \quad \forall\mathcal{O}\in\mathcal{B}_\mathcal{X} \qquad (14)$$

Indeed, the countable additivity property of the measure for any finite collection $\{\mathcal{A}_i\}_{i=1}^n$ of disjoint sets in $\mathcal{B}_\mathcal{X}$, follows from the statements preceding Lemma 3.2. Thus, each $x\in\mathcal{X}$ is associate with $I_\mathcal{X}(x)$, which is also called the self-rate functional. The more likely $x\in\mathcal{X}$ is the larger the value of $I_\mathcal{X}(x)\in[-\infty,0]$.

Theorem 2.5 *Let $\Big\{\Big(\mathcal{X},\mathcal{B}_\mathcal{X},P^\epsilon\Big)\Big\}_{\epsilon>0} \sim I_\mathcal{X}(x)$. Then $\forall\mathcal{O}\in\mathcal{B}_\mathcal{X}$*

$$\begin{aligned}I_\mathcal{X}(\mathcal{O}) &\triangleq \lim_{\epsilon\to 0}\epsilon\log P^\epsilon(\mathcal{O}) = \sup_{x\in\mathcal{O}} I_\mathcal{X}(x)\\ &= \sup\Big\{\chi_\mathcal{O}(x)+I_\mathcal{X}(x); x\in\mathcal{X}\Big\},\end{aligned} \qquad (15)$$

is a finite-additive probability measure.

Expectation with Respect to Deterministic Measures. Clearly, for max-plus probability measures the rate functional $I_\mathcal{X}:\mathcal{X}\to[-\infty,0]$ is the analog of the density function for usual probability measures. Therefore, by Laplace-Varadhan lemma we have the following definition of expectation with respect to the max-plus measure.

Definition 2.6 *Suppose $\Big\{\Big(\mathcal{X},\mathcal{B}_\mathcal{X},P^\epsilon\Big)\Big\}_{\epsilon>0} \sim I_\mathcal{X}(x)$ and $F:\mathcal{X}\to\Re$ is any bounded measurable function.*
The expectation of $F(\cdot)$ with respect to the deterministic measure of $I_\mathcal{X}(\cdot)$ is defined by

$$\begin{aligned}E_{I_\mathcal{X}}(F) &\triangleq \lim_{\epsilon\to 0}\epsilon\log\int_\mathcal{X}\exp\Big(\frac{F(x)}{\epsilon}\Big)dP^\epsilon(x)\\ &= \sup_{x\in\mathcal{X}}\Big\{F(x)+I_\mathcal{X}(x)\Big\}.\end{aligned} \qquad (16)$$

The definition of the expectation given by (16) can be extended to measurable functions F which are u.s.c..

Induced Rate Functionals. Induced (max,plus) measures are obtained by the contraction principle.

Joint and Marginal Rate Functionals. The next theorem extends these results to a pair of R.V.'s., and therefore to an arbitrary number of R.V.'s.

Lemma 2.7 *Let $\Big\{\Big(\mathcal{X}\times\mathcal{Y},\mathcal{B}_{\mathcal{X}\times\mathcal{Y}},P^\epsilon_{X,Y}\Big)\Big\}_{\epsilon>0} \sim I_{\mathcal{X},\mathcal{Y}}:\mathcal{X}\times\mathcal{Y}\to[-\infty,0]$. Then*

$$\begin{aligned}I_\mathcal{X}(\mathcal{O}_x) &\triangleq \lim_{\epsilon\to 0}\epsilon\log\int_{\mathcal{O}_x}\int_\mathcal{Y} dP^\epsilon_{X,Y}(x,y)\\ &= \sup_{x\in\mathcal{O}_x}\sup_{y\in\mathcal{Y}} I_{\mathcal{X},\mathcal{Y}}(x,y), \quad \forall\mathcal{O}_x\in\mathcal{B}_\mathcal{X}\\ &= \sup\Big\{\chi_{\mathcal{O}_x}(x)+\chi_\mathcal{Y}(y)+I_{\mathcal{X},\mathcal{Y}}(x,y);\\ &(x,y)\in\mathcal{O}_x\times\mathcal{Y}\Big\}, \quad \forall\mathcal{O}_x\in\mathcal{B}_\mathcal{X}\end{aligned} \qquad (17)$$

in the marginal measure on $(\mathcal{X},\mathcal{B}_\mathcal{X})$ and $I_{\mathcal{X},\mathcal{Y}}(x,y)$ is the joint-rate functional.

Letting $\mathcal{O}_x = \{x\}$ and $\mathcal{O}_y = \{y\}$ in (17), (??), respectively, we obtain the marginal rate functionals from the joint-rate functions as follows.

$$I_{\mathcal{X}}(x) = \sup_{y\in\mathcal{Y}} I_{\mathcal{X},\mathcal{Y}}(x,y); \quad I_{\mathcal{Y}}(y) = \sup_{x\in\mathcal{X}} I_{\mathcal{X},\mathcal{Y}}(x,y) \quad (18)$$

Conditional Rate Functionals. We can also define the conditional rate functional that emerges from the definition of the conditional expectation of two R.V.'s $(X,Y) : \Omega \to \mathcal{X} \times \mathcal{Y}$.

Lemma 2.8 *Let* $\left\{\left(\mathcal{X} \times \mathcal{Y}, \mathcal{B}_{\mathcal{X}\times\mathcal{Y}}, P^{\epsilon}_{X,Y}\right)\right\}_{\epsilon>0} \sim I_{\mathcal{X},\mathcal{Y}} : \mathcal{X} \times \mathcal{Y} \to [-\infty, 0]$.
Then for all $\mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}$, $y \in \mathcal{Y}$

$$\begin{aligned} I_{\mathcal{X}|\mathcal{Y}}(\mathcal{O}_x|y) &\stackrel{\triangle}{=} \lim_{\epsilon\to 0} \epsilon \log P^{\epsilon}_{X|Y}\big(X \in \mathcal{O}_x | Y = y\big) \\ &= \sup\Big\{I_{\mathcal{X},\mathcal{Y}}(x,y) + \chi_{\mathcal{O}_{\mathcal{X}}}(x); x \in \mathcal{X}\Big\} - I_{\mathcal{Y}}(y) \\ &= \sup\Big\{I_{\mathcal{X}|\mathcal{Y}}(x|y) + \chi_{\mathcal{O}_{\mathcal{X}}}(x); x \in \mathcal{X}\Big\}. \qquad (19) \end{aligned}$$

The next statement relates the joint-rate functional and the conditional rate functional. It is equivalent to the Bayes rule with respect to the usual probability measure.

Corollary 2.9 *Let* $\left\{\left(\mathcal{X} \times \mathcal{Y}, \mathcal{B}_{\mathcal{X}\times\mathcal{Y}}, P^{\epsilon}_{X,Y}\right)\right\}_{\epsilon>0} \sim I_{\mathcal{X},\mathcal{Y}} : \mathcal{X} \times \mathcal{Y} \to [-\infty, 0]$. *Then the joint-rate functional of x and y is the sum of the conditional rate-functional of x given y and the marginal rate functional of y:*

$$I_{\mathcal{X},\mathcal{Y}}(x,y) = I_{\mathcal{X}|\mathcal{Y}}(x|y) + I_{\mathcal{Y}}(y) \qquad (20)$$

and $\forall \mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}, \forall y \in \mathcal{Y}$

$$I_{\mathcal{X}|\mathcal{Y}}(\mathcal{O}_x|y) = \sup_{x\in\mathcal{O}_x}\Big\{I_{\mathcal{X},\mathcal{Y}}(x,y)\Big\} - I_{\mathcal{Y}}(y) \le 0 \quad (21)$$

Note that independence of the R.V.'s X and Y implies that $I_{\mathcal{X}|\mathcal{Y}}(x|y) = I_{\mathcal{X}}(x)$ and $I_{\mathcal{X},\mathcal{Y}}(x,y) = I_{\mathcal{X}}(x) + I_{\mathcal{Y}}(y)$, suggesting that knowledge of the values of y will result in a reduction of uncertainty about x, that is, $I_{\mathcal{X}|\mathcal{Y}}(\mathcal{O}_x|y), \mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}$ denotes the uncertainly reduction of $\mathcal{O}_x$ given the observed y. Therefore we have the following definition of independence.

Definition 2.10 *We say that the variables* $x \in \mathcal{X}$ *and* $y \in \mathcal{Y}$ *are independent if and only if* $I_{\mathcal{X}|\mathcal{Y}}(x|y) = I_{\mathcal{X}}(x) \iff I_{\mathcal{X},\mathcal{Y}}(x,y) = I_{\mathcal{X}}(x) + I_{\mathcal{Y}}(y)$, $\forall (x,y) \in (\mathcal{X}, \mathcal{Y})$.

2.2 Information Theoretic Measures

In this section we illustrate the importance of the rate functionals in defining entropy, conditional entropy, mutual information, etc., and we investigate some of their properties.

Entropy of Rate Functionals. Let $\left\{\left(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, P^{\epsilon}_X\right)\right\}_{\epsilon>0}$ be a family of probability spaces, and $X : (\Omega, \mathcal{F}) \to (\mathcal{X}, B_{\mathcal{X}})$ a R.V. defined on it. Suppose that X is the output of a discrete information source having a finite alphabet containing M symbols, $\mathcal{X} = \Big\{x_1, x_2, \ldots, x_M\Big\}$, and each x_i is produced according to the probability $P^{\epsilon}_X(\{x_i\}), 1 \le i \le M$. If x_i occurs then the amount of information associated with the known occurrence of x_i is defined by $-\log P^{\epsilon}_X(\{x_i\})$. If X is a discrete memoryless source, then the information generated each time a symbol x_i is selected is $-\log P^{\epsilon}_X(\{x_i\})$ bits. Moreover, the average amount of information per source output symbol, known as the average information, uncertainty or entropy is

$$H(P^{\epsilon}_X) = -\sum_{i=1}^{M} P^{\epsilon}_X(\{x_1\}) \log P^{\epsilon}_X(\{x_1\}), \qquad (22)$$

in bits/symbol.

Definition 2.11 *Let* $\left\{\left(\mathcal{X} \times \mathcal{Y}, \mathcal{B}_{\mathcal{X}\times\mathcal{Y}}, P^{\epsilon}_{X,Y}\right)\right\}_{\epsilon>0} \sim I_{\mathcal{X},\mathcal{Y}} : \mathcal{X} \times \mathcal{Y} \to [-\infty, 0]$.

1. *The Entropy Rate Functional of any event* $\mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}$ *denoted by* $H_{\mathcal{X}}(\mathcal{O}_x)$ *is defined by*

$$\begin{aligned} H_{\mathcal{X}}(\mathcal{O}_x) &\stackrel{\triangle}{=} \lim_{\epsilon\to 0} \epsilon \log \frac{1}{P^{\epsilon}_X\Big(X \in \mathcal{O}_x\Big)} \\ &= -\mu_{\mathcal{X}}(\mathcal{O}_x), \ \mathcal{O}_x \in \mathcal{B}_{\mathcal{X}} \qquad (23) \end{aligned}$$

2. *The Joint-Entropy Rate Functional of any events* $\mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}, \mathcal{O}_y \in \mathcal{B}_{\mathcal{Y}}$ *is defined by*

$$H_{\mathcal{X},\mathcal{Y}}(\mathcal{O}_x, \mathcal{O}_y) \stackrel{\triangle}{=} -\mu_{\mathcal{X},\mathcal{Y}}(\mathcal{O}_x, \mathcal{O}_y) \qquad (24)$$

3. *The Conditional Entropy Rate Functional of any events* $\mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}$ *given another event* $\mathcal{O}_y \in \mathcal{B}_{\mathcal{Y}}$ *is defined by*

$$H_{\mathcal{X}|\mathcal{Y}}(\mathcal{O}_x|\mathcal{O}_y) = -\mu_{\mathcal{X},\mathcal{Y}}(\mathcal{O}_x, \mathcal{O}_y) + \mu_{\mathcal{Y}}(\mathcal{O}_y) \ (25)$$

According to the above definition the entropy rate functional stated in (23) enjoys analogous properties as the entropy $H(P^{\epsilon}_X)$, defined by (22). Suppose $\mu_{\mathcal{X}}(A) = \sup\Big\{I_{\mathcal{X}}(x); x \in \mathcal{O}_x\Big\}$, $\mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}$ is a (max,plus) measure induced by a variable X taking values in the discrete space, $\mathcal{X} = \Big\{x_1, x_2, \ldots, x_M\Big\}$, in which X models the output of a discrete information source, producing symbols according to the (max,plus) law $\Big\{I_{\mathcal{X}}(\{x_i\})\Big\}$. If symbols x_i occurs then the amount of information associated with the known occurrence of x_i is defined by $-I_{\mathcal{X}}(\{x_i\}) \geq 0$. Moreover, we have the following properties. i) The entropy rate functional is nonnegative, $H_{\mathcal{X}}(\mathcal{O}_x) \geq 0$, $\forall \mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}$, and equal to zero, $H_{\mathcal{X}}(\mathcal{O}_x) = 0$, if and only if at least one $I_{\mathcal{X}}(\{x_i\})$, is equal to zero. Moreover, unlike the entropy function $H(P^{\epsilon}_X)$ which can be negative for continuous R.V., the entropy rate functional is never negative, because the rate functional $I_{\mathcal{X}} : \mathcal{X} \to [-\infty, 0]$. ii) The entropy rate functional $H_{\mathcal{X}}(\mathcal{O}_x\Big)$ is a continuous function of the rate functional $I_{\mathcal{X}}$. iii) The entropy rate functional, $H_{\mathcal{X}}(\mathcal{O}_x)$, is a concave function of the rate functional, $I_{\mathcal{X}}(x)$. iv) The entropy rate functional of a set of variables $\{X_i\}$ is less than or equal to the sum of the entropy rate functional of the individual variables, that is, $H_{\mathcal{X}_1,\ldots,\mathcal{X}_n}(\mathcal{O}_{x_1}, \ldots, \mathcal{O}_{x_n}) \leq \sum_{i=1}^{n} H_{\mathcal{X}_i}(\mathcal{O}_{x_i})$, and equality holds of the variable are independent (by Corollary 2.9).

Moreover, it is easily seen that the following symmetry holds.

$$
\begin{aligned}
H_{\mathcal{X},\mathcal{Y}}(\mathcal{O}_x, \mathcal{O}_y) &= H_{\mathcal{X}}(\mathcal{O}_x) + H_{\mathcal{Y}|\mathcal{X}}(\mathcal{O}_y|\mathcal{O}_x) \\
&= H_{\mathcal{Y}}(\mathcal{O}_y) + H_{\mathcal{X}|\mathcal{Y}}(\mathcal{O}_x|\mathcal{O}_y). \qquad (26)
\end{aligned}
$$

Thus, joint-Entropy rate functional of any events $\mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}, \mathcal{O}_y \in \mathcal{B}_{\mathcal{Y}}$ is the entropy of one event plus the conditional entropy rate functional of the other event.

Mutual Information Rate Functionals. Next, we introduce the mutual information rate functional, which is a measure of the amount of information that one variable contains about another variable.

Lemma 2.12 *Let* $\Big\{\Big(\mathcal{X} \times \mathcal{Y}, \mathcal{B}_{\mathcal{X}\times\mathcal{Y}}, P^{\epsilon}_{X,Y}\Big)\Big\}_{\epsilon>0} \sim I_{\mathcal{X},\mathcal{Y}} : \mathcal{X} \times \mathcal{Y} \to [-\infty, 0]$.

The Mutual Information Rate Functional of two events $\mathcal{O}_x \in \mathcal{B}_{\mathcal{X}}, \mathcal{O}_y \in \mathcal{B}_{\mathcal{Y}}$ *is defined by*

$$
\begin{aligned}
I_{\mathcal{X};\mathcal{Y}}\Big(\mathcal{O}_x; \mathcal{O}_y\Big) &= \sup_{(x,y)\in\big(\mathcal{O}_x, \mathcal{O}_y\big)} I_{\mathcal{X},\mathcal{Y}}(x, y) \\
&- \sup_{x\in\mathcal{O}_x} I_{\mathcal{X}}(x) - \sup_{y\in\mathcal{O}_y} I_{\mathcal{Y}}(y),
\end{aligned}
$$

and, if X *and* Y *are independent,* $I_{\mathcal{X};\mathcal{Y}}(\mathcal{O}_x; \mathcal{O}_y) = 0$.

The above lemma suggests that the self-mutual information of x with respect to y should be defined by $I_{\mathcal{X};\mathcal{Y}}(x; y) \stackrel{\triangle}{=} -I_{\mathcal{X}}(x) - I_{\mathcal{Y}}(y) + I_{\mathcal{X},\mathcal{Y}}(x, y)$.

References

[1] P. Bernhard, "A separation theorem for expected value and feared value discrete time control", ESIAM: Control Optim. and Calc. of Vars., *1* (1996), 191–206.

[2] V.P. Maslov, "On a new principle of superposition for optimization problems", Russian Math. Surveys, *42* (1987) 43–54.

[3] J.-D. Deuschel and D.W. Stroock, *Large Deviations,* Academic Press Inc., (1989).

[4] S.R.S. Varadhan, *Large Deviations and Applications,* Society for Industrial and Applied Mathematics, (1984).

[5] D.W. Stroock, *An Introduction to the Theory of Large Deviations,* Springer-Verlag, (1984).

[6] M. Akian *Theory of Cost Measures: Convergence of Decision Variables,* Rapport de Recherche 2611, INRIA, (1995).

[7] M. Akian, "Densities of Idempotent Measures and Large Deviations," Transactions of the American Mathematical Society, *Vol.351, Num.11*, 4515-4543, (1999).

[8] W.M. McEneaney and C.D. Charalambous, "Large Deviations Theory, Induced Log-Plus and Max-Plus Measures and their Applications," *in Mathematical Theory of Network and Systems 2000,* Perpignan, France, June 19-23, 2000.

[9] A.V. Balakrishnan, *Applied Functional Analysis,* Springer-Verlag, (1976).