PAPER *Special Section on Information Theory and Its Applications*

Error Exponent for Coding of Memoryless Gaussian Sources with a Fidelity Criterion

Shunsuke IHARA†, ***Regular Member and*** **Masashi KUBO**††, ***Nonmember***

SUMMARY We are interesting in the error exponent for source coding with fidelity criterion. For each fixed distortion level Δ, the maximum attainable error exponent at rate R, as a function of R, is called the reliability function. The minimum rate achieving the given error exponent is called the minimum achievable rate. For memoryless sources with finite alphabet, Marton (1974) gave an expression of the reliability function. The aim of the paper is to derive formulas for the reliability function and the minimum achievable rate for memoryless Gaussian sources.

key words: *error exponent, minimum achievable rate, reliability function, memoryless Gaussian source, large deviation property*

1. Introduction

In this paper we study the error exponent for source coding with fidelity criterion.

Let $X = \{X_n\}$ be an information source. We denote by $\mathcal{X}$ and $\mathcal{Y}$ the input alphabet space and the output alphabet space, respectively. Each X_n is an $\mathcal{X}$ valued random variable. An encoder φ_n and a decoder ψ_n are given by mappings

$$\varphi_n : \mathcal{X}^n \longrightarrow \mathcal{M}_n, \qquad \psi_n : \mathcal{M}_n \longrightarrow \mathcal{Y}^n,$$

where $\mathcal{M}_n = \{1, 2, ..., M_n\}$ and M_n is an integer. A sequence $\{(\varphi_n, \psi_n)\}_{n=1,2,...}$ of pairs of encoders and decoders is simply called a code. Without loss of generality, we may identify $\mathcal{Y}$ with $\mathcal{X}$ and $\psi_n \circ \varphi_n$ with φ_n, respectively. Hence, a code $\varphi \equiv \{\varphi_n\}_{n=1,2,...}$ is a sequence of mappings

$$\varphi_n : \mathcal{X}^n \longrightarrow \mathcal{X}^n, \quad n = 1, 2, ...$$

The rate $\|\varphi\|$ of the code $\varphi = \{\varphi_n\}$ is defined as

$$\|\varphi\| = \limsup_{n\to\infty} \frac{1}{n} \log |\varphi_n|,$$

where $|\varphi_n|$ denotes the cardinality of the set $\varphi_n(\mathcal{X}^n)$.

The fidelity criterion is defined by a distortion function $\rho : \mathcal{X} \times \mathcal{X} \longrightarrow [0, \infty)$. On $\mathcal{X}^n \times \mathcal{X}^n$ we define a distortion function by

$$\rho_n(x_1^n, y_1^n) = \frac{1}{n} \sum_{k=1}^{n} \rho(x_k, y_k),$$

where $x_1^n \equiv (x_1, ..., x_n) \in \mathcal{X}^n$. For each distortion level $\Delta > 0$, we are interesting in the asymptotic behavior of the error probability

$$e_n(\varphi_n, \Delta) = P(\rho_n(X_1^n, \varphi_n(X_1^n)) > \Delta).$$

Let $R(\Delta; X)$ be the rate-distortion function of X. The rate distortion theorem says that if a rate greater than $R(\Delta; X)$ is available then there exists a code $\varphi = \{\varphi_n\}$ such that the error probability $e_n(\varphi_n, \Delta)$ goes to zero as $n \to \infty$. On the other hand, if the converse theorem holds, $e_n(\varphi_n, \Delta)$ goes to one provided that the rate of φ is less than $R(\Delta; X)$.

For a stationary memoryless source (SMS) with finite alphabet Marton [1] (see also [2]) showed that, if the rate is greater than $R(\Delta; X)$, the minimum error probability $e_n(\varphi_n, \Delta)$ converges to zero exponentially and gave an explicit expression for the error exponent.

To study the error exponent we define the minimum achievable rate and the reliabibility function.

Definition 1: For each non-negative number r, a code $\varphi = \{\varphi_n\}$ is said to be (Δ, r)-achievable if

$$\limsup_{n\to\infty} \frac{1}{n} \log e_n(\varphi_n, \Delta) \le -r.$$

We denote by $\mathcal{C}(\Delta, r)$ the family of all (Δ, r)-achievable codes. The minimum (Δ, r)-achievable rate $R_e(\Delta, r)$ is defined by

$$R_e(\Delta, r) = \inf\{\|\varphi\|; \varphi \in \mathcal{C}(\Delta, r)\}.$$

For each non-negative number R, the minimum error exponent is defined by

$$r_e(\Delta, R) = -\inf_{\varphi\,:\,\|\varphi\|\le R} \limsup_{n\to\infty} \frac{1}{n} \log e_n(\varphi_n, \Delta).$$

As a function of R, $r_e(\Delta, R)$ is called the reliability function.

Remark: In case of without distortion the minimum achievable rate has been studied by Han [3].

Note that, for each fixed distortion level Δ, the reliability function $r_e(\Delta, \cdot)$ is the inverse function of $R_e(\Delta, \cdot)$, and vice versa.

In the case where the rate is less than $R(\Delta; X)$, under some assumptions, the error probability $e_n(\varphi_n, \Delta)$ goes to one exponentially as $n \to \infty$. In other words,

Manuscript received January 18, 2000.
Manuscript revised April 12, 2000.

†The author is with the School of Informatics and Sciences, Nagoya University, Nagoya-shi, 464-8601 Japan.
††The author is with the Faculty of Education, Tokoha Gakuen University, Shizuoka-shi, 420-0911 Japan.

the probability of correct decoding converges to zero exponentially. To study the speed of the convergence we introduce the minimum achievable rate and the reliability functoin for correct decoding.

Definition 2: For each non-negative number r, a code $\varphi = \{\varphi_n\}$ is said to be (Δ, r)-achievable for correct decoding if

$$\liminf_{n\to\infty} \frac{1}{n} \log(1 - e_n(\varphi_n, \Delta)) \geq -r.$$

We denote by $\mathcal{C}^*(\Delta, r)$ the family of all (Δ, r)-achievable codes for correct decoding. The minimum (Δ, r)-achievable rate (for correct decoding) $R_e^*(\Delta, r)$ is defined by

$$R_e^*(\Delta, r) = \inf\{\|\varphi\|; \varphi \in \mathcal{C}^*(\Delta, r)\}.$$

The reliability function $r_e^*(\Delta, R)$ (for correct decoding) is defined by

$$r_e^*(\Delta, R) = - \sup_{\varphi\,:\,\|\varphi\| \leq R} \liminf_{n\to\infty} \frac{1}{n} \log(1 - e_n(\varphi_n, \Delta)).$$

The reliability functon was first investigated by Marton [1] (see also [2]) for a SMS with finite alphabet. She showed that the reliability function is expressed in terms of the divergence and the rate-distortion function. For probability distributions μ and ν, we denote by $D(\nu\|\mu)$ the divergence (or Kullback-Leibler information number) of ν with respect to μ. For a random variable ξ with distribution ν, the rate-distortion function $R(\Delta; \xi) \equiv R(\Delta; \nu)$ is defined as

$$R(\Delta; \xi) = \inf_{\eta} \{I(\xi, \eta); E[\rho(\xi, \eta)] \leq \Delta\},$$

where $I(\xi, \eta)$ denotes the mutual information between ξ and η. Marton [1] proved the following theorem.

Theorem 1: Let $X = \{X_n\}$ be a SMS with finite alphabet and denote by μ the probability distribution of X_n. Then the reliability function is given by

$$r_e(\Delta, R) = \inf_{\nu} \{D(\nu\|\mu); R(\Delta; \nu) > R\}. \tag{1}$$

For the sources with continuous plphabet, as far as we know, no formulas have been obtained for the minimum achievable rate or the reliability function. In this paper we study the Gaussian source which is one of the most important sources with continuous alphabet.

The aim of the paper is to derive explicit formulas for the minimum achievable rate and the reliability function for a stationary memoryless Gaussian source (SMGS), where the fidelity criterion is defined by mean squared error (Theorem 2 and Theorem 3). It will also be shown that the same formula (1) holds for the SMGS (Theorem 4). For the SMS with finite alphabet (1) has been proved by using the standard method of type sets or typical sets (cf. [1], [2]). We should note that the method of types does not work for the sources with continuous plphabet. We shall prove our results by using large deviation properties of Gaussian random sequences, and the known results on the rate-distortion function and the divergence of Gaussian random variables as well.

Our main theorems will be stated in Sect. 2, and the proofs will be given in Sect. 3.

2. Error Exponent for Coding of Memoryless Gaussian Sources

We consider a SMGS $X = \{X_n\}$ with distribution $N(0, \sigma^2)$. Namely, $X = \{X_n\}$ is a sequence of i.i.d. Gaussian random variables with distribution $N(0, \sigma^2)$. We assume that the fidelity criterion is defined by

$$\rho(x, y) = |x - y|^2, \quad x, y \in \boldsymbol{R}.$$

For simplicity we put

$$\|x_1^n\|_n^2 = \frac{1}{n} \sum_{k=1}^{n} |x_k|^2, \quad x_1^n \in \boldsymbol{R}^n.$$

If ν is a Gaussian distribution with mean a and variance γ^2, we simply denote as $\nu \sim N(a, \gamma^2)$. The following properties concerning the divergence and the rate-distortion function of Gaussian distributions are well known (see e.g. [4]). The first inequality in (3) is due to Binia-Zakai-Ziv [5].

Lemma 1: Let $\mu \sim N(0, \sigma^2)$. Assume that ν is a probability distribution on $\boldsymbol{R}$ with mean a and variance γ^2, $\nu_a \sim N(a, \gamma^2)$ and $\nu_0 \sim N(0, \gamma^2)$. Then

$$\begin{aligned} D(\nu\|\mu) &\geq D(\nu_a\|\mu) \\ &= \frac{1}{2}\left(\frac{\gamma^2}{\sigma^2} - 1 - \log\frac{\gamma^2}{\sigma^2} + \frac{a^2}{\sigma^2}\right), \end{aligned} \tag{2}$$

and

$$\begin{aligned} R(\Delta; \nu_a) - D(\nu\|\nu_a) &\leq R(\Delta; \nu) \leq R(\Delta; \nu_a) \\ &= R(\Delta; \nu_0) = \frac{1}{2} \log \max\left(\frac{\gamma^2}{\Delta}, 1\right). \end{aligned} \tag{3}$$

We now state our first main result to give the explicit formulas for the minimum achievable rate and the reliability function of the SMGS.

Theorem 2: Let $X = \{X_n\}$ be the SMGS with distribution $N(0, \sigma^2)$. Then

$$R_e(\Delta, r) = \frac{1}{2} \log \max\left(\frac{\alpha^2}{\Delta}, 1\right), \tag{4}$$

$$r_e(\Delta, R) = \frac{1}{2}\left(\frac{\alpha^2}{\sigma^2} - 1 - \log\frac{\alpha^2}{\sigma^2}\right) \quad \text{if} \quad R > R(\Delta; X_1), \tag{5}$$

$$r_e(\Delta, R) = 0 \quad \text{if} \quad R \leq R(\Delta; X_1), \tag{6}$$

where, in (4) $\alpha \geq \sigma$ is determined by

$$r = \frac{1}{2}\left(\frac{\alpha^2}{\sigma^2} - 1 - \log\frac{\alpha^2}{\sigma^2}\right), \tag{7}$$

and, in (5) α is determined by

$$R = \frac{1}{2}\log\frac{\alpha^2}{\Delta}. \tag{8}$$

Our second result provides the formulas for the minimum achievable rate and the minimum error exponent for correct decoding.

Theorem 3: For the same SMGS as in Theorem 2 we have

$$R_e^*(\Delta, r) = \frac{1}{2}\log\max\left(\frac{\beta^2}{\Delta}, 1\right), \tag{9}$$

$$r_e^*(\Delta, R) = \frac{1}{2}\left(\frac{\beta^2}{\sigma^2} - 1 - \log\frac{\beta^2}{\sigma^2}\right) \quad \text{if} \quad R \leq R(\Delta; X_1), \tag{10}$$

$$r_e^*(\Delta, R) = 0 \quad \text{if} \quad R > R(\Delta; X_1), \tag{11}$$

where $0 < \beta \leq \sigma$ in (9) (resp. (10)) is determined by (7) (resp. (8)), by replacing α with β.

It seems to be interesting to show that the minimum achievable rates $R_e(\Delta, r)$ and $R_e^*(\Delta, r)$ and the reliability functions $r_e(\Delta, R)$ and $r_e^*(\Delta, R)$ can be expressed in terms of the rate-distortion function and the divergence, and show that the formula (1) remains true for the SMGS.

Theorem 4: For the SMGS we have

$$R_e(\Delta, r) = \max_\nu\{R(\Delta; \nu); D(\nu\|\mu) \leq r\}, \tag{12}$$

$$r_e(\Delta, R) = \min_\nu\{D(\nu\|\mu); R(\Delta; \nu) \geq R\}, \tag{13}$$

$$R_e^*(\Delta, r) = \min_\nu\{R(\Delta; \nu); D(\nu\|\mu) \leq r\}, \tag{14}$$

and

$$r_e^*(\Delta, R) = \min_\nu\{D(\nu\|\mu); R(\Delta; \nu) \leq R\}, \tag{15}$$

where μ denotes the distribution of X_n.

Remark: For the SMGS, as will be shown in later, the maximum or the minimums in (12)–(15) are attained by Gaussian distributions.

3. Proof of Theorems

In this section we prove our theorems.

We denote by

$$B_n(c_1^n, \gamma) = \{x_1^n \in \boldsymbol{R}^n; \|x_1^n - c_1^n\|_n^2 \leq \gamma^2\}$$

the n-dimensional sphere with center c_1^n and radius $\sqrt{n}\gamma$. Note that the volume $|B_n(c_1^n, \gamma)|$ of $B_n(c_1^n, \gamma)$ is equal to

$$|B_n(c_1^n, \gamma)| = \frac{(\sqrt{\pi}\sqrt{n}\gamma)^n}{\Gamma\left(\frac{n}{2} + 1\right)}. \tag{16}$$

Let $X = \{X_n\}$ be the SMGS with distribution $\mu \sim N(0, \sigma^2)$. To prove the theorems we need a lemma concerning the asymptotic behavior of $X = \{X_n\}$. For each $a > 0$, we define $D(\mathcal{P}_a\|\mu)$ by

$$D(\mathcal{P}_a\|\mu) = \inf\{D(\nu\|\mu); \nu \in \mathcal{P}_a\},$$

where $\mathcal{P}_a$ denotes the class of all probability distributions ν satisfying

$$\int_{\boldsymbol{R}} |x - a|^2 d\nu(x) = \Delta.$$

Lemma 2: Let $X = \{X_n\}$ be the SMGS.
(i) For any $a > 0$ we have

$$\lim_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}\sum_{k=1}^n |X_k - a|^2 \leq \Delta\right) = -D(\mathcal{P}_a\|\mu). \tag{17}$$

In particular, if $a^2 = \sigma^2 + \Delta$, then

$$\lim_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}\sum_{k=1}^n |X_k - a|^2 \leq \Delta\right) = -\frac{1}{2}\log\frac{a^2}{\Delta}. \tag{18}$$

(ii) If $0 < \beta < \sigma$, then

$$\lim_{n\to\infty}\frac{1}{n}\log P\left(\|X_1^n\|_n < \beta\right) = -\frac{1}{2}\left(\frac{\beta^2}{\sigma^2} - 1 - \log\frac{\beta^2}{\sigma^2}\right). \tag{19}$$

(iii) If $\alpha > \sigma$, then

$$\lim_{n\to\infty}\frac{1}{n}\log P\left(\|X_1^n\|_n > \alpha\right) = -\frac{1}{2}\left(\frac{\alpha^2}{\sigma^2} - 1 - \log\frac{\alpha^2}{\sigma^2}\right). \tag{20}$$

The proof will be given in Appendix.

We now proceed to prove the theorems.

Proof of Theorem 2: We start with the proof of (4). If $\Delta \geq \alpha^2$, by encoding as $\varphi_n(x_1^n) = 0$, $\forall x_1^n \in \boldsymbol{R}^n$, we have $\varphi = \{\varphi_n\} \in \mathcal{C}(\Delta, r)$. Since $\|\varphi\| = 0$, we have $R_e(\Delta, r) = 0$ and (4). We assume $0 < \Delta < \alpha^2$ in the sequel.

Proof of Converse Part: We shall prove the inequality

$$R_e(\Delta, r) \geq \frac{1}{2}\log\frac{\alpha^2}{\Delta}. \tag{21}$$

For this end it sufficies to prove

$$\|\varphi\| \geq \frac{1}{2}\log\frac{\alpha^2}{\Delta}, \tag{22}$$

for any (Δ, r)-achievable code $\varphi = \{\varphi_n\} \in \mathcal{C}(\Delta, r)$. We put $\varphi_n(\mathcal{X}^n) \equiv \{(y^{(j)})_1^n; j = 1, ..., |\varphi_n|\}$ and $A_n = \cup_{j=1}^{|\varphi_n|} B_n((y^{(j)})_1^n, \sqrt{\Delta})$. Let $\gamma_n > 0$ be the number such that

$$|B_n(0, \gamma_n)| = |A_n|,$$

where $|A_n|$ denotes the volume of A_n, and let $\gamma = \limsup_{n\to\infty} \gamma_n$. It is clear that

$$|A_n| \leq |\varphi_n||B_n(0, \sqrt{\Delta})|.$$

Therefore, using (16), we have

$$|\varphi_n| \geq \frac{|B_n(0, \gamma_n)|}{|B_n(0, \sqrt{\Delta})|} = \left(\frac{\gamma_n}{\sqrt{\Delta}}\right)^n$$

and

$$\|\varphi\| \geq \frac{1}{2}\log\frac{\gamma^2}{\Delta}. \tag{23}$$

Since the probability density function $p_n(x_1^n)$ of X_1^n depends only on $\|x_1^n\|_n$ and is stricty decreasing with respect to $\|x_1^n\|_n$, it is clear that

$$P(X_1^n \notin B_n(0, \gamma_n)) \leq P(X_1^n \notin A_n) = e_n(\varphi_n, \Delta),$$

so that

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\log P(X_1^n \notin B_n(0, \gamma_n))\\ &\quad\leq \limsup_{n\to\infty}\frac{1}{n}\log e_n(\varphi_n, \Delta) \leq -r.\end{aligned} \tag{24}$$

Using (iii) of Lemma 2 and noting that the function $f(x) = x - \log x$ is increasing on $[1, \infty)$, we have

$$\begin{aligned}&-\frac{1}{2}\left(\frac{\gamma^2}{\sigma^2} - 1 - \log\frac{\gamma^2}{\sigma^2}\right)\\ &\quad\leq \limsup_{n\to\infty}\frac{1}{n}\log P(X_1^n \notin B_n(0, \gamma_n)).\end{aligned} \tag{25}$$

It follows from (7), (24) and (25) that

$$\frac{1}{2}\left(\frac{\gamma^2}{\sigma^2} - 1 - \log\frac{\gamma^2}{\sigma^2}\right) \geq \frac{1}{2}\left(\frac{\alpha^2}{\sigma^2} - 1 - \log\frac{\alpha^2}{\sigma^2}\right),$$

meaning that $\gamma \geq \alpha$. Therefore the desired inequality (22) follows from (23).

Proof of Direct Part: For each

$$R > \frac{1}{2}\log\frac{\alpha^2}{\Delta},$$

by using the random coding method, we shall show that there exists a code $\varphi \in \mathcal{C}(\Delta, r)$ with a rate less than R. Let $Y = \{Y_n\}$ be a SMGS with distribution $N(0, \tau^2)$, where

$$\tau^2 = \alpha^2 - \Delta.$$

Applying (i) of Lemm 2 to $Y = \{Y_n\}$, we have

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}\sum_{k=1}^n |Y_k - \alpha|^2 \leq \Delta\right)\\ &\quad= -\frac{1}{2}\log\frac{\alpha^2}{\Delta}.\end{aligned}$$

Therefore, for any $\frac{1}{2}\log(\alpha^2/\Delta) < R_0 < R$, there exists an integer n_0 such that

$$P\left(\frac{1}{n}\sum_{k=1}^n |Y_k - \alpha|^2 \leq \Delta\right) > e^{-nR_0}, \quad n \geq n_0.$$

Since the distribution of Y_1^n is rotation invariant, for any $x_1^n \in B_n(0, \alpha)$, it is clear that

$$P\left(\|Y_1^n - x_1^n\|_n^2 \leq \Delta\right) \geq P\left(\frac{1}{n}\sum_{k=1}^n |Y_k - \alpha|^2 \leq \Delta\right),$$

so that

$$P\left(\|Y_1^n - x_1^n\|_n^2 \leq \Delta\right) > e^{-nR_0}, \quad n \geq n_0.$$

Suppose that a family $\boldsymbol{Y} = \{Y^{(m)}\}$ of independent copies $Y^{(m)} = \{Y_n^{(m)}\}$, $m = 1, 2, ...$, of $Y = \{Y_n\}$ is independent of X. We denote by

$$W_n(\boldsymbol{Y}, x, \Delta) = \inf\{m; \|(Y^{(m)})_1^n - x_1^n\|_n^2 \leq \Delta\}$$

the waiting time for string matching. Then, for any sequence $x = \{x_n\}$ satisfying $\|x_1^n\|_n \leq \alpha$ $(n \geq n_0)$, we have

$$\begin{aligned}&P\left(W_n(\boldsymbol{Y}, x, \Delta) > \lceil e^{nR}\rceil\right)\\ &\quad= \prod_{m=1}^{\lceil e^{nR}\rceil} P(\|(Y^{(m)})_1^n - x_1^n\|_n^2 > \Delta)\\ &\quad\leq (1 - e^{-nR_0})^{e^{nR}}\\ &\quad= (1 - e^{-nR_0})^{e^{nR_0}e^{n(R-R_0)}}\\ &\quad< e^{-e^{n(R-R_0)}}, \quad n \geq n_0,\end{aligned}$$

where $\lceil x\rceil$ is the smallest integer $\geq x$, and we have used the inequality $(1 - x^{-1})^x < e^{-1}$ $(x > 1)$. Therefore

$$P(\|X_1^n\|_n \leq \alpha,\, W_n(\boldsymbol{Y}, X, \Delta) > \lceil e^{nR}\rceil) < e^{-e^{n(R-R_0)}}$$

and

$$\begin{aligned}&P(W_n(\boldsymbol{Y}, X, \Delta) > \lceil e^{nR}\rceil)\\ &\quad< P(\|X_1^n\|_n > \alpha) + e^{-e^{n(R-R_0)}}, \quad n \geq n_0.\end{aligned}$$

Since $\boldsymbol{Y}$ is independent of X, this implies that there exists at least a realization $\boldsymbol{y} = \{y^{(m)}\}$ of $\boldsymbol{Y} = \{Y^{(m)}\}$ for which

$$\begin{aligned}&P(W_n(\boldsymbol{y}, X, \Delta) > \lceil e^{nR}\rceil)\\ &\quad< P(\|X_1^n\|_n > \alpha) + e^{-e^{n(R-R_0)}}, \quad n \geq n_0,\end{aligned} \tag{26}$$

holds. Using $y^{(m)} = \{y_n^{(m)}\}$, $m = 1, 2, ...$, we define a code $\varphi = \{\varphi_n\}$ as follows. If $\|x_1^n - (y^{(i)})_1^n\|_n^2 > \Delta$, $i = 1, ..., k-1$, and $\|x_1^n - (y^{(k)})_1^n\|_n^2 \leq \Delta$, then $\varphi_n(x_1^n) = (y^{(k)})_1^n$ $(k = 1, 2, ..., \lceil e^{nR} \rceil)$; otherwise $\varphi_n(x_1^n) = (y^{(1)})_1^n$. Then clearly

$$\|\varphi\| = \lim_{n\to\infty} \frac{1}{n} \log |\varphi_n| = R \tag{27}$$

and

$$\begin{aligned} &P(\|X_1^n - \varphi_n(X_1^n)\|_n^2 > \Delta) \\ &\quad \leq P(W_n(\boldsymbol{y}, X, \Delta) > \lceil e^{nR} \rceil). \end{aligned}$$

Thus, it follows from (7), (20) and (26) that

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n} \log P(\|X_1^n - \varphi_n(X_1^n)\|_n^2 > \Delta) \\ &\quad \leq \lim_{n\to\infty} \frac{1}{n} \log P(\|X_1^n\|_n > \alpha) \\ &\quad = -r, \end{aligned}$$

meaning that φ is (Δ, r)-achievable. Therefore from (27) we have

$$R_e(\Delta, r) \leq R.$$

Since $R > \frac{1}{2} \log(\alpha^2/\Delta)$ is arbitrary,

$$R_e(\Delta, r) \leq \frac{1}{2} \log \frac{\alpha^2}{\Delta}. \tag{28}$$

The desired equation (4) follows from (28) and (21).

Since $r_e(\Delta, \cdot)$ is the inverse function of $R_e(\Delta, \cdot)$, (5) and (6) follow from (4). q.e.d.

Proof of Theorem 3: If $\Delta \geq \sigma^2$, then it is clear that $R_e^*(\Delta, r) \equiv 0$ and $r_e(\Delta, R) \equiv 0$. We assume $0 < \Delta < \sigma^2$ in the following.

We shall prove (9). By the definition of β and (19) we have

$$\lim_{n\to\infty} \frac{1}{n} \log P(X_1^n \in B_n(0, \beta)) = -r. \tag{29}$$

Therefore, when $\beta^2 < \Delta$, ovbiously $R^*(\Delta, r) = 0$ and (9) holds. We assume $\beta^2 > \Delta$ in the sequel. In the same manner as in the proof of (21) we can prove the inequality

$$R_e^*(\Delta, r) \geq \frac{1}{2} \log \frac{\beta^2}{\Delta}.$$

Hence it is enough to prove the converse inequality

$$R_e^*(\Delta, r) \leq \frac{1}{2} \log \frac{\beta^2}{\Delta}. \tag{30}$$

Let $Y = \{Y_n\}$ be an i.i.d. sequence with distribution $N(0, \tau^2)$, where

$$\tau^2 = \beta^2 - \Delta,$$

and $\boldsymbol{Y} = \{Y^{(m)}\}$ be a family of independent copies $Y^{(m)} = \{Y_n^{(m)}\}$, $m = 1, 2, ...$, of $Y = \{Y_n\}$. We assume that $\boldsymbol{Y}$ is independent of X. Let R and R_0 be real numbers such that

$$\frac{1}{2} \log \frac{\beta^2}{\Delta} < R_0 < R.$$

In the same way as in the proof of Theorem 2 one can show that there is an integer n_0 such that

$$P\left(W_n(\boldsymbol{Y}, x, \Delta) > \lceil e^{nR} \rceil\right) < e^{-e^{n(R-R_0)}}, \tag{31}$$

for any $n \geq n_0$ and $x_1^n \in B_n(0, \beta)$. Since

$$\begin{aligned} &P(W_n(\boldsymbol{Y}, X, \Delta) \leq \lceil e^{nR} \rceil) \\ &\quad \geq P(W_n(\boldsymbol{Y}, X, \Delta) \leq \lceil e^{nR} \rceil, X_1^n \in B_n(0, \beta)) \\ &\quad \geq P(X_1^n \in B_n(0, \beta)) \\ &\qquad - P(W_n(\boldsymbol{Y}, X, \Delta) > \lceil e^{nR} \rceil), \end{aligned}$$

it follows from (29) and (31) that

$$\liminf_{n\to\infty} \frac{1}{n} \log P(W_n(\boldsymbol{Y}, X, \Delta) \leq \lceil e^{nR} \rceil) \geq -r.$$

This means that there exists at least a realization $\boldsymbol{y} = \{y^{(m)}\}$ of $\boldsymbol{Y} = \{Y^{(m)}\}$ which satisfies

$$\liminf_{n\to\infty} \frac{1}{n} \log P(W_n(\boldsymbol{y}, X, \Delta) \leq \lceil e^{nR} \rceil) \geq -r.$$

We now define a code $\varphi = \{\varphi_n\}$ in the same way as in the proof of Theorem 2. Then we have $\|\varphi\| = R$ and, since

$$\begin{aligned} &P(\|X_1^n - \varphi_n(X_1^n)\|_n^2 \leq \Delta) \\ &\quad = P(W_n(y, X, \Delta) \leq \lceil e^{nR} \rceil), \end{aligned}$$

we see that

$$\liminf_{n\to\infty} \frac{1}{n} \log P(\|X_1^n - \varphi_n(X_1^n)\|_n^2 \leq \Delta) \geq -r$$

and $\varphi \in \mathcal{C}^*(\Delta, r)$. Therefore

$$R_e^*(\Delta, r) \leq R.$$

Thus we have obtained (30).

Since $r_e^*(\Delta, \cdot)$ is the inverse function of $R_e^*(\Delta, \cdot)$, (10) and (11) follow from (9). q.e.d.

Proof of Theorem 4: We shall prove (12). Let $\mu \sim N(0, \sigma^2)$ and $\nu^* \sim N(0, \alpha^2)$. Then from (3) we know that (4) is rewritten as $R_e(\Delta, r) = R(\Delta; \nu^*)$. Therefore (12) is equivalent to

$$R(\Delta; \nu^*) = \max_{\nu} \{R(\Delta; \nu); D(\nu\|\mu) \leq r\}.$$

It follows from (2) and (7) that $D(\nu^*\|\mu) = r$. Thus we have only to show

$$R(\Delta; \nu) \leq R(\Delta; \nu^*), \tag{32}$$

for any ν such that $D(\nu\|\mu) \leq r$. We may assume that the expectation of ν is 0. Denoting by γ^2 the variance of ν, let $\nu_0 \sim N(0,\gamma^2)$. Since

$$D(\nu_0\|\mu) \le D(\nu\|\mu) \le r = D(\nu^*\|\mu)$$

and $\sigma < \alpha$, we know that $\gamma \le \alpha$, so that

$$R(\Delta;\nu) \le R(\Delta;\nu_0) \le R(\Delta;\nu^*).$$

Thus we have obtained (32). We can prove that (13) is equivalent to Eqs. (5) and (6).

We turn to prove (14). We first assume that $\Delta < \beta^2$. Let $\nu^* \sim N(0,\beta^2)$. Then it follows from (3) and (9) that (14) is equivalent to

$$R(\Delta;\nu^*) = \min_{\nu}\{R(\Delta;\nu); D(\nu\|\mu) \le r\}. \tag{33}$$

Since

$$D(\nu^*\|\mu) = \frac{1}{2}\left(\frac{\beta^2}{\sigma^2} - 1 - \log\frac{\beta^2}{\sigma^2}\right) = r, \tag{34}$$

to demonstrate (33) it is sufficient to prove

$$R(\Delta;\nu^*) \le R(\Delta;\nu) \tag{35}$$

for any ν such that $D(\nu\|\mu) \le r$. We may assume that the expectation of ν is 0. Let $\nu_0 \sim N(0,\gamma^2)$, where γ^2 is the variance of ν. By Lemma 1, if $\gamma < \beta$, then $D(\nu\|\mu) \ge D(\nu_0\|\mu) > D(\nu^*\|\mu) = r$. This contradicts with the assumption $D(\nu\|\mu) \le r$. Therefore $\gamma \ge \beta$. Using (3) we have

$$\begin{aligned} R(\Delta;\nu) &\ge R(\Delta;\nu_0) - D(\nu\|\nu_0) \\ &= \frac{1}{2}\log\frac{\gamma^2}{\Delta} - D(\nu\|\nu_0). \end{aligned} \tag{36}$$

It is easily seen that

$$\begin{aligned} &D(\nu\|\nu_0) \\ &\quad = \int_{\boldsymbol{R}} \log\frac{d\nu}{d\mu}(x)d\nu(x) - \int_{\boldsymbol{R}} \log\frac{d\nu_0}{d\mu}(x)d\nu(x) \\ &\quad = D(\nu\|\mu) - \int_{\boldsymbol{R}} \log\frac{d\nu_0}{d\mu}(x)d\nu_0(x) \\ &\quad \le r - D(\nu_0\|\mu). \end{aligned} \tag{37}$$

It follows from (34), (36) and (37) that

$$\begin{aligned} R(\Delta;\nu) \ge\ & \frac{1}{2}\log\frac{\gamma^2}{\Delta} - r + D(\nu_0\|\mu) \\ \ge\ & \frac{1}{2}\log\frac{\gamma^2}{\Delta} - \frac{1}{2}\left(\frac{\beta^2}{\sigma^2} - 1 - \log\frac{\beta^2}{\sigma^2}\right) \\ & + \frac{1}{2}\left(\frac{\gamma^2}{\sigma^2} - 1 - \log\frac{\gamma^2}{\sigma^2}\right) \\ =\ & \frac{1}{2}\log\frac{\beta^2}{\Delta} + \frac{1}{2\sigma^2}(\gamma^2 - \beta^2) \\ \ge\ & R(\Delta;\nu^*). \end{aligned}$$

Thus we have obtained (35). Secondary we assume that $\Delta \ge \beta^2$. Then we can easily show that the right hand side of (14) is equal to zero. We can prove that (15) is equivalent to Eqs. (10) and (11). q.e.d.

4. Concluding Remarks

We have studied the error exponent for source coding with fidelity criterion and obtained explicit formulas for the minimum achievable rate $R_e(\Delta,r)$ and the reliability function $r_e(\Delta,R)$ for the SMGS. It has been shown that the expression of the error exponent, obtained by Marton, for the SMS with finite alphabet remains true for the SMGS.

Although we have treated only memoryless sources in this paper, we may expect that our results can be extended for information sources with memory. Let $X = \{X_n\}$ be a stationary ergodic sources and μ be the probability distribution of X. It is conjectured that, under some assumptions, we have the following formulas:

$$R_e(\Delta,r) = \sup_{\nu}\{\overline{R}(\Delta;\nu); \overline{D}(\nu\|\mu) \le r\},$$

$$r_e(\Delta,R) = \inf_{\nu}\{\overline{D}(\nu\|\mu); \overline{R}(\Delta;\nu) \ge R\},$$

$$R_e^*(\Delta,r) = \inf_{\nu}\{\overline{R}(\Delta;\nu); \overline{D}(\nu\|\mu) \le r\},$$

$$r_e^*(\Delta,R) = \inf_{\nu}\{\overline{D}(\nu\|\mu); \overline{R}(D;\nu) \le R\},$$

where we denote by $\overline{D}(\nu\|\mu)$ the divergence per unit time and by $\overline{R}(\Delta;\nu)$ the rate-distortion function per unit time.

References

[1] K. Marton, "Error exponent for source coding with a fidelity criterion," IEEE Trans. Inf. Theory, vol.IT-20, pp.197–199, 1974.
[2] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic Press, New York, 1981.
[3] T.S. Han, Information Spectrum Methods in Information Theory, Baihukan, Tokyo, 1998.
[4] S. Ihara, Information Theory for Continuous Systems, World Scientific, Singapore, 1993.
[5] J. Binia, M. Zakai, and J. Ziv, "On the ε-entropy and the rete-distortion function of certain non-Gaussian random variables," IEEE Trans. Inf. Theory, vol.IT-20, pp.517–524, 1974.
[6] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applicaions, Second Ed., Springer, New York, 1998.

Appendix: Proof of Lemma 2

Proof of Lemma 2: (i) Using a large deviation theorem (cf. [4], [6]) for the i.i.d. sequence $X = \{X_n\}$, we have (17). Let $a = \sqrt{\sigma^2+\Delta}$. We consider a Gaussian distribution

$$\nu^* \sim N\left(\frac{\sigma^2}{a}, \frac{\sigma^2\Delta}{a^2}\right).$$

Then $\nu^* \in \mathcal{P}_a$. In fact, by elementary caluculation, we can show that

$$\begin{aligned}
&\int_{\boldsymbol{R}} (x-a)^2 d\nu^*(x) \\
&\quad = \int_{\boldsymbol{R}} \left(x - \frac{\sigma^2}{a}\right)^2 d\nu^*(x) + \left(\frac{\sigma^2}{a} - a\right)^2 \\
&\quad = \frac{\sigma^2 \Delta}{a^2} + \left(\frac{\sigma^2}{a} - a\right)^2 \\
&\quad = \Delta.
\end{aligned}$$

Furthermore, by Lemma 1,

$$\begin{aligned}
D(\nu^* \| \mu) &= \frac{1}{2}\left(\frac{\Delta}{a^2} - 1 - \log\frac{\Delta}{a^2} + \frac{\sigma^2}{a^2}\right) \\
&= \frac{1}{2}\log\frac{a^2}{\Delta}.
\end{aligned}$$

Therefore to prove (18) it is sufficient to show that

$$D(\nu^* \| \mu) \le D(\nu \| \mu), \quad \forall \nu \in \mathcal{P}_a. \tag{A$\cdot$1}$$

We can easily show that

$$\begin{aligned}
\log\frac{d\nu^*}{d\mu}(x) &= -\frac{1}{2}\log\frac{\sigma^2\Delta}{a^2} + \frac{1}{2}\log\sigma^2 \\
&\quad - \frac{a^2}{2\sigma^2\Delta}\left(x - \frac{\sigma^2}{a}\right)^2 + \frac{x^2}{2\sigma^2}
\end{aligned} \tag{A$\cdot$2}$$

and that

$$\begin{aligned}
&-\frac{a^2}{2\sigma^2\Delta}\left(x - \frac{\sigma^2}{a}\right)^2 + \frac{x^2}{2\sigma^2} \\
&\quad = -\frac{1}{2\Delta}(x^2 - 2ax + \sigma^2) \\
&\quad = -\frac{1}{2\Delta}(x-a)^2 + \frac{1}{2}.
\end{aligned} \tag{A$\cdot$3}$$

Let $\nu \in \mathcal{P}_a$ and ν_0 be a Gaussian distribution with the same mean and variance as ν. By Lemma 1

$$D(\nu_0 \| \mu) \le D(\nu \| \mu). \tag{A$\cdot$4}$$

Moreover it is easily seen that

$$\begin{aligned}
D(\nu_0 \| \mu) &= \int_{\boldsymbol{R}} \log\frac{d\nu^*}{d\mu}(x) d\nu_0(x) + D(\nu_0 \| \nu^*) \\
&\ge \int_{\boldsymbol{R}} \log\frac{d\nu^*}{d\mu}(x) d\nu_0(x).
\end{aligned} \tag{A$\cdot$5}$$

By (A$\cdot$2) and (A$\cdot$3)

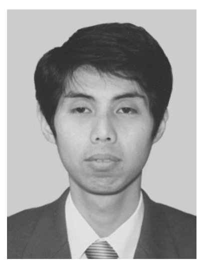
$$\begin{aligned}
&\int_{\boldsymbol{R}} \log\frac{d\nu^*}{d\mu}(x) d\nu_0(x) \\
&\quad = -\frac{1}{2}\log\frac{\sigma^2\Delta}{a^2} + \frac{1}{2}\log\sigma^2 \\
&\qquad - \frac{1}{2\Delta}\int_{\boldsymbol{R}} (x-a)^2 d\nu_0(x) + \frac{1}{2} \\
&\quad = -\frac{1}{2}\log\frac{\sigma^2\Delta}{a^2} + \frac{1}{2}\log\sigma^2 \\
&\quad = D(\nu^* \| \mu).
\end{aligned} \tag{A$\cdot$6}$$

The desired inequality (A$\cdot$1) follows from (A$\cdot$4), (A$\cdot$5) and (A$\cdot$6).

(ii) and (iii) can be proved in the same way.

q.e.d.

Shunsuke Ihara received the B.S., M.S. and Ph.D. degrees, all in mathematics, from Nagoya University, Ngoya, Japan, in 1965, 1967 and 1975, respectively. From 1967 to 1969, he was with Nagoya University and from 1969 to 1974, with Nagoya City University. He was an Associate Professor at Ehime University from 1974 to 1985, and a Professor of Mathematics at Kochi University, Kochi, Japan, from 1985 to 1989. Since 1989, he has been a Professor of Nagoya University. He is now with School of Informatics and Sciences, Nagoya University, Nagoya, Japan. He is the author of *Information Theory for Continuous Systems* (World Scientific, 1993). His research interests include basic problems in Shannon theory, large deviation theoretic approaches to source and channel coding theorems and applications of probability theory in information theory.



Masashi Kubo was born in Aichi Prefecture, Japan in 1968. He received the B.S. degree in Mathematics from Yamaguchi University, Yamaguchi, Japan in 1992, and the M.S. degree in mathematics, from Nagoya University, Ngoya, Japan in 1994. Since 1999 he has been a Lecturer of the Faculty of Education at Tokoha Gakuen University, Shizuoka, Japan. His current research interests are Shannon theory and computer sciences.