Sharper Lower Bounds for Discrimination Information in Terms of Variation

GODFRIED T. TOUSSAINT, MEMBER, IEEE

Abstract—The lower bound for discrimination information in terms of variation, derived recently by Kullback [7] for the distribution-free case, is sharpened. Furthermore, under a restriction, a lower bound is derived that is sharper than all other existing bounds.

Given two probability distributions $f_1(x)$ and $f_2(x)$, there are two well-known measures of the “distance” or difference between $f_1(x)$ and $f_2(x)$. One is the discrimination information given by

$$ I = \int f_1(x) \log \left( \frac{f_1(x)}{f_2(x)} \right) \, dx. $$

(1)

The other is the variation given by

$$ V = \int |f_1(x) - f_2(x)| \, dx. $$

(2)

In the past there has been a great deal of interest in bounding $I$ in terms of $V$. Volokonskii and Rozanov [1] showed that

$$ I \geq V - \log (1 + V). $$

(3)

 Pinsker [2] improved (3) by showing that

$$ I \geq \frac{V^2}{\Gamma}, $$

(4)

where $\Gamma$ is a constant greater than two. Csiszar [3] proved (4) with $\Gamma = 16$ while McKeen [4] established (4) with $\Gamma = 4e$.

Csiszar [5] and Kemperman [6], apparently independently, sharpened these results by proving that

$$ I \geq \frac{V^2}{2}. $$

(5)

Kullback [7], [8] sharpened (5) by showing that

$$ I \geq \frac{V^2}{2} + \frac{V^4}{36}. $$

(6)

The disadvantage of the bounds (3)–(6) is that for $V$ close to two they are loose, and for $V = 2$ the equality does not hold. In an attempt to improve these bounds, at least for $V$ close to two, Vajda [9] proved that

$$ I \geq \log \left( \frac{2 + V}{2 - V} \right) \frac{2V}{2 + V}. $$

(7)

The bound given by (7) is slightly looser than (6) for $V$ less than approximately 1.75, but much sharper than (6) for $V \geq 1.75$. Furthermore, it has the added nice property that the equality holds for both $V = 0$ and $V = 2$.

In this correspondence Kullback’s bound (6) is sharpened further. In fact, it will be shown that

$$ I \geq \frac{V^2}{2} + \frac{V^4}{36} + \frac{V^6}{288}. $$

(8)

Thus the maximum of (8) and (7) provides the sharpest lower bound available for $I$ in terms of $V$ for arbitrary distributions.

Let $\Omega_i$ denote the space where $f_i(x) > f_j(x)$, $i = 1, 2$, $i \neq j$. For the set of distribution pairs such that

$$ \int_{\Omega_1} f_2(x) \, dx = \int_{\Omega_2} f_1(x) \, dx, $$

(9)

which holds, for example, for the important case of Gaussian distributions with equal covariance matrices, it will be shown that

$$ I \geq \frac{V}{2} \log \left( \frac{2 + V}{2 - V} \right), $$

(10)

where the equality holds for both $V = 0$ and $V = 2$. Furthermore, it will be shown that (10) is sharper than both (7) and (8), for every $V \in [0,2]$.

Proof of (8): Let $L(u,t)$ be a function given by

$$ L(u,t) = (u + t) \log \left( \frac{1 + t}{u} \right) $$

$$ + (1 - u - t) \log \left( 1 - \frac{1 - t}{u} \right), $$

(11)

where $u$ and $t$ are real numbers. Also let

$$ \int_{\Omega_2} f_2(x) \, dx = \alpha_2 $$

(12)

and

$$ \int_{\Omega_1} f_1(x) \, dx = \alpha_1. $$

(13)

It follows from (12) and (13), see [11], that

$$ V = 2(1 - \alpha_1 - \alpha_2). $$

(14)

Krafft and Schmitz [10] showed that for

$$ 0 < u < 1 $$

(15)

and

$$ -u < t < 1 - u $$

(16)

it holds that

$$ L(u,t) \geq 2t^2 + \frac{4t^4}{9} + \frac{2t^6}{9}. $$

(17)

It is easy to verify that for $u = 1 - \alpha_2$ and $t = \alpha_1 + \alpha_2 - 1$, (15) and (16) are satisfied and

$$ L(1 - \alpha_2, \alpha_1 + \alpha_2 - 1) = \alpha_1 \log \left( \frac{\alpha_1}{1 - \alpha_2} \right) $$

$$ + (1 - \alpha_2) \log \left( \frac{1 - \alpha_1}{\alpha_2} \right). $$

(18)
Kullback [12] has shown that

\[ I \geq L(1 - \alpha_2, x_1 + \alpha_2 - 1). \] (19)

It also follows from (14) that

\[ t = -\frac{V}{2}. \] (20)

Substituting (20) into (17) and combining the latter with (19) yields (8), the desired result.

**Proof of (10):** Substitute \( x_1 = \alpha_2 = \frac{1}{2} - V/4 \) into (19).

Equation (10) can be used to obtain useful bounds in pattern recognition [13]. To prove that (10) is sharper than all the other bounds in this correspondence, for all \( V \in [0,2] \), it is sufficient to prove that (10) is sharper than (7) and (8).

Substitute \( u = \frac{1}{2} + V/4 \) and \( t = -V/2 \) into (17). Then the left side of (17) becomes the right side of (10), and the right side of (17) becomes the right side of (8), thus proving that (10) is sharper than (8).

In order to prove that (10) is sharper than (7) it must be shown that

\[ \frac{2x}{1 - x^2} = \log \left( \frac{1 + x}{1 - x} \right) \geq 0, \]

where \( x = V/2 \), which follows from the fact that, for \( 0 \leq x \leq 1 \), we have

\[ \frac{1}{2} \log \left( \frac{1 + x}{1 - x} \right) = \int_0^x (1 - y^2)^{-1} \, dy \]

\[ \leq (1 - x^2)^{-1} \int_0^x dy = \frac{x}{1 - x^2}. \]

Acknowledgment

The author is indebted to one of the referees for suggesting some of the preceding proofs, which are simpler than the original ones.

**REFERENCES**


**The Capacity Region of a Multiple-Access Discrete Memoryless Channel Can Increase with Feedback**

N. THOMAS GAARDER, MEMBER, IEEE, AND JACK K. WOLF, FELLOW, IEEE

**Abstract**—The capacity of a single-input single-output discrete memoryless channel is not increased even if the encoder could observe the output of the channel via a noiseless delayless feedback link. Recently, Liao [2], and then, Slepian and Wolf [3] gave formulas for the capacity region of a two-input single-output discrete memoryless channel with independent encoding of two source messages. After summarizing their results, we evaluate the performance of a transmission scheme for this channel, which makes use of noiseless feedback links from the output to the two encoders. We show that this scheme yields a vanishingly small error probability for a pair of rates that lie outside the capacity region.

**Introduction**

Shannon [1] proved that the capacity of a single-input single-output discrete memoryless channel is not increased even if the encoder could observe the output of the channel via a noiseless delayless feedback link. Recently, Liao [2], and then, Slepian and Wolf [3] gave formulas for the capacity region of a two-input single-output discrete memoryless channel with independent encoding of two source messages. After summarizing their results, we evaluate the performance of a transmission scheme for this channel, which makes use of noiseless feedback links from the output to the two encoders. We show that this scheme yields a vanishingly small error probability for a pair of rates that lie outside the capacity region.

**Capacity Regions without Feedback**

In this section we summarize the previously published results concerned with the capacity region of a multiple-access discrete memoryless channel without feedback. Consider the block diagram shown in Fig. 1. Two sources are described by a two-dimensional rate vector \( R = (R_1, R_2) \) with nonnegative components. Let \( N \) be a fixed positive integer. Every \( N \) time units, the sources\(^1\) produce a pair of statistically independent random variables \((U_1, U_2)\), where \( U_i \) is uniformly distributed over the set of integers \( \{1, 2, \cdots, M_i = [2^{N^i}] \} \) \( \mathbb{F} \). Here \( [x] \) is the smallest integer greater than or equal to \( x \).

The channel is described by a conditional probability distribution of the output random variable \( Y \) (which takes values \( y \in \mathbb{F} \)) given the inputs \( X_1 - x_1 \in \mathbb{F}_1 \) and \( X_2 - x_2 \in \mathbb{F}_2 \). We denote this conditional probability distribution \( P_{Y|X_1,X_2}(y|x_1,x_2) \). The channel is assumed memoryless in the usual sense. That is, the conditional probability distribution for \( N \)-vectors is equal to the product of the marginal conditional probability distributions. The encoders are a pair of deterministic mappings from the source outputs to channel input \( N \)-vectors. The mappings are such that if the sources produce the pair \((U_1 = i, U_2 = j)\), encoder 1 produces the \( N \)-vector \( x_1 \in (\mathbb{F}_1)^N \), which depends only on \( i \), and encoder 2 produces the \( N \)-vector \( x_2 \in (\mathbb{F}_2)^N \), which depends only on \( j \).

The decoder is a deterministic mapping from the channel output \( N \)-vector \( y \) to the pair \((i^*,j^*)\), where \( i^* \in \mathbb{F}, j^* \in \mathbb{F} \). We denote the decoder outputs by the pair of random variables \((U_1^*, U_2^*)\).