Mathematical
Inequalities
& Applications
Volume 2, Number 4 (1999), 599–611

FURTHER IMPROVEMENTS OF SOME BOUNDS ON ENTROPY MEASURES IN INFORMATION THEORY

M. MATIĆ, C. E. M. PEARCE AND J. PEČARIĆ

(*communicated by G. Peskir*)

Abstract. Recently Dragomir and Goh have produced some interesting new bounds on entropy measures in information theory. We strengthen further their results.

1. Introduction

The entropy function plays a key role in information theory. A key property is concavity, by virtue of which Jensen's inequality provides upper bounds for entropy measures. Recently Dragomir and Goh [1,2] have addressed the question of establishing lower bounds for the entropy measures of discrete–valued random variables and shown that these may also be provided by a suitable extension of Jensen's theorem. Dragomir and Goh derive several interesting bounds from their extension of Jensen's theorem and a corollary to it. Improvements of their results were given in [3] and [4]. In this paper, we shall give some further improvements of such results. First let us give an improvement of Dragomir-Goh key lemma:

LEMMA 1. (i) *Suppose* $\xi_k > 0,\ p_k > 0,\ (k = 1, \ldots, n)$ *with* $\sum_{k=1}^n p_k = 1$. *Then*

$$\begin{aligned} 0 &\leqslant \log\left(\sum_{k=1}^n p_k\xi_k\right) - \sum_{k=1}^n p_k \log \xi_k \\ &\leqslant \log\left(\sum_{k=1}^n p_k\xi_k \sum_{j=1}^n \frac{p_j}{\xi_j}\right) \\ &\leqslant \frac{1}{\ln b}\left(\sum_{k=1}^n p_k\xi_k \sum_{j=1}^n \frac{pj}{\xi_j} - 1\right) \end{aligned} \tag{1.1}$$

where all logarithms are of base $b > 1$.

Mathematics subject classification (1991): 94A17, 26D15.

Key words and phrases: Inequalities, bounds, entropy, information.

(*ii*) *If also* $\rho := \max_{i,k} \xi_i/\xi_k$, *then*

$$\begin{aligned} 0 &\leqslant \log\left(\sum_{k=1}^{n} p_k \xi_k\right) - \sum_{k=1}^{n} p_k \log \xi_k \\ &\leqslant \log\left(\frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2\right) \\ &\leqslant \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2. \end{aligned} \tag{1.2}$$

(*iii*) *If*

$$\rho \leqslant \Phi(\varepsilon) := 2b^{\varepsilon} - 1 + 2\sqrt{b^{\varepsilon}(b^{\varepsilon} - 1)} \tag{1.3}$$

for $\varepsilon > 0$, *then*

$$0 \leqslant \log\left(\sum_{k=1}^{n} p_k \xi_k\right) - \sum_{k=1}^{n} p_k \log \xi_k \leqslant \varepsilon. \tag{1.4}$$

Proof. (*i*) Jensen's inequality for concave function $f(x) = \log x$ states:

$$\log\left(\sum_{k=1}^{n} p_k x_k\right) \geqslant \sum_{k=1}^{n} p_k \log x_k \tag{1.5}$$

By using $x_k = \xi_k$ and $x_k = 1/\xi_k$, we have

$$\log\left(\sum_{k=1}^{n} p_k \xi_k\right) \geqslant \sum_{k=1}^{n} p_k \log \xi_k \tag{1.6}$$

and

$$\log\left(\sum_{k=1}^{n} \frac{p_k}{\xi_k}\right) \geqslant \sum_{k=1}^{n} p_k \log \frac{1}{\xi_k},$$

that is

$$-\sum_{k=1}^{n} p_k \log \xi_k \leqslant \log \sum_{k=1}^{n} \frac{p_k}{\xi_k}. \tag{1.7}$$

(1.6) gives the first inequality in (1.1), while (1.7) gives the second inequality in (1.1). Moreover, we can also obtain these inequalities from the well-known inequality between the arithmetic, the geometric and the harmonic means, that is

$$\left(\sum_{i=1}^{n} \frac{p_i}{\xi_i}\right)^{-1} \leqslant \prod_{i=1}^{n} \xi_i^{p_i} \leqslant \sum_{i=1}^{n} p_i \xi_i$$

wherefrom we have

$$-\log\left(\sum_{i=1}^{n} \frac{p_i}{\xi_i}\right) \leqslant \sum_{i=1}^{n} p_i \log \xi_i \leqslant \log \sum_{i=1}^{n} p_i \xi_i. \tag{1.8}$$

The first inequality in (1.8) is our inequality (1.7), while the second inequality in (1.8) is (1.6).The last inequality in (1.1) is a simple consequence of the following elementary inequality

$$\log x \leqslant \frac{1}{\ln b}(x-1) \tag{1.9}$$

(*ii*) Note that the following inequality was proved in [3]:

$$\sum_{k=1}^{n} p_k \xi_k \sum_{j=1}^{n} \frac{p_j}{\xi j} - 1 \leqslant \frac{1}{4}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2$$

wherefrom we have

$$\sum_{k=1}^{n} p_k \xi_k \sum_{j=1}^{n} \frac{pj}{\xi j} \leqslant \frac{1}{4}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2 + 1 = \frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2,$$

that is

$$\log\left(\sum_{k=1}^{n} p_k \xi_k \sum_{j=1}^{n} \frac{p_j}{\xi j}\right) \leqslant \log\left(\frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2\right), \tag{1.10}$$

and (1.2) is a simple consequence of (1.1), (1.10) and (1.9).

(*iii*) Set

$$\log\left(\frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2 j\right) \leqslant \varepsilon,$$

so that

$$\rho^2 - 2\rho(2b^\varepsilon - 1) + 1 \leqslant 0.$$

This holds if and only if

$$2b^\varepsilon - 1 - 2\sqrt{b^\varepsilon(b^\varepsilon - 1)} \leqslant \rho \leqslant 2b^\varepsilon - 1 + 2\sqrt{b^\varepsilon(b^\varepsilon - 1)}.$$

Since

$$2b^\varepsilon - 1 - 2\sqrt{b^\varepsilon(b^\varepsilon - 1)} = \left(2b^\varepsilon - 1 + 2\sqrt{b^\varepsilon(b^\varepsilon - 1)}\right)^{-1},$$

(1.4) now follows from (1.2), that is, (1.4) holds for all ρ satisfying (1.3). □

Remark 1. The third part of Lemma 1 gives an improvement of the key result in [1] and [3], where the conditions

$$\rho \leqslant \phi(\varepsilon) := 1 + \varepsilon \ln b + \sqrt{(2 + \varepsilon \ln b)\varepsilon \ln b} \text{ and } \rho \leqslant \phi(2\varepsilon)$$

were used (respectively) in place of (1.3). Also $\phi(\varepsilon)$ is obviously strictly increasing for $\varepsilon > 0$ so that we have

$$\begin{aligned} \phi(\varepsilon) &< \phi(2\varepsilon) = 1 + 2\ln b^\varepsilon + 2\sqrt{(1 + \ln b^\varepsilon)\ln b^\varepsilon} \\ &< 1 + 2\left(b^\varepsilon - 1\right) + 2\sqrt{\left(1 + b^\varepsilon - 1\right)\left(b^\varepsilon - 1\right)} \\ &= \Phi(\varepsilon), \end{aligned}$$

since $b^\varepsilon > 1$.

2. Bounds on the entropy of a random variable

Let X be a discrete–valued random variable with finite range $\{x_1, \dots, x_r\}$. Assume $p_i = P\{X = x_i\} > 0$ ($i = 1, \dots, r$). The b-entropy of X is defined by

$$H_b(X) := \sum_{i=1}^{r} p_i \log(1/p_i).$$

The following bounds on the entropy function give further improvement of Theorem 1 of [1].

THEOREM 1. *(i) With X as above*

$$\begin{aligned} 0 &\leqslant \log r - H_b(X) \\ &\leqslant \log\left(r\sum_{k=1}^{r} p_k^2\right) \\ &\leqslant \frac{1}{\ln b}\left(r\sum_{k=1}^{r} p_k^2 - 1\right). \end{aligned} \tag{2.1}$$

(ii) Define $\rho := \max_{i,k} p_i/p_k$. We have

$$\begin{aligned} 0 &\leqslant \log r - H_b(X) \\ &\leqslant \log\left(\frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2\right) \\ &\leqslant \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2. \end{aligned} \tag{2.2}$$

(iii) If $\rho \leqslant \Phi(\varepsilon)$ for $\varepsilon > 0$, then

$$0 \leqslant \log r - H_b(X) \leqslant \varepsilon. \tag{2.3}$$

Proof. Set $n = r$ and $\xi_k = 1/p_k$ in Lemma 1. □

Remark 2. Since $H_b(X) \geqslant 0$, the second inequality in (2.2) is nontrivial if

$$\frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2 < r$$

that is, if

$$\rho^2 - 2\rho(2r-1) + 1 < 0. \tag{2.4}$$

Inequality (2.4) holds if (since $\rho \geqslant 1$):

$$\rho < 2r - 1 + 2\sqrt{r(r-1)}. \tag{2.5}$$

Nontriviallity of the last bound in (2.2) was considered in [3].

THEOREM 2. *With X as above, define $M = \max_i p_i$ and $m = \min_i p_i$. Then*

$$0 \leqslant \log r - H_b(X) \leqslant \log\left(\left[\frac{r^2}{4}\right](M-m)^2 + 1\right) \leqslant \frac{(M-m)^2}{\ln b}\left[\frac{r^2}{4}\right]. \tag{2.6}$$

If

$$\max_{1 \leqslant i < j \leqslant r} |p_i - p_j| \leqslant \sqrt{\frac{b^\varepsilon - 1}{[r^2/4]}}, \tag{2.7}$$

then (2.3) is valid. (For $a \in \mathbf{R}$, $[a]$ denotes the greatest integer less than or equal to a.)

Proof. We have proved in [4]

$$r\sum_{i=1}^{r} p_i^2 - 1 \leqslant \left[\frac{r}{2}\right]\left(r - \left[\frac{r}{2}\right]\right)(M-m)^2 .$$

It is easy to check that $\left[\frac{r}{2}\right]\left(r - \left[\frac{r}{2}\right]\right) = \left[\frac{r^2}{4}\right]$ holds for all $r \in \mathbf{N}$, so that we have

$$\log\left(r\sum_{i=1}^{r} p_i^2\right] \leqslant \log\left\{\left[\frac{r^2}{4}\right)(M-m)^2 + 1\right\}$$

and (2.6) follows from (2.1) and (1.9). □

Remark 3. A similar result was proved in [4] but condition (2.7) was with $\varepsilon \ln b$ instead $b^\varepsilon - 1$ on the right-hand side. Note that by (1.9) we have $\varepsilon \ln b = \ln b^\varepsilon < b^\varepsilon - 1$, since $b^\varepsilon > 1$, so result is better than this from [4].

Remark 4. The second inequality in (2.6) is nontrivial (since $H_b(X) \geqslant 0$), if $\left[\frac{r^2}{4}\right](M-m)^2 + 1 < r$, that is if $M - m < \sqrt{(r-1)/[r^2/4]}$. If we compare it with (2.7), it is clear that we should have $b^\varepsilon < r$ for nontriviality of (2.3). For nontriviality of the last bound in (2.6) we should have $M - m < \sqrt{(\ln r)/[r^2/4]}$.

3. Bounds on conditional entropy

Let X, Y be a pair of random variables with respective ranges $\{x_1, x_2, \ldots, x_r\}$ and $\{y_1, y_2, \ldots, y_s\}$. The conditional entropy of X given Y is defined by

$$H_b(X \mid Y) := \sum_{i,j} p(x_i, y_j)\log(1/p(x_i|y_j)),$$

where

$$p(x_i, y_j) := P\{X = x_i, Y = y_j\}$$

and

$$p(x_i|y_j) := P\{X = x_i \mid Y = y_j\} = p(x_i, y_j)/p(y_j).$$

(See, for example, [5, p. 22].) Without loss of generality we need to define these quantities only for those (i,j) for which $p(x_i,y_j) > 0$. There will be $n(\leqslant rs)$ such pairs. The conditional entropy can be interpreted as the amount of uncertainty remaining about X after Y has been observed.

THEOREM 3. *Let X and Y be as above. For $1 \leqslant j \leqslant s$, define $V_j := \{i : p(x_i,y_j) > 0\}$ and $U := \{(i,j) : i \in V_j\}$ and let $r' = \sum_{j=1}^{s} p(y_j)\,|V_j|$. Then we have*

$$\begin{aligned} 0 &\leqslant \log r' - H_b(X|Y) \\ &\leqslant \log\left(r' \sum_{(i,j)\in U} p(y_j)p^2\,(x_i|y_j)\right) \\ &\leqslant \frac{1}{\ln b}\left(r' \sum_{(i,j)\in U} p(y_j)p^2\,(x_i|y_j) - 1\right). \end{aligned} \tag{3.1}$$

If $\rho := \max_{(i,j),(u,v)\in U} p(x_i|y_j)/p(x_u|y_v)$, then

$$\begin{aligned} 0 &\leqslant \log r' - H_b(X \mid Y) \\ &\leqslant \log\left(\frac{1}{4}\left(\sqrt{\rho}+\frac{1}{\sqrt{\rho}}\right)^2\right) \\ &\leqslant \frac{1}{4\ln b}\left(\sqrt{\rho}-\frac{1}{\sqrt{\rho}}\right)^2. \end{aligned} \tag{3.2}$$

If

$$\rho \leqslant \Phi(\varepsilon), \quad \varepsilon > 0, \tag{3.3}$$

then

$$0 \leqslant \log r' - H_b(X \mid Y) \leqslant \varepsilon. \tag{3.4}$$

Proof. We may label those pairs (i,j) for which $p(x_i,y_j) > 0$, that is the pairs $(i,j) \in U$, as $k = 1,2,\cdots,n$. We then put $p_k = p(x_i,y_j)$ and $\xi_k = 1/p(x_i|y_j) = p(y_j)/p\,(x_i,y_j)$ in Lemma 1. This gives:

$$\begin{aligned} 0 &\leqslant \log\left(\sum_{(i,j)\in U} p(y_j)\right) - \sum_{(i,j)\in U} p\,(x_i|y_j)\log\frac{1}{p\,(x_i|y_j)} \\ &\leqslant \log\left(\sum_{(i,j)\in U} p(y_j) \sum_{(u,v)\in U} p(y_v)p^2\,(x_u|y_v)\right). \end{aligned}$$

This is equivalent to the first two inequalities in (3.1) since

$$\sum_{(i,j)\in U} p(y_j) = \sum_j p(y_j) \sum_{i\in V_j} 1 = r'.$$

The rest of theorem is a similar consequence of Lemma 1. □

Let us note that in [4] the following result was proved:

$$\begin{aligned}&\sum_{(i,j)\in U} p(y_j) \sum_{(u,v)\in U} p(y_v)p^2\left(x_u|y_v\right)\\ \leqslant\ & \frac{1}{4}\left(\max_{(i,j)\in U} p\left(x_i|y_j\right) - \min_{(i,j)\in U} p\left(x_i|y_j\right)\right)^2 r'^2 + 1.\end{aligned}$$

Therefore, from Theorem 3, using (1.9).we get the following result:

THEOREM 4. *Let X, Y and U be as in Theorem 3. Define $M = \max_{(i,j)\in U} p\left(x_i|y_j\right)$ and $m = \min_{(i,j)\in U} p\left(x_i|y_j\right)$. Then*

$$0 \leqslant \log r' - H_b(X|Y) \leqslant \log\left(\frac{1}{4}(M-m)^2 r'^2 + 1\right) \leqslant \frac{(M-m)^2}{4\ln b} r'^2. \tag{3.5}$$

If

$$\max_{(i,j),(u,v)\in U} \left|p\left(x_i|y_j\right) - p\left(x_u|y_v\right)\right| \leqslant \frac{2}{r'}\sqrt{b^{\varepsilon}-1}, \tag{3.6}$$

then (3.4) is valid.

Remark 5. Note that the second inequality in (3.2) is not trivial if we have $\rho < 2r' - 1 + 2\sqrt{r'(r'-1)}$. Also, the second inequality in (3.5) is not trivial if $M - m < (2/r')\sqrt{r'-1}$.

We now introduce a third discrete–valued random variable Z, assuming values $z_1, \cdots, z_t$, each with positive probability. As in [2, Theorem 1.2], we define an associated random variable A which takes on the value $\sum_{i,j} p(x_i, y_j, z_k)/p(x_i|y_j)$ with probability $p(z_k)$ $(k = 1, \cdots, t)$. The following theorem gives improvements of Theorem 3.2 from [3].

THEOREM 5. *With ρ defined as in Theorem 3, we have*

$$\begin{aligned}0 &\leqslant H_b(Z) + E(\log A) - H_b(X \mid Y)\\ &\leqslant \log\left(\frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2\right)\\ &\leqslant \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2.\end{aligned}$$

If condition (3.3) *holds, then*

$$0 \leqslant H_b(Z) + E(\log A) - H_b(X \mid Y) \leqslant \varepsilon.$$

Proof. For fixed z_ℓ, put $p_k = p(x_i, y_j, z_\ell)/p(z_\ell)$ and $\xi_k = 1/p(x_i|y_j)$, where much as in Theorem 3 we relabel $k = (i,j)$ for those (i,j) for which $p(x_i, y_j, z_\ell) > 0$. We derive from Lemma 1 that

$$\begin{aligned} 0 &\leqslant \log\left(\sum_k \frac{p(x_i, y_j, z_\ell)}{p(z_\ell)} \frac{1}{p(x_i|y_j)}\right) - \sum_k \frac{p(x_i, y_j, z_\ell)}{p(z_\ell)} \log \frac{1}{p(x_i|y_j)} \\ &\leqslant \log\left(\frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2\right) \\ &\leqslant \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2. \end{aligned}$$

Multiplication by $p(z_\ell)$ and summation over $\ell = 1, \cdots, t$ yields

$$\begin{aligned} 0 &\leqslant H_b(Z) + \sum_{\ell=1}^{t} p(z_\ell) \log\left(\sum_k \frac{p(x_i, y_j, z_\ell)}{p(x_i|y_j)}\right) - H_b(X|Y) \\ &\leqslant \log\left(\frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2\right) \\ &\leqslant \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2. \end{aligned}$$

The desired results follow. □

We may use the preceding result for further improvements to Fano's inequality, which states the following:

If X, Y have a common range and $P_e = P(X \neq Y)$, then

$$H_b(X|Y) \leqslant H_b(P_e) + P_e \log(r-1).$$

We note that it is also tacit in Fano's inequality that $p(x_i, y_j) > 0 \ \forall i, j$.

The following result extends related results from [1] and [3]:

COROLLARY 1. *Suppose X, Y have the same range. Define Z by $Z = 0$ if $X = Y$ and $Z = 1$ if $X \neq Y$. Further, define*

$$T_j := |\{i : i \neq j, p(x_i, y_j) > 0\}|,$$

$$R_j := |V_j| - T_j = \begin{cases} 1 & \text{if } p(x_j, y_j) > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Then

$$\begin{aligned} 0 &\leqslant H_b(P_e) + P_e \log\left(\sum_j p(y_j) T_j\right) + (1 - P_e) \log\left(\sum_j p(y_j) R_j\right) - H_b(X \mid Y) \\ &\leqslant \log\left(\frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2\right) \\ &\leqslant \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2. \end{aligned}$$

If (3.3) *holds, then*

$$0 \leqslant H_b(P_e) + P_e \log\left(\sum_j p(y_j)T_j\right) + (1-P_e)\log\left(\sum_j p(y_j)R_j\right) - H_b(X \mid Y) \leqslant \varepsilon.$$

We may interpret this result in terms of the transmission of discrete characters. If X is sent and Y received, then P_e is the probability of erroneous reception and $H_b(Z) = H_b(P_e) = -P_e \log P_e - (1-P_e)\log(1-P_e)$.

Similarly we can obtain the following improvement of the related results from [4]:

THEOREM 6. *Let* X, Y, Z *and* A *be as in Theorem 5. Along with the notations from Theorem 5 we define also*

$$K := \sum_{\ell=1}^{t} \frac{1}{p(z_\ell)} A^2(z_\ell),$$

where $A(z_\ell) = \sum_{(i,j)\in U} p(x_i, y_j, z_\ell)/p(x_i|y_j)$ *for* $\ell = 1, \cdots, t$. *Finally we put*

$$M = \max_{(i,j)\in U} p(x_i|y_j) \quad \text{and} \quad m = \min_{(i,j)\in U} p(x_i|y_j).$$

Then we have

$$\begin{aligned} 0 &\leqslant H_b(Z) + E(\log A) - H_b(X|Y) \\ &\leqslant \log\left(\frac{1}{4}(M-m)^2 \min\left\{K, \frac{1}{Mm}\right\} + 1\right) \\ &\leqslant \frac{(M-m)^2}{4\ln b} \min\left\{K, \frac{1}{Mm}\right\}. \end{aligned}$$

If $\varepsilon > 0$ *is given and*

$$\max_{(i,j),(u,v)\in U} |p(x_i|y_j) - p(x_u|y_v)| \leqslant 2\sqrt{\frac{b^\varepsilon - 1}{K}},$$

then

$$0 \leqslant H_b(Z) + E(\log A) - H_b(X|Y) \leqslant \varepsilon.$$

COROLLARY 2. *Suppose* X, Y *and* Z *are as in Corollary 1. If* T_j, R_j $(j = 1, \cdots, s)$ *and* P_e *are defined as in Corollary 1, then*

$$A(0) = \sum_{j=1}^{s} p(y_j)R_j,\ A(1) = \sum_{j=1}^{s} p(y_j)T_j.$$

For given $\varepsilon > 0$, *if*

$$\max_{(i,j),(u,v)\in U} |p(x_i|y_j) - p(x_u|y_v)| \leqslant 2\sqrt{\frac{(1-P_e)P_e(b^\varepsilon - 1)}{A^2(0)P_e + A^2(1)(1-P_e)}},$$

then

$$0 \leqslant H_b(P_e) + P_e \log A(1) + (1-P_e)\log A(0) - H_b(X|Y) \leqslant \varepsilon.$$

4. Bounds on mutual information

The b–mutual information between random variables X, Y is defined by

$$I_b(X;Y) := H_b(X) - H_b(X \mid Y) = \sum_{i,j} p(x_i,y_j) \log \frac{p(x_i,y_j)}{p(x_i)p(y_j)}.$$

The following results improve the bounds on mutual information given in [1-4].

THEOREM 7. *Let* $V := \{(i,j) : p(x_i,y_j) > 0\}$ *and*

$$K := \sum_{(i,j)\in V} \frac{p^2(x_i,y_j)}{p(x_i)p(y_j)}, \quad S := \sum_{(i,j)\in V} p(x_i)p(y_j).$$

Then

$$0 \leqslant \log S + I_b(X;Y) \leqslant \log(SK) \leqslant \frac{1}{\ln b}(SK-1). \tag{4.1}$$

Suppose

$$\rho := \max_{(i,j),(u,v)\in V} \frac{p(x_i)p(y_j)p(x_u,y_v)}{p(x_u)p(y_v)p(x_i,y_j)}.$$

Then

$$\begin{aligned} 0 &\leqslant \log S + I_b(X;Y) \\ &\leqslant \log\left(\frac{1}{4}\left(\sqrt{\rho}+\frac{1}{\sqrt{\rho}}\right)^2\right) \\ &\leqslant \frac{1}{4\ln b}\left(\sqrt{\rho}-\frac{1}{\sqrt{\rho}}\right)^2. \end{aligned} \tag{4.2}$$

If $\rho \leqslant \Phi(\varepsilon)$ *for* $\varepsilon > 0$, *then*

$$0 \leqslant \log S + I_b(X;Y) \leqslant \varepsilon. \tag{4.3}$$

Proof. This follows the lines of our earlier proofs, setting $p_k = p(x_i,y_j)$ and $\xi_k = p(x_i)p(y_j)/p(x_i,y_j)$ in Lemma 1 after suitable relabelling. □

THEOREM 8. *If*

$$\max_{(i,j),(u,v)\in V} \left| \frac{p(x_i)p(y_j)}{p(x_i,y_j)} - \frac{p(x_u)p(y_v)}{p(x_u,y_v)} \right| \leqslant \frac{2}{K}\sqrt{b^{\varepsilon}-1}$$

then (4.3) *is valid.*

THEOREM 9. *Suppose* $W := \{(i,j,k) : p(x_i,y_j,z_k) > 0\}$ *and define*

$$L := \sum_{(i,j,k)\in W} \frac{p^2(x_i,y_j,z_k)}{p(x_i,y_j)p(z_k|y_j)}, \quad T := \sum_{(i,j,k)\in W} p(x_i,y_j)p(z_k|y_j).$$

Then

$$0 \leqslant \log T + I_b(X,Y;Z) - I_b(Y;Z) \leqslant \log(TL) \leqslant \frac{1}{\ln b}(TL-1).$$

If

$$\rho := \max_{(i,j,k),(u,v,w)\in W} \frac{p(z_k|x_i,y_j)p(z_w|y_v)}{p(z_w|x_u,y_v)p(z_k|y_j)},$$

then

$$\begin{aligned} 0 &\leqslant \log T + I_b(X,Y;Z) - I_b(Y;Z) \\ &\leqslant \log\left(\frac{1}{4}\left(\sqrt{\rho}+\frac{1}{\sqrt{\rho}}\right)^2\right) \\ &\leqslant \frac{1}{4\ln b}\left(\sqrt{\rho}-\frac{1}{\sqrt{\rho}}\right)^2. \end{aligned}$$

If $\rho \leqslant \Phi(\varepsilon)$ *for* $\varepsilon > 0$*, then*

$$0 \leqslant \log T + I_b(X,Y;Z) - I_b(Y;Z) \leqslant \varepsilon.$$

THEOREM 10. *Suppose the conditions of Theorem 9 are satisfied and*

$$M := \max_{(i,j,k)\in W} \frac{p(z_k|x_i,y_j)}{p(z_k|y_j)}, \quad m := \min_{(i,j,k)\in W} \frac{p(z_k|x_i,y_j)}{p(z_k|y_j)}.$$

Then

$$\begin{aligned} 0 &\leqslant \log T + I_b(X,Y;Z) - I_b(Y;Z) \\ &\leqslant \log\left(\frac{1}{4}(M-m)^2 \min\left\{T^2, \frac{1}{Mm}\right\}+1\right) \\ &\leqslant \frac{(M-m)^2}{4\ln b} \min\left\{T^2, \frac{1}{Mm}\right\}. \end{aligned}$$

5. Remarks and further improvements

As we can see, the main improvement in our result are based on new term

$$\log\left(\sum_{k=1}^{n} p_k\xi_k \sum_{j=1}^{n} \frac{p_j}{\xi_j}\right)$$

in our Lemma 1. As in [3, Remarks and further improvements] we can give improvements of some previous results in the case when (ξ_k) is monotonic sequence by using an result from [8]. Namely, as in [3] we can prove the following result:

LEMMA 2. *Suppose* $\xi_k \in (0,\infty)$ *and* $p_k \geqslant 0$ *with* $\sum_{k=1}^{n} p_k = 1$*. Let* σ *be a permutation of* $(1,\ldots,n)$ *such that* $(\xi_{\sigma(k)})_1^n$ *is monotone and set* $P_k := \sum_{i=1}^{k} p_{\sigma(i)}$ *and*

$\rho := \max_{i,k} \xi_i/\xi_k$. *Define* $M := \max_{1 \leqslant k < n} P_k(1-P_k)$. *Then for* $b > 1$ *we have*

$$\begin{aligned} 0 &\leqslant \log\left(\sum_{k=1}^{n} p_k \xi_k\right) - \sum_{k=1}^{n} p_k \log \xi_k \\ &\leqslant \log\left(M\left(\sqrt{\rho} - 1/\sqrt{\rho}\right)^2 + 1\right) \\ &\leqslant \frac{M}{\ln b}\left(\sqrt{\rho} - 1/\sqrt{\rho}\right)^2. \end{aligned} \tag{5.1}$$

If

$$\rho \leqslant \Phi_M(\varepsilon) := 1 + \frac{b^{\varepsilon} - 1}{2M} + \frac{1}{2M}\sqrt{(4M + b^{\varepsilon} - 1)(b^{\varepsilon} - 1)} \tag{5.2}$$

for some $\varepsilon > 0$, *then*

$$0 \leqslant \log\left(\sum_{k=1}^{n} p_k \xi_k\right) - \sum_{k=1}^{n} p_k \log \xi_k \leqslant \varepsilon. \tag{5.3}$$

THEOREM 11. *Let* X *be a discrete–valued random variable with finite range* $\{x_1, \ldots, x_r\}$ *and probability distribution* $p_k = P\{X = x_k\}$ $(1 \leqslant k \leqslant r)$, *and set* $\rho := \max_{i,k} p_i/p_k$. *Let* σ *be a permutation of* $(1, \ldots, n)$ *such that* $(p_{\sigma(k)})_1^n$ *is monotone. Define* $P_k := \sum_{i=1}^{k} p_{\sigma(i)}$ *and* $M := \max_{1 \leqslant k < n} P_k(1-P_k)$. *Then*

$$0 \leqslant \log r - H_b(X) \leqslant \log\left(M\left(\sqrt{\rho} - 1/\sqrt{\rho}\right)^2 + 1\right) \leqslant \frac{M}{\ln b}\left(\sqrt{\rho} - 1/\sqrt{\rho}\right)^2.$$

If $\rho \leqslant \Phi_M(\varepsilon)$ *for some* $\varepsilon > 0$ *then*

$$0 \leqslant \log r - H_b(X) \leqslant \varepsilon.$$

Proof. Set $n = r$ and $\xi_k = 1/p_k$ in Lemma 2. □

Remark 6. Again we have the situation in which Theorem 11 gives a nontrivial upper bound for the difference $\log r - H_b(X)$ if and only if

$$M\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2 + 1 < r,$$

which is equivalent to

$$\rho^2 - 2\left(1 + \frac{r-1}{2M}\right)\rho + 1 < 0$$

and (since $\rho \geqslant 1$) to

$$\rho < 1 + \frac{r-1}{2M} + \frac{1}{2M}\sqrt{(4M + r - 1)(r-1)}.$$

Analogous improvements can be given for the results in Sections 3 and 4.

REFERENCES

[1] S. S. Dragomir and C. J. Goh Some bounds on entropy measures in information theory Appl. Math. Lett. **10 (3)** (1997) 23–28.
[2] S. S. Dragomir and C. J. Goh A counterpart of Jensen's discrete inequality for differentiable convex mappings and applications in information theory Math. Comput. Modelling **24 (2)** (1996) 1–11.
[3] M. Matić, C. E. M. Pearce and J. Pečarić Improvements of some bounds on entropy measures in information theory Math. Inequal. and Appl. 1 (1998) 295–304.
[4] M. Matić, C. E. M. Pearce and J. Pečarić Refinements of some bounds in information theory J. Austr. Math. Soc. Ser. B, to appear.
[5] R. J. McEliece The theory of information and coding Addison-Wesley, Reading, Mass. 1977.
[6] J. Pečarić On an inequality of T. Popoviciu Bul. Si. Tehn. Inst. Politechn. Timişoara 2, **24 (38)** (1979) 9–15.

(Received January 13, 1999)

M. Matić
FESB, Mathematics Department
University of Split
R. Boškovića bb
21000 Split
Croatia
e-mail: mmatic@fesb.hr

C. E. M. Pearce
Applied Mathematics Department
University of Adelaide
Adelaide SA 5005, Australia
e-mail: cpearce@maths.adelaide.edu.au

J. Pečarić
Faculty of Textile Technology
University of Zagreb
Pierottijeva 6
Zagreb
Croatia
e-mail: jpecaric@maths.adelaide.edu.au