Mathematical
Inequalities
& Applications
Volume 1, Number 2 (1998), 295–304

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 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] 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.

In this note we show that the key results of [1] can be strengthened by use of a result relating to Popoviciu's inequality and derived by one of us in [4]. We deduce consequent tighter bounds than those found by Dragomir and Goh in [1] for the entropy, conditional entropy and mutual information for discrete–valued random variables. We note also that the proofs of Theorem 3 and Corollary 2 of [1] involve a tacit assumption as to the joint values that may be taken with positive probability by the random variables involved. In fact if this assumption is not satisfied, the enunciations need modification.

We begin with the following lemma.

LEMMA 1.1. *Suppose* $0 < \xi_1 \leqslant \xi_2 \leqslant \cdots \leqslant \xi_n < \infty$ *or* $0 < \xi_n \leqslant \xi_{n-1} \leqslant \cdots \leqslant \xi_1 < \infty$. *Suppose also* $p_k \geqslant 0$ *with* $\sum_{k=1}^{n} p_k = 1$. *Take* $b > 1$. *Then*

$$0 \leqslant \log_b\left(\sum_{k=1}^{n} p_k \xi_k\right) - \sum_{k=1}^{n} p_k \log_b \xi_k \leqslant \frac{1}{\ln b} \frac{(\xi_n - \xi_1)^2}{\xi_n \xi_1} \max_{1 \leqslant k < n} P_k(1 - P_k), \qquad (1.1)$$

where $P_k = \sum_{i=1}^{k} p_i$.

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

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

Proof. Since the mapping $f : (0,\infty) \to \mathbf{R}$ given by $f(x) = \log_b x$ is differentiable and concave on $(0,\infty)$, we have

$$f(x) - f(y) \leqslant f'(y)(x-y)$$

for all $x, y > 0$, that is,

$$\log_b x - \log_b y \leqslant \frac{1}{\ln b} \cdot \frac{x-y}{y} \qquad \forall x, y > 0.$$

Put $x = \sum_{j=1}^n p_j\xi_j$ and $y = \xi_i$ $(i = 1, 2, \ldots, n)$. Then

$$\log_b\left(\sum_{j=1}^n p_j\xi_j\right) - \log_b \xi_i \leqslant \frac{1}{\ln b} \cdot \frac{1}{\xi_i}\left(\sum_{j=1}^n p_j\xi_j - \xi_i\right) \qquad (i = 1, 2, \ldots, n).$$

Multiplying by p_i and summing over i gives

$$0 \leqslant \log_b\left(\sum_{j=1}^n p_j\xi_j\right) - \sum_{i=1}^n p_i \log_b \xi_i \leqslant \frac{1}{\ln b}\left[\left(\sum_{i=1}^n \frac{p_i}{\xi_i}\right)\left(\sum_{j=1}^n p_j\xi_j\right) - 1\right]. \tag{1.2}$$

It has been shown in [4] (see also [3, p. 302]) that if (a_i) is a positive nondecreasing sequence and (p_i) a nonnegative sequence with $P_n = 1$, then

$$\sum_{i=1}^n p_i a_i \sum_{j=1}^n \frac{p_j}{a_j} - 1 \leqslant \frac{(a_n - a_1)^2}{a_n a_1} \max_{1 \leqslant k < n} P_k(1 - P_k). \tag{1.3}$$

It is easy to see that (1.3) remains true if (a_i) is a positive nonincreasing sequence (simply take $(1/a_i)$ instead of (a_i)). Coupling this with (1.2) gives the second inequality in (1.1). The first is a simple consequence of Jensen's inequality for concave functions. □

The second inequality in Lemma 1.1 may be put to use to give the following useful corollary which does not require the monotonicity of the sequence $(\xi_i)_1^n$ and in which the upper bound does not depend on the numerical values $(p_i)_1^n$.

LEMMA 1.2. *Let* $\xi_k \in (0,\infty)$ $(1 \leqslant k \leqslant n)$ *and* $\rho := \max_{i,k} \xi_i/\xi_k$. *Suppose* $p_k \geqslant 0$ *with* $\sum_{k=1}^n p_k = 1$ *and* $b > 1$. *We have*

$$0 \leqslant \log_b\left(\sum_{k=1}^n p_k\xi_k\right) - \sum_{k=1}^n p_k \log_b \xi_k \leqslant \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2. \tag{1.4}$$

If

$$\rho \leqslant \Phi(\varepsilon) := 1 + 2\varepsilon \ln b + 2\sqrt{(1 + \varepsilon \ln b)\varepsilon \ln b} \tag{1.5}$$

for $\varepsilon > 0$, *then*

$$0 \leqslant \log_b\left(\sum_{k=1}^n p_k\xi_k\right) - \sum_{k=1}^n p_k \log_b \xi_k \leqslant \varepsilon. \tag{1.6}$$

Proof. As the two sums in (1.4) are independent of subscript permutations, we can suppose without loss in generality that $\xi_1 \leqslant \cdots \leqslant \xi_n$ or $\xi_n \leqslant \cdots \leqslant \xi_1$, so $\rho = \xi_n/\xi_1$ or ξ_1/ξ_n. Because

$$\max_{1 \leqslant k < n} P_k(1 - P_k) \leqslant 1/4,$$

we have by (1.1) that (1.4) holds.

Set

$$\frac{1}{4 \ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2 \leqslant \varepsilon,$$

so that

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

This holds if and only if

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

Since

$$1 + 2\varepsilon \ln b - 2\sqrt{(1 + \varepsilon \ln b)\varepsilon \ln b} = [1 + 2\varepsilon \ln b + 2\sqrt{(1 + \varepsilon \ln b)\varepsilon \ln b}]^{-1},$$

(1.6) now follows from (1.4), that is, (1.6) holds for all ρ satisfying (1.5). □

REMARK. The second part of Lemma 1.2 gives an improvement of the key result in [1], where the condition

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

is used in place of (1.5). It is easy to show that $\Phi(\varepsilon) = \phi(2\varepsilon)$. Also $\phi(\varepsilon)$ is obviously strictly increasing for $\varepsilon > 0$ so that we have $\phi(\varepsilon) < \phi(2\varepsilon) = \Phi(\varepsilon)$. The first part gives a more general result not involving any constraints on ρ.

2. Bounds on the entropy of a random variable

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

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

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

THEOREM 2.1. *With X as above, define $\rho := \max_{i,k} p_i/p_k$. We have*

$$0 \leqslant \log_b r - H_b(X) \leqslant \frac{1}{4 \ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2.$$

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

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

Proof. Set $n = r$ and $\xi_k = 1/p_k$ in Lemma 1.2. □

3. Bounds on conditional entropy

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

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

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, [2, p. 22].) Without loss of generality we need 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.

The following theorem gives improvements of Theorem 3 of [1].

THEOREM 3.1. *For* $1 \leqslant j \leqslant s$, *define* $V_j := \{i : p(x_i, y_j) > 0\}$ *and* $U := \{(i,j) : i \in V_j\}$. *If* $\rho := \max_{(i,j),(u,v) \in U} p(x_i|y_j)/p(x_u|y_v)$, *then*

$$0 \leqslant \log_b \left[\sum_j p(y_j)|V_j| \right] - H_b(X \mid Y) \leqslant \frac{1}{4 \ln b} \left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}} \right)^2. \tag{3.1}$$

If

$$\rho \leqslant \Phi(\varepsilon), \qquad \varepsilon > 0, \tag{3.2}$$

then

$$0 \leqslant \log_b \left[\sum_j p(y_j)|V_j| \right] - H_b(X \mid Y) \leqslant \varepsilon. \tag{3.3}$$

Proof. We may label those pairs (i,j) for which $p(x_i, y_j) > 0$ as $k = 1, 2, \cdots, n$. We then put $p_k = p(x_i, y_j)$ and $\xi_k = 1/p(x_i|y_j)$ in Lemma 1.2. This gives

$$0 \leqslant \log_b \left[\sum_k \frac{p(x_i, y_j)}{p(x_i|y_j)} \right] - H_b(X \mid Y) \leqslant \frac{1}{4 \ln b} \left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}} \right)^2.$$

The desired results follow, since

$$\frac{p(x_i, y_j)}{p(x_i|y_j)} = p(y_j). \quad \square$$

In [1, Theorem 3] it is implicit that $p(x_i, y_j) > 0$ for all $i = 1, \cdots, r$ and $j = 1, \cdots, s$. However it is clear from our argument that if this condition fails, the first term on the right in the first inequality in (3.1) and (3.3) is strictly less than $\log_b r$. This is important for applications, as sometimes the value taken by one random variable will restrict the set of possible values that the other can assume simultaneously with positive probability. Suppose, in particular, that the value of Y uniquely determines that of X. Then $|V_j| = 1 \ \forall j$ and so $\log_b \sum_k p(y_j)|V_j| = 0$. Since Y determines X uniquely, we have also $H_b(X|Y) = 0$. Further, $\rho = 1$. Thus in this case, (3.1) states that $0 \leqslant 0 \leqslant 0$.

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 [1, Corollary 1].

THEOREM 3.2. *With ρ defined as in Theorem 3.1, we have*

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

If condition (3.2) holds, then

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

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 3.1 we relabel $k = (i,j)$ for those (i,j) for which $p(x_i, y_j, z_\ell) > 0$. We derive from Lemma 1.2 that

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

Multiplication by $p(z_\ell)$ and summation over ℓ yields

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

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

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

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 [1, Corollary 2].

COROLLARY 3.3. *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

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

If (3.2) holds, then

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

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 $H_b(Z) = H_b(P_e) = -P_e \log_b P_e - (1 - P_e) \log_b (1 - P_e)$.

4. Bounds on mutual information

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

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

The two following results improve the bounds on mutual information given in [1, Corollary 3].

THEOREM 4.1. *Suppose*

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

Then

$$0 \leqslant I_b(X, Y) \leqslant \frac{1}{4 \ln b} \left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}} \right)^2.$$

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

$$0 \leqslant I_b(X, Y) \leqslant \varepsilon.$$

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 1.2 after suitable relabelling. □

We conclude by stating a result involving bounds on mutual information involving three random variables. These are improvements of Corollary 4 from [1].

THEOREM 4.2. *Suppose* $W := \{(i,j,k) : p(x_i,y_j,z_k) > 0\}$ *and define*

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

Then

$$0 \leqslant I_b(X,Y,Z) - I_b(Y,Z) \leqslant \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2.$$

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

$$0 \leqslant I_b(X,Y,Z) - I_b(Y,Z) \leqslant \varepsilon.$$

5. Remarks and further improvements

Our central result Theorem 2.1 (which improves the result from [1]) is based upon the inequality (1.3) using the fact that $\max_{1\leqslant k<n} P_k(1-P_k) \leqslant 1/4$. We observe that if the sequence $(p_i)_1^r$ is monotone, we can set $\xi_i = 1/p_i$ in (1.2) to derive

$$0 \leqslant \log_b r - H_b(X) \leqslant \frac{1}{\ln b}\left[r\sum_{k=1}^{r} p_k^2 - 1\right].$$

On the other hand, if $0 < p_1 \leqslant p_2 \leqslant \cdots \leqslant p_r$ and $\sum_{k=1}^r p_k = 1$, then we have the simpler bound

$$r\sum_{k=1}^{r} p_k^2 \leqslant rp_r \sum_{k=1}^{r} p_k = rp_r,$$

and since $p_1 \leqslant 1/r$ we have $r\sum_{k=1}^r p_k^2 \leqslant p_r/p_1$. So we can obtain the simple result

$$0 \leqslant \log_b r - H_b(X) \leqslant \frac{1}{\ln b}\left(\frac{p_r}{p_1} - 1\right) =: \beta_1. \tag{5.1}$$

The upper bound

$$\beta_2 := \frac{1}{4\ln b}\frac{(p_r - p_1)^2}{p_1 p_r}$$

from Theorem 2.1 satisfies

$$\beta_2 = \frac{1}{4}\left(1 - \frac{p_1}{p_r}\right)\beta_1.$$

This shows that our result in Theorem 2.1 is much better than (5.1). However both inequalities are optimal in the sense that equality obtains when $p_1 = \ldots p_r = 1/r$. Now consider the following simple example.

EXAMPLE 5.1. Let $r = 2$, $p_1 = \varepsilon \leqslant 1/2$ and $p_2 = 1-\varepsilon$. Define $H_b(\varepsilon, 1-\varepsilon) := H_b(X)$. Then $\rho = (1-\varepsilon)/\varepsilon$ and $\beta_2 \to \infty$ as $\varepsilon \to 0$. This shows that Theorem 2.1 doesn't give a nontrivial upper bound for the difference $\log_b 2 - H_b(\varepsilon, 1-\varepsilon)$ when ε is sufficiently small.

More generally, since $H_b(X) \geqslant 0$, Theorem 2.1 gives a nontrivial upper bound for the difference $\log_b r - H_b(X)$ if and only if

$$\frac{1}{4\ln b}\left(\sqrt{\rho} - 1/\sqrt{\rho}\right)^2 < \log_b r.$$

This is equivalent to $\rho - 2 + 1/\rho < 4\ln r$ or

$$\rho^2 - 2(1 + 2\ln r)\rho + 1 < 0. \tag{5.2}$$

Inequality (5.2) holds if and only if

$$\left[1 + 2\ln r + 2\sqrt{(1+\ln r)\ln r}\right]^{-1} < \rho < 1 + 2\ln r + 2\sqrt{(1+\ln r)\ln r}.$$

Since $\rho = \max_{i,k} p_i/p_k \geqslant 1$, we may conclude that Theorem 2.1 gives a nontrivial upper bound for the difference $\log_b r - H_b(X)$ only for those random variables X for which we have

$$\rho < R(\ln r),$$

where

$$R(u) := 1 + 2u + 2\sqrt{(1+u)u} \quad (u > 0).$$

In Example 5.1 this condition is $(1-\varepsilon)/\varepsilon < 1 + 2\ln 2 + 2\sqrt{(1+\ln 2)\ln 2}$, or

$$\varepsilon > \varepsilon_1 := \frac{1}{2}\left(1 - \sqrt{\frac{\ln 2}{1+\ln 2}}\right).$$

So we have nontrivial upper bound for $\log_b 2 - H_b(\varepsilon, 1-\varepsilon)$ only for $\varepsilon_1 < \varepsilon \leqslant 1/2$.

What happens if we use (1.3) and keep $\max_{1\leqslant k<n} P_k(1-P_k)$ instead of constant $1/4$? It turns out that we get better results! In Example 5.1, if we use Lemma 1.1 with $\xi_1 = 1/\varepsilon$ and $\xi_2 = 1/(1-\varepsilon)$ we get

$$0 \leqslant \log_b 2 - H_b(\varepsilon, 1-\varepsilon) \leqslant \frac{(1-2\varepsilon)^2}{\ln b}.$$

The term on the right-hand side in the second inequality tends to $1/\ln b$ as ε tends to zero. Also the upper bound is nontrivial if and only if $(1-2\varepsilon)^2/\ln b < \log_b 2$ and this condition is equivalent to

$$\varepsilon_2 := \left(1 - \sqrt{\ln 2}\right)/2 < \varepsilon < \left(1 + \sqrt{\ln 2}\right)/2.$$

Since we must have $\varepsilon \leqslant 1/2$ to ensure $\rho \geqslant 1$, the upper bound for the difference $\log_b 2 - H_b(\varepsilon, 1-\varepsilon)$ is nontrivial if and only if $\varepsilon_2 < \varepsilon \leqslant 1/2$. This result is better than the previous one since $\varepsilon_2 < \varepsilon_1$. More generally, we have the following.

LEMMA 5.2. *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* $(1,\dots,n)$ *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*

$$0 \leqslant \log_b\left(\sum_{k=1}^n p_k\xi_k\right) - \sum_{k=1}^n p_k \log_b \xi_k \leqslant \frac{M}{\ln b}\left(\sqrt{\rho}-1/\sqrt{\rho}\right)^2. \tag{5.3}$$

If

$$\rho \leqslant \Phi_M(\varepsilon) := 1 + \frac{\varepsilon \ln b}{2M} + \frac{1}{2M}\sqrt{(4M+\varepsilon\ln b)\varepsilon\ln b} \tag{5.4}$$

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

$$0 \leqslant \log_b\left(\sum_{k=1}^n p_k\xi_k\right) - \sum_{k=1}^n p_k \log_b \xi_k \leqslant \varepsilon. \tag{5.5}$$

Proof. The proof for (5.3) is evidently the same as for (1.4) with relabelling and M in place of $1/4$. Furthermore, the inequality

$$\frac{M}{\ln b}\left(\sqrt{\rho}-1/\sqrt{\rho}\right)^2 \leqslant \varepsilon$$

is equivalent to

$$\rho^2 - 2\rho\left(1+\frac{\varepsilon\ln b}{2M}\right)+1 \leqslant 0,$$

which holds if and only if

$$\left[1+\frac{\varepsilon\ln b}{2M}+\frac{1}{2M}\sqrt{(4M+\varepsilon\ln b)\varepsilon\ln b}\right]^{-1} \leqslant \rho \leqslant 1+\frac{\varepsilon\ln b}{2M}+\frac{1}{2M}\sqrt{(4M+\varepsilon\ln b)\varepsilon\ln b}.$$

Since $\rho \geqslant 1$, we see that (5.5) holds if (5.4) is satisfied. $\square$

REMARK. We have

$$\Phi_M(\varepsilon) = \Phi(\varepsilon/(4M))$$

and $\Phi(u)$ is strictly increasing for $u>0$. Since $4M \leqslant 1$, this implies that $\Phi_M(\varepsilon) \geqslant \Phi(\varepsilon)$. So Lemma 5.2 gives a better result than Lemma 1.2.

THEOREM 5.3. *Let* X *be a discrete–valued random variable with finite range* $\{x_1,\dots,x_r\}$ *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* $(1,\dots,n)$ *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*

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

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

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

Proof. Set $n=r$ and $\xi_k = 1/p_k$ in Lemma 5.2. $\square$

Again we have the situation in which Theorem 5.3 gives a nontrivial upper bound for the difference $\log_b r - H_b(X)$ if and only if

$$\frac{M}{\ln b}(\sqrt{\rho} - 1/\sqrt{\rho})^2 < \log_b r,$$

which is equivalent to

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

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

$$\rho < R_M(\ln r),$$

where

$$R_M(u) := 1 + \frac{u}{2M} + \frac{1}{2M}\sqrt{(4M+u)u}$$

for $u > 0$. We have that $R(u)$ is strictly increasing for $u > 0$ and

$$R_M(u) = R(u/(4M)) \geqslant R(u),$$

since $4M \leqslant 1$. So Theorem 5.3 gives a nontrivial upper bound for $\log_b r - H_b(X)$ for a wider class of random variables X than Theorem 2.1 does.

REMARK. 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] R. J. MCELIECE, *The theory of information and coding*, Addison-Wesley, Reading, Mass., 1977.
[3] D. S. MITRINOVIĆ, J. E. PEČARIĆ AND A. M. FINK, *Classical and new inequalities in analysis*, Kluwer Academ. Publ., Dordrecht, 1993.
[4] J. PEČARIĆ, *On an inequality of T. Popoviciu*, Bul. Si. Tehn. Inst. Politechn. Timişoara **2, 24 (38)** (1979), 9–15.

(Received January 1, 1998)

Marko Matić
Mathematics Department, FESB
R. Boškovića b.b.
Split, Croatia
e-mail: mmatic@fesb.hr

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

Josip Pečarić
Faculty of Textile Technology
University of Zagreb
Pierottijeva 6, Zagreb
Croatia
e-mail: pecaric@hazu.hr