ANZIAM J. 42(2001), 387–398

REFINEMENTS OF SOME BOUNDS IN INFORMATION THEORY

M. MATIĆ[1], C. E. M. PEARCE[2] and J. PEČARIĆ[2]

(Received 14 October 1997)

Abstract

Recently Dragomir and Goh have produced some interesting new bounds relating to entropy measures in information theory. We establish several refinements of their results.

1. Introduction

Entropy, conditional entropy and mutual information for discrete-valued random variables play important roles in information theory (see, for example, Ash [1] and McEliece [5]). A number of simple bounds have long been known for key quantities.

Suppose X is a discrete random variable assuming value x_i with probability $p_i > 0$ $(1 \le i \le n)$. For $b > 1$, the b-entropy of X is defined by

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

It is well-known (see Ash [1, Theorem 1.4.2]) that $H(X)$ is maximized when all its values occur with equal probability $1/n$, in which case $H(X) = \log_b n$.

Recently refinements have been provided for some of these results by Dragomir and Goh [3]. Thus the above-mentioned result is sharpened by Theorem A below.

THEOREM A. *If*

$$\max_{1 \le i < j \le n} |p_i - p_j| \le \sqrt{\frac{2\varepsilon \ln b}{n(n-1)}},$$

then

$$0 \le \log_b n - H(X) \le \varepsilon.$$

[1]Mathematics Department, FESB, R. Boškovića B. B., Split, Croatia.
[2]Applied Mathematics Department, The University of Adelaide, SA 5005, Australia.

https://doi.org/10.1017/S1446181100012013 Published online by Cambridge University Press

For a pair of discrete random variables X and Y with finite ranges $\{x_i\}$, $\{y_j\}$, the conditional b–entropy of X given Y is defined by

$$H(X \mid Y) := \sum_{i,j} p_{i,j} \log_b 1/p_{i|j}, \tag{1.1}$$

where, as in the sequel, $p_{i,j} := P\{X = x_i, Y = y_j\}$ and $p_{i|j} := P\{X = x_i \mid Y = y_j\}$ (see [5, p. 22]). The following result relating to conditional entropy in the context of three discrete random variables X, Y, Z was proved in [3]. As subsequently in the paper we denote the range of Z by $\{z_\ell\}$ and the three marginal distributions by (p_i), (q_j), (r_ℓ), respectively. Further we put $p_{i,j,\ell} := P(X = x_i, Y = y_j, Z = z_\ell)$, $r_{\ell|i,j} := P(Z = z_\ell \mid X = x_i, Y = y_j)$ and $r_{\ell|j} := P(Z = z_\ell \mid Y = y_j)$.

THEOREM B. *Let X, Y, Z be discrete random variables with finite ranges and let $\varepsilon > 0$ be given. If*

$$\max_{(i,j),(u,v)} |p_{i|j} - p_{u|v}| \le \sqrt{2\varepsilon \ln b/K}, \tag{1.2}$$

then we have

$$0 \le H(Z) + E(\log_b A) - H(X \mid Y) \le \varepsilon, \tag{1.3}$$

where $H(Z)$ is the b–entropy of Z and

$$A_\ell := \sum_{i,j} \alpha_{i,j,\ell}\,, \quad \alpha_{i,j,\ell} := q_j r_{\ell|i,j} = \frac{p_{i,j,\ell}}{p_{i|j}}\,, \quad \forall i, j, \ell,$$

$$K := \sum_\ell \frac{1}{r_\ell} \sum_{i,j} \alpha_{i,j,\ell} \sum_{u,v} \alpha_{u,v,\ell} = \sum_\ell \frac{1}{r_\ell} A_\ell^2.$$

The expectation in (1.3) *is taken over the sample space of Z.*

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

$$I(X;Y) := H(X) - H(X \mid Y) = \sum_{i,j} p_{i,j} \log_b \frac{p_{i,j}}{p_i q_j}. \tag{1.4}$$

If X, Y, Z are given random variables, then the mutual information $I(X, Y; Z)$, which may be interpreted as the amount of information X and Y provide about Z, is defined by

$$I(X, Y; Z) := \sum_{i,j,\ell} p_{i,j,\ell} \log_b \frac{r_{\ell|i,j}}{r_\ell} \tag{1.5}$$

(see [5, p. 26]). It is implicit that $p_{i,j} > 0$ for all pairs (i, j).

In [3] the following result was given.

THEOREM C. *We have*

$$0 \le I(X, Y; Z) - I(Y; Z)$$
$$\le \frac{1}{2\ln b} \sum_{i,j,\ell} \sum_{u,v,w} \frac{p_{i,j}}{r_{\ell|j}} \frac{p_{u,v}}{r_{w|v}} \left[r_{\ell|j} r_{w|u,v} - r_{w|v} r_{\ell|i,j} \right]^2. \tag{1.6}$$

Both inequalities become equalities if and only if $r_{\ell|i,j} = r_{\ell|j}$ *for all* (i, j, ℓ) *with* $p_{i,j,\ell} > 0$.

In this paper we show how further improvements can be given for these results. In Section 2 we give a rather more general form of Theorem A, in which $\log_b n - H(X)$ is shown to be less than or equal to each of two bounds. One of these gives a refinement of the bound given by Theorem A.

In Section 3 we give several bounds pertaining to conditional entropy, one of which provides an improvement to Theorem B. Our arguments have the simple but apparently novel feature of exploiting the fact that $p_i > 0$ and $q_j > 0$ can coexist with $p_{i,j} = 0$.

Finally, in Section 4, we give some results relating to mutual information.

2. Bounds on the entropy of a random variable

We now proceed to a strengthened version of Theorem A. We shall need the following result of Biernacki, Pidek and Ryll–Nardzewski [2].

THEOREM D. *Let* (a_i) *and* (b_i) *be n-tuples such that* $c_1 \le a_i \le c_2$ *and* $d_1 \le b_i \le d_2$ *for* $1 \le i \le n$. *Then*

$$\left| \frac{1}{n} \sum_{i=1}^{n} a_i b_i - \frac{1}{n^2} \sum_{i=1}^{n} a_i \sum_{j=1}^{n} b_j \right| \le \frac{1}{n} \left[\frac{n}{2}\right] \left(1 - \frac{1}{n}\left[\frac{n}{2}\right]\right) (c_2 - c_1)(d_2 - d_1).$$

Here, as in the following theorem, $[x]$ denotes the largest integer less than or equal to x.

THEOREM 2.1. *Suppose the random variable X admits values* x_i *with respective probabilities* $p_i > 0$ $(i = 1, \dots, n)$ *and let* $M = \max_i p_i$ *and* $m = \min_i p_i$. *Then*

$$0 \le \log_b n - H(X) \le \frac{(M - m)^2}{\ln b} \min \left\{ \left[\frac{n}{2}\right] \left(n - \left[\frac{n}{2}\right]\right), \frac{1}{4Mm} \right\}.$$

If

$$\max_{1 \le i < j \le n} |p_i - p_j| \le \sqrt{\frac{\varepsilon \ln b}{\left[\frac{n}{2}\right]\left(n - \left[\frac{n}{2}\right]\right)}}, \tag{2.1}$$

https://doi.org/10.1017/S1446181100012013 Published online by Cambridge University Press

then

$$0 \leq \log_b n - H(X) \leq \varepsilon.$$

PROOF. From [3, Theorem 4.3] we have

$$0 \leq \log_b n - H(X) \leq \frac{1}{\ln b}\left[n \sum_{i=1}^{n} p_i^2 - 1\right]. \tag{2.2}$$

Putting $a_i = b_i = p_i$ in Theorem D provides

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

or

$$n\sum_{i=1}^{n} p_i^2 - 1 \leq \left[\frac{n}{2}\right]\left(n - \left[\frac{n}{2}\right]\right)(M-m)^2. \tag{2.3}$$

Combining (2.2) and (2.3) yields

$$0 \leq \log_b n - H(X) \leq \frac{1}{\ln b}\left[\frac{n}{2}\right]\left(n - \left[\frac{n}{2}\right]\right)(M-m)^2. \tag{2.4}$$

In [4, Theorem 2.1] we showed that if $\rho := \max_{i,k} p_i/p_k$, then

$$0 \leq \log_b n - H(X) \leq \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2.$$

Here $\rho = M/m$, so

$$0 \leq \log_b n - H(X) \leq \frac{1}{4\ln b}\frac{(M-m)^2}{Mm}. \tag{2.5}$$

Combining (2.4) and (2.5) gives the first part of the enunciation.

Since

$$\max_{1 \leq i < j \leq n} |p_i - p_j| = M - m,$$

the second part follows at once from (2.1) and (2.4).

Since $2\left[\frac{n}{2}\right]\left(n - \left[\frac{n}{2}\right]\right) < n(n-1)$ for $n > 2$, the second part of the theorem is clearly stronger than Theorem A.

3. Bounds on conditional entropy

We shall need the following preliminary result established in [3].

LEMMA 3.1. *Let* $\xi_k \in (0, \infty)$ *and* $s_k \geq 0$ $(1 \leq k \leq n)$ *with* $\sum_{k=1}^n s_k = 1$ *and suppose* $b > 1$. *Then*

$$0 \leq \log_b \left(\sum_{k=1}^n s_k \xi_k \right) - \sum_{k=1}^n s_k \log_b \xi_k \leq \frac{1}{\ln b} \left[\sum_{j=1}^n \frac{s_j}{\xi_j} \sum_{k=1}^n s_k \xi_k - 1 \right].$$

We shall need also the following discrete version of the Grüss inequality (see [6, Chapter 10]).

LEMMA 3.2. *Suppose* $s_k \geq 0$ *with* $u \leq a_k \leq U$ *and* $v \leq b_k \leq V$ $(1 \leq k \leq n)$. *Then*

$$\left| \sum_{i=1}^n s_i \sum_{j=1}^n s_j a_j b_j - \sum_{i=1}^n s_i a_i \sum_{j=1}^n s_j b_j \right| \leq \frac{1}{4}(U - u)(V - v) \left(\sum_{i=1}^n s_i \right)^2. \tag{3.1}$$

We now proceed to provide an upper bound on the conditional entropy of a pair of discrete-valued random variables. We shall take advantage of the fact that in the definition (1.1) of conditional entropy, the summation is only over those pairs (i, j) for which $p_{i,j} > 0$.

THEOREM 3.3. *Suppose* X, Y *are random variables each with a finite range. Define* $V_j := \{i : p_{i,j} > 0\}$, *put* $U := \{(i, j) : i \in V_j\}$ *and let*

$$r' := \sum_j q_j \,|V_j|.$$

Further define $M = \max_{(i,j)\in U} p_{i|j}$ *and* $m = \min_{(i,j)\in U} p_{i|j}$. *Then*

$$0 \leq \log_b r' - H(X \mid Y) \leq \frac{(M - m)^2}{4 \ln b} \min \left\{ \frac{1}{Mm}, r'^2 \right\}.$$

If

$$\max_{(i,j),(u,v)\in U} |p_{i|j} - p_{u|v}| \leq \frac{2}{r'} \sqrt{\varepsilon \ln b}, \tag{3.2}$$

then

$$0 \leq \log_b r' - H(X \mid Y) \leq \varepsilon. \tag{3.3}$$

https://doi.org/10.1017/S1446181100012013 Published online by Cambridge University Press

PROOF. We may label pairs $(i, j) \in U$ as $k = 1, \ldots, n$, say, and then apply Lemma 3.1 with $s_k = p_{i,j}$ and $\xi_k = 1/p_{i|j} = q_j/p_{i,j}$. This gives

$$\begin{aligned}
0 &\le \log_b \left[\sum_{(i,j)\in U} p_{i,j} \frac{q_j}{p_{i,j}} \right] - \sum_{(i,j)\in U} p_{i,j} \log_b \frac{q_j}{p_{i,j}} \\
&\le \frac{1}{\ln b} \left[\sum_{(i,j)\in U} p_{i,j} \frac{p_{i,j}}{q_j} \sum_{(k,\ell)\in U} p_{k,\ell} \frac{q_\ell}{p_{k,\ell}} - 1 \right] \\
&= \frac{1}{\ln b} \left[\sum_{(i,j)\in U} q_j p_{i|j}^2 \sum_{(k,\ell)\in U} q_\ell - \left(\sum_{(i,j)\in U} q_j p_{i|j} \right)^2 \right] \\
&\le \frac{1}{4\ln b} \left[\max_{(i,j)\in U} p_{i|j} - \min_{(i,j)\in U} p_{i|j} \right]^2 \left(\sum_{(i,j)\in U} q_j \right)^2 . \qquad (3.4)
\end{aligned}$$

The last inequality follows from Lemma 3.2 with $s_k = q_j$ and $a_k = b_k = p_{i|j}$ in (3.1). Now

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

so from (3.4) we get

$$0 \le \log_b r' - H(X \mid Y) \le \frac{1}{4\ln b}(M - m)^2 r'^2. \qquad (3.5)$$

In [4, Theorem 3.1] we proved that

$$0 \le \log_b r' - H(X \mid Y) \le \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2,$$

where $\rho := \max_{(i,j),(u,v)\in U} p_{i|j}/p_{u|v} = M/m$. So we have

$$0 \le \log_b r' - H(X \mid Y) \le \frac{1}{4\ln b}\frac{(M - m)^2}{Mm}. \qquad (3.6)$$

The first part of the enunciation follows from (3.5) and (3.6).

For the second we need only note that if (3.2) holds, then (3.3) follows from (3.5).

Our next theorem gives an improvement of Theorem B.

THEOREM 3.4. *Let X, Y be a pair of random variables as in Theorem 3.3 and let Z be a discrete-valued random variable assuming values z_k $(1 \le k \le t)$ each with* *positive probability* r_k $(1 \leq k \leq t)$. *We define an associated random variable* A *which assumes the value*

$$A_k := \sum_{(i,j)\in U} p_{i,j,k}/p_{i|j}$$

with probability r_k $(1 \leq k \leq t)$. *We define also*

$$K := \sum_{k=1}^{t} \frac{1}{r_k}\left[\sum_{(i,j)\in U} \alpha_{i,j,k}\right]^2 = \sum_{k=1}^{t} \frac{1}{r_k} A_k^2,$$

where $\alpha_{i,j,k} := p_{i,j,k}/p_{i|j}$ *for* $(i,j) \in U$. *Finally we put*

$$M = \max_{(i,j)\in U} p_{i|j}, \qquad m = \min_{(i,j)\in U} p_{i|j}.$$

Then we have

$$0 \leq H(Z) + E(\log_b A) - H(X \mid Y) \leq \frac{(M-m)^2}{4\ln b} \min\left\{K, \frac{1}{Mm}\right\}.$$

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

$$\max_{(i,j),(u,v)\in U} |p_{i|j} - p_{u|v}| \leq 2\sqrt{\frac{\varepsilon \ln b}{K}}, \tag{3.7}$$

then

$$0 \leq H(Z) + E(\log_b A) - H(X \mid Y) \leq \varepsilon. \tag{3.8}$$

PROOF. By Lemma 3.2 with $a_k = b_k = p_{i|j}$ and $s_k = \alpha_{i,j,\ell}$ $(k = (i,j) \in U)$, we have for fixed ℓ that

$$\begin{aligned}
&\sum_{(i,j)\in U} \frac{p_{i,j,\ell}}{r_\ell} p_{i|j} \sum_{(k,h)\in U} \frac{p_{k,h,\ell}}{r_\ell} \frac{1}{p_{k|h}} - 1 \\
&\quad = \frac{1}{r_\ell^2}\left[\sum_{(i,j)\in U} \frac{p_{i,j,\ell}}{p_{i|j}} p_{i|j}^2 \sum_{(k,h)\in U} \frac{p_{k,h,\ell}}{p_{k|h}} - r_\ell^2\right] \\
&\quad = \frac{1}{r_\ell^2}\left[\sum_{(i,j)\in U} \alpha_{i,j,\ell} p_{i|j}^2 \sum_{(k,h)\in U} \alpha_{k,h,\ell} - \left(\sum_{(i,j)\in U} \alpha_{i,j,\ell} p_{i|j}\right)^2\right] \\
&\quad \leq \frac{1}{4r_\ell^2}\left(\max_{(i,j)\in U} p_{i|j} - \min_{(i,j)\in U} p_{i|j}\right)^2 \left(\sum_{(k,h)\in U} \alpha_{k,h,\ell}\right)^2 \\
&\quad = \frac{1}{4r_\ell^2}(M-m)^2 A_\ell^2.
\end{aligned}$$

By Lemma 3.1 with $s_k = p_{i,j,\ell}/r_\ell$ and $\xi_k = 1/p_{i|j}$ ($k = (i,j) \in U$) we have

$$\begin{aligned} 0 &\le \log_b\left[\sum_{(i,j)\in U}\frac{p_{i,j,\ell}}{r_\ell}\frac{1}{p_{i|j}}\right] - \sum_{(i,j)\in U}\frac{p_{i,j,\ell}}{r_\ell}\log_b\frac{1}{p_{i|j}} \\ &\le \frac{1}{\ln b}\left[\sum_{(i,j)\in U}\frac{p_{i,j,\ell}}{r_\ell}p_{i|j}\sum_{(k,h)\in U}\frac{p_{k,h,\ell}}{r_\ell}\frac{1}{p_{k|h}} - 1\right] \\ &\le \frac{1}{4r_\ell^2\ln b}(M-m)^2A_\ell^2. \end{aligned}$$

Multiplication by r_ℓ and summation over $\ell = 1, \dots, t$ yields

$$\begin{aligned} 0 &\le H(Z) + E(\log_b A) - H(X \mid Y) \\ &\le \frac{1}{4\ln b}(M-m)^2\sum_{l=1}^{t}\frac{A_\ell^2}{r_\ell} = \frac{(M-m)^2K}{4\ln b}. \end{aligned} \tag{3.9}$$

We deduce that (3.8) follows if (3.7) holds.

In [4, Theorem 3.2] we proved that

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

where $\rho := \max_{(i,j),(u,v)\in U} p_{i|j}/p_{u|v}$. Here $\rho = M/m$ so

$$0 \le H(Z) + E(\log_b A) - H(X \mid Y) \le \frac{1}{4\ln b}\frac{(M-m)^2}{Mm}.$$

Combining this with (3.9) gives the main part of the theorem.

The factor 2 in (3.7) replaces the $\sqrt{2}$ of (1.2), so the second part of Theorem 3.4 represents a strengthening of Theorem B.

We can use Theorem 3.4 to improve [3, Corollary 5.4].

COROLLARY 3.5. *Suppose X, Y have the same range. Define*

$$Z := \begin{cases} 0 & \text{if } X = Y, \\ 1 & \text{if } X \ne Y. \end{cases}$$

Further, define

$$\begin{aligned} T_j &:= |\{i : i \ne j, p_{i,j} > 0\}|, \\ R_j &:= |V_j| - T_j = \begin{cases} 1 & \text{if } p_{j,j} > 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

https://doi.org/10.1017/S1446181100012013 Published online by Cambridge University Press

Let $P_e := P\{X \neq Y\} = P\{Z = 1\}$ *and let* $\varepsilon > 0$ *be given. If*

$$\max_{(i,j),(u,v)\in U} |p_{i|j} - p_{u|v}| \leq 2\sqrt{\frac{\varepsilon(1-P_e)P_e \ln b}{A^2(0)P_e + A^2(1)(1-P_e)}}, \tag{3.10}$$

then

$$0 \leq H(P_e) + P_e \log_b A(1) + (1-P_e)\log_b A(0) - H(X \mid Y) \leq \varepsilon,$$

where $A(0) = \sum_j q_j R_j$, $A(1) = \sum_j q_j T_j$ *and*

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

PROOF. We may take $y_j = x_j$ for all j, since X and Y have the same range. The random variable A assumes the values

$$A(0) = \sum_{(i,j)\in U} \alpha_{i,j,0} = \sum_j q_j \sum_{i\in V_j} r_{0|i,j} = \sum_j q_j R_j$$

and

$$A(1) = \sum_{(i,j)\in U} \alpha_{i,j,1} = \sum_j q_j \sum_{i\in V_j} r_{1|i,j} = \sum_j q_j T_j.$$

We derive

$$K = \frac{A^2(0)}{1-P_e} + \frac{A^2(1)}{P_e} = \frac{A^2(0)P_e + A^2(1)(1-P_e)}{P_e(1-P_e)}.$$

As in the proof of Theorem 3.4 we have

$$0 \leq H(P_e) + P_e \log_b A(1) + (1-P_e)\log_b A(0) - H(X \mid Y) \leq \frac{(M-m)^2 K}{4\ln b}. \tag{3.11}$$

Under (3.10), the result follows from (3.11).

REMARK 3.6. We have $0 \leq A(0) \leq 1$ and $0 \leq A(1) \leq r-1$. Hence

$$P_e A^2(0) + (1-P_e)A^2(1) \leq P_e + (1-P_e)(r-1)^2 = 1 + (1-P_e)(r^2 - 2r)$$

and our condition (3.10) is better than the corresponding condition

$$M - m \leq \sqrt{\frac{2\varepsilon P_e(1-P_e)\ln b}{1 + (1-P_e)(r^2-2r)}}$$

in [3].

REMARK 3.7. It can happen that $p_{j,j} = 0$ for all j. In this case $R_j = 0$ for all j and so $A(0) = 0$, and also $P\{X = Y\} = 0$ and hence $P_e = 1$. Then $(1-P_e)\log_b A(0) = 0$, since $0 \log 0 := 0$. Also $A(0) + A(1) = r' \leq r$.

4. Bounds on mutual information

As with Theorem 3.3, our analysis in the following result hangs on the fact that in the definition of mutual information (1.4), the summation is over those pairs (i, j) for which $p_{i,j} > 0$.

THEOREM 4.1. *Let* $V := \{(i, j) : p_{i,j} > 0\}$ *and*

$$K := \sum_{(i,j)\in V} \frac{p_{i,j}^2}{p_i q_j}, \qquad S := \sum_{(i,j)\in V} p_i q_j .$$

If

$$\max_{(i,j),(u,v)\in V} \left| \frac{p_i q_j}{p_{i,j}} - \frac{p_u q_v}{p_{u,v}} \right| \le \frac{2}{K}\sqrt{\varepsilon \ln b}, \tag{4.1}$$

then

$$0 \le \log_b S + I(X; Y) \le \varepsilon.$$

PROOF. Take s_k as $p_{i,j}$ and ξ_k as $p_i q_j / p_{i,j}$, for $(i, j) \in V$, in Lemma 3.1. This gives

$$\begin{aligned}
0 &\le \log_b \left(\sum_{(i,j)\in V} p_{i,j} \frac{p_i q_j}{p_{i,j}} \right) - \sum_{(i,j)\in V} p_{i,j} \log_b \frac{p_i q_j}{p_{i,j}} \\
&\le \frac{1}{\ln b} \left[\sum_{(i,j)\in V} p_{i,j} \frac{p_i q_j}{p_{i,j}} \sum_{(u,v)\in V} p_{u,v} \frac{p_{u,v}}{p_u q_v} - 1 \right] \\
&= \frac{1}{\ln b} \left[\sum_{(i,j)\in V} \frac{p_{i,j}^2}{p_i q_j} \left(\frac{p_i q_j}{p_{i,j}} \right)^2 \sum_{(u,v)\in V} \frac{p_{u,v}^2}{p_u q_v} - \left(\sum_{(i,j)\in V} \frac{p_{i,j}^2}{p_i q_j} \frac{p_i q_j}{p_{i,j}} \right)^2 \right] \\
&\le \frac{1}{4 \ln b} \left[\max_{(i,j)\in V} \frac{p_i q_j}{p_{i,j}} - \min_{(i,j)\in V} \frac{p_i q_j}{p_{i,j}} \right]^2 \left(\sum_{(u,v)\in V} \frac{p_{u,v}^2}{p_u q_v} \right)^2 \le \varepsilon,
\end{aligned}$$

since

$$\max_{(i,j),(u,v)\in V} \left| \frac{p_i q_j}{p_{i,j}} - \frac{p_u q_v}{p_{u,v}} \right| = M - m,$$

where $M = \max_{(i,j)\in V} p_i q_j / p_{i,j}$ and $m = \min_{(i,j)\in V} p_i q_j / p_{i,j}$. The third inequality in the proof follows from Lemma 3.2 with $a_k = b_k = p_i q_j / p_{i,j}$ and $s_k = p_{i,j}^2 / p_i q_j$ for $(i, j) \in V$.

REMARK 4.2. We have $0 < S \le 1$ and it can happen that $S < 1$. For example, if X and Y both have range $\{0, 1\}$ and

$$p_{0,0} = p_{1,1} = 0, \quad p_{1,0} = p_{0,1} = 1/2,$$

then we have $S = 1/2$. Of course $S = 1$ if $p_{i,j} > 0$ for all pairs (i, j).

REMARK 4.3. In [3] the same conclusion is given but without the improvement provided by the term $\log_b S$ and under the condition $M - m \leq (1/K)\sqrt{2\varepsilon \ln b}$. Our condition (4.1) is weaker. Also, in [4] it was shown that $I(X;Y) \leq (1/4\ln b)(\sqrt{\rho} - 1/\sqrt{\rho})^2$, where $\rho = M/m$. The improvement provided by the term $\log_b S$ was missed but follows at once from the argument, so that in fact we have

$$\log_b S + I(X;Y) \leq \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2.$$

Combining this with the proof above we get

$$0 \leq \log_b S + I(X;Y) \leq \frac{(M-m)^2}{4\ln b}\left\{K^2, \frac{1}{Mm}\right\}.$$

Finally we address strengthening Theorem C. Definition 1.5 requires that $p_{i,j} > 0$ hold for all pairs (i, j). The summations need only be over the set of triples

$$W := \{(i, j, \ell) : p_{i,j,\ell} > 0\}.$$

Let

$$T := \sum_{(i,j,\ell)\in W} p_{i,j} r_{\ell|j}.$$

Then $0 < T \leq 1$ and it can happen that $T < 1$. For example, if X, Y, Z have the range $\{0, 1\}$ and

$$p_{1,0,0} = p_{0,1,0} = p_{1,1,0} = p_{0,0,1} = 1/4$$

with $p_{i,j,\ell} = 0$ for other all triples, then $T = 3/4$. As with our discussion in Remark 4.3, the proof of Theorem C actually provides the improvement that the middle quantity in (1.6) be replaced by $\log_b T + I(X, Y; Z) - I(Y; Z)$. We can make a further improvement.

THEOREM 4.4. *Suppose the conditions of Theorem* C *are satisfied and*

$$M := \max_{(i,j,\ell)\in W} \frac{r_{\ell|i,j}}{r_{\ell|j}}, \quad m := \min_{(i,j,\ell)\in W} \frac{r_{\ell|i,j}}{r_{\ell|j}}.$$

Then

$$0 \leq \log_b T + I(X, Y; Z) - I(Y; Z) \leq \frac{(M-m)^2}{4\ln b} \min\left\{T^2, \frac{1}{Mm}\right\}. \tag{4.2}$$

PROOF. Relation (1.6) can be rewritten in the form

$$\log_b T + I(X, Y; Z) - I(Y; Z)$$
$$\leq \frac{1}{\ln b}\left[\sum_{(i,j,\ell)\in W} p_{i,j} r_{\ell|j} \sum_{(u,v,w)\in W} p_{u,v} r_{w|v}\left(\frac{r_{w|u,v}}{r_{w|v}}\right)^2 - \left(\sum_{(i,j,\ell)\in W} p_{i,j} r_{\ell|j} \frac{r_{\ell|i,j}}{r_{\ell|j}}\right)^2\right].$$

https://doi.org/10.1017/S1446181100012013 Published online by Cambridge University Press

Now we can apply Lemma 3.2 with $s_k = p_{i,j} r_{\ell|j}$ and $a_k = b_k = r_{\ell|i,j}/r_{\ell|j}$, for $(i, j, \ell) \in W$, to obtain

$$\begin{aligned} 0 &\le \log_b T + I(X, Y; Z) - I(Y; Z) \\ &\le \frac{1}{4\ln b}(M-m)^2 \left(\sum_{(i,j,\ell)\in W} p_{i,j} r_{\ell|j} \right)^2 \\ &= \frac{1}{4\ln b}(M-m)^2 T^2. \end{aligned} \tag{4.3}$$

In [3, Theorem 4.2] it was proved that

$$I(X, Y; Z) - I(Y; Z) \le (1/4\ln b)(\sqrt{\rho} - 1/\sqrt{\rho})^2$$

and again we may insert the term $\log_b T$ on the left. Since $\rho = M/m$, we have

$$0 \le \log_b T + I(X, Y; Z) - I(Y; Z) \le \frac{1}{4\ln b}\frac{(M-m)^2}{Mm}. \tag{4.4}$$

Combining (4.3) and (4.4) provides the desired result.

References

[1] R. Ash, *Information theory* (Interscience, New York, 1965).

[2] M. Biernacki, H. Pidek and C. Ryll-Nardzewski, "Sur une inégalité entre des intégrales définies", *Ann. Univ. Mariae Curie–Skłodowska Sect. A Math.* 4 (1950) 1–4.

[3] 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 (1996) 1–11.

[4] M. Matić, C. E. M. Pearce and J. Pečarić, "Improvements of some bounds on entropy measures in information theory", *Math. Ineq. and Applic.* 1 (1998) 295–304.

[5] R. J. McEliece, *The theory of information and coding* (Addison–Wesley, Reading, Mass., 1977).

[6] D. S. Mitrinović, J. E. Pečarić and A. M. Fink, *Classical and new inequalities in analysis* (Kluwer Academ. Publ., Dordrecht, 1993).