Mathematical
Inequalities
& Applications
Volume 4, Number 2 (2001), 209–214

ON AN INEQUALITY FOR THE ENTROPY OF A PROBABILITY DISTRIBUTION

CVETAN JARDAS, JOSIP PEČARIĆ, RAJKO ROKI AND NIKOLA SARAPA

(*communicated by I. Pinelis*)

Abstract. In this paper we prove that Alzer's inequality for the entropy of a probability distribution (see [3]) is valid with reverse sign of the inequality.

Introduction

Let $(p_k > 0,\ k = 0, 1, 2, \dots)$ be a discrete probability distribution (finite or countable infinite). The Shannon entropy of this distribution is defined by

$$H = -\sum_k p_k \log p_k.$$

The quantity H is a measure of information of the distribution $(p_k, k = 0, 1, 2, \dots)$ and it plays a key role in information theory. For this reason in many applications of discrete probability distributions, it is important to find lower and upper bounds for the entropy H. Some basic properties of entropies of probability distributions can be found in [5].

J.–P. Allouche, M. Mendes France and G. Tenenbaum [2] have proved the following result:

THEOREM A. *Let* $(p_k, k \geqslant 0)$ *be a countable infinite probability distribution such that*

$$\lambda = \sup_{n \geqslant 0}(p_n^{-1} \sum_{k=n+1}^{\infty} p_k) < \infty.$$

Then

$$H \leqslant F(\lambda),$$

Mathematics subject classification (2000): 26D15, 94A17, 28D20.

Key words and phrases: Entropy of probability distribution, inequality for infinite series, convex function.

© ELEMENT, Zagreb
Paper MIA-04-19

where $F(x) = (x+1)\log(x+1) - x\log x$, $x > 0$, *with equality if* $p_k = \dfrac{\lambda^k}{(\lambda+1)^{k+1}}$ $(k \geqslant 0)$.

H. Alzer [3] tried to give an improvement of the above inequality, i.e. he has considered the following inequality:

$$H + \sum_{n=0}^{\infty}(\lambda p_n - P_{n+1})(\log P_n - \log P_{n+1}) \leqslant F(\lambda), \qquad (*)$$

where H, λ and $F(\lambda)$ are defined as in Theorem A, and $P_n = \sum_{k=n}^{\infty} p_k$. However, this inequality is not true, and a correction of the second term on the left-hand side of $(*)$ was given in [4].

In this paper we can prove that the reverse inequality in $(*)$ is valid. It is an unexpected result because Alzer had used in his "proof" of $(*)$ two inequalities which, in fact, had opposite signs.

Results

We first prove the following result (we suppose that all series in the next theorem converge), which gives some further extensions of results in [2] and [4]:

THEOREM 1. *Let* $\lambda > 0$ *and* $a_n > 0$ $(n = 0, 1, 2, \dots)$ *be real numbers and let* $A_n = \sum_{k=n}^{\infty} a_k < \infty$ $(n = 0, 1, 2, \dots)$. *Then, for all* p, $0 < p < 1$ *or* $p > 2$, *we have*

$$\begin{aligned} &p[(\lambda+1)^{p-1} - \lambda^{p-1}]\sum_{n=0}^{\infty}(A_{n+1} - \lambda a_n)a_n^{p-1} + \\ &+[(\lambda+1)^p - \lambda^p]\sum_{n=0}^{\infty} a_n^p \leqslant \Big(\sum_{n=0}^{\infty} a_n\Big)^p \leqslant \\ &\leqslant p\sum_{n=0}^{\infty}(A_{n+1} - \lambda a_n)(A_n^{p-1} - A_{n+1}^{p-1}) + [(\lambda+1)^p - \lambda^p]\sum_{n=0}^{\infty} a_n^p. \end{aligned} \qquad (1)$$

For all $p, 1 < p < 2$, *the opposite inequalities hold in (1). Equalities hold in (1) if and only if* $a_n = a_0\left(\dfrac{\lambda}{\lambda+1}\right)^n$ $(n = 0, 1, 2, \dots)$.

Proof. Let $c > 0$ be a fixed real number and $p \in (0,1)$ or $p \in (2,\infty)$. Then the function

$$f(x) = (x+c)^p - x^p,$$

is strictly convex on $(0, +\infty)$. Therefore, for all $x, y > 0$ we have

$$f(y) + (x-y)f'(y) \leqslant f(x) \leqslant f(y) + (x-y)f'(x). \qquad (2)$$

Setting in (2)

$$x = A_{n+1}, \quad y = \lambda a_n \text{ and } \quad c = a_n,$$

and using the fact that $A_{n+1} + a_n = A_n (n = 0, 1, 2, \dots)$ it follows

$$\begin{aligned} &[(\lambda+1)^p - \lambda^p]a_n^p + p(A_{n+1} - \lambda a_n)[(\lambda+1)^{p-1} - \lambda^{p-1}]a_n^{p-1} \leqslant \\ &\leqslant A_n^p - A_{n+1}^p \leqslant [(\lambda+1)^p - \lambda^p]a_n^p + p(A_{n+1} - \lambda a_n)(A_n^{p-1} - A_{n+1}^{p-1}). \end{aligned} \tag{3}$$

Summing up inequalities in (3) for all $n = 0, 1, 2, \dots$ leads to (1) since

$$\sum_{n=0}^{\infty}(A_n^p - A_{n+1}^p) = A_0^p = \left(\sum_{n=0}^{\infty} a_n\right)^p.$$

If $p \in (1, 2)$ the function f is strictly concave, so we have opposite inequalities in (2) and in the same way we get opposite inequalities in (1).

Since equalities hold in (2) if and only if $x = y$, and since all series in (1) converge, we conclude that equalities hold in (1) if and only if $A_{n+1} = \lambda a_n$, for all $n \geqslant 0$ and it is easy to verify that this is equivalent to $a_n = a_0\left(\frac{\lambda}{\lambda+1}\right)^n$ for all $n \geqslant 0$. □

REMARK 1. Let $0 < p < 1$. If we suppose in Theorem 1 that in addition we have

$$A_{n+1} = \sum_{k=n+1}^{\infty} a_k \leqslant \lambda\, a_n, \quad n = 0, 1, 2, \dots, \tag{4}$$

then the first expression on the left–hand side of the first inequality in (1) is non-negative, hence

$$[(\lambda+1)^p - \lambda^p]\sum_{n=0}^{\infty} a_n^p \leqslant \left(\sum_{n=0}^{\infty} a_n\right)^p. \tag{5}$$

This inequality has been proved in [2].

REMARK 2. It is easy to check that (4) implies $A_n \leqslant \left(\frac{\lambda}{\lambda+1}\right)^n A_0 \ (n \in \mathbf{N})$ and therefore $a_n \leqslant \left(\frac{\lambda}{\lambda+1}\right)^n A_0 \ (n \in \mathbf{N})$. Since A_0 is finite by (4), we conclude that $\sum_{n=0}^{\infty} a_n^p$ and $\sum_{n=0}^{\infty} A_n^p$ are convergent series for all $p > 0$. Also, if (4) holds, then for all $p > 0$ we have $\sum_{n=0}^{\infty} A_{n+1}a_n^{p-1} \leqslant \lambda \sum_{n=0}^{\infty} a_n^p < \infty$.

REMARK 3. It is easy to check that for $p = 1$ and $p = 2$ equalities hold in (1).

Now, using inequalities of Theorem 1, similarly as in Remark 1 we obtain a generalization of (5):

COROLLARY 1. *Let* $\lambda > 0$ *and* $a_n > 0$ $(n = 0, 1, 2, \dots)$ *be real numbers.*

(i) Suppose that (4) holds. If $0 < p < 1$*, then (5) holds. If* $p > 1$*, then we have the opposite inequality in (5).*

(ii) Suppose that $A_{n+1} \geqslant \lambda a_n$ $(n = 0, 1, 2, \dots)$*. If* $0 < p < 1$*, then we have the opposite inequality in (5). If* $p > 1$*, then (5) holds.*

The following theorem is a consequence of Theorem 1.

THEOREM 2. *Let* $(p_n > 0, n \geqslant 0)$ *be a probability distribution such that*

$$\sup_{n \geqslant 0}(p_n^{-1} P_{n+1}) \leqslant \lambda < \infty, \tag{6}$$

where $P_n = \sum_{k=n}^{\infty} p_k (n \geqslant 0)$*. Then, the Shannon entropy*

$$H = -\sum_{n=0}^{\infty} p_n \log p_n,$$

of $(p_n, n \geqslant 0)$ *exists and we have*

$$H + [\log(\lambda + 1) - \log \lambda] \sum_{n=0}^{\infty} (\lambda p_n - P_{n+1}) \leqslant F(\lambda) \leqslant \tag{7}$$

$$\leqslant H + \sum_{n=0}^{\infty} (\lambda p_n - P_{n+1})(\log P_n - \log P_{n+1}),$$

where $F(x) = (x + 1)\log(x + 1) - x \log x$ $(x > 0)$*. Equalities hold in (7) if* $p_n = \dfrac{\lambda^n}{(\lambda + 1)^{n+1}} (n \geqslant 0)$.

Proof. We can use Theorem 1 for $a_n = p_n$ and $A_n = P_n (n = 0, 1, 2, \dots)$. Set

$$a = a(p) = [(\lambda + 1)^p - \lambda^p] \sum_{n=0}^{\infty} p_n^p,$$

$$b = b(p) = p[(\lambda + 1)^{p-1} - \lambda^{p-1}] \sum_{n=0}^{\infty} (P_{n+1} - \lambda p_n) p_n^{p-1}.$$

Then, if $1 < p < 2$, it follows from Theorem 1:

$$a + b \geqslant \left(\sum_{n=0}^{\infty} p_n \right)^p = 1. \tag{8}$$

The condition (6) implies $\lambda p_n - P_{n+1} \geqslant 0$, for all $n \geqslant 0$, so using Remark 2 we conclude that the series in definition of a and b are uniformly convergent for $p \in [1, 2)$ and that the derivatives of these series are equal to the sums of the derivatives of corresponding terms. Therefore, we have $\lim_{p\to 1+} a(p) = 1$, $\lim_{p\to 1+} b(p) = 0$. It follows easily from (8)

$$\frac{1}{1-p}\log\sum_{n=0}^{\infty} p_n^p \leqslant \frac{1}{1-p}\log(1-b) + \frac{1}{p-1}\log[(\lambda+1)^p - \lambda^p]. \tag{9}$$

The quantity $H(p) = \frac{1}{1-p}\log\sum_{n=0}^{\infty} p_n^p$ is the Rényi's entropy of order p $(p \neq 1)$ of the distribution $(p_n, n \geqslant 0)$ (see [6]).

In limiting case $p \longrightarrow 1+$, the left–hand side of (9) gives $-\sum_{n=0}^{\infty} p_n \log p_n = H$, while the right–hand side of (9) becomes

$$[\log(\lambda+1) - \log\lambda]\sum_{n=0}^{\infty}(P_{n+1} - \lambda p_n) + F(\lambda).$$

Together, it is the first inequality in (7).

Let us prove the second inequality in (7). Set

$$c = c(p) = p\sum_{n=0}^{\infty}(P_{n+1} - \lambda p_n)(P_n^{p-1} - P_{n+1}^{p-1}).$$

Using Remark 2 we conclude $\lim_{p\to 1+} c(p) = 0$. If $1 < p < 2$, it follows from Theorem 1:

$$a + c \leqslant \left(\sum_{n=0}^{\infty} p_n\right)^p = 1, \tag{10}$$

and therefore

$$\frac{1}{1-p}\log\sum_{n=0}^{\infty} p_n^p \geqslant \frac{1}{1-p}\log(1-c) + \frac{1}{p-1}\log[(\lambda+1)^p - \lambda^p]. \tag{11}$$

As we have already seen, in limiting case $p \longrightarrow 1+$, the left–hand side of (11) gives H, while the right–hand side of (11) becomes

$$\sum_{n=0}^{\infty}(P_{n+1} - \lambda p_n)(\log P_n - \log P_{n+1}) + F(\lambda).$$

Together, it is the second inequality in (7).

The last statement follows easily from the fact that $p_n = \dfrac{\lambda^n}{(\lambda+1)^{n+1}}$ $(n = 0, 1, 2, \dots)$ is equivalent to $\lambda p_n = P_{n+1}$ $(n = 0, 1, 2, \dots)$, so (7) becomes $H = F(\lambda)$. $\square$

REMARK 4. It is easy to check that the condition (6) implies that the series in definition of a, b and c in the proof of Theorem 2 are uniformly convergent for

$p \in [p_0, 1]$ for each fixed p_0, $0 < p_0 < 1$. Thus, we can get inequalities (7) by use of inequalities (1) for $0 < p < 1$ and by taking limits $p \longrightarrow 1-$.

REMARK 5. In fact, the first inequality in (7) was proved by Daróczy in [4], while the second is $(*)$ but with reverse sign.

REFERENCES

[1] J.–P. ALLOUCHE AND M. MENDES FRANCE, *On an extremal property of the Rudin–Shapiro sequence*, Mathematica **42** (1985), 33–38.

[2] J.–P. ALLOUCHE, M. MENDES FRANCE AND G. TENENBAUM, *Entropy: An inequality*, Tokyo J. Math. **11, 2** (1988), 323–328.

[3] H. ALZER, *Note on an inequality for infinite series*, Acta Math. Hungar. **67(3)** (1995), 203–206.

[4] Z. DARÓCZY, *Inequalities for some infinite series*, Acta Math. Hungar. **75(1–2)** (1997), 27–30.

[5] Ž. PAUŠE, *Uvod u teoriju informacije, 2. izd.*, Školska knjiga, Zagreb, 1989.

[6] A. RÉNYI, *Wahrscheinlichkeitsrechnung (mit einem Anhang über Informationstheorie)*, VEB Deutscher Verlag der Wisssenschaften, Berlin, 1966.

(Received April 13, 2000)

Cvetan Jardas
Department of Mathematics
University of Rijeka
51000 Rijeka
Omladinska 14
Croatia

Josip Pečarić
Faculty of Textile Technology
University of Zagreb
10000 Zagreb
Pierottijeva 6
Croatia

Rajko Roki
Faculty of Food Technology and Biotechnology
University of Zagreb
10000 Zagreb
Pierottijeva 6
Croatia

Nikola Sarapa
Department of Mathematics
University of Zagreb
10000 Zagreb
Bijenička 30
Croatia