Mathematical
Inequalities
& Applications
Volume 8, Number 4 (2005), 743–748

AN INEQUALITY FOR MIXED L^p –NORMS

HARI BERCOVICI AND DIRK VAN GUCHT

(*communicated by G. Bennett*)

Abstract. Consider a nonnegative measurable function f defined on $\Omega_1 \times \Omega_2$, where Ω_j is a probability space with probability measure μ_j. We prove the inequality

$$\left[\iint_{\Omega_1\times\Omega_2} f\,d\mu_1 d\mu_2\right]^p + \iint_{\Omega_1\times\Omega_2} f^p\,d\mu_1 d\mu_2 \geqslant \int_{\Omega_1}\left[\int_{\Omega_2} f\,d\mu_2\right]^p d\mu_1 + \int_{\Omega_2}\left[\int_{\Omega_1} f\,d\mu_1\right]^p d\mu_2$$

provided that $1 \leqslant p \leqslant 2$. The inequality fails in general if $p > 2$. It also fails if one of the measures μ_j has total mass greater than one. Curiously however, the inequality is true for all $p \in [1,\infty)$ if the measures μ_j are counting measures. This last fact follows from a subadditivity result proved by G. A. Raggio for p-entropies. Our inequality also has a formulation in terms of p-entropies.

1. Counting measure

Consider a partition $\mathscr{A} = \{A_1, A_2, \ldots, A_n\}$ of a probability space Ω, with probability measure μ. For $p > 0$ one defines the p-entropy of this partition by the formula

$$h_p(\mathscr{A}) = (p-1)^{-1}\left[1 - \sum_{j=1}^{n} \mu(A_j)^p\right].$$

This quantity was introduced by Z. Daróczy [1] (with a different normalization) and C. Tsallis[3], and it approximates the classical Shannon entropy as $p \to 1$. It was shown by G. A. Raggio [2] that

$$h_p(\mathscr{A} \vee \mathscr{B}) \leqslant h_p(\mathscr{A}) + h_p(\mathscr{B})$$

provided that $p \geqslant 1$; here $\mathscr{A} \vee B$ denotes the common refinement of $\mathscr{A}$ and $\mathscr{B}$. As noted in [2], this inequality does not generally hold for $p < 1$. Denoting $a_{ij} = \mu(A_i \cap B_j)$, the subadditivity of h_p can be rewritten as

$$1 + \sum_{i,j} a_{ij}^p \geqslant \sum_i \left[\sum_j a_{ij}\right]^p + \sum_j \left[\sum_i a_{ij}\right]^p$$

Mathematics subject classification (2000): 94A17, 26D17.

Key words and phrases: entropy, partition, refinement, probability measure.

The authors were supported by NSF grants DMS-0307166 and IIS-0082407.

Paper MIA-08-69

provided that $a_{ij} \geqslant 0$, $\sum a_{ij} = 1$, and $p \geqslant 1$. This inequality can be extended to arbitrary arrays of nonnegative numbers a_{ij} replacing each a_{mn} by $a_{mn}/\sum a_{ij}$:

$$\left[\sum_{i,j} a_{ij}\right]^p + \sum_{i,j} a_{ij}^p \geqslant \sum_i \left[\sum_j a_{ij}\right]^p + \sum_j \left[\sum_i a_{ij}\right]^p$$

for $p \geqslant 1$.

The main result of this section is that, at least for $p \in [1,2]$, the sums in this inequality can be replaced by averages.

THEOREM 1.1. *For any $N \times N$ matrix $[a_{ij}]_{i,j=1}^N$ with nonnegative entries, and every $p \in [1,2]$, we have*

$$\left[\frac{1}{N^2} \sum_{i,j} a_{ij}\right]^p + \frac{1}{N^2} \sum_{i,j} a_{ij}^p \geqslant \frac{1}{N} \sum_i \left[\frac{1}{N} \sum_j a_{ij}\right]^p + \frac{1}{N} \sum_j \left[\frac{1}{N} \sum_i a_{ij}\right]^p.$$

For the proof it will be enough to assume that $1 < p < 2$. Denoting by A the $N \times N$ matrix in the statement, consider the function

$$f_t(A) = t^2 \left[\sum_{i,j} a_{ij}\right]^p + \sum_{i,j} a_{ij}^p - t \left\{\sum_i \left[\sum_j a_{ij}\right]^p + \sum_j \left[\sum_i a_{ij}\right]^p\right\}$$

defined for real values of t. The proposition asserts that $f_t(A) \geqslant 0$ for $t = N^{1-p}$. The subadditivity inequality discussed above amounts to $f_1(A) \geqslant 0$.

LEMMA 1.2. *Fix an $N \times N$ matrix A with nonnegative entries, and $p \geqslant 1$. The function $f_t(A)$ attains its minimum in the interval $[N^{1-p}, 1]$.*

Proof. The minimum of f_t occurs at

$$t_0 = \frac{\sum_i \left[\sum_j a_{ij}\right]^p + \sum_j \left[\sum_i a_{ij}\right]^p}{2\left[\sum_{i,j} a_{ij}\right]^p}.$$

Since $p \geqslant 1$, we have

$$\sum_i \left[\sum_j a_{ij}\right]^p \leqslant \left[\sum_i \left[\sum_j a_{ij}\right]\right]^p.$$

This, and a similar inequality with the roles of i and j reversed, yields $t_0 \leqslant 1$. On the other hand, Hölder's inequality yields

$$\left[\sum_i \left[\sum_j a_{ij}\right]\right]^p \leqslant N^{p-1} \sum_i \left[\sum_j a_{ij}\right]^p.$$

This easily implies $t_0 \geqslant N^{1-p}$. □

Once the theorem above is proved, it follows easily that $f_t(A) \geqslant 0$ if $t \notin (N^{1-p}, 1)$ and $p \in (1,2)$. Note that the interval where f_t can be negative increases with N.

We need one more observation.

LEMMA 1.3. *For fixed real numbers* $p \in (1,2)$ *and* $c > 0$, *the function* $u(t) = (t^{p-1} + c)^{1/(p-1)}$ *is strictly concave on the interval* $(0,\infty)$.

Proof. $u''(t) = c(p-2)t^{p-3}(t^{p-1}+c)^{(1-2p)/(p-1)}$. □

We are now ready to prove the theorem.

Proof. We assume that $p \in (1,2)$, and proceed by induction noting that the result is trivial for $N = 1$. By the preceding remarks, the induction hypothesis can be reformulated as follows: if A is an $N \times N$ matrix with nonnegative entries and it has a zero row and a zero column, then $f_t(A) \geqslant 0$ for $t \notin ((N-1)^{1-p}, 1)$, and consequently $f_t(A) \geqslant 0$ for $t = N^{1-p}$. To prove the result we only need to show that, for $t = N^{1-p}$, the minimum values of the function $f_t(A)$ on the simplex defined by $a_{ij} \geqslant 0$, $\sum_{i,j=1}^{N} a_{ij} = 1$ are nonnegative. From this point on we will fix $t = N^{1-p}$, and assume that $B = [b_{ij}]_{i,j=1}^{N}$ is a minimum point for f_t on this simplex. If $b_{k\ell} \neq 0 \neq b_{mn}$, the matrix obtained replacing $b_{k\ell}$ by $b_{k\ell} + \varepsilon$ and b_{mn} by $b_{mn} - \varepsilon$ also belongs to the simplex if $|\varepsilon|$ is small enough. We deduce that

$$\frac{\partial f_t}{\partial a_{k\ell}}(B) = \frac{\partial f_t}{\partial a_{mn}}(B)$$

for such k, ℓ, m, n. The same argument applies if $b_{k\ell} = 0 \neq b_{mn}$, except that in this case ε must be restricted to positive values. This yields the inequalities

$$\frac{\partial f_t}{\partial a_{k\ell}}(B) \geqslant \frac{\partial f_t}{\partial a_{mn}}(B), \quad b_{mn} \neq 0.$$

Calculating the partial derivatives, and taking into account the fact that $\sum_{i,j} b_{ij} = 1$, this can be rewritten as

$$b_{k\ell}^{p-1} - tR_k^{p-1} - tC_\ell^{p-1} \geqslant b_{mn}^{p-1} - tR_m^{p-1} - tC_n^{p-1}$$

whenever $b_{mn} \neq 0$, with equality if $b_{k\ell} \neq 0$ as well, where we denote by R_i the sum of the entries in the ith row of B, and by C_j the sum of the entries in the jth column. We claim that the following pattern of zero entries is impossible: $b_{11} \neq 0 \neq b_{22}$, $b_{12} = b_{21} = 0$ Indeed, if that pattern occurs, we deduce

$$b_{11}^{p-1} - tR_1^{p-1} - tC_1^{p-1} = b_{22}^{p-1} - tR_2^{p-1} - tC_2^{p-1} \leqslant -tR_1^{p-1} - tC_2^{p-1}, -tR_2^{p-1} - tC_1^{p-1},$$

hence

$$b_{11}^{p-1} + b_{22}^{p-1} - tR_1^{p-1} - tC_1^{p-1} - tR_2^{p-1} - tC_2^{p-1} \leqslant -tR_1^{p-1} - tC_1^{p-1} - tR_2^{p-1} - tC_2^{p-1},$$

which is impossible. The claim just proved can be restated as follows: the index sets $S_i = \{j : b_{ij} \neq 0\}$ are totally ordered by inclusion. Without loss of generality, we may assume that S_1 is the largest of these sets. Similarly, we may assume that among the sets $S'_j = \{i : b_{ij} \neq 0\}$, S'_1 is the largest. If both S_1 and S'_1 have fewer than N elements then the last row and last column of B are zero, and the inequality $f_t(B) \geqslant 0$

follows from the induction hypothesis. We may therefore assume that one of these sets, say S_1, contains N elements. In other words, $b_{1j} \neq 0$ for $j = 1, 2, \ldots, N$. Let us consider the entries in the second row. We have

$$b_{1j}^{p-1} - tR_1^{p-1} - tC_j^{p-1} \leqslant b_{2j}^{p-1} - tR_2^{p-1} - tC_j^{p-1}$$

or, equivalently

$$b_{1j}^{p-1} - b_{2j}^{p-1} \leqslant tR_1^{p-1} - tR_2^{p-1},$$

with equality whenever $b_{2j} \neq 0$. If all the entries in the second row are equal to zero, we deduce that the entries in the first row are all equal; indeed, because of the special value of t, the inequalities above reduce to $b_{1j}^{p-1} \leqslant (R_1/N)^{p-1}$. If the second row contains both zero and nonzero entries, the quantity $c = tR_1^{p-1} - tR_2^{p-1}$ is strictly positive, and the above inequalities can be rewritten as

$$b_{1j} \leqslant u(b_{2j}),$$

where $u(x) = (c + x^{p-1})^{1/(p-1)}$. Since $t = N^{1-p}$ and the function u is strictly concave, we also have

$$\frac{R_1}{N} = u\left(\frac{R_2}{N}\right) > \frac{1}{N}\sum_{j=1}^{N} u(b_{2j}) \geqslant \frac{R_1}{N}$$

unless all the entries in the second row are equal, and therefore so are the entries in the first row. In summary, if the second row contains entries different from zero, then all of these entries are different from zero, and either they are constant, or they are identical with the ones in the first row. This is because $c > 0$ leads to the above contradiction, and so does $c < 0$ by symmetry.

These considerations can be applied to the other rows of B, and the conclusion is that this matrix must have constant columns or constant rows. If the columns are constant, we can now calculate

$$f_t(B) = N(tN^{p-1} - 1)\left(t\left[\sum_{j=1}^{N} b_{1j}\right]^p - \sum_{j=1}^{N} b_{1j}^p\right),$$

so that $f_t(B) = 0$ for $t = N^{1-p}$. $\square$

2. Probability measures

In this section we prove our main result.

THEOREM 2.1. *Consider probability spaces Ω_1, Ω_2 with probability measures μ_1, μ_2, and a nonnegative measurable function f defined on $\Omega_1 \times \Omega_2$. The following inequality holds*

$$\left[\iint_{\Omega_1\times\Omega_2} f\, d\mu_1 d\mu_2\right]^p + \iint_{\Omega_1\times\Omega_2} f^p\, d\mu_1 d\mu_2 \geqslant \int_{\Omega_1}\left[\int_{\Omega_2} f\, d\mu_2\right]^p d\mu_1 + \int_{\Omega_2}\left[\int_{\Omega_1} f\, d\mu_1\right]^p d\mu_2$$

for $p \in [1, 2]$.

Proof. Every nonnegative measurable function is the supremum of a sequence of simple functions, so the monotone convergence theorem allows us to assume that f is simple. A measurable subset of $\Omega_1 \times \Omega_2$ can be approximated arbitrarily well by a finite union of rectangles, and this allows us to further restrict ourselves to the case in which f is constant on sets of the form $A_i \times B_j$, where $A_1, A_2, \ldots, A_n$ form a partition of Ω_1, and $B_1, B_2, \ldots, B_m$ form a partition of Ω_2. If we denote by f_{ij} the value of this constant, and we set $\alpha_i = \mu_1(A_i), \beta_j = \mu_1(B_j)$, the inequality to be proved is

$$\left[\sum_{i,j} \alpha_i \beta_j f_{ij}\right]^p + \sum_{i,j} \alpha_i \beta_j f_{ij}^p \geqslant \sum_i \alpha_i \left[\sum_j \beta_j f_{ij}\right]^p + \sum_j \beta_j \left[\sum_i \alpha_i f_{ij}\right]^p$$

whenever $f_{ij} \geqslant 0$, $\alpha_i, \beta_j \geqslant 0$, and $\sum_i \alpha_i = \sum_j \beta_j = 1$. Clearly it suffice to prove this inequality under the additional assumption that α_i and β_j are rational numbers, so that $p_i = N\alpha_i, q_j = N\beta_j$ are integers for some integer N. Write now the set $\{1, 2, \ldots, N\}$ as the union of disjoint sets $C_1, C_2, \ldots, C_n$ (resp. $D_1, D_2, \ldots, D_m$) with $p_1, p_2, \ldots, p_n$ (resp. $q_1, q_2, \ldots, q_m$) elements, and set $a_{ij} = f_{k\ell}$ if $i \in C_k$ and $j \in D_\ell$. The inequality to be proved reduces then to the inequality in Theorem 1.1. $\square$

As mentioned earlier, the above theorem is not true if one of the measures μ_j has mass greater than one. The simplest example is obtained by considering a singleton Ω_1 with mass $1/2$, and a set Ω_2 with two elements and endowed with counting measure. If f is defined to be equal to 1 on $\Omega_1 \times \Omega_2$, the difference between the left- and right-hand sides of the inequality in the above theorem is

$$2 - 2^{p-1} - 2^{1-p} = -(2^{(p-1)/2} - 2^{(1-p)/2})^2,$$

and this is negative unless $p = 1$.

To see that the results in this paper do not hold for $p > 2$, consider an $N \times N$ matrix with entries equal to one in the first row and in the first column, and zero elsewhere. The difference between the left- and right-hand sides of the inequality in Theorem 1.1 is now

$$\left[\frac{2N-1}{N^2}\right]^p + \frac{2N-1}{N^2} - \frac{2}{N}\left[1 + (N-1)\left[\frac{1}{N}\right]^p\right] = \frac{1}{N^2}\left[\frac{(2N-1)^p}{N^{2p-2}} - 1 - \frac{2(N-1)}{N^{p-1}}\right].$$

Clearly this expression is negative for large values of N, provided that $p > 2$.

We would like to formulate the result of Theorem 1.1 in terms of p-entropy. The proof of the following result is a simple calculation.

PROPOSITION 2.2. *Given two partition $\mathscr{A}, B$ with N elements of a probability space, we have*

$$h_p(\mathscr{A} \vee \mathscr{B}) \leqslant N^{1-p}[h_p(\mathscr{A}) + h_p(\mathscr{B})] + \frac{1}{p-1}(1 - N^{1-p})^2$$

for all $p \in (1, 2]$.

We would like to remark that one can make a similar analysis for Raggio's inequality to conclude that the equality $h_p(\mathscr{A} \vee \mathscr{B}) = h_p(\mathscr{A}) + h_p(\mathscr{B})$ only occurs when one of the two partitions is trivial (i.e., it contains a set of probability one). Equality in the preceding proposition holds precisely when $\mu(A_i \cap B_j)$ depends only on i or only on j. If, for instance, $\mu(A_i \cap B_j)$ depends only on i, then clearly $\mu(B_j) = \sum_i \mu(A_i \cap B_j)$ does not depend on j, whence $\mu(B_j) = 1/N$ for all j. Thus $\mu(A_i) = N\mu(A_i \cap B_j)$ so that $\mu(A_i \cap B_j) = \mu(A_i)\mu(B_j)$, i.e., the partitions $\mathscr{A}$ and $\mathscr{B}$ are independent.

We conclude with a result obtained from Theorem 2.1 as $p \to 1$.

PROPOSITION 2.3. *Consider probability spaces Ω_1, Ω_2 with probability measures μ_1, μ_2, and a nonnegative measurable function f defined on $\Omega_1 \times \Omega_2$. Assume that $\iint_{\Omega_1 \times \Omega_2} f \, d\mu_1 d\mu_2 = 1$, and define $f_1(\omega_1) = \int_{\Omega_2} f(\omega_1, \omega_2)\, d\mu_2(\omega_2)$, $f_2(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2)\, d\mu_1(\omega_1)$ for $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$. We have then*

$$-\iint_{\Omega_1 \times \Omega_2} f \log f \, d\mu_1 d\mu_2 \leqslant -\int_{\Omega_1} f_1 \log f_1 \, d\mu_1 - \int_{\Omega_2} f_2 \log f_2 \, d\mu_2.$$

Proof. Standard approximation arguments allow us to assume that f and $1/f$ are bounded, so that

$$\lim_{p \downarrow 1} \frac{f - f^p}{p-1} = -f \log f$$

uniformly. The result follows because the inequality in Theorem 2.1 can be rewritten as

$$\iint_{\Omega_1 \times \Omega_2} \frac{f - f^p}{p-1} \, d\mu_1 d\mu_2 \leqslant \int_{\Omega_1} \frac{f_1 - f_1^p}{p-1} \, d\mu_1 + \int_{\Omega_2} \frac{f_2 - f_2^p}{p-1} \, d\mu_2.$$

□

REFERENCES

[1] Z. DARÓCZY, *Generalized information functions*, Information and Control, **16**, (1970), 36–51.

[2] G. A. RAGGIO, *Properties of q-entropies*, Journal of Mathematical Physics, **36**, (1995), 4785–4791.

[3] C. TSALLIS, *Possible generalizations of Bolzmann-Gibbs statistics*, Journal of Statistical Physics, **52**, (1988), 479–487.

(Received April 26, 2004)

Hari Bercovici
Department of Mathematics
Indiana University
Bloomington, IN 47405
USA
e-mail: bercovic@indiana.edu

Dirk Van Gucht
Department of Computer Science
Indiana University
Bloomington, IN 47405
USA
e-mail: vgucht@cs.indiana.edu