THE CONVEXITY METHOD OF PROVING MOMENT-TYPE INEQUALITIES

Villő Csiszár and Tamás F. Móri[†]

Department of Probability Theory and Statistics, Loránd Eötvös University, Budapest

Abstract. The power of the convexity method is demonstrated by proving Qi and Diaz–Metcalf type inequalities.

1. The convexity method in probability theory

The essence of this method is very simple. Let $(\Omega, \mathcal{F})$ be a measurable space, $\mathcal{P}$ a convex family of probability measures defined on it, and $f = (f_1, \dots, f_n) : \Omega \to \mathbb{R}^n$, a Borel measurable function. Let $\mathcal{Q} \subset \mathcal{P}$ be a subfamily, with $\mathcal{P}$ as its (weakly) closed convex hull: $\mathcal{P} = \operatorname{conv}(\mathcal{Q})$; one can think of the set of extremal points of $\mathcal{P}$, as the most important example (by the Krein–Milman theorem [9]). Then the n-dimensional set of f-moments $\mathcal{M}(\mathcal{P}) := \left\{\int_\Omega f\,dP : P \in \mathcal{P}\right\}$ coincides with the closed convex hull of the set $\mathcal{M}(\mathcal{Q})$. This makes it easier to characterize the moment set $\mathcal{M}(\mathcal{P})$, which can then be used for deriving moment-type inequalities in the following way. Let X be a random variable with distribution in $\mathcal{P}$, and suppose that the moments $Ef_1(X) = \mu_1, Ef_2(X) = \mu_2, \dots, Ef_{n-1}(X) = \mu_{n-1}$ are all known. We wish to find exact bounds for $Ef_n(X)$. Then clearly

$$\inf\left\{\mu_n : (\mu_1, \dots \mu_n) \in \mathcal{M}(\mathcal{P})\right\} \le Ef_n(X) \le \sup\left\{\mu_n : (\mu_1, \dots \mu_n) \in \mathcal{M}(\mathcal{P})\right\},$$

and these bounds cannot be improved: the whole interval is obtained as the distribution of X runs through $\mathcal{P}$. As a corollary we get that an inequality of the form

$$h_1\big(Ef_1(X), \dots, Ef_{n-1}(X)\big) \le Ef_n(X) \le h_2\big(Ef_1(X), \dots, Ef_{n-1}(X)\big),$$

where h_1 and h_2 are convex, resp. concave $(n-1)$-variate functions, holds for every X with distribution in $\mathcal{P}$, if and only if it holds in $\mathcal{Q}$.

The following convincing examples illustrate the applicability of the method.

2000 *Mathematics Subject Classification.* Primary 60E15, secondary 26D15.

Key words and phrases. Convexity, Diaz–Metcalf inequality, Qi inequality.

[†]Research supported by the Hungarian National Foundation for Scientific Research, Grant No. T-043037

1. Let $\mathcal{P}$ be the set of k-dimensional probability distributions concentrated onto the Borel set $H \subset \mathbb{R}^k$. Then $\mathcal{Q}$ consists of all degenerate distributions (Dirac measures) δ_x with $x \in H$.
2. Let $\mathcal{P}$ be the set of probability distributions with fixed expectation $\mu \in \mathbb{R}$. Then $\mathcal{Q} = \{P_{xy} : x < \mu < y\} \cup \{\delta_\mu\}$, where $P_{xy} = \dfrac{y-\mu}{y-x}\,\delta_x + \dfrac{\mu-x}{y-x}\,\delta_y$.
3. Let $\mathcal{P}$ consist of all symmetric unimodal distributions. By Khinchin's well-known result [2] $\mathcal{Q}$ is the family of symmetric uniform distributions $U[-x,x]$, $x \geq 0$.
4. If $\mathcal{P}$ is the family of unimodal distributions with mode μ, then
$$\mathcal{Q} = \{U[x,\mu] : x < \mu\} \cup \{U[\mu,x] : \mu \leq x\}.$$
5. Let $\mathcal{P}$ be the family of probability distributions concentrated onto the unit interval $[0,1]$, with concave density function $f : [0,1] \to \mathbb{R}$.
   Define $\varphi_0(x) = 2(1-x)$, $\varphi_1(x) = 2x$, $0 \leq x \leq 1$; and for $0 < t < 1$ let
$$\varphi_t(x) = \begin{cases} 2\,\dfrac{x}{t}, & \text{if } 0 \leq x < t, \\ 2\,\dfrac{1-x}{1-t}, & \text{if } t \leq x \leq 1. \end{cases}$$

These triangle functions are concave probability density functions, and every concave probability density function $f : [0,1] \to \mathbb{R}$ can be expressed as their mixture, that is, there exists a probability measure λ defined on the Borel subsets of $[0,1]$, such that
$$f(x) = \int_0^1 \varphi_t(x)\,\lambda(dt), \quad 0 < x < 1. \tag{1.1}$$

Namely,
$$f(x) = \frac{f(0+)}{2}\,\varphi_0(x) + \frac{f(1-)}{2}\,\varphi_1(x) + \int_{(0,1)} \varphi_t(x)\,\frac{t(1-t)}{2}\,d(-f'_+(t)), \tag{1.2}$$

where f'_+ denotes the right derivative of f (decreasing and right continuous). This formula can be checked easily by integrating in parts.

This method was applied e.g. in [4] to obtain sharp Bonferroni–Galambos type inequalities, and in [5], to show that the correlation coefficient of the sample mean and the sample variance is at most $\frac{15}{16}$, if the sample comes from a unimodal distribution.

Here we apply the method for proving and generalizing some moment-type inequalities studied recently by T. K. Pogány [6] and [7].

2. A problem of Qi revisited

Feng Qi [8] asked about conditions on f under which the inequality
$$\int_a^b [f(x)]^t dx \geq \left(\int_a^b f(x)dx\right)^{t-1}$$

holds for some $t > 1$. Different answers can be found in [10] and [11]. In [6] Pogány found conditions sufficient for the more general inequality

$$\int_a^b [f(x)]^\alpha dx \geq \left(\int_a^b f(x) dx \right)^\beta . \tag{2.1}$$

For another set of conditions see [10].

Consider the interval $[a, b]$, with the Borel sets and the uniform distribution (i.e., $(b-a)^{-1}$ times the Lebesgue measure) defined on it. This is a probability space and $X = f$ is a random variable. Inequality (2.1) can be rewritten as

$$E(X^\alpha) \geq C\,(EX)^\beta, \tag{2.2}$$

where $C = (b-a)^{\beta-1}$. We are going to demonstrate the power of the method by finding sharp conditions on the range of X, under which (2.2) or the converse inequality

$$E(X^\alpha) \leq C\,(EX)^\beta, \tag{2.3}$$

holds for fixed $0 < \beta < \alpha$.

Let $H = [x_1, x_2]$ be a (finite or infinite) interval, $0 \leq x_1 < x_2 \leq \infty$. Suppose $P(X \in H) = 1$, that is, $\mathcal{P}$ is the family of probability distributions P such that $P(H) = 1$. The set of extremal points is the set of degenerate (Dirac) distributions $\mathcal{Q} = \{\delta_x : x \in H\}$, hence the moment set $\mathcal{M} = \big\{(EX, E(X^\alpha)) : P(X \in H) = 1\big\}$ is equal to the convex hull $\operatorname{conv}\big\{(x, x^\alpha) : x \in H\big\}$. Now, (2.2) holds if and only if $\mathcal{M}$ lies entirely above the graph of the function $x \in H$, $x \mapsto c\,x^\beta$.

Definition. $H = [x_1, x_2]$ is called *a maximal interval of type I* (*type II*, resp.) if inequality (2.2) (inequality (2.3), resp.) holds for every random variable X such that $P(X \in H) = 1$, and no larger interval exists with the same property.

Let x_0 be the only positive solution of the equation $x^\alpha = Cx^\beta$, i.e., $x_0 = C^{\frac{1}{\alpha-\beta}}$.

Theorem 2.1. *Characterization of maximal intervals of type I.*

Suppose $\alpha < 1$. Then all maximal intervals $H = [x_1, x_2]$ of type I can be obtain in the following way: let $z \geq x_0$ be arbitrary, and let $x_1 < x_2$ be the positive solutions of the equation

$$x^\alpha = C(1-\beta)z^\beta + C\beta z^{\beta-1}x. \tag{2.4}$$

On the other hand, if $1 \leq \alpha$, the only maximal interval of type I is $H = [x_0, +\infty)$.

Theorem 2.2. *Characterization of maximal intervals of type II.*

Suppose $\beta \leq 1$. Then there is only one maximal interval of type II, namely $H = [0, x_0]$. On the other hand, if $1 < \beta$, then all maximal intervals of type II can be constructed by taking an arbitrary number z, $0 < z \leq x_0$, then $H = [x_1, x_2]$, where x_1 and x_2 are the positive solutions of (2.4).

Remarks. When $\alpha < 1$, solutions of (2.4) satisfy $x_0 \leq x_1 < x_2$, and every $x_1 \geq x_0$ can be obtained in this way. Similarly, when $1 < \beta$, solutions of (2.4) satisfy $0 < x_1 < x_2 \leq x_0$, and every x_2, $0 < x_2 \leq x_0$, can be obtained in this way.

The particular case $\alpha > 1$ and $C = (b-a)^{\beta-1}$ in our Theorem 2.1 yields Theorem 2.1 of [6]. The weighted inequality of [5, Theorem 4.1] can also be derived from our Theorem 2.1 in the same way, but for the basic probability space the uniform distribution on $[a,b]$ is to be replaced by the probability measure $\mu([a,b])^{-1}\mu(\cdot)$. Our Theorem 2.2 corresponds to Theorems 3.1 and 3.2 of [6]. There α was greater than 1, the random variable X was defined as a function on $[a,b]$ and it was supposed to be concave. It is easy to see that the best (i.e., largest) bound was attained as $q \to 1$ in Theorem 3.2 there. Namely, in our terminology Pogány obtained that the interval

$$\left[0,\ x_0\left(\frac{1+\alpha}{3^\alpha}\right)^{\frac{1}{\alpha-\beta}}\right]$$

was of type II. Note that the multiplier of x_0 is always less than 1; this yields a smaller interval than ours when $\beta \le 1$. When $\beta > 1$, the multiplier of x_0 is still less than 0.4, but, due to the concavity of the random variable, no lower bound is needed, unlike our Theorem 2.2 (for the concave case see our Theorem 2.3).

Proof of Theorem 2.1. Since the function $x \mapsto x^\alpha$ is either convex or concave, the moment set over H is just the area between the curve of the function and its chord. When $\alpha < 1$, the function is concave, hence the chord lies below the curve. Since $x^\alpha < Cx^\beta$ for $x < x_0$ and $x^\alpha > Cx^\beta$ for $x > x_0$, a maximal interval of type I has to be a subset of $[x_0, +\infty)$, and by the maximality, the chord must be tangential to the curve $x \mapsto Cx^\beta$. The straight line described on the right-hand side of (2.4) is just the tangent to the curve $x \mapsto Cx^\beta$ at point $x = z$, see Fig.1.

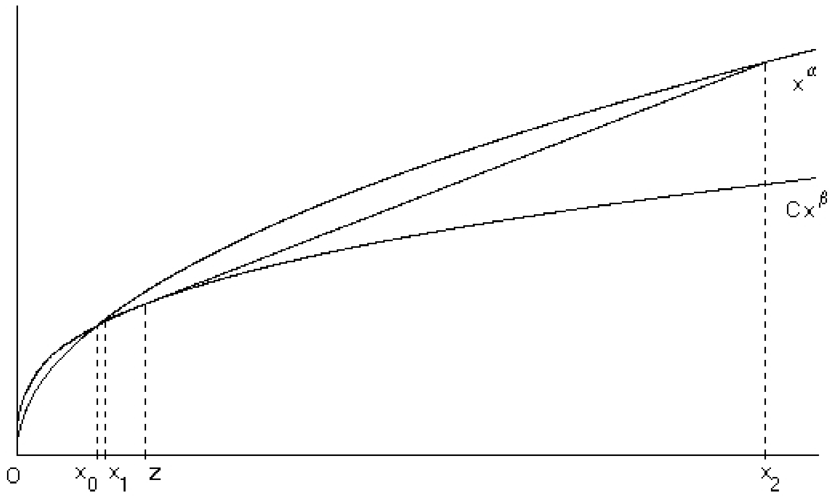

Figure 1

If $\alpha \ge 1$, the function $x \mapsto x^\alpha$ is convex, hence the moment sets (closed convex hulls) lie above the curve. Since $x^\alpha \ge Cx^\beta$ for $x \ge x_0$, $H = [x_0, +\infty)$ is a type I interval and it cannot be extended, because $x^\alpha < Cx^\beta$ for $0 < x < x_0$. □

Proof of Theorem 2.2. It follows the lines of the proof of Theorem 2.1 with obvious changes (see Fig.2), therefore all details are omitted. □

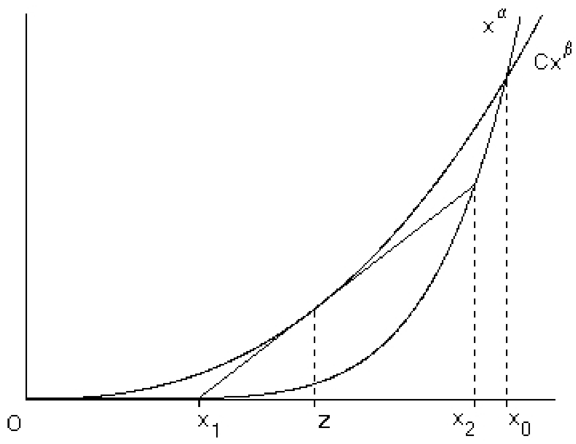

Figure 2

In the last part of this section we slightly improve [5, Theorem 3.1] by increasing the upper bound imposed on the concave function f. The proof is based on the same principles as the convexity method is.

Theorem 2.3. *Let f be nonnegative, concave and integrable on $[a,b]$, and $0<\beta<\alpha$. Suppose*

$$f(x) \le \Big(\frac{1+\alpha}{2^\alpha (b-a)^{1-\beta}}\Big)^{\frac{1}{\alpha-\beta}}, \quad x\in[a,b]. \tag{2.5}$$

Then

$$\int_a^b f^\alpha(x)\,dx \le \Big(\int_a^b f(x)\,dx\Big)^\beta. \tag{2.6}$$

Note that our constraint (2.5) is larger than those of Pogány [6], by a factor of $(1.5)^{\frac{\alpha}{\alpha-\beta}}$ at least.

Proof. We shall apply an integral representation, which is an immediate consequence of (1.1).

Define $\varphi_a(x) = \dfrac{b-x}{b-a}$, $\varphi_b(x) = \dfrac{x-a}{b-a}$, $a\le x\le b$; and for $a<t<b$ let

$$\varphi_t(x) = \begin{cases} \dfrac{x-a}{t-a}, & \text{if } a\le x<t,\\ \dfrac{b-x}{b-t}, & \text{if } t\le x\le b. \end{cases} \tag{2.7}$$

Then for every nonnegative concave function $f : [a,b] \to \mathbb{R}$ there exists a finite measure λ defined on the Borel subsets of $[a,b]$, such that

$$f(x) = \int_a^b \varphi_t(x)\,\lambda(dt), \quad a < x < b.$$

By using this representation we obtain

$$\int_a^b f(x)^\alpha\,dx = \int_a^b \left(\int_a^b \varphi_t(x)\,\lambda(dt)\right)^\alpha dx \le \lambda([a,b])^{\alpha-1} \int_a^b \int_a^b \varphi_t^\alpha(x)\,\lambda(dt)\,dx$$
$$= \lambda([a,b])^{\alpha-1} \int_a^b \int_a^b \varphi_t^\alpha(x)\,dx\,\lambda(dt).$$

Here $\int_a^b \varphi_t(x)^\alpha\,dx = \dfrac{b-a}{\alpha+1}$ for every $t \in [a,b]$, and $\lambda([a,b]) = \dfrac{2}{b-a} \int_a^b f(x)\,dx$ by Fubini's theorem. Thus we obtain that

$$\int_a^b f(x)^\alpha\,dx \le \frac{b-a}{\alpha+1}\,\lambda([a,b])^\alpha = \frac{2^\alpha}{(\alpha+1)(b-a)^{\alpha-1}} \left(\int_a^b f(x)\,dx\right)^\alpha$$
$$\le \frac{2^\alpha}{(\alpha+1)(b-a)^{\beta-1}} \left(\max_{[a,b]} f\right)^{\alpha-\beta} \left(\int_a^b f(x)\,dx\right)^\beta.$$

By applying (2.5) we immediately get (2.6). □

Remark. Let us emphasize separately the following byproduct which may be of independent interest.

Let $f : [a,b] \to \mathbb{R}$ be a concave probability density function, and $\alpha > 1$, then

$$\int_a^b f^\alpha(x)dx \le \frac{2^\alpha}{(\alpha+1)(b-a)^{\alpha-1}}. \tag{2.8}$$

Equality holds if and only if $f(x) = \dfrac{2}{b-a}\,\varphi_t(x)$ a.e. for some t, $a \le t \le b$, where φ_t is defined in (2.7).

3. Diaz–Metcalf type inequalities

When reformulated in a probabilistic setting, the Diaz–Metcalf inequality reads as follows (cf. [1], [7]).

Let ξ and η be bounded random variables,

$$P(m_1 \le \xi \le M_1) = 1, \quad P(m_2 \le \eta \le M_2) = 1, \tag{3.1}$$

where $m_2 > 0$. Then

$$m_2 M_2 E\xi^2 + m_1 M_1 E\eta^2 \le (m_1 m_2 + M_1 M_2) E\xi\eta, \tag{3.2}$$

and equality holds if and only if $P(m_2\xi = M_1\eta \text{ or } M_2\xi = m_1\eta) = 1$.

Let us generalize the problem slightly. Let a and b be fixed. We are interested in the smallest positive c such that

$$aE\xi^2 + bE\eta^2 \le cE\xi\eta \tag{3.3}$$

whenever condition (3.1) holds. Since our main goal is to demonstrate how the convexity method works, we will suppose that m_1, m_2, a, and b are all positive. This is just to avoid discussion; the same proof with minor changes could be repeated in all the other cases.

Now, $\mathcal{P}$ is the family of two dimensional probability measures concentrated onto the rectangle $[m_1, M_1] \times [m_2, M_2]$. The set of extremal points is the subfamily of degenerate distributions. Inequality (3.3) holds for every pair ξ, η with joint distribution in $\mathcal{P}$, if and only if

$$ax^2 + by^2 \le cxy \tag{3.4}$$

whenever $m_1 \le x \le M_1$, and $m_2 \le y \le M_2$.

Divide by xy in (3.4), and denote x/y by t to obtain

$$c = \max\Big\{ at + \frac{b}{t} : \frac{m_1}{M_2} \le t \le \frac{M_1}{m_2} \Big\}.$$

Since the function $t \mapsto at + b/t$ is convex, the maximum is attained at one of the endpoints of the interval. That is,

$$c = \max\Big\{ a\,\frac{m_1}{M_2} + b\,\frac{M_2}{m_1}, \quad a\,\frac{M_1}{m_2} + b\,\frac{m_2}{M_1} \Big\},$$

thus

$$c = \begin{cases} a\,\dfrac{m_1}{M_2} + b\,\dfrac{M_2}{m_1}, & \text{if} \quad \dfrac{a}{b} \le \dfrac{m_2 M_2}{m_1 M_1}, \\ a\,\dfrac{M_1}{m_2} + b\,\dfrac{m_2}{M_1}, & \text{if} \quad \dfrac{a}{b} > \dfrac{m_2 M_2}{m_1 M_1}. \end{cases} \tag{3.5}$$

As a particular case, we obtain (3.2).

At this point the following problem arises naturally. Let ξ and η be bounded random variables; suppose $E\xi^2$ and $E\eta^2$ are known. What is the range of $E\xi\eta$?

In order to find a straightforward generalization to the problem, let us denote ξ^2 and η^2 by X and Y resp.; then $\xi\eta = \sqrt{XY}$, a concave function of X and Y. Thus, let (X, Y) be a random vector, $P\big[(X, Y) \in \mathcal{D}\big] = 1$, where $\mathcal{D} = [a, A] \times [b, B]$, $a < A$, $b < B$. Let $\varphi : \mathcal{D} \to \mathbb{R}$ be a concave function; we are interested in the range of $E\varphi(X, Y)$ when EX and EY are fixed.

Theorem 3.1.

(*a*) *Suppose that* $\varphi(a, b) + \varphi(A, B) \ge \varphi(a, B) + \varphi(A, b)$.

If $(B - b)EX + (A - a)EY \le AB - ab$, *then*

$$pEX + qEY + r \le E\varphi(X, Y) \le \varphi(EX, EY),$$

where

$$p=\frac{\varphi(A,b)-\varphi(a,b)}{A-a},\qquad q=\frac{\varphi(a,B)-\varphi(a,b)}{B-b},$$
$$r=\frac{AB-ab}{(A-a)(B-b)}\,\varphi(a,b)-\frac{b}{B-b}\,\varphi(a,B)-\frac{a}{A-a}\,\varphi(A,b).$$

If $(B-b)EX+(A-a)EY>AB-ab,$ *then*

$$pEX+qEY+r\le E\varphi(X,Y)\le\varphi(EX,EY),$$

where

$$p=\frac{\varphi(A,B)-\varphi(a,B)}{A-a},\qquad q=\frac{\varphi(A,B)-\varphi(A,b)}{B-b},$$
$$r=\frac{A}{A-a}\,\varphi(a,B)+\frac{B}{B-b}\,\varphi(A,b)-\frac{AB-ab}{(A-a)(B-b)}\,\varphi(A,B).$$

(*b*) *Suppose that* $\varphi(a,b)+\varphi(A,B)<\varphi(a,B)+\varphi(A,b)$.
If $(B-b)EX-(A-a)EY\le aB-Ab,$ *then*

$$pEX+qEY+r\le E\varphi(X,Y)\le\varphi(EX,EY),$$

where

$$p=\frac{\varphi(A,B)-\varphi(a,B)}{A-a},\qquad q=\frac{\varphi(a,B)-\varphi(a,b)}{B-b},$$
$$r=\frac{B}{B-b}\,\varphi(a,b)+\frac{aB-Ab}{(A-a)(B-b)}\,\varphi(a,B)-\frac{a}{A-a}\,\varphi(A,B).$$

If $(B-b)EX-(A-a)EY>aB-Ab,$ *then*

$$pEX+qEY+r\le E\varphi(X,Y)\le\varphi(EX,EY),$$

where

$$p=\frac{\varphi(A,b)-\varphi(a,b)}{A-a},\qquad q=\frac{\varphi(A,B)-\varphi(A,b)}{B-b},$$
$$r=\frac{A}{A-a}\,\varphi(a,b)-\frac{aB-Ab}{(A-a)(B-b)}\,\varphi(A,b)-\frac{b}{B-b}\,\varphi(A,B).$$

Proof. By introducing

$$X'=\frac{X-a}{A-a},\quad Y'=\frac{Y-b}{B-b},\quad \varphi'(x,y)=\varphi\big(a+(A-a)x,b+(B-b)y\big)$$

one can reduce the problem to the case where $a=b=0$, $A=B=1$.

Again, $\mathcal{P}$ is the family of probability distributions concentrated onto the unit square $\mathcal{D}$, and the moment set $\mathcal{M} = \big\{(EX, EY, E\varphi(X,Y)) : P\big[(X,Y) \in \mathcal{D}\big] = 1\big\}$ is equal to the convex hull of $\big\{(x,y,z) : (x,y) \in \mathcal{D}, z = \varphi(x,y)\big\}$, the graph of φ.

Since φ is concave, the convex hull is bordered from above by the graph itself. Introduce the notation $\varphi_{ij} = \varphi(i,j)$, $i,j \in \{0,1\}$. Then the moment set is bordered from below by the lower faces of the tetrahedron with vertices $V_{ij} = (i,j,\varphi_{ij})$, $i,j \in \{0,1\}$. If edge $V_{00}V_{11}$ lies below edge $V_{01}V_{10}$, the lower faces are the triangles $V_{00}V_{10}V_{11}$ and $V_{00}V_{11}V_{01}$, otherwise the lower faces are triangles $V_{00}V_{10}V_{01}$ and $V_{01}V_{10}V_{11}$. The edges to be compared have their bisectors differ just in the third coordinate, thus the two cases above take place according as $\varphi_{00} + \varphi_{11}$ is less or greater than $\varphi_{01} + \varphi_{10}$.

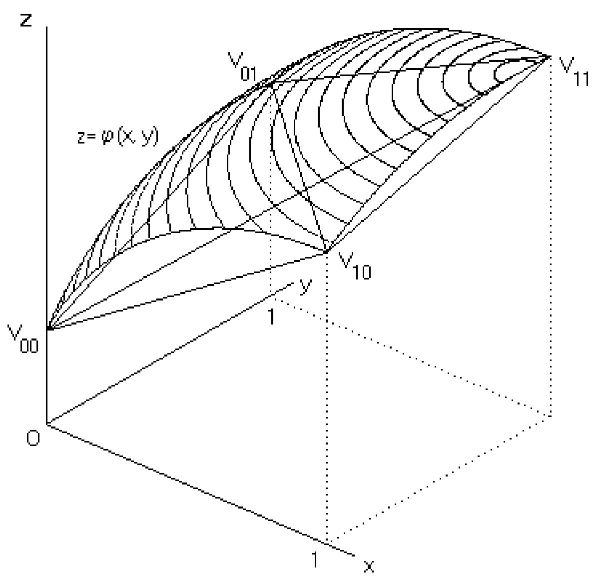

Figure 3

The equations of the planes through the faces of the tetrahedron are:

$$V_{00}V_{10}V_{11}: \begin{vmatrix} x & y & z & 1 \\ 0 & 0 & \varphi_{00} & 1 \\ 1 & 0 & \varphi_{10} & 1 \\ 1 & 1 & \varphi_{11} & 1 \end{vmatrix} = 0, \qquad V_{00}V_{11}V_{01}: \begin{vmatrix} x & y & z & 1 \\ 0 & 0 & \varphi_{00} & 1 \\ 1 & 1 & \varphi_{11} & 1 \\ 0 & 1 & \varphi_{01} & 1 \end{vmatrix} = 0,$$

$$V_{00}V_{10}V_{01}: \begin{vmatrix} x & y & z & 1 \\ 0 & 0 & \varphi_{00} & 1 \\ 1 & 0 & \varphi_{10} & 1 \\ 0 & 1 & \varphi_{01} & 1 \end{vmatrix} = 0, \qquad V_{01}V_{10}V_{11}: \begin{vmatrix} x & y & z & 1 \\ 0 & 1 & \varphi_{01} & 1 \\ 1 & 0 & \varphi_{10} & 1 \\ 1 & 1 & \varphi_{11} & 1 \end{vmatrix} = 0.$$

In each case the points (x,y,z) *above* the plane satisfy the equation with $\geq$ in place of the equality sign.

Suppose first that $\varphi_{00}+\varphi_{11}\ge\varphi_{01}+\varphi_{10}$. Then the lower bound for $E\varphi(X,Y)$ comes from the matrix corresponding to the face $V_{00}V_{10}V_{01}$ or $V_{01}V_{10}V_{11}$, according as $EX+EY\le 1$ or $EX+EY\ge 1$. Thus,

$$(\varphi_{00}-\varphi_{10})EX+(\varphi_{11}-\varphi_{00})EY+\varphi_{00}\le E\varphi(X,Y)\le\varphi(EX,EY),$$

if $EX\le EY$, and

$$(\varphi_{11}-\varphi_{01})EX+(\varphi_{11}-\varphi_{10})EY+(\varphi_{10}+\varphi_{01}-\varphi_{11})\le E\varphi(X,Y)\le\varphi(EX,EY),$$

if $EX>EY$. This proves part (a) of Theorem 3.1. Part (b) can be proved similarly, details are left to the reader. □

Remark. Conditions of equality in the upper bound depend on the function φ. If φ is strictly concave, equality holds if and only if both X and Y are degenerate.

In the lower bound equality holds if and only if the joint distribution of X and Y is concentrated onto the projection of the vertices of one of the lower faces. In details:

If $\varphi(a,b)+\varphi(A,B)=\varphi(a,B)+\varphi(A,b)$, then $P\big(X\in\{a,A\},Y\in\{b,B\}\big)=1$.

If $\varphi(a,b)+\varphi(A,B)>\varphi(a,B)+\varphi(A,b)$, then

– either $P\big[(X,Y)\in\{(a,b),(a,B),(A,B)\}\big]=1$,

– or $P\big[(X,Y)\in\{(a,b),(A,b),(A,B)\}\big]=1$.

If $\varphi(a,b)+\varphi(A,B)<\varphi(a,B)+\varphi(A,b)$, then

– either $P\big[(X,Y)\in\{a,b),(a,B),(A,b)\}\big]=1$,

– or $P\big[(X,Y)\in\{(A,B),(a,B),(A,b)\}\big]=1$.

Finally, let us turn back to the Diaz–Metcalf inequality. Let $0<\gamma\le 1/2$, then $\varphi(x,y)=(xy)^\gamma$ is concave for nonnegative x,y, and

$$\varphi(a,b)+\varphi(A,B)-\varphi(a,B)+\varphi(A,b)=(A^\gamma-a^\gamma)(B^\gamma-b^\gamma)\ge 0,$$

thus Theorem 3.1(a) applies. By substitution $\gamma=1/2$, $X=\xi^2$, $Y=\eta^2$, $a=m_1^2$, $A=M_1^2$, $b=m_2^2$, $B=M_2^2$, we obtain that the exact lower bound to $E\xi\eta$ in terms of $E\xi^2$ and $E\eta^2$ is of the form $pE\xi^2+qE\eta^2+r$, where

$$p=\frac{m_2}{m_1+M_1},\quad q=\frac{m_1}{m_2+M_2},\quad r=(M_1M_2-m_1m_2)pq,$$

if $\quad(M_2^2-m_2^2)E\xi^2+(M_1^2-m_1^2)E\eta^2\le M_1^2M_2^2-m_1^2m_2^2,\quad$ and

$$p=\frac{M_2}{m_1+M_1},\quad q=\frac{M_1}{m_2+M_2},\quad r=-(M_1M_2-m_1m_2)pq$$

otherwise. Note that both choices provide lower bounds, the exact bound is always the greater one. Multiply these inequalities by M_1M_2 and m_1m_2, respectively, then add them together to obtain the Diaz–Metcalf inequality. For equality both bounds have to be strict, that is, we are on the edge shared by the two lower faces. Thus the condition, necessary and sufficient for the equality, is

$$P\big[(X,Y)\in\{(m_1,M_2),(M_1,m_2)\}\big]=1.$$

References

1. J. B. Diaz and F. T. Metcalf, *Stronger forms of a class of inequalities of G. Pólya–G. Szegő and L. V. Kantorovich*, Bull. Amer. Math. Soc. **69** (1963), 415–418.
2. A. Ja. Hinčin, Izv. Naučn.-Issled. Inst. Mat. Meh. Tomsk Gos. Univ. **2** (1938), 1–7.
3. S. Mazouzi and F. Qi, *On an open problem regarding an integral inequality.*, J. Inequal. Pure Appl. Math. **4** (2003), no. 2, Art. 31 (electronic).
4. T. F. Móri and G. J. Székely, *A note on the background of several Bonferroni–Galambos type inequalities*, J. Appl. Probab. **22** (1985), 836–843.
5. T. F. Móri and G. J. Székely, *On the correlation between the sample mean and sample variance*, Second Conference of Program Designers. (A. Iványi, ed.), ELTE, Budapest, 1986, pp. 207–209.
6. T. K. Pogány, *On an open problem of F. Qi*, J. Inequal. Pure Appl. Math. **3** (2002), no. 4, Art. 54 (electronic).
7. T. K. Pogány, *New (probabilistic) derivation of Diaz–Metcalf and Pólya–Szegő inequalities and consequences*, manuscript (2003).
8. F. Qi, *Several integral inequalities*, J. Inequal. Pure Appl. Math. **1** (2000), no. 2, Art. 19 (electronic).
9. H. H. Schaefer, *Topological vector spaces*, Macmillan, New York, 1966.
10. N. Towghi, *Notes on integral inequalities*, RGMIA Res. Rep. Coll. **4** (2001), no. 2, Art. 12 (electronic), 277–278.
11. K. W. Yu and F. Qi, *A short note on an integral inequality*, RGMIA Res. Rep. Coll. **4** (2001), no. 1, Art. 4 (electronic), 23–25.

Department of Probability Theory and Statistics,
Loránd Eötvös University,
Pázmány P. s. 1/C,
H-1117 Budapest, Hungary

E-mail address: villo@ludens.elte.hu, moritamas@ludens.elte.hu