Delfos Project for mathematically interested youngsters to extend the proof to cover the remaining main facts in [AY]. The suggestion was taken up by the 18 years old second author who did the bulk of the mathematics of Section 2.

2. Proofs for the finite valued cases

Martin [M, Theorem 1.4] shows that a real symmetric matrix is copositive if and only if every of its principal submatrices passes his 'Test 1cop' ; in other words Martin establishes the following. Let $1_k = (1, 1, ..., 1)' \in \mathbb{R}^k$.

Proposition 1 (Martin). *A real symmetric matrix C is copositive if and only if each principal submatrix $\bar{C}$ of C either is non-invertible or it is invertible and $\bar{C}y = -1_k$, implies $y \not\geq 0$.* □

Let $s = \lceil b/a \rceil - 1$. It will be convenient to note that the matrix $C(\underline{a})$ referred to above in connection with Theorem 1 has the alternative definition

$$c_{ij} = \begin{cases} 2s & \text{if } |a_i - a_j| \leq a \\ -1 & \text{if } a < |a_i - a_j| \leq b \\ 0 & \text{if } b < |a_i - a_j|. \end{cases}$$

Example. For $\underline{a}_0 = (.3, .7, 1.2, 1.3, 2.0, 2.5, 2.8) \in \mathbb{R}^7$ and $a = 1, b = 2$, the associated matrix is

$$C(\underline{a}_0) = \begin{pmatrix} 2 & 2 & 2 & 2 & -1 & 0 & 0 \\ 2 & 2 & 2 & 2 & -1 & -1 & 0 \\ 2 & 2 & 2 & 2 & 2 & -1 & -1 \\ 2 & 2 & 2 & 2 & 2 & -1 & -1 \\ -1 & -1 & 2 & 2 & 2 & 2 & 2 \\ 0 & -1 & -1 & -1 & 2 & 2 & 2 \\ 0 & 0 & -1 & -1 & 2 & 2 & 2 \end{pmatrix}.$$

We can now prove the following.

Proposition 2. *The matrix $C(\underline{a})$ defined in* (1) *is copositive.*

Proof. The proof is by induction on the number m of points a_i on the real line. The base case $m = 1$ is trivial since then $a = (a_1)$ and the matrix $C = C(\underline{a}) = [2s]$. In the case $n = 2$, $\underline{a} = (a_1, a_2)$ and the matrix $C = C(\underline{a})$ has one of the forms $\begin{pmatrix} 2s & 0 \\ 0 & 2s \end{pmatrix}$, $\begin{pmatrix} 2s & -1 \\ -1 & 2s \end{pmatrix}$, or $\begin{pmatrix} 2s & 2s \\ 2s & 2s \end{pmatrix}$ according to if the cases $|a_1 - a_2| > b$, $a < |a_1 - a_2| \leq b$, or $|a_1 - a_2| \leq a$ hold. The associated quadratic forms are $2sx^2 + 2sy^2$, $2sx^2 - 2xy + 2sy^2$ and $2sx^2 + 4sxy + 2sy^2$. As $s \geq 3$, they all give nonnegative values if $(x, y) > (0, 0)$ (in fact these

Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 14–05

THE 123 THEOREM OF PROBABILITY THEORY AND COPOSITIVE MATRICES

ALEXANDER KOVAČEC AND MIGUEL M. R. MOREIRA

Abstract: Alon and Yuster give for independent identically distributed real or vector valued random variables X, Y combinatorially proved estimates of the form $\text{Prob}(\|X - Y\| \leq b) \leq c\,\text{Prob}(\|X - Y\| \leq a)$. We derive these using copositive matrices instead. We also formulate a version of this inequality as an integral inequality for monotone functions.

Keywords: probabilistic inequalities, copositivity, integral inequality.
AMS Subject Classification (2000): 60E15, 26D10.

1. Introduction

In [AY], Alon and Yuster prove theorems 1 and 2 below.

Theorem 1 (generalized 123-theorem). *Let $b > a > 0$ be two reals and let X and Y be independent identically distributed (iid) real random variables. Then*

$$\text{Prob}(|X - Y| \leq b) \leq (2\lceil b/a \rceil - 1)\text{Prob}(|X - Y| \leq a)$$

and the multiplicative constant at the right cannot be improved.

In the case $a = 1, b = 2$, the inequality takes the form

$$\text{Prob}(|X - Y| \leq 2) \leq 3\,\text{Prob}(|X - Y| \leq 1),$$

explaining the name of the theorem found in response to a question of G. A. Margoulis. He had conjectured an inequality of this type for some constant in place of the 3 at the right hand side.

Theorem 2 (which is [AY, Corollary 3.3]) is a version of Theorem 1 for higher dimensional Euclidean spaces. For fixed dimension $d \geq 2$ endow $\mathbb{R}^d$ with the Euclidean norm. Define a (b, n)-*configuration* to be a pair (B, F) consisting of a closed Euclidean ball $B = B(a_0, b)$ of radius $b > 1$ centered at $a_0 \in \mathbb{R}^d$ and a set F of n points in B containing a_0 and having $\binom{n}{2}$ mutual distances > 1.

Received January 31, 2014.

Clearly the non-existence of such a configuration happens for large enough n, although only for special cases it is known what the smallest such n is as a function of dimension d and radius b.

Theorem 2. *Assume $n \in \mathbb{Z}_{\geq 2}$ and $b \in \mathbb{R}_{>1}$ such that there exists no $(b, n+1)$-configuration in $\mathbb{R}^d$, $d \geq 2$. Then for any two $\mathbb{R}^d$-valued iid random variables X, Y, there holds*

$$\mathrm{Prob}(\|X - Y\| \leq b) \leq n\,\mathrm{Prob}(\|X - Y\| \leq 1).$$

The case that at the right we have $\|X - Y\| \leq a$ is dealt with by applying the theorem with b/a in place of b.

In [AY] actually it is shown via an additional argument that the inequality of Theorem 1 is strict and a simple probabilistic argument also shows that $2\lceil b/a \rceil - 1$ is the best constant. Concerning Theorem 2 it is shown that if there is a lattice of minimum distance 1 in $\mathbb{R}^d$ such that n points of it are contained in a ball of radius b then n is the best constant. The famous Newton - Gregory debate of 1694 concerning the maximum number of points that can be placed on the unit sphere so that any two points have distance at least 1, was decided by researchers in the nineteenth century in favor of Newton's conjecture that the number is 12. This together with the existence of a suitable lattice yields in case of dimension $d = 3$ that there exists an $\varepsilon > 0$ so that for all $1 < b < 1 + \varepsilon$ we have $n = 13$ as the best constant. This is one of the few cases in which in Theorem 2 one knows the best possible n; some more are related in [AY]. Concerning the quest for best possible constants we have nothing to add in this paper.

The proofs in [AY] are combinatorial. Our purpose here is to give for the case X, Y are finite valued hopefully attractive alternative proofs based on the theory of real symmetric matrices C with the property that for all real columns $x > 0$ (i.e. $x \geq 0$ entrywise and $x \neq 0$) of appropriate size there holds $x'Cx \geq 0$, where $'$ denotes transposition. Such a matrix C is called *copositive*; if the hypothesis implies even $x'Cx > 0$, it is *strictly copositive*.

To see the connection between probability theory and copositive matrices, assume X, Y are iid random variables assuming finitely many real values $a_1, a_2, ..., a_m$ with respective probabilities $\xi_1, \xi_2, ..., \xi_m > 0$. Define $\chi(P)$ to be 1 or 0 according to whether property P holds or not. Then for any real r,

$$\begin{aligned}
\mathrm{Prob}(|X - Y| \leq r) &\overset{1}{=} \sum_{i,j=1}^{m} \mathrm{Prob}(|a_i - a_j| \leq r, X = a_i, Y = a_j) \\
&\overset{2}{=} \sum_{i,j=1}^{m} \chi(|a_i - a_j| \leq r)\mathrm{Prob}(X = a_i, Y = a_j) \\
&\overset{3}{=} \sum_{i,j=1}^{m} \chi(|a_i - a_j| \leq r)\xi_i\xi_j \\
&= \xi'(\chi(|a_i - a_j| \leq r))\xi,
\end{aligned}$$

where $\xi = (\xi_1, ..., \xi_m)'$ is the m-column of probabilities. Here '$\overset{1}{=}$' holds since the events $(X, Y) = (a_i, a_j)$, $i, j = 1, ..., m$, are mutually exclusive and cover all possibilities; to see '$\overset{2}{=}$' fix temporarily i, j. If $|a_i - a_j| > r$, then the probability in line 1 is evidently zero; if $|a_i - a_j| \leq r$, then the event in line 1 happens if and only if $(X, Y) = (a_i, a_j)$ happens; finally '$\overset{3}{=}$' follows from the definition of independence of the random variables and their identical distribution. Almost all probability theoretic and measure theoretic material we shall later need can be found in Loève's or Bauer's books [L], [Ba].

Evidently an analogous computation holds in the vector valued case. It then follows that the inequality of Theorem 1 can for the case that X, Y are iid random variables assuming values only in $\{a_1, ..., a_m\} \subseteq \mathbb{R}$ be established by showing that the matrix $C = C(\underline{a}) = C(a_1, \ldots, a_m) = (c_{ij})$ given by

$$c_{ij} = (2\lceil b/a \rceil - 1)\chi(|a_i - a_j| \leq a) - \chi(|a_i - a_j| \leq b) \tag{1}$$

is copositive. We prove Theorem 2 similarly showing that if X, Y assume values only in $\{a_1, ..., a_m\} \subseteq \mathbb{R}^d$, then the matrix $C(\underline{a}) = C(a_1, ..., a_m)$ defined by

$$c_{ij} = (n\chi(\|a_i - a_j\| \leq 1) - \chi(\|a_i - a_j\| \leq b) \tag{2}$$

is strictly copositive.

These proofs are given in Section 2 based on characterizations of (strict) copositivity given by Cottle, Habetler and Lemke [CHL] and Martin [M]. In Section 3 we give the arguments that extend the inequalities to arbitrary iid real or vector valued random variables. In Section 4 we derive from Theorem 1 an integral inequality for increasing bounded functions on $\mathbb{R}$ of a possibly novel type.

A proof of the original 123 theorem via the theory of copositive matrices is due to the first author who suggested to the students of Coimbra University's

have $p_{i_{\bar{k}}} \neq \mu = p_{i_0}$, and hence $p_{i_{\bar{k}}} < p_{i_0}$, contradicting that by the previous paragraph $p_{i_{\bar{k}}} = p_{i_0}$. □

We have herewith proved for $b > a > 0$ and iid real random variables X, Y that take only finitely many values, the inequality

$$\mathrm{Prob}(|X - Y| \leq b) \leq (2\lceil b/a \rceil - 1)\mathrm{Prob}(|X - Y| \leq a),$$

and similarly for the case $d \geq 2$ and $\mathbb{R}^d$-valued iid random variables X, Y that take only finitely many values, the inequality

$$\mathrm{Prob}(\|X - Y\| \leq b) \leq n\mathrm{Prob}(|X - Y| \leq 1),$$

provided Euclidean d-space does not permit a $(b, n+1)$-configuration.

3. Extension to arbitrary random vectors

In this section the results obtained for finitely valued random vectors are extended to arbitrary real valued random vectors. We prove only the extension of Theorem 2; it will be clear that the Theorem 1 can be extended similarly.

To the extent that our considerations in Section 2 where probabilistic we used only the elementary theory devoid of measure theoretic and limit considerations. To treat the general case we have to go back to the notions of more advanced probability theory and actually the discussion is essentially measure theoretic. All we need can be found in [L]. For the convenience of the reader who may have these notions not present we sometimes give exact page references to that book in forms like 'p123c-4' meaning 'page 123, about 4cm from last text row'.

A triple $(\Omega, \mathcal{A}, P)$ composed from a space Ω, a σ-algebra $\mathcal{A}$ (called σ-field in Loève) of subsets of Ω, called events and a probability measure P on Ω assigning a real value $P(A)$ to each $A \in \mathcal{A}$ is a probability space, p152c1. On the reals one defines as the standard the Borel σ-algebra $\mathcal{B}$, on $\mathbb{R}^d$ the σ-algebra $\mathcal{B}^d$. A random variable on Ω is simply a function $X : \Omega \to \mathbb{R}$ which is measurable, p152c10; see p107 for different characterizations of measurability. For $A \subseteq \Omega$ let $1_A : \Omega \to \{0, 1\} \subseteq \mathbb{R}$ be the indicator function o A. As done in Loève, p106c-1, if X is a real valued random variable, and $S \subseteq \mathbb{R}$, write $[X \in S]$ for $\{\omega \in \Omega : X(\omega) \in S\}$. Below a similar notation for vector valued random variables on Ω is used.

Given $j \in \mathbb{Z}_{\geq 1}$ define the function given on p108,

$$E_j^X := -j1_{[X<j]} + \sum_{k=-j2^j+1}^{j2^j} \frac{k-1}{2^j} 1_{[\frac{k-1}{2^j} \leq X < \frac{k}{2^j}]} + j1_{[X \geq j]}.$$

forms are even positive semidefinite). Assume now copositivity of matrices C already established for the case that $\underline{a}$ consists of up to $m-1$ reals. Fix an $\underline{a} = (a_1, ..., a_m)$ consisting of m real entries. Any principal submatrix of C, say $\bar{C} = C[i_1, ..., i_k]$ (made from rows and columns with indices i_ν as indicated), is actually the matrix $C(\tilde{\underline{a}})$ associated to $\tilde{\underline{a}} = (a_{i_1}, ..., a_{i_k})$. By induction assumption, if this matrix is not equal to C, i.e. if $k < m$, it is copositive. Thus by Martin's criterion given in Proposition 1, it is sufficient to show that the equation $Cy = C(\underline{a})y = -1_m$ is not solvable with a vector $y > 0$. Assume otherwise and assume that $y > 0$ solves $Cy = -1_m$.

For each row index i, define $P_i = \{j : c_{ij} = 2s\}$ and $N_i = \{j : c_{ij} = -1\}$ as the indices of positive and negative entries, respectively, of the matrix C. Now let $p_i = \sum_{j \in P_i} y_j$ and let i_0 be a row index so that $p_{i_0} = \max_i p_i$. By hypothesis we have $-1 = \sum_j c_{i_0 j} y_j = \sum_{j \in P_{i_0}} 2sy_j - \sum_{j \in N_{i_0}} y_j$, and hence $\sum_{j \in N_{i_0}} y_j > 2sp_{i_0}$.

Clearly $N_{i_0} = \{j : a_j \in [a_{i_0} - b, a_{i_0} - a[\cup]a_{i_0} + a, a_{i_0} + b]\}$. So N_{i_0} is partitioned into $2s$ sets, all defined by translations of two intervals,

$$\{j : a_j \in -a - t\tfrac{b-a}{s} + [a_{i_0}, a_{i_0} + \tfrac{b-a}{s}[\}, \quad \{j : a_j \in a + t\tfrac{b-a}{s} +]a_{i_0} - \tfrac{b-a}{s}, a_{i_0}]\};$$

$t = 1, 2, 3, ..., s$. Each of the specified intervals has length $\frac{b-a}{s} \leq \frac{b-a}{\frac{b}{a}-1} = a$. Thus for each set N' of the partition and any $l, l' \in N'$ we have $|a_l - a_{l'}| \leq a$. Hence $N' \subseteq P_l$ and thus $\sum_{j \in N'} y_j \leq p_l \leq p_{i_0}$. Consequently, denoting by '$\sum_{N'}$' the sum over the sets of the partition, we get $\sum_{j \in N_{i_0}} y_j = \sum_{N'} \sum_{j \in N'} y_j \leq 2sp_{i_0}$, contradicting the strict inequality above. □

This establishes Theorem 1 for random variables that assume only finitely many values. We now take up the vector valued case.

According to [CHL, Theorem 3.2], for a real symmetric $m \times m$ matrix M that itself is not strictly copositive but all whose principal $(m-1) \times (m-1)$ submatrices are strictly copositive (that is, M is strictly copositive of order $m-1$ but not of order m), there exist $\lambda \in \mathbb{R}_{\leq 0}$, and $y \in \mathbb{R}^m_{\geq 0} - \{0\}$, so that $My = \lambda y$.

From this we find the following criterion for strict copositivity.

Lemma 3. *A real symmetric matrix C is strictly copositive if and only if for every principal submatrix $\bar{C}$, $y > 0$ implies $\bar{C}y \not\leq 0$.*

Proof. $\Rightarrow$: Assume there exists $y > 0$ so that $\bar{C}y \leq 0$. Construct the vector $\dot{y}$ by putting $\dot{y}_i = y_l$ if i is the index of the lth column of $\bar{C}$ as a submatrix of C; put $\dot{y}_i = 0$ otherwise. Then $\dot{y} > 0$ while $\dot{y}' C \dot{y} = y' \bar{C} y \leq 0$.

This contradicts strict copositivity of C. $\Leftarrow$: Assume C is not strictly copositive. Then there exists a principal submatrix $\bar{C}$ of order $k \geq 1$ so that $\bar{C}$ is strictly copositive of order $k-1$ but not of order k. So by the fact in [CHL] cited, there exists a real $\lambda \leq 0$ associated to $y > 0$ so that $\bar{C}y = \lambda y$. Hence $\bar{C}y \leq 0$, a contradiction. □

For proving an analogue to Proposition 2 for the higher dimensional case, we need a lemma.

Lemma 4. *Let $b > 1$ be a real and assume that there does not exist a $(b, n+1)$-configuration in $\mathbb{R}^d$, $d \geq 2$. Then:*
a. Given a ball $B = B(a_0, b)$ and a set $\mathcal{P}$ of points so that $a_0 \in \mathcal{P} \subseteq B$, there exist n (not necessarily distinct) points $a_0, a_1, a_2, ..., a_{n-1} \in \mathcal{P}$ so that any two distinct of these points have distance > 1, and every point $x \in \mathcal{P}$ is near *one of them: i.e. there exists i, $0 \leq i < n$, so that $\|x - a_i\| \leq 1$.*
b. If the points $a_0, a_1, ..., a_{n-1}$ referred in (a) are well distributed*, that is if $\|a_i - a_j\| > 1$ for $0 \leq i < j \leq n-1$, then the first coordinate of one of the points $a_1, ..., a_{n-1}$ is smaller than the first coordinate of $a_0 = \text{center}(B)$.*

Proof. a. Let i_0 be the largest number i with the property that there exist points $a_0, a_1, ..., a_i$, all in $\mathcal{P}$, that have mutual distance > 1. Then necessarily $i_0 \leq n-1$, for otherwise we had found a set of $n+1$ distinct points within B, centered at one of them, that have mutual distances > 1, contradicting the general hypothesis of the lemma. If $i_0 < n-1$, then define points $a_{i_0+1}, a_{i_0+2}, ..., a_{n-1}$ all to be equal to one of the points a_i, $i \leq i_0$. Then if $x \in \mathcal{P}$ is any point, the definition of i_0 and the a_i implies that there exists an $i \in \{0, 1, 2, ..., n-1\}$ so that $\|x - a_i\| \leq 1$.

b. Assume without loss of generality that $\text{center}(B) = (0, 0, \ldots, 0) = a_0$. If the claim is false, then distinct points $a_1, ..., a_{n-1}$, chosen to satisfy the hypothesis of (b), have nonnegative first coordinate. Select a positive $\varepsilon < b-1$ and define $q = (-(1+\varepsilon), 0, ..., 0)$. By definition of the Euclidean norm it is clear that $\|a_i - q\| > 1$, for $i = 1, ..., n-1$. Then $\{a_0, \cdots, a_{n-1}, q\}$ is a set of $n+1$ points of the type that by the hypothesis of the lemma is forbidden. □

To prove Theorem 2 note that the $m \times m$ matrix C defined in (2) in connection with it has the alternative definition

$$c_{ij} = \begin{cases} (n-1) & \text{if } \|a_i - a_j\| \leq 1 \\ -1 & \text{if } 1 < \|a_i - a_j\| \leq b \\ 0 & \text{if } b < \|a_i - a_j\| \end{cases} .$$

Proposition 5. *This matrix C is strictly copositive.*

Proof. We use induction on m. The cases $m = 1, 2$ are clear. We assume the proposition proved for the matrix associated to any set of $< m$ points. We have to show that for $y > 0$ it is impossible that $Cy \leq 0$. Suppose not and take a specific $0 \neq y = (y_1, ..., y_m)^{\mathsf{T}} \geq 0$ so that $Cy \leq 0$. Note that in such a supposed y *every* entry must be positive. For if, say, $y_l = 0$ then remove column and rows l from the symmetric matrix C. We get an $(m-1) \times (m-1)$ matrix $\bar{C}$ subordinated to the $m-1$ points $a_1, ..., a_{l-1}, a_{l+1}, ..., a_m$ such that $\bar{y} = (y_1, ..., y_{l-1}, y_{l+1}, ..., y_m)'$ is so that $\bar{C}\bar{y} \leq 0$, $\bar{y} > 0$. But by hypothesis of induction, this is impossible. Thus the function $2^{\{1,...,m\}} \ni I \mapsto \sum_{j \in I} y_j$ defines a positive measure on $\{1, 2, ..., m\}$ whose only null set is the empty set.

Let $P_i = \{j : c_{ij} = n-1\}$ and $N_i = \{j : c_{ij} = -1\}$ be the sets of column indices of line i, where c_{ij} is positive or negative, respectively. Let $p_i = \sum_{j \in P_i} y_j$ and let $\mu = \max_i p_i$. Among all i for which $p_i = \mu$, take i_0 to be an i such that a_i has minimal first coordinate. By hypothesis, $0 \geq \sum_j c_{i_0 j} y_j = \sum_{j \in P_{i_0}} (n-1) y_j - \sum_{j \in N_{i_0}} y_j$, hence $\sum_{j \in N_{i_0}} y_j \geq (n-1) p_{i_0}$.

The set $\mathcal{P} = \{a_j : j \in N_{i_0}\} \cup \{a_{i_0}\}$ of points is contained in the ball $B = B(a_{i_0}, b)$ which is centered at one of them. By Lemma 4a, we can find n not necessarily distinct points in $\mathcal{P}$ so that any two distinct ones have distance > 1 and so that each point in $\mathcal{P} \setminus \{a_{i_0}\} = \{a_j : j \in N_{i_0}\} \subset \mathcal{P}$ has distance ≤ 1 to one of them; that is, 'is near' to one of them.

Note $i_0 \notin N_{i_0}$. Let $a_j \in \mathcal{P}$, with $j \in N_{i_0}$. Then $\|a_j - a_{i_0}\| > 1$ which means that a_j is *not* near to a_{i_0}; the only point in $\mathcal{P}$ near to a_{i_0} is a_{i_0} itself. So we find not necessarily distinct indices $i_1, i_2, ..., i_{n-1} \in N_{i_0}$ so that for every $j \in N_{i_0}$, there exists a $\nu \in \{1, ..., n-1\}$ such that $\|a_j - a_{i_\nu}\| \leq 1$, that is $j \in P_{i_\nu}$. In other words, $N_{i_0} \subseteq P_{i_1} \cup \cdots \cup P_{i_{n-1}}$. But then, by the definition of i_0, and the inequality above,

$$(n-1)p_{i_0} \leq \sum_{j \in N_{i_0}} y_j \leq \sum_{k=1}^{n-1} \sum_{j \in P_{i_k}} y_j = \sum_{k=1}^{n-1} p_{i_k} \leq \sum_{k=1}^{n-1} p_{i_0} = (n-1)p_{i_0}.$$

So it follows that these inequalities have to be equalities. Hence $p_{i_k} = p_{i_0}$, for $k = 1, ..., n-1$. Furthermore having equality in the second inequality of the above chain implies by virtue of that all y_j are positive, that for $1 \leq k < k' \leq n-1$ we have $P_{i_k} \cap P_{i_{k'}} = \emptyset$, which in turn says $\|a_{i_k} - a_{i_{k'}}\| > 1$.

Now by Lemma 4b, we have that one of these points, say $a_{i_{\bar{k}}}$, has its first coordinate less than the first coordinate of a_{i_0}. By definition of μ and i_0 we

Rua Luís de Camões, Nr. 102, 1300-360, Lisboa, PORTUGAL
E-mail address: mmrm1996@hotmail.com

Note that for every $A \subseteq \mathbb{R}$ we have $1_{[X\in A]}(\omega) = (1_A \circ X)(\omega) = 1_A(X(\omega))$, so that E_j^X is a Borel function of the measurable function X in the sense of p111c6. Clearly E_j^X is finitely valued.

Now let X be an $\mathbb{R}^d$-valued random *vector*; that is assume $X = (X_1, ..., X_d)$, where each component function X_i is a real random variable, pp110c-4 and 152c-2. Then to X and j associate the function $E_j^X = (E_j^{X_1}, ..., E_j^{X_d})$. Again E_j^X is a finitely valued Borel function *of* X. To see this note that the event $[\alpha \leq X_i < \beta]$ could be written as $[X \in (\mathbb{R}^{i-1} \times [\alpha, \beta[\times\mathbb{R}^{d-i})]$. By a general theorem for Borel functions of distributions, p168c7 and p171c2, the distribution of E_j^X, that is, the function $\mathcal{A} \ni S \mapsto P(E_j^X \in S)$, depends only on the distribution of X.

Finally let X and Y be any two independent, identically distributed $\mathbb{R}^d$ valued random variables. By the made remarks E_j^X, E_j^Y are identically distributed. Random vectors E_j^X and E_j^Y are also independent since they are Borel functions of independent random vectors X, Y; see, p236c6. Therefore, since E_j^X, E_j^Y are finitely valued, Theorem 2 tells us that under its hypothesis the inequality

$$P(\|E_j^X - E_j^Y\| \leq b) \leq nP(\|E_j^X - E_j^Y\| \leq 1)$$

is valid for all $j = 1, 2, 3, ...$. Here we abbreviated the previous 'Prob' to P.

Now, by construction, p108, we know pointwise convergences $E_j^X \to X$, and $E_j^Y \to Y$, from the corresponding facts for the coordinates. Since the norm $\|.\|$ is continuous real valued on $\mathbb{R}^d$ it is Borel and we have the pointwise convergence of the real valued nonnegative random variables $E_j := \|E_j^X - E_j^Y\|$ towards the nonnegative variable $E := \|X - Y\|$. Then the functions $\mathbb{R} \ni r \mapsto P(E_j \leq r) =: F_{E_j}(r)$ and $\mathbb{R} \ni r \mapsto P(E \leq r) =: F_E(r)$ are the distribution functions of the random variables E_j and E. According to p170c2, we conclude from the convergence $E_j \to E$ that $F_{E_j} \to F_E$ on the set $C(F_E)$ of continuity points of F_E. So if b and 1 are in $C(F_E)$, the inequalities above, saying $F_{E_j}(b) \leq nF_{E_j}(1)$ for all j tell us directly that $F_E(b) \leq nF_E(1)$. (Loève defines distribution functions for a random variable X using the definition $F_X(x) = P(X < x)$, but his proof can be easily adapted to our definition $F_X(x) = P(X \leq x)$, chosen to be closer to [AY]. Loèves definition leads to left continuous functions, the one here to right continuous ones.)

In the case that $1 \notin C(F_E)$ or $b \notin C(F_E)$, proceed as follows to reach the same conclusion. For given j select an $\varepsilon'_j > 0$ so that $1+\varepsilon'_j \in C(F_E)$. Since functions F_{E_j} and F_E are evidently nondecreasing they have at most countably many discontinuities [B, p149] So we may choose a sequence $\varepsilon'_j \downarrow 0$. Also if necessity be, choose an $\varepsilon_j > 0$ so that $b+\varepsilon_j \in C(F_E)$, $0 \leq \varepsilon_j \leq \varepsilon'_j$, and $*_1$: $F_{E_j}(b+\varepsilon_j) \leq (n+\varepsilon'_j)F_{E_j}(1+\varepsilon'_j)$. This is possible by monotonicity and right continuity of F_{E_j}.

Then, as $l \to \infty$, $F_{E_l}(b+\varepsilon_j) \to F(b+\varepsilon_j)$ and $F_{E_l}(1+\varepsilon'_j) \to F(1+\varepsilon'_j)$. Writing F_{E_l} instead of F_{E_j} in $*_1$ we may violate the inequality, but not by much: by Cauchy's criterium for convergence, for every $\bar{\varepsilon} > 0$ there exists a j such that for all $l \geq j$, $|F_{E_l}(b+\varepsilon_j) - F_{E_j}(b+\varepsilon_j)| \leq \bar{\varepsilon}$, and similarly $|F_{E_l}(1+\varepsilon'_j) - F_{E_j}(1+\varepsilon_j)| \leq \bar{\varepsilon}$. From this and $*_1$ one derives $F_{E_l}(b+\varepsilon_j) \leq (n+\varepsilon'_j)F_{E_l}(1+\varepsilon'_j) + (n+1)\bar{\varepsilon}$. Doing $l \to \infty$ we obtain that for all $\bar{\varepsilon} > 0$ there exists j so that $F_E(b+\varepsilon_j) \leq (n+\varepsilon'_j)F_E(1+\varepsilon'_j) + (n+1)\bar{\varepsilon}$. Now choosing a sequence of j with $j \uparrow \infty$, right continuity of F_E, yields $F_E(b) \leq nF_E(1) + (n+1)\bar{\varepsilon}$. The arbitraryness of $\bar{\varepsilon}$ finally yields $F_E(b) \leq nF_E(1)$, proving Theorem 2 also in the case that 1 or b are not continuity points of F_E.

4. An integral inequality

The 123 theorem can be casted into an integral inequality of an apparently new type. To this end write P (instead of Prob) for the probability measure defined on the space of real random variables X, Y. For X, Y independent we have the following computation which we justify below using facts from mostly from the book of Bauer [Ba].

$$\begin{aligned} P(|X-Y| \leq a) &= P(X-Y \in [-a,a]) \stackrel{1}{=} P_{X-Y}([-a,a]) \\ &\stackrel{2}{=} (P_X * P_{-Y})([-a,a]) \stackrel{3}{=} (A_2(P_X \otimes P_{-Y}))([-a,a]) \\ &\stackrel{4}{=} (P_X \otimes P_{-Y})(S) \stackrel{5}{=} \int P_{-Y}(S_{\omega_1}) dP_X(\omega_1). \end{aligned}$$

Here $S = \{(\omega_1,\omega_2) : \omega_1+\omega_2 \in [-a,a]\} = \{(\omega_1,\omega_2) : \omega_2 \in [-\omega_1-a, -\omega_1+a]\}$, and $S_{\omega_1} = \{\omega_2 : (\omega_1,\omega_2) \in S\}$ is the ω_1-section of S, see p112c-4. For '$\stackrel{1}{=}$' see the notational convention p140c3; in '$\stackrel{2}{=}$' we use that the measure of the sum of two independent random variables induces by p159c-0 the convolution of their individual image measures; '$\stackrel{3}{=}$' follows from the definition of convolution on (p122c-8ff); for '$\stackrel{4}{=}$' see the definition of image measure on p33c3 and the definition of S and $A_2(x,y) = x+y$; finally, for '$\stackrel{5}{=}$' see the relevant version of the Fubini theorem p114c6 and c13.

Now if in addition X and Y have the same distribution, they have the same distribution function $f = F_X = F_Y$, as defined in the previous section. So $f(x) = P_X(]-\infty, x]) = P(X \leq x) = P(Y \leq x)$. From this it easily follows that random variable $-Y$ has the distribution function $x \mapsto 1-f(-x)$. Consequently $P_{-Y}(S_{\omega_1}) = F_{-Y}(-\omega_1+a) - F_{-Y}(-\omega_1-a) = f(\omega_1+a) - f(\omega_1-a)$ and therefore finally $P(|X-Y| \leq a) = \int_{-\infty}^{+\infty}(f(\omega_1+a)-f(\omega_1-a))df(\omega_1)$. Via these observations, Theorem 1 yields in the differentiable case the following:

Theorem 3. *Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded increasing differentiable function, and let $0 < a < b$ be reals. Then*

$$\int_{-\infty}^{+\infty} (\,(2\lceil b/a \rceil - 1)(f(x+a) - f(x-a)) \;-\; f(x+b) + f(x-b)\,) f'(x)dx \geq 0.$$

Proof. The theorem is trivial if f is constant. If f is not constant we can choose adequate positive constants α and a real β, so that the function $f_1(x) = \alpha f(x) + \beta$ is increasing with $f_1(-\infty) = 0$, and $f_1(\infty) = 1$ and the theorem is true iff it is true with this f_1 in place of f. So we may now suppose f itself as a differentiable function increasing from 0 to 1. By the characterization given in [Ba, p146c-3], such a function - besides satisfying $df(x) = f'(x)dx$ - is certainly the distribution function of a probability measure. Then Theorem 1 and above formula for $P(|X-Y| \leq a)$ and a similar for b in place of a yield the claim after changing notation to x instead of ω_1.

□

References

[AY] N. Alon and R. Yuster, *The 123 Theorem and Its Extensions*, J. of Combin. Theory Ser. A 72, 321-331 (1995).

[Ba] H. Bauer, *Probability theory and elements of measure theory*, Academic Press, 1981.

[B] R.P. Boas, *A Primer of Real Functions*, 3rd Edition, MAA, 1981.

[CHL] R.W. Cottle, C.E. Habetler and G.J. Lemke, *On classes of copositive matrices*, Linear Algebra Appl. 3, 295-310 (1970).

[L] M. Loève, *Probability Theory I*, 4th Edition, Springer 1977.

[M] D.H. Martin, *Finite criteria for conditional positivity of a quadratic form*, Linear Algebra Appl. 39, 9-21 (1981).

Alexander Kovačec
Department of Mathematics, University of Coimbra, 3001-501, Coimbra, PORTUGAL
E-mail address: kovacec@mat.uc.pt

Miguel M. R. Moreira