Research Article Open Access

Alexander Kovačec, Miguel M. R. Moreira, and David P. Martins

The 123 theorem of Probability Theory and Copositive Matrices

Abstract: Alon and Yuster give for independent identically distributed real or vector valued random variables X, Y combinatorially proved estimates of the form $\text{Prob}(\|X-Y\| \le b) \le c\,\text{Prob}(\|X-Y\| \le a)$. We derive these using copositive matrices instead. By the same method we also give estimates for the real valued case, involving $X+Y$ and $X-Y$, due to Siegmund-Schultze and von Weizsäcker as generalized by Dong, Li and Li. Furthermore, we formulate a version of the above inequalities as an integral inequality for monotone functions.

Keywords: probabilistic inequalities, copositivity, integral inequality

MSC: 60E15, 15B48, 26D10

DOI 10.2478/spma-2014-0016
Received April 14, 2014; accepted October 10, 2014

1 Introduction

In [2], Alon and Yuster prove Theorems 1 and 3 below, while Theorem 2 is, for the case $a = b > 0$, due to Siegmund-Schultze and von Weizsäcker [9] in work on random walks, and to Dong, Li and Li [6] in the general case.

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*

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

and the multiplicative constant at the right cannot be improved.

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

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

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. *Let X, Y be independent, identically distributed real random variables and assume reals $0 < a \le b$. Then there holds the inequality*

$$\text{Prob}(|X+Y| \le b) \le \lceil 2b/a \rceil \text{Prob}(|X-Y| \le a).$$

Alexander Kovačec: Department of Mathematics, University of Coimbra, 3001-501, Coimbra, Portugal, E-mail: kovacec@mat.uc.pt
Miguel M. R. Moreira: Rua Luís de Camões, Nr. 102, 1300-360, Lisboa, Portugal, E-mail: miguel.mrm@hotmail.com
David P. Martins: Rua D. Manuel I, Edif. Império Porta 2-2D, 5370-412 Mirandela, Portugal, E-mail: davidmartins.chess@gmail.com

Theorem 3 (which is [2, 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.

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 3. *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*

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

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. One of the referees asked whether Theorem 1 is a corollary to Theorem 3. Concerning this, it is not clear how to tweak the proof of Theorem 3 to go through with dimension $d = 1$. But even if this can be done, given $b > 1$, in the interval $[-b, b]$ there exist at most $\lceil 2b \rceil$ points having mutual distances > 1. So there exists no $(b, \lceil 2b \rceil + 1)$-configuration on the real line. This means by Theorem 3 that we get Theorem 1 with the weaker constant $\lceil 2b \rceil$ in place of $\lceil 2b \rceil - 1$.

In [2] 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; similar remarks hold for the version of Theorem 2 given in [9].

Concerning Theorem 3 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 discussed in [2]. Concerning the quest for best possible constants we have nothing to add in this paper.

The proofs in [2] are combinatorial. Our purpose here is to give for the case that X, Y assume only finitely many values 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 $x'Cx > 0$ for all $x > 0$, 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 $a_1, a_2, \ldots, a_m$ with respective probabilities $\xi_1, \xi_2, \ldots, \xi_m > 0$. Define $\chi(P)$ to be 1 or 0 according to whether property P holds or not. Then for any real r,

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

where $\xi = (\xi_1, \ldots, \xi_m)'$ is the m-column of probabilities.

Almost all probability theoretic and measure theoretic material we shall later need can be found in Loève's or Bauer's books [7], [3].

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 $\{a_1, \ldots, a_m\} \subseteq \mathbb{R}$, be established by showing that the matrix $C = C(\underline{a}) = C(a_1, \ldots, a_m) = (c_{i,j})$ given by

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

is copositive; the proof of Theorem 2 can for the finitely many valued case similarly be reduced to the proof of copositivity of the matrix given by

$$c_{i,j} = \lceil 2b/a \rceil \chi(|a_i - a_j| \leq a) - \chi(|a_i + a_j| \leq b); \tag{2}$$

and the proof of Theorem 3 reduces to showing that if X, Y assume values only in $\{a_1, ..., a_m\} \subseteq \mathbb{R}^d$, then the matrix defined by

$$c_{i,j} = n\chi(\|a_i - a_j\| \leq 1) - \chi(\|a_i - a_j\| \leq b) \tag{3}$$

is copositive. Indeed we will show that all these matrices are strictly copositive and therefore in the finitely many valued cases we get strict inequalities for the probabilities.

These proofs are given in Section 2 based on characterizations of strict copositivity given by Cottle, Habetler and Lemke [5]. 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 and end with some comments.

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 Delfos Project for mathematically interested pre-university youngsters to extend the proof to cover the remaining main facts in [2]. The suggestion was taken up by the two other authors who did a good part of the mathematics of Section 2.

For convenience of the reader we have chosen consecutive numbering of the the statements.

2 Proofs for the cases that X, Y assume only finitely many values

According to [5, 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 criterion for strict copositivity given in part a of the following lemma.

Lemma 4. *Let C be a real symmetric matrix. Then:*
a. C is strictly copositive if and only if for every one of its principal submatrices $\bar{C}$, $\bar{y} > 0$ implies $\bar{C}\bar{y} \nleq 0$.
b. If each proper *principal submatrix of C is strictly copositive, then $y > 0$ and $Cy \leq 0$ imply y is entrywise positive (while C is not strictly copositive).*

Proof. a. $\Rightarrow$: Assume there exists $\bar{y} > 0$ so that $\bar{C}\bar{y} \leq 0$. Construct the vector y by putting $y_i = \bar{y}_l$ if i is the index of the lth column of $\bar{C}$ as a submatrix of C; put $y_i = 0$ otherwise. Then $y > 0$ while $y'Cy = \bar{y}'\bar{C}\bar{y} \leq 0$. 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 [5] cited, there exists a real $\lambda \leq 0$ associated to $\bar{y} > 0$ so that $\bar{C}\bar{y} = \lambda\bar{y}$. Hence $\bar{C}\bar{y} \leq 0$, a contradiction.

b. Assume there is a y so that $y > 0$ and $Cy \leq 0$, but y is not entrywise positive. Then there exists a k so that the kth entry of y, $y_k = 0$. Let $\bar{y}$ be the vector obtained from y by removing its k-th coordinate; and let $\bar{C}$ be the matrix obtained by removing the row and the column of indices k. Then $\bar{y} > 0$ and $\bar{C}$ is a principal submatrix of C and we get the contradiction $0 \geq y'Cy = \bar{y}'\bar{C}\bar{y} > 0$. □

We also shall need two simple facts for 'well distributed' sets on the real line.

Lemma 5. *Let S be a set of points on the real line that have mutual distances all larger than 1 and let $p \in \mathbb{R}$ and $b \in \mathbb{R}_{>1}$. Then:*
a. There are at most $\lceil b \rceil - 1$ points $s \in S$ satisfying $1 < p - s \leq b$ and at most $\lceil b \rceil - 1$ points $s \in S$ satisfying $1 < s - p \leq b$.

b. An interval I of length $\lambda(I)$ contains at most $\lceil \lambda(I) \rceil$ points from S.

Proof. a. Assume points $s_1 < s_2 < \cdots < s_{\lceil b \rceil}$ in S satisfy $1 < s_i - p \le b$. Since the distances between successive s_i are larger than 1, we have $s_{\lceil b \rceil} - p = \sum_{i=1}^{\lceil b \rceil -1}(s_{i+1} - s_i) + (s_1 - p) > (\lceil b \rceil - 1) + 1 \ge b$, a contradiction. The other claim is proved similarly.

b. A set of $1 + \lceil \lambda(I) \rceil$ points in S defines $\lceil \lambda(I) \rceil$ sucessive distances and hence the leftmost and rightmost of the points define a distance strictly larger than $\lceil \lambda(I) \rceil$ and cannot both lie in I. This yields the claim. □

Let now $b > a > 0$. It will be convenient to note that the $m \times m$ matrix $C(\underline{a}) = C(a_1, \ldots, a_m)$ referred to in (1) in connection with Theorem 1 has the alternative definition

$$c_{i,j} = \begin{cases} 2\lceil b/a \rceil - 2 & \text{if } |a_i - a_j| \le a \\ -1 & \text{if } a < |a_i - a_j| \le b \\ 0 & \text{if } b < |a_i - a_j|. \end{cases}$$

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

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

Proposition 6. *The matrix $C(\underline{a})$ in (1) is strictly copositive.*

Proof. By scaling we may assume that $a = 1$. 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 $\underline{a} = (a_1)$ and the matrix $C = C(a_1) = (2\lceil b \rceil - 2)$ has only one entry and this is positive. In the case $n = 2$, $\underline{a} = (a_1, a_2)$ and the matrix $C = C(\underline{a})$ has one of the forms $\begin{pmatrix} 2\lceil b \rceil - 2 & 0 \\ 0 & 2\lceil b \rceil - 2 \end{pmatrix}$, $\begin{pmatrix} 2\lceil b \rceil - 2 & -1 \\ -1 & 2\lceil b \rceil - 2 \end{pmatrix}$, or $\begin{pmatrix} 2\lceil b \rceil - 2 & 2\lceil b \rceil - 2 \\ 2\lceil b \rceil - 2 & 2\lceil b \rceil - 2 \end{pmatrix}$ according to if the cases $|a_1 - a_2| > b$, $1 < |a_1 - a_2| \le b$, or $|a_1 - a_2| \le 1$ hold. As $2\lceil b \rceil - 2 \ge 2$, the quadratic forms $y'Cy$ associated to these matrices assume on variable vectors $y > 0$ only (strictly) positive values.

Assume the claim already established for all matrices C associated with up to $m-1$ points on the real line. Fix an $\underline{a} = (a_1, \ldots, a_m)$ consisting of m real entries. Since the property of (strict) copositivity remains evidently unaltered under permutation equivalence, we may assume without loss of generality that $a_1 \le a_2 \le \cdots \le a_m$.

Define $k_1 = 1$ and inductively $k_{l+1} = \min\{i : k_l < i \le m, a_i - a_{k_l} > 1\}$, if the set used here is nonempty. Only finitely many ks can be defined, say $1 = k_1 < k_2 < \cdots < k_e \le m$.

Claim: The sum of the rows of C with indices $k_1, k_2, \ldots, k_e$ is a nonnegative nonzero m-row.

⁅ The claim says that $c_{k_1,j} + \cdots + c_{k_e,j} \ge 0$ for $j = 1, 2, \ldots, m$, with strict inequality for at least one j. There exists a unique s such that $k_s \le j < k_{s+1}$ or $k_s \le j \le m$ holds. These cases correspond to $a_{k_s} \le a_j < a_{k_{s+1}}$ or $a_{k_s} \le a_j \le a_m$, respectively. In either case $|a_{k_s} - a_j| \le 1$, and so $c_{k_s,j} = 2\lceil b \rceil - 2$.

By definition of the k_i, the reals $a_{k_1} < a_{k_2} < \cdots < a_{k_e}$ constitute a set of points that have mutual distances > 1. By Lemma 5a there exist at most $2\lceil b \rceil - 2$ indices k_i so that $1 < |a_{k_i} - a_j| \le b$ or, equivalently, so that $c_{k_i,j} = -1$.

Consequently $\sum_{i=1}^{e} c_{k_i,j} \ge -(2\lceil b \rceil - 2) + (2\lceil b \rceil - 2) = 0$ for all $j = 1, \ldots, m$. In the case $j = 1 = k_1$, we have $a_j < a_{k_i}$ for all $i = 2, \ldots, e$, and hence at most $\lceil b \rceil - 1$ of the $c_{k_i,1}$ are equal to -1. Since $c_{k_1,k_1} = 2\lceil b \rceil - 2$, we get $\sum_{i=1}^{e} c_{k_i,1} \ge -\lceil b \rceil + 1 + 2\lceil b \rceil - 2 = \lceil b \rceil - 1 \ge 1$. The claim is proved. ⁆

Let 1_K be the m-column that has 1s in coordinates $k_1, \ldots, k_e$ and 0s elsewhere. The claim then says $1_K' C > 0$. Now consider a column $y \in \mathbb{R}^m$, $y > 0$ and assume $Cy \le 0$. Then by Lemma 4b we know that y must be entrywise positive. But then, $0 \ge 1_K'(Cy) = (1_K' C)y > 0$, which is impossible. So $Cy \nleq 0$ and Lemma 4a gives that C is strictly copositive. □

This establishes Theorem 1 for random variables that assume only finitely many values.

To prove Theorem 2 via considerations similar to the previous ones, in the case that X, Y assume only finitely many values $a_1, a_2, \ldots, a_m$, after scaling, the fact to establish is that if $b \geq 1$, the matrix $C = C(\underline{a}) = (c_{i,j})$, defined by

$$c_{i,j} = \begin{cases} 0 & \text{if } |a_i - a_j| > 1, |a_i + a_j| > b, \\ -1 & \text{if } |a_i - a_j| > 1, |a_i + a_j| \leq b, \\ \lceil 2b \rceil & \text{if } |a_i - a_j| \leq 1, |a_i + a_j| > b, \\ \lceil 2b \rceil - 1 & \text{if } |a_i - a_j| \leq 1, |a_i + a_j| \leq b. \end{cases}$$

is copositive. In fact we have somewhat more.

Proposition 7. *This matrix C is strictly copositive.*

Proof. As $b \geq 1$, $\lceil 2b \rceil - 1 \geq 1$. Thus if the number of points $m = 1$, the matrix C consists of a single entry ≥ 1 and is hence strictly copositive. If $m = 2$, the matrix is of the form $C = \begin{pmatrix} c_{1,1} & u \\ u & c_{2,2} \end{pmatrix}$ with $c_{1,1}, c_{2,2} \geq 1$ and $u \geq 0$ or $u = -1$. If $u \geq 0$, C is trivially strictly copositive. If $u = -1$ the quadratic form defined by C is $(c_{1,1} - 1)x^2 + (c_{2,2} - 1)y^2 + (x - y)^2$. The only case that this form assumes for $(x, y) > (0, 0)$ a value ≤ 0 is when $x = y > 0$ and $c_{1,1} = c_{2,2} = 1$. In this case $b = 1$ and $|a_i + a_i| \leq b = 1$, that is, $-1/2 \leq a_i \leq 1/2$, for $i = 1, 2$. But then $|a_1 - a_2| \leq 1$ which contradicts that $u = -1$. Hence the matrix is C is strictly copositive also in this case.

Assume now established that every proper principal submatrix of C is strictly copositive.

Define, as in the previous proposition $k_1 = 1$, and inductively $k_{l+1} = \min\{i : k_l < i \leq m, a_i - a_{k_l} > 1\}$. We obtain integers $1 = k_1 < k_2 < \cdots < k_e \leq m$ so that $a_1 = a_{k_1} < a_{k_2} < \cdots < a_{k_e} \leq a_m$ is a sequence of points any two successive of which have distance > 1, except that $|a_{k_e} - a_m| \leq 1$. Also note that, if $e = 1$, then the matrix has only positive entries and then is trivially strictly copositive.

Claim: For all $j = 1, \ldots, n$, $c_{k_1,j} + c_{k_2,j} + \cdots + c_{k_e,j} \geq 0$ with strict inequality for at least one j.

⁅ Fix a j and define s as the unique integer satisfying $k_s \leq j < k_{s+1}$ or $k_s \leq j \leq m$. Then $|a_{k_s} - a_j| \leq 1$, and therefore

$$c_{k_s,j} = \begin{cases} \lceil 2b \rceil & \text{if } |a_{k_s} + a_j| > b \\ \lceil 2b \rceil - 1 & \text{if } |a_{k_s} + a_j| \leq b. \end{cases}$$

Let $I = \{i : c_{k_i,j} = -1\}$ and let $I' = \{i : -b \leq a_{k_i} + a_j \leq b\}$. By definition of C, $I \subseteq I'$. Furthermore, as any two distinct of the points $a_{k_i} + a_j$, $i \in I'$ have distance larger than 1 while the interval $[-b, b]$ has length $2b \leq \lceil 2b \rceil$, Lemma 5b yields $|I| \leq |I'| \leq \lceil 2b \rceil$. If $s \in I'$, then $c_{k_s,j} = \lceil 2b \rceil - 1 \neq -1$, $s \notin I$, $|I| \leq \lceil 2b \rceil - 1$ and hence $\sum_{i=1}^{e} c_{k_i,j} \geq -|I| + \lceil 2b \rceil - 1 \geq 0$. If $s \notin I'$ then $c_{k_s,j} = \lceil 2b \rceil$ and hence again $\sum_{i=1}^{e} c_{k_i,j} \geq -|I| + \lceil 2b \rceil \geq 0$.

Finally we show that the sum cannot be 0 for all j. If $a_{k_e} + a_1 \leq 0$, choose $j = 1 = k_1$. Then $s = 1$ and $c_{k_s,j} = c_{1,1}$ is $\lceil 2b \rceil$ if $2|a_1| > b$ and $\lceil 2b \rceil - 1$ if $2|a_1| \leq b$. Consequently, $|I| = 0$ implies a positive sum and we need only to look at the cases $|I| \geq 1$. For all i, $1 \leq i < e$, we have $a_{k_i} + a_1 < i - e$. In particular if $i \leq e - \lceil b \rceil$, we have $|a_{k_i} + a_1| > b$ and hence $c_{k_i,1} \geq 0$. So in order that $c_{k_i,1} = -1$, it is necessary that $i \in \{e - \lceil b \rceil + 1, \ldots, e\}$. This means $|I| \leq \lceil b \rceil$. If $b > 1$, then $\lceil 2b \rceil - 1 > \lceil b \rceil$, and so $\sum_{i=1}^{e} c_{k_i,1} \geq 1$. For the case $b = 1$ we have the subcase $2|a_1| > 1$ in which $c_{1,1} = 2$ and $|I| \leq 1$, so $\sum_{i=1}^{e} c_{k_i,1} \geq 1$; and the subcase $2|a_1| \leq 1$ with $c_{1,1} = 1$. If then $|I| = 1$, then $I = \{e\}$. Suppose $c_{k_e,1} = -1$. Then $a_{k_e} - a_1 > 1$ and so $a_1 + a_1 < a_{k_e} + a_1 - 1 \leq -1$, contradicting $2|a_1| \leq 1$. The case $0 < a_{k_e} + a_1$ can be handled similarly, selecting $j = k_e$ instead of k_1. ⁆

As in the proof of Proposition 6 we now conclude that C is strictly copositive. □

We now take up the vector valued case. For proving an analogue to above propositions for the higher dimensional case, we need a lemma.

Lemma 8. *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 exists a set of $n' \leq n - 1$ further (distinct) points $a_i \in \mathcal{P}$, $i = 1, \ldots, n'$ so that $\{a_0, a_1, a_2, \ldots, a_{n'}\} \subseteq \mathcal{P}$ is a well distributed *point set, and every point in $\mathcal{P}$ is* near *one of its points: that is, for $0 \leq i < j \leq n'$, $\|a_i - a_j\| > 1$ and for all $x \in \mathcal{P}$, there is an i, $0 \leq i \leq n'$ so that $\|x - a_i\| \leq 1$.*

b. If in part a, $n' = n-1$, then the first coordinate of one of the points $a_1, \ldots, a_{n-1}$ is smaller than the first coordinate of $a_0 = \text{center}(B)$.

Proof. a. Choose points, $a_1, a_2, \ldots$ in $\mathcal{P}$ such that for each i all points in $a_0, a_1, \ldots, a_i$ have mutual distances > 1. This process necessarily comes to a halt at an $i = n' \le n-1$, for otherwise we would contradict the general hypothesis of the lemma. Also, if $x \in \mathcal{P}$ is any point, the construction evidently implies that there exists an $i \in \{0, 1, 2, \ldots, n'\}$ so that $\|x - a_i\| \le 1$.

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

To prove Theorem 3 note that the $m \times m$ matrix C defined in (3) in connection with it has the alternative definition

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

Proposition 9. *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 up to $m-1$ points. We have to show that for $y > 0$ it is impossible that $Cy \le 0$. Supposing that such a y exists, by Lemma 4b, y has strictly positive entries. Thus the function $2^{\{1,\ldots,m\}} \ni I \mapsto \sum_{j\in I} y_j$ defines a positive measure on $\{1, 2, \ldots, m\}$ whose only null set is the empty set.

Let $P_i = \{j : c_{i,j} = n-1\}$ and $N_i = \{j : c_{i,j} = -1\}$ be the sets of column indices of line i, where $c_{i,j}$ 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, $0 \ge \sum_j c_{i_0,j} y_j = \sum_{j\in P_{i_0}} (n-1) y_j - \sum_{j\in N_{i_0}} y_j$, hence $\sum_{j\in N_{i_0}} y_j \ge (n-1) p_{i_0}$.

The set $\mathcal{P} = \{a_j : j \in N_{i_0}\} \uplus \{a_{i_0}\}$ of points is contained in the ball $B = B(a_{i_0}, b)$ which is centered at one of them. By Lemma 8, we can find $n' \le n-1$ indices $i_1, \ldots, i_{n'} \in N_{i_0}$ so that $\{a_{i_0}, a_{i_1}, \ldots, a_{i_{n'}}\}$ is a well distributed point set and each point in $\mathcal{P}$ is near to one of the points contained in it.

No point in $\mathcal{P} \setminus \{a_{i_0}\}$ is near to a_{i_0}; hence $N_{i_0} \subseteq P_{i_1} \cup \cdots \cup P_{i_{n'}}$. But then, by the definition of i_0, and the inequality above,

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

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

Now by Lemma 8b, 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 have $p_{i_{\bar{k}}} \ne \mu = p_{i_0}$, and hence $p_{i_{\bar{k}}} < p_{i_0}$, contradicting that by the previous paragraph $p_{i_{\bar{k}}} = p_{i_0}$. □

The strict copositivities of our matrices imply that we have herewith proved for iid real random variables X, Y that take only finitely many values, the strict inequalities

$$\text{Prob}(|X-Y| \le b) < (2\lceil b/a \rceil - 1)\text{Prob}(|X-Y| \le a), \text{ if } b > a > 0;$$

$$\text{Prob}(|X+Y| \le b) < \lceil 2b/a \rceil \text{Prob}(|X-Y| \le a), \text{ if } b \ge a > 0,$$

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

$$\text{Prob}(\|X-Y\| \le b) < n\text{Prob}(\|X-Y\| \le 1), \text{ if } b > 1,$$

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 3; it will be clear that Theorems 1 and 2 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 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 [7]. 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'. The second edition of Loève's book [L2nd] has the same material usually one or two pages earlier.

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. For lighter notation we use, in this and the next section, P instead of 'Prob'.

On the reals one defines as the standard the Borel σ-algebra $\mathcal{B}$ and 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 of A. As done in Loève, p106c1, 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,

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

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 $X = (X_1, \ldots, X_d)$, where each component function X_i is a real random variable, pp110c-4 and 152c-2. Then to X and j associate the function $E_j^X = (E_j^{X_1}, \ldots, E_j^{X_d})$. 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 and 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 (p224c6 in [L2nd]). Therefore, since E_j^X, E_j^Y are finitely valued, Theorem 3 tells us that under its hypotheses 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 $j = 1, 2, 3, \ldots$. Now by construction, p108, we know for each coordinate X_i of X the following: for each $\varepsilon > 0$ and each $\omega \in \Omega$ there exists an $n = n(\omega, \varepsilon, i)$ so that for all $j \geq n$ there holds $|X_i(\omega) - E_j^{X_i}(\omega)| \leq \varepsilon$ if $|X_i(\omega)| \leq n$ and $E_j^{X_i}(\omega) \begin{cases} = & n \\ = & -n \end{cases}$ if $X_i(\omega) \begin{cases} \geq & n \\ \leq & -n \end{cases}$, respectively.

According to Exercise 2 to Ch. 2 on p139, this implies that for every $\varepsilon > 0$ there exists a set $A \subseteq \Omega$ with $P(A) \geq 1 - \varepsilon$ such that $E_j^{X_i} \overset{\text{u.}}{\to} X_i$; i.e. uniformly on A, a fact Loève calls almost uniformly and abbreviates with '$\overset{\text{a.u.}}{\to}$.' As we are working in finite dimensional space, we thus get that $\|E_j^X - X\| \overset{\text{a.u.}}{\to} 0$ on A. Now Exercise 3, p140, says that almost uniform convergence of a real random variable implies convergence in measure; in our context we therefore have $E_j^X \overset{P}{\to} X$, that is, for all $\varepsilon > 0$, $P([\|E_j^X - X\| \geq \varepsilon]) \to 0$; and similarly for Y. Now define $E_j := E_j(\omega) = \|E_j^X(\omega) - E_j^Y(\omega)\|$, and $E := E(\omega) = \|X(\omega) - Y(\omega)\|$. Functions E_j and E are real random variables and we have the estimate $|E_j - E| = |\|E_j^X - E_j^Y\| - \|X - Y\|| \leq \|(E_j^X - E_j^Y) - (X - Y)\| = \|E_j^X - X + Y - E_j^Y\| \leq \|E_j^X - X\| + \|Y - E_j^Y\|$, which allows us to infer from the above that $E_j \overset{P}{\to} E$.

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. For what follows in a moment we note that Loève defines distribution functions for a random variable X using the definition $F_X(x) = P(X < x)$, but the proof of the fact we cite below from [L] can be easily adapted to our definition $F_X(x) = P(X \leq x)$, chosen to be closer to [2]. Loève's definition leads to left continuous functions, the one here to right continuous ones.

By p170c2 we can conclude from $E_j \overset{P}{\to} E$ that $F_{E_j} \to F_E$ on the set $C(F_E)$ of continuity points of F_E and this is sufficient to prove the desired inequality $F_E(b) \leq nF_E(1)$ if $b, 1 \in C(F_E)$.

But in fact the proof of the cited fact in Loève shows that for any v and any $v^{'}, v^{''}$ with $v^{'} < v < v^{''}$ there holds $F_E(v^{'}) \leq \liminf_j F_{E_j}(v) \leq \limsup_j F_{E_j}(v) \leq F_E(v^{''})$, and hence by definition of lim inf and lim sup for any $\varepsilon > 0$ and almost all integers $j > 0$, $F_E(v^{'}) - \varepsilon \leq F_{E_j}(v) \leq F_E(v^{''}) + \varepsilon$. If we use this for $v = b$ and $v = 1$ then we may conclude that for all $x^{'}, y^{''}, \varepsilon$ satisfying $x^{'} < b$ and $1 < y^{''}$ and $\varepsilon > 0$, there holds for almost all j the estimate $F_E(x^{'}) - \varepsilon \leq F_{E_j}(b) \leq nF_{E_j}(1) \leq nF_E(y^{''}) + \varepsilon$, and hence $F_E(x^{'}) - \varepsilon \leq nF_E(y^{''}) + \varepsilon$. From right continuity of F_E, we can now infer the same inequality with 1 in place of $y^{''}$. By the arbitraryness of $\varepsilon > 0$ we finally find $F_E(x^{'}) \leq nF_E(1)$, that is, $P(E \leq x^{'}) \leq nP(E \leq 1)$ for all $x^{'} < b$. By left continuity of $t \mapsto P(E < t)$ this means $P(E < b) \leq nP(E \leq 1)$.

It might not be true that the nonexistence of a $(b, n + 1)$-configuration entails for small enough $\varepsilon > 0$ the nonexistence of a $(b + \varepsilon, n + 1)$-configuration. This in turn means a slightly involved argument is necessary to obtain the slightly stronger inequality $P(E \leq b) \leq nP(E \leq 1)$. Define a real $1^+ \geq 1$ (called so instead of say $1 + \varepsilon$ with $\varepsilon \geq 0$ for simplicity of notation). Modify the definition of a (b, n)-configuration to that of a $(b, n, 1^+)$-configuration by replacing the radius and distances 1 in that definition by 1^+. The proof of Lemma 8 remains valid if the distances 1 there occurring are replaced by 1^+. The matrix C of Proposition 9 with distance limits 1 replaced by 1^+ is strictly copositive again since the proof of the proposition goes through with these alterations. These facts allow to say that if $d \geq 2$ and $\mathbb{R}^d$-valued iid random variables X, Y take only finitely many values, there holds the inequality

$$P(\|X - Y\| \leq b^{'}) < nP(\|X - Y\| \leq 1^+),$$

provided Euclidean d-space does not permit a $(b^{'}, n+1, 1^+)$-configuration. Now the considerations before let us say $P(E < b^{'}) \leq nP(E \leq 1^+)$ with a $b^{'} > 1^+$ which we may assume larger than the original b. As $[E \leq b] \subseteq [E \leq b^{'}]$, we have $P(E \leq b) \leq nP(E \leq 1^+)$. As $1^+ \geq 1$ was arbitrary, we get by right continuity of the distribution function, that $P(E \leq b) \leq nP(E \leq 1)$, i.e. $F_E(b) \leq nF_E(1)$ also in the cases that b or 1 are possibly not in $C(F_E)$.

One can obtain by similar (in fact easier) considerations the inequalities of Theorems 1 and 2 once that they are proved for strict inequality '$< b$' at the left.

4 An integral inequality and some remarks

The 123 theorem can be casted into an integral inequality which after consulting several books and an authority [Al] in the field of inequalities, we venture to guess is of an apparently new type. As in the section before, let P stand for a probability measure defined on the space of real random variables X, Y. Denote by $\star$ the convolution of measures, by $\otimes$ the product of two measures, and by A_2 the map $\mathbb{R} \times \mathbb{R} \ni (x, y) \mapsto x + y$. The measures P_X, P_Y are the image measures of P induced by X, Y respectively, and $A_2(P_X \otimes P_Y)$ is the image measure of $P_X \otimes P_Y$ induced on $(\mathbb{R}, \mathcal{B})$ by A_2. Also for fixed real $a > 0$, let $S = \{(\omega_1, \omega_2) : \omega_1 + \omega_2 \in [-a, a]\} = \{(\omega_1, \omega_2) : \omega_2 \in [-\omega_1 - a, -\omega_1 + a]\}$, and $S_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \in S\}$ be the ω_1-section of S, see p112c-4 of [Ba].

For X, Y independent we have the following computation which we justify below citing additional facts mostly from the book of Bauer [3].

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

For '$\overset{1}{=}$' see the notational convention p140c3; in '$\overset{2}{=}$' we use that the measure of the sum of two independent random variables induces by p159c-0 the convolution of their individual image measures; '$\overset{3}{=}$' follows from the definition of convolution in Bauer's book on p122c-8ff ; for '$\overset{4}{=}$' see the definition of image measure on p33c3 and the definition of S; finally, for ‘$\overset{5}{=}$’ see the relevant version of the Fubini theorem p114c6 and c13 and the notational convention p57c8.

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 in case F is continuous, 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 10. *Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded increasing differentiable function, and let $0 < a < b$ be reals. Then*

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

Proof. The theorem is trivial if f is constant. If f is not constant we can choose suitable 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 [3, 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 the 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. □

We conclude with some remarks.

a. We learned of the articles [9] and [6] in the time between submission of this paper and reception of the referee’s reports. The proof of Theorem 2 in these papers follows very different lines but it is noteworthy that in proving a dichotomy for integrals over measurable functions, the authors of [9] unwittingly establish on pages 677-8 the following: Let A be a real symmetric matrix. Assume that for every $y > 0$ there is an index i such that $y_i > 0$ and $(Ay)_i > 0$. Then A is strictly copositive.

The dichotomy result itself says that for a $\mathcal{B} \otimes \mathcal{B}$ measurable bounded symmetric function $f : \Omega^2 \to \mathbb{R}$ one either has i or ii:

i. For some probability measure P on $\mathcal{B}$, $\int_\Omega f(\cdot, y)P(dy) \leq 0$ P-a.e.

ii. For all probability measures P on $\mathcal{B}$, $\int_{\Omega^2} f(x, y)P(dx)P(dy) > 0$.

To establish their Theorem 1 (our Theorem 2) with strict inequality, they apply their dichotomy theorem to the function $f(x, y) = 2\chi(|x - y| \leq 1) - \chi(|x + y| \leq 1)$ after having ruled out - with some hard work - the first alternative. In spite of a serious search, we did not find further papers giving probability inequalities resembling those we treated in this article. Dong, Li and Li [6] generalize versions of Theorem 2 to the Banach space setting.

b. An article by Martin [8] gives necessary and sufficient conditions for nonnegativity and positivity of quadratic forms $x'Cx$ under the condition that x satisfies $Ax \geq 0$, where A is any real rectangular matrix. Classical copositivity arises when $A = I$. It might be worthwhile to exploit his results to obtain further probabilistic inequalities.

Acknowledgement: We thank one of the referees for the kind reception which he or she conceded this paper and the other referee in particular for the doubts he expressed about our original treatment of the inequality $F_E(b) \leq nF_E(1)$ at points that are not necessarily continuity points of F_E, the suggestion to formulate and prove Theorems 1 and 2 for strict copositive matrices (instead of simple copositivity), and the request to clarify the convergence $E_j \to E$. These requests led, we think, to a substantial improvement of the paper.

References

[1] H. Alzer, Private communication.

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

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

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

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

[6] Z. Dong, J. Li and W.V. Li, *A Note on Distribution-Free Symmetrization Inequalities*, J. Theor. Probab. 2014 (DOI 10.1007/s10959-014-0538-z)

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

[8] D.H. Martin, *Finite criteria for conditional definiteness of quadratic forms*, Linear Algebra Appl. 39, 9-21 (1981).

[9] R. Siegmund-Schultze and H. von Weizsäcker, *Level crossing probabilities I: One-dimensional random walks and symmetrization*, Adv. Math. 208, 672-679 (2007).