THE TOPOLOGY OF NEGATIVELY ASSOCIATED DISTRIBUTIONS

BY

JONATHAN ROOT (BOSTON, MA) AND MARK KON* (BOSTON, MA)

Abstract. We consider the sets of negatively associated (NA) and negatively correlated (NC) distributions as subsets of the space $\mathcal{M}$ of all probability distributions on $\mathbb{R}^n$, in terms of their relative topological structures within the topological space of all measures on a given measurable space. We prove that the class of NA distributions has a non-empty interior with respect to the topology of the total variation metric on $\mathcal{M}$. We show, however, that this is not the case in the weak topology (i.e. the topology of convergence in distribution), unless the underlying probability space is finite. We consider both the convexity and the connectedness of these classes of probability measures, and also consider the two classes on their (widely studied) restrictions to the Boolean cube in $\mathbb{R}^n$.

2020 Mathematics Subject Classification: Primary 60A10; Secondary 60E05, 60E99.

Key words and phrases: negative association, positive association, probability measure, topology.

1. INTRODUCTION

In recent years negatively associated probability distributions have been studied as potential generalizations of independent random variables [15, 7, 5, 28, 1]. However, the characterization of such probability measures on $\mathbb{R}^n$ has been elusive. In many cases just the specialization of such a characterization to Boolean cube measures, i.e. probability measures whose marginals are simple variations of Bernoulli measures, has generated a great deal of interest [24, 4]. The characterization of the set of negatively associated measures on $\mathbb{R}^n$ can involve even simpler questions regarding the topological structure of this set within the space of all measures. This question may have different answers under different topologies on the space of measures, which include the total variation topology and the standard weak (distributional) topology. Simple versions of this question include whether the space of such distributions is connected, convex, closed, and whether it has an interior with respect to a given topology.

Denote by $\mathcal{M}(\mathbb{R}^n)$ the set of all Borel probability measures on $\mathbb{R}^n$. A probability measure $\mu \in \mathcal{M}(\mathbb{R}^n)$ is said to be *negatively correlated* (NC) if

$$(1.1) \qquad \int\limits_{\mathbb{R}^n} x_i x_j \, d\mu(x) \leqslant \int\limits_{\mathbb{R}^n} x_i \, d\mu(x) \int\limits_{\mathbb{R}^n} x_j \, d\mu(x), \quad \forall 1 \leqslant i \neq j \leqslant n,$$

assuming the integrals exist. We say that μ is *strictly* NC if strict inequality holds in (1.1). We denote the class of NC distributions by $\mathcal{M}_{\mathrm{NC}}$ or $\mathcal{M}_{\mathrm{NC}}(\mathbb{R}^n)$. In this context, it is also common and equivalent to say that the variables themselves, $X_1, \ldots, X_n$, are negatively correlated.

The functions $f_i(x) = x_i$, $i = 1, \ldots, n$, are non-decreasing on $\mathbb{R}^n$. In general, we say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is *non-decreasing* if $f(x) \geqslant f(y)$ whenever $x \geqslant y$ in the product ordering ($x \geqslant y$ if and only if $x_i \geqslant y_i$ for each $i = 1, \ldots, n$). We say that $f : \mathbb{R}^n \to \mathbb{R}$ is *non-increasing* if $f(x) \leqslant f(y)$ whenever $x \geqslant y$ (again in the product ordering). We denote subsets of the index set $\{1, \ldots, n\}$ by I, J, and define $x_I \in \mathbb{R}^{|I|}$ to be the restriction of $x \in \mathbb{R}^n$ to the index set I; here $|I|$ denotes the cardinality of I. Moreover, we denote by $\mu^{(I)}$ the marginal distributions of μ: for $A \subset \mathbb{R}^{|I|}$,

$$(1.2) \qquad \mu^{(I)}(A) := \int\limits_{A} \int\limits_{\mathbb{R}^{-I}} d\mu(x),$$

where $\mathbb{R}^{-I}$ denotes all $x = (x_i)_{i \notin I}$. Then μ is said to be *negatively associated* (NA) if for all disjoint $I, J \subset \{1, \ldots, n\}$ and all non-decreasing and integrable $f : \mathbb{R}^{|I|} \to \mathbb{R}$ and $g : \mathbb{R}^{|J|} \to \mathbb{R}$, we have

$$(1.3) \qquad \int\limits_{\mathbb{R}^n} f(x_I) g(x_J) \, d\mu(x) \leqslant \int\limits_{\mathbb{R}^n} f(x_I) \, d\mu(x) \int\limits_{\mathbb{R}^n} g(x_J) \, d\mu(x),$$

assuming the integrals exist. Equivalently,

$$(1.4) \qquad \mathrm{Cov}_\mu(f(x_I), g(x_J)) \leqslant 0,$$

where $\mathrm{Cov}_\mu : L^1(\mathbb{R}^n, \mu) \times L^1(\mathbb{R}^n, \mu) \to \mathbb{R}$ denotes the covariance operator. Note that if f or g is constant, then we have trivial equality in (1.3). With that said, we say that μ is *strictly* NA if strict inequality holds in (1.3) for all (μ-almost surely) non-constant $f(x_I)$ and $g(x_J)$ and disjoint $I, J \subset \{1, \ldots, n\}$. If we specify $f(x) = f_i(x) = x_i$ and $g(x) = g_j(x) = x_j$, $i \neq j$, then we arrive at (1.1). Thus negative association is stronger than negative correlation. We denote the class of NA distributions by $\mathcal{M}_{\mathrm{NA}}$ or $\mathcal{M}_{\mathrm{NA}}(\mathbb{R}^n)$. As in the case of negative correlation, it is common and equivalent to consider negatively associated variables. That is, if the variables $X_1, \ldots, X_n$ are distributed according to a negatively associated distribution, then one may say that the variables themselves are negatively associated.

If μ is the standard Cauchy distribution, then the second moment is not defined, nor is the first moment, whence the first integral on the left-hand side of (1.3) is

not always defined, nor is it defined in (1.1). We therefore assert that indeed, the definition of negative association and negative correlation requires the assumption that the product fg is L^1, that is, the integral of the product fg exists and is finite.

To clarify this, we put forth the assertion that the variables X and $-X$ are always negatively associated. We supply the proof.

Fix variables X and $-X$, and let Y be a distinct variable. It is to be noted that whenever $X \geqslant Y$, we have $-X \leqslant -Y$. With that in mind, suppose f and g are increasing functions in one variable, and consider

$$(f(X) - f(Y))(g(-X) - g(-Y)).$$

Whenever $X > Y$ we have $-X < -Y$, whence we obtain the inequality

$$(f(X) - f(Y))(g(-X) - g(-Y)) \leqslant 0.$$

Expanding the left hand side gives us

$$f(X)g(-X) - f(X)g(-Y) - f(Y)g(-X) + f(Y)g(-Y) \leqslant 0.$$

Integrating in X and Y yields

$$2 \int f(X)g(-X)\,d\mu(X) \leqslant 2 \int f(X)\,d\mu(X) \int f(-X)\,d\mu(X).$$

It follows that $X, -X$ are negatively associated assuming the above integrals are finite and exist.

It is stated in [15] that pairs of random variables X and Y satisfy the condition of negative association (NA) if and only if they satisfy the condition of negative quadrant dependence (NQD). The latter condition reads, for all real x, y,

$$P(X \leqslant x, Y \leqslant y) \leqslant P(X \leqslant x)P(Y \leqslant y).$$

An infinite family of NA random variables is defined by demanding that every finite subfamily is negatively associated. Fix a family $(X_n)_{n \in \mathbb{N}}$ of pairwise negatively quadrant dependent random variables with the same distribution. It was proven by Matuła [20] that $S_n/n \to a$ almost surely if $S_n = \sum_{i=1}^n X_i$ and $a = \mathbf{E}X_1$. He further proves that if $(X_n)_{n \in \mathbb{N}}$ is a sequence of negatively associated random variables with finite second moment and $\sum_{n=1}^\infty \text{Var}(X_n)/n^2 < \infty$, then $(S_n - \mathbf{E}S_n)/n \to 0$ almost surely as $n \to \infty$.

Examples of infinite sequences of NA random variables include infinite sequences of independent random variables. It was shown in [15] that a permutation distribution defined on a fixed vector $x = (x_1, \ldots, x_k)$ of k real numbers is NA. The permutation distribution is the joint distribution of the vector $X = (X_1, \ldots, X_k)$ which takes as values the $k!$ permutations of x with equal probability $1/k!$. We can extend a permutation distribution to sequences of infinite length as follows.

Property P_7 in [15] asserts that the union of independent sets of NA random variables is NA. Conduct an experiment where a vector $(x_1, \ldots, x_n)$ of real numbers is fixed, and a permutation distribution is given on this set of n real numbers. Repeat the permutation independently without bound, yielding an infinite family of NA random variables.

Besides negative association, other attempts to quantify and conceptualize dependences among random variables appear in, for example, [2, 3, 10, 16, 17]. A concept closely related to NA, known as positive association (PA), sheds some light on the class of NA distributions. The notion of positive association was introduced into the statistical literature prior to negative association, in [11]. We say that μ is *positively associated* if $\mathrm{Cov}_\mu(f, g) \geqslant 0$ for all pairs of non-decreasing, real-valued functions f and g. We note that we no longer assume that f and g are defined on disjoint subsets of variables, as we did with negative association. Remarkably (or not), significantly greater progress has been made in the theory of positive association than in the theory of negative association. This, in part, is due to an elegant result known as the FKG inequality [13], which gives a sufficient criterion for PA. At its most basic level, the FKG inequality is known as Chebyshev's inequality [9] (distinct from the standard Chebyshev inequality in elementary probability). This theorem states that if X is a random variable on $\mathbb{R}$ (as opposed to $\mathbb{R}^n$), and $f, g : \mathbb{R} \to \mathbb{R}$ are both non-decreasing, then

$$(1.5) \qquad\qquad \mathbf{E}(f(X)g(X)) \geqslant \mathbf{E}f(X)\mathbf{E}g(X).$$

This holds for any probability distribution on the real line, so long as f and g are non-decreasing (or non-increasing). The proof of (1.5) is straightforward, and follows from the basic pointwise inequality

$$(1.6) \qquad\qquad (f(x) - f(y))(g(x) - g(y)) \geqslant 0,$$

which holds for all non-decreasing (or non-increasing) $f, g : \mathbb{R} \to \mathbb{R}$. Indeed, assuming x and y are independent and identically distributed, upon expanding (1.6) and double integrating (over x and y) we obtain Chebyshev's inequality (1.5).

The FKG inequality is essentially a generalization of (1.5) to the product setting, $\mathbb{R}^n$ equipped with the product ordering. To state the result, we first define the functions $\wedge$ (meet, or greatest lower bound) and $\vee$ (join, or least upper bound) by

$$x \wedge y := \max\{z \in \mathbb{R}^n : z \leqslant x, z \leqslant y\},$$
$$x \vee y := \min\{z \in \mathbb{R}^n : z \geqslant x, z \geqslant y\}.$$

Then the FKG theorem states that if a discrete probability measure μ on $\mathbb{R}^n$ satisfies

$$(1.7) \qquad\qquad \mu(x \vee y)\mu(x \wedge y) \geqslant \mu(x)\mu(y)$$

then μ is positively associated.

Unfortunately, a criterion as simple as (1.7) does not (yet) exists for negative association. As pointed out by Pemantle [24], the notion of negative association is not nearly as "robust" as positive association. Since Chebyshev's inequality (1.5) implies that any random variable is positively associated with itself, we cannot incorporate every non-decreasing function in the definition of negative association; but rather non-decreasing functions defined on disjoint coordinate subsets. And again, as noted by Pemantle [24], there is a bound on how far $\mathbf{E}x_i x_j$ can lie below $\mathbf{E}x_i \mathbf{E}x_j$, due to the inequality $\text{Var}(\sum x_i) = \sum \text{Cov}(x_i, x_j) \geqslant 0$.

The study of the class of negatively associated random variables dates back to [15], where basic properties of NA random variables are derived, and examples of NA random variables are given: multinomial, convolution of unlike multinomials, multivariate hypergeometric, Dirichlet, and Dirichlet compound multinomial variables. Though the notion of negative association has existed for some time, the literature on these distributions is still quite sparse [15, 7, 24, 4]. But interest in them is growing, due in part to the ease with which sums of NA (even NC) random variables satisfy sub-Gaussian tail bounds. Specifically, if μ is negatively correlated on $\mathbb{R}^n$, then

$$\mu\Big(x \in \mathbb{R}^n : \Big|\sum_{i=1}^n x_i - \mathbf{E}_\mu \sum_{i=1}^n x_i\Big| \geqslant \lambda\Big) \leqslant Ce^{-c\lambda^2}$$

for some absolute constants $c, C > 0$. It is conjectured that the same may be true for the replacement of sums $\sum_i x_i$ by more general Lipschitz functions of such variables; the only work on this question seems to come from [25, 8, 14]. (The notion that Lipschitz functions on a probability space concentrate about their mean, in the sense that their tails are (in the best case) sub-Gaussian, is called the concentration of measure phenomenon [18].) The work in [25] seems to be inspired by the recent article [4], in which the authors develop a novel notion of negative dependence known as the strong Rayleigh property. Their approach is via the geometry of associated generating polynomials, and they prove several conjectures put forth in this area of research.

We emphasize here, however, that nowhere in the literature has the structure of the space of negatively associated or negatively correlated distributions been studied. It is therefore natural to ask about the topological or geometric properties of the space of NA, or even NC, distributions. This question is the major impetus behind our work.

We consider these two classes of measures (NA and NC) broadly, from a topological perspective. We view them as subsets of the general space of measures $\mathcal{M}(\mathbb{R}^n) = C_0(\mathbb{R}^n)^*$ (the dual of the space of continuous real-valued functions which vanish at infinity) endowed with the weak topology (technically this should be denoted as the weak-* topology). This is the weakest topology ensuring the continuity of the maps $f \mapsto \int_{\mathbb{R}^n} f \, d\mu$ for $f \in C_0(\mathbb{R}^n)$, and coincides with the standard topology of convergence in distribution for measures. Thus we say that a sequence

of distributions μ_n *converges weakly* to a distribution μ if

$$\int\limits_{\mathbb{R}^n} f \, d\mu_n \to \int\limits_{\mathbb{R}^n} f \, d\mu$$

for all $f \in C_0(\mathbb{R}^n)$. When $X \subset \mathbb{R}^n$ is compact in the standard topology, we may define the weak topology on $\mathcal{M}(X)$ as follows. A basic open set in the weak topology is given by [23, 27]

$$(1.8) \quad V_\mu(f_1, \ldots, f_k; \epsilon_1, \ldots, \epsilon_k)$$
$$:= \Big\{ \nu \in \mathcal{M}(X) : \Big| \int f_i \, d\nu - \int f_i \, d\mu \Big| < \epsilon_i, \, i = 1, \ldots, k \Big\},$$

where $f_1, \ldots, f_k$ are continuous real-valued functions on X. The family of sets obtained by varying $\mu, k, f_1, \ldots, f_k, \epsilon_1, \ldots, \epsilon_k$ form a basis for the weak topology, i.e. a collection of sets whose unions form all open sets. Thus a sequence of distributions μ_n converges weakly to a distribution μ if and only if

$$\int\limits_X f \, d\mu_n \to \int\limits_X f \, d\mu$$

for every $f \in C(X)$ (now bounded due to compactness).

We may in addition view the NC and NA families as subsets of the space $\mathcal{M}(\mathbb{R}^n)$ of all measures, but now endowed with the total variation topology. This topology is induced from the *total variation distance*

$$\|\mu - \nu\|_{\text{TV}} := \sup_{|f| \leqslant 1} \Big| \int\limits_{\mathbb{R}^n} f \, d\mu - \int\limits_{\mathbb{R}^n} f \, d\nu \Big|.$$

In particular, in the setting of a discrete probability space (i.e. with support on a countable number of points), the total variation distance may be expressed as

$$(1.9) \qquad \qquad \|\mu - \nu\|_{\text{TV}} = \sum_{x \in \mathbb{R}^n} |\mu\{x\} - \nu\{x\}|.$$

Unless μ has finite support, the total variation distance induces a stronger topology than the weak topology. This will be discussed later on in this paper.

An outline of the paper is as follows. We begin by showing that the general class of NA distributions on a compact subspace of $\mathbb{R}^n$ has a non-empty interior in the total variation topology, but not in the weak topology. We next specialize to the subspace of measures concentrated on $I_n = \{0, 1\}^n$, the Boolean cube, and consider the interior and boundary of these distributions. This simple case affords intuitive arguments and constructive proofs. But it is still of great interest, and much is unknown about negative association on the Boolean cube [24, 25].

Next we address the question of the convexity of the spaces of negatively associated and negatively correlated distributions. We show that these spaces are not convex for distributions on $\mathbb{R}^n$, and they are similarly non-convex when restricted to the Boolean cube. We then address whether or not these spaces are connected in the weak or total variation topology.

2. THE TOPOLOGY OF $\mathcal{M}_{\text{NC}}$ AND $\mathcal{M}_{\text{NA}}$

2.1. The interior of $\mathcal{M}_{\text{NC}}(\mathbb{R}^n)$ and $\mathcal{M}_{\text{NA}}(\mathbb{R}^n)$. The interiors of both the space of negatively correlated and negatively associated distributions are intimately connected to their *strict* counterparts. Recall that a distribution μ on $\mathbb{R}^n$ is *strictly NC* if

$$\text{Cov}_\mu(x_i, x_j) < 0$$

for all $1 \leqslant i < j \leqslant n$, and a distribution μ on $\mathbb{R}^n$ is *strictly NA* if

$$\text{Cov}_\mu(f(x_I), g(x_J)) < 0$$

for all strictly non-decreasing (not almost surely constant) f, g and disjoint $I, J \subset \{1, \ldots, n\}$. If a distribution μ is strictly NA then it must be strictly NC. It is not *a priori* clear that strictly NA distributions even exist. However, we prove below in Lemma 2.1 that they indeed exist on the Boolean cube $\{0, 1\}^n$ (and thus by extension on $\mathbb{R}^n$).

We note that here and elsewhere, the notion of a strictly non-decreasing function f means by implication that f is *strictly* non-decreasing, i.e. that it is not essentially constant with respect to the measure μ under consideration. We define the total variation of a non-decreasing function f with respect to a measure μ to be $\sup f - \inf f$, where sup denotes essential sup (i.e. modulo sets of measure 0) and inf denotes essential inf. Note however that on the cube these two notions (sup and essential sup) coincide for non-decreasing and non-increasing functions, as do inf and essential inf.

LEMMA 2.1. *There exist strictly NA distributions μ on $\{0, 1\}^n$ such that for all non-decreasing $f(x_I), g(x_J)$ (with I, J disjoint sets of indices) having total variation 1 on $\{0, 1\}^n$, there is an $\epsilon > 0$ such that*

$$\int fg \, d\mu \leqslant \int f \, d\mu \int g \, d\mu - \epsilon.$$

Consequently, there exist strictly NA distributions on the whole of $\mathbb{R}^n$ satisfying the above inequality (under the same measure μ supported on $\{0, 1\}^n$ viewed as a subset of $\mathbb{R}^n$).

Proof. Let $I_{n,1}$ denote the collection of vectors $(x_1, \ldots, x_n) \in \{0, 1\}^n$ such that $\sum_i x_i = 1$, and let μ be any probability distribution supported on $I_{n,1}$. Thus for some $\epsilon > 0$, we have $\mu(x) > \sqrt{\epsilon}$ for all $x \in I_{n,1}$. Now assume that, as stated, $f(x_I)$ and $g(x_J)$ are non-decreasing functions of total variation 1, with I and J disjoint subsets of $1, \ldots, n$. First note that to check the condition

$$\text{Cov}_\mu(f(x_I), g(x_J)) < 0,$$

it suffices to assume that $f(0, \ldots, 0) = 0$ and $g(0, \ldots, 0) = 0$. Indeed, we may replace $f(x_I)$ with $f(x_I) - f(0, \ldots, 0)$ and $g(x_J)$ with $g(x_J) - g(0, \ldots, 0)$ without

changing $\mathrm{Cov}_\mu(f(x_I), g(x_J))$. It follows that

$$\mathbf{E}_\mu[f(x_I)g(x_J)] = 0,$$

since for x in the support of μ, we have $\sum_i x_i = 1$ and so necessarily $x_I = (0, \ldots, 0)$ or $x_J = (0, \ldots, 0)$. On the other hand, since f and g are strictly non-decreasing (i.e. non-constant), zero at the zero vector, and of total variation 1, each must equal 1 at one or more points in the support of our measure, i.e., in $I_{n,1}$. Thus $\mathbf{E}_\mu[f(x_I)] > \sqrt{\epsilon}$ and $\mathbf{E}_\mu[g(x_J)] > \sqrt{\epsilon}$, and so

$$\mathbf{E}_\mu[f(x_I)g(x_J)] = 0 < (\sqrt{\epsilon})^2 = \epsilon \leqslant \mathbf{E}_\mu[f(x_I)]\mathbf{E}_\mu[g(x_J)].$$

Therefore μ is strictly negatively associated satisfying the lower ϵ-bound in the statement of the theorem. We note that since μ satisfies this bound as a measure on the cube I_n, it also satisfies this bound when viewed as a measure on $\mathbb{R}^n$ (concentrated on $I_n \subset \mathbb{R}^n$). ∎

We now move to the main results of this section. We begin by considering the weak interior (i.e. interior in the weak topology on measures) of the space of negatively associated distributions on $\mathbb{R}^n$ or on any fixed open subset $G \subset \mathbb{R}^n$. The following proposition shows that the weak interior of the NA distributions on G is in fact empty.

PROPOSITION 2.1. *Consider the space of probability distributions supported on a fixed open set $G \subset \mathbb{R}^n$ (or all of $\mathbb{R}^n$). If μ is strictly NA on G, then every weak neighborhood of μ contains a non-NA distribution ν.*

Proof. Let μ be strictly NA and let ϵ and continuous bounded $f_1, \ldots, f_k$ be given. Then ν' will be in the weak neighborhood $V_\mu(\epsilon; f_1, \ldots, f_k)$ if and only if

$$\left| \int f_i \, d\mu - \int f_i \, d\nu' \right| \leqslant \epsilon$$

for every $i = 1, \ldots, k$. We construct a non-negatively associated distribution ν' in the weak neighborhood $V_\mu(\epsilon; f_1, \ldots, f_k)$. This is done by way of a discrete distribution ν' satisfying $\int f_i \, d\mu = \int f_i \, d\nu'$ for each $i = 1, \ldots, k$, but which itself is not negatively associated.

Let ν be a positively associated distribution on G such that

$$\int f_{k+1} f_{k+2} \, d\nu - \int f_{k+1} \, d\nu \int f_{k+2} \, d\nu > 0$$

for some non-decreasing f_{k+1}, f_{k+2}, and append the latter two functions to the above sequence, yielding $f_1, \ldots, f_{k+2}$. (See [13] for existence theorems and examples of positively associated measures.) Assume without loss of generality that

$f_1, \ldots, f_{k+1}, f_{k+2}, f_{k+1}f_{k+2}$ are linearly independent. For each x consider the vector

$$\mathbf{f}_x = (f_1(x), \ldots, f_k(x), f_{k+1}(x), f_{k+2}(x), f_{k+1}(x)f_{k+2}(x)) \in \mathbb{R}^{k+3},$$

with the last entry the product of f_{k+1} and f_{k+2}. The collection $\{\mathbf{f}_x\}_{x \in G}$ spans $\mathbb{R}^{k+3}$. Thus we can find $\alpha_1, \ldots, \alpha_{k+3}$ and $x_1, \ldots, x_{k+3}$ such that

$$(\mu(f_1), \ldots, \mu(f_k), \nu(f_{k+1}), \nu(f_{k+2}), \nu(f_{k+1}f_{k+2})) = \sum_{j=1}^{k+3} \alpha_j \mathbf{f}_{x_j},$$

where $\mu(f_i) := \int f_i \, d\mu$. Therefore

$$\mu(f_i) = \sum_{j=1}^{k+3} \alpha_j f_i(x_j)$$

for each $i = 1, \ldots, k$, and

$$\nu(f_{k+1}) = \sum_{j=1}^{k+3} \alpha_j f_{k+1}(x_j),$$

$$\nu(f_{k+2}) = \sum_{j=1}^{k+3} \alpha_j f_{k+2}(x_j),$$

$$\nu(f_{k+1}f_{k+2}) = \sum_{j=1}^{k+3} \alpha_j f_{k+1}(x_j)f_{k+2}(x_j).$$

So the discrete distribution $\nu' = \sum_{j=1}^{k+3} \alpha_j \delta_{x_j}$ is in the weak neighborhood

$$V_\mu(\epsilon; f_1, \ldots, f_k),$$

but it is not negatively associated. ∎

The study of the interior of the set of negatively associated distributions is complicated by the fact that the covariance condition for negative association must be checked on infinitely many functions (in order to establish negative association for a single measure). This situation can however be avoided when the distribution is supported on a finite subset of $\mathbb{R}^n$ (by a compactness argument in Section 2.2 below). On a finite product probability space X^n, the space of probability distributions is finite-dimensional, and we may conclude as will be done in Section 2.2 that the interior of the collection of NA distributions is non-empty. Note that since the set of distributions on a finite space is finite-dimensional, the two topologies (weak and total variation) discussed here coincide.

On the other hand, a probability measure is negatively correlated (on $\mathbb{R}^n$) if finitely many covariance conditions (1.1) are satisfied. Because of this, we may prove that the weak interior of the class of NC distributions is non-empty in the space of distributions supported on a fixed compact set $X \subset \mathbb{R}^n$.

PROPOSITION 2.2. *Let $X \subset \mathbb{R}^n$ be a compact subset. Then $\mathcal{M}_{\mathrm{NC}}(X) \subset \mathcal{M}(\mathbb{R}^n)$ has a non-empty interior in the weak topology.*

Proof. It suffices to show that $\int x_i x_j \, d\mu - \int x_i \, d\mu \int x_j \, d\mu$ is continuous in μ, which follows from the fact that the functions $f_{ij}(x) = x_i x_j$, $1 \leqslant i, j \leqslant n$, and $f_k(x) = x_k$, $1 \leqslant k \leqslant n$, are bounded. ∎

This result however does not hold when we consider the class of negatively correlated distributions supported on all of $\mathbb{R}^n$.

PROPOSITION 2.3. *The collection of negatively correlated distributions on $\mathbb{R}^n$ has no interior in the total variation topology (and hence in the weak topology).*

Proof. Let μ be a negatively correlated distribution. Consider the distribution $\nu_c = \frac{1}{2}(\delta_{-c\mathbf{1}} + \delta_{c\mathbf{1}})$ (a sum of point masses at two points), where $c > 0$ is large and $\mathbf{1} = (1, \ldots, 1)$. We note that ν_c has total variation 1. We claim that for any neighborhood V of μ (in the TV metric), there is a distribution in V that is positively correlated, of the form $\mu_{\alpha,c} = \alpha\mu + (1-\alpha)\nu_c$ for some $\alpha \in (0,1)$ close to 1 and $c > 0$.

The idea here is that the distribution ν_c has a positive correlation that is arbitrarily large as c becomes large, so that adding only a small multiple $(1-\alpha)\nu_c$ (if c is large) will cause a distribution to become positively correlated.

Note first that

$$\begin{aligned}
\|\mu - \mu_{\alpha,c}\|_{\mathrm{TV}} &= \|(1-\alpha)\mu + (\alpha-1)\nu_c\|_{\mathrm{TV}} \\
&\leqslant (1-\alpha)\|\mu\|_{\mathrm{TV}} + (1-\alpha)\|\nu_c\|_{\mathrm{TV}} = 2(1-\alpha),
\end{aligned}$$

which is arbitrarily small for α close to 1 (uniformly in c). Hence, uniformly in c, the measure $\mu_{\alpha,c}$ is in V for α sufficiently close to 1, which we assume is the case. However, for this value of α, we now allow c to grow larger. Note that the covariance equals

$$\begin{aligned}
\int x_i x_j \, d\mu_{\alpha,c} &= \int x_i x_j \, d[\alpha\mu + (1-\alpha)\nu_c] \\
&= \alpha \int x_i x_j \, d\mu + \tfrac{1}{2}(1-\alpha) \int x_i x_j \, d[\delta_{-c\mathbf{1}} + \delta_{c\mathbf{1}}] \\
&= \alpha \int x_i x_j \, d\mu + \tfrac{1}{2}(1-\alpha)[c^2 + c^2],
\end{aligned}$$

which for sufficiently large c is clearly positive. Thus any TV neighborhood V of μ has a positively correlated distribution in it. ∎

As a consequence of Proposition 2.3 we have

PROPOSITION 2.4. *The space $\mathcal{M}_{\mathrm{NA}}$ on all of $\mathbb{R}^n$ has an empty interior with respect to both the TV and weak topologies.*

However, on a compact set $X \subset \mathbb{R}^n$, the space $\mathcal{M}_{\mathrm{NA}}(X)$ has a non-empty TV interior:

PROPOSITION 2.5. *Let $X \subset \mathbb{R}^n$ be a compact subset. Then $\mathcal{M}_{\mathrm{NA}}(X)$ has a non-empty interior with respect to the total variation metric.*

Proof. It is not hard to show that if $\|\mu - \nu\|_{\mathrm{TV}} < \epsilon$, then we have $|\mathrm{Cov}_\mu(f, g) - \mathrm{Cov}_\nu(f, g)| \leqslant 3\epsilon$ for all f, g satisfying $\|f\|_\infty, \|g\|_\infty \leqslant 1$.

According to Lemma 2.1, let μ be strictly NA so that $\mathrm{Cov}_\mu(f, g) < -\epsilon$ for all strictly non-decreasing $f(x_I), g(x_J)$ of total variation 1 defined on disjoint index subsets $I, J \subset \{1, \dots, n\}$, and some $\epsilon > 0$. Choose μ' such that $\|\mu - \mu'\|_{\mathrm{TV}} < \epsilon/6$. Then

$$(2.1) \qquad \mathrm{Cov}_{\mu'}(f, g) < \epsilon/2 + \mathrm{Cov}_\mu(f, g) < -\epsilon/2 < 0.$$

For general f and g (not of total variation 1), (2.1) holds by multiplying these by constants, without changing the negative covariance. This completes the proof. ∎

For a compact X, since $\mathcal{M}_{\mathrm{NA}}(X) \subset \mathcal{M}_{\mathrm{NC}}(X)$, it immediately follows that the interior of the collection of NC distributions is non-empty in the total variation topology (this also follows from the fact that it is non-empty in the weak topology):

COROLLARY 2.1. *For a compact $X \subset \mathbb{R}^n$, $\mathcal{M}_{\mathrm{NC}}(X)$ has a non-empty interior with respect to the total variation metric.*

2.2. $\mathcal{M}_{\mathrm{NC}}$ and $\mathcal{M}_{\mathrm{NA}}$ on the Boolean cube. We reformulate the conditions for negative correlation and negative association on the Boolean cube $I_n = \{0, 1\}^n$ as polynomial inequalities. We consider a compactness argument in the case of strict negative association in order to get a handle on the infinity of conditions contained within definition (1.3). Restricting these measures to the Boolean cube affords great flexibility due to the topological properties of both the space of probability measures and the space of continuous functions on the said cube.

Denote by $\mu^{(i)}$, $i = 1, \dots, n$, and $\mu^{(i,j)}$, $1 \leqslant i, j \leqslant n$, respectively, the one- and two-dimensional marginals of μ. On the Boolean cube $I_n = \{0, 1\}^n$ we have $\mathbf{E}x_i = \mu^{(i)}(1)$ for each $i = 1, \dots, n$, and $\mathbf{E}x_i x_j = \mu^{(i,j)}(1, 1)$ for all $1 \leqslant i, j \leqslant n$. The condition of negative correlation therefore reduces to

$$(2.2) \qquad \mu^{(i,j)}(1, 1) \leqslant \mu^{(i)}(1)\mu^{(j)}(1), \quad \forall 1 \leqslant i, j \leqslant n.$$

Any probability measure μ on the Boolean cube is uniquely determined by a vector of length 2^n, $\mu = (\mu_1, \dots, \mu_{2^n})$, such that $\sum_i \mu_i = 1$. Thus (2.2) may be written in the form

$$(2.3) \qquad p_\mu(\mu_1, \dots, \mu_{2^n}) \leqslant 0,$$

where p_μ is a polynomial in $\mu_1, \dots, \mu_{2^n}$.

Now consider definition (1.3) of negative association, specialized to the Boolean cube. Denote by I and J disjoint subsets of indices in $\{1, \dots, n\}$, and by x_I, x_J vectors restricted to the indices of I and J respectively. Further, let

$\mu^{(I)}, \mu^{(J)}$ denote the respective marginal distributions as defined in (1.2). Then the condition of negative association may be written

$$\sum_{x_I, x_J} f(x_I)g(x_J)\mu(x_I, x_J) \leqslant \sum_{x_I, x_J} f(x_I)g(x_J)\mu^{(I)}(x_I)\mu^{(J)}(x_J)$$

or

$$(2.4) \qquad \sum_{x_I, x_J} f(x_I)g(x_J)\big(\mu(x_I, x_J) - \mu^{(I)}(x_I)\mu^{(J)}(x_J)\big) \leqslant 0.$$

As in (2.3), this may be reformulated as

$$(2.5) \qquad p_{f,g}(\mu_1, \ldots, \mu_{2^n}) \leqslant 0,$$

where $p_{f,g}(\mu_1, \ldots, \mu_{2^n})$ is a polynomial in $\mu_1, \ldots, \mu_{2^n}$ depending on f and g.

Inequality (2.2) must hold for $1 \leqslant i, j \leqslant n$, which is a finite number of constraints. Inequality (2.4) must hold for all non-decreasing f and g, and disjoint index sets I and J – an infinite number of constraints. Let us however restrict our attention to the set of all *strictly* NA distributions, i.e. those μ for which strict inequality holds in (2.4):

$$\mathrm{Cov}_\mu(f(x_I), g(x_J)) < 0$$

for all monotone f and g. Multiplying (2.4) by a constant, we may assume such f and g are uniformly bounded. Note also that every function $f : \{0,1\}^n \to \mathbb{R}$ is a polynomial of bounded degree. Consider the space of non-decreasing, uniformly bounded polynomials of degree at most n (in the space of continuous functions $C(\{0,1\}^n)$ equipped with the ∞-norm). Such polynomials of a finite degree compose a finite-dimensional space. Moreover, the supremum norm on this space is equivalent to the supremum norm on the coefficients, which is equivalent to the equicontinuity of the space. Thus by the Arzelà–Ascoli theorem, this space is compact in the ∞-norm. Thus given a covering of the space by balls in the ∞-norm, we may reduce to a finite subcover, and in choosing the centers of these balls, we may reduce a covariance condition as follows. As the covariance operator is continuous, there exists $\epsilon > 0$ and finitely many $f_1, \ldots, f_m$ and $g_1, \ldots, g_m$ such that if

$$\mathrm{Cov}_\mu(f_i(x_I), g_i(x_J)) < -\epsilon$$

for all $i, j = 1, \ldots, m$, then

$$\mathrm{Cov}_\mu(f(x_I), g(x_J)) < 0$$

for all f, g non-decreasing. Or in the language of (2.5), with the same $\epsilon > 0$ and $f_1, \ldots, f_m$ and $g_1, \ldots, g_m$ such that

$$(2.6) \qquad p_{f_i, g_i}(\mu_1, \ldots, \mu_{2^n}) < -\epsilon, \quad i = 1, \ldots, m,$$

we have

$$\mathrm{Cov}_\mu(f(x_I), g(x_J)) < 0$$

for all f, g non-decreasing.

From the viewpoint of (2.2) and (2.6), the conditions of strict negative correlation and strict negative association on the Boolean cube are continuous in the parameters $\mu_1, \ldots, \mu_{2^n}$ of a given distribution μ. That is, the condition will still be satisfied under small perturbations of $\mu_1, \ldots, \mu_{2^n}$.

Of course, the space of probability measures on the Boolean cube is finite-dimensional, and all Hausdorff vector topologies on a finite-dimensional space are equivalent. Thus one may define basic open sets by (1.8) or by (1.9). Or, equivalently, one may choose the Euclidean topology induced by the coordinate system $\mu = (\mu_1, \ldots, \mu_{2^n})$. As the conditions defining both the class of NC and NA distributions are continuous in this Euclidean topology, we moreover obtain

THEOREM 2.1. *Let $\mathcal{M}_{\mathrm{NC}}$ and $\mathcal{M}_{\mathrm{NA}}$ denote the spaces of NC and NA distributions on the Boolean cube. We have*

$$\partial\mathcal{M}_{\mathrm{NC}} \subset \{\mu \in \mathcal{M}_{\mathrm{NC}} : \mu^{(i,j)}(1,1) = \mu^{(i)}(1)\mu^{(j)}(1) \text{ for some } i, j\}$$

and

$$\partial\mathcal{M}_{\mathrm{NA}} \subset \{\mu \in \mathcal{M}_{\mathrm{NA}} : \exists f, g \text{ non-constant, non-decreasing,}$$
$$\mathrm{Cov}_\mu(f(x_I), g(x_J)) = 0 \text{ for some } I, J\},$$

where I and J are disjoint subsets of $\{1, \ldots, n\}$. Moreover, the interior of $\mathcal{M}_{\mathrm{NC}}$ and the interior of $\mathcal{M}_{\mathrm{NA}}$ are non-empty.

2.3. Convexity and connectedness. We further our study of the topological properties of the spaces of negatively associated and negatively correlated distributions by considering properties of convexity and connectedness. We consider such questions on both the Boolean cube and on all of $\mathbb{R}^n$.

2.3.1. *Convexity properties of the space of negatively associated distributions*

THEOREM 2.2. *The space of negatively associated distributions on $\mathbb{R}^n$ is not convex.*

Proof. We consider strictly negatively associated distributions μ and ν, which exist by Lemma 2.1. We show that there exist increasing functions f, g and a $\lambda \in (0, 1)$ for which the condition of negative association fails for the measure $\lambda\mu + (1 - \lambda)\nu$.

We begin with a general algebraic manipulation. Given increasing f and g defined on disjoint index sets, there exist ϵ_1 and ϵ_2 such that

$$\int fg \, d\mu = \int f \, d\mu \int g \, d\mu - \epsilon_1,$$
$$\int fg \, d\nu = \int f \, d\nu \int g \, d\nu - \epsilon_2.$$

Setting $A := \int f\, d\mu$, $B := \int f\, d\nu$, $C := \int g\, d\mu$, and $D := \int g\, d\nu$, it follows that

$$
\begin{aligned}
\int fg\, d(\lambda\mu + (1-\lambda)\nu) &= \lambda \int fg\, d\mu + (1-\lambda) \int fg\, d\nu \\
&= \lambda \int f\, d\mu \int g\, d\mu + (1-\lambda) \int f\, d\nu \int g\, d\nu \\
&\quad - (\lambda\epsilon_1 + (1-\lambda)\epsilon_2) \\
&= \lambda AC + (1-\lambda)BD - (\lambda\epsilon_1 + (1-\lambda)\epsilon_2).
\end{aligned}
$$

Further, we have

$$
\begin{aligned}
\int f\, d(\lambda\mu + (1-\lambda)\nu) &\int g\, d(\lambda\mu + (1-\lambda)\nu) \\
&= (\lambda A + (1-\lambda)B)(\lambda C + (1-\lambda)D) \\
&= \lambda^2 AC + \lambda(1-\lambda)AD + \lambda(1-\lambda)BC + (1-\lambda)^2 BD.
\end{aligned}
$$

The condition of convexity therefore becomes

$$
\begin{aligned}
\lambda AC + (1-\lambda)BD - (\lambda\epsilon_1 + (1-\lambda)\epsilon_2) \\
\leqslant \lambda^2 AC + \lambda(1-\lambda)AD + \lambda(1-\lambda)BC + (1-\lambda)^2 BD
\end{aligned}
$$

for $0 \leqslant \lambda \leqslant 1$. Simplifying, we obtain

$$
\lambda^2 (A-B)(C-D) - \lambda(A-B)(C-D) \geqslant -(\lambda\epsilon_1 + (1-\lambda)\epsilon_2).
$$

Upon setting $\tilde{C}_1 = (A-B)(C-D)$ this becomes

$$
(2.7) \qquad\qquad \tilde{C}_1\lambda^2 - \tilde{C}_1\lambda \geqslant -(\lambda\epsilon_1 + (1-\lambda)\epsilon_2).
$$

Thus we must show that (2.7) fails for certain increasing f, g and $\lambda \in (0,1)$. If $\tilde{C}_1 < 0$, then the quadratic $\tilde{C}_1\lambda^2 - \tilde{C}_1\lambda = \tilde{C}_1\lambda(\lambda - 1)$ is non-negative for all $0 \leqslant \lambda \leqslant 1$, thus satisfying (2.7). However, we claim that if $\tilde{C}_1 > 0$, then (2.7) will not hold for certain $\lambda \in (0,1)$, as is shown below (this would complete the proof of non-convexity).

To this end, we will need to guarantee $\tilde{C}_1 > 0$; this can be accomplished by choosing μ and ν in such a way that $A = \int f\, d\mu > \int f\, d\nu = B$, and $C = \int g\, d\mu > \int g\, d\nu = D$.

Given real numbers $p_1, \ldots, p_n$ we can translate the mean of a probability measure μ in each variable by p_i without changing the covariance structure of μ. Specifically, map $\mu \mapsto \mu \circ T^{-1}$, where the transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$
T(x_1, \ldots, x_n) = (x_i + p_i)_{i=1,\ldots,n} = (y_i)_{i=1,\ldots,n}.
$$

By the change of variables formula, for any integrable f,

$$
\int_{\mathbb{R}^n} f(y)\, d(\mu \circ T^{-1})(y) = \int_{\mathbb{R}_n} f(Tx)\, d\mu(x)
$$

In particular,

$$(2.8) \qquad \int\limits_{\mathbb{R}^n} y_i \, d(\mu \circ T^{-1})(y) = \int\limits_{\mathbb{R}^n} x_i \, d\mu(x) + p_i.$$

Thus a short calculation shows that $\mathrm{Cov}_{\mu \circ T^{-1}}(y_i, y_j) = \mathrm{Cov}_\mu(x_i, x_j)$. What is more, since T is mere translation, it preserves the product ordering on $\mathbb{R}^n$, that is, $x \geqslant y$ if and only if $Tx \geqslant Ty$. Therefore a function f is non-decreasing on $\mathbb{R}^n$ if and only if $f \circ T$ is non-decreasing on $\mathbb{R}^n$, whence μ is negatively associated if and only if $\mu \circ T^{-1}$ is negatively associated. We may thus assume that $A = \int f \, d\mu > \int f \, d\nu = B$, and $C = \int g \, d\mu > \int g \, d\nu = D$. Thus $\tilde{C}_1 > 0$. In this case the quadratic $\tilde{C}_1 \lambda^2 - \tilde{C}_1 \lambda$ is bounded above by 0 for all $0 \leqslant \lambda \leqslant 1$, and its minimum value is attained at $\lambda = 1/2$. If convexity is to hold, then (2.7) must be valid when $\lambda = 1/2$. Setting $\lambda = 1/2$ in (2.7) we obtain

$$\frac{\tilde{C}_1}{4} \leqslant \frac{1}{2}(\epsilon_1 + \epsilon_2).$$

This is evidently false if ϵ_1 and ϵ_2 can be made arbitrarily small independent of $\tilde{C}_1$, or $\tilde{C}_1$ can be made large independent of ϵ_1 and ϵ_2.

We can alter the value of $\tilde{C}_1$ via a translation operator T independent of ϵ_1 and ϵ_2. Indeed, as was also considered in the proof of Lemma 2.1, we may replace $f(x_I)$ with $f(x_I) - c_1$ for some constant c_1, and $g(x_J)$ with $g(x_J) - c_2$ for some constant c_2, without changing $\mathrm{Cov}_\mu(f(x_I), g(x_J)) = -\epsilon_1$. The same can be said when considering ν and ϵ_2. Thus, $\tilde{C}_1$ can be made large independent of ϵ_1 and ϵ_2. This proves the result.

Thus it is shown that the space of negatively associated distributions on $\mathbb{R}^n$ is non-convex. ∎

2.3.2. *Non-convexity of* $\mathcal{M}_{\mathrm{NC}}(\mathbb{R}^n)$ Define the sets

$$(2.9) \qquad E_{p_1,\ldots,p_n} := \{\mu \in \mathcal{M}_{\mathrm{NC}}(\mathbb{R}^n) : \mathbf{E}_\mu x_i = p_i, \ i = 1, \ldots, n\}.$$

COROLLARY 2.2. *The space of negatively correlated distributions on $\mathbb{R}^n$ is not convex. However, for any fixed $p_1, \ldots, p_n \in \mathbb{R}$, the collection $E_{p_1,\ldots,p_n}$ of measures is convex.*

Proof. The non-convexity follows as in the proof of Theorem 2.2. Specifically, given strictly negatively correlated distributions μ and ν, fix i, j and note that

$$\int x_i x_j \, d\mu = \int x_i \, d\mu \int x_j \, d\mu - \epsilon_1$$

for some $\epsilon_1 > 0$, and

$$\int x_i x_j \, d\nu = \int x_i \, d\nu \int x_j \, d\nu - \epsilon_2$$

for some $\epsilon_2 > 0$. Set $A = \int x_i \, d\mu$, $B = \int x_i \, d\nu$, $C = \int x_j \, d\mu$, and $D = \int x_j \, d\nu$. Then if convexity is to hold, we once again must have

$$\lambda AC + (1 - \lambda)BD - (\lambda\epsilon_1 + (1 - \lambda)\epsilon_2)$$
$$\leqslant \lambda^2 AC + \lambda(1 - \lambda)AD + \lambda(1 - \lambda)BC + (1 - \lambda)^2 BD$$

for $0 \leqslant \lambda \leqslant 1$. Now with $\tilde{C}_2 = (A - B)(C - D)$ this simplifies to

$$(2.10) \qquad \tilde{C}_2\lambda^2 - \tilde{C}_2\lambda \geqslant -(\lambda\epsilon_1 + (1 - \lambda)\epsilon_2).$$

If we can show that there exist NC distributions μ and ν such that $\tilde{C}_2 > 0$, then upon setting $\lambda = 1/2$ in (2.10), we will arrive at the condition

$$\frac{\tilde{C}_2}{4} \leqslant \frac{1}{2}(\epsilon_1 + \epsilon_2),$$

which will fail for small enough ϵ_1, ϵ_2 if we can make $\tilde{C}_2$ large enough independent of ϵ_1, ϵ_2.

Thus we must show that there exist NC μ and ν such that $\tilde{C}_2 > 0$, and such that ϵ_1 and ϵ_2 are sufficiently small. That is, $\int x_i \, d\mu > \int x_i \, d\nu$ and $\int x_j \, d\mu > \int x_j \, d\nu$, for $i \neq j$. Given real numbers $p_1, \ldots, p_n$ we can translate the mean of a probability measure μ in each variable by p_i without changing the covariance structure of μ, as described in (2.8). Specifically, map $\mu \mapsto \mu \circ T^{-1}$, where the transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$T(x_1, \ldots, x_n) = (x_i + p_i)_{i=1,\ldots,n} = (y_i)_{i=1,\ldots,n}.$$

By the change of variables, for any integrable f,

$$\int\limits_{\mathbb{R}^n} f(y) \, d(\mu \circ T^{-1})(y) = \int\limits_{\mathbb{R}_n} f(Tx) \, d\mu(x).$$

In particular,

$$\int\limits_{\mathbb{R}^n} y_i \, d(\mu \circ T^{-1})(y) = \int\limits_{\mathbb{R}^n} x_i \, d\mu(x) + p_i.$$

Thus $\mathrm{Cov}_{\mu \circ T^{-1}}(y_i, y_j) = \mathrm{Cov}_\mu(x_i, x_j)$. Thus given NC distributions μ and ν, we may always translate μ until its mean values in each coordinate, i.e. $\int x_i \, d\mu(x)$, dominate the mean values of ν in each coordinate. This does not change the covariance structure of μ, and therefore preserves negative correlation, and in particular ϵ_1 and ϵ_2.

We now prove that $E_{p_1,\ldots,p_n}$ is convex for each fixed $p_1, \ldots, p_n \in \mathbb{R}$. This follows from (2.10). Indeed, if $\tilde{C}_2 = 0$ then certainly (2.10) holds for each $0 \leqslant \lambda \leqslant 1$, and therefore the collection of NC distributions which satisfy $\tilde{C}_2 = 0$ for each $1 \leqslant i < j \leqslant n$ will be convex. We have $\tilde{C}_2 = 0$ whenever $\int x_i \, d\mu = \int x_i \, d\nu$, and therefore if $\int x_i \, d\mu = \int x_i \, d\nu = p_i$ for each $i = 1, \ldots, n$, then their convex combination $\lambda\mu + (1 - \lambda)\nu$ will be NC for each $0 \leqslant \lambda \leqslant 1$. The result follows. ∎

2.3.3. *Non-convexity of* $\mathcal{M}_{\mathrm{NC}}(I_n)$

LEMMA 2.2. *For any* $0 < \epsilon < 1$ *and* $1 \leqslant i \leqslant n$, *there exists a negatively correlated distribution on the Boolean cube* $I_n = \{0,1\}^n$ *satisfying* $\mu^{(i)}(1) = \epsilon$. *In fact, given* $0 < \epsilon_i < 1$, $i = 1, \ldots, n$, *satisfying* $\sum_i \epsilon_i = 1$, *there exist a negatively correlated distribution on* I_n *satisfying* $\mu^{(i)}(1) = \epsilon_i$, $i = 1, \ldots, n$.

Proof. Given $1 \leqslant i \leqslant n$, define a distribution as follows:

$$\mu_{i,\epsilon} := \epsilon \delta_{\mathbf{e}_i} + (1 - \epsilon) \delta_{\mathbf{e}_j},$$

where $i \neq j$ and $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0)$ is the vector with a single 1 in the ith component. We see that $\mu_{i,\epsilon}^{(i)}(1) = \epsilon$. What is more,

$$\mathbf{E}_{\mu_{i,\epsilon}} x_k x_\ell = \mu_{i,\epsilon}^{(k,\ell)}(1,1) = 0$$

for all $1 \leqslant k, \ell \leqslant n$, and thus

$$\mathbf{E}_{\mu_{i,\epsilon}} x_k x_\ell - \mathbf{E}_{\mu_{i,\epsilon}} x_k \mathbf{E}_{\mu_{i,\epsilon}} x_\ell \leqslant 0.$$

The first result follows.

Define a measure μ as the convex combination of point masses centered at each $\mathbf{e}_i$:

$$\mu := \sum_i \epsilon_i \delta_{\mathbf{e}_i}, \qquad \sum \epsilon_i = 1.$$

Evidently, for each $j = 1, \ldots, n$, $\mu^{(j)}(1) = \epsilon_j$, and for each $i \neq j$, $\mu^{(i,j)}(1,1) = 0$. Thus μ is strictly negatively correlated:

$$\mu^{(i,j)}(1,1) - \mu^{(i)}(1)\mu^{(j)}(1) = -\epsilon_i \epsilon_j.$$

Further, μ satisfies the requirements on the one-dimensional marginals. The result follows. ∎

COROLLARY 2.3. *The space of negatively correlated distributions on* I_n *is non-convex. However, for any fixed* $p_1, \ldots, p_n \in \mathbb{R}$, *the collection of measures*

$$E_{p_1,\ldots,p_n} = \{\mu \in \mathcal{M}_{\mathrm{NC}}(I_n) : \mu^{(i)}(1) = p_i, \ i = 1, \ldots, n\}$$

is convex.

Proof. We begin with strictly negatively correlated distributions μ and ν, which exist by Lemma 2.1. We derive conditions under which the convex combination $\lambda \mu + (1 - \lambda)\nu$ fails to be negatively correlated. We then produce strictly negatively correlated distributions whose convex combination fails to satisfy the above-mentioned condition.

Note that $\mathbf{E}_\mu x_i = \mu^{(i)}(1)$ and $\mathbf{E}_\mu x_i x_j = \mu^{(i,j)}(1,1)$ on the Boolean cube. Thus given strictly negatively correlated distributions μ and ν, fix i, j and note that

$$\mu^{(i,j)}(1,1) = \mu^{(i)}(1)\mu^{(j)}(1) - \epsilon_1$$

for some $\epsilon_1 > 0$, and

$$\nu^{(i,j)}(1,1) = \nu^{(i)}(1)\nu^{(j)}(1) - \epsilon_2$$

for some $\epsilon_2 > 0$. Set $A = \mu^{(i)}(1)$, $B = \nu^{(i)}(1)$, $C = \mu^{(j)}(1)$, and $D = \nu^{(j)}(1)$. Then if convexity is to hold, we once again must have (see proof of Theorem 2.2 above, where the roles of f and g are played by x_i and x_j)

$$\lambda AC + (1-\lambda)BD - (\lambda\epsilon_1 + (1-\lambda)\epsilon_2)$$
$$\leqslant \lambda^2 AC + \lambda(1-\lambda)AD + \lambda(1-\lambda)BC + (1-\lambda)^2 BD$$

for $0 \leqslant \lambda \leqslant 1$. Now with $\tilde{C}_3 = (A-B)(C-D)$ this simplifies to

$$\tilde{C}_3\lambda^2 - \tilde{C}_3\lambda \geqslant -(\lambda\epsilon_1 + (1-\lambda)\epsilon_2).$$

As this holds for $\tilde{C}_3 = 0$, the second statement of the corollary holds.

Now set $\lambda = 1/2$. We arrive at the condition

$$\frac{\tilde{C}_3}{4} \leqslant \frac{1}{2}(\epsilon_1 + \epsilon_2).$$

Our strategy is as follows. We introduce measures μ and ν as convex combinations of point masses at the standard basis vectors. We demonstrate that the above inequality is valid. We then define a perturbation of this measure under which the inequality is violated, whereby we obtain a convex combination of measures which fail to be negatively correlated.

According to Lemma 2.2, there exist strictly negatively correlated distributions with $\tilde{C}_3 > 0$. Specifically, define

$$\mu = \sum_k \beta_k \delta_{\mathbf{e}_k}, \quad \nu = \sum_k \beta'_k \delta_{\mathbf{e}_k},$$

where $\sum_k \beta_k = 1$ and $\sum_k \beta'_k = 1$. Then

$$\tilde{C}_3 = (\beta_i - \beta'_i)(\beta_j - \beta'_j),$$

where we have used that $\tilde{C}_3 = (A-B)(C-D)$ and $A = \mu^{(i)}(1)$, $B = \nu^{(i)}(1)$, $C = \mu^{(j)}(1)$, and

$$\epsilon_1 = \beta_i\beta_j, \quad \epsilon_2 = \beta'_i\beta'_j.$$

Thus the condition

$$\frac{\tilde{C}_3}{4} \leqslant \frac{1}{2}(\epsilon_1 + \epsilon_2)$$

becomes

$$\frac{(\beta_i - \beta'_i)(\beta_j - \beta'_j)}{4} \leqslant \frac{1}{2}(\beta_i\beta_j + \beta'_i\beta'_j).$$

This inequality reduces to

$$\beta_i\beta_j + \beta'_i\beta'_j + \beta_i\beta'_j + \beta'_i\beta_j \geqslant 0,$$

which holds for all non-negative reals.

Since strict negative correlation is continuous in the Euclidean parameters of the distribution, we may perturb the above convex measure as follows. Define μ as above, but perturbed with a small additional weight given to $(1, 1, 0, \ldots, 0)$: $\mu(1, 1, 0, \ldots, 0) = \epsilon$. Here $\epsilon > 0$ is small enough so that μ is still negatively correlated. Compensating this increased weight by a total decrease in the other decoupled positive weights totaling ϵ to keep normalization will still not affect negative correlation if ϵ is sufficiently small. Then $\mu^{(1,2)}(1, 1) = \epsilon$ and $\mu^{(i)}(1) = \beta_i + \epsilon$ for $i = 1, 2$. If we do the same for ν, with the same ϵ perturbation, then $\tilde{C}_3$ does not change: indeed, for $i = 1, j = 2$,

$$
\begin{aligned}
\tilde{C}_3 &= (A - B)(C - D) \\
&= (\beta_1 + \epsilon - \beta_1' - \epsilon)(\beta_2 + \epsilon - \beta_2' - \epsilon) \\
&= (\beta_1 - \beta_1')(\beta_2 - \beta_2').
\end{aligned}
$$

We may assume that $\tilde{C}_3 > 0$, by virtue of choosing $\beta_1 > \beta_1'$ and $\beta_2 > \beta_2'$. On the other hand, for $\epsilon > 0$ small enough,

$$
-\epsilon_1 \equiv \mu^{(1,2)}(1, 1) - \mu^{(1)}(1)\mu^{(2)}(1) < 0
$$

as the quantity

$$
\mu^{(1,2)}(1, 1) - \mu^{(1)}(1)\mu^{(2)}(1) = \epsilon - (\beta_1 + \epsilon)(\beta_2 + \epsilon)
$$

is continuous in ϵ, and approaches $-\beta_1\beta_2$ from above as $\epsilon \to 0$. For β_1 and β_2 small enough, the quantity

$$
\epsilon - (\beta_1 + \epsilon)(\beta_2 + \epsilon)
$$

will be positive; letting ϵ tend to 0 shows that this quantity will pass through zero, towards $-\beta_1\beta_2$. This shows that ϵ_1 will move through 0, and thus the inequality

$$
\frac{\tilde{C}_3}{4} \leqslant \frac{1}{2}(\epsilon_1 + \epsilon_2)
$$

will be violated, as consideration of ϵ_2 will be analogous.

It follows that the right hand side of the above inequality will decrease. And thus we can, by continuity, decrease the right hand side until the inequality is violated. Thus it is shown that the space of negatively correlated distributions on the Boolean cube is non-convex. ∎

2.3.4. *Non-convexity of $\mathcal{M}_{\mathrm{NA}}(I_n)$*

LEMMA 2.3. *For any $1 \leqslant i \leqslant n$ and $0 < \epsilon < 1$, there exists a negatively associated distribution μ on $I_n = \{0, 1\}^n$ satisfying $\mu^{(i)}(1) = \epsilon$.*

Proof. As in the proof of Lemma 2.1, any distribution supported on

$$
I_{n,1} = \left\{ (x_1, \ldots, x_n) \in I_n : \sum_j x_j = 1 \right\}
$$

is strictly negatively associated. Such a measure will be of the form

$$\mu = \sum_{\mathbf{x}_k \in I_{n,1}} \alpha_k \delta_{\mathbf{x}_k}$$

where $\sum_k \alpha_k = 1$. Given $1 \leqslant i \leqslant n$ and $0 < \epsilon < 1$, on setting $\alpha_i = \epsilon$, and determining the remaining coefficients by the condition $\sum_k \alpha_k = 1$, we see that $\mu^{(i)}(1) = \epsilon$. ■

COROLLARY 2.4. *The space of negatively associated distributions on I_n is non-convex.*

Proof. Following once again the proof of Theorem 2.2, we set $A := \int_{I_n} f \, d\mu$, $B := \int_{I_n} f \, d\nu$, $C := \int_{I_n} g \, d\mu$, and $D := \int_{I_n} g \, d\nu$, and obtain the same inequality dictating convexity:

$$\lambda^2(A - B)(C - D) - \lambda(A - B)(C - D) \geqslant -(\lambda\epsilon_1 + (1 - \lambda)\epsilon_2)$$

for all $0 \leqslant \lambda \leqslant 1$. Recall once again that ϵ_1 and ϵ_2 are defined by the initial assumption: Given increasing f and g defined on disjoint index sets, there exist ϵ_1 and ϵ_2 such that

$$\int fg \, d\mu = \int f \, d\mu \int g \, d\mu - \epsilon_1,$$
$$\int fg \, d\nu = \int f \, d\nu \int g \, d\nu - \epsilon_2.$$

Upon setting $\tilde{C}_4 = (A - B)(C - D)$ this becomes

(2.11) $\tilde{C}_4\lambda^2 - \tilde{C}_4\lambda \geqslant -(\lambda\epsilon_1 + (1 - \lambda)\epsilon_2).$

This condition is satisfied whenever $\tilde{C}_4 = 0$ or $\tilde{C}_4 < 0$, and fails when $\tilde{C}_4 > 0$ for small enough $\epsilon > 0$.

Following the proof of Lemma 2.1, we define strictly negatively associated distributions μ and ν. We consider once again the inequality

(2.12) $\dfrac{\tilde{C}_4}{4} \leqslant \dfrac{1}{2}(\epsilon_1 + \epsilon_2)$

relative to the context at hand. It suffices to verify that for certain f, g and measures μ, ν, the inequality is invalid. We may translate the means without changing the covariance structure, thus $\tilde{C}_1$ may be translated, or increased, independent of ϵ_1 and ϵ_2. Indeed, as was also considered in the proof of Lemma 2.1, we may replace $f(x_I)$ with $f(x_I) - c_1$ for some constant c_1, and $g(x_J)$ with $g(x_J) - c_2$, without changing $\mathrm{Cov}_\mu(f(x_I), g(x_J)) = -\epsilon_1$. The same can be said when considering ν and ϵ_2. Thus, $\tilde{C}_4$ can be made large independent of ϵ_1 and ϵ_2. This proves the result.

This shows that (2.11) fails for $\lambda = 1/2$, whereby convexity is violated for the corresponding distribution. Thus the space of negatively associated distributions on the Boolean cube is non-convex. ■

2.3.5. *Connectedness properties of the spaces of negatively correlated and negatively associated distributions*

THEOREM 2.3. *The space of negatively correlated and the space of negatively associated distributions on the Boolean cube, and on $\mathbb{R}^n$, are path connected in the weak topology.*

Proof. For any negatively associated measure μ, consider the family of measures μ_t such that for any set A, $\mu_t(A) = \mu(A/t)$; here $0 < t \leqslant 1$. Each measure μ_t is negatively associated. For $t = 0$ we define μ_0 to be the point mass at the origin. We have defined A/t to be the set of all points in A divided by the constant t. As we scale t from 1 to 0, this effectively concentrates the measure μ through this scaling into a point mass at the origin, while preserving negative association in the process. Indeed, assuming A is a ball away from the origin, the mass of the set A/t approaches zero as $t \to 0$, as the distance of A/t from the origin approaches infinity as $t \to 0$. It follows that $\mu(A/t)$ converges weakly to the point mass at the origin. This provides a path connecting any two negatively associated distributions to the point mass at 0, proving that the family is path connected in the weak topology. ∎

REFERENCES

[1] V. T. N. Anh, H.T. Nguyen, L. V. Thanh, and V. T. H. Van, *The Marcinkiewicz–Zygmund-type strong law of large numbers with general normalizing sequences*, J. Theoret. Probab. 34 (2021), 331–348.

[2] I. Bernou and F. Boukhari, *Limit theorems for dependent random variables with infinite means*, Statist. Probab. Lett. 189 (2022), art. 109563, 9 pp.

[3] H. W. Block, T. H. Savits, and M. Shaked, *Some concepts of negative dependence*, Ann. Probab. 10 (1982), 765–772.

[4] J. Borcea, P. Brändén, and T. Liggett, *Negative dependence and the geometry of polynomials*, J. Amer. Math. Soc. 22 (2009), 521–567.

[5] M. V. Boutsikas and M. V. Koutras, *A bound for the distribution of the sum of discrete associated or negatively associated random variables*, Ann. Appl. Probab. 10 (2000), 1137–1150.

[6] A. V. Bulinski and A. Shaskin, *Limit Theorems for Associated Random Fields and Related Systems*, World Sci., Singapore 2007.

[7] D. Dubhashi and D. Ranjan, *Balls and bins: a study in negative dependence*, Random Structures Algorithms 13 (1998), 99–124.

[8] S. Duppala, G. Z. Li, J. Luque, A. Srinivasan, and R. Valieva, *Concentration of submodular functions under negative dependence*, arXiv:2309.05554 (2023).

[9] M. L. Eaton, *Lectures on Topics in Probability Inequalities*, CWI Tracts in Math. 35, Amsterdam, 1982.

[10] N. Ebrahami and M. Ghosh, *Multivariate negative dependence*, Comm. Statist. Theory Methods A10 (1981), 307–337.

[11] J. D. Esary, F. Proschan, and D. W. Walkup, *Association of random variables, with applications*, Ann. Math. Statist. 44 (1967), 1466–1474.

[12] N. Etemadi, *On the strong law of large numbers for nonnegative random variables*, J. Multivariate Anal. 13 (1983), 187–193.

[13] C. M. Fortuin, J. Ginibre, and P. N. Kasteleyn, *Correlation inequalities for partially ordered sets*, Comm. Math. Phys. 22 (1971), 89–103.

[14] K. Garbe and J. Vondrák, *Concentration of Lipschitz functions of negatively dependent variables*, arXiv:1804.10084 (2018).

[15] K. Joag-Dev and F. Proschan, *Negative association of random variables, with applications*, Ann. Statist. 11 (1983), 286–295.

[16] K. Jogdeo and G. P. Patil, *Probability inequalities for certain multivariate discrete distributions*, Sankhya B 37 (1975), 158–164.

[17] S. Karlin and Y. Rinott, *Classes of orderings of measures and related correlation inequalities*, J. Multivariate Anal. 10 (1980), 499–516.

[18] M. Ledoux, *The Concentration of Measure Phenomenon*, Math. Surveys Monogr., 89, Amer. Math. Soc., Providence, RI, 2001.

[19] E. L. Lehmann, *Some concepts of dependence*, Ann. Math. Statist. 37 (1966), 1137–1153.

[20] P. Matuła, *A note on the almost sure convergence of sums of negatively dependent random variables*, Statist. Probab. Lett. 15 (1992), 209–213.

[21] C. M. Newman, *Asymptotic independence and limit theorems for positively and negatively dependent random variables*, in: Inequalities in Statistics and Probability, Inst. Math. Statist., Hayward, CA, 1984, 127–140.

[22] P. E. Oliveira, *Asymptotics for Associated Random Variables*, Springer, Berlin, 2012.

[23] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, AMS Chelsea Publ., 2005.

[24] R. Pemantle, *Towards a theory of negative dependence*, J. Math. Phys. 41 (2000), 1371–1390.

[25] R. Pemantle and Y. Peres, *Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures*, Combin. Probab. Comput. 23 (2014), 140–160.

[26] B. L. S. Prakasa Rao, *Associated Sequences, Demimartingales and Nonparametric Inference*, Springer, Basel 2012.

[27] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis*, Academic Press, San Diego, CA, 1980.

[28] K. Wang and X. Wang, *Strong convergence properties for partial sums of asymptotically negatively associated random vectors in Hilbert spaces*, Comm. Statist. Theory Methods 49 (2020), 5578–5586.

Jonathan Root
Department of Mathematics and Statistics
Boston University
Boston, MA 02215, USA
E-mail: jroot@bu.edu

Mark Kon
Department of Mathematics and Statistics
Boston University
Boston, MA 02215, USA
E-mail: mkon@bu.edu