Negative Association - Definition, Properties, and Applications

David Wajc

Last Updated: April 11, 2017

Abstract

In these notes we present the notion of Negative Association, discuss some of its useful properties, and end with some example applications. The slogan to bear in mind here is "independent, or better".

1 Negative Association - Definition

In randomized algorithms, our randomness often takes on the form of independent random variables, allowing us to apply powerful theorems concerning such variables, a prominent example being Chernoff-Hoeffding bounds. However, we can't always expect random variables we observe (or generate during the run of our algorithms) to be independent. Nonetheless, these variables may satisfy some form of negative dependence, in which case useful properties of independence may carry over. This talk focuses on one such notion of negative dependence, namely Negative Association.

Intuition Consider a set of random variables $X_1, X_2, \ldots, X_n$, satisfying the following: if a subset S of these variables is "high", then a disjoint subset T must be "low". This property can be formalized as follows.

Definition 1 (Negative Association [10, 8])**.** *A set of random variables $X_1, X_2, \ldots, X_n$ is said to be* negatively associated (NA) *if for any two disjoint index sets $I, J \subseteq [n]$ and two functions f, g both monotone increasing or both monotone decreasing, it holds*

$$\mathbb{E}[f(X_i : i \in I) \cdot g(X_j : j \in J)] \leq \mathbb{E}[f(X_i : i \in I)] \cdot \mathbb{E}[g(X_j : j \in J)].$$

In order to simplify notation later, we will think of monotone functions f and g and disjoint subsets $I, J \subseteq [n]$ as defining functions in n variables $f_I, g_J : \mathbb{R}^n \to \mathbb{R}$, applying them to vectors $\vec{X} = (X_1, X_2, \ldots, X_n)$ given by sets of (NA) random variables, and stipulate that the values of $f(\vec{X})$ and $g(\vec{X})$ be determined by disjoint subsets of the X_i. In this notation, the above definition can be restated as follows.

Definition 2 (Negative Association [10, 8])**.** *A set of random variables $X_1, X_2, \ldots, X_n$ is said to be* negatively associated *(NA) if for any two n-dimensional functions $f, g : \mathbb{R}^n \to \mathbb{R}$, depending on disjoint subsets of indices and both monotone increasing or both monotone decreasing in their respective indices, it holds*

$$\mathbb{E}[f(\vec{X}) \cdot g(\vec{X})] \leq \mathbb{E}[f(\vec{X})] \cdot \mathbb{E}[g(\vec{X})].$$

2 Useful Properties - Part 1

As a special case of the definition of NA, taking $f_i(\vec{X}) = X_i$, we find that NA variables are negatively correlated.

Corollary 1 (NA implies Negative Correlation)**.** *Let $X_1, X_2, \ldots, X_n$ be NA random variables. Then, for all $i \neq j$, the following holds: $\mathbb{E}[X_i X_j] \leq \mathbb{E}[X_i] \cdot \mathbb{E}[X_j]$. That is, $Cov(X_i, X_j) \leq 0$.*

Another useful property of NA variables is Negative Orthant Dependence (NOD), given below.

Corollary 2 (NA implies NOD)**.** *If $X_1, X_2, \ldots, X_n$ are NA random variables, then for any set of values $x_1, x_2, \ldots, x_n$ and disjoint subsets $I, J \subseteq [n]$, it holds*

$$\Pr[X_i \geq x_i,\ \forall i \in I \cup J] \leq \Pr[X_i \geq x_i,\ \forall i \in I] \cdot \Pr[X_j \geq x_j,\ \forall j \in J]$$

$$\Pr[X_i \leq x_i,\ \forall i \in I \cup J] \leq \Pr[X_i \leq x_i,\ \forall i \in I] \cdot \Pr[X_j \leq x_j,\ \forall j \in J].$$

Proof. This is a direct application of the definition, by taking f and g to be indicator random variables for the relevant events, which are clearly both monotone increasing/decreasing functions in $\vec{X}$. □

Relying on Corollary 2, we obtain the following.

Corollary 3 (Marginal Probability Bounds)**.** *For any NA variables $X_1, \ldots, X_n$ and real values $x_1, \ldots, x_n$,*

$$\Pr[\bigwedge_i X_i \geq x_i] \leq \prod_i \Pr[X_i \geq x_i] \text{ and } \Pr[\bigwedge_i X_i \leq x_i] \leq \prod_i \Pr[X_i \leq x_i].$$

The following corollary of NA will prove useful shortly.

Corollary 4. *Let $X_1, X_2, \ldots, X_n$ be NA random variables. Then, for every set of k positive monotone increasing functions $f_1, \ldots, f_k$ depending on disjoint subsets of the X_i, it holds*

$$\mathbb{E}[\prod_i f_i(\vec{X})] \leq \prod_i \mathbb{E}[f_i(\vec{X})].$$

Proof. Induction on k, relying on the observation that the product of positive monotone increasing functions is itself a positive monotone increasing function. □

Chernoff-Hoeffding Bounds - Expect the Expected. One appealing property of NA variables is the applicability of Chernoff-Hoeffding bounds to such variables, despite the variables' potential dependences.

Theorem 5 (Chernoff-Hoeffding bounds for NA variables [4])**.** *Let $X_1, X_2, \ldots, X_n$ be NA random variables with $X_i \in [a_i, b_i]$ always. Then $Y = \sum_i X_i$ satisfies Hoeffding's upper tail bound. Namely,*

$$\Pr[|Y - \mathbb{E}[Y]| \geq t] \leq 2 \exp\left(-\frac{2t^2}{\sum_i (b_i - a_i)^2}\right).$$

If $X_i \in \{0, 1\}$ always, we have

$$\Pr[Y \geq (1+\delta)\mathbb{E}[Y]] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}[Y]}$$

$$\Pr[Y \leq (1-\delta)\mathbb{E}[Y]] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mathbb{E}[Y]}.$$

Proof. For the upper tail bounds, the standard proof of Chernoff/Hoeffding's Inequality goes through almost unchanged, with one small change: for any $s > 0$, we cannot claim $\mathbb{E}[e^{\sum_i sX_i}] = \prod_i \mathbb{E}[e^{sX_i}]$, as the X_i are not independent; but, by Corollary 4 since the X_i are NA, they satisfy $\mathbb{E}[e^{\sum_i sX_i}] = \mathbb{E}[\prod_i e^{sX_i}] \leq \prod_i \mathbb{E}[e^{sX_i}]$, which suffices to prove the desired bounds. For the lower tail bounds, we rely on $\{b_i - X_i\}_i$ being NA, a fact which we do not prove here, but will follow from Lemma 9, Proposition ii, below. □

3 Some example NA distributions

Knowing a joint distribution is NA may be useful, but do any such distributions even exist? In this section we present a few such joint distributions on n variables which are NA.

3.1 Toy Example - Independent Random Variables

Observation 6. *[([8])] Let $X_1, X_2, \ldots, X_n$ be independent random variables. Then $X_1, X_2, \ldots, X_n$ are NA.*

Ok, this example is a little unimpressive. Let's consider some more interesting examples.

3.2 More Interesting Examples - The 0-1 Principle

Lemma 7. *[The Zero-One Principle ([4])] Let $X_1, X_2, \ldots, X_n$ be zero-one random variables such that $\sum_i X_i \leq 1$ always. Then $X_1, X_2, \ldots, X_n$ are NA.*

Proof. Consider two monotonically increasing functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ depending on disjoint subsets of indices. We want to show that $\mathbb{E}[f(\vec{X}) \cdot g(\vec{X})] \leq \mathbb{E}[f(\vec{X})] \cdot \mathbb{E}[f(g(\vec{X})]$. WLOG, $f(\vec{0}) = g(\vec{0}) = 0$.[1] Now, for f and g monotone n-dimensional functions determined by disjoint subsets of $[n]$ and satisfying $f(\vec{0}) = g(\vec{0}) = 0$ we have, as $\vec{X} \in \{0,1\}^n$ and $\sum_i X_i \leq 1$, that

$$\mathbb{E}[f(\vec{X}) \cdot g(\vec{X})] = 0 \leq \mathbb{E}[f(\vec{X})] \cdot \mathbb{E}[g(\vec{X})]$$

□

3.3 More Interesting Examples - Permutation Distributions

Many problems involving permutations claim some form of negative dependence, without proof. The following is a simple proof implying many such claims.

Lemma 8. *[Permutation Distributions are NA ([8])] Let $x_1 \leq x_2 \leq \cdots \leq x_n$ be n values and let $X_1, X_2, \ldots, X_n$ be random variables such that $\{X_1, X_2, \ldots, X_n\} = \{x_1, x_2, \ldots, x_n\}$ always, with all possible assignments equally likely. Then $X_1, X_2, \ldots, X_n$ are NA.*

Proof. The proof is by induction on n. For $n = 2$ this is easy to check or prove similarly to the 0-1 Lemma. The condition $\mathbb{E}[f(\vec{X} \cdot g(\vec{X}))] \leq \mathbb{E}[f(\vec{X})] \cdot \mathbb{E}[g(\vec{X})]$ can be written more succinctly as $Cov(f(\vec{X}), g(\vec{X}) \leq 0$ (indeed, this is exactly how [8, 10] define NA). For the inductive step, we use two properties of covariances: (1) Let X, Y, Z be three random variables. Then, $Cov(X,Y) = \mathbb{E}[Cov(X,Y) \mid Z] + Cov(\mathbb{E}[X \mid Z], \mathbb{E}[Y \mid Z])$.[2][3] (2) Chebyshev's (less famous) Inequality: Let X be a random variable and f and g be monotone increasing and decreasing functions, respectively; then, $Cov(f(X), g(X)) \leq 0$.

Now, to show that permutation distributions are NA for general n, we consider two monotone increasing functions f_1 and f_2 defined over disjoint subsets of the indices, S_1 and S_2. First, as we are dealing with permutations and all permutations have equal probability, we may assume that f_1 and f_2 are permutation invariant. Next, let I be an indicator for the location of the smallest value of the x_i's, x_1. I takes on all values in $[n]$ with equal probability. By (1) above, we have

$$Cov(f_1(\vec{X}), f_2(\vec{X})) = \mathbb{E}[Cov(f_1(\vec{X}), f_2(\vec{X}) \mid I)] + Cov(\mathbb{E}[f_1(\vec{X}) \mid I], \mathbb{E}[f_2(\vec{X}) \mid I])$$

Conditioning on the location of the smallest value's place leaves us with $n-1$ variables in random order, or in other words a permutation distribution on $n-1$ elements. By the inductive hypothesis, the first term of the rhs is at most zero. Next, by the permutation-invariance of f_1 and f_2, we find that $\mathbb{E}[f_1(\vec{X} \mid I)]$ takes on exactly two values, corresponding to whether or not $I \in S_1$ (that is, some $i \in S_1$ satisfies $X_i = x_1$). In particular, this value is smaller when $I \in S_1$ and larger otherwise. Therefore $\mathbb{E}[f_1(\vec{X}) \mid I]$ and $\mathbb{E}[f_2(\vec{X}) \mid I]$ are respectively monotone increasing and decreasing functions of a binary random variable, namely $\mathbb{1}[I \in S_1]$. By Chebyshev's Inequality the second term on the rhs is at most zero, too. □

[1]Otherwise, the functions $f'(\vec{X}) = f(\vec{X}) - f(\vec{0})$ and $g'(\vec{X}) = g(\vec{X}) - g(\vec{0})$ vanish at $\vec{0}$ and satisfy the required inequality if and only if f and g do. That is, f' an g' depend on disjoint subsets of indices if and only if f and g do and f' and g' also satisfy $\mathbb{E}[f'(\vec{X}) \cdot g'(\vec{X})] \leq \mathbb{E}[f'(\vec{X})] \cdot \mathbb{E}[g'(\vec{X})]$ if and only if $\mathbb{E}[f(\vec{X}) \cdot g(\vec{X})] \leq \mathbb{E}[f(\vec{X})] \cdot \mathbb{E}[g(\vec{X})]$.

[2]Throughout these notes, for discrete variables X, Z we denote by $\mathbb{E}[Z \mid X]$ the conditional expectation of Z given X, which is itself a random variable which for all $x \in support(X)$ takes on value $\mathbb{E}[Z \mid X = x]$ with probability $\Pr[X = x]$, and similarly for continuous X and Z.

[3]See, e.g. `https://en.wikipedia.org/wiki/Law_of_total_covariance` for a proof of this identity.

4 Closure Properties

An attractive feature of NA is that NA can be proven to hold for sets of random variables if they can be obtained by sets of NA variables (e.g., the distributions discussed in Section 3) using the following operations. This allows for "calculation-free" proofs of NA.

Lemma 9 (NA closure properties [10, 8, 4])**.**

(i) *The union of independent sets of NA random variables is NA.*
That is, if $X_1, X_2, \ldots, X_n$ are NA, $Y_1, Y_2, \ldots, Y_m$ are NA, and $\{X_i\}_i$ are independent of $\{Y_j\}_j$, then $X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m$ are NA.

(ii) *Concordant monotone functions defined on disjoint subsets of a set of NA random variables are NA.*
That is, suppose $f_1, f_2, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$ are all monotonically increasing or all monotone decreasing, with each f_i depending on disjoint subsets of $[n]$, $S_1, S_2, \ldots, S_k \subseteq [n]$. In that case, if $X_1, X_2, \ldots, X_n$ are NA, then the set of random variables $Y_1 = f_1(\vec{X}), Y_2 = f_2(\vec{X}), \ldots, Y_k = f_k(\vec{X})$ are NA.

Proof. We start by proving Property i. Denote by X_A, X_B and Y_A, Y_B some arbitrary partition of $\vec{X}$ and $\vec{Y}$. Let f_A and f_B be increasing functions. Then,

$$\mathbb{E}_X[f_A(X_A, Y_A) \mid Y_A = \vec{y_A}, Y_B = \vec{y_B}] = \mathbb{E}_X[f_A(X_A, Y_A) \mid Y_A = \vec{y_A}]$$
$$\mathbb{E}_X[f_B(X_B, Y_B) \mid Y_A = \vec{y_A}, Y_B = \vec{y_B}] = \mathbb{E}_X[f_B(X_B, Y_B) \mid Y_B = \vec{y_B}],$$

as $f_A(X_A, Y_A)$ and $g_B(X_B, Y_B)$ do not depend on Y_B and Y_A, respectively. Denote the conditional expectations

$$h_A(Y_A) = \mathbb{E}_X[f_A(X_A, Y_A) \mid Y_A]$$
$$h_B(Y_B) = \mathbb{E}_X[f_B(X_B, Y_B) \mid Y_B].$$

Note that h_A and h_B are increasing functions in their arguments. Therefore, by NA of $\vec{X}$ and $\vec{Y}$, as $\vec{X}$ and $\vec{Y}$ are independent, we have

$$\mathbb{E}[f(X_A, Y_A) \cdot g(X_B, Y_B)] = \mathbb{E}[\mathbb{E}[f_A(X_A, Y_A) \cdot f_B(X_B, Y_B) \mid Y_A, Y_B]]$$

Now, for all values Y_A and Y_B can take, as $\vec{X}$ and $\vec{Y}$ are independent, the expectation of the products is no greater than the product of the expectations, by NA of $\vec{X}$. So, by total probability, the above is at most

$$\begin{aligned}
&\leq \mathbb{E}[\mathbb{E}[f_A(X_A, Y_A) \mid Y_A, Y_B] \cdot \mathbb{E}[f_B(X_B, Y_B) \mid Y_A, Y_B] \\
&= \mathbb{E}[\mathbb{E}[f_A(X_A, Y_A) \mid Y_A] \cdot \mathbb{E}[f_B(X_B, Y_B) \mid Y_B] \\
&= \mathbb{E}[h_A(Y_A) \cdot h_B(Y_B)] \\
&\leq \mathbb{E}[h_A(Y_A)] \cdot \mathbb{E}[h_B(Y_B)] \\
&= \mathbb{E}[f_A(X_A, Y_A)] \cdot \mathbb{E}[f_B(X_B, Y_B)].
\end{aligned}$$

where the last inequality follows from the independence of $\vec{X}$ and $\vec{Y}$ and the NA of $\vec{X}$ and $\vec{Y}$, respectively.

Property ii is nearly immediate. Consider $f_1, \ldots, f_k$ all monotone increasing/decreasing and g and h monotonically increasing functions determined by disjoint subsets of the Y_i. Then g and h applied to the Y_j are compositions of monotone increasing functions with monotone increasing/decreasing functions and are therefore increasing/decreasing, as the case may be, determined by disjoint subsets of the X_i. But then, as the X_i are NA, we have $\mathbb{E}[g(\vec{Y}) \cdot h(\vec{Y})] \leq \mathbb{E}[g(\vec{Y})] \cdot \mathbb{E}[h(\vec{Y})]$. □

Note: Joag-Dev and Proschan [8] show that property ii is unique to NA among a wide class of notions of negative correlation.

5 Applications

5.1 Random Sampling Without Replacement

Suppose we have N items to pick from and we sample $n \leq N$ items of these without replacement, $X_1, X_2, \ldots, X_n$. These X_i can be seen as the first n items of a permutation distribution, and are therefore NA.

5.2 Fermi-Dirac Occupancy Numbers

The Fermi-Dirac distribution is given by the following process. Bins contain at most one ball each, with each distribution of the m balls among the n bins equally likely. That is, if $B_1, B_2, \ldots, B_n$ are indicator variables for bin i having a ball in it, then for any values $m_1, m_2, \ldots$ satisfying $\sum_i m_i = m$ this distribution satisfies

$$\Pr[\bigwedge_{i=1}^{n} B_i = m_i] = \binom{n}{m}^{-1}.$$

Proof. Follows immediately from Permutation Distributions being NA, as Dermi-Firac distribution is precisely a Permutation Distribution on m ones and $n-m$ zeros. □

5.3 Balls and Bins, and Balls and Bins

Consider the standard balls and bins process. The process consists of m balls and n bins, with each ball b placed in bin i with probability $p_{b,i}$, independently of other ball placements. Let B_i be the number of balls placed in bin i. This is often called the *occupancy number* of bin i. These variables are clearly not independent, as in particular $\sum_i B_i = m$ always. They are, however, NA.

Theorem 10. *The occupancy numbers, $B_1, B_2, \ldots, B_n$, in a balls and bin process are NA.*

Proof. Let $X_{b,i}$ be an indicator random variable for ball b being placed in bin i. By the Zero-One Lemma the set of variables $\{X_{b,i} \mid i \in [n]\}$ is NA. By closure of NA sets under union of independent NA sets (Property i), as each ball is placed in dependently of all other balls, the set $\{X_{b,i} \mid b \in [m], i \in [n]\}$ is NA. Finally, by closure of NA under monotone increasing functions on disjoint subsets (Property ii), noting that $B_i = \sum_b X_{b,i}$, we find that the B_i are indeed NA. □

A consequence of the above, together with the fact that indicator r.v.s for a bin i being empty are monotone decreasing functions in $\vec{B}$ determined by disjoint subsets of indices, is the following.

Corollary 11. *Consider a balls and bins process with m balls and n bins. Then, if $p_{b,i} = \frac{1}{n}$ for all b, i, then w.h.p the number of empty bins, N satisfies $N = \mathbb{E}[N] \pm O(\sqrt{n \log n}) = n \cdot e^{-m/n} \pm O(\sqrt{n \log n})$. In addition, provided $n \cdot e^{-m/n} = \Omega(\log n)$, then w.h.p $N = \Theta(\mathbb{E}[N]) = \Theta(n \cdot e^{-m/n})$.*

These consequences are disturbingly simple to prove when contrasted with the work needed to prove these results without NA (see [9, 12]).

5.4 Sampling Graphs

In this section, we will consider a few properties of graphs and the results of randomly subsampling the edges/vertices of a graph to these properties.

For the first example, recall that the *chromatic number of a graph,* $\chi(G)$, is the minimum number of colors with which you can color the vertices of G such that no two adjacent vertices share the same color.

Lemma 12. *Let G be a graph and H be obtained from G by sampling every edge of G with probability $\frac{1}{2}$ independently of all other edges. Then, $\mathbb{E}[\chi(H)] \geq \sqrt{\chi(G)}$.*

We will use the following simple fact, implied by the cartesian product of two subgraphs' colorings.

Fact 13. *For any two graphs on the same vertex set H_1, H_2, it holds*

$$\chi(H_1 \cup H_2) \leq \chi(H_1) \cdot \chi(H_2) \tag{1}$$

Now, consider $H_1 = H$ (the sampled subgraph) and $H_2 = G \setminus H$ (the subgraph of non-sampled edges). Clearly H_1 and H_2 have the same distribution. As $H_1 \cup H_2 = G$ we might hope that applying the expectation operator to both sides of Equation 1 would give $\chi(G) \leq \mathbb{E}[\chi(H_1) \cdot \chi(H_2)] \stackrel{?}{=} \mathbb{E}[\chi(H_1)] \cdot \mathbb{E}[\chi(H_2)] = \mathbb{E}[\chi(H)]^2$, yielding our desired result. Unfortunately, the equality with a question mark over it is sketchy, at best, as $\chi(H_1)$ and $\chi(H_2)$ are clearly not independent. They are, however, negatively associated, which will give us the desired inequality.[4]

Lemma 14. *[[1]] The random variables $\chi(H)$ and $\chi(G \setminus H)$ are negatively associated.*

Proof. Let $H_1 = H$ and $H_2 = G \setminus H$. Consider the pairs of indicator random variables $X_{e,1}, X_{e,2}$ defined by $X_{e,i} = 1$ if $e \in H_i$ and 0 else. By the Zero-One Principle every such pair is NA. But, as every edge $e \in G$ is put in H_1 or in H_2 independently of all other edges, by Property i the set of variables $\{X_{e,i} \mid e \in E[G], i = 1, 2\}$ are NA. Finally, we note that $\chi(H_1)$ and $\chi(H_2)$ are increasing functions determined by $\{X_{e,1} \mid e \in E[G]\}$ and $\{X_{e,2} \mid e \in E[G]\}$ respectively. Thus, by Property ii, $\chi(H_1) = \chi(H)$ and $\chi(H_2) = \chi(G \setminus H)$ are NA. □

Proof of Lemma 12. As $H_1, H_2 \sim H$, we have

$$\chi(G) \leq \mathbb{E}[\chi(H_1) \cdot \chi(H_2)] \leq \mathbb{E}[\chi(H_1)] \cdot \mathbb{E}[\chi(H_2)] = \mathbb{E}[\chi(H)]^2$$

Where the second inequality follows by NA of $\chi(H_1)$ and $\chi(H_2)$. Consequently, $\sqrt{\chi(G)} \leq \mathbb{E}[\chi(H)]$. □

The following is a generalization of Lemma 14 to multiple sub-sampled sub-graphs and general monotone increasing graph functions.

Theorem 15. *[[1]] Let G be some graph, and $H_1, H_2, \ldots, H_k$ be k subgraphs of G, with each edge e of G randomly and independently placed in graph H_i with probability $p_{e,i}$, where $\sum_i p_{e,i} = 1$. Let $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$ be a monotone increasing function on graphs. Then, $f(H_1), f(H_2), \ldots, f(H_k)$ are NA.*

Proof. The proof is identical to the proof for the chromatic number in Lemma 14 (nothing magical about the number 2, the uniform distribution ($p_{e,i} = \frac{1}{k}$), or chromatic number in this context). □

This NA property for a sampled and non-sampled subgraph implies many natural consequences. Consider for example the following process: a graph's edges are randomly partitioned into blue and red edges, each edge e colored blue with probability p_e and red with probability $1 - p_e$. Let G_R and G_B be the red and blue subgraphs. Let p_R and p_B be the probability that G_R and G_B are connected, respectively. Let p_{RB} be the probability that both subgraphs are connected. Then, as we would expect, we have

$$p_{RB} \leq p_R \cdot p_B$$

[4]As it so happens, for this particular problem application of NA is overkill, as the AM-GM inequality applied to 1 yields $\sqrt{\chi(H_1 \cup H_2)} \leq \sqrt{\chi(H_1) \cdot \chi(H_2)} \leq \frac{\chi(H_1)+\chi(H_2)}{2}$, following which linearity of expectation, together with $\mathbb{E}[\chi(H_1)] = \mathbb{E}[\chi(H_2)]$ yield the desired result. This example is instructive, however, so we present it nonetheless.

6 Other Useful Properties of NA

Theorem 16 (Coupling with Independent Variables [13]). *Let $X_1, X_2, \ldots, X_n$ be NA r.v.s and $X_1^*, X_2^*, \ldots, X_n^*$ be independent r.v.s such that X_i and X_i^* have the same distribution for each $i = 1, 2 \ldots, n$. Then for all convex functions f,*

$$\mathbb{E}[f(\sum_{i=1}^{n} X_i)] \leq \mathbb{E}[f(\sum_{i=1}^{n} X_i^*)]$$

If f is further a non-decreasing function then

$$\mathbb{E}[f(\max_k \sum_{i=1}^{k} X_i)] \leq \mathbb{E}[f(\max_k \sum_{i=1}^{k} X_i^*)]$$

Corollary 17 (Variance is sub-additive for NA variables). *Let $X_1, X_2, \ldots, X_n$ be NA r.v.s. Then,*

$$Var(\sum_i X_i) \leq \sum_i Var(X_i)$$

Proof. Recall that $Var(\sum_i X_i) = \mathbb{E}[(\sum_i X_i)^2] - \mathbb{E}[\sum_i X_i]^2$. By Theorem 16, if $X_1^*, X_2^*, \ldots, X_n^*$ are independent r.v.s with each X_i^* and X_i having the same individual distribution, then $\mathbb{E}[(\sum_i X_i)^2] \leq \mathbb{E}[(\sum_i X_i^*)^2]$ (and clearly $\mathbb{E}[\sum_i X_i]^2 = \mathbb{E}[\sum_i X_i^*]^2$.) In particular, we have

$$Var(\sum_i X_i) \leq Var(\sum_i X_i^*) = \sum_i Var(X_i^*) = \sum_i Var(X_i)$$ □

The above corollary facilitates the use of the second moment method (Chebyshev's more famous inequality, namely $\Pr[|X - \mathbb{E}[X]| \geq k] \leq \frac{Var(X)}{k^2}$) for sums of NA variables.[5] More interestingly, the above theorem implies that Kolmogorov's Inequality, a strengthening of Chebyshev's Inequality which holds for independent random variables, also holds for NA variables. This inequality, stated below, is nice in that it allows us to bound the probability of a partial sum deviating from the mean (zero) by some λ for *any* k by the same bound we obtain using Chebyshev for the probability of the *last* partial sum deviating from the mean.

Corollary 18 (Kolmogorov's Inequality holds for NA variables [13, 11]). *Let $X_1, X_2, \ldots, X_n$ be NA r.v.s with expected value $\mathbb{E}[X_k] = 0$ and finite variance $Var(X_k) < \infty$ for all k. Then, if $S_k = \sum_{i=1}^{k} X_i$, we have*

$$\Pr[\max_{1 \leq k \leq n} S_k \geq \lambda] \leq \frac{1}{\lambda^2} \cdot \sum_{i=1}^{n} Var(X_i)$$

Proof. Kolmogorov's inequality states that the above bound holds if we replace the $X_1, X_2, \ldots, X_n$ by independent $X_1^*, X_2^*, \ldots, X_n^*$ with each X_i^* having the same individual distribution as X_i. That is, if we denote by $S_k^* = \sum_{i=1}^{k} X_i^*$ the partial sums of X_i^*'s, then

$$\Pr[\max_{1 \leq k \leq n} S_k^* \geq \lambda] \leq \frac{1}{\lambda} \cdot \sum_i Var(X_i^*)$$

The rhs of this inequality, is exactly $\sum_i Var(X_i)$, as X_i^* and X_i are identically distributed. The lhs, by Theorem 16, can be lower bounded by the equivalent term with $S_k = \sum_{i=1}^{k} X_i$ replacing S_k^*. □

[5]Though why you would do so, given that you could rely on Chernoff bounds and use all moments, is a mystery.

Uses for Optimization. Two ideas come to mind:

First, any one-dimensional random walk with NA steps of mean zero and bounded variance, is unlikely to deviate by more than $O(\sqrt{n})$ from zero. Note that Azuma's Inequality gives us concentration w.h.p in a range of size $\tilde{O}(\sqrt{n})$, but if we're worried about polylog terms, this is good enough to give us constant probability for staying within a similar-sized range, without "paying" for union bound.

The second use of Theorem 16 (and corollaries) is the following. Suppose we want to minimize a convex function of the form $g(x) = f(\sum_i x_i)$ (f convex) over some subdomain of $\mathbb{R}^n$. Suppose furthermore that we can output a NA sequence $X_1, X_2, \ldots, X_n$ which is a feasible solution (or perhaps only feasible with high probability, provable using Chernoff-Hoeffding bounds). Then the expected value of this solution $\vec{X}$ is upper bounded by the expected cost of the "same" solution with $\vec{X}$ replaced by independent variables $X_1^*, X_2^*, \ldots, X_n^*$ following the same individual distributions as $X_1, X_2, \ldots, X_n$, an expectation which is likely to be easier to compute, due to independence of these variables.

7 FKG Inequality

Finally, we discuss the FKG inequality, a powerful theorem which can be used to prove NA of distributions. To state it, we need the following definitions.

Definition 3 (Distributive Lattice). *We say a set and order* $(\Gamma, <)$ *constitute a* lattice *if they form a partially-ordered set in which each two elements* $x, y \in \Gamma$ *have a unique least common upper bound, denoted by* $x \vee y$, *and a unique greatest upper bound, denoted by* $x \wedge y$.

A lattice is called distributive *if for each* $x, y, z \in \Gamma$ *the functions* $\vee$ *and* $\wedge$ *satisfy the following:*

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z),$$

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z).$$

As an example, consider the distributive lattice $(\Gamma, <)$ with Γ a collection of subsets of some ground set S, with the partial order of set inclusion (i.e., $x < y$ corresponds to $x \subset y$). In that case, for $x, y \in \Gamma$, we have $x \wedge y = x \cup y$ and $x \vee y = x \cap y$, and the distributive laws are the standard distributive laws for set union and intersection. In fact, this example is a **complete characterization** (up to isomorphism) of all distributive lattices, so thinking of this kind of distributive lattice is without any major loss of generality. Nonetheless, we will consider yet another example, which will prove useful shortly.

As a second example, consider the set Γ of k-element ordered subsets of some ground set of n reals, where we denote by x_i the i-th largest element of x, and let $x \leq y$ iff $x_i \leq y_i$. In this case, $(x \vee y)_i = \max\{x_i, y_i\}$ and $(x \wedge y)_i = \min\{x_i, y_i\}$ for all $i \in [k]$. This is a distributive lattice due to the distributive properties of min and max.

Theorem 19 (FKG Inequality [7]). *Let* $(\Gamma, <)$ *be a finite distributive lattice, and let* $\mu : \Gamma \to \mathbb{R}^+$ *be a log super-modular function on* $(\Gamma, <)$*; i.e.,*

$$\mu(x \wedge y) \cdot \mu(x \vee y) \geq \mu(x) \cdot \mu(y).$$

Then, for any two monotonically increasing functions f *and* g *on* $(\Gamma, <)$*, (that is,* $x \leq y$ *implies* $f(x) \leq f(y)$*) the following holds:*

$$\left(\sum_{x\in\Gamma} f(x)g(x)\mu(x)\right) \cdot \left(\sum_{x\in\Gamma} \mu(x)\right) \geq \left(\sum_{x\in\Gamma} f(x)\mu(x)\right) \cdot \left(\sum_{x\in\Gamma} g(x)\mu(x)\right).$$

Reading the above with μ as a probability function over sets, the above gives us a condition under which monotone increasing functions f and g on a randomly sampled set X from some distribution satisfy

$$\mathbb{E}[f(X) \cdot g(X)] \geq \mathbb{E}[f(X)] \cdot \mathbb{E}[g(X)].$$

7.1 Application: Alternative Proof of NA for Permutation Distributions

This example is taken from [5]. Let $X_1, \ldots, X_n$ have the permutation distribution over $[n]$ with each permutation appearing with probability $\frac{1}{n!}$. We re-prove this distribution is NA using the FKG inequality.

Lemma 20 (Permutation Distributions are NA ([5]))**.** *Let $X_1, X_2, \ldots, X_n$ be random variables such that $\{X_1, X_2, \ldots, X_n\} = [n]$ always, with all $n!$ possible assignments equally likely. Then $X_1, X_2, \ldots, X_n$ are NA.*

Proof. Consider the finite distributive lattice $(\Gamma, <)$ of k-element ordered sets of $[n]$ with the component-wise order, discussed above. In order to apply the FKG inequality, we need to define some μ. Take $\mu(S) = \binom{n}{k}^{-1}$ for all $S \subseteq [n]$. This function is trivially log super-modular, satisfying the log super-modular inequality with equality. Now, in order to leverage FKG to prove NA of $X_1, \ldots, X_n$, let f and g be two monotone increasing functions over $\vec{X}$ which depend on disjoint subsets of the indices. Fix some $k \in [n]$. For any k-element ordered subset $S \subseteq [n]$, $S = (S_1, S_2, \ldots, S_k)$ and permutation on k elements, σ, we let $\sigma(S) = (\sigma(S_1), \ldots, \sigma(S_k))$. Using this notation, define the functions $f'(S) = \frac{1}{k!} \sum_\tau f(\tau(S))$ and $g'(S) = \frac{1}{(n-k)!} \sum_\sigma f(\sigma([n] \setminus S))$. In that case, f' and g' are easily verified to be monotone increasing and decreasing respectively (i.e., $S \leq T$ satisfies $f'(S) \leq f'(T)$ and satisfies $g'(S) \geq g'(T)$), by considering the sums in these functions' definition term-wise. Consequently, by the FKG inequality, we have that

$$\sum_{S \in \Gamma} f'(S) g'(S) \mu(S) \cdot \sum_{S \in \Gamma} \mu(S) \leq \sum_{S \in \Gamma} f'(S) \mu(S) \cdot \sum_{S \in \Gamma} g'(S) \mu(S)$$

Now, by our choice of $\mu = \binom{n}{k}^{-1}$, the left hand side of the above is precisely $\mathbb{E}[f(S)g(S)] = \sum_{S \in \Gamma} f(S)g(S) \cdot \frac{1}{n!}$, whereas the right hand side is precisely equal $\mathbb{E}[f(S)] \cdot \mathbb{E}[g([n] \setminus S)] = \sum_{S \in \Gamma} f(S) \cdot \frac{(n-k)!}{n!} \cdot \sum_{S \in \Gamma} g([n] \setminus S) \cdot \frac{k!}{n!}$. We conclude that $X_1, \ldots, X_n$ are NA (as g need not depend on all its $n-k$ inputs.) ☐

Proving that for any n (not necessarily distinct) numbers $x_1 \leq x_2 \leq \cdots \leq x_n$, a permutation distribution over the x_i is also NA follows easily from the above and closure properties of NA.

8 Summary

We discussed a strong notion of negative dependence, called negative association. This notion has several appealing properties. The first and foremost one is that it allows us to apply several strong theorems which are normally reserved for independent random variables to variables which are clearly not independent. These include Chernoff-Hoeffding bounds, NOD, Kolmogorov's Inequality, and many more we didn't cover (see, e.g. [14] for an extension of the law of the iterated logarithm to NA variables[6]) Another appealing property of negative association is its closure properties, allowing for "calculation-free" proofs that distributions are NA, by plugging in known NA distributions and applying closure properties to obtain the relevant distribution. We have discussed a few simple NA distributions which can be used to build more elaborate NA distributions by means of these closure properties.

We note that we have omitted some more elaborate and general NA distributions, such as regular matroids and strong Ralyeigh distributions (see Feder and Mihail [6] and Borcea et al. [2]), as well as some interesting distributions for algorithmic applications, for example the output of Srinivissan's dependent rounding procedure for level sets (see [15, 3]). Generally, the Feder-Mihail Theorem (see [6]) gives conditions by which a distribution is NA, and in fact is NA even conditioned on setting any subsets of the variables to fixed values (see [3], pg. 5 for a simple exposition of this theorem.)

[6]The law of the iterated logarithm states that for any sequence $\{X_i\}$ of i.i.d r.v.s with mean zero and unit variance, if $S_n = \sum_{i=1}^n X_i$, then $\limsup_{n \to \infty} \frac{S_n}{\sqrt{n \ln \ln n}} = \sqrt{2}$ almost surely.

References

[1] Vijay Bhattiprolu and David Wajc, unpublished, 2015.

[2] Julius Borcea, Petter Brändén, and Thomas Liggett, *Negative dependence and the geometry of polynomials*, Journal of the American Mathematical Society **22** (2009), no. 2, 521–567.

[3] Devdatt Dubhashi, Johan Jonasson, and Desh Ranjan, *Positive influence and negative dependence*, Combinatorics, Probability and Computing **16** (2007), no. 01, 29–41.

[4] Devdatt Dubhashi and Desh Ranjan, *Balls and bins: A study in negative dependence*, BRICS Report Series **3** (1996), no. 25.

[5] Devdatt P Dubhashi, Volker Priebe, and Desh Ranjan, *Negative dependence through the fkg inequality*, BRICS Report Series **3** (1996), no. 27.

[6] Tomás Feder and Milena Mihail, *Balanced matroids*, Proceedings of the twenty-fourth annual ACM symposium on Theory of computing, ACM, 1992, pp. 26–38.

[7] Cees M Fortuin, Pieter W Kasteleyn, and Jean Ginibre, *Correlation inequalities on some partially ordered sets*, Communications in Mathematical Physics **22** (1971), no. 2, 89–103.

[8] Kumar Joag-Dev and Frank Proschan, *Negative association of random variables with applications*, The Annals of Statistics (1983), 286–295.

[9] Anil Kamath, Rajeev Motwani, Krishna Palem, and Paul Spirakis, *Tail bounds for occupancy and the satisfiability threshold conjecture*, Random Structures & Algorithms **7** (1995), no. 1, 59–80.

[10] Alam Khursheed and KM Lai Saxena, *Positive dependence in multivariate distributions*, Communications in Statistics-Theory and Methods **10** (1981), no. 12, 1183–1196.

[11] Przemysław Matuła, *A note on the almost sure convergence of sums of negatively dependent random variables*, Statistics & Probability Letters **15** (1992), no. 3, 209–213.

[12] Alessandro Panconesi and Aravind Srinivasan, *Fast randomized algorithms for distributed edge coloring*, Proceedings of the Eleventh Annual ACM Symposium on Principles of Distributed Computing (New York, NY, USA), PODC '92, ACM, 1992, pp. 251–262.

[13] Qi-Man Shao, *A comparison theorem on moment inequalities between negatively associated and independent random variables*, Journal of Theoretical Probability **13** (2000), no. 2, 343–356.

[14] Qi-Man Shao and Chun Su, *The law of the iterated logarithm for negatively associated random variables*, Stochastic Processes and Their Applications **83** (1999), no. 1, 139–148.

[15] Aravind Srinivasan, *Distributions on level-sets with applications to approximation algorithms*, Proceedings. of the 42nd IEEE Symposium on Foundations of Computer Science, 2001., IEEE, 2001, pp. 588–597.