MAX-PLANCK-INSTITUT FÜR INFORMATIK

Some Correlation Inequalities for Probabilistic Analysis of Algorithms

Devdatt Dubhashi, Desh Ranjan

MPI–I–94–143 August 1994

INFORMATIK

Im Stadtwald
66123 Saarbrücken
Germany

Some Correlation Inequalities for Probabilistic Analysis of Algorithms

Devdatt Dubhashi, Desh Ranjan

MPI–I–94–143 August 1994

Some Correlation Inequalities for Probabilistic Analysis of Algorithms

Devdatt Dubhashi *
Max–Planck–Institute für Informatik
Im Stadtwald, 66123 Saarbrücken, Germany
dubhashi@mpi-sb.mpg.de

Desh Ranjan†
Department of Computer Science
New Mexico State University, Las Cruces
New Mexico 88003, USA
dranjan@cs.nmsu.edu

August 1, 1994

Abstract

The analysis of many randomized algorithms, for example in dynamic load balancing, probabilistic divide-and-conquer paradigm and distributed edge-coloring, requires ascertaining the precise nature of the correlation between the random variables arising in the following prototypical "balls-and-bins" experiment. Suppose a certain number of balls are thrown uniformly and independently at random into n bins. Let X_i be the random variable denoting the number of balls in the ith bin, $i \in [n]$. These variables are clearly not independent and are intuitively negatively related. We make this mathematically precise by proving the following type of correlation inequalities:

- For index sets $I, J \subseteq [n]$ such that $I \cap J = \emptyset$ or $I \cup J = [n]$, and any non-negative integers t_I, t_J,

$$\Pr[\sum_{i \in I} X_i \geq t_I \mid \sum_{j \in J} X_j \geq t_J] \leq \Pr[\sum_{i \in I} X_i \geq t_I].$$

- For any disjoint index sets $I, J \subseteq [n]$, any $I' \subseteq I, J' \subseteq J$ and any non-negative integers $t_i, i \in I$ and $t_j, j \in J$,

$$\Pr[\bigwedge_{i \in I} X_i \geq t_i \mid \bigwedge_{j \in J} X_j \geq t_j] \leq \Pr[\bigwedge_{i \in I'} X_i \geq t_i \mid \bigwedge_{j \in J'} X_j \geq t_j].$$

Although these inequalities are intuitively appealing, establishing them is non-trivial; in particular, direct counting arguments become intractable very fast. We prove the inequalities of the first type by an application of the celebrated FKG Correlation Inequality. The proof for the second uses only elementary methods and hinges on some *monotonicity* properties.

More importantly, we then introduce a general methodology that may be applicable whenever the random variables involved are negatively related. Precisely, we invoke a general notion of *negative assocation* of random variables and show that:

- The variables X_i are negatively associated. This yields most of the previous results in a uniform way.
- For a set of negatively associated variables, one can apply the Chernoff-Hoeffding bounds to the sum of these variables. This provides a tool that facilitates analysis of many randomized algorithms, for example, the ones mentioned above.

*Supported by the ESPRIT Basic Research Actions Program of the EC under contract No. 7141 (project ALCOM II).
†Work done while the author was visiting the Max-Planck-Institut für Informatik

1 Introduction

The analysis of many randomized algorithms involves random variables that are not independent. This renders many of the standard tools of probability theory that are valid for independent variables, such as the Chernoff bounds, inapplicable. In this case it is desirable to ascertain the precise nature of the correlation between the random variables involved and determine sufficient conditions in which these probabilistic tools can be reinstated. Some previous work of this type can be found in [11, 12, 16, 17].

In many cases, these variables are not independent, but related, intuitively, in the following negative way: for any two of these variables, given that one of them is "large" one would expect the other to be "small". Typical examples of this kind arise when analysing algorithms for dynamic load balancing, in the probabilistic divide-and-conquer paradigm and distributed graph algorithms[8, 7, 11, 12, 16]. The essence of the correlation in these examples is captured by the following "balls and bins" experiment, which is a classical paradigm example.

Suppose we throw a certain number of balls into a certain number, n of bins uniformly and independently at random. For $i \in [n]$, let X_i be the random variable denoting the number of balls in the ith bin. The question is: how are the X_i related? Although the balls are thrown uniformly and independently at random, these variables are *not* independent; in particular, their sum is fixed, being equal to the number of balls thrown. Intuitively, the variables are ***negatively correlated*** in the manner indicated by the following innocous looking statements:

1. $\Pr[X_1 + X_2 \geq 5 \mid X_3 + X_6 + X_{17} \geq 6] \leq \Pr[X_1 + X_2 \geq 5]$.
2. $\Pr[X_1 \geq 3, X_2 \geq 4 \mid X_3 \geq 5, X_4 \geq 6] \leq \Pr[X_1 \geq 3, X_2 \geq 4 \mid X_3 \geq 5] \leq \Pr[X_1 \geq 3, X_2 \geq 4]$.

Although these statements appear almost self-evident, they seem to be surprisingly hard to prove. In particular, a direct counting argument with binomial coefficients seems to lead nowhere.

1.1 Statement of Results

In this paper, we first prove the following types of correlation inequalities:

- For index sets $I, J \subseteq [n]$ such that $I \cap J = \emptyset$ or $I \cup J = [n]$, and any non-negative integers t_I, t_J,

$$\Pr[\sum_{i \in I} X_i \geq t_I \mid \sum_{j \in J} X_j \geq t_J] \leq \Pr[\sum_{i \in I} X_i \geq t_I].$$

- For any disjoint index sets $I, J \subseteq [n]$, any $I' \subseteq I, J' \subseteq J$ and any non-negative integers $t_i, i \in I$ and $t_j, j \in J$,

$$\Pr[\bigwedge_{i \in I} X_i \geq t_i \mid \bigwedge_{j \in J} X_j \geq t_j] \leq \Pr[\bigwedge_{i \in I'} X_i \geq t_i \mid \bigwedge_{j \in J'} X_j \geq t_j].$$

We prove the inequalities of the first type by an application of the celebrated FKG Correlation Inequality. Good accounts of the FKG inequality and its previous applications can be found in [1, 6]. The proof for the second uses only elementary methods and hinges on some *monotonicity* properties.

More importantly, we then introduce a general methodology that may be applicable whenever the random variables involved are negatively related. Precisely, we invoke a general notion of ***negative assocation*** of random variables[3]. Negative association provides a unifying framework to make precise a notion of strong negative dependence between variablese. We show that:

- The variables X_i are negatively associated. This yields most of the previous results in a uniform way.
- For a set of negatively associated variables, one can apply the Chernoff-Hoeffding bounds to the sum of these variables. This provides a tool that facilitates analysis of many randomized algorithms, for example, the ones mentioned above.

1.2 Applications

1.2.1 Dynamic Load Balancing

Consider a scenario in which one has to allocate various jobs to available servers, for eaxmple, programs requesting data from disc drives, or user queries to a database system. It is desirable to perform the allocation dynamically in such a way that the load is relatively balanced across the servers. Dynamic load balancing is a well-studied problem and several strategies for load balancing have been proposed and analysed. In a recent work, Lauer describes a new dynamic load balancing strategy [8]. The analysis of this algorithm requires establishing correlation inequalities such as those described in the introduction. Our solution enables this analysis to be brought to completion.

1.2.2 Occupancy Problems in Statistical Physics

In Statistical Physics, one has an *ensemble* of m particles, distributed in a *phase space* which is divided into n regions or *cells*, in such a way that "all configurations are equally likely". In order to calculate various random quantities of interest, it is necessary to carefully specify in what sense one intends this last qualification. There are two key dichotomies: whether the particles are regarded as indistinguishable and whether multiple occupancy of a cell is permitted. There are three well known models in use in Statistical Physics:

1. **[Maxwell–Boltzmann Model]** The particles are distinguishable and multiple occupancy is allowed.
2. **[Fermi–Dirac Model]** The particles are indistinguishable and multiple occupancy is forbidden (owing to the so called *exclusion principle*).
3. **[Bose–Einstein Model]** The particles are indistinguishable but multiple occupancy is allowed.

Although the Maxwell–Boltzmann model appears at first to be the most natural one, empirical and theoretical studies have showed that various classes of elementary particles actually obey one of the other two distributions.

Let X_i denote the *occupancy number* i.e. the number of balls in the ith cell, $i \in [n]$. Note that the statistics in the balls and bins example from the introduction is of the Maxwell–Boltzmann type. The joint distribution of $X_1, \cdots, X_n$ is well known under all three distributions:

Proposition 1 *For any non-negative integers $m_1, \cdots, m_n$ such that $m_1 + \cdots + m_n = m$, we have,*

1. *For the Maxwell-Boltzmann statistics,*

$$\Pr[X_1 = m_1, \cdots, X_n = m_n] = \frac{m!}{m_1! \cdots m_n!} n^{-m}.$$

2. *For the Fermi-Dirac statistics,*

$$\Pr[X_1 = m_1, \cdots, X_n = m_n] = \binom{n}{m}^{-1}.$$

3. *For the Bose-Einstein statistics,*

$$\Pr[X_1 = m_1, \cdots, X_n = m_n] = \binom{n+m-1}{m}^{-1}.$$

In principle, one can deduce from this joint distribution, all other quantities and relationships of interest. For the Fermi–Dirac statistics, it is easy to show that one has

$$\Pr[\bigwedge_{i \in I} X_i \geq t_i \mid \bigwedge_{j \in J} X_j \geq t_j] \leq \Pr[\bigwedge_{i \in I} X_i \geq t_i]$$

for disjoint index sets I, J and any reals $t_i, i \in I$ and $t_j, j \in J$. However, it does not seems easy to draw similar conclusions for the other models. Below we resolve questions of this type in the other two models as well.

1.2.3 Probabilistic Recurrences

A *probabilistic recurrence relation* is a recurrence of the form:

$$T(x) = a(x) + \sum_{1 \leq i \leq k} T(H_i(x)), \quad (1)$$

where $a(x)$ is a fixed function, $k \geq 1$ is an integer and each $H_i(x)$ for $i \in [k]$ is a random variable in the range $[0, x]$. Such a recurrence describes succinctly the performance of certain divide–and–conquer randomised algorithms, or certain recursively produced random structures [7].

Under this circumstance, T is itself a random variable whose distribution is determined by that of the random variables H_i. A weak but useful set of assumptions to make on the distribution of the H_i's is that for each $i \in [k]$, there exists a fixed function $m_i(x)$ with $0 \leq m_i(x) \leq x$ such that $E[H_i(x)] \leq m_i(x)$. Under this set of assumptions, we would like to give tail probability bounds on the random variable $T(x)$.

The probabilistic recurrence (1) determines, in a natural way, a labelling of the infinite k–ary tree, T_k, by a random vector $\mathbf{X}$ determined or described by the recurrence. Namely, label the root with x, and having labelled a vertex z, label its children by the values $H_i(z)$ determined by the joint distribution of the H_i vaiables. The components of this random vector might not be independent, causing complications in the analysis. However, in many situations, the random vector $\mathbf{X}$ regarded as a labelling of the tree T_k has the following property:

For each node $v \in T_k$ with children $w_1, \cdots, w_k$, and for each $t, t_1 \in \mathsf{R}$, the joint distribution of the variables $X(w_1) \cdots X(w_k)$ is such that

$$\Pr[X(w_1) \geq t_1 \mid X(v) = t, X(w_2) = t_2, \cdots, X(w_k) = t_k]$$

is non–increasing in $t_2, \ldots, t_k$.

In this situation, we would like to claim that the variables corresponding to different paths in the tree are negatively related in some sense. Hence that we can upper bound stochastic properties of the tree by treating the variables corresponding to different paths as truly independent. The concepts and results of § 5 provide tools to enable us to accomplish this task. The analysis of the recurrence is thereby considerably simplified, as the analysis of a single path in the tree can be carried out as in [7, 4].

1.2.4 Edge Colouring of Graphs

Panconesi and Srinivasan [11, 12, 16] describe edge colouring algorithms for graphs in the distributed model of computation (these can also be implemented directly in the PRAM model). The analysis of the algorithm requires stochastic bounds on the sum of random variables that are not independent, but negatively related. Our results give an efficient proof of the required bounds, see § 5.3 below.

2 Preliminaries: The FKG Inequality

We shall use the following elementary equalities involving conditional probabilities quite extensively:

1. For any events A, B, $\Pr[A \bigwedge B] = \Pr[A|B]\Pr[B]$.

2. Let A and B be arbitray events, and let $C_i, i \in I$ be a partition of the universe of events. Then

$$\Pr[A \mid B] = \sum_{i \in I} \Pr[A \mid B, C_i]\Pr[C_i \mid B].$$

We start by recalling some concepts from the theory of partial orders. A *lattice* is a partially ordered set in which every two elements x, y have a unique minimal upper bound, denoted $x \vee y$ and called the *join* of x and y, and a unique maximal lower bound denoted $x \wedge y$ and called the *meet* of x and y. A lattice L, is *distributive* if for all $x, y, z \in L$, we have the following two equivalent (dual) *distributive laws*:

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

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

REMARK: The standard example of a distributive lattice is the lattice of all subsets of a given set ordered by set inclusion. The join and meet in this lattice are given by set union and set intersection respectively. A somewhat striking but easy result is that *any* finite distributive lattice is isomorphic to a sublattice of the lattice of all subsets of some finite set. Given a distributive lattice L, a function $f : L \rightarrow \mathsf{R}$ is said to be non-decreasing (non-increasing) if $x \leq_L y$ implies $f(x) \leq f(y)$ (respectively, $x \leq_L y$ implies $f(x) \geq f(y)$). A function $\mu : L \rightarrow \mathsf{R}^+$ on a distributive lattice L, is called *log-supermodular* if

$$\mu(x)\mu(y) \leq \mu(x \vee y)\mu(x \wedge y)$$

for all $x, y \in L$.

Motivated by a problem in statistical mechanics, Fortuin, Kasteleyn and Ginibre, and independently Sarkar, proved the following celebrated correlation inequality [5, 13, 6, 1]:

Theorem 2 (FKG Inequality) *Let L be a finite distributive lattice and let $\mu : L \rightarrow \mathsf{R}^+$ be a log-supermodular function. Then if $f, g : L \rightarrow \mathsf{R}^+$ are both non-decreasing or both non-increasing, we have*

$$(\sum_{x \in L} \mu(x)f(x)) \cdot (\sum_{x \in L} \mu(x)g(x)) \leq (\sum_{x \in L} \mu(x)f(x)g(x)) \cdot (\sum_{x \in L} \mu(x)).$$

If one of the functions is non-decreasing and the other is non-increasing then the reverse inequality holds.

REMARK: It is helpful to view μ as a measure on L. Assuming μ is not identically zero, we can define, for any $f : L \rightarrow \mathsf{R}^+$, its expectation,

$$\langle f \rangle := \frac{\sum_{x \in L} f(x)\mu(x)}{\sum_{x \in L} \mu(x)}.$$

In this notation, the FKG Inequality asserts that for any log-supermodular μ, and functions $f, g : L \rightarrow \mathsf{R}^+$,

$$\langle f \rangle \cdot \langle g \rangle \leq \langle fg \rangle$$

if f, g are both non-decreasing or both non-increasing, and

$$\langle f \rangle \cdot \langle g \rangle \geq \langle fg \rangle$$

if one of the functions is non-decreasing and the other is non-increasing. This formulation clearly brings out the probabilistic nature of the FKG Inequality.

3 The Proof via FKG

We apply the FKG Inequality to give a quick proof of certain correlation statements about the random variables in the balls and bins example. Suppose we throw m balls into n bins uniformly and independently at random, for positive integers m, n. As before X_i is the random variable denoting the number of balls in the ith bin.

A possible *configuration* of the experiment can be represented by a vector $\mathbf{a} := (a_1, \cdots, a_m)$, with $a_i \in [n]$ for each $i \in [m]$. This is the configuration where ball i goes into bin a_i for each $i \in [m]$. Define the lattice L to be the set of all such configurations ordered component-wise:

$$\mathbf{a} \leq_L \mathbf{b} \quad \leftrightarrow \quad a_i \leq b_i, \text{ for each } i \in [m].$$

It turns out that this in fact defines a distributive lattice, with join and meet given by the following equation on the components:

$$(a \vee b)_i := \max(a_i, b_i) \quad \text{and} \quad (a \wedge b)_i := \min(a_i, b_i).$$

Distributivity follows because of the following property of the integers:

$$\min(a, \max(b, c)) = \max(\min(a, b), \min(a, c)),$$

$$\max(a, \min(b, c)) = \min(\max(a, b), \max(a, c)).$$

Now define $\mu : L \rightarrow \mathsf{R}^+$ by $\mu(\mathbf{a}) := 1/n^m$ for each $\mathbf{a} \in L$. So defined, μ is trivially log-supermodular, and crucially, it makes each configuration equally likely, representing the fact that the balls are thrown uniformly and independently at random into the bins.

For a configuration $\mathbf{a}$, we introduce naturally, for $i \in [n]$,

$$X_i(\mathbf{a}) := |\{j \mid a_j = i\}|.$$

This gives the number of balls in the ith bin in configuration $\mathbf{a}$.

Let $I, J \subseteq [n]$ be two index sets such that either $I \cap J = \emptyset$ or $I \cup J = [n]$; with no loss of generality, we can arrange it by renumbering, so that $J := \{1, \cdots, |J|\}$ and $I := \{n - |I| + 1, \cdots, n\}$. Let t_I, t_J be arbitrary non-negative integers. Define $f, g : L \rightarrow \mathsf{R}^+$ as indicator functions by

$$f(\mathbf{a}) := \begin{cases} 1, & \text{if } \sum_{i \in I} X_i(\mathbf{a}) \geq t_I; \\ 0, & \text{otherwise.} \end{cases}$$

$$g(\mathbf{a}) := \begin{cases} 1, & \text{if } \sum_{j \in J} X_j(\mathbf{a}) \geq t_J; \\ 0, & \text{otherwise.} \end{cases}$$

The definition of the lattice order ensures that f is non-decreasing while g is non-increasing. Also, note that

$$\begin{aligned} \sum_{\mathbf{a} \in L} \mu(\mathbf{a}) f(\mathbf{a}) &= \frac{1}{n^m} \cdot |\{\mathbf{a} \in L : \sum_{i \in I} X_i(\mathbf{a}) \geq t_I\}| \\ &= \Pr[\sum_{i \in I} X_i \geq t_I]. \end{aligned}$$

Similarly

$$\sum_{\mathbf{a} \in L} \mu(\mathbf{a}) g(\mathbf{a}) = \Pr[\sum_{j \in J} X_j \geq t_J]$$

$$\sum_{\mathbf{a} \in L} \mu(\mathbf{a}) f(\mathbf{a}) g(\mathbf{a}) = \Pr[\sum_{i \in I} X_i \geq t_I, \sum_{j \in J} X_j \geq t_J].$$

Applying the FKG Inequality, we get the following correlation inequality on the random variables $X_i, i \in [n]$:

Theorem 3 *Let $I, J \subseteq [n]$ be index sets such that $I \cap J = \emptyset$ or $I \cup J = [n]$, and let t_I, t_J be arbitrary non-negative integers. Then*

$$\Pr[\sum_{i \in I} X_i \geq t_I \mid \sum_{j \in J} X_j \geq t_J] \leq \Pr[\sum_{i \in I} X_i \geq t_I].$$

REMARK 1: By taking I, J to be singletons, this also implies that $\Pr[X_i \geq t_i \mid X_j \geq t_j] \leq \Pr[X_i \geq t_i]$ for any distinct $i, j \in [n]$ and any non-negative integers t_i, t_j.
REMARK 2: In fact a closer examination of the proof shows that it yields somewhat more. We invoke the notion of *majorisation*. Given two vectors $\mathbf{x}, \mathbf{y} \in \mathsf{R}^n$, $\mathbf{x}$ is said to be majorised by $\mathbf{y}$, denoted $\mathbf{x} \preceq \mathbf{y}$ if $\sum_{i \in [n]} x_i = \sum_{i \in [n]} y_i$ and for each $k \in [n]$, $\sum_{k \leq i \leq n} x_i \leq \sum_{k \leq i \leq n} y_i$. (This is a slightly modified version of the usual definition, [9].) A function $f : \mathsf{R}^n \rightarrow \mathsf{R}$ is called *Schur non-decreasing* if it respects this ordering, that is $\mathbf{x} \preceq \mathbf{y}$ implies $f(\mathbf{x}) \leq f(\mathbf{y})$, and similarly one defines a Schur non-increasing function. The proof above shows directly that for any Schur non-decreasing function ψ and any Schur non-increasing function ϕ, and any reals t, t',

$$\Pr[\psi(X_1, \cdots, X_n) \geq t \mid \phi(X_1, \cdots, X_n) \geq t'] \leq \Pr[\psi(X_1, \cdots, X_n) \geq t].$$

4 An Elementary Proof

In this section, we give an elementary proof of the following type of correlation inequalites:

$$\Pr[X_1 \geq t_1, X_2 \geq t_2 \mid X_3 \geq t_3, X_4 \geq t_4] \leq \Pr[X_1 \geq t_1, X_2 \geq \mid X_3 \geq t_3]$$

for any non–negative integers $t_1, \cdots, t_4$. More precisely we establish the following theorem.

Theorem 4 *Let $I, J \subseteq [n]$ be disjoint index sets, $t_i, i \in I$ and $t_j, j \in J$ any non-negative integers and $I' \subseteq I, J' \subseteq J$. Then*

$$\Pr[\bigwedge_{i \in I} X_i \geq t_i \mid \bigwedge_{j \in J} X_j \geq t_j] \leq \Pr[\bigwedge_{i \in I'} X_i \geq t_i \mid \bigwedge_{j \in J'} X_j \geq t_j].$$

We begin by making some observations and proving some *monotonicity* lemmas.

Notation 5 *$\Pr^n_m[E]$ denotes the probability of the event E when m indistinguishable balls are thrown into n distinguished bins.*

Observation 6 *Let $I, J \subseteq [n]$, $I \cap J = \emptyset$. Let $a_i, i \in I$ and $a_j, j \in J$ be arbitrary non-negative integers. Let E_1 be the event $\bigwedge_{i \in I} X_i = a_i$ and E_2 be the event $\bigwedge_{j \in J} X_j = a_j$. Then,*

$$\Pr^n_m[E_1 | E_2] = \Pr^{n-N}_{m-M}[E_1]$$

where $M = \sum_{j \in J} a_j$ and $N = |J|$.

In other words, if the number of balls in certain bins is fixed, then the remaining balls get distributed in the remaining bins independently and uniformly at random. Using this lemma we can establish the following:

Lemma 7 *Let $I, J \subseteq [n]$, $I \cap J = \emptyset$. Let $a_i, i \in I$ and $a_j, j \in J$ be arbitrary non-negative integers. Let E_1 be the event $\bigwedge_{i \in I} X_i \geq a_i$ and E_2 be the event $\bigwedge_{j \in J} X_j = a_j$. Then,*

$$\Pr^n_m[E_1 | E_2] = \Pr^{n-N}_{m-M}[E_1]$$

where $M = \sum_{j \in J} a_j$ and $N = |J|$

To maintain continuity of the exposition, We defer the inductive proof of the following crucial monotonicity lemma to the appendix.

Lemma 8 *Let $I, J \subseteq [n]$, $I \cap J = \emptyset$. Let $l \in [n] \setminus I \cup J$. Let $a_i, i \in I$ and $a_j, j \in J$ and be arbitrary non-negative integers. Let E_1 be the event $\bigwedge_{i \in I} X_i \geq a_i$ and E_2 be the event $\bigwedge_{j \in J} X_j \geq a_j$. Let*

$$f(a) := \Pr^n_m[E_1 | E_2, X_l = a].$$

Then f is non–increasing in a.

An immediate application of this lemma gives the following.

Lemma 9 *Let $I, J \subseteq [n]$, $I \cap J = \emptyset$.Let $l \in [n] \setminus I \cup J$. Let $a_i, i \in I, a_j, j \in J$ and be arbitrary non-negative integers. Let E_1 be the event $\bigwedge_{i \in I} X_i \geq a_i$ and E_2 be the event $\bigwedge_{j \in J} X_j \geq a_j$. Let*

$$g(a) := \Pr^n_m[E_1 | E_2, X_l \geq a].$$

Then g is non–increasing in a.

Proof. We have,

$$\Pr^n_m[E_1|E_2, X_l \geq a] = \frac{\Pr^n_m[E_1, E_2, X_l \geq a+1] + \Pr^n_m[E_1, E_2, X_l = a]}{\Pr^n_m[E_2, X_l \geq a+1] + \Pr^n_m[E_2, X_l = a]}$$

Let

$$\begin{aligned} A &= \Pr^n_m[E_1, E_2, X_l \geq a+1] \\ B &= \Pr^n_m[E_2, X_l \geq a+1] \\ C &= \Pr^n_m[E_1, E_2, X_l = a] \\ D &= \Pr^n_m[E_2, X_l = a] \end{aligned}$$

Then

$$\begin{aligned} \frac{A}{B} &= \Pr^n_m[E_1 \mid E_2, X_l \geq a+1] \\ &= \sum_{t \geq a+1} \Pr^n_m[E_1 \mid E_2, X_l = t] \cdot \Pr^n_m[X_l = t \mid E_2, X_l \geq a+1] \end{aligned}$$

which, using the monotonicity lemma gives

$$\begin{aligned} \frac{A}{B} &\leq \Pr^n_m[E_1 \mid E_2, X_l = a] \cdot \sum_{t \geq a+1} \Pr^n_m[X_l = t \mid E_2, X_l \geq a+1] \\ &= \frac{C}{D}. \end{aligned}$$

So, we get that, $\frac{A+C}{B+D} \geq \frac{A}{B}$ or that

$$\Pr^n_m[E_1|E_2, X_l \geq a] \geq \Pr^n_m[E_1 \mid E_2, X_l \geq a+1].$$

establishing the lemma. ∎

The proof of the theorem now follows easily by noting that,

$$\Pr[\bigwedge_{i \in I} X_i \geq t_i \mid \bigwedge_{j \in J} X_j \geq t_j] \leq \Pr[\bigwedge_{i \in I'} X_i \geq t_i \mid \bigwedge_{j \in J} X_j \geq t_j].$$

and that,

$$\begin{aligned} &\Pr\left[\bigwedge_{i \in I'} X_i \geq t_i \mid \bigwedge_{j \in J'} X_j \geq t_j\right] \\ &= \Pr[\bigwedge_{i \in I'} X_i \geq t_i \mid \bigwedge_{j \in J'} X_j \geq t_j, \bigwedge_{j \in J \setminus J'} X_j \geq 0] \\ &\geq \Pr[\bigwedge_{i \in I'} X_i \geq t_i \mid \bigwedge_{j \in J} X_j \geq t_j] \end{aligned}$$

by the last monotonicity lemma.

5 A General Methodology

Although the proof of the previous section has the advantage of being elementary, it has the major drawback of being too specific. One would like a more general theory that captures the underlying notion of negative dependence of random variables of which the bins and balls example is only a special case. In this section, we invoke a general concept of ***negative association*** of random variables and related techniques from the theory of multi–variate probability inequalities, [2, 3], to show that the random variables X_i from the balls and bins experiment are in fact negatively related in a very strong sense. Most of our results from the previous sections will then follow as easy corollaries. Next we show that the well-known Chernoff-Hoeffding bounds for independent variables can also be applied to negatively associated variables. We use this to give a simple proof of a result of Panconesi and Srinivasan[12, 16, 11]. In fact, our formulation brings out clearly why the CH-bounds hold in their case. Moreover, this methodology has potentially much wider applicability in the analysis of probabilistic algorithms as suggested in § 1.2.3. We also introduce a notion of *strong* negative association and point out an application.

In this section, n and m are integers with $n, m \geq 2$. For a positive integer k, we shall regard R^k as a poset with the component–wise ordering. A function $f : \mathsf{R}^k \rightarrow \mathsf{R}$ is non–decreasing if it respects this ordering, i.e. is a poset–homomorphism.

5.1 Negative Association

In [3], the following notion of negatively related random variables is proposed:

Definition 10 *The vector of random variables,* $\mathbf{X} := (X_1, \cdots, X_n)$ *is said to be* **negatively associated** *if for every two disjoint index sets,* $I, J \subseteq [n]$,

$$\mathrm{cov}(f(X_i, i \in I), g(X_j, j \in J)) \leq 0$$

that is,

$$E[f(X_i, i \in I)g(X_j, j \in J)] \leq E[f(X_i, i \in I)]E[g(X_j, j \in J)]$$

for all non–decreasing functions $f : \mathsf{R}^{|I|} \rightarrow \mathsf{R}$ *and* $g : \mathsf{R}^{|J|} \rightarrow \mathsf{R}$.

OBSERVATIONS:

1. Equivalent formulations are: $\mathbf{X} := (X_1, \cdots, X_n)$ is negatively associated if
   (a) For all disjoint index sets $I, J \subseteq [n]$, and non–decreasing $f : \mathsf{R}^{|I|} \rightarrow \mathsf{R}$, $g : \mathsf{R}^{|J|} \rightarrow \mathsf{R}$ and $s, t \in \mathsf{R}$,
   $$\Pr[f(X_i, i \in I) \geq s, g(X_j, j \in J) \geq t] \leq \Pr[f(X_i, i \in I) \geq s] \cdot \Pr[g(X_j, j \in J) \geq t].$$
   (b) For all disjoint index sets $I, J \subseteq [n]$, and filters $\mathcal{A} \subseteq \mathsf{R}^{|I|}, \mathcal{B} \subseteq \mathsf{R}^{|J|}$,
   $$\Pr[(X_i, i \in I) \in \mathcal{A}, (X_j, j \in J) \in \mathcal{B}] \leq \Pr[(X_i, i \in I) \in \mathcal{A}] \cdot \Pr[(X_j, j \in J) \in \mathcal{B}].$$
2. For two binary–valued variables X, Y, it is necessary and sufficient that $\mathrm{cov}(X, Y) \leq 0$ for X, Y to be negatively associated. However, in general the condition that $\mathbf{X}$ is negatively associated is strictly stronger than the condition $\mathrm{cov}(X_i, X_j) \leq 0$ for each $i, j \in [n]$. In general, it can be quite difficult to establish that a set of variables is negatively associated. Below, we give some properties that can be used to establish negative association in certain situations.
3. If $\mathbf{X} = (X_1, \cdots, X_n)$ is negatively associated, then
   $$\Pr[\bigwedge_{i \in I} X_i \geq t_i, \bigwedge_{j \in J} X_j \geq t_j] \leq \Pr[\bigwedge_{i \in I} X_i \geq t_i] \cdot \Pr[\bigwedge_{j \in J} X_i \geq t_j]$$
   for all disjoint index sets $I, J \subseteq [n]$, and all reals $t_i, i \in$, and $t_j, j \in J$. This follows by taking f, g to be the indicator functions for the sets $\{\mathbf{a} \in \mathsf{R}^n \mid \bigwedge_{i \in I} a_i \geq t_i\}$ and $\{\mathbf{a} \in \mathsf{R}^n \mid \bigwedge_{j \in J} a_j \geq t_j\}$ respectively. By induction, this also implies that
   $$\Pr[\bigwedge_{i \in [n]} X_i \geq t_i] \leq \prod_{i \in [n]} \Pr[X_i \geq t_i]$$
   for any reals $t_i, i \in [n]$.

The following proposition lists some basic properties enabling one to construct negatively associated variables [3]:

Proposition 11 *1. A single variable by itself is negatively associated.*

2. *If $\mathbf{X} := (X_1, \cdots, X_n)$ and $\mathbf{Y} := (Y_1, \cdots, Y_m)$ are negatively associated, and $\mathbf{X}, \mathbf{Y}$ are independent, then the augmented vector $(\mathbf{X}, \mathbf{Y}) = (X_1, \cdots, X_n, Y_1, \cdots, Y_m)$ is negatively associated.*

3. *If $U_1, \cdots, U_n$ are independent, then $\mathbf{U} := (U_1, \cdots, U_n)$ is negatively associated.*

4. *If $X_1, \cdots, X_n$ are negatively associated, then so is any subset of these variables.*

5. *Let $\mathbf{X} := (X_1, \cdots, X_n)$ be negatively associated. Let $I_1, \cdots, I_k \subseteq [n]$ be disjoint index sets, for some positive integer k. For $j \in [k]$, let $h_j : \mathrm{R}^{|I_k|} \rightarrow \mathrm{R}$ be non-decreasing (or non-increasing) functions, and define $Y_j := h_j(X_i, i \in I_j)$. Then the vector $\mathbf{Y} := (Y_1, \cdots, Y_k)$ is also negatively associated. That is, non-decreasing (or non-increasing) functions of disjoint subsets of negatively associated variables are also negatively associated.*

5.2 Negative Association in Balls and Bins

We use Proposition 11 to give a simple yet completely general solution to the balls and bins problem. For the situation where m balls are throw into n bins, define indicator random variables $X_{i,j}$ for $i \in [n], j \in [m]$ by

$$X_{i,j} := \begin{cases} 1, & \text{if ball } j \text{ goes into bin } i; \\ 0, & \text{otherwise;} \end{cases}$$

We start with the following intuitively appealing result.

Lemma 12 *For any $j \in [m]$, the variables $X_{1,j}, \cdots, X_{n,j}$ are negatively associated.*

Proof. Let $I, J \subseteq [n]$ be any two disjoint index sets; without loss of generality, we can by renumbering arrange it so that $I := \{1, \ldots, |I|\}$ and $J := \{n-|J|+1, \ldots, n\}$. We would like to show that for any two non–decreasing functions $f : \mathrm{R}^{|I|} \rightarrow \mathrm{R}$ and $g : \mathrm{R}^{|J|} \rightarrow \mathrm{R}$, $E[f(X_i, i \in I)g(X_j, j \in J)] \leq E[f(X_i, i \in I)]E[g(X_j, j \in J)]$. Without loss of generality, we can, once again by renumbering, arrange it so that $f(0, \cdots 0, 0) \leq f(0, \cdots 0, 1) \leq \cdots, \leq f(1, 0, \cdots, 0)$ and $g(0, 0, \cdots, 0) \leq g(1, 0, \cdots, 0) \leq \cdots \leq g(0, 0, \cdots, 1)$. Consider the linear order L, on n–tuples consisting of exactly one 1 and $n-1$ zeroes, whose order is given by:

$$(0, 0, \ldots, 0, 1) < (0, 0, \ldots, 1, 0) < \cdots < (1, 0, \ldots, 0).$$

Define functions $\hat{f}, \hat{g} : L \rightarrow \mathrm{R}$ by

$$\hat{f}(a_1, \cdots, a_n) := f(a_i, i \in I)$$
$$\hat{g}(a_1, \cdots, a_n) := g(a_j, j \in J).$$

By the definition of the ordering on L, and using the fact that f, g are non–decreasing (in the component-wise ordering), we observe that $\hat{f}$ is non–decreasing while $\hat{g}$ is non–increasing in L. By the FKG inequality, with the measure $\mu(\mathbf{a}) := \Pr[X_{1,j} = a_1, \cdots, X_{n,j} = a_n] = \frac{1}{n}$, we get

$$\sum_{\mathbf{a} \in L} \hat{f}(\mathbf{a})\hat{g}(\mathbf{a})\mu(\mathbf{a}) \leq \sum_{\mathbf{a} \in L} \hat{f}(\mathbf{a})\mu(\mathbf{a}) \cdot \sum_{\mathbf{a} \in L} \hat{g}(\mathbf{a})\mu(\mathbf{a})$$

which is exactly the desired inequality. ∎

Since the balls are thrown independently of each other, the variables $(X_{1,j}, \cdots, X_{n,j})$ are independent of the rest of the variables. Hence from property (1) of Proposition 11, we deduce that the full vector $(X_{i,j})_{i \in [n], j \in [m]}$ is also negatively associated. Now, note that $X_i = \sum_{j \in [m]} X_{i,j}$ for each $i \in [n]$. For each $i \in [n]$, $f_i : \mathrm{R}^{nm} \rightarrow \mathrm{R}$, with $f_i(\mathbf{a}) := \sum_{j \in [m]} a_{i,j}$ is non–decreasing. Finally using property (2) from Proposition 11 above we establish the negative relation between the number of balls in each bin in full generality:

Theorem 13 *Let* $\mathbf{X} := (X_1, \cdots, X_n)$ *be the vector of the number of balls in the bins. Then* $\mathbf{X}$ *is negatively associated.*

This general result includes most of the statements proved in the preceding sections as corollaries. For instance, the result that

$$\Pr[\bigwedge_{i\in I} X_i \geq t_i \mid \bigwedge_{j\in J} X_j \geq t_j] \leq \Pr[\bigwedge_{i\in I} X_i \geq t_i]$$

is a corollary to the negative association as observed above. However results such as

$$\Pr[X_1 + \cdots + X_{n-1} \geq t \mid X_2 + \cdots + X_n \geq s] \leq \Pr[X_1 + \cdots + X_{n-1} \geq t]$$

which are implied by the FKG–proof of § 3 do not follow from this result. Nor do the results of the form $\Pr[X_1 \geq t_1 \mid X_2 \geq t_2, X_3 \geq t_3] \leq \Pr[X_1 \geq t_1 \mid X_2 \geq t_2]$. For the latter, however, one can prove that the variables $X_{i,j}$ are negatively related in the stronger sense of the following definition:

Definition 14 *The vector* $\mathbf{X} := (X_1, \cdots, X_n)$ *is* **strongly negatively associated** *if for all disjoint index sets* $I_1, \cdots, I_{k+1} \subseteq [n]$*, and all non-descreasing functions* $f_j : \mathsf{R}^{|I_j|} \rightarrow \mathsf{R}$ *for* $j \in [k+1]$*,*

$$\begin{aligned} &E[f_1(X_i, i \in I_1) \cdots f_{k+1}(X_i, i \in I_{k+1})] \cdot E[f_2(X_i, i \in I_2) \cdots f_k(X_i, i \in I_k)] \\ \leq\ &E[f_1(X_i, i \in I_1) \cdots f_k(X_i, i \in I_k)] \cdot E[f_2(X_i, i \in I_2) \cdots f_{k+1}(X_i, i \in I_{k+1})]. \end{aligned}$$

Moreover, the properties of negative association from Proposition 11 carry over to strong negative association as well. Hence, the variables X_i are actually strongly negatively associated. For strongly negatively associated variables $\mathbf{X}$, the following holds: For any disjoint index sets $I, J \subseteq [n]$ and any $I' \subseteq I, J' \subseteq J$, and any reals $t_i, i \in I$ and $t_j, j \in J$

$$\Pr[\bigwedge_{i\in I} X_i \geq t_i \mid \bigwedge_{j\in J} X_j \geq t_j] \leq \Pr[\bigwedge_{i\in I'} X_i \geq t_i \mid \bigwedge_{j\in J'} X_j \geq t_j].$$

This yields all the results of the previous section.

5.3 Negative Association and Chernoff–Hoeffding Bounds

A property of negatively associated random variables that is very useful is that one can apply the Chernoff–Hoeffding(CH) bounds to give tail estimates on their sum; in effect, for purposes of stochastic bounds on the sum, one can treat the variables as if they were independent.

Proposition 15 *Let* $\mathbf{X} := (X_1, \cdots, X_n)$ *be negatively associated (where all variables* X_i *are bounded) and let* $X := X_1 + \cdots + X_n$*. Then, for any* $\delta > 0$*, we have*

$$\Pr[|X - E[X]| > \delta E[X]] < 2 \cdot \exp(\frac{-\delta^2}{3} E[X]).$$

Proof. We use the standard proof of the CH–bound, see for example, [1, 10]. The only change needed is in a crucial step, where one uses the fact that for *independent* variables, $E[e^{tX}] = E[\prod_i e^{tX_i}] = \prod_i E[e^{tX_i}]$. For negatively associated variables, we have, for $t > 0$, $E[e^{tX}] = E[\prod_i e^{tX_i}] \leq \prod_i E[e^{tX_i}]$, because the functions e^{tX_i} are non-decreasing functions of disjoint argument sets. The rest of the proof is unchanged, and gives the upper tail bound. For the lower tail, we apply the same argument to the variables $b_i - X_i$, where b_i is a given bound on the variable X_i. Note that if the X_i variables are negatively associated, then so are the variables $b_i - X_i$. ∎

We give an application taken from [12, 16, 11]. There a certain notion of *self–weakening* or 1*–correlated* variables is defined, and it is shown that the CH–bound extends to sums of such variables. This extention is useful but somewhat *ad hoc*. Here we can see clearly that it is no co-incidence that CH-bounds can be applied in their case, for such self–weakening variables are special cases of negative association. Note that in the proof above, we avoid any expansion and manipulations of Taylor series, as in [11, 12, 16].

In their analysis of an edge colouring algorithm, Panconesi and Srinivasan [12, 16, 11] have to analyse the balls and bins experiment. Specifically, they define indicator variables $Y_i := 1$ iff the ith bin is empty, and seek to stochastically bound the sum $Y_1 + \cdots + Y_n$. These variables are not independent, preventing a direct application of the CH–bounds. However, we note that $Y_i = 1 - \text{sgn}(X_i)$, for $i \in [n]$, where X_i is the number of balls in the ith bin. As shown above, the X_i variables are negatively associated. Now the Y_i variables are non–increasing functions of disjoint variables, which themselves are negatively associated; hence the vector $(Y_1, \cdots, Y_n)$ is also negatively associated. One can now apply the CH–bound to get tail estimates for $\Pr[Y_1 + \cdots + Y_n > s]$.

6 Unresolved Issues

Since in general, it can be quite difficult to establish that a certain set of variables is negatively associated, we would like to have useful sufficient conditions under which this obtains. The random variables $\mathbf{X} := (X_1, \cdots, X_n)$ are said to satisfy the *negative monotone regression property* if: For each $i \in [n]$, and for each non–decreasing function $f : \mathsf{R} \to \mathsf{R}$,

$$E[f(X_i) \mid X_1 = t_1, \cdots, X_i = t_i]$$

is non–increasing in $(t_1, \cdots, t_i) \in \mathsf{R}^i$. Our indicator variables $X_{i,j}, i \in [n], j \in [m]$ also have satisfy the negative monotone regression property.

Lemma 16 *For any $(i,j) \in [n] \times [m]$ and index set $K \subseteq [n] \times [m]$ such that $(i,j) \notin K$, and any $f : \{0,1\} \to \mathsf{R}$ such that $f(0) \leq f(1)$, we have*

$$E[f(X_{i,j}) \mid \bigwedge_{(k,l) \in K} X_{k,l} = t_{k,l}]$$

is non–increasing in $(t_{k,l})_{(k,l)\in K} \in \{0,1\}^{|K|}$.

Proof. First observe that

$$E[f(X_{i,j}) \mid \bigwedge_{(k,l) \in K} X_{k,l} = t_{k,l}] = E[f(X_{i,j}) \mid \bigwedge_{(k,l) \in K'} X_{k,l} = t_{k,l}]$$

for $K' := K \cap [n] \times \{j\}$, since each ball is thrown independently of the others. So, if $(t_{k,l})_{(k,l)\in K} \leq (t'_{k,l})_{(k,l)\in K}$ with $t_{k,l} = t'_{k,l}$ for $(k,l) \in K'$, then

$$E[f(X_{i,j}) \mid \bigwedge_{(k,l) \in K} X_{k,l} = t_{k,l}] = E[f(X_{i,j}) \mid \bigwedge_{(k,l) \in K} X_{k,l} = t'_{k,l}].$$

In general,

$$E[f(X_{i,j}) \mid \bigwedge_{(k,l) \in K} X_{k,l} = t_{k,l}] = f(0)p(0) + f(1)p(1)$$

where, for $i = 0, 1$,

$$p(i) := \Pr[X_{i,j} = 1 \mid \bigwedge_{(k,l) \in K'} X_{k,l} = t_{k,l}].$$

On the other hand, if $(t_{k,l})_{(k,l)\in K'} < (t'_{k,l})_{(k,l)\in K'}$, then

$$\Pr[X_{i,j} = 1 \mid \bigwedge_{(k,l) \in K'} X_{k,l} = t'_{k,l}] = 0$$

and so

$$E[f(X_{i,j}) \mid \bigwedge_{(k,l) \in K'} X_{k,l} = t'_{k,l}] = f(0).$$

Since $f(0) \leq f(1)$, we have $f(0)p(0) + f(1)p(1) \geq f(0)$ for any probabilities $p(0), p(1)$ summing to 1. The result is established. ∎

We conjecture that the negative monotone regression condition implies the negative association of the variables. Besides giving an alternative proof of our result, this would be an extremely useful tool in establishing the negative association of variables in applications such as that in § 1.2.3. Having used monotone regression to show that the variables corrseponding to the children of a node in the recurrence tree are negatively associated (conditioned on the value of the parent node), a straightforward induction shows that the variables corresponding to different paths are also negatively associated. One can then analyse a single path by itself and subsequently apply large–deviation bounds as if they were independent.

Another, rather ambitious, task would be to resolve the following kind of mixed conditions. We know that $\Pr[X_1 \geq t_1 \mid X_2 \geq t_2] \leq \Pr[X_1 \geq t_1]$ and also that $\Pr[X_1 \geq t_1 \mid X_3 \leq t_3] \geq \Pr[X_1 \geq t_1]$. What can one say about $\Pr[X_1 \geq t_1 \mid X_2 \geq t_2, X_3 \leq t_3]$? What one would really want is a *calculus of correlations* that enables one, in a general way, under certain circumstances, to combine several such correlations into one. That is, given $\Pr[A \mid B] \leq \Pr[A]$, and $\Pr[A \mid C] \geq \Pr[A]$, under which circumstances can one also obtain $\Pr[A \mid B, C] \leq \Pr[A]$? In this context the work of Shepp[14, 15] and Winkler, [18] might be relevant.

7 Acknowledgements

We would like to thank Thomas Lauer and Torben Hagerup for bringing the non–triviality of the problem to our attention and Shiva Chaudhuri, Torben Hagerup and Kurt Mehlhorn for valuable discussions. We thank Suresh Chari for quickly faxing us a valuable but elusive reference. We thank Alessandro Panconesi and Volker Priebe for commenting on an earlier draft. Last but not the least, we thank Juris Hartmanis for asking us in disbelief "Didn't Bernoulli do this already?" which induced us to look further into the relevant literature.

References

[1] N. Alon, P. Erdös, J. Spencer, *The Probabilistic Method*, John Wiley, 1992.

[2] M.L. Eaton, *Lectures on Topics in Probability Inequalities*, CWI Tracts in Mathematics, 35. 1982.

[3] K. Joag-Dev and F. Proschan, "Negative Association of Random Variables with Applications", *Annals of Statistics*, 11:4, pp. 286–295, 1983.

[4] S. Choudhuri, D. Dubhashi, A. Panconesi, "Probabilistic Recurrence Relations (Revisited)", unpublished, 1994.

[5] C.M. Fortuin, J. Ginibre and P.N. Kasteleyn, "Corelational Inequalities for Partially Ordered Sets", *Communications of Mathematical Physics* 22, pp. 89–103, 1971.

[6] R.L. Graham, "Application of the FKG Inequality and its Relatives" in A. Bachem, M. Grötschel and B. Korte Ed., *Mathematical Programming: The State of the Art*, Springer–Verlag, 1983.

[7] Richard M. Karp, "Probabilistic Recurrence Relations", *Proc. 23 ACM Symp. on The Theory of Computing*, pp. 190–197, 1991.

[8] T. Lauer, Ph.D. Thesis, Universität des Saarlandes, in progress.

[9] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, Academic Press, New York, 1979.

[10] R. Motwani and P. Raghavan, *Randomized Algorithms*, forthcoming.

[11] A. Panconesi and A. Srinivasan, "Fast Randomized Algorithms for Distributed Edge Coloring", in *Proceedings of the ACM Symposium on Principles of Distributed Computing*, 1992, pp. 251–262.

[12] A. Panconesi, *Locality in Distributed Computing*, Ph.D. Thesis, Cornell University, 1993.

[13] T.K. Sarkar, "Some Lower Bounds of Reliability", Tech. Report. 124, Dept. of Operations Research and Statistics, Stanford University, 1969.

[14] L.A. Shepp, "The FKG Inequality and Some Monotonicity Properties of Partial Orders", *SIAM Journal on Algebraic and Discrete Methods* 1, pp. 295–299, 1980.

[15] L.A. Shepp, "The XYZ Conjecture and the FKG Inequality" *Annals of Probability* 10, pp. 824-827, 1982.

[16] A. Srinivasan, *Techniques for probabilistic analysis and randomness- efficient computation*, Ph.D. Thesis, Cornell University, 1993.

[17] J.P. Schmidt, A.Seigel and A. Srinivasan, "Chernoff–Hoeffding Bounds for Applications with Limited Independence", SODA, pp. 331–340, 1993.

[18] P.M Winkler, "Correlations among Partial Orders", *SIAM Journal on Algebraic and Discrete Methods*, 4(1), pp. 1–7, 1983.

A The Monotonicity Lemma

We give a proof of the crucial monotonicity lemma here.

Lemma 17 *Let $I = \{1, 2 \cdots k_1\}, J = \{n, n-1, ..n-k_2\}$ such that $I \cap J = \emptyset$. Let $l \in [n] \setminus I \cup J$. Let $a_i, i \in I$ and $a_j, j \in J$ and be arbitrary non-negative integers. Let E_1 be the event $\bigwedge_{i \in I} X_i \geq a_i$ and E_2 be the event $\bigwedge_{j \in J} X_j \geq a_j$. Let*

$$f(a) := \Pr_m^n[E_1 | E_2 \bigwedge X_l = a].$$

Then f is non-increasing in a.

Proof. First, note that we can assume without any loss of generality that each a_i is positive because if any a_i is zero we can remove the corresponding condition. It suffices to show that $f(a+1) \leq f(a)$ for arbitrary a. We begin by noting that,

$$f(a) := \Pr_{m-a}^{n-1}[E_1 | E_2]$$

and that

$$f(a+1) := \Pr_{m-a-1}^{n-1}[E_1 | E_2]$$

If we let $m_1 = m - a - 1$ and $n_1 = n - 1$ our task reduces to proving that

$$\Pr_{m_1}^{n_1}[E_1 | E_2] \leq \Pr_{m_1+1}^{n_1}[E_1 | E_2]$$

We prove this by induction on $|J|$.
Base Case: $|J| = 0$: In this case,

$$\begin{aligned}
\Pr_{m_1+1}^{n_1}[E_1 | E_2] &= \Pr_{m_1+1}^{n_1}[\bigwedge_{i \in I} X_i \geq a_i] \\
&= \Pr_{m_1}^{n_1}[\bigwedge_{i \in I} X_i \geq a_i] + \frac{1}{n_1} \sum_{u \in I} \Pr_{m_1}^{n_1}[\bigwedge_{i \in I \setminus \{u\}} X_i \geq a_i, X_u = a_u - 1] \\
&\geq \Pr_{m_1}^{n_1}[\bigwedge_{i \in I} X_i \geq a_i].
\end{aligned}$$

Induction Hypothesis: Assume that the statement is true for all J with $|J| \leq s$.
Induction Step: Let $|J| = s+1$.

$$\begin{aligned}
\Pr^{n_1}_{m_1+1}[E_1|E_2] &= \Pr^{n_1}_{m_1+1}[\bigwedge_{i\in I} X_i \geq a_i | \bigwedge_{j\in J} X_j = a_j] \\
&= \frac{\Pr^{n_1}_{m_1+1}[\bigwedge_{i\in I} X_i \geq a_i, \bigwedge_{j\in J} X_j \geq a_j]}{\Pr^{n_1}_{m_1+1}[\bigwedge_{j\in J} X_j \geq a_j]} \\
&= \frac{\Pr^{n_1}_{m_1}[\bigwedge_{i\in I} X_i \geq a_i, \bigwedge_{j\in J} X_j \geq a_j] + \sum_{l\in J} A_l + \Delta}{\Pr^{n_1}_{m_1}[\bigwedge_{j\in J} X_j \geq a_j] + \sum_{l\in J} B_l}
\end{aligned}$$

where,

$$\begin{aligned}
A_l &= \frac{1}{n_1}\Pr^{n_1}_{m_1}[\bigwedge_{i\in I} X_i \geq a_i, \bigwedge_{j\in J\setminus\{l\}} X_j \geq a_j, X_l = a_l - 1], \\
B_l &= \frac{1}{n_1}\Pr^{n_1}_{m_1}[\bigwedge_{j\in J\setminus\{l\}} X_j \geq a_j, X_l = a_l - 1],
\end{aligned}$$

and

$$\Delta = \frac{1}{n_1}\sum_{u\in I}\Pr^{n_1}_{m_1}[\bigwedge_{j\in J} X_j \geq a_j, \bigwedge_{i\in I\setminus\{u\}} X_i \geq a_i, X_u = t_u - 1].$$

Now, let

$$\begin{aligned}
A &= \Pr^{n_1}_{m_1}[\bigwedge_{i\in I} X_i \geq a_i, \bigwedge_{j\in J} X_j \geq a_j], \\
B &= \Pr^{n_1}_{m_1}[\bigwedge_{j\in J} X_j \geq a_j].
\end{aligned}$$

Then, we claim that for each $l \in I, A_l/B_l \leq A/B$. This follows becuase

$$\begin{aligned}
\frac{A}{B} &= \Pr^{n_1}_{m_1}[\bigwedge_{i\in I} X_i \geq a_i \mid \bigwedge_{j\in J} X_j \geq a_j] \\
&= \sum_{a\geq a_l} \Pr^{n_1}_{m_1}[\bigwedge_{i\in I} X_i \geq a_i \mid \bigwedge_{j\in J\setminus\{l\}} X_j \geq a_j, X_l = a] \cdot \Pr^{n_1}_{m_1}[X_l = a \mid \bigwedge_{j\in J} X_j \geq a_j]
\end{aligned}$$

Now, since $|J \setminus \{l\}| = s$, we can use the induction hypothesis to get that

$$\begin{aligned}
\frac{A}{B} &\leq \Pr^{n_1}_{m_1}[\bigwedge_{j\in J\setminus\{l\}} X_j \geq a_j, X_l = a_l - 1] \cdot \sum_{a\geq a_l} \Pr^{n_1}_{m_1}[X_l = a \mid \bigwedge_{j\in J} X_j \geq a_j] \\
&= A_l/B_l
\end{aligned}$$

Now, let

$$C = \sum_{l\in J} A_l + \Delta, \quad \text{and} \quad D = \sum_{l\in J} B_l.$$

Then, $\frac{C}{D} \geq \frac{A}{B}$. Noting that,

$$\frac{A+C}{B+D} \geq \frac{A}{B} \quad \Leftrightarrow \quad \frac{C}{D} \geq \frac{A}{B},$$

we get that $\frac{A+C}{B+D} \geq \frac{A}{B}$. In other words,

$$\Pr^{n_1}_{m_1+1}[\bigwedge_{i\in I} X_i \geq a_i | \bigwedge_{j\in J} X_j \geq a_j] \geq \Pr^{n_1}_{m_1}[\bigwedge_{i\in I} X_i \geq a_i | \bigwedge_{j\in J} X_j \geq a_j].$$

establishing the lemma. ∎