BRICS

Basic Research in Computer Science

BRICS RS-96-25 Dubhashi & Ranjan: Balls and Bins: A Study in Negative Dependence

Balls and Bins:
A Study in Negative Dependence

Devdatt Dubhashi
Desh Ranjan

BRICS Report Series **RS-96-25**

ISSN 0909-0878 **July 1996**

See back inner page for a list of recent publications in the BRICS Report Series. Copies may be obtained by contacting:

BRICS
Department of Computer Science
University of Aarhus
Ny Munkegade, building 540
DK - 8000 Aarhus C
Denmark

Telephone: +45 8942 3360
Telefax: +45 8942 3255
Internet: BRICS@brics.dk

BRICS publications are in general accessible through WWW and anonymous FTP:

```
http://www.brics.dk/
ftp ftp.brics.dk (cd pub/BRICS)
```

Balls and Bins: A Study in Negative Dependence *

Devdatt Dubhashi
BRICS†,
Department of Computer Science,
University of Aarhus,
Ny Munkegade,
DK-8000 Aarhus C, Denmark
Email: `dubhashi@daimi.aau.dk`

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

August 28, 1996

1 Introduction

This paper investigates the notion of negative dependence amongst random variables and attempts to advocate its use as a simple and unifying paradigm for the analysis of random structures and algorithms.

The assumption of independence between random variables is often very convenient for the several reasons. Firstly, it makes analyses and calculations much simpler. Secondly, one has at hand a whole array of powerful mathematical concepts and tools from classical probability theory for the analysis, such as laws of

*Work done partly at the Max–Planck–Institut für Informatik, Saarbrücken, Germany, and partially supported by the ESPRIT Basic Research Actions Program of the EC under contract No. 7141 (project ALCOM II)

†Basic Research in Computer Science,
Centre of the Danish National Research Foundation.

‡Work done while the author was visiting the Max–Planck–Institut für Informatik and BRICS.

large numbers, central limit theorems and large deviation bounds which are usually derived under the assumption of independence.

Unfortunately, the analysis of most randomized algorithms involves random variables that are *not* independent. In this case, classical tools from standard probability theory like large deviation theorems, that are valid under the assumption of independence between the random variables involved, cannot be used as such. It is then necessary to determine under what conditions of dependence one can still use the classical tools.

It has been observed before [32, 33, 38, 8], that in some situations, even though the variables involved are not independent, one can still apply some of the standard tools that are valid for independent variables (directly or in suitably modified form), provided that the variables are dependent in specific ways. Unfortunately, it appears that in most cases somewhat *ad hoc* strategems have been devised, tailored to the specific situation at hand, and that a unifying underlying theory that delves deeper into the nature of dependence amongst the variables involved is lacking.

A frequently occuring scenario underlying the analysis of many randomised algorithms and processes involves random variables that are, intuitively, dependent in the following *negative* way: if one subset of the variables is "high" then a disjoint subset of the variables is "low". In this paper, we bring to the forefront and systemize some precise notions of negative dependence in the literature, analyse their properties, compare them relative to each other, and illustrate them with several applications.

One specific paradigm involving negative dependence is the classical "balls and bins" experiment. Suppose we throw m balls into n bins independently at random. For $i \in [n]$, let B_i be the random variable denoting the number of balls in the ith bin. We will often refer to these variables as *occupancy numbers*. This is a classical probabilistic paradigm [16, 22, 26] (see also [31, § 3.1]) that underlies the analysis of many probabilistic algorithms and processes. In the case when the balls are identical, this gives rise to the well–known *multinomial distribution* [16, §VI.9]: there are m repeated independent trials (balls) where each trial (ball) can result in one of the outcomes $E_1, \ldots, E_n$ (bins). The probability of the realisation of event E_i is p_i for $i \in [n]$ for each trial. (Of course the probabilities are subject to the condition $\sum_i p_i = 1$.) Under the multinomial distribution, for any integers $m_1, \ldots, m_n$ such that $\sum_i m_i = m$ the probability that for each $i \in [n]$, event E_i occurs m_i times is

$$\frac{m!}{m_1! \ldots m_n!} p_1^{m_1} \ldots p_n^{m_n}.$$

The balls and bins experiment is a generalisation of the multinomial distribution: in the general case, one can have an arbitrary set of probabilities for each ball: the probability that ball k goes into bin i is $p_{i,k}$, subject only to the natural restriction that for each ball k, $\sum_i p_{i,k} = 1$. The joint distribution function correspondingly has a more complicated form.

A fundamental natural question of interest is: how are these B_i related? Note that even though the balls are thrown independently of each other, the B_i variables are *not* independent; in particular, their sum is fixed to m. Intuitively, the B_i's are negatively dependent on each other in the manner described above: if one set of variables is "high", a disjoint set is "low". However, establishing such assertions precisely by a direct calculation from the joint distribution function, though possible in principle, appears to be quite a formidable task, even in the case where the balls are assumed to be identical.

One of the major contributions of this paper is establishing that the the B_i are negatively dependent in a very strong sense. In particular, we show that the B_i variables satisfy *negative association* and *negative regression*, two strong notions of negative dependence that we define precisely below. All the intuitively obvious assertions of negative dependence in the balls and bins experiment follow as easy corollaries. We illustrate the usefulness of these results by showing how to streamline and simplify many existing probabilistic analyses in literature.

1.1 Organization

In § 2, we discuss discuss the notion of negative association. We examine its basic properties and relation to other better–known (but weaker) notions of negative dependence. Then we apply it in the context of the balls and bins experiment. We give a simple proof of a very simple assertion involving certain natural indicator variables that describe the balls and bins experiment. Though extremely simple, this result turns out to constitute a powerful and versatile technique for deriving various correlation inequalities in a deft and "calculation–free" manner. In particular, it follows that the occupancy numbers in the balls and bins expriment are negatively associated. In § 3 we discuss the notion of negative regression, and some of its variants. After discussing some general properties and relationships between these different notions of regression, we turn once again to apply it to the context of the balls and bins experiment. The major result of this section is that even in the most general balls and bins experiment, the occupancy numbers satisfy the negative regression property. The proof again is "calculation-free", but surprisingly non-trivial. (We actually prove a stronger result from which this is an easy consequence.) In § 4, we illustrate the usefulness of our results by applications of our results to probabilistic analyses in areas as diverse as simulation of parallel computers [8], dynamic load balancing [1],distributed graph algorithms [32, 33],and in random graphs and percolation theory [15, 29].

We shall restrict our attention exclusively to *discrete, non–negative integer–valued* random variables, as these are the ones of principal interest for the applications we have in mind. When we write conditional probabilities $\Pr[E \mid E']$, we are tacitly assuming that E' is an event of non–zero probability to avoid triviality.

2 Negative Association

A strong notion of negative dependence from the theory of multi–variate probability inequalities [12, 13, 39, 40] is that of *negative association*. The intuitive idea behind the definition of this strong notion of negative dependence is as follows: if a set of random variables is negatively related then if *any* monotone increasing function f of one subset of variables increases then *any* other monotone increasing function g of a disjoint set of variables must decrease. This is what is made formal below.

Definition 1 (Negative Association) *Let $\mathbf{X} := (X_1, \ldots, X_n)$ be a vector of random variables.*

$(-A)$ *The random variables, $\mathbf{X}$ are* **negatively associated** *if for every two disjoint index sets, $I, J \subseteq [n]$,*

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

for all functions $f : \mathsf{R}^{|I|} \rightarrow \mathsf{R}$ and $g : \mathsf{R}^{|J|} \rightarrow \mathsf{R}$ that are both non–decreasing or both non–increasing.

2.1 Properties of Negative Association

In this section, we collect together some useful properties of negatively associated variables.

Lemma 2 *Let $X_1, \ldots, X_n$ satisfy the negative association condition $(-A)$. Then for any non–decreasing functions $f_i, i \in [n]$,*

$$E[\prod_{i \in [n]} f_i(X_i)] \leq \prod_{i \in [n]} E[f_i(X_i)].$$

Proof. Take the non–decreasing functions $f(X_i, i < n) := \prod_{i<n} f_i(x_i)$ and $g(x_n) := f_n(x_n)$ to deduce that $E[\prod_{i \in [n]} f_i(X_i)] \leq E[\prod_{i<n} f_i(X_i)]E[f_n(X_n)]$ and now use induction. ∎

Many useful consequences of the $(-A)$ condition flow out of this simple lemma.

Proposition 3 *The negative association property $(-A)$ on a set of variables $X_1, \ldots, X_n$ implies the following notions of negative dependence:*

$(-COV)$ **Negative Covariance**: *for any $I \subseteq [n]$,*

$$E[\prod_{i \in I} X_i] \leq \prod_{i \in I} E[X_i].$$

$(-OD)$ Negative Right Orthant Dependence: *For any two disjoint subsets $I, J \subseteq [n]$,*

$$\Pr[X_i \geq t_i, i \in I \mid X_j \geq t_j, j \in J] \leq \Pr[X_i \geq t_i, i \in I].$$

Proof. For $(-COV)$, apply Lemma 2 with each f_i being the identity. For $(-OD)$, apply the definition of $(-A)$ with $f(a_i, i \in I) := \prod_{i \in I}[a_i \geq t_i]$, and $g(a_j, j \in J) := \prod_{j \in J}[a_j \geq t_j]$, the indicator functions of the two events $(X_i \geq t_i, i \in I)$ and $(X_j \geq t_j, j \in J)$, respectively. ∎

A very useful property of negative association is that the joint probability can be upper–bounded by the product of the marginals. This is another simple consequence of Lemma 2 applied with each $f_i(a_i) := [a_i \geq t_i]$, the indicator function of the event $X_i \geq t_i$.

Proposition 4 (Marginal Probability Bounds) *Let $X_1, \ldots, X_n$ satisfy $(-A)$. Then*

$$\Pr[X_i \geq t_i, i \in [n]] \leq \prod_{i \in [n]} \Pr[X_i \geq t_i].$$

A property of negatively associated random variables that is very useful in applications to the analysis of algorithms 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 5 ($-A$ and Chernoff–Hoeffding Bounds) *The Chernoff–Hoeffding bounds are applicable to sums of variables that satisfy the negative association condition $(-A)$.*

Proof. Let $X_1, \cdots, X_n$ be negatively associated (and bounded) variables. To show that the Chernoff–Hoeffding bounds apply to the sum $X := X_1 + \cdots + X_n$, we use the standard proof of the CH–bound, see for example, [3, 31]. 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 $t > 0$, $E[e^{tX}] = E[\prod_i e^{tX_i}] \leq \prod_i E[e^{tX_i}]$, by Lemma 2 applied with each $f_i(x) := e^{tx}$. 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 an upper bound on the variable X_i. Note that if the X_i variables are negatively associated, then so are the variables $b_i - X_i$. ∎

Remark 6 Colin McDiarmid (personal communication) has independently observed results in a similar vein.

Finally, the following proposition lists two simple but extremely useful properties of negative association [13]:

Proposition 7
1. *If* $\mathbf{X}$ *and* $\mathbf{Y}$ *satisfy* $(-A)$ *and are mutually independent, then the augmented vector* $(\mathbf{X}, \mathbf{Y}) = (X_1, \cdots, X_n, Y_1, \cdots, Y_m)$ *satisfies* $(-A)$.
2. *Let* $\mathbf{X} := (X_1, \cdots, X_n)$ *satisfy* $(-A)$. *Let* $I_1, \cdots, I_k \subseteq [n]$ *be disjoint index sets, for some positive integer* k. *For* $j \in [k]$, *let* $h_j : \mathsf{R}^{|I_k|} \to \mathsf{R}$ *be functions that are all non–decreasing or all non–increasing, and define* $Y_j := h_j(X_i, i \in I_j)$. *Then the vector* $\mathbf{Y} := (Y_1, \cdots, Y_k)$ *also satisfies* $(-A)$. *That is, non–decreasing (or non–increasing) functions of disjoint subsets of negatively associated variables are also negatively associated.*

2.2 Negative Association in Balls and Bins

We use Proposition 7 to give a simple "calculation–free" proof that the variables $B_1, \ldots, B_n$ are negatively associated. It is most expedient to introduce the indicator random variables $B_{i,k}$ for $i \in [n], k \in [m]$:

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

We start with the following intuitively appealing result which will turn out to be surprisingly powerful.

Lemma 8 (Zero–One Lemma for $(-A)$) *If* $X_1, \ldots, X_n$ *are zero-one random variables such that* $\sum_i X_i = 1$, *then* $X_1, \ldots, X_n$ *satisfy* $(-A)$.

We shall prove this by using the one–dimensional case of the famous FKG inequality [17, 3, 19], also known as Chebyshev's inequality [12, 39, 40] or as Harris' Lemma [20]:

Theorem 9 (Chebyshev, FKG, Harris) *Let* X *be a random variable on the real line, and let* $f, g : R \to R$ *be two functions.*

- *If* f, g *are both non-decreasing then*

$$E[f(X)g(X)] \geq E[f(X)]E[g(X)].$$

- *If* f *is non-decreasing and* g *is non-increasing then*

$$E[f(X)g(X)] \leq E[f(X)]E[g(X)].$$

Proof. (Of Zero–One Lemma): Let $X_1, X_2, .., X_n$ be zero-one random variables with exactly one $X_i = 1$. Let I and J be disjoint subsets of $[n]$ and let $f(a_i, i \in I)$ and $g(a_j, j \in J)$ be non-decreasing functions. Suppose by renumbering if necessary that $I := \{1, \ldots, |I|\}$, $J := \{n - |J| + 1, \ldots, n\}$ and that

$$f(0, .., 0) \leq f(0, .., 1) \leq .. \leq f(1, 0, ..0)$$

and

$$g(0,..,0) \leq g(1,..,0) \leq .. \leq g(0,..,1).$$

Note that since I and J are disjoint sets, this can always be arranged by renumbering. Define

$$X := i \quad \leftrightarrow \quad X_i = 1.$$

Thus X is a random variable taking values in $[n]$ with $\Pr[X = i] = p_i$ for some probabilities p_i summing to 1.

Set for $i \in [n]$,

$$f'(i) = \begin{cases} f(0,\ldots,0,\ldots,0), & i \notin I; \\ f(0,\ldots,1,\ldots,0), & i \in I. \end{cases}$$

and

$$g'(i) = \begin{cases} g(0,\ldots,0,\ldots,0), & i \notin J; \\ g(0,\ldots,1,\ldots,0), & i \in J. \end{cases}$$

where the 1 appears in the ith position. Observe that f' is non-increasing and g' is non-decreasing. Hence

$$E[f'(X)g'(X)] \leq E[f'(X)]E[g'(X)],$$

by the FKG–inequality. Finally observe that

$$\begin{aligned} E[f'(X)] &= E[f(X_i, i \in I)] \\ E[g'(X)] &= E[g(X_j, j \in J)] \\ E[f'(X)g'(X)] &= E[f(X_i, i \in I)g(x_j, j \in J)] \end{aligned}$$

and hence the conclusion of the Zero–One Lemma. ∎

Remark 10 The following simple proof of the Zero–One Lemma for $(-A)$ was communicated to us by Colin McDiarmid. By considering the non–negative functions $f(a_i, i \in I) - f(0,\ldots,0)$ and $g(a_j, j \in J) - g(0,\ldots,0)$ instead, we may assume that $f(0,\ldots,0) = 0 = g(0,\ldots,0)$. Then

$$E[f(X_i, i \in I)g(X_j, j \in J)] = 0 \leq E[f(X_i, i \in I)]E[g(X_j, j \in J)].$$

This completely elementary proof does not require the use of any inequality at all!

For any fixed $k \in [m]$, take $X_i := B_{i,k}, i \in [n]$ and use the Zero–One lemma to conclude that the indicator variables $(B_{i,k}, i \in [n])$ for any fixed $k \in [m]$ satisfy $(-A)$. Since the balls are thrown independently of each other, we obtain immediately from Proposition 7 the following consequence:

Proposition 11 *The full vector $(B_{i,j}, i \in [n], j \in [m])$ is negatively associated.*

Remark 12 Proposition 11 taken in conjunction with Proposition 7 will turn out to constitute a simple but extremely potent and versatile technique. We shall see many examples of how it can be used to provide deft "calculation–free" proofs of various correlation statements starting with the main result of this subsection, namely that the variables $B_1, \ldots, B_n$ are negatively associated (Proposition 13 below) and continuing with applications in the next sub–section. We thank Martin Dietzfelbinger for impressing this upon us, in particular for sharing some results of his own [7] which are intermediate in strength between some of our results.

Theorem 13 *Let* $\mathbf{B} := (B_1, \cdots, B_n)$ *be the vector of the number of balls in the bins. Then* $\mathbf{B}$ *is negatively associated.*

Proof. Apply Proposition 11 and Proposition 7 (2) together with the non–decreasing functions $B_i = \sum_{j \in [m]} B_{i,j}$ for each $i \in [n]$. ∎

Remark 14 Joag–Dev and Proschan [13] also prove Theorem 13 for the multinomial distribution (§ 3.1(a)) although their proof is a bit cryptic. They also claim without proof the same result for the general balls and bins experiment ("convolution of unlike multinomials").

Remark 15 Immediate consequences of this theorem are that the occupany numbers $B_1, \ldots, B_n$ satisfy the negative orthant dependence conditions, $(-OD)$,

$$\Pr[B_i \geq t_i, i \in I \mid B_j \geq t_j, j \in J] \leq \Pr[B_i \geq t_i, i \in I],$$

for any disjoint index sets $I, J \subseteq [n]$. However results such as

$$\Pr[B_i \geq t_i, i \in I \mid B_j \geq t_j, j \in J] \leq \Pr[B_i \geq t_i, i \in I \mid B_j \geq t'_j, j \in J],$$

for any disjoint index sets $I, J \subseteq [n]$ and for any reals $t'_j \leq t_j, j \in J$ do not follow. For this we turn to an apparently stronger notion of dependence in the next section.

2.3 Negative Association and the BK Inequality

In this subsection we try to relate the concept of negative association to the concept of "disjointly–occuring events" and the associated BK inequality which is widely used in Percolation Theory [20]. Consider the space (Ω, μ), where $\Omega := \{0,1\}^n$ for a positive integer n, endowed with the component–wise order and $\mu : \Omega \to \mathsf{R}$ is a measure, not necessarily the product measure. Denote for each $\omega \in \Omega$, $1(\omega) := \{i \mid \omega_i = 1\}$. Likewise, conversely, for $K \subseteq [n]$, denote $\Omega(K) := \{\omega \in \Omega \mid \omega_i = 1, i \in K\}$. For non–decreasing events $A, B \subseteq \Omega$, define

$$A \otimes B := \{\omega \in \Omega \mid \exists H \subseteq 1(\omega), \Omega(H) \subseteq A \text{ and } \Omega(1(\omega) \setminus H) \subseteq B\}. \qquad (1)$$

Definition 16 *The space (Ω, μ) is a* **BK space** *if*

$$\mu(A \otimes B) \leq \mu(A)\mu(B),$$

for all non–decreasing events $A, B \subseteq \Omega$.

The following result due to van den Berg and Kesten [6, 20] is widely used in Percolation Theory to complement the FKG inequality:

Theorem 17 (BK Inequality) *Let (Ω, μ) be a product space, that is, μ is a product measure, $\mu(\omega) = \prod_{i \in [n]} \mu_i(\omega_i)$, for probabilities $\mu_i(1) = p_i = 1 - \mu_i(0)$ for each $i \in [n]$. Then (Ω, μ) is a BK space.*

Remark 18 To see what this connective $\otimes$ means, it is helpful to view each co–ordinate ω_i as standing for a *resource*. Thus $\omega_i = 1$ iff resource i is available. A non–decreasing event A is enabled or established as soon as all the resources necessary for it are available. To establish two different non–decreasing events A, B, the resources necessary for both should be available. However, resources are consumed and cannot be reused. Thus to establish both events together, there must be partition of the available resources, one set enabling event A and the other the event B. The resource intuition is the basic intuition behind linear logic and the connective $\otimes$ is exactly the linear logic connective, [18] (see also [5] for a very readable account stressing the resource interpretation). In the literature in Percolation Theory [20, Chap. 2] (and the references therein) the connective is denoted $\circ$ and is discussed as "disjoint occurences of events" .

Let (Ω, μ) be a BK space with $\Omega := \prod_{i \in [n]} \Omega_i$, and each $\Omega_i := \{0, 1\}$. Let $I \subseteq [n]$ be fixed, and consider two cylindrical non–decreasing events $A = A_I \times \prod_{i \in [n] \setminus I} \Omega_i$ and $B = B_{[n] \setminus I} \times \prod_{i \in I} \Omega_i$ with $A_I \subseteq \prod_{i \in I} \Omega_i$ and $B_{[n] \setminus I} \subseteq \prod_{i \in [n] \setminus I} \Omega_i$. Note that in this case, $A \otimes B = A \wedge B$. Hence for such events in a BK–space, $\mu[A \wedge B] = \mu[A \otimes B] \leq \mu(A)\mu(B)$.

It is easily seen that

Observation 19 *Let $X_1, \ldots, X_n$ be 0/1 variables with $\sum_i X_i = 1$. Then their distribution forms a BK space.*

Further, we conjecture that

Conjecture 20 *BK spaces are preserved under direct products.*

If true, the conjecture together with Observation 19, would establish that the product space
$(B_{i,k}, i \in [n].k \in [m]) = \prod_{k \in [m]} (B_{i,k}, i \in [n])$, would also be a BK space. Actually one can verify directly that this product space is in fact also a BK space, but it would be neater to apply the conjecture.

Let $I, J \subseteq [n]$ be disjoint, and let E_I, E_J be non–decreasing events that depend only on the variables $(B_i, i \in I)$ and $(B_j, j \in J)$ respectively. Observe that these are disjoint cylindrical events in the BK–space of the underlying indicator variables. Hence by the remarks above,

$$\Pr[E_I \wedge E_J] \leq \Pr[E_I]\Pr[E_J].$$

This puts the results on negative association in balls and bins in the perspective of "disjointly–occuring events" from percolation theory [20, 6].

For some more remarks on the relation between the two notions, and an outline of how negative association can be applied to derive the BK inequality, see [9].

3 Negative Regression

Negative regression is possibly the most direct and compelling formulation of the intuition that when one set of variables is "high", a disjoint set is "low".

3.1 Negative Regression Conditions

Definition 21 *Let* $\mathbf{X} := (X_1, \ldots, X_n)$ *be a vector of random variables.* $\mathbf{X}$ *satisfies*

$(-R)$ *the* **negative regression** *condition if* $E[f(X_i, i \in I) \mid X_j = t_j, j \in J]$ *is non–increasing in each* $t_j, j \in J$ *for any disjoint* $I, J \subseteq [n]$ *and any non–decreasing function* f*.*

$(-LTR)$ *the* **negative left tail regression** *condition if* $E[f(X_i, i \in I) \mid X_j \leq t_j, j \in J]$ *is non–increasing in each* $t_j, j \in J$ *for any disjoint* $I, J \subseteq [n]$ *and any non–decreasing function* f*.*

$(-RTR)$ *the* **negative right tail regression** *condition if* $E[f(X_i, i \in I) \mid X_j \geq t_j, j \in J]$ *is non–increasing in each* $t_j, j \in J$ *for any disjoint* $I, J \subseteq [n]$ *and any non–decreasing function* f*.*

Remark 22 The negative regression condition $(-R)$ yields some stronger correlation inequalities in some cases than negative association. This, and the fact that it is highly intuitive, might make it a more appealing notion of negative dependence. Unfortunately, as we shall also see below, it does not seem as robust and versatile as negative association under monotone transformations of variables. This limits its applicability rather severly. A judicious combination of the two appears to be the optimal strategy.

3.2 Properties of Regression

We collect together some useful properties of the regression conditions.

We begin with the following proposition, which is intuitive and perhaps folklore, but we include a oomplete proof since the proof is tricky and instructive and we are unaware of another source where it has been published in detail.

Proposition 23 (Mixed Regression) *Let $X_1, \ldots, X_n$ be random variables satisfying the negative regression condition $(-R)$. Let $I, J, K \subseteq [n]$ be disjoint index sets. Then*

$$E[f(X_i, i \in I) \mid (X_j = t_j, j \in J), (X_k \geq t_k, k \in K)]$$

is non–increasing in each of $t_j, j \in J$ and $t_k, k \in K$ for an arbitrary non–decreasing function f.

Proof. We shall proceed by induction on on the size of K. If $K = \emptyset$, this is simply the condition $(-R)$. For the inductive step, let $l \in [n] \setminus I \cup J \cup K$ and consider

$$E[f(X_i, i \in I) \mid (X_j = t_j, j \in J), (X_k \geq t_k, k \in K), X_l \geq t_l].$$

It suffices to show that this is non–increasing in t_l. Fix integers $t_j, j \in J$ and $t_k, k \in K$ and let us abbreviate $X_j = t_j, j \in J$ by $X_J = t_J$ and similarly $X_k \geq t_k, k \in K$ by $X_K \geq t_K$ and $f(X_i, i \in I)$ by $f(X_I)$. It suffices now to show that for any integer a,

$$E[f(X_I) \mid X_J = t_J, X_K \geq t_K, X_l \geq a] \geq E[f(X_I) \mid X_J = t_J, X_K \geq t_K, X_l \geq a+1].$$

For this in turn, it suffices to prove that for any non–decreasing f, and any integer t_I,

$$\Pr[f(X_I) \geq t_I \mid X_J = t_J, X_K \geq t_K, X_l \geq a] \geq \Pr[f(X_I) \geq t_I \mid X_J = t_J, X_K \geq t_K, X_l \geq a+1].$$

Denote $\mathcal{C} := X_J = t_J, X_K \geq t_K$. We have,

$$\begin{aligned} \Pr[f(X_I) \geq t_I \mid \mathcal{C}, X_l \geq a] &= \frac{\Pr[f(X_I) \geq t_I, \mathcal{C}, X_l \geq a]}{\Pr[\mathcal{C}, X_l \geq a]} \\ &= \frac{A + C}{B + D}, \end{aligned}$$

where we put

$$\begin{aligned} A &:= \Pr[f(X_I) \geq t_I, X_J = t_J, X_K \geq t_K, X_l \geq a + 1] \\ B &:= \Pr[X_J = t_J, X_K \geq t_K, X_l \geq a + 1] \\ C &:= \Pr[f(X_I) \geq t_I, X_J = t_J, X_K \geq t_K, X_l = a] \\ D &:= \Pr[X_J = t_J, X_K \geq t_K, X_l = a] \end{aligned}$$

Then

$$
\begin{aligned}
\frac{A}{B} &= \Pr[f(X_I) \geq t_I \mid \mathcal{C}, X_l \geq a+1] \\
&= \sum_{t \geq a+1} \Pr[f(X_I) \geq t_I \mid \mathcal{C}, X_l = t] \cdot \Pr[X_l = t \mid \mathcal{C}, X_l \geq a+1] \\
&\quad \text{by induction} \\
&\leq \Pr[f(X_I) \geq t_I \mid \mathcal{C}, X_l = a] \cdot \sum_{t \geq a+1} \Pr[X_l = t \mid \mathcal{C}, X_l \geq a+1] \\
&= \frac{C}{D}.
\end{aligned}
$$

Hence $\frac{A+C}{B+D} \geq \frac{A}{B}$ which is what we needed to prove. ∎

Corollary 24 *The regression condition $(-R)$ implies both the tail regression conditions $(-RTR)$ and $(-LTR)$.*

Proof. Take $J := \emptyset$ in Proposition 23. ∎

Let the comparsion operators $\{<, \leq, =, \geq, >\}$ be ordered as follows:

$$< \ \preceq \ \leq \ \preceq \ = \ \preceq \ \geq \ \preceq \ >,$$

and let $?_i, i \in I$ stand for a sequence of comparison operators. The technique used in the proof of Proposition 23 can be used to prove the following intuitive assertion about a compound regression condition on the variable values and the comparison operators ordered by $\preceq$:

Corollary 25 (Compound Regression) *Let $I, J \subseteq [n]$ be disjoint, and let f be non–decreasing and $t_j, j \in J$ be arbitrary reals. If $X_1, \ldots, X_n$ satisfy $(-R)$, then*

$$E[f(X_i, i \in I) \mid X_j \quad ?_j \quad t_j, j \in J],$$

is non–increasing in each $t_j, j \in J$ and in each $?_j, j \in J$.

Next we state a sequence of properties analogous to those that obtained for the negative association condition.

Lemma 26 *Let $X_1, \ldots, X_n$ satisfy the negative regression condition (-R). Then for any index set $I \subseteq [n]$ and any non–decreasing functions $f_i, i \in I$,*

$$E[\prod_{i \in I} f_i(X_i)] \leq \prod_{i \in I} E[f_i(X_i)].$$

Proof. Without loss of generality, suppose $I := \{1, \ldots, |I|\}$ and denote $X_I := X_{|I|}, f_I := f_{|I|}$. Then

$$\begin{aligned}
E[\prod_{i \in I} f_i(X_i)] &= E[E[\prod_{i \in I} f_i(X_i) \mid X_I]] \\
&= E[E[\prod_{i \in I \setminus |I|} f_i(X_i) \mid X_I] f_I(X_I)] \\
&= \sum_a E[\prod_{i \in I \setminus |I|} f_i(X_i) \mid X_I = a] \cdot f_I(a) \Pr[X_I = a] \\
&\leq E[\prod_{i \in I \setminus |I|} f_i(X_i)] E[f_I(X_I)]
\end{aligned}$$

In the penultimate line we used the regression condition to apply the Chebyshev–FKG–Harris inequality, Theorem 9. Now the result follows by induction. ∎

Analogous to $(-A)$, the regression condition $(-R)$ also implies some other notions of negative dependence:

Proposition 27 *The negative regression property $(-R)$ on a set of variables $X_1, \ldots, X_n$ implies the following notions of negative dependence: negative covariance, $(-COV)$, and negative orthant dependence, $(-OD)$.*

Proof. The first assertion is proved by by applying Lemma 26. The second follows from Corollary 24. ∎

Again, like $(-A)$, the regression condition $(-R)$ has the very useful that the joint probability distribution can be upper–bounded by the product of the marginals:

Proposition 28 (Marginal Probability Bounds) *Let $X_1, \ldots, X_n$ be distributed to satisfy $(-R)$. Then*

$$\Pr[X_1 \leq t_1, \ldots, X_n \leq t_n] \leq \prod_{i \in [n]} \Pr[X_i \leq t_i].$$

Finally, we get Chernoff–Hoeffding bounds on sums of variables which satisfy the negative regression condition:

Proposition 29 ($-R$ and Chernoff–Hoeffding Bounds) *The Chernoff–Hoeffding bounds apply to sums of variables that satisfy the negative regression condition (-R).*

The proof, as in § 2 follows the standard route with Lemma 26 used (taking each $f_i(x) := e^{tx}$) to replace the equality $E[e^{t(X_1 + \cdots + X_n)}] = \prod_{i \in [n]} E[e^{tX_i}]$ (which holds for independent variables) by the inequality $E[e^{t(X_1 + \cdots + X_n)}] \leq \prod_{i \in [n]} E[e^{tX_i}]$, applying Lemma 26 with each $f_i(x) := e^{tx}$.

Remark 30 Colin McDiarmid (personal communication) has independently observed results in a similar vein.

3.3 Negative Regression in Balls and Bins

In this sub–section, we show that the variables $B_1, \ldots, B_n$ from the most general balls and bins experiment satisfy the negative regression condition, $(-R)$.

Theorem 31 *The vector* $\mathbf{B} := (B_1, \ldots, B_n)$ *satisfies the negative regression condition* $(-R)$*.*

Corollary 32 *The variables* $B_1, \ldots, B_n$ *satisfy the negative right and left tail regression conditions,* $(-RTR)$ *and* $(-LTR)$*.*

Proof. Apply Corollary 24. ∎

Let us start by considering the special case of Theorem 31 when all balls are identical (the bins need not be identical). This is the situation of the Multinomial Distribution. In this case, by symmetry between any two subsets of the balls of the same size, the conditioning $B_j = t_j, j \in J$ is equivalent to the simple unconditional balls and bins experiment with fewer balls and bins – precisely with $m' := m - \sum_{j \in J} t_j$ balls thrown into the bins labelled by the set $\bar{J} := [n] \setminus J$. Let us use superscripts to denote the variables in the experiment corresponding to throwing m balls into bins labelled by $I \subseteq [n]$ by $B_i^{m,I}, i \in I$. Then, our observation can be phrased as:

$$E[f(B_i^{m,[n]}, i \in I) \mid B_j^{m,[n]} = t_j, j \in J] = E[f(B_i^{m',\bar{J}}, i \in I)].$$

Finally, we conclude that this is a monotone increasing function in m' by noting that for each $i \in I$,

$$B_i^{m+1,I} = B_i^{m,I} + B_{i,m+1}.$$

Thus the $(-R)$ property holds easily in the case when all balls are identical.

Remark 33 A weaker form of this result was proved by Mallows [28]: he shows that in the case of identical balls, the joint probability distribution can be bounded by the product of the marginal distributions:

$$\Pr[B_1 \leq t_1, \ldots, B_n \leq t_n] \leq \prod_{i \in [n]} \Pr[B_i \leq t_i].$$

By Proposition 28, this is simple consequence of the regression property $(-R)$. Of course the regression condition $(-R)$ yields much more. Mallows claims the analogous result for the general case (balls not identical) but does not supply a proof. We shall prove a stronger version of the $(-R)$ property for the general case, when neither the bins nor the balls need be identical.

The general case appears to be surprisingly non–trivial by comparison, with many subtle technical difficulties. As a first indication of this, let us comment on why another plausible simple approach, analogous to that used in the proof of negative association, is not applicable.

Proposition 34 *The variables $B_{i,j}, i \in [n], j \in [m]$ satisfy the negative regression condition $(-R)$.*

As with negative association, it is true that the union of independent families of random variables satisfies $(-R)$ if each family satisfies it separately. Hence it suffices, as in the negative association case, to prove

Lemma 35 (Zero One Lemma for $(-R)$) *Let $X_1, \ldots, X_n$ be $0/1$ variables with $\sum_i X_i = 1$. Then they satisfy $(-R)$.*

Proof. Let $I, J \subseteq [n]$ be disjoint subsets and assume, without loss of generality that $n \in J$, It suffices to prove that

$$E[f(X_i, i \in I) \mid X_j = 0, j \in J] \geq E[f(X_i, i \in I) \mid X_n = 1, X_j = 0, n \neq j \in J].$$

Let $f_0 := f(0, \ldots, 0)$ and for $i \in I$, denote $f_i := f(0, \ldots, 1, \ldots, 0)$ (with the 1 in the ith place). Note that $f_0 \leq f_i$ for each $i \in I$. Then, for some probabilities $p_0, p_i, i \in I$ summing to 1,

$$\begin{aligned}
E[f(X_i, i \in I) \mid X_j = 0, j \in J] &= \sum_i f_i p_i \\
&\geq \sum_i f_0 p_i \\
&= f_0 \\
&= E[f(X_i, i \in I) \mid X_n = 1, X_j = 0, n \neq j \in J].
\end{aligned}$$

∎

Now, observing that for each $i \in [n]$, $B_i = \sum_{k \in [m]} B_{i,k}$, the $(-R)$ property would hold for $B_1, \ldots, B_n$ if we could, in analogy to the negative association property $(-A)$, transfer the property to disjoint sums of variables. Unfortunately, this is not true in general. There is a simple counter–example to the following plausible–sounding conjectures, see [11].

Conjecture 36

- *Sums of disjoint subsets of variables satisfying $(-R)$ also satisfy $(-R)$.*
- *Let $X_1, \ldots, X_n$ satisfy $(-R)$ and suppose $Y_1, \ldots, Y_n$ are a set of $0/1$ variables independent of the X variables, such that $\sum_i Y_i = 1$. Then $X_1 + Y_1, \ldots, X_n + Y_n$ also satisfy $(-R)$.*

Instead, we shall prove the following statement about a "mixed" negative regression condition involving both the indicator variables $B_{i,j}$ and the occupancy numbers $B_1, \ldots, B_n$.

Theorem 37 *Let I and J be disjoint subsets of $[n]$ and let f be a non–decreasing function. Then*

$$E[f(B_{i,k}, i \in I, k \in [m]) \mid B_j = t_j, j \in J],$$

is non–increasing in each $t_j, j \in J$.

Remark 38 Note that the variables $B_{i,k}, i \in I, k \in [m]$ are disjoint from the indicator variables involved in the condition on the right. By considering $f(\sum_k B_{i,k}, i \in I)$, we get Theorem 31, the $(-R)$ condition for the occupancy numbers $B_1, \ldots, B_n$ as an immediate corollary.

We shall now embark on the proof of Theorem 37. For a start, let us introduce some notation.

Notation 39 Let $S_i \subseteq [m]$ denote the set of balls in bin i for $i \in [n]$ Thus $\bigcup_i S_i = [m]$ and $|S_i| = B_i$ for $i \in [n]$. For a subset $J \subseteq [n]$, we use the abbreviations $S_J := (S_j, j \in J)$ and $S(J) := \bigcup_{j \in J} S_j$. As usual, let I and J be disjoint subsets of $[n]$, and let $f(B_{i,k}, i \in I, k \in [m])$ be a an arbitrary non–decreasing function.

Recall, that in the case of identical balls, conditioning on the event $B_J = t_J$ was equivalent to an unconditional experiment involving the remaining balls and bins. The analogue of this assertion in the general case is stated next. Let us use the subscripts in $E_{I,K}$ etc. to denote the statistics of the balls and bins experiment restricted to the subset K of balls distributed independently into the subset I of bins with probabilities proportional to the original ones. That is, for $I \subseteq [n]$, and $K \subseteq [m]$,

$$\Pr_{I,K}[B_{i,k} = 1] = \begin{cases} \frac{p_{i,k}}{1 - \sum_{j \notin I} p_{j,k}}, & \text{if } i \in I, k \in K; \\ 0, & \text{otherwise.} \end{cases}$$

Proposition 40 *Let $K \subseteq [m]$. Then for any event E_I involving the variables $B_{i,k}, i \in I, k \in [m]$,*

$$\Pr[E_I \mid S(J) = K] = \Pr_{\overline{J},\overline{K}}[E_I].$$

That is, conditioning on the event $S(J) = K$ is equivalent to an unconditional balls and bins experiment involving the subset $\overline{K}$ of balls distributed in the subset $\overline{J}$ of bins with probabilities that are proportional to the original ones.

Proof. First, we compute $\Pr[B_{i,k^*} = 1 \mid S(J) = K]$ for $k^* \notin K$ and $i \in I$. Let $K' := [m] \setminus (K \cup \{k^*\})$. Then,

$$\begin{aligned} \Pr[B_{i,k^*} = 1 \mid S(J) = K] &= \frac{\Pr[B_{i,k^*} = 1, S(J) = K]}{\Pr[S(J) = K]} \\ &= \frac{\Pr[k^* \in S_i, (k \in S(J), k \in K), (k \notin S(J), k \in K')]}{\Pr[(k \in S(J), k \in K), (k \notin S(J), k \notin K)]} \\ &= \frac{\Pr[k^* \in S_i]}{\Pr[k^* \notin S(J)]}, \quad \text{by independence of the balls} \quad (2) \\ &= \frac{p_{i,k^*}}{1 - \sum_{j \in J} p_{j,k^*}} \end{aligned}$$

Thus each remaining ball is thrown into the remaining bins with probabilities proportional to the original ones.

Next we verify that the remaing balls are also thrown independently of each other even under the conditioning $S(J) = K$. It suffices to show for a set of pairs $P := \{(i, k) \mid i \in I, k \notin K\}$, that $\Pr[B_{i,k} = 1, (i, k) \in P \mid S(J) = K] = \prod_{(i,k)\in P} \Pr[B_{i,k} = 1 \mid S(J) = K]$. We have by a computation similar to the one above,

$$\begin{aligned}
\Pr[B_{i,k} = 1, (i, k) \in P \mid S(J) = K] &= \prod_{(i,k)\in P} \frac{\Pr[k \in S_i]}{\Pr[k \notin S(J)]} \\
&= \prod_{(i,k)\in P} \Pr[B_{i,k} = 1 \mid S(J) = K], \quad \text{using (2).}
\end{aligned}$$

∎

Corollary 41 *Let*

$$\hat{f}(K) := E[f(B_{i,k}, i \in I, k \in [m]) \mid S(J) = K].$$

Then f is non–increasing in K.

Proof. By Proposition 40, we have

$$\hat{f}(K) = E_{\overline{J,K}}[f(B_{i,k}, i \in I, k \in [m])],$$

and the result follows by a trivial coupling, since $\Pr_{\overline{J,K}}[B_{i,k} = 1] = 0$ for $k \in K$ and $i \in I$ while $\Pr_{\overline{J,K}}[B_{i,k} = 1] = \Pr_{\overline{J,K'}}[B_{i,k} = 1]$ for any $i \in I$, any $K, K' \subseteq [m]$ and $k \notin K, K'$. ∎

Remark 42 Corollary 41 does not follow readily from Proposition 34. The additional information from Proposition 40 relating a conditional experiment to an unconditional one is used in an essential manner.

Now we are ready to prove Theorem 37.
Proof. (of Theorem 37): We need to show that for disjoint index sets $I, J \subset [n]$, any non–decreasing f, and any fixed integers $t_j \leq t'_j, j \in J$,

$$E[f(B_{i,k}, i \in I, k \in [m]) \mid (B_j = t_j, j \in J)] \geq E[f(B_{i,k}, i \in I, k \in [m]) \mid (B_j = t'_j, j \in J)],$$

that is, with the abbreviation $f(B_I) := f(B_{i,k}, i \in I, k \in [m])$, and abbreviations in Notation 39,

$$E[f(B_I) \mid B_J = t_J] \geq E[f(B_I) \mid B_J = t'_J].$$

By partitioning the probability space, we can write, for K ranging over all subsets of $[m]$ of size $\sum_{j\in J} t_j$,

$$E[f(B_I) \mid B_J = t_J] \quad = \quad \sum_K E[f(B_I) \mid S(J) = K]\Pr[S(J) = K \mid B_J = t_J]$$

$$= \frac{\sum_K \hat{f}(K)\Pr[S(J) = K]}{\Pr[B_J = t_J]}$$

$$= \frac{\sum_K \hat{f}(K)\mu(K)}{\sum_K \mu(K)} \tag{3}$$

where we put $\hat{f}(K) := E[f(B_I) \mid S_J = K]$, and

$$\mu(K) := \Pr[S(J) = K].$$

Similarly, with K' ranging over all subsets of $[m]$ of sizes $\sum_{j \in J} t'_j$,

$$E[f(B_I) \mid B_J = t'_J] = \frac{\sum_{K'} \hat{f}(K')\mu(K')}{\sum_{K'} \mu(K')}. \tag{4}$$

Interrupting for a check, let us return to the case where all the balls are identical (the bins may not be identical). In this case, $\Pr[S_J = K \mid B_J = t_J]$, and $\hat{f}(K)$ depend only on $|K|$. Let us denote these quantities by p_k and f_k respectively. Then, by Lemma 41 $f_k \geq f_{k'}$ if $k \leq k'$ and the inequality follows immediately by comparing (3) and (4).

Let's get back to the general case.Observe that for $K \subseteq [m]$,

$$\mu(K) = \prod_{k \in K} (\sum_{j \in J} p_{j,k}) \prod_{k \in [m] \setminus K} (1 - \sum_{i \notin J} p_{i,k}).$$

By Lemma 41, for $K \subseteq K'$, $\hat{f}(K) \geq \hat{f}(K')$. Thus we conclude the proof by comparing (3) and (4) using the following Expectation Levels Lemma applied to $-f$ (note that f is non–increasing iff $-f$ is non–decreasing). ∎

Lemma 43 (Expectation Levels Lemma) *Let μ be a product measure on the lattice of all subsets of $[m]$ defined by*

$$\mu(K) := \prod_{k \in K} p_k \prod_{k \notin K} q_k,$$

for arbitrary reals $p_k, q_k, k \in [m]$. Let f be a non–decreasing function on the lattice. Then,

$$\frac{\sum_{|K|=t} f(K)\mu(K)}{\sum_{|K|=t} \mu(K)},$$

is non–decreasing in t.

Proof. It suffices to show, for any $a \geq 0$ that

$$\frac{\sum_{|K|=a} f(K)\mu(K)}{\sum_{|K|=a} \mu(K)} \leq \frac{\sum_{|K'|=a+1} f(K')\mu(K')}{\sum_{|K'|=a} \mu(K')}.$$

By cross–multiplying, let us rewrite this as:

$$\sum_{K,K'} f(K)\mu(K)\mu(K') \quad \leq \quad \sum_{K,K'} f(K')\mu(K)\mu(K')$$

Here K ranges over all subsets of size a and K' over all subsets of size $a+1$. Think of $(p_k, q_k, k \in [m])$ as independent indeterminates, and hence regard this an inequality over the polynomial ring $N[p_k, q_k, k \in [m]]$. Then, of course, it is natural to compare the two sides term–wise. Pick a fixed *monomial* t, and let

$$S_t := \{(K, K') \mid \mu(K)\mu(K') = t\},$$

be the set of pairs producing this monomial. Then, it suffices to prove that

$$\sum_{(K,K')\in S_t} f(K) \leq \sum_{(K,K')\in S_t} f(K'). \tag{5}$$

Let us take a closer look at the structure of the set S_t. Let (K, K') be a pair of sets producing the monomial t. Note that for each $i \in [m]$, the factor $p_i^\alpha q_i^\beta$ occurs in t with exponent

- $\alpha = 2, \beta = 0$, exactly if i is in both K and K';
- $\alpha = 1 = \beta$, exactly if i is in one of K or K'.
- $\alpha = 0, \beta = 2$ exactly if i is in neither K nor K'.

Thus, the monomial t records exactly the multi–set $U_t := K + K'$. What other pairs of sets could produce the monomial t? Exactly those that produce the same multiset U_t as their multiset–union. Note that U_t is of size $2a+1$ counting multiplicity. Let I_t denote the intersection $K \cap K'$. Then S_t consists exactly of the pairs (K, K') with $K \cap K' = I_t$ and the remaining elements in $U_t - (I_t + I_t)$ partitioned in all possible ways into K and K' with exactly one more element in K'. Let U'_t denote the multi–set difference $U_t - (I_t + I_t)$. Note that U'_t is a *set* of odd size. Note also that each K can be paired with exactly one K' and vice–versa to produce the monomial t.

Thus (5) reduces to showing:

$$\sum_{K\subseteq U'_t, |K|=a-|I_t|} f(K \cup I_t) \quad \leq \quad \sum_{K'\subseteq U'_t, |K'|=a-|I_t|+1} f(K' \cup I_t) \tag{6}$$

This follows from the following lemma with $S := U'_t$ and $g(K) := f(K \cup I_t)$. ∎

Lemma 44 *Let S be a set of size $2a+1$ for a non–negative integer a let g be any real–valued function on sets such that $K \subseteq K'$ implies $g(K) \leq g(K')$. Then,*

$$\sum_{K\subseteq S, |K|=a} g(K) \quad \leq \quad \sum_{K'\subseteq S, |K'|=a+1} g(K').$$

Proof. Consider the bipartite graph $G := (A, B, E)$ where $A := \{K \subseteq S \mid |K| = a\}$ and $B := \{K' \subseteq S \mid |K'| = a + 1\}$ with an edge from K to K' exactly if $K \subseteq K'$. This is a regular graph of degree $a + 1$, hence by Hall's Marriage Theorem, there is a matching saturating A. For any K and the matching K', we have $g(K) \leq g(K')$. Hence the result. ∎

4 Applications

4.1 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.

The joint distribution of the occupancy numbers $B_1, \cdots, B_n$ is well known under all three distributions:

Proposition 45 *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,*

$$\Pr[B_1 = m_1, \cdots, B_n = m_n] = \frac{m!}{m_1! \cdots m_n!} n^{-m}.$$

This is just the multinomial distribution with equal cell probabilities.

2. *For the Fermi–Dirac statistics, for $m_i = 0, 1, i \in [n]$,*

$$\Pr[B_1 = m_1, \cdots, B_n = m_n] = \binom{n}{m}^{-1}.$$

3. For the Bose–Einstein statistics,

$$\Pr[B_1 = m_1, \cdots, B_n = m_n] = \binom{n+m-1}{m}^{-1}.$$

In principle, one can deduce from this joint distribution, all other quantities and relationships of interest. But establishing even such innocuous–looking correlation inequalities like

1. $\Pr[B_1 \geq 3 \mid B_2 \geq 5] \leq \Pr[B_1 \geq 3]$,
2. $\Pr[B_1 \geq 3 \mid B_2 \geq 5, B_3 \geq 4] \leq \Pr[B_1 \geq 3 \mid B_2 \geq 5] \leq \Pr[B_1 \geq 3]$,
3. $\Pr[B_1 \geq 3 \mid B_2 \geq 5, B_3 \geq 4] \leq \Pr[B_1 \geq 3 \mid B_2 \geq 4, B_3 \geq 3]$,
4. $\Pr[B_1 + B_2 \geq 5 \mid B_3 + B_6 + B_{17} \geq 6] \leq \Pr[B_1 + B_2 \geq 5]$,

directly by calculation appears to be a rather formidable matter. Of course, for the Maxwell–Boltzmann and Bose–Einstein statistics, these are easy deductions from our results showing that the ocupancy numbers satisfy $(-A)$ as well as $(-R)$. Some interesting correlation inequalities on the sums of variables, as in the last inequality above, can also be deduced directly via the full FKG inequality, see [10].

It is shown in [10] that the occupancy numbers in the Fremi–Dirac statistics also satisfy both the dependence conditions $(-A)$ and $(-R)$; in this case, curiously, it is much easier to show that $(-R)$ holds than $(-A)$.

4.2 Occupancy and Distributed Edge Coloring

In their analysis of an edge coloring algorithm, Panconesi and Srinivasan [33, 32] have to analyze the balls and bins experiment. Specifically, they define indicator variables $E_i := 1$ iff the ith bin is empty, and seek to stochastically bound the sum $E_1 + \cdots + E_n$. The variables E_i are not independent, preventing a direct application of the CH–bounds. The authors overcome this problem by defining a certain notion of *self–weakening* or 1*–correlated* variables and showing that the CH–bound extends to sums of such variables. This extension 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.

The same indicator variables for empty bins also underlie results related to the Satisfiability Threshold in [23].

The analysis in both these papers can be streamlined and simplified. The key idea is that the variables E_i satisfy in fact the much stronger properties of negative dependence, negative association and negative regression:

Theorem 46 *The empty–bins indicator variables $E_1, \ldots, E_n$ satisfy both $(-A)$ as well as $(-R)$.*

Proof. We note that $E_i = [B_i \leq 0]$, for $i \in [n]$ are non–increasing functions of disjoint variables. Applying Proposiiton 7(2), we conclude that $(E_1, \cdots, E_n)$ also satisfy $(-A)$.

For $(-R)$. we note that $E_i = 0$ iff $B_i > 0$ and $E_i = 1$ iff $B_i = 0$ for each $i \in [n]$. Then the $(-R)$ property for the occupancy numbers transfers to $E_1, \ldots, E_n$ via Corollary 25. ∎

One can now apply the CH–bound to get tail estimates for $\Pr[E_1 + \cdots + E_n > s]$. Note that in this proof, we avoid any expansion and manipulations of Taylor series, as in [32, 33]. The Occupancy bounds Theorems 2 and 3 in [23] follow directly as well.

4.3 Load Balancing

Consider a scenario in which one has to allocate various jobs to available servers, for example, 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 analyzed. In a recent work, Lauer describes a new dynamic load balancing strategy [27]. The analysis of this algorithm requires establishing correlation inequalities of the type mentioned in section § 4.1.

In [1], several parallel greedy strategies are presented for load–balancing. In the analysis of these algorithms, the authors use a Poisson approximation to stochastically majorise the variables of interest by a set of independent Poisson variables. The Poisson approximation is simple but incurs a loss by a factor $\sqrt{2\pi em}$ from the probability for independent variables. By the observation that the events in question are negatively associated, we can directly employ the Chernoff bounds getting the *same* bounds as if the variables were in fact independent.

4.4 Correlation Inequalities of Farr and McDiarmid

Given a graph with vertex set V and a positive integer k, consider a random k–coloring of V where each vertex independently chooses a color from the set $[k]$. For each $i \in [k]$, let S_i be the random set of vertices colored i. For the special case when the colors are chosen uniformly from $[k]$, Farr [15] gives the following correlation inequality:

$$\Pr[\bigwedge_{i \in [k]} \mathrm{Stable}(S_i)] \leq \prod_{i \in [k]} \Pr[\mathrm{Stable}(S_i)],$$

where Stable(S) denotes that S is a stable or independent set in the graph. That is, the left hand side in the above inequality is the probability that the random coloring is proper for the graph. The above inequality immediately gives a bound on the chromatic polynomial of the graph in terms of the stability polynomial, see [15].

The inequality is attractive and intuitively plausible though by no means obvious or easy to prove by direct computation. In this connection, Farr states in the introduction [15, p. 15]:

> I found this surprisingly hard to prove, and indeed the proof given uses the considerable power of the Ahslwede–Daykin Theorem.

Farr's proof for the case of uniformly chosen colours, and using the Ahlswede–Daykin Four Functions Theorem [3] occupies pages 17–19 of his paper and can be contrasted with the totally elementary and "calculation–free" 5 line proof of a stronger inequalilty that we give next.

Farr's correlation inequality is valid even in the case where the colours are not necessarily chosen uniformly from $[k]$, and can be deduced almost effortlessly from Proposition 11. Let bin i correspond to colour i for $i \in [k]$ and let ball v correspond to vertex v for $v \in V$. Since each vertex chooses its colour independently of the others, the balls $v, v \in V$ are thrown independently into the bins $i, i \in [k]$. Further, the indicator $B_{i,v} = 1$ iff vertex v is coloured i. Now Stable(S_i) is a non–increasing function of the variables $B_{i,v}, v \in V$ for each $i \in [k]$, and the inequality follows from Proposition 11 and Proposition 7.

McDiarmid [29] gives a general lemma (which was originally proved via his General Clutter Percolation Theorem) and a proof via Harris' inequality [20]. The correlation inequality of Farr is a direct consequence. In McDiarmid's general lemma, there is a finite set I, a finite set V and a collection of independent random variables $X_v, v \in V$ taking values in a set containing I. This is easily recognized as a balls and bins experiment where balls $b_v, v \in V$ are tossed independently (with possibly unequal probabilities) into a number of bins and we focus on bins $B_i, i \in I$. Notice that the variables X_v are related to our indicator variables $B_{i,k}$ as follows: $X_v = i$ iff $B_{i,v} = 1$. All the applications of McDiarmid's general lemma can now be viewed in a much more transparent manner in this framework.

4.5 Simulation of Parallel Computers

Dietzfelbinger and Meyer auf der Heide [8] present several algorithms for the simulation of parallel random access machines (PRAMS) on more realistic models called the *distributed memory machines*. In the analysis of the algorithms for this simulation, one encounters random variables which are not independent but related in exactly the manner of the balls and bins experiment. For the purposes of upper bounds on certain probabilities our results show that these variables may be treated as if they were independent. Thus their combinatorial lemma A.2 in the appendix can be obtained directly from our results. We thank Martin Dietzfelbinger for bringing their work to our attention.

5 Unresolved Issues

- We conjecture that negative regression implies negative association. This would be an interesting result in itself. In addition, it could be a very useful device in establishing negative association. In [11], we give a simple counter–example to show that the two notions of negative dependence are not the same.

- Shepp[35, 36] conjectures that there must be a way to apply the FKG inequality systematically in many different situations. Can one apply it directly to deduce also the results on association and regression? We also believe there is a strong connection between the notion of negative association and the notion of "disjoint occurences of events" in Percolation Theory [20], in particular the so–called BK–Inequality and its relatives. We have hinted at some of these relationships in § 2.3.

- Another, rather ambitious, task would be to resolve the following kind of mixed conditions. We know that $\Pr[B_1 \geq t_1 \mid B_2 \geq t_2] \leq \Pr[B_1 \geq t_1]$ and also that $\Pr[B_1 \geq t_1 \mid B_3 \leq t_3] \geq \Pr[B_1 \geq t_1]$. What can one say about $\Pr[B_1 \geq t_1 \mid B_2 \geq t_2, B_3 \leq t_3]$? 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[35, 36] and Winkler [41] might be relevant.

6 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. We are especially indebted to Martin Dietzfelbinger for sharing with us his own investigations [7] on the subject, and for drawing our attention to [8, 29, 28]. We are grateful to Colin McDiarmid for reading a previous version of this paper and for many technical remarks and suggestions, including the short proof of the Zero–One Lemma for $(-A)$.. 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] M. Adler, S. Chakrabarti, M. Mitzenmacher and L. Rasmussen, "Parallel Randomise Load Balancing (Preliminary Version)", STOC 1995.

[2] R. Ahlswede and D.E. Daykin, "An inequality for the weights of two families of sets, their unions and intersections", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 43, 183–185, 1978.

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

[4] A.D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, Oxford Studies in Probability 2, Clarendon Press, Oxford 1992.

[5] A. Blass, "Is Game Semantics Necessary?", *Computer Science Logic*, 7th Workshop, LNCS 832, pp. 66–77, Springer Verlag 1994.

[6] J. van den Berg and H. Kesten, "Inequalities with Applications to Percolation and Reliability", *J. Appl. Prob.*, 22:556–569, 1985.

[7] M. Dietzfelbinger, unpublished notes.

[8] M. Dietzfelbinger and F. Meyer auf der Heide, "Simple Efficient Shared Memory Simulations (Extended Abstract)", SPAA 1993.

[9] D. Dubhashi, "On the BK inequality", manuscript, 1995.

[10] D. Dubhashi, V. Priebe and D. Ranjan, "Two Applications of the FKG Inequality", manuscript 1995.

[11] D. Dubhashi and D. Ranjan, "Counter–Examples for Negative Regression", in preparation.

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

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

[14] S. Chaudhuri, D. Dubhashi, "Probabilistic Recurrence Relations (Revisited)", in R. Baeza–Yates, E. Goles and P. Poblete (eds): LATIN'95: Theoretical Informatics, Second Latin American Symposium, Valparamso, Chile, LNCS 911, 207–219, 1995.

[15] G. Farr, "A Correlation Inequality involving Stable Set and Chromatic Polynomials", *J. Comb. Theory, Series B* 58, 14–21 1993.

[16] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol I, John Wiley & Sons Inc. 1968.

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

[18] J.–Y. Girard, "Linear Logic", *Theor. Comp. Sci.*, 59 pp. 1–102, 1987.

[19] 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.

[20] G. Grimmett, *Percolation*, Springer–Verlagm 1989.

[21] R. Holley, "Remarks on the FKG Inequality", *Comm. in Math. Physics*, 36, pp. 227–231, 1974.

[22] N.L. Johnson and S. Kotz, *Urn Models and their Applications*, John Wiley& Sons, 1977.

[23] A. Kamath, R. Motwani, K. Palem and P. Spirakis, "Tail Bounds for Occupancy and the Satisfiability Threshold Conjecture", FOCS 1994.

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

[25] J.H.B. Kemperman, "On the FKG Inequality for Measures on a Partially Ordered Space", *Indag. Math.*, 39, pp. 313–331, 1977.

[26] V.F. Kolchin, B.A. Sevastyanov and V.P. Chistyakov, *Random Allocations*, John Wiley& Sons, 1978.

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

[28] C. L. Mallows, "An inequality involving multinomial probabilities", *Biometrika* 55, 422–424, 1968.

[29] C. McDiarmid, "On a Correlation Inequality of Farr", *Combinatorics, Probability and Computing*, 1, 157–160, 1992.

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

[31] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.

[32] 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.

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

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

[35] 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.

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

[37] A. Srinivasan, *Techniques for probabilistic Analysis and Randomness–Efficient Computation*, Ph.D. Thesis, Cornell University, 1993.

[38] J. Schmidt, A. Siegel and A. Srinivasan, "Chernoff–Hoeffding Bounds for Applicationswith Limited Independence", *SIAM J. Discrete Mathematics* 1995.

[39] R. Szekli, *Stochastic Ordering and Dependence in Applied Probability*, Lecture Notes in Statistics 97, Springer–Verlag 1995

[40] Y.L. Tong, *Probability Inequalities in Multivariate Distributions*, Academic Press, 1980.

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

Recent Publications in the BRICS Report Series

RS-96-25 **Devdatt Dubhashi and Desh Ranjan.** ***Balls and Bins: A Study in Negative Dependence*****. July 1996. 27 pp.**

RS-96-24 **Henrik Ejersbo Jensen, Kim G. Larsen, and Arne Skou.** ***Modelling and Analysis of a Collision Avoidance Protocol using SPIN and UPPAAL*****. July 1996. 20 pp.**

RS-96-23 **Luca Aceto, Wan J. Fokkink, and Anna Ingólfsdóttir.** ***A Menagerie of Non-Finitely Based Process Semantics over BPA*: From Ready Simulation Semantics to Completed Tracs*****. July 1996. 38 pp.**

RS-96-22 **Luca Aceto and Wan J. Fokkink.** ***An Equational Axiomatization for Multi-Exit Iteration*****. June 1996. 30 pp.**

RS-96-21 **Dany Breslauer, Tao Jiang, and Zhigen Jiang.** ***Rotation of Periodic Strings and Short Superstrings*****. June 1996. 14 pp.**

RS-96-20 **Olivier Danvy and Julia L. Lawall.** ***Back to Direct Style II: First-Class Continuations*****. June 1996. 36 pp. A preliminary version of this paper appeared in the proceedings of the 1992 ACM Conference on Lisp and Functional Programming, William Clinger, editor, LISP Pointers, Vol. V, No. 1, pages 299–310, San Francisco, California, June 1992. ACM Press.**

RS-96-19 **John Hatcliff and Olivier Danvy.** ***Thunks and the λ-Calculus*****. June 1996. 22 pp. To appear in** ***Journal of Functional Programming*****.**

RS-96-18 **Thomas Troels Hildebrandt and Vladimiro Sassone.** ***Comparing Transition Systems with Independence and Asynchronous Transition Systems*****. June 1996. 14 pp. To appear in Montanari and Sassone, editors,** ***Concurrency Theory: 7th International Conference*****, CONCUR '96 Proceedings, LNCS 1119, 1996.**

RS-96-17 **Olivier Danvy, Karoline Malmkjær, and Jens Palsberg.** ***Eta-Expansion Does The Trick (Revised Version)*****. May 1996. 29 pp. To appear in** ***ACM Transactions on Programming Languages and Systems (TOPLAS)*****.**