MAX-PLANCK-INSTITUT FÜR INFORMATIK

Stochastic Majorisation: Exploding some Myths

Devdatt Dubhashi, Desh Ranjan

MPI–I–94–144 August 1994

INFORMATIK

Im Stadtwald
66123 Saarbrücken
Germany

Stochastic Majorisation: Exploding some Myths

Devdatt Dubhashi, Desh Ranjan

MPI–I–94–144 August 1994

Stochastic Majorisation: Exploding some Myths

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 5, 1994

1 Introduction

The analysis of many randomised algorithms involves random variables that are not independent, and hence many of the standard tools from classical probability theory that would be useful in the analysis, such as the Chernoff–Hoeffding bounds are rendered inapplicable. However, in many instances, the random variables involved are, nevertheless *negatively related* in the intuitive sense that when one of the variables is "large", another is likely to be "small". (this notion is made precise and analysed in [1].) In such situations, one is tempted to conjecture that these variables are in some sense *stochastically dominated* by a set of *independent* random variables with the same marginals. Thereby, one hopes to salvage tools such as the Chernoff–Hoeffding bound also for analysis involving the dependent set of variables. The analysis in [6, 7, 8] seems to strongly hint in this direction. In this note, we explode myths of this kind, and argue that stochastic majorisation in conjunction with an independent set of variables is actually much less useful a notion than it might have appeared.

2 Stochastic Majorisation

To quote Marshall and Olkin, [5]:

> The notion of stochastic ordering for random variables is a familiar and useful concept. In spite of this, references to even the basic results are not easy to find.

In this section, we collect together some relevant facts from the theory of stochastic majorisation.

Definition 1 *The random variable X is said to be* **stochastically majorised** *by the random variable Y, $X \leq_{st} Y$ if for all $t \in \mathbb{R}$,*

$$\Pr[X \geq t] \leq \Pr[Y \geq t].$$

*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

To see how this ordering can be extended to random vectors, the following equivalences are useful:

Proposition 2 *The following conditions are equivalent:*

1. $X \leq_{\rm st} Y$.
2. $E[\phi(X)] \leq E[\phi(Y)]$ *for all non-decreasing functions for which the expectations exist.*
3. $\phi(X) \leq_{\rm st} \phi(Y)$ *for all non-decreasing functions* ϕ.
4. $\Pr[X \in A] \leq \Pr[Y \in A]$ *for all sets* A *with non-decreasing indicator functions, (in other words, for an interval* A *which is semi-infinite to the right).*

For a fixed $n > 0$, regard R^n as a poset with the component-wise ordering. The extensions to random variables in R^n i.e. to real vectors is the following

Proposition 3 *Let* $\mathbf{X} := (X_1, \cdots, X_n)$ *and* $\mathbf{Y} := (Y_1, \cdots, Y_n)$ *two random vectors. Consider the following conditions:*

1. *For all reals* $t_1, \cdots, t_n$,
$$\Pr[X_1 \geq t_1, \cdots, X_n \geq t_n] \leq \Pr[Y_1 \geq t_1, \cdots, Y_n \geq t_n]$$
i.e. $\Pr[\mathbf{X} \in A] \leq \Pr[\mathbf{Y} \in A]$ *for every principal ideal* $A \in \mathsf{R}^n$.
2. $E[\phi(X_1, \cdots, X_n)] \leq E[\phi(Y_1, \cdots, Y_n)]$, *for all order-preserving functions* $\phi : \mathsf{R}^n \longrightarrow \mathsf{R}$ *for which the expectations exist.*
3. $\phi(X_1, \cdots, X_n) \leq_{\rm st} \phi(Y_1, \cdots, Y_n)$, *for all order-preserving functions* $\phi : \mathsf{R}^n \longrightarrow \mathsf{R}$.
4. $\Pr[\mathbf{X} \in A] \leq \Pr[\mathbf{Y} \in A]$ *for all measurable sets* $A \subseteq \mathsf{R}^n$ *with non-decreasing indicator functions, i.e. ideals* $A \subseteq \mathsf{R}^n$.

Then conditions (2),(3) and (4) are all equivalent and imply condition (1).

The equivalence of the conditions (2)–(4) of Proposition 3, motivates the following definition.

Definition 4 *A random vector* $\mathbf{X}$ *is said to be* **stochastically majorised** *by a random vector* $\mathbf{Y}$, $\mathbf{X} \leq_{\rm st} \mathbf{Y}$, *if* $E[\phi(\mathbf{X})] \leq E[\phi(\mathbf{Y})]$ *for all order-preserving functions* $\phi : \mathsf{R}^n \longrightarrow \mathsf{R}$, *for which the expectations exist.*

3 Exploding some Myths: Counterexamples

As stated before, it is intuitively appealing and tempting to conjecture that whenever random variables are negatively associated, *i.e.* when increasing one of them tends to decrease one of the others, that a set of independent variables with the same marginal distribution will stochastically dominate these variables. In what follows, we give counterexamples to this and some other intuitively appealing statements and show that stochastic majorisation by an independent set of variables is a really elusive condition to obtain.

3.1 Some Counter–examples

EXAMPLE 1: Consider the joint distribution on two binary valued variables, X_1, X_2 with

$$\Pr[X_1 = 0, X_2 = 1] = \frac{1}{2} = \Pr[X_1 = 1, X_2 = 0].$$

The marginals are for $i = 1, 2$,

$$\Pr[X_i = 0] = \frac{1}{2} = \Pr[X_i = 1].$$

Now, for independent variables Y_1, Y_2 with these marginals,

$$\Pr[Y_1 + Y_2 \geq 1] = \frac{3}{4}$$

whereas

$$\Pr[X_1 + X_2 \geq 1] = 1!$$

So $\mathbf{X} \not\leq_{\rm st} \mathbf{Y}$. This happens even though the variables X_1, X_2 are very strongly negatively associated, in the sense of [1].

EXAMPLE 2: Consider the experiment where $n > 0$ balls are thrown uniformly and independently at random into $m > 0$ bins. For $i \in [n]$, let $Y_i = 1$ if the ith bin is empty and 0 otherwise. This is the set of variables considered in [6, 7, 8]. This set of variables is also strongly negatively associated, [1]. However, as the following calculation with 3 balls and 3 bins shows, the Y_i variables are not majorised by independent variables with the same marginals.

- We have $\Pr[\mathbf{Y} = (0,0,0)] = 3!/3^3 = 2/9$.
- For $\mathbf{a} \in \{(0,1,1),(1,0,1),(1,1,0)\}$, we have $\Pr[\mathbf{Y} = \mathbf{a}] = 1/27$.
- For $\mathbf{a} \in \{(1,0,0),(0,1,0),(0,0,1)\}$, we have $\Pr[\mathbf{Y} = \mathbf{a}] = (1 - (2/9 + 3/27))/3 = 2/9$.

The marginals are, for $i = 1, 2, 3$, $\Pr[Y_i = 1] = 2/27 + 2/9 = 8/27$ and $\Pr[Y_i = 0] = 19/27$. Now, $\Pr[Y_1 + Y_2 + Y_3 \geq 1] = 1 - \Pr[\mathbf{Y} = (0,0,0)] = 1 - 2/9 = 7/9$, whereas, the similar probability for an independent set with the same marginals is $1 - \Pr[Y_1 = 0]\Pr[Y_2 = 0]\Pr[Y_3 = 0] = 1 - (19/27)^3 \approx .65 < 7/9$.

EXAMPLE 3: Consider the same experiment as in Example 2. This time, for $i \in [n]$, let X_i denote the number of balls in the ith bin. Once again, these variables are strongly negatively associated, [1]. However, stochastic majorisation fails, as shown by the following simple argument. Consider the monotone function $f(X_1, \cdots, X_n) = X_1 + \cdots + X_n$. Then, $\Pr[f(X_1, \cdots, X_n) \geq n'] = 1$ for any $n' \leq n$, whereas for the independent random variables Y_i's with the same marginal distributions $\Pr[f(X_1, \cdots, X_n) \geq n' < 1$. Hence, Y_i's do not stochastically majorise X_i's.

In connection with Examples 2 and 3, it is interesting to note that the Chernoff-Hoeffding bound nevertheless holds for sums of the variables involved, so that, for purposes on stochastic bounds on the sum, we may treat them as if they were independent. However, the real reason for the bounds being valid is not stochastic majorisation, but negative association as shown in [1].

3.2 Useful Criteria?

A random vector $\mathbf{X} := (X_1, \cdots, X_n)$ is said to satisfy the *negative monotone regression property* if: For each $i \in [n]$, and for each non-decreasing function $f : \mathsf{R} \rightarrow \mathsf{R}$,

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

is non-increasing in $(t_1, \cdots, t_i) \in \mathsf{R}^i$. Naively, one might conjecture that if $\mathbf{X}$ satisfies negative monotone regression, then it is stochastically dominated by an independent set of variables with the same marginals. However, Example 1 from the previous subsection is a simple counter–example.

The following proposition appears in Marshall and Olkin, [5] where it is attributed to several authors independently.

Proposition 5 *Let* $\mathbf{X} := (X_1, \cdots, X_n)$ *and* $\mathbf{Y} := (Y_1, \cdots, Y_n)$ *be random vectors such that for* $1 < j \leq n$, *all* $t \in \mathsf{R}$ *and all* $\mathbf{u} \leq \mathbf{v}, \mathbf{u}, \mathbf{v} \in \mathsf{R}^{j-1}$,

$$\Pr[X_1 \geq t] \leq \Pr[Y_1 \geq t]$$

and

$$\Pr[X_j \geq t \mid X_1 = u_1, \cdots, X_{j-1} = u_{j-1}] \leq \Pr[Y_j \geq t \mid Y_1 = v_1, \cdots, Y_{j-1} = v_{j-1}].$$

Then $\mathbf{X} \leq_{\text{st}} \mathbf{Y}$.

How useful is this proposition when taken in conjunction with an independent set of variables? Taking $\mathbf{Y}$ to be an independent set of variables such that the marginal distribution of Y_i is identical to that of X_i for each $i \in [n]$, we get that $\mathbf{X} \leq_{\text{st}} \mathbf{Y}$ if for $1 < j \leq n$, all $t \in \mathsf{R}$ and all $\mathbf{u} \in \mathsf{R}^{j-1}$,

$$\Pr[X_j \geq t \mid X_1 = u_1, \cdots, X_{j-1} = u_{j-1}] \leq \Pr[X_j \geq t].$$

This destroys any hope of actually being able to use this result in our context!

Another criterion that is mentioned in [3, 4, 2], is the following relative of the celebrated FKG Inequality. (We only give a version valid for discrete variables.)

Proposition 6 *For random vectors* $\mathbf{X}, \mathbf{Y} \in \mathsf{R}^n$, *we have* $\mathbf{X} \leq_{\text{st}} \mathbf{Y}$ *if for all* $\mathbf{a}, \mathbf{b} \in \mathsf{R}^n$,

$$\Pr[\mathbf{X} = \mathbf{a}] \cdot \Pr[\mathbf{Y} = \mathbf{b}] \leq \Pr[\mathbf{X} = \mathbf{a} \wedge \mathbf{b}] \cdot \Pr[\mathbf{Y} = \mathbf{a} \vee \mathbf{b}].$$

Taking $\mathbf{Y}$ as above to be a vector of independent variables with the same marginals as $\mathbf{X}$, we get the following curious criterion on the distribution of $\mathbf{X}$ that is sufficient to ensure that $\mathbf{X}$ is dominated stochastically by a set of independent variables with the same marginals.

Proposition 7 *Let* $\mathbf{X} := (X_1, \cdots, X_n)$ *be a random vector such that the joint distribution satisfies the condition: for any* $\mathbf{a}, \mathbf{b} \in \mathsf{R}^n$,

$$\Pr[\mathbf{X} \geq \mathbf{a} \mid \mathbf{X} \geq \mathbf{a} \wedge \mathbf{b}] \leq \prod_{a_i > b_i} \frac{\Pr[X_i = a_i]}{\Pr[X_i = b_i]}$$

that is,

$$\Pr[\bigwedge_{i \in [n]} X_i \geq a_i \mid \bigwedge_{i \in [n]} X_i \geq \min(a_i, b_i)] \leq \prod_{a_i > b_i} \frac{\Pr[X_i = a_i]}{\Pr[X_i = b_i]}.$$

Then $\mathbf{X} \leq_{\text{st}} \mathbf{Y}$ *where* $\mathbf{Y} := (Y_1, \cdots Y_n)$ *is a set of independent random variables with* $\Pr[Y_i = a_i] = \Pr[X_i = a_i]$ *for all reals* a_i *and* $i \in [n]$.

We do not know of a single natural distribution where this condition obtains for it to be useful.

4 Open Questions

The really interesting problem is finding useful, sufficient conditions for applying Chernoff-Hoeffding bounds[6, 9]. As shown in [1], negative association is one of them. An interesting open question is if negative monotone regression implies that the CH-Bounds can be applied. Another interesting question is if negative monotone regression condition implies negative association. Of course, a positive answer to the second question will imply a positive answer to the first.

References

[1] D. Dubhashi, D. Ranjan, "Some Correlation Inequalities for Probabilistic Analysis of Algorithms", manuscripts, 1994.

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

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

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

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

[6] 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.

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

[8] A. Srinivasan, *Techniques for Probabilistic Analysis and Randomness-Efficient Computation*, Ph.D. Thesis, Cornell University, 1993.

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