BRICS RS-98-25 D. P. Dubhashi: Talagrand's Inequality in Hereditary Settings

BRICS

Basic Research in Computer Science

Talagrand's Inequality in Hereditary Settings

Devdatt P. Dubhashi

BRICS Report Series **RS-98-25**

ISSN 0909-0878 **October 1998**

See back inner page for a list of recent BRICS Report Series publications.
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 the World Wide Web and anonymous FTP through these URLs:

`http://www.brics.dk`
`ftp://ftp.brics.dk`
This document in subdirectory `RS/98/25/`

Talagrand's Inequality in Hereditary Settings *

Devdatt P. Dubhashi †
Department of Computer Science and Engg.
Indian Institute of Technology, Delhi
Hauz Khas, New Delhi 110016, INDIA.
email: dubhashi@cse.iitd.ernet.in

October 6, 1998

Abstract

We develop a nicely packaged form of Talagrand's inequality that can be applied to prove concentration of measure for functions defined by hereditary properties. We illustrate the framework with several applications from combinatorics and algorithms. We also give an extension of the inequality valid in spaces satisfying a certain negative dependence property and give some applications.

1 Talagrand's Inequality

Talagrand's inequality is an isoperimetric inequality that applies in the setting where $\Omega = \prod_{i \in I} \Omega_i$ is a product space indexed by some finite index set I with the product measure.

*Submitted to *Random Structures and Algorithms* and under revision.
†Work done while at the SPIC Mathematical Institute, Chennai, India and while visiting **BRICS**, Basic Research in Computer Science, Centre of the Danish National Research Foundation, Department of Computer Science, University of Aarhus, Denmark.

The crucial matter however is the distance. Talagrand's *convex distance* is defined by[1]

$$d_T(x, A) := \max_{0 \neq \alpha \geq 0} \min_{y \in A} \frac{\sum_{x_i \neq y_i} \alpha_i}{\left(\sum_i \alpha_i^2\right)^{1/2}}. \tag{1}$$

The max is taken over all non–negative reals α_i which are not all zero. By normalising the α_is we can write this also as:

$$d_T(x, A) := \max_{\sum_i \alpha_i^2 = 1} \min_{y \in A} \sum_{x_i \neq y_i} \alpha_i.$$

Define the t–neighbourhood of a set in the Talagrand distance in the usual way:

$$A_t := \{x \in \Omega \mid d_T(x, A) \leq t\}.$$

Theorem 1 (Talagrand's Inequality) *Let A be a subset in a product space. Then for any $t > 0$,*

$$\mathtt{Pr}[A]\mathtt{Pr}[\overline{A_t^T}] \leq e^{-t^2/4}. \tag{2}$$

In this note, we shall supplement Talagrand's inequality in hereditary settings:

- We develop a general framework for applying Talagrand's inequality to functions defined on product spaces by hereditary properties. The precise definition is in § 2 below. This gives a nicely packaged and easy to use framework for proving concentration of measure. We should point out that the germs of this are already implicit in the original paper of Talagrand [6], see also the account in Steele's monograph [5]. Spencer [4] makes it explicit, but we believe our formulation is of independent interest.

- We extend Talagrand's inequality to certain settings where the underlying measure is not the product measure (i.e. independence is not available), but satisfies a certain negative dependence condition.

[1] We should have sup and inf in place of max and min but in the cases of interest to us, where A is finite, the replacement is justified by compactness.

This extension is incomparable with Marton's recent extension of Talagrand's inequality to dependent variables [3]. The extension is stated precisely and proved in § 3

We illustrate the frameworks developed with several examples; still these only scratch the surface of the wide applicability of these tools.

2 Hereditary Properties

We will develop a general framework to analyse a certain class of functions on product spaces which are defined by hereditary (i.e. monotone) properties of index sets. This framework generalises slightly the results implicit in Talagrand [6] and explicit in Steele [5] and Spencer [4]. We then illustrate the versatality of this framework by several examples.

2.1 A General Framework

The starting point of our framework comes from the original example that motivated Talagrand to develop his theory: increasing subsequences. Given a sequence of reals $x = x_1, \ldots, x_n$, an increasing subsequence is an index set $J \subseteq [n]$ such that for each $j < k \in J$, we have $x_j \leq x_k$. Consider the boolean property $\phi(x, J)$ which is true iff J is an increasing subsequence in x. One observes that

- If $x_j = y_j$ for each $j \in J$, then $\phi(x, J) = \phi(y, J)$.
- If $J \subseteq J'$, then $\phi(x, J) \geq \phi(x, J')$ (where we are identifying `true` with 1 and `false` with 0).

These are the two essential properties that are needed in order to apply Talagrand's inequality. We will actually generalise even further to consider properties defined by *families* of index sets. However, it will be helpful to keep in mind the example and the formulation of the two properties above.

We shall need some notations. For $x, y \in \Omega$ and $J \subseteq I$, we use the notation $x_J = y_J$ to mean $x_j = y_j, j \in J$, and for a family $\mathcal{J}$ of subsets of I, by $x_{\mathcal{J}} = y_{\mathcal{J}}$ we mean $x_J = y_J, J \in \mathcal{J}$. Let $\mathcal{J}_{x=y} := \{J \in \mathcal{J} \mid x_J = y_J\}$ and note that $x_{\mathcal{J}_{x=y}} = y_{\mathcal{J}_{x=y}}$.

Let $\phi(x, \mathcal{J})$ be a boolean property such that it is

- a *property of index sets* i.e. if $x_{\mathcal{J}} = y_{\mathcal{J}}$, then $\phi(x, \mathcal{J}) = \phi(y, \mathcal{J})$, and
- *hereditary non–increasing on the index sets*, i.e. if $\mathcal{J} \subseteq \mathcal{J}'$ then $\phi(x, \mathcal{J}) \geq \phi(x, \mathcal{J}')$ (where again, we . identify `true` with 1 and `false` with 0).

We shall say that ϕ is a *hereditary property of index sets.*

Let $w(i, J), i \in I, J \subseteq I$ be a matrix of non–negative weights. Set $w(J) := \sum_i w(i, J)$. The function f_ϕ defined by a hereditary property ϕ of index sets is given by:

$$f_\phi(x) := \max_{\mathcal{J}:\phi(x,\mathcal{J})} \sum_{J \in \mathcal{J}} w(J). \tag{3}$$

By $\mathcal{J}(x)$, we shall denote a family achieving the maximum in (3). Thus,

$$f_\phi(x) := \sum_{J \in \mathcal{J}(x)} w(J). \tag{4}$$

A function f such that $f = f_\phi$ for some hereditary property ϕ of index sets will be called a *hereditary function of index sets.*

Theorem 2 *Let f be a hereditary function defined by a hereditary property ϕ and weights $w(i, J), i \in I, J \subseteq I$. Then for all $t > 0$,*

$$\mathtt{Pr}[f > \mathtt{M}[f] + t] \leq 2 \exp\left(\frac{-t^2}{4wk(M[f] + t)}\right),$$

and

$$\mathtt{Pr}[f < \mathtt{M}[f] - t] \leq 2 \exp\left(\frac{-t^2}{4wkM[f]}\right),$$

where $\mathtt{M}[f]$ *is a median of* f *and*

$$k := \max_{x \in \Omega} \max_{J \in \mathcal{J}(x)} |J|,$$

and

$$w := \max_{x \in \Omega} \max_{i \in I} \sum_{\substack{J \in \mathcal{J}(x) \\ i \in J}} w(J).$$

Proof. First note that for any $x, y \in \Omega$,

$$\begin{aligned}
1 &= \phi(x, \mathcal{J}(x)) \\
&\leq \phi(x, \mathcal{J}(x)_{x=y}), \quad \text{by the herditary property} \\
&= \phi(y, \mathcal{J}(x)_{x=y}), \quad \text{index set property of } \phi.
\end{aligned}$$

Hence,

$$\begin{aligned}
f(y) &= \max_{\substack{\mathcal{J} \\ \phi(y,\mathcal{J})}} \sum_{J \in \mathcal{J}} w(J) \\
&\geq \sum_{J \in \mathcal{J}(x)_{x=y}} w(J) \\
&= \sum_{J \in \mathcal{J}(x)} w(J) - \sum_{J \in \mathcal{J}(x)_{x \neq y}} w(J) \\
&= f(x) - \sum_{\substack{J \in \mathcal{J}(x) \\ x_J \neq y_J}} w(J) \qquad (5)
\end{aligned}$$

Define $\alpha_i = \alpha(x)_i := \sum_{J \in \mathcal{J}(x)} 1[i \in J] w(J)$ for $i \in I$. Then,

$$\begin{aligned}
\sum_{x_i \neq y_i} \alpha_i &= \sum_{x_i \neq y_i} \sum_{J \in \mathcal{J}(x)} 1[i \in J] w(J) \\
&= \sum_{J \in \mathcal{J}(x)} w(J) \sum_{i \in J} 1[x_i \neq y_i] y_i] \\
&\geq \sum_{J \in \mathcal{J}(x)} w(J) 1[x_J \neq y_J].
\end{aligned}$$

Hence, by (5), we have,

$$f(y) \geq f(x) - \sum_{x_i \neq y_i} \alpha_i. \qquad (6)$$

Now set $A = A(a) := \{y \in \Omega \mid f(y) \leq a\}$. Then we have by the definition of Talagrand's distance,

$$
\begin{aligned}
d_T(x, A) &\geq \frac{\min_{y \in A} \sum_{x_i \neq y_i} \alpha(x)_i}{(\sum_i \alpha_i^2)} \\
&= \frac{\sum_{x_i \neq y_i^*} \alpha_i}{(\sum_i \alpha_i^2)},
\end{aligned}
$$

for some $y^* \in A$. Further note that

$$
\begin{aligned}
\sum_i \alpha_i^2 &= \sum_i \left(\sum_{J \in \mathcal{J}(x)} 1[i \in J] w(J) \right)^2 \\
&\leq \sum_i w \sum_{J \in \mathcal{J}(x)} 1[i \in J] w(J) \\
&= w \sum_{J \in \mathcal{J}(x)} w(J) \sum_{i \in J} 1 \\
&\leq wk \sum_{J \in \mathcal{J}(x)} w(J) \\
&= wkf(x).
\end{aligned}
$$

Thus,

$$
\sum_{x_i \neq y_i} \alpha(x)_i \leq \sqrt{wkf(x)}\, d_T(x, A),
$$

and plugging into (6), we get:

$$
d_T(x, A) \geq \frac{f(x) - a}{\sqrt{wkf(x)}}.
$$

Now applying Talagrand's inequality, we have,

$$
\begin{aligned}
\mathtt{Pr}[f(X) > a + t] &= \mathtt{Pr}\left[\frac{f(X) - a}{\sqrt{wk(a+t)}} > \frac{t}{\sqrt{wk(a+t)}} \right] \\
&\leq \mathtt{Pr}\left[\frac{f(X) - a}{\sqrt{wkf(x)}} > \frac{t}{\sqrt{wk(a+t)}} \right]
\end{aligned}
$$

$$
\begin{aligned}
&\le \mathtt{Pr}\left[d_T(x,A) > \frac{t}{\sqrt{k(a+t)}}\right] \\
&\le \frac{1}{\mathtt{Pr}[A]}\exp\left(\frac{-t^2}{4wk(a+t)}\right).
\end{aligned}
$$

Let us rewite this as

$$
\mathtt{Pr}[f(X) \le a]\mathtt{Pr}[f(X) > a+t] \le \exp\left(\frac{-t^2}{4wk(a+t)}\right).
$$

Setting $a := \mathtt{M}[f]$ and $a := \mathtt{M}[f] - t$ successively,we get the result. ∎

Setting $t := \epsilon\mathtt{M}[f]$ in Proposition 2, we get

$$
\mathtt{Pr}[f > (1+\epsilon)\mathtt{M}[f]] \le 2\exp\left(\frac{-\epsilon^2}{4k(1+\epsilon)}\mathtt{M}[f]\right),
$$

$$
\mathtt{Pr}[f < (1-\epsilon)\mathtt{M}[f]] \le 2\exp\left(\frac{-\epsilon^2}{4k}\mathtt{M}[f]\right).
$$

One can obtain a concentration of measure inequality around the mean rather than the median via the following genral result:

Proposition 3 *The following are equivalent for an arbitrary function f and random variables $X_1, \ldots, X_n$:*

- *For all $t > 0$, there are positive constants c and α such that*

$$
\mathtt{Pr}[|f - \mathtt{M}[f]| > t] \le ce^{-\alpha t}.
$$

- *For all $t > 0$, there are positive constants c' and α' such that*

$$
\mathtt{Pr}[|f - \mathtt{E}[f]| > t] \le c'e^{-\alpha' t}.
$$

2.2 Examples

We illustrate the general framework with several examples. The example of increasing subsequences in § 2.2.1 was one of the motivating examples for Talagrand's inequality [6], see also [5]. The examples involving counting extensions in § 2.2.3 and counting representations in § 2.2.4 were pointed out by Spencer [4].

In the examples below, we will always identify a singleton set with the element itself. To indicate how the framework of Theorem 2 applies, we shall give the weight system $w(i, J)$, the hereditary property ϕ and the parameters k and w that appear in the bound, without writing the probability bound explicitly each time. The unweighted case will easily follow by setting all weights to be unity.

2.2.1 Increasing subsequences

Let $w_1, \ldots, w_n$ be given non–negative weights and let $I(x_1, \ldots, x_n)$ denote the weight of the heaviest increasing subsequence from $x_1, \ldots, x_n$. Let $X_1, \ldots, X_n$ be chosen independently at random from $[0, 1]$. Theorem 2 can be applied immediately to give a sharp concentration result on $I(X_1, \ldots, X_n)$.

Set $w(i, J) := 1[i \in J]w_i$. Take the hereditary property $\phi(x, \mathcal{J})$ to be:

- $|J| = 1$ for all $J \in \mathcal{J}$.
- For $j, j' \in \mathcal{J}$, if $j < j'$ then $x(j) \leq x(j')$.

Here $k = 1$ and $w = \max_i w_i$. Taking $w_i = 1$ for all $i \in [n]$ gives the original result.

2.2.2 Balls and Bins

Consider the probabilistic experiment where m balls are thrown independently at random into n bins and we are interested in a sharp concentration result on the number of empty bins. Equivalently, we can give a sharp concentration result on the number of non–empty bins.

To cast this in the framework of configuration functions, consider the product space $[n]^m$ with the product measure where $\texttt{Pr}[X_k = i]$ is the probability that ball k is thown into bin i. We shall give a sharp concentration result on the number of non–empty bins. Take $w(i, J) = 1[i \in J]$ and the hereditary property $\phi(x, \mathcal{J})$ to be:

- $|J| = 1$ for all $J \in \mathcal{J}$.
- For distinct $j, j' \in \mathcal{J}$, $x(j) \neq x(j')$.

2.2.3 Counting Extensions

Consider the random graph $G(n, p)$ with non–negative weights $w_e, e \in E$ on the edges. A subset E' of the edges has weight $w(E') := \sum_{e \in E'} w_e$. For a fixed vertex u, let let $T(u)$ denote the weight of the triangles containing a given vertex u. One obtains a sharp concentration result on $T(u)$ as follows. Consider the product space $\{0,1\}^E$, take $w(e, J) := 1[e \in J]w_e$ and take the hereditary property $\phi(x, \mathcal{J})$ to be:

- $|J| = 3$ for all $J \in \mathcal{J}$.
- The edges in $\{e_1, e_2, e_3\}$ in $\mathcal{J}$ form a triangle including the vertex u.

With all weights $w_e = 1$, it is easy to see that the number of triangles containing u is $\texttt{E}[N(u)] = \binom{n-1}{2}p^3$ and we get sharp concentration around this value. Similarly one can prove generalisations due to Spencer for the number of extensions $N(x_1, \ldots, x_r)$ of a given set of vertices to a copy of a given graph H.

A number of other results about random graphs can also be proved in a similar fashion.

2.2.4 Counting Representations

For a given set S of natural numbers, let $f_S(n)$ denote the number of representations $n = x + y$ for distinct $x, y \in S$. An old result of Erdös shows that there is a set S for which $f_S(n) = \theta(\ln n)$. This is obtained by a probabilistic construction: define S randomly setting $\mathtt{Pr}[x \in S] := \min\left(1, 10\sqrt{\frac{\ln x}{x}}\right)$. One can show that $\mathtt{E}[f_S(n)] \sim 100\pi \ln n$. A sharp concentration result about this value is obtained by taking unit weights $w_i = 1, i \geq 1$ and the hereditary property $\phi(x, \mathcal{J})$ to be:

- $|J| = 2$ for all $J \in \mathcal{J}$.
- $x + y = n$ for all $\{x, y\} \in \mathcal{J}$.

Note that in this example, $k = 2$ and also $w = 2$ since an element x can be in at most one subset in $\mathcal{J}$.

Extending this let $g_S(n)$ denote the number of representations $n = x + y + z$ for distinct $x, y, z \in S$. Erdös and Tetali choose S randomly by letting $\mathtt{Pr}[i \in S] := \min(10 \left(\frac{\ln i}{i^2}\right)^{1/3} 1/2)$ and show that $\mathtt{E}[g_S(n)] = K \ln n$ for some constant $K > 0$. In fact one obtains a sharp concentration result around this value by bootstrapping the previous result. Set all weights to be unity as before and take the hereditary property $\phi(x, \mathcal{J})$ to be:

- $|J| = 3$ for all $J \in \mathcal{J}$.
- $x + y + z = n$ for all $\{x, y, z\} \in \mathcal{J}$.

This time $k = 3$ and w is bounded by $f_S(n)$ before.

Similarly, one can bootstrap upwards to k–ary representations for $k > 3$.

2.2.5 Discrete Isoperimetric Inequalities

Let A be a downward closed subset of the cube $\{0, 1\}^n$ equipped with the product measure, and let us consider the Hamming distance $d_H(x, A)$ from a point x to the set A. This is in fact a function of hereditary index sets. Take the weight system $w(i, J) := 1[i \in J]$ and the hereditary property $\phi(x, \mathcal{J})$ to be:

- $|J| = 1$ for all $J \in \mathcal{J}$.
- For all $j \in \mathcal{J}$, $x(j) = 1$ and $y(j) = 0$ for all $y \in A$.

We have $k = 1 = w$ and the bound obtained is comparable with the bounds obtained directly by isoperimetric inequalities in the theory of hereditary sets [1, Theorem 14] (see also [5, p. 132].

2.2.6 Edge Colouring

In this example, we shall consider some simple randomised algorithms for edge colouring a graph and illustrate the flexibility of our framework for giving sharp concentration results for different colouring schemes.

Given a graph G and a palette Δ of colours, we would like to assign colours to the edges in such a way that no two edges incident on a vertex have the same colour. We would also like the algorithm to be truly distributed, and hence to have a local character. This leads naturally to randomised algorithms of the type considered below. These algorithms run in stages. At each stage, some edges are successfully coloured by some simple local process. The others pass on to the next stage. Typically one analyses the algorithm stage by stage; in each stage, we would like to show that a significant number of edges are successfully coloured, so that the graph passed to the next stage is significantly smaller.

Algorithm 1: each edge picks a colour independently from the common palette $[\Delta]$. Conflicts are resolved in the simplest fashion: all edges incident on a given vertex which recive the same colour are decoloured and remain to be passed to the next stage.

We are interested in a sharp concentration result on the number of edges around a fixed vertex u that are successfuly coloured. Alternatively, we can give a sharp concentration result on the number of edges around u that are *not* successfully coloured.

The underlying product space is $[\Delta]^E(u)$ where $E(u)$ is the set of edges that are incident to u or to a neighbour of u. Let $w_(e, J) := [e \in J][u \in e]$ (thus only edges incident on u carry non–zero weights). Take the hereditary property $\phi(x, \mathcal{J})$ to be:

- The sets in $\mathcal{J}$ are all disjoint and $|J| = 2$ or $|J| = 3$ for all $J \in \mathcal{J}$.
- All edges in each $J \in \mathcal{J}$ are incident on a common vertex, and
- For each $J \in \mathcal{J}$ is monochromatic with respect to x i.e. $x(e) = x(e')$ for each $e, e' \in J \in \mathcal{J}$.

Some comments are in order about how to find $\mathcal{J}(x)$ in this case.

- First, for each edge $e = (u, v)$ which is unsuccessful beacuse of an edge $e' = (u', v)$ which received the same colour, pick the set $\{e, e'\}$.
- This leaves, for each colour, a bunch of edges incident on u with the same colour. Group these in disjoint pairs. Either this exhaust all of the bunch or leaves one. In the latter case, pick the triple that results by adding the remaining odd edge to the last pair.

In this case $k = 3$ and $w \leq 3w_{max}$; in the unweighted case, $w \leq 3$.

Next we consider another variant of the colouring algorithm. In **Algorithm 2**, we assume that edges are numbered in some canonical order and that after all edges have chosen colours tentatively, conflicts are resolved by decolouring higher numbered edges in favour of lower numbered edges that received the same colour. Thus, for each bunch of conflicting edges, a "winner" is chosen in some canonical fashion.

Let $w_(e, J) := [e \in J][u \in e]$ if e is not the lowest numbered edged in J (thus the lowest numbered edge in a set is not counted)). Take the hereditary property $\phi(x, \mathcal{J})$ to be:

- $|J| = 2$ for all $J \in \mathcal{J}$ and the higher numbered edge in two different $J, J' \in \mathcal{J}$ are different (i.e. the sets in $\mathcal{J}$ are "disjoint" with respect to their higher numbered edges).
- The two edges in each $J \in \mathcal{J}$ are incident on a common vertex, and
- For each $J \in \mathcal{J}$ is monochromatic with respect to x i.e. $x(e) = x(e')$ for each $e, e' \in J \in \mathcal{J}$.

Some comments are in order about how to find $\mathcal{J}(x)$ in this case.

- First, for each edge $e = (u, v)$ which is unsuccessful beacuse of an edge $e' = (u', v)$ which received the same colour, pick the set $\{e, e'\}$.
- This leaves, for each colour, a bunch of edges, A incident on u with the same colour. Let e^* be the lowest numbered edge in A. Pick the sets $\{e^*, e\}$ for $e^* \neq e \in A$.

In this case $k = 2$ and $w \leq 2w_{max}$; in the unweighted case, $w \leq 3$.

3 Talagrand's Inequality with Negative Dependence

In this section we give an extension of Talagrand's inequality to a setting where the the underlying measure in the product space is not the product measure but satisfies a negative dependence property. We will consider product spaces of the form $\Omega = \prod_i \Omega_i$ where each Ω_i is ordered.

3.1 A General Framework

A function $f(x_i, i \in i)$ is said to be non–decreasing (or non–increasing) if it is non–decreasing (non–increasing) in each co–ordinate. A subset A is non–decreasing (non–increasing) if its characteristic function $\chi(x_i, i \in I) := 1[(x_i, i \in I) \in A]$ is.

A set of variables $X_1, \ldots, X_n$ satisfies the *negative regression* condition $(-R)$ if for all disjoint index sets I and J of $[n]$ and all non–decreasing functions $f(x_i, i \in I)$,

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

is non–increasing in each $t_j, j \in J$. A measure in a product space $\Omega = \prod_i \Omega_i$ is said to satisfy negative regression if the co–ordinate functions $X_k(\omega_i, i \in I) = \omega_k$ satisfy negative regression.

A set of variables $X_1, \ldots, X_n$ is said to be *exchangeable* or *symmetric* if for all permutations $\sigma : [n] \to [n]$ and all $a_1, \ldots, a_n$,

$$\texttt{Pr}[X_i = a_i, i \in [n]] = \texttt{Pr}[X_i = a_{\sigma(i)}, i \in [n]].$$

A product space is symmetric if the co–ordinate functions are symmetric.

Theorem 4 (Talagrand's Inequality with Negative Dependence) *Let $\Omega = \{0,1\}^I$ be a symmetric product space satisfying $(-R)$. Then for any non–increasing subset $A \subseteq \Omega$,*

$$\texttt{Pr}[A]\texttt{Pr}[\overline{A_t}] \le e^{-t^2/4}.$$

Remark 5 Marton [3] has extended Talagrand's inequality to dependent variables in a different way. Our result is incomparable with Marton's: Our extension applies in situations that are *qualitatively* different from independence (i.e. negative) whereas Marton's inequality applies in situations *quantitatively* different from independence (i.e. one has a handle on the amount of dependence). Thus, while Marton's result covers a general kind of dependence (negative or otherwise), our result is stronger when applied in a situation of negative dependence on two counts: first, we do not require any handle on the amount of dependence (qualitative negative dependence in the form $(-R)$ suffices) and second, our estimate is sharper in general since it is the same as Talagrand's inequality itself.

The proof follows very much along the lines of the usual proof of Talagrand's inequality with some simple additional observations.

There are two main ingredients of the standard proof. The first is a general probabilistic fact, namely *Holder's inequality*: for any two random variables X and Y (whether independent or not), and any two reals p, q with $1/p + 1/q = 1$,

$$\mathtt{E}[|XY|] \leq \mathtt{E}[X^p]^{1/p}\mathtt{E}[Y^q]^{1/q}. \tag{7}$$

The second is a key inequality having to do with the convexity of Talagrand's distance function. For a subset $A \subseteq \Omega = \prod_{i\in[n]} \Omega_i$, denote by

$$\pi(A) := \{x \mid (x, x_n) \in A\},$$

the *projection* of A on all coordinates except the last, and for $x_n \in \Omega_n$, let

$$A(x_n) := \{x \mid (x, x_n) \in A\},$$

denote the x_n*–section* of A. Then for $(x, x_n) \in \Omega$, and all $0 \leq \lambda \leq 1$,

$$d_T^2((x, x_n), A) \leq (1-\lambda)d_T^2(x, \pi(A)) + \lambda d_T^2(x, A(x)) + (1-\lambda)^2. \tag{8}$$

We will also use crucially, two more simple observations. First are some general properties of the Talagrand distance valid in any product space of ordered spaces.

Proposition 6 *Let $A \subseteq \Omega$ be a non–increasing subset. Then*

- *Any projection $\pi(A)$ as well as any section $A(\omega)$ are also non–increasing.*
- *The Talagrand distance $d_T(x, A)$ is a non–decreasing function of x.*
- *For any $x = (x_1, \ldots, x_n)$, $d_T(x_1, \ldots, x_{n-1}, A(x_n))$ is a non–decreasing function of x.*

Proof. The first part is immediate. We will prove that $d_T(x, A)$ is non–decreasing and a similar proof applies to the last part.

Let $x \leq x'$ and suppose consider some arbitrary non–negative $\alpha \neq 0$. We shall show that

$$\min_{y \in A} \sum_{x_i \neq y_i} \alpha_i \leq \min_{y' \in A} \sum_{x_i \neq y'_i} \alpha_i,$$

which will complete the proof. For each $y' \in A$, consider $\min(x, y')$ which is in A since A is non–increasing. Hence,

$$\begin{aligned}
\min_{y \in A} \sum_{x_i \neq y_i} \alpha_i &\leq \sum_{x_i \neq \min(x,y')_i} \alpha_i \\
&\leq \sum_{x_i \neq y'_i} \alpha_i,
\end{aligned}$$

which completes the proof. ∎

The second is a property of the "influence" of Boolean variables from [2] that is of independent interest and has other applications. Let $f(x_i, i \in [n])$ be a Boolean function. A variable $x_i, i \in [n]$ is said to have *positive influence* for f and a given distribution if

$$\mathtt{E}[f(X_1, \ldots, X_n) \mid X_i = x],$$

is non–decreasing in x.

Proposition 7 *For any Boolean function $f(x_1, \ldots, x_n)$ and any distribution of its arguments, there is always a variable of positive influence i.e. an $i \in [n]$ such that*

$$\mathtt{E}[f(X_1, \ldots, X_n) \mid X_i = 0] \leq \mathtt{E}[f(X_1, \ldots, X_n) \mid X_i = 1].$$

Proof. (Of Theorem 4): We have,

$$\begin{aligned}
\mathtt{Pr}[d_T(X, A) > t] &= \mathtt{Pr}[e^{\frac{1}{4}d_T^2(X,A)} > e^{\frac{1}{4}t^2}] \\
&\leq \mathtt{E}[e^{\frac{1}{4}d_T^2(X,A)}]e^{-\frac{1}{4}t^2}.
\end{aligned}$$

Thus to complete the proof, we only need to show that $\mathtt{E}[e^{\frac{1}{4}d_T^2(X,A)}] \leq \frac{1}{\mathtt{Pr}[A]}$.

This will be proved by induction on the dimension n. For $n = 1$, we have $d_T(x, A) = 1[x \notin A]$ and hence

$$\begin{aligned} \mathtt{E}[e^{\frac{1}{4}d_T^2(X,A)}] &= e^{1/4}(1 - \mathtt{Pr}[A]) + \mathtt{Pr}[A] \\ &\leq 1/\mathtt{Pr}[A], \end{aligned}$$

since $e^{1/4}(1-u) + u \leq 1/u$ for $0 \leq u \leq 1$ by elementary calculus.

For the induction step, we shall use the key convexity inequality (8) to write

$$\begin{aligned} &\mathtt{E}[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_{n+1},A)}] \leq \\ &\qquad \mathtt{E}\left[e^{\frac{1}{4}(1-\lambda)^2} e^{(1-\lambda)\frac{1}{4}d_T^2(X_1,\ldots,X_n,\pi(A))} e^{\lambda\frac{1}{4}d_T^2(X_1,\ldots,X_n,A(X_{n+1}))}\right] \\ &= \mathtt{E}\left[e^{\frac{1}{4}(1-\lambda)^2}\mathtt{E}\left[e^{(1-\lambda)\frac{1}{4}d_T^2(X_1,\ldots,X_n,\pi(A))} e^{\lambda\frac{1}{4}d_T^2(X_1,\ldots,X_n,A(X_{n+1}))} \mid X_{n+1}\right]\right] \\ &\leq \mathtt{E}\left[e^{\frac{1}{4}(1-\lambda)^2}\left(\mathtt{E}\left[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_n,\pi(A))}|X_{n+1}\right]\right)^{1-\lambda}\left(\mathtt{E}\left[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_n,A(X_{N+1}))}|X_{n+1}\right]\right)^{\lambda}\right], \end{aligned}$$

using Holder's inequality in the last line.

Now, using Proposition 6 and the $(-R)$ property, we have that

$$\mathtt{E}\left[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_n,\pi(A))} \mid X_{n+1} = x_{n+1}\right],$$

is a non–increasing function of x_{n+1} whereas by Proposition 7,

$$\mathtt{E}\left[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_n,A(x_{n+1}))} \mid X_{n+1} = x_{n+1}\right]$$

is a non–decreasing function of x_{n+1}. Hence,

$$\begin{aligned} &\mathtt{E}[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_{n+1},A)}] \leq \\ &\qquad \mathtt{E}\left[e^{\frac{1}{4}(1-\lambda)^2}\mathtt{E}\left[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_n,\pi(A))} \mid X_{n+1}\right]\mathtt{E}\left[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_n,A(X_{n+1}))} \mid X_{n+1}\right]\right] \\ &= e^{\frac{1}{4}(1-\lambda)^2}\mathtt{E}\left[\mathtt{E}\left[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_n,\pi(A))} \mid X_{n+1}\right]\right]\mathtt{E}\left[\mathtt{E}\left[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_n,A(X_{n+1}))} \mid X_{n+1}\right]\right] \\ &= e^{\frac{1}{4}(1-\lambda)^2}\mathtt{E}\left[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_n,\pi(A))}\right]\mathtt{E}\left[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_n,A(X_{n+1}))}\right] \\ &\leq \mathtt{E}\left[e^{\frac{1}{4}(1-\lambda)^2}\left(\frac{1}{\mathtt{Pr}[\pi(A)]}\right)^{1-\lambda}\left(\frac{1}{\mathtt{Pr}[A(X_{n+1})]}\right)^{\lambda}\right], \end{aligned}$$

where the last line follows by induction. applied to the sets $\pi(A)$ and $A(X_n)$.

Now, we appeal to simple calculus once again to show that for $0 \leq r \leq 1$, $\inf_{0 \leq \lambda \leq 1} r^{-\lambda} e^{\frac{1}{4}(1-\lambda)^2} \leq 2 - r$. Hence, continuing the inequalities from above, we have

$$\begin{aligned}
\mathtt{E}[e^{\frac{1}{4}d_T^2(X_1,\ldots,X_{n+1},A)}] &\leq \mathtt{E}\left[e^{\frac{1}{4}(1-\lambda)^2}\left(\frac{1}{\mathtt{Pr}[\pi(A)]}\right)^{1-\lambda}\left(\frac{1}{\mathtt{Pr}[A(X_{n+1})]}\right)^{\lambda}\right] \\
&= \frac{1}{\mathtt{Pr}[\pi(A)]}\mathtt{E}\left[\left(\frac{\mathtt{Pr}[A(X_{n+1})]}{\mathtt{Pr}[\pi(A)]}\right)^{-\lambda} e^{\frac{1}{4}(1-\lambda)^2}\right] \\
&\leq \frac{1}{\mathtt{Pr}[\pi(A)]}\mathtt{E}\left[2 - \frac{\mathtt{Pr}[A(X_n)]}{\mathtt{Pr}[\pi(A)]}\right] \\
&= \frac{1}{\mathtt{Pr}[\pi(A)]}\left(2 - \frac{\mathtt{Pr}[A]}{\mathtt{Pr}[\pi(A)]}\right) \\
&= \frac{1}{\mathtt{Pr}[A]}\frac{\mathtt{Pr}[A]}{\mathtt{Pr}[\pi(A)]}\left(2 - \frac{\mathtt{Pr}[A]}{\mathtt{Pr}[\pi(A)]}\right) \\
&\leq \frac{1}{\mathtt{Pr}[A]},
\end{aligned}$$

since $x(2-x) \leq 1$ for all reals x. ∎

We can apply this to derive an analogue of Theorem 2 in a negative dependence situation. Call a hereditary property $\phi(x, \mathcal{J})$ of index sets *bi–hereditary* if it is also non–decreasing in x i.e. if $x \leq y$, then for any $\mathcal{J}$, $\phi(x, \mathcal{J}) \leq \phi(y, \mathcal{J})$.

Theorem 8 *Let $\Omega := \{0,1\}^n$ be a symmetric space satisfying $(-R)$ and let f be a function defined by a bi-hereditary property ϕ and weights $w(i, J), i \in I, J \subseteq I$. Then for all $t > 0$,*

$$\mathtt{Pr}[f > \mathtt{M}[f] + t] \leq 2\exp\left(\frac{-t^2}{4wk(M[f]+t)}\right),$$

and

$$\mathtt{Pr}[f < \mathtt{M}[f] - t] \leq 2\exp\left(\frac{-t^2}{4wkM[f]}\right),$$

where $\mathtt{M}[f]$ *is a median of* f *and* $k := \max_{x\in\Omega}\max_{J\in\mathcal{J}(x)}|J|$, *and* $w := \max_{x\in\Omega}\max_{i\in I}\sum_{\substack{J\in\mathcal{J}(x)\\ i\in J}}\sum_J w(J)$.

Proof. The proof is exactly analogous to that of Theorem 2 noting that if ϕ is bi–hereditary then the set $A(a) := \{y \mid f(y) \leq a\}$ is non–increasing, and hence Theorem 4 is applicable. ∎

3.2 Examples

We give examples to illustrate how to obtain concentration of measure results by applying Theorem 8 in settings where the underlying space is not a product space but is symmetric and satisfies $(-R)$.

3.2.1 Hypergeometric Distribution

Consider a sample of size n drawn from an urn containing $N \geq n$ balls, $M \leq N$ of which are red. We are interested in the number of red balls drawn in the sample, $H(N, M, n)$ – this is the hypergeometric distribution.

The underlying product space is $\{0,1\}^n$ where we identify 1 with a red ball and 0 with a non–red ball. The measure here is *not* the product measure but is easily shown to satisfy $(-R)$. To apply Theorem 8, take the hereditary property $\phi(x, \mathcal{J})$ to be

- $|J| = 1$ for all $J \in \mathcal{J}$.
- For $j \in \mathcal{J}$, $x(j) = 1$.

This is easily verified to be a bi–hereditary property and we get the following concentration of measure result on the hypergeometric distribution:

$$\mathtt{Pr}[H(N,M,n) > m + t] \leq 2\exp(\frac{-t^2}{m+t}),$$
$$\mathtt{Pr}[H(N,M,n) < m - t] \leq 2\exp(\frac{-t^2}{m}),$$

which is comparable to the Chernoff bound. Taking arbitrary weights $w_i, i \in [n]$ and setting $w(i, J) := [i \in J]w_i$ gives a weighted version of the result.

3.2.2 Balls and Bins

Let us return to the balls ad bins experiment, considered this time as the product space $\{0,1\}^{[n]\times[m]}$ where a 1 in co–ordinate (i,k) indicates that ball k is put into bin i. This measure (for any probabilities $p_{i,k}$ of ball k going into bin i) is easily shown to satisfy $(-R)$. Take $\phi(x, \mathcal{J})$ to be the same as in the previous example: $|J| = 1, J \in \mathcal{J}$ and $x(i,k) = 1$ for $(i,k) \in \mathcal{J}$. This gives the same concentration result for the number of non–empty bins as derived in the previous section.

3.2.3 Fermi–Dirac statistics

Consider an ensemble of particles occupying a set of n states. There are m types of particles, with $c_k \leq n$ particles of type $k \in [m]$. Each type is independent of any other type, but the particles of a given type satisfy the *Pauli exclusion principle*: no two particles of the same type may occupy the same state. In particular, the particles of a fixed type obey the Fermi–Dirac statistics: for each $k \in [m]$, all c_k subsets of the n–states are equally likely to be occupied. We are interested in the number of unoccupied states. Equivalently we acn focus on the number of occupied states. (This example generalises the balls and bins experiment above where each $c_k = 1$ and each ball is uniformly distributed in the bins.)

The underlying product space is $\{0,1\}^{[n]\times[m]}$ and the measure can be shown to satisfy $(-R)$. (the measure is not a product measure). Take the same bi–hereditary property $\phi(x, \mathcal{J})$ as in the previous balls and bins example to get a concentration result for the number of non–empty cells.

3.2.4 Discrete Isoperimetric Inequalities

One can extend the discrete isoperimetric inequalty from § 2.2.5 to the case when te underlying space is not a product measure but is symmetric and satisfies the $(-R)$ condition. The same inequality obtains in this setting as well.

4 Acknowledgments

Thanks to all the participants in the **BRICS** minicourse "Concentration of Measure with Applications to Ananlysis of Algorithms" that I taught along with Alessandro Panconesi at the Department of Computer Science at the University of Aarhus in September 1997. This work developed out of my efforts to understand Talagrand's inequality in the course of those lectures. Thanks also to Volker Priebe for faxing me material for the course including, invaluably, the chapter on Talagrand's inequality from Steele's brand new monograph [5].

References

[1] B. Bollobás and I. Leader: "Isoperimetric Inequalities and Fractional Set Systems", *J. Comb. Theory*, Seier A, 56, pp. 63–74, 1991.

[2] D. Dubhashi, P.B. Miltersen, D. Ranjan and V. Priebe, "On Positive Influence and Applications", manuscript, 1998.

[3] K. Marton "On a Measure Concentration Inequality of Talagrand for Dependent Random Variables", manuscript.

[4] J. Spencer, "Applications of Talagrand's Inequality", Unpublished manuscript.

[5] J.M. Steele, *Probability Theory and Combinatorial Optimization*, SIAM Monographs 1997.

[6] M. Talagrand, "Concentration of Measure and Isoperimetric Inequalities in Product Spaces", *Publ. Math. IHES* 81, 73–205, 1995.

Recent BRICS Report Series Publications

RS-98-25 Devdatt P. Dubhashi. ***Talagrand's Inequality in Hereditary Settings.*** **October 1998. 22 pp.**

RS-98-24 Devdatt P. Dubhashi. ***Talagrand's Inequality and Locality in Distributed Computing.*** **October 1998. 14 pp.**

RS-98-23 Devdatt P. Dubhashi. ***Martingales and Locality in Distributed Computing.*** **October 1998. 19 pp.**

RS-98-22 Gian Luca Cattani, John Power, and Glynn Winskel. ***A Categorical Axiomatics for Bisimulation.*** **September 1998. ii+21 pp. Appears in Sangiorgi and de Simone, editors,** ***Concurrency Theory: 9th International Conference*****, CONCUR '98 Proceedings, LNCS 1466, 1998, pages 581–596.**

RS-98-21 John Power, Gian Luca Cattani, and Glynn Winskel. ***A Representation Result for Free Cocompletions.*** **September 1998. 16 pp.**

RS-98-20 Søren Riis and Meera Sitharam. ***Uniformly Generated Submodules of Permutation Modules.*** **September 1998. 35 pp.**

RS-98-19 Søren Riis and Meera Sitharam. ***Generating Hard Tautologies Using Predicate Logic and the Symmetric Group.*** **September 1998. 13 pp.**

RS-98-18 Ulrich Kohlenbach. ***Things that can and things that can't be done in PRA.*** **September 1998. 24 pp.**

RS-98-17 Roberto Bruni, José Meseguer, Ugo Montanari, and Vladimiro Sassone. ***A Comparison of Petri Net Semantics under the Collective Token Philosophy.*** **September 1998. 20 pp. To appear in** ***4th Asian Computing Science Conference*****, ASIAN '98 Proceedings, LNCS, 1998.**

RS-98-16 Stephen Alstrup, Thore Husfeldt, and Theis Rauhe. ***Marked Ancestor Problems.*** **September 1998.**

RS-98-15 Jung-taek Kim, Kwangkeun Yi, and Olivier Danvy. ***Assessing the Overhead of ML Exceptions by Selective CPS Transformation.*** **September 1998. 31 pp. To appear in the proceedings of the** ***1998 ACM SIGPLAN Workshop on ML*****, Baltimore, Maryland, September 26, 1998.**