A Tail Bound for Read-k Families of Functions

Dmitry Gavinsky
NEC Laboratories America, Inc.
Princeton, NJ 08540, U.S.A.

Shachar Lovett
Institute of Advanced Study
Princeton, NJ 08540, U.S.A.

Michael Saks
Department of Mathematics, Rutgers University
Piscataway, NJ 08854, U.S.A.

Srikanth Srinivasan
DIMACS, Rutgers University
Piscataway, NJ 08854, U.S.A.

Abstract

We prove a Chernoff-like large deviation bound on the sum of non-independent random variables that have the following dependence structure. The variables $Y_1, \ldots, Y_r$ are arbitrary Boolean functions of independent random variables $X_1, \ldots, X_m$, modulo a restriction that every X_i influences at most k of the variables $Y_1, \ldots, Y_r$.

1 Introduction

Let $Y_1, \ldots, Y_r$ be independent indicator random variables with $\mathbf{Pr}\,[Y_i = 1] = p$ and let $Y = Y_1 + \ldots + Y_r$ denote their sum. The average number of variables which are set is $\mathbf{E}\,[Y] = pr$. A basic question is how concentrated is the sum around its average. The answer is given by Chernoff bound: for any $\varepsilon > 0$, the probability that the fraction of variables that occur exceeds the expectation by more than ε is bounded by

$$\mathbf{Pr}\,[Y_1 + \ldots + Y_r \geq (p + \varepsilon)r] \leq e^{-d_{KL}(p+\varepsilon||p)\cdot r} \leq e^{-2\varepsilon^2 r},$$

and similarly, the probability that it is below the expectation by more than ε is bounded by

$$\mathbf{Pr}\,[Y_1 + \ldots + Y_r \leq (p - \varepsilon)r] \leq e^{-d_{KL}(p-\varepsilon||p)\cdot r} \leq e^{-2\varepsilon^2 r}.$$

Here, $d_{KL}(q||p)$ is Kullback-Leibler divergence defined as

$$d_{KL}(q||p) = q \log\left(\frac{q}{p}\right) + (1-q)\log\left(\frac{1-q}{1-p}\right),$$

and the latter estimate (which is the one more commonly used in applications) follows from $d_{KL}(q||p) \geq 2(q-p)^2$.

Our focus in this paper is on deriving similar estimates when the variables $Y_1, \ldots, Y_r$ are not independent, but the level of dependence between them is in some sense 'weak'. These type of questions are commonly studied in probability theory, as they allow one to apply Chernoff-like bounds to more general scenarios. For example, one case which is used in various applications is where the variables $Y_1, \ldots, Y_r$ are assumed to be k-wise independent. In this case, Bellare and Rompel [BR94] obtained Chernoff-like tail bounds on the probability

that the number of set variables deviates from its expectation. Another well studied case is when $Y_1, \ldots, Y_r$ form a martingale, in which case Azuma inequality and its generalizations give bounds which are comparable to Chernoff bound.

We consider in this paper another model of weak dependence. Assume that the variables $Y_1, \ldots, Y_r$ can be factored as functions of independent random variables $X_1, \ldots, X_m$. More concretely, each Y_j is a function of a subset of the variables $X_1, \ldots, X_m$. One extreme case is that these subsets are disjoint, in which case the variables $Y_1, \ldots, Y_r$ are independent. The other extreme case is when these subsets all share a common element, in which case $Y_1, \ldots, Y_r$ can be arbitrary. We say that $Y_1, \ldots, Y_r$ are a *read-k family* if there exists such a factorization where each X_i influences at most k of the variables $Y_1, \ldots, Y_r$.

Definition 1. (Read-k families) *Let $X_1, \ldots, X_m$ be independent random variables. For $j \in [r]$, let $P_j \subseteq [m]$ and let f_j be a Boolean function of $(X_i)_{i \in P_j}$. Assume that $|\{j \mid i \in P_j\}| \leq k$ for every $i \in [m]$. Then the random variables $Y_j = f_j((X_i)_{i \in P_j})$ are called a read-k family.*

There are several motivations for studying read-k families of functions: for example, they arise naturally when studying subgraphs counts in random graphs; or in generalizations of read-once models in computational complexity. We will not discuss applications in this paper, but instead focus on the basic properties of read-k families. Our main result is that Chernoff-like tail bounds hold for read-k families, where the bounds degrade as k increases.

Theorem 1.1. *Let $Y_1, \ldots, Y_r$ be a family of read-k indicator variables with $\mathbf{Pr}\,[Y_i = 1] = p_i$, and let p be the average of $p_1, \ldots, p_r$. Then for any $\varepsilon > 0$,*

$$\mathbf{Pr}\,[Y_1 + \ldots + Y_r \geq (p + \varepsilon)r] \leq e^{-d_{KL}(p+\varepsilon||p) \cdot r/k}$$

and

$$\mathbf{Pr}\,[Y_1 + \ldots + Y_r \leq (p - \varepsilon)r] \leq e^{-d_{KL}(p-\varepsilon||p) \cdot r/k}.$$

That is, we obtain the same bound as that of the standard Chernoff bound, except that the exponent is divided by k. This is clearly tight: let $Y_1 = \ldots = Y_k = X_1$, $Y_{k+1} = \ldots = Y_{2k} = X_2$, etc, where $X_1, X_2, \ldots, X_{r/k}$ are independent with $\Pr[X_i = 1] = p$. Then for example

$$\mathbf{Pr}\,[Y_1 + \ldots + Y_r \geq (p + \varepsilon)r] = \mathbf{Pr}\,[X_1 + \ldots + X_{r/k} \geq (p + \varepsilon)r/k].$$

We note that concentration bounds for $Y = \sum_i Y_i$ may also be obtained by observing that Y is a k-Lipschitz function of the underlying independent random variables $X_1, \ldots, X_m$ and then applying standard Martingale-based arguments, such as Azuma's inequality (see, for example, [DP09]). However, these techniques are limited in the sense that they only yield concentration bounds when the deviation $\varepsilon \cdot r$ above is at least $\sqrt{m}$ or so, whereas our results yield concentration bounds even when the deviation is roughly the square root of the mean $p \cdot r$, which may be much smaller. It is also known (see [Von10, Section 4]) that such 'dimension-free' concentration bounds cannot be obtained for Lipschitz functions in general.

1.1 Proof overview

Let $Y_1, \ldots, Y_r$ be a read-k family with $\Pr[Y_i = 1] = p_i$. Let us consider for a moment a more basic question: what is the maximal probability that $Y_1 = \ldots = Y_r = 1$? The answer to this question is given by Shearer's lemma [CGFS86], which is a simple yet extremely powerful tool.

Theorem 1.2. *Let $Y_1, \ldots, Y_r$ be a family of read-k indicator variables with $\mathbf{Pr}\left[Y_i = 1\right] = p$. Then*

$$\mathbf{Pr}\left[Y_1 = \ldots = Y_r = 1\right] \leq p^{r/k}.$$

Note that the answer is the k-th root of the answer in the case where the variables are independent, similar to what we get for the tail bounds. The result is tight, as can be seen by taking $Y_1 = \ldots = Y_k = X_1$, $Y_{k+1} = \ldots = Y_{2k} = X_2$, etc, where $X_1, X_2, \ldots, X_{r/k}$ are independent with $\mathbf{Pr}\left[X_i = 1\right] = p$. The proof of Shearer's lemma is based on the entropy method. We refer the interested reader to a survey of Radhakrishnan [Rad03] on applications of Shearer's lemma; and to continuous analogs by Finner [Fin92] and Friedgut [Fri04].

We derive Theorem 1.1 by constructing an information-theoretic proof of Chernoff bound and applying Shearer's lemma to make the proof robust for the case of non-independent random variables forming a read-k family. In our proof we use the "entropy method" that was introduced by Ledoux [Led96], and more recently has been used to prove a number of concentration inequalities (cf. [BLM03]). From a technical point of view, we construct analogs of Shearer's lemma for Kullback-Leibler divergence.

2   Preliminaries

Let X be a random variable and A be a subset of its support. We write "$X \in A$" to address the event "the value of X belongs to A".

2.1   Entropy

If X is distributed according to μ, then we will write both $\mathbf{H}\left[X\right]$ and $\mathbf{H}\left[\mu\right]$ to denote the entropy $\mathbf{E}_{X=x}\left[\log(1/\mu(x))\right]$. All logarithms in this paper are natural. If X is supported in the set A then $\mathbf{H}\left[X\right] \leq \log|A|$, where equality holds when X is the uniform distribution over A. For two random variables X, Y we denote their conditional entropy by $\mathbf{H}\left[X|Y\right] = \mathbf{E}_{Y=y}\left[\mathbf{H}\left[X|Y=y\right]\right]$. We note that always $\mathbf{H}\left[X\right] \geq \mathbf{H}\left[X|Y\right]$. If $X_1, \ldots, X_m$ are random variables then $\mathbf{H}\left[X_1, \ldots, X_m\right]$ denotes the entropy of their joint distribution. The chain rule of entropy is

$$\mathbf{H}\left[X_1, \ldots, X_m\right] = \mathbf{H}\left[X_1\right] + \mathbf{H}\left[X_2|X_1\right] + \ldots + \mathbf{H}\left[X_m|X_1, \ldots, X_{m-1}\right].$$

2.2   Relative entropy

Let μ' and μ be distributions defined over the same discrete domain A. The relative entropy (or Kullback-Leibler divergence) between μ' and μ is defined as

$$d_{KL}\left(\mu'||\mu\right) \overset{\text{def}}{=} \sum_{a \in A} \mu'(a) \log \frac{\mu'(a)}{\mu(a)}.$$

If random variables X, X' are distributed according to μ, μ', accordingly, then we write $d_{KL}\left(X'||X\right) = d_{KL}\left(\mu'||\mu\right)$. Kullback-Leibler divergence also satisfies a chain rule. We will only need the following simple corollary of it.

Fact 2.1. *Let X, X' be random variables defined on a domain A. Let $\phi : A \to B$ be some function. Then*

$$d_{KL}\left(X'||X\right) \geq d_{KL}\left(\phi(X')||\phi(X)\right).$$

We will usually use relative entropy in the case when the second operand is the uniform distribution over some set. The following observations will be useful.

Fact 2.2. *Let μ be the uniform distribution over some set A, and let μ' be any distribution over A. Then*

$$d_{KL}\left(\mu'||\mu\right) = \mathbf{H}\left[\mu\right] - \mathbf{H}\left[\mu'\right].$$

Claim 2.3. *Let μ be the uniform distribution over some set A, and let $A' \subseteq A$ satisfy $\mu(A') = p$. Let μ' be any distribution over A, satisfying $\mu'(A') = q$. Then*

$$d_{KL}\left(\mu'||\mu\right) \geq d_{KL}\left(q||p\right).$$

Here we use the standard convention $d_{KL}\left(q||p\right)$ to denote the relative entropy between Bernoulli distributions whose probabilities of positive outcomes are, respectively, q and p.

Proof. Let X be a random variable taking values in A according to the distribution μ', and let I be the indicator of the event "$X \in A'$". Let $\mathbf{H}\left[q\right]$ denote the entropy of Bernoulli distribution whose probability of positive outcome is q, then

$$\mathbf{H}\left[\mu'\right] = \mathbf{H}\left[I\right] + \mathbf{H}\left[X|I\right] \leq \mathbf{H}\left[q\right] + q\log\left(\left|A'\right|\right) + (1-q)\log\left(\left|A\right| - \left|A'\right|\right)$$

$$= \mathbf{H}\left[q\right] + q\log\left(p\left|A\right|\right) + (1-q)\log\left((1-p)\left|A\right|\right)$$

$$= \mathbf{H}\left[q\right] + \log\left|A\right| + q\log p + (1-q)\log(1-p)$$

$$= \mathbf{H}\left[\mu\right] - d_{KL}\left(q||p\right),$$

and the result follows by Fact 2.2. ∎

We will need the following facts on $d_{KL}\left(q||p\right)$. The first is the convexity of Kullback-Leibler divergence. The last two relate to monotonicity of Kullback-Leibler divergence.

Fact 2.4. *Let $0 \leq p_1, p_2, q_1, q_2, \lambda \leq 1$. Then*

$$d_{KL}\left(\lambda p_1 + (1-\lambda)p_2||\lambda q_1 + (1-\lambda)q_2\right) \leq \lambda \cdot d_{KL}\left(p_1||q_1\right) + (1-\lambda)\cdot d_{KL}\left(p_2||q_2\right).$$

Fact 2.5. *Let $0 \leq p \leq q \leq q' \leq 1$. Then $d_{KL}\left(q'||p\right) \geq d_{KL}\left(q||p\right)$.*

Fact 2.6. *Let $0 \leq q' \leq q \leq p \leq 1$. Then $d_{KL}\left(q'||p\right) \geq d_{KL}\left(q||p\right)$.*

2.3   Shearer's Lemma

In order to derive bounds on mutual entropy we will use an elegant tool of amazing universality, *Shearer's Lemma* [CGFS86]. The following formulation of the lemma will be convenient for us.

Lemma 2.7. [Shearer's Lemma] *Let $X_1, \ldots, X_m$ be random variables and let $P_1, \ldots, P_r \subseteq [m]$ be subsets such that for every $i \in [m]$ it holds that $|\{j \mid i \in P_j\}| \geq k$. Then*

$$k \cdot \mathbf{H}\left[(X_1, \ldots, X_m)\right] \leq \sum_{j=1}^{r} \mathbf{H}\left[(X_i)_{i \in P_j}\right].$$

For completeness we include a short proof, based on the idea commonly attributed to Jaikumar Radhakrishnan.

Proof. Let us denote $s_j = |P_j|$ and let $P_j = \{i_{j,1}, \ldots, i_{j,s_j}\}$, where we order the elements $i_{j,1} < \ldots < i_{j,s_j}$. We apply the chain rule for entropy:

$$\mathbf{H}\left[(X_i)_{i \in P_j}\right] = \sum_{t=1}^{s_j} \mathbf{H}\left[X_{i_{j,t}} \mid X_{i_{j,1}}, \ldots, X_{i_{j,t-1}}\right] \geq \sum_{t=1}^{s_j} \mathbf{H}\left[X_{i_{j,t}} \mid X_1, \ldots, X_{i_{j,t}-1}\right],$$

where the second inequality follows from non-increasing of entropy as a result of conditioning. Summing over all $1 \leq j \leq r$ we obtain that

$$\sum_{j=1}^{r} \mathbf{H}\left[(X_i)_{i \in P_j}\right] \geq \sum_{i=1}^{m} |\{j \mid i \in P_j\}| \cdot \mathbf{H}\left[X_i \mid X_1, \ldots, X_{i-1}\right] \geq k \cdot \sum_{i=1}^{m} \mathbf{H}\left[X_i \mid X_1, \ldots, X_{i-1}\right],$$

and the result follows. ∎

We need the following simple corollary of Shearer's Lemma:

Corollary 2.8. *Let $X_1, \ldots, X_m$ be random variables and let $P_1, \ldots, P_r \subseteq [m]$ be such that for every $i \in [m]$ it holds that $|\{j \mid i \in P_j\}| \leq k$. For each $i \in [m]$, let Y_i be an independent random variable, uniformly distributed over a set that includes the support of X_i. Then*

$$k \cdot d_{KL}\left((X_i)_{i=1}^{m} \| (Y_i)_{i=1}^{m}\right) \geq \sum_{j=1}^{r} d_{KL}\left((X_i)_{i \in P_j} \| (Y_i)_{i \in P_j}\right).$$

Note the difference in the conditions put upon $|\{j \mid i \in P_j\}|$: in Shearer's Lemma it was "$\geq k$", and in the corollary it is "$\leq k$".

Proof. First observe that w.l.o.g we can assume that $|\{j \mid i \in P_j\}| = k$ for every i (adding elements to some P_j can only increase the right-hand side of the stated inequality by Fact 2.1). Under this assumption we can apply Lemma 2.7, concluding that

$$k \cdot \mathbf{H}\left[(X_i)_{i=1}^{m}\right] \leq \sum_{j=1}^{r} \mathbf{H}\left[(X_i)_{i \in P_j}\right].$$

Applying Claim 2.3 leads to

$$k \cdot d_{KL}\left((X_i)_{i=1}^{m} \| (Y_i)_{i=1}^{m}\right) = k \, \mathbf{H}\left[(Y_i)_{i=1}^{m}\right] - k \, \mathbf{H}\left[(X_i)_{i=1}^{m}\right]$$

$$\geq k \, \mathbf{H}\left[(Y_i)_{i=1}^{m}\right] - \sum_{j=1}^{r} \mathbf{H}\left[(X_i)_{i \in P_j}\right]$$

$$= \sum_{j=1}^{r} \left(\mathbf{H}\left[(Y_i)_{i \in P_j}\right] - \mathbf{H}\left[(X_i)_{i \in P_j}\right]\right),$$

where the last equality follows from mutual independence of Y_i's and the assumption that every i belongs to exactly k sets among $P_1, \ldots, P_r$. The result follows. ∎

3   Read-k families of functions

We prove Theorem 1.1 in this section. We first fix notations. Let $X_1, \ldots, X_m$ be independent random variables. For $j \in [r]$, let $P_j \subseteq [m]$ be a subset and let $Y_j = f_j((X_i)_{i \in P_j})$. We assume that $Y_1, \ldots, Y_r$ are a read-k family, that is, $|\{j \mid i \in P_j\}| \leq k$ for every $i \in [m]$. Let $\mathbf{Pr}[Y_i = 1] = p_i$, and let p be the average of $p_1, \ldots, p_r$. Let $\varepsilon > 0$ be fixed. We will show that

$$\mathbf{Pr}[Y_1 + \ldots + Y_r \geq (p + \varepsilon)r] \leq e^{-d_{KL}(p+\varepsilon||p) \cdot r/k} \tag{1}$$

and

$$\mathbf{Pr}[Y_1 + \ldots + Y_r \leq (p - \varepsilon)r] \leq e^{-d_{KL}(p-\varepsilon||p) \cdot r/k}. \tag{2}$$

We first note that it suffices to prove the theorem in the case where each X_i is uniform over a finite set A_i. This is since we can assume w.l.o.g that each X_i is a discrete random variable, and any discrete distribution can be approximated to arbitrary accuracy by a function of a uniform distribution over a large enough finite set. We denote by $A \overset{\text{def}}{=} A_1 \times \ldots \times A_m$ the set of all possible inputs, and let μ denote the uniform distribution over A.

We start by proving (1). For $t = (p + \varepsilon)r$ let us denote by $A^{\geq t}$ the set of inputs for which $\sum f_i \geq t$,

$$A^{\geq t} \overset{\text{def}}{=} \left\{ (a_1, \ldots, a_m) \in A_1 \times \cdots \times A_m \; \middle| \; \sum_{j=1}^{r} f_j(a_1, \ldots, a_m) \geq t \right\}.$$

We denote by $\mu^{\geq t}$ be uniform distribution over $A^{\geq t}$. We next define restrictions of these distributions to the sets $P_1, \ldots, P_m$. For a set $P_j \subseteq [m]$ we denote by μ_j the restriction of μ to the coordinates of P_j (it is the uniform distribution over $\prod_{i \in P_j} A_i$), and by $\mu_j^{\geq t}$ the restriction of $\mu^{\geq t}$ to the coordinates of P_j.

We wish to upper bound the probability that $\sum f_j \geq t$. Equivalently, we wish to upper bound $|A^{\geq t}|/|A|$. Taking logarithms, this is equal to $\mathbf{H}\left[\mu^{\geq t}\right] - \mathbf{H}[\mu]$. By Fact 2.2, this is equivalent to $d_{KL}\left(\mu^{\geq t}||\mu\right)$. That is,

$$\mathbf{Pr}_{\mu}\left[\sum f_j \geq t\right] = \frac{|A^{\geq t}|}{|A|} = \exp\left(\mathbf{H}\left[\mu^{\geq t}\right] - \mathbf{H}[\mu]\right) = \exp\left(-d_{KL}\left(\mu^{\geq t}||\mu\right)\right). \tag{3}$$

So, we need to obtain a lower bound on $d_{KL}\left(\mu^{\geq t}||\mu\right)$. Naturally, to do that we will use Corollary 2.8. We thus have

$$\mathbf{Pr}_{\mu}\left[\sum f_j \geq t\right] \leq \exp\left(-\frac{1}{k} \sum_{j=1}^{r} d_{KL}\left(\mu_j^{\geq t}||\mu_j\right)\right). \tag{4}$$

Recall that $p_j = \mathbf{Pr}_{\mu}[f_j = 1]$ is the probability that $f_j = 1$ under the uniform distribution. Let us denote by $q_j = \mathbf{Pr}_{\mu^{\geq t}}[f_j = 1]$ the probability that $f_j = 1$ conditioned on $\sum f_j \geq t$.

6

By Fact 2.1 (using $\phi = f_j$) we have that $d_{KL}\left(\mu_j^{\geq t}\middle\|\mu_j\right) \geq d_{KL}\left(q_j\|p_j\right)$, hence

$$\Pr_{\mu}\left[\sum f_j \geq t\right] \leq \exp(-\frac{1}{k}\sum_{j=1}^{r} d_{KL}\left(q_j\|p_j\right)). \tag{5}$$

By convexity of Kullback-Leibler divergence (Fact 2.4) we have that

$$\frac{1}{r}\sum_{j=1}^{r} d_{KL}\left(q_j\|p_j\right) \geq d_{KL}\left(q\|p\right),$$

where $q = \frac{1}{r}\sum_{j=1}^{r} q_j$ and $p = \frac{1}{r}\sum_{j=1}^{r} p_j$. To conclude, recall that q_j is the probability that $f_j = 1$ given that $f_1 + \ldots + f_r \geq t$. Hence the sum $\sum_{j=1}^{r} q_j$ is the expected number of f_j for which $f_j = 1$, which by definition is at least t. Hence $q \geq t/r = p + \varepsilon$ and $d_{KL}\left(q\|p\right) \geq d_{KL}\left(p + \varepsilon\|p\right)$ by Fact 2.5. We have thus shown that

$$\Pr_{\mu}\left[f_1 + \ldots + f_r \geq (p + \varepsilon)r\right] \leq \exp(-d_{KL}\left(p + \varepsilon\|p\right) \cdot r/k).$$

which establishes (1).

The proof of (2) is very similar. Define for $t = (p - \varepsilon)r$ the set $A^{\leq t}$ as

$$A^{\leq t} \stackrel{\text{def}}{=} \left\{(a_1, \ldots, a_m) \in A_1 \times \cdots \times A_m \;\middle|\; \sum_{j=1}^{r} f_j(a_1, \ldots, a_m) \leq t\right\}.$$

Let $\mu^{\leq t}$ be the uniform distribution over $A^{\leq t}$, and let $\mu_j^{\leq t}$ be its projection to $(X_i)_{i \in P_j}$. Then analogously to (4) we have

$$\Pr_{\mu}\left[\sum f_j \leq t\right] \leq \exp\left(-\frac{1}{k}\sum_{j=1}^{r} d_{KL}\left(\mu_j^{\leq t}\middle\|\mu_j\right)\right). \tag{6}$$

Let $q_j' = \Pr_{\mu^{\leq t}}\left[f_j = 1\right]$ be the probability that $f_j = 1$ conditioned on $\sum f_j \leq t$. Then analogously to (5) we have

$$\Pr_{\mu}\left[\sum f_j \leq t\right] \leq \exp(-\frac{1}{k}\sum_{j=1}^{r} d_{KL}\left(q_j'\|p_j\right)). \tag{7}$$

Let $q' = \frac{1}{r}\sum_{j=1}^{r} q_j'$. Then by convexity of Kullback-Leibler divergence, we have

$$\frac{1}{r}\sum_{j=1}^{r} d_{KL}\left(q_j\|p_j\right) \geq d_{KL}\left(q'\|p\right).$$

Now, $q' \leq t/r = p - \varepsilon$ by definition, hence by Fact 2.6 $d_{KL}\left(q'\|p\right) \geq d_{KL}\left(p - \varepsilon\|p\right)$. So, we conclude that

$$\Pr_{\mu}\left[f_1 + \ldots + f_r \leq (p - \varepsilon)r\right] \leq \exp(-d_{KL}\left(p - \varepsilon\|p\right) \cdot r/k).$$

which establishes (2).

7

References

[BLM03]   S. Boucheron, G. Lugosi, and P. Massart. Concentration Inequalities using the Entropy Method. *The Annals of Probability 31(3)*, pages 1583–1614, 2003.

[BR94]   M. Bellare and J. Rompel. Randomness-efficient oblivious sampling. In *Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on*, pages 276 –287, nov 1994.

[CGFS86] F R Chung, R L Graham, P Frankl, and J B Shearer. Some intersection theorems for ordered sets and graphs. *J. Comb. Theory Ser. A*, 43(1):23–37, September 1986.

[DP09]   Devdatt P. Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

[Fin92]   Helmut Finner. A generalization of holder's inequality and some probability inequalities. *The Annals of Probability*, 20(4):pp. 1893–1901, 1992.

[Fri04]   Ehud Friedgut. Hypergraphs, entropy, and inequalities. *The American Mathematical Monthly*, 111(9):pp. 749–760, 2004.

[Led96]   M. Ledoux. On Talagrand's Deviation Inequalities for Product Measures. *ESAIM: Probability and Statistics 1*, pages 63–87, 1996.

[Rad03]   Jaikumar Radhakrishnan. Entropy and counting. *IIT Kharagpur Golden Jubilee Volume*, pages 1–25, 2003.

[Von10]   Jan Vondrák. A note on concentration of submodular functions. *CoRR*, abs/1005.2791, 2010.