Electron. Commun. Probab. **18** (2013), no. 44, 1–11.
DOI: 10.1214/ECP.v18-2659
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

On McDiarmid's concentration inequality

Emmanuel Rio*

Abstract

In this paper we improve the rate function in the McDiarmid concentration inequality for separately Lipschitz functions of independent random variables. In particular the rate function tends to infinity at the boundary. We also prove that in some cases the usual normalization factor is not adequate and may be improved.

Keywords: McDiarmid inequality ; Concentration inequality ; Hoeffding inequality ; Vajda's tight lower bound .
AMS MSC 2010: 60E15.
Submitted to ECP on March 8, 2013, final version accepted on June 1, 2013.

1 Introduction

Throughout the paper $(E_1, d_1), \ldots, (E_n, d_n)$ is a finite sequence of separable metric spaces with positive finite diameters $\Delta_1, \ldots, \Delta_n$. Let $E^n = E_1 \times \cdots \times E_n$. A function f from E^n into $\mathbb{R}$ is said to be separately 1-Lipschitz if

$$|f(x_1, \ldots, x_n) - f(y_1, \ldots, y_n)| \leq d_1(x_1, y_1) + \cdots + d_n(x_n, y_n).$$

Let $(\Omega, \mathcal{T}, \mathbb{P})$ be a probability space and $X = (X_1, \ldots, X_n)$ be a random vector with independent components, with values in E^n. Let f be any separately 1-Lipschitz function from E^n into $\mathbb{R}$. Set

$$Z = f(X) = f(X_1, \ldots, X_n). \tag{1.1}$$

Let the McDiarmid diameter σ_n be defined by

$$\sigma_n^2 = \Delta_1^2 + \Delta_2^2 + \cdots + \Delta_n^2. \tag{1.2}$$

McDiarmid [9], [10] proved that, for any positive x,

$$\mathbb{P}(Z - \mathbb{E}(Z) \geq \sigma_n x) \leq \exp(-2x^2). \tag{1.3}$$

This inequality is an extension of Theorem 2 in Hoeffding [6]. We refer to [4], Chapter 2, for more about concentration inequalities. Later Bentkus [3] (paper submitted on August 17, 2001) and Pinelis [12] replaced the upper bound in (1.3) by a Gaussian tail function. They proved that

$$\mathbb{P}(Z - \mathbb{E}(Z) \geq \sigma_n x) \leq c\,\mathbb{P}(Y \geq 2x), \quad \text{with } Y \stackrel{D}{=} N(0,1). \tag{1.4}$$

*UMR 8100 CNRS, Laboratoire de Mathématiques de Versailles, France. E-mail: emmanuel.rio@uvsq.fr

The best known constant in (1.4) is $c = 5.70$, due to [12]. In the real-valued case, the bounds may be much better in the moderate deviations area when the standard deviations of the random variables are significantly smaller than the diameters Δ_i. Furthermore the random variables do not need to be bounded from below. We refer to [1], [2], [7], [11] and [12] for more about this subject, which is essentially outside the scope of this paper. Here we do not impose conditions on the variances of the random variables. More precisely, our aim is to get upper bounds for the quantity $P_{McD}(z, \Delta)$ introduced before Inequality (1.9) below.

We now comment on the results (1.3) and (1.4). Since f is separately 1-Lipschitz and the spaces E_i have a finite diameter Δ_i, the function f is uniformly bounded over E^n. Furthermore if $M = \sup_{E^n} f$ and $m = \inf_{E^n} f$, then

$$m \leq Z \leq M \text{ and } M - m \leq \Delta_1 + \Delta_2 + \cdots + \Delta_n := D_n. \tag{1.5}$$

From (1.5) it follows that

$$\mathbb{P}(Z - \mathbb{E}(Z) \geq D_n) = \mathbb{P}(Z = M \text{ and } \mathbb{E}(Z) = m) = 0. \tag{1.6}$$

Now (1.6) cannot be deduced from either (1.3) or (1.4). Hence it seems clear that the rate function $2x^2$ in the McDiarmid inequality (1.3) is suboptimal for large values of x. One of the goals of this paper is to improve the rate function appearing in (1.3). In Section 2, we give a more efficient large deviations rate function in the case $\Delta_1 = \Delta_2 = \cdots = \Delta_n = 1$. In particular we prove that, for any x in $[0, 1]$,

$$\mathbb{P}\big(Z - \mathbb{E}(Z) \geq n(1 - x)\big) \leq x^{n(1-x^2)}. \tag{1.7}$$

This inequality implies (1.3) and yields (1.6). Next, in Section 3, we extend the results of Section 2 to the case of distinct diameters, for small values or large values of the deviation. In Theorem 3.1 we give the following extension of (1.7): for any x in $[0, 1]$,

$$\mathbb{P}\big(Z - \mathbb{E}(Z) \geq D_n(1 - x)\big) \leq x^{(1-x^2)D_n^2/\sigma_n^2}. \tag{1.8}$$

We now recall the known lower bounds for large values of the deviations. Take $E_i = [0, \Delta_i]$. Let $\Delta = (\Delta_1, \Delta_2, \ldots, \Delta_n)$ and let $P_{McD}(z, \Delta)$ be defined as the maximal value of $\mathbb{P}(Z - \mathbb{E}(Z) \geq z)$ over all the separately 1-Lipschitz functions and all the random vectors X with values in E and with independent components. By Proposition 5.7 in Ohwadi et al. [11],

$$P_{McD}(D_n - nx, \Delta) \geq x^n / (\Delta_1 \Delta_2 \ldots \Delta_n) \text{ for any } x \leq \min(\Delta_1, \Delta_2, \ldots, \Delta_n). \tag{1.9}$$

As shown by the converse inequality (1.9), (1.8) is suitable for large values of the deviation when $\sigma_n^2 \sim D_n^2/n$. Nevertheless (1.8) has to be improved when $\sigma_n \gg n^{-1/2} D_n$. In Theorem 3.2 of Section 3, we prove the converse inequality of (1.9) with $D_n - (56/67)nx$ instead of $D_n - nx$. Finally we give a more general inequality in Section 5. This inequality, based on partitions of the set of diameters, provides better numerical estimates than the results of Section 3 for intermediate values of the deviation. Section 4 is devoted to the proofs of the results of Sections 2 and 3.

2 The case of equality of the diameters

In this section we assume that $\Delta_1 = \Delta_2 = \cdots = \Delta_n = 1$. Then (1.3) yields

$$\mathbb{P}(Z - \mathbb{E}(Z) \geq nx) \leq \exp\big(-n\varphi_0(x)\big) \text{ with } \varphi_0(x) = 2x^2.$$

For $x = 1$, $\varphi_0(1) = 2 < \infty$. Hence Inequality (1.3) does not imply (1.6). In Theorem 2.1 below, we give a better large deviations rate function for large values of x.

Theorem 2.1. *Let Z be defined by (1.1). For any positive t,*

$$n^{-1}\log \mathbb{E}\Big(\exp(tZ - t\mathbb{E}(Z))\Big) \le (t - \log t - 1) + t(e^t - 1)^{-1} + \log(1 - e^{-t}) := \ell(t). \quad (a)$$

Let the rate functions ψ_1 and ψ_2 be defined by

$$\psi_1(x) = 2x^2 + 4x^4/9,\ \psi_2(x) = (x^2 - 2x)\log(1 - x) \text{ for } x \in [0, 1[\quad (2.1)$$

and $\psi_2(x) = +\infty$ for $x \ge 1$. Let ℓ^* denote the Young transform of ℓ, which is defined by $\ell^*(x) = \sup_{t>0}(xt - \ell(t))$. For any positive x,

$$\ell^*(x) \ge \max(\psi_1(x), \psi_2(x)). \quad (b)$$

Consequently, by the usual Chernoff calculation, for any x in $[0, 1]$,

$$\mathbb{P}(Z - \mathbb{E}(Z) \ge nx) \le \exp\big(-n \max(\psi_1(x), \psi_2(x))\big) \le (1 - x)^{n(2x - x^2)}. \quad (c)$$

Remark 2.1. *In Section 4, it is shown that $\psi_2(x) \ge 2x^2 + x^4/6$. Consequently the second part of Theorem 2.1(c) also improves (1.3). Now, by (1.9) and Theorem 2.1(b),*

$$-\log(1 - x) + (1 - x)^2 \log(1 - x) \le \ell^*(x) \le -\log(1 - x) \text{ for any } x \in [0, 1[\,. \quad (2.2)$$

Hence $\lim_{x\uparrow 1}(\ell^*(x) + \log(1 - x)) = 0$, which gives the asymptotics of ℓ^* as $x \uparrow 1$.

Remark 2.2. *It comes from Lemma 4.3(a) in Section 4 that $\ell^*(x) = L_V(2x)$, where L_V is the information function defined in Equation (3) in Vajda [16]. Vajda proved that $L_V(2x) \ge -\log(1 - x) + \log(1 + x) - 2x/(1 + x)$. Theorem 2.1 in Gilardoni [5] gives the better lower bound $L_V(2x) \ge -\log(1 - x) - (1 - x)\log(1 + x) := L_2(2x)$. Using the concavity of the logarithm function, it can easily be proven that $\psi_2(x) > L_2(2x)$. Hence the lower bound $\ell^* \ge \psi_2$ improves Gilardoni's lower bound.*

Remark 2.3. *The expansion of ℓ at point 0 of order 5 is $\ell(t) = t^2/8 - t^4/576 + O(t^6)$. It follows that $\ell^*(x) = 2x^2 + (4/9)x^4 + O(x^6)$ as x tends to 0. Hence ψ_1 is the exact expansion of ℓ^* of order 5. The lower bound $\ell^* \ge \psi_1$ is based on Inequality (2) in Krafft [8]. Using Corollary 1.4 in Topsøe [15], one can obtain the slightly better lower bound $\ell^* \ge \psi_3$, where ψ_3 is defined by $\psi_3(x) = \psi_1(x) + (32/135)x^6 + (7072/42525)x^8$.*

3 The general case: moderate and large deviations

Here we assume that the diameters Δ_i do not satisfy $\Delta_1 = \Delta_2 = \cdots = \Delta_n$. Let us introduce the quantities below, which will be used to state our bounds:

$$D_n = \Delta_1 + \Delta_2 + \cdots + \Delta_n,\ A_n = D_n/n \text{ and } G_n = (\Delta_1\Delta_2\ldots\Delta_n)^{1/n}. \quad (3.1)$$

Then $G_n < A_n$. Our first result is an extension of Theorem 2.1, which preserves the variance factor σ_n^2. This result is suitable for moderate deviations. Here ℓ denotes the function already defined in Theorem 2.1(a) and ℓ^* is the Young transform of ℓ.

Theorem 3.1. *Let Z be defined by (1.1). For any positive t,*

$$\log \mathbb{E}\big(\exp(tZ - t\mathbb{E}(Z))\big) \le (D_n/\sigma_n)^2 \ell(\sigma_n^2 t/D_n). \quad (a)$$

Consequently, for any x in $[0, 1]$,

$$\mathbb{P}\big(Z - \mathbb{E}(Z) \ge D_n x\big) \le \exp\big(-(D_n/\sigma_n)^2 \ell^*(x)\big). \quad (b)$$

Contrary to the McDiarmid inequality, the upper bound in Theorem 3.1(b) converges to 0 as x tends to 1. Now, by the Cauchy-Schwarz inequality, $(D_n/\sigma_n)^2 \leq n$ in the general case. Moreover, in some cases $(D_n/\sigma_n)^2 = o(n)$ as n tends to ∞. In that case Theorem 3.2 below provides better results for large values of x. In order to state this result we need to introduce a second rate function. This is done in Proposition 3.1. below.

Proposition 3.1. *Let $\eta(t) = \ell(t) - (t - \log t - 1)$ and let $t_0 \simeq 1.5936$ be the solution of the equation $1 - e^{-t} = t/2$. Then η is concave and increasing on $]0, t_0]$ and decreasing on $[t_0, \infty[$. Furthermore $\eta'(t_0) = 0$ and consequently $\ell'(t_0) = 1 - t_0^{-1}$. Define the function η_c be defined by $\eta_c(t) = \eta(t)$ for t in $]0, t_0]$ and $\eta_c(t) = \eta(t_0)$ for $t \geq t_0$. Let ℓ_c be defined by $\ell_c(t) = (t - \log t - 1) + \eta_c(t)$. Then ℓ_c is a convex, continuously differentiable and increasing function on $\mathbb{R}^+$, and*

$$\ell_c^*(x) = \ell^*(x) \text{ for } x \leq \ell'(t_0) \text{ and } \ell_c^*(x) = -\eta(t_0) - \log(1-x) \text{ for } x \in [\ell'(t_0), 1[. \qquad (a)$$

The numerical value of $\eta(t_0)$ is $\eta(t_0) \simeq 0.17924$. Furthermore

$$\ell_c(t) \leq t^2/8 \text{ for any } t > 0 \text{ and } \ell_c^*(x) \geq 2x^2 \text{ for any } x > 0. \qquad (b)$$

We now state our second result.

Theorem 3.2. *Let Z be defined by (1.1). For any positive t,*

$$n^{-1} \log \mathbb{E}\big(\exp(tZ - t\mathbb{E}(Z))\big) \leq \log(A_n/G_n) + \ell_c(A_n t). \qquad (a)$$

Let t_0 be defined as in Proposition 3.1 and let $x_0 = \ell'(t_0) = 1 - t_0^{-1}$. The numerical value of x_0 is $x_0 \simeq 0.3725$. For x in $[0, x_0]$,

$$\mathbb{P}\big(Z - \mathbb{E}(Z) \geq D_n x\big) \leq \exp\big(n \log(A_n/G_n) - n\ell^*(x)\big) \qquad (b)$$

and, for x in $[x_0, 1]$,

$$\mathbb{P}\big(Z - \mathbb{E}(Z) \geq D_n x\big) \leq \exp\big(n \log(A_n/G_n) + n\eta(t_0)\big)\ (1-x)^n. \qquad (c)$$

Remark 3.1. *Since the maximum value of η_c is $\eta(t_0)$, $\ell_c(t) \leq t - \log t - 1 + \eta(t_0)$ for any positive t. Hence, for any x in $[0, 1[$,*

$$\ell_c^*(x) \geq -\eta(t_0) - \log(1-x) \geq \log(56/67) - \log(1-x). \qquad (3.2)$$

It follows that, for any positive y,

$$\mathbb{P}\big(Z - \mathbb{E}(Z) \geq D_n - (56/67)ny\big) \leq y^n/(\Delta_1\Delta_2 \ldots \Delta_n). \qquad (3.3)$$

The factor $1/(\Delta_1\Delta_2 \ldots \Delta_n)$ appearing in (3.3) cannot be removed, as shown by (1.9). For sake of completeness, we give here the proof of (1.9). let $\Delta_1 \geq \Delta_2 \geq \cdots \geq \Delta_n$ be positive reals and y be any positive real in $[0, \Delta_n]$. Let $b_1, \ldots, b_k, \ldots, b_n$ be independent random variables with Bernoulli laws $b(y/\Delta_k)$. Set $T_n = \Delta_1 b_1 + \Delta_2 b_2 + \cdots + \Delta_n b_n$. Then $\mathbb{P}(T_n - \mathbb{E}(T_n) \geq D_n - ny) = y^n/(\Delta_1\Delta_2 \ldots \Delta_n)$.

Example 3.1. *Take $n = 100$, $\Delta_1 = 49$ and $\Delta_k = 1$ for $k \geq 2$. Then $\sigma_n = 50$, $D_n = 148$ and $A_n = 1.48$. Let $p = \mathbb{P}(Z - \mathbb{E}(Z) \geq 75)$. The McDiarmid inequality (1.3) applied with $x = 3/2$ yields $p \leq e^{-9/2} \simeq 1.1\,10^{-2}$ and (1.4) yields $p \leq 7.7\,10^{-3}$. Theorem 3.1(b) together with the lower bound $\ell^* \geq \psi_3$ (see Remark 2.3) yields $p \leq 8.3\,10^{-3}$. Theorem 3.2(c) applied with $x = 75/148$ ($x > x_0$) yields $p \leq 2.7\,10^{-8}$.*

4 Proofs of the results of Sections 2 and 3

We start by proving an upper bound on the Laplace transform of Z which implies Theorem 2.1(a) in the case $\Delta_1 = \Delta_2 = \ldots = \Delta_n$.

Lemma 4.1. *Let ℓ be the function already defined in Theorem 2.1(a). Then, for any positive t, $\log \mathbb{E}\big(\exp(tZ - t\mathbb{E}(Z))\big) \leq \ell(\Delta_1 t) + \ell(\Delta_2 t) + \cdots + \ell(\Delta_n t) := L(t)$.*

Proof of Lemma 4.1. Let us briefly recall the martingale decomposition of Z. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$. Set $Z_k = \mathbb{E}(Z \mid \mathcal{F}_k)$. Then $Z = Z_n$ and $Z_0 = \mathbb{E}(Z)$. Furthermore $(Z_k)_k$ is a martingale sequence adapted to the above filtration. Now, set $Y_k = Z_k - Z_{k-1}$. Define the $\mathcal{F}_{k-1}$-measurable random variable W_{k-1} by

$$W_{k-1} = \mathbb{E}\Big(\inf_{x \in E_k} f(X_1, \ldots X_{k-1}, x, X_{k+1}, \ldots, X_n) \mid \mathcal{F}_{k-1}\Big) - Z_{k-1}. \tag{4.1}$$

By the Lipschitz condition on f, $W_{k-1} \leq Y_k \leq W_{k-1} + \Delta_k$ (see [10]). From this inequality and the convexity of the exponential function,

$$\Delta_k e^{tY_k} \leq (Y_k - W_{k-1}) e^{t(W_{k-1}+\Delta_k)} + (\Delta_k + W_{k-1} - Y_k) e^{tW_{k-1}}. \tag{4.2}$$

Hence, using the martingale property,

$$\Delta_k \mathbb{E}\big(e^{tY_k} \mid \mathcal{F}_{k-1}\big) \leq -W_{k-1} e^{t(W_{k-1}+\Delta_k)} + (\Delta_k + W_{k-1}) e^{tW_{k-1}}. \tag{4.3}$$

Set then

$$\gamma(r,t) = \log(1 + r(e^t - 1)) - tr \text{ and } r_{k-1} = -(W_{k-1}/\Delta_k). \tag{4.4}$$

Since (Z_k) is a martingale sequence, $\mathbb{E}(Y_k \mid \mathcal{F}_{k-1}) = 0$. Hence, from (4.2), $W_{k-1} \leq 0$ and $0 \leq W_{k-1} + \Delta_k$. Consequently r_{k-1} belongs to $[0,1]$. Now, starting from (4.3) and using the definitions (4.4), we get that $\log \mathbb{E}\big(e^{tY_k} \mid \mathcal{F}_{k-1}\big) \leq \gamma(r_{k-1}, \Delta_k t)$. Define now

$$\ell(t) = \sup_{r \in [0,1]} \gamma(r,t) = \sup_{r \in]0,1[} (\log(1 + r(e^t - 1)) - tr). \tag{4.5}$$

From the above inequalities $\log \mathbb{E}\big(e^{tY_k} \mid \mathcal{F}_{k-1}\big) \leq \ell(\Delta_k t)$ almost surely, which implies Lemma 4.1 for the function ℓ defined in (4.5). It remains to prove that ℓ is equal to the function already defined in Theorem 2.1(a). Now

$$\frac{\partial \gamma}{\partial r}(r,t) = \frac{e^t - t - 1 - rt(e^t - 1)}{1 + r(e^t - 1)},$$

and consequently the function $\gamma(r,t)$ has an unique maximum with respect to r in the interval $[0,1]$. This maximum is obtained for $r = r_t = (e^t - t - 1)/(t(e^t - 1))$, whence

$$\ell(t) = \log((e^t - 1)/t) - 1 + t/(e^t - 1) = (t - \log t - 1) + t(e^t - 1)^{-1} + \log(1 - e^{-t}).$$

□

We now prove Theorem 2.1(b). The first step is to compare the functions ψ_1 and ψ_2.

Lemma 4.2. *There exists a unique real x_0 in $[0.6670, 0.6675]$ such that $\psi_1(x) \geq \psi_2(x)$ for any $x \leq x_0$ and $\psi_1(x) < \psi_2(x)$ for $x > x_0$.*

Proof of Lemma 4.2. For any real x in $[0,1[$, $\psi_2(x) = 2x^2 + (x^4/6) + \sum_{k>4} a_k x^k$ with $a_k = (k-3)/(k^2 - 3k + 2)$. Define now f by $f(x) = x^{-4}(\psi_2(x) - \psi_1(x))$ for x in $[0,1[$. Then $f(x) = (-5/18) + \sum_{k>4} a_k x^{k-4}$, which implies that f is increasing on $[0,1]$. Lemma 4.2 follows then from the facts that $f(0.6670) < 0$ and $f(0.6675) > 0$. □

The second step is to prove that $\ell^*(x) \geq \psi_1(x)$ for any x in $[0,1]$.

Lemma 4.3. *For any r in $]0,1[$, let the function h_r be defined by*

$$h_r(x) = (r+x)\log(1+x/r) + (1-r-x)\log(1-x/(1-r)) \text{ for } x \text{ in } [0, 1-r[,$$

$h_r((1-r) = -\log r$ and $h_r(x) = +\infty$ for $x > 1-r$. Then

$$\ell^*(x) = \inf_{r\in]0,1[} h_r(x) \text{ for any } x > 0. \tag{a}$$

Consequently $\ell^*(x) \geq \psi_1(x)$ for any x in $[0,1]$.

Proof of Lemma 4.3. From (4.5), for any positive x,

$$\ell^*(x) = \sup_{t>0} \inf_{r\in[0,1-x]} (tx - \gamma(r,t)).$$

Now the function $(r,t) \to tx - \gamma(r,t)$ is convex in r and concave in t. Hence, the minimax theorem (see Corollary 3.3 in [14], for example) applies and yields

$$\ell^*(x) = \inf_{r\in[0,1-x]} \sup_{t>0} (tx - \gamma(r,t)). \tag{4.6}$$

Let then $\gamma_r^*(x) = \sup_{t>0}(tx - \gamma(r,t))$. As proved by Hoeffding [6],

$$\gamma_r^*(x) = (r+x)\log(1+x/r) + (1-r-x)\log(1-x/(1-r)) \text{ for } x \text{ in } [0, 1-r[\tag{4.7}$$

and $\gamma_r^*(1-r) = -\log r$. Moreover $\gamma_r^*(x) = +\infty$ for $x > 1-r$. Hence (4.6) and (4.7) yield Lemma 4.3(a). Now, by Inequality (2) in Krafft [8], $\gamma_r^*(x) \geq \psi_1(x)$ for any r in $]0, 1-x]$, which completes the proof of Lemma 4.3. □

To prove Theorem 2.1(b), it remains to prove that

$$\ell^*(x) \geq \psi_2(x) \text{ for any } x \geq x_0. \tag{4.8}$$

This inequality is a direct consequence of Lemma 4.4 below together with the fact that $-\log(1-x) \geq 1$ for $x \geq 2/3$.

Lemma 4.4. *$\ell^*(x) \geq -\log(1-x) - (1-x)^2$ for any $x \geq 2/3$.*

Proof of Lemma 4.4. Let η be the function which is defined in Proposition 3.1. Set $t_x = 1/(1-x)$. By definition of ℓ^*, $\ell^*(x) \geq xt_x - \ell(t_x) = \log t_x - \eta(t_x)$. Now, if $x \geq 2/3$, then $t_x \geq 3$. Consequently the proof of Lemma 4.4 will be complete if we prove that

$$t^2\eta(t) \leq 1 \text{ for any } t \geq 3. \tag{4.9}$$

By concavity of the logarithm,

$$t^2\eta(t) \leq t^2(t(e^t-1)^{-1} - e^{-t}) = (t^2 + (t^3 - t^2)e^t)/(e^{2t} - e^t).$$

Hence the inequality $t^2\eta(t) \leq 1$ holds true if $\delta(t) := (e^t + t^2 - t^3 - 1)e^t - t^2 \geq 0$ for $t \geq 3$. Let $\beta(t) := e^t + t^2 - t^3 - 1$. β is strictly convex on $[3, \infty[$ and has a unique minimum at $t_0 \simeq 3.1699$. Now $\beta(t_0) \simeq 1.00137 > 1$, whence $\delta(t) > e^t - t^2 > 0$ for $t \geq 3$. Hence (4.9) holds true, which implies Lemma 4.4. □

Proof of Theorem 2.1(b). Theorem 2.1(b) follows from Lemmas 4.2 and 4.3 together with (4.8). □

Proof of Theorem 3.1. The proof of Theorem 3.1 is based on the concavity property below.

Lemma 4.5. *The function ℓ' is concave on $\mathbb{R}^+$.*

Proof of Lemma 4.5. Set $v = 1/(e^t - 1)$. Then $\ell(t) = vt - \log v - \log t - 1$. Since $v' = -v(1+v)$,

$$\ell' = 1 + 2v - tv - tv^2 - (1/t),\ \ell'' = -3(v + v^2) + tv(1+v)(1+2v) + (1/t^2) \tag{4.10}$$

and

$$-\ell''' = (2/t^3) - 4v(1+v)(1+2v) + tv(1+v)(1+6v(1+v)). \tag{4.11}$$

Let $f(t) := -\ell'''(t)/(tv^2(1+v)^2)$. We prove that $f \geq 0$. Since $2v(1+v)(\cosh t - 1) = 1$, the function f can be decomposed as follows:

$$f(t) = f_1(t) + f_2(t) \text{ with } f_1(t) = 8t^{-4}(\cosh t - 1)^2 \text{ and } f_2(t) = 2\cosh t + 4 - 8(\sinh t/t).$$

Now f_1 and f_2 are analytic. First $f_2(t) = -2-(t^2/3)+\sum_{k\geq 2} a_k t^{2k}$, for positive coefficients a_k. More precisely $a_k = 2(2k-3)/(2k+1)!$. Consequently $f_2(t) \geq -2-(t^2/3)$. And second $2(\cosh t - 1) \geq t^2(1 + t^2/12)$, whence

$$f_1(t) + f_2(t) \geq 2(1 + t^2/12)^2 - 2 - t^2/3 = t^4/72 > 0.$$

Hence $f(t) > 0$ for any positive t, which ensures that ℓ' is concave. □

We now complete the proof of Theorem 3.1. According to Lemma 4.1, we have to prove that

$$L(t) := \ell(\Delta_1 t) + \ell(\Delta_2 t) + \cdots + \ell(\Delta_n t) \leq (D_n/\sigma_n)^2 \ell(\sigma_n^2 t/D_n). \tag{4.12}$$

Now

$$L(t) = \int_0^t L'(u)du = \int_0^t \Big(\Delta_1 \ell'(\Delta_1 u) \cdots + \Delta_n \ell'(\Delta_n u)\Big) du.$$

Next, by Lemma 4.5,

$$\Delta_1 \ell'(\Delta_1 u) \cdots + \Delta_n \ell'(\Delta_n u) \leq D_n\, \ell'(\sigma_n^2 u/D_n).$$

Hence

$$L(t) \leq D_n \int_0^t \ell'(\sigma_n^2 u/D_n) du = (D_n/\sigma_n)^2\, \ell(\sigma_n^2 t/D_n).$$

Hence (4.12) holds, which implies Theorem 3.1(a). Theorem 3.1(b) follows from the usual Chernoff calculation. □

Proof of Proposition 3.1. With the notations of the proof of Lemma 4.5,

$$\eta' = v(2 - (1+v)t) \text{ and } \eta'' = v(1+v)(t(1+2v) - 3).$$

Therefrom $\eta'(t) > 0$ if and only if $2 > te^t/(e^t - 1)$, which holds for $t > 0$ if and only $t < t_0$. Moreover $\eta'(t_0) = 0$ and $\eta'(t) < 0$ for $t < t_0$. Hence η is increasing on $[0, t_0]$ and decreasing on $[t_0, \infty[$. Now $\eta''(t) < 0$ if and only if $t(e^t + 1) < 3(e^t - 1)$. This condition holds if and only if $t < t_1$, where t_1 is the unique positive solution of the equation $t = 3\tanh(t/2)$. Consequently η is strictly concave on $]0, t_1]$ and convex on $[t_1, \infty[$. Since $t_1 \simeq 2.5757 > 2 > t_0$, it follows that η is strictly concave on $]0, t_0]$, increasing on $]0, t_0]$ and decreasing on $[t_0, \infty[$.

We now prove that ℓ_c is convex, increasing and continuously differentiable. Since $\ell_c(t) = \ell(t)$ for $t \le t_0$, the function ℓ_c is strictly convex, increasing and continuously differentiable on $[0, t_0]$. Next $\ell_c(t) = t - \log t - 1 + \eta(t_0)$ for $t \ge t_0$, which ensures that ℓ_c is continuous on $[t_0, \infty[$. Hence ℓ_c is continuous at the point t_0. Next the right derivative of ℓ_c at point t_0 is equal to $1 - t_0^{-1}$ and the left derivative is equal to $\ell'(t_0)$. Since $\eta'(t_0) = 0$, $\ell'(t_0) = 1 - t_0^{-1}$. Hence ℓ_c is differentiable at point t_0. If follows that ℓ_c is continuously differentiable on $[0, \infty[$. Now the function $t \to t - \log t - 1$ is strictly convex and increasing on $[1, \infty[$ and $t_0 > 1$. From the above facts, we get that ℓ_c is continuously differentiable, strictly convex and increasing on $[0, \infty[$. Furthermore ℓ_c' is a one to one continuous and increasing map from $[0, \infty[$ onto $[0, 1[$.

We now prove (a). From the definition of ℓ_c, $\ell_c'(t) = \ell'(t) \le \ell'(t_0)$ for $t \le t_0$ and $\ell'(t) = 1 - t^{-1} > \ell'(t_0)$ for $t > t_0$. Hence, for $x \le \ell'(t_0)$ the maximum of $xt - \ell_c(t)$ over all positive reals t is reached at $t_x = \ell'^{-1}(x) \le t_0$. Then $\ell_c^*(x) = xt_x - \ell(t_x) = \ell^*(x)$. For $x \ge \ell'(t_0)$, the maximum of $xt - \ell_c(t)$ over all positive reals t is reached at the unique point $t_x \ge t_0$ such that $1 - t_x^{-1} = x$. Then $t_x = 1/(1-x)$ and

$$\ell_c^*(x) = xt_x - \ell_c(t_x) = \frac{x}{1-x} - \frac{1}{1-x} + 1 - \log(1-x) - \eta(t_0) = -\log(1-x) - \eta(t_0),$$

which completes the proof of Proposition 3.1(a).

To prove (b), we note that, for any t in $[0, t_0]$, $\ell_c(t) = \ell(t) \le t^2/8$, since $\ell^*(x) \ge 2x^2$ for any positive x. Now, for any $t \ge t_0$, $\ell_c(t) = t - \log t - 1 + \eta(t_0)$. Deriving this equality, we get that $(t/4) - \ell_c'(t) = (t-2)^2/(4t) \ge 0$. Consequently $t^2/8 - \ell_c$ is nondecreasing on $[t_0, \infty[$, whence $t^2/8 - \ell_c(t) \ge (t_0^2/8) - \ell_c(t_0) \ge 0$ for $t \ge t_0$. Proposition 3.1(b) holds. □

Proof of Theorem 3.2. By definition, η_c is concave. Hence

$$\eta_c(\Delta_1 t) + \eta_c(\Delta_2 t) + \cdots + \eta_c(\Delta_n t) \le n\eta_c(A_n t).$$

Since $\ell \le \ell_c$,

$$\ell(\Delta_1 t) + \ell(\Delta_2 t) + \cdots + \ell(\Delta_n t) \le n(A_n t - 1 - \log t) - \log(\Delta_1 \ldots \Delta_n) + n\eta_c(A_n t).$$

It follows that

$$\ell(\Delta_1 t) + \ell(\Delta_2 t) + \cdots + \ell(\Delta_n t) \le n \log(A_n/G_n) + n\ell_c(A_n t), \tag{4.13}$$

which, together with Lemma 4.1, implies Theorem 3.2(a). (b) and (c) follow from the usual Chernoff calculation together with Proposition 3.1(a). □

5 An inequality involving partitions

In this section we are interested in intermediate values of the deviation x. In the sketchy Example 3.1, it appears that the McDiarmid diameter σ_n defined in (1.2) is too big for intermediates values of the deviation. In this section, we introduce a method which minimizes the effect of variations of the values of the individual diameters $\Delta_1, \Delta_2, \ldots, \Delta_n$.

Definition 5.1. *A set $\mathcal{P}$ of subsets of $\{1, 2, \ldots, n\}$ is called partition of $\{1, 2, \ldots, n\}$ iff: (i) for any I in $\mathcal{P}$, I is nonempty; (ii) for any I and for any J in $\mathcal{P}$, either $I \cap J = \emptyset$ or $I = J$; (iii) $\bigcup_{I \in \mathcal{P}} I = \{1, 2, \ldots, n\}$.*

We now define the diameter $\sigma(\mathcal{P})$ and the entropy $H(\mathcal{P})$ of a partition $\mathcal{P}$ as follows. Let $|J|$ denote the cardinality of a finite set J. We set

$$D_J = \sum_{j \in J} \Delta_j, \; A_J = |J|^{-1} D_J \quad \text{and} \quad \sigma^2(\mathcal{P}) = \sum_{J \in \mathcal{P}} |J| \, A_J^2. \tag{5.1}$$

Let the geometric means G_J and the entropy be defined by

$$G_J = \Big(\prod_{j\in J} \Delta_j\Big)^{1/|J|} \quad \text{and} \quad H(\mathcal{P}) = \sum_{J\in\mathcal{P}} |J| \log(A_J/G_J). \tag{5.2}$$

The so defined quantities satisfy $\sigma^2(\mathcal{P}) \le \sigma_n^2$ and $H(\mathcal{P}) \ge 0$. Furthermore $H(\mathcal{P}) = 0$ if and only if $\sigma^2(\mathcal{P}) = \sigma_n^2$.

Theorem 5.1. *Let the convex and differentiable function ℓ_0 be defined by*

$$\ell_0(t) = t^2/8 \ \text{ for } \ t \in [0,2] \ \text{and} \ \ \ell_0(t) = t - \log t - (3/2) + \log 2 \ \text{ for } \ t \ge 2.$$

Let Z be defined by (1.1). For any positive t and any partition $\mathcal{P}$ of $\{1,2,\ldots,n\}$,

$$\log \mathbb{E}\big(\exp(tZ - t\mathbb{E}(Z))\big) \le H(\mathcal{P}) + (D_n^2/\sigma^2(\mathcal{P}))\ell_0(\sigma^2(\mathcal{P})t/D_n). \tag{a}$$

Consequently, for any x in $[0,1]$,

$$\mathbb{P}\big(Z - \mathbb{E}(Z) \ge D_n x\big) \le \exp\big(H(\mathcal{P}) - (D_n^2/\sigma^2(\mathcal{P}))\ell_0^*(x)\big) \tag{b}$$

and, for any positive y,

$$\mathbb{P}\big(Z - \mathbb{E}(Z) \ge D_n \ell_0^{*-1}\big(D_n^{-2} \min_{\mathcal{P}} \sigma^2(\mathcal{P})(H(\mathcal{P}) + y)\big)\big) \le e^{-y}. \tag{c}$$

Remark 5.1. *In Theorem 5.1(c), for small values of y, the optimal partition has a small entropy and a large diameter, while, for large values of y, the optimal partition has a small diameter and a large entropy.*

Remark 5.2. *The functions ℓ_0^* and ℓ_0^{*-1} are explicit. More precisely*

$$\ell_0^*(x) = 2x^2 \ \text{ for } \ x \in [0,1/2] \ \text{and} \ \ \ell_0^*(x) = -\log(1-x) + (1/2) - \log 2 \ \text{ for } \ x \in [1/2,1],$$

$$\ell_0^{*-1}(y) = \sqrt{y/2} \ \text{ for } \ y \in [0,1/2] \ \text{and} \ \ \ell_0^{*-1}(y) = 1 - (\sqrt{e}/2)e^{-y} \ \text{ for } \ y \ge 1/2.$$

Example 3.1 (continued). *Let Q denote the quantile function of $Z - \mathbb{E}(Z)$. For $p = e^{-9/2}$, Theorem 5.1(c) applied with $\mathcal{P} = \{[1,13],[14,100]\}$ (the optimal partition) yields $Q(p) \le 62.18$. The McDiarmid inequality (1.3) yields $Q(p) \le 75$, and Theorem 3.2 yields $Q(p) \le 64.93$. For small values of p, the optimal partition is $\mathcal{P} = \{[1,100]\}$. In this case Theorem 5.1 is less efficient than Theorem 3.2, since $\ell_0^{*-1}(y) > \ell_c^{*-1}(y)$. For example, let $q = \mathbb{P}(Z - \mathbb{E}(Z) \ge 75)$. Theorem 5.1(b) yields $q \le 3.0\,10^{-7}$ instead of $q \le 2.7\,10^{-8}$ with Theorem 3.2. Recall that (1.3) yields $q \le 1.1\,10^{-2}$.*

Proof of Theorem 5.1. By Lemma 4.1 together with (4.13),

$$\log \mathbb{E}\big(\exp(tZ - t\mathbb{E}(Z))\big) \le \sum_{J\in\mathcal{P}} \sum_{j\in J} \ell(\Delta_j t) \le H(\mathcal{P}) + \sum_{J\in\mathcal{P}} |J|\, \ell_c(A_J t).$$

Now $\ell_c(t) \le \min(t^2/8, \eta(t_0) + t - \log t - 1) \le \ell_0(t)$ for any positive t. Hence

$$\log \mathbb{E}\big(\exp(tZ - t\mathbb{E}(Z))\big) \le H(\mathcal{P}) + \sum_{J\in\mathcal{P}} |J|\, \ell_0(A_J t). \tag{5.3}$$

To complete the proof of Theorem 5.1, we proceed exactly as in the proof of Theorem 3.1: since $\ell_0(0) = 0$,

$$\sum_{J\in\mathcal{P}} |J|\, \ell_0(A_J t) = \int_0^t \Big(\sum_{J\in\mathcal{P}} D_J \ell_0'(A_J u)\Big) du. \tag{5.4}$$

Now $\ell_0'(t) = t/4$ for $t \le 2$ and $\ell_0'(t) = 1 - (1/t)$ for $t \ge 2$, which ensures that ℓ_0' is continuous and increasing. $\ell_0''(t) = 1/4$ for $t < 2$ and $\ell''(0) = t^{-2}$ for $t > 2$, which ensures that $\lim_{t\downarrow 2} \ell_0''(t) = 1/4$. Hence, by L'Hospital's rule, ℓ_0' is differentiable at point 2, and $\ell_0''(2) = 4$. Consequently ℓ_0'' is continuous and nonincreasing, which ensures that ℓ_0' is concave. It follows that

$$\sum_{J\in\mathcal{P}} D_J \ell_0'(A_J u) \le D_n \ell_0'(\sigma^2(\mathcal{P})u/D_n).$$

Integrating this inequality, we then get that

$$\int_0^t \Big(\sum_{J\in\mathcal{P}} D_J \ell_0'(A_J u)\Big) du \le (D_n^2/\sigma^2(\mathcal{P}))\ell_0(\sigma^2(\mathcal{P})t/D_n), \tag{5.5}$$

which, together with (5.3) and (5.4), implies Theorem 5.1(a). Theorem 5.1(b) follows from the usual Chernoff calculation and Theorem 5.1(c) is an immediate consequence of Theorem 5.1(b). □

References

[1] Bentkus, V. A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand. *Lithuanian Math. J.* **42**, no. 3 (2002), 262–269. MR-1947624

[2] Bentkus, Vidmantas. On Hoeffding's inequalities. *Ann. Probab.* **32**, no. 2 (2004), 1650–1673. MR-2060313

[3] Bentkus, V. On measure concentration for separately Lipschitz functions in product spaces. *Israel J. Math.* **158**, (2007) 1–17. MR-2342455

[4] Devroye, L., Lugosi, G. Combinatorial methods in density estimation. Springer Series in Statistics. Springer-Verlag, New York, 2001. MR-1843146

[5] Gilardoni, G. An improvement on Vajda's inequality. In and out of equilibrium. 2, 299–304, Progr. Probab., 60, Birkhäuser, Basel, 2008. MR-2477387

[6] Hoeffding, W. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, (1963), 13–30. MR-0144363

[7] Kearns, M. J. and Saul, L. K. Large deviation methods for approximate probabilistic inference. In UAI, 1998.

[8] Krafft, O. A note on exponential bounds for binomial probabilities. *Annals of the Institute of Statistical Mathematics* **21**, (1969), 219–220.

[9] McDiarmid, C. On the method of bounded differences, dans Surveys of combinatorics, ed J. Siemons, London Mathematical Society lecture notes series 141, 1989. MR-1036755

[10] McDiarmid, C. "Concentration," Probabilistic methods for algorithmic discrete mathematics. 195–248. M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed (Editors), Springer, New York, 1998. MR-1678578

[11] Owhadi, H., Sullivan, T. J., McKerns, M., Ortiz, M. and Scovel, C. Optimal Uncertainty Quantification. arXiv: 1009.0679v3. Math. PR. 2012.

[12] Pinelis, I. On normal domination of (super)martingales. *Electron. J. Probab.* **11**, (2006), 1049–1070. MR-2268536

[13] Pinelis, Iosif. Binomial upper bounds on generalized moments and tail probabilities of (super)martingales with differences bounded from above. High dimensional probability, 33–52, IMS Lecture Notes Monogr. Ser., 51, Inst. Math. Statist., Beachwood, OH, 2006. MR-2387759

[14] Sion, Maurice On general minimax theorems. *Pacific J. Math.* **8**, (1958), 171–176. MR-0097026

[15] Topsøe, F. Bounds for entropy and divergence for distributions over a two-element set. *JIPAM. J. Inequal. Pure Appl. Math.* **2**, no. 2, (2001), Article 25. MR-1873865

[16] Vajda, I. Note on discrimination information and variation. *IEEE Trans. Information Theory* **16**, (1970), 771–773. MR-0275575

Acknowledgments. I would like to thank the referee for pointing out a gap in the proof of Lemma 4.3 and filling this gap. Thanks are also due to the referee for a careful reading of the initial version.