

On McDiarmid's concentration inequality Emmanuel Rio

▶ To cite this version:

Emmanuel Rio. On McDiarmid's concentration inequality. 2013. hal-00798275

HAL Id: hal-00798275 https://hal.science/hal-00798275

Preprint submitted on 8 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On McDiarmid's concentration inequality

Emmanuel RIO, UMR 8100 CNRS, Laboratoire de mathématiques de Versailles, Bât. Fermat, 45 Av. des Etats Unis, 78035 VERSAILLES Cedex. France.

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.

Mathematics subject classification (2000) : 60E15.

Keywords and phrases : McDiarmid inequality, Concentration inequality, Hoeffding inequality.

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

$$|f(x_1, \dots, x_n) - f(y_1, \dots, y_n)| \le d_1(x_1, y_1) + \dots + d_n(x_n, y_n)$$

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). Let the McDiarmid diameter σ_n be defined by

(1.1)
$$\sigma_n^2 = \Delta_1^2 + \Delta_2^2 + \dots + \Delta_n^2.$$

McDiarmid (1989, 1998) proved that, for any positive x,

(1.2)
$$\mathbb{P}(Z - \mathbb{E}(Z) \ge \sigma_n x) \le \exp(-2x^2).$$

This inequality is an extension of Theorem 2 in Hoeffding (1963). We refer to Devroye and Lugosi (2001), Chapter 2, for more about concentration inequalities. Later Bentkus (2007, paper submitted on August 17, 2001) and Pinelis (2006) replaced the upper bound in (1.2) by a Gaussian tail function. They proved that

(1.3)
$$\mathbb{P}(Z - \mathbb{E}(Z) \ge \sigma_n x) \le c \mathbb{P}(Y \ge 2x), \text{ with } Y \stackrel{D}{=} N(0, 1).$$

The best known constant in (1.3) is c = 5.70, due to Pinelis (2006).

We now comment these results. 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

(1.4)
$$m \leq Z \leq M$$
 and $M - m \leq \Delta_1 + \Delta_2 + \dots + \Delta_n := D_n$.

From (1.4) it follows that

(1.5)
$$\mathbb{P}(Z - \mathbb{E}(Z) \ge D_n) = \mathbb{P}(Z = M \text{ and } \mathbb{E}(Z) = m) = 0.$$

Now (1.5) cannot be deduced from either (1.2) or (1.3). Hence it seems clear that the rate function $2x^2$ in the Mc-Diarmid inequality (1.2) is suboptimal for large values of x. The aim of this paper is to improve the rate function in (1.2). 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],

(1.6)
$$\mathbb{P}(Z - \mathbb{E}(Z) \ge n(1-x)) \le x^{n(1-x^2)}.$$

This inequality implies (1.2) and yields (1.5). 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. Let us recall the known lower bounds for large values of the deviations. Take $E_i = [0, \Delta_i]$ and set $\Delta = (\Delta_1, \Delta_2, \ldots, \Delta_n)$. Let $P_{McD}(z, \Delta)$ denote 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. (2012),

(1.7)
$$P_{McD}(D_n - nx, \Delta) \ge x^n / (\Delta_1 \Delta_2 \dots \Delta_n) \text{ for any } x \le \min(\Delta_1, \Delta_2, \dots, \Delta_n).$$

In Theorem 3.2 of Section 3, we prove the converse inequality, with $D_n - (56/67)nx$ instead of $D_n - nx$. For small values of the deviation we obtain in Theorem 3.1 the following extension of (1.6): for any x in [0, 1],

(1.8)
$$\mathbb{P}(Z - \mathbb{E}(Z) \ge D_n(1-x)) \le x^{(1-x^2)D_n^2/\sigma_n^2}.$$

This extension is also suitable for large values of the deviation when $\sigma_n^2 \sim D_n^2/n$. However, as shown by the converse inequality (1.7), (1.8) has be improved when $\sigma_n \gg n^{-1/2}D_n$. We

give a first result in this direction in Section 3. The results of Sections 2 and 3 are proved in Section 4. To conclude this paper, 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.

2. The case $\Delta_1 = \Delta_2 = \cdots = \Delta_n$

In this section we assume that $\Delta_1 = \Delta_2 = \cdots = \Delta_n = 1$. Then (1.2) yields

(2.1)
$$\mathbb{P}(Z - \mathbb{E}(Z) \ge nx) \le \exp(-n\varphi(x)) \text{ with } \varphi(x) = \varphi_0(x) = 2x^2.$$

From Theorem A in Rio (2001), (2.1) still holds true with the greater function $\varphi(x) = \varphi_1(x) = 2((1+x)\log(1+x) + (1-x)\log(1-x))$ (here log is the Neper logarithm). Nevertheless $\varphi_1(1) = 4\log 2 < \infty$, so that this result still is suboptimal for x = 1. In Theorem 2.1 below, we give a better large deviations rate function.

Theorem 2.1. For any positive t,

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

Let then $\psi_1(x) = 2x^2 + (4x^4/9)$ and $\psi_2(x) = (x^2 - 2x)\log(1 - x)$, with the convention that $\psi_2(1) = +\infty$. For any x in [0, 1],

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

Remark 2.1. Let ℓ^* denotes the Young transform of ℓ . (b) follows from the fact that

(2.2)
$$\ell^*(x) \ge \max(\psi_1(x), \psi_2(x)).$$

Note that $\psi_2(x) > 2x^2$. Consequently, the two upper bounds in Theorem 1.1(b) improve McDiarmid's result. Furthermore it can be proven that $\max(\psi_1(x), \psi_2(x)) > \varphi_1(x)$ for any x in [0, 1]. Consequently (b) also improves Theorem A in Rio (2001).

By (1.7) and (2.2), $-\log(1-x) + (1-x)^2 \log(1-x) \le \ell^*(x) \le -\log(1-x)$. It follows that $\lim_{x\uparrow 1} \ell^*(x) + \log(1-x) = 0$, which gives the asymptotics of l^* as $x \uparrow 1$.

Remark 2.2. The expansion of ℓ at point 0 is $\ell(t) = \frac{1}{8}t^2 - \frac{1}{576}t^4 + O(t^5)$. It follows that $\ell^*(x) = 2x^2 + (4/9)x^4 + O(x^5)$ as x tends to 0. Hence ψ_1 is the exact expansion of ℓ^* at order 4.

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:

(3.1)
$$A_n = (D_n/n), \ B_n = n^{-1/2} \sigma_n \text{ and } G_n = (\Delta_1 \Delta_2 \dots \Delta_n)^{1/n}.$$

Then $G_n < A_n < B_n$. Our first result is an extension of Theorem 3.1 below gives 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. For any positive t,

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

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

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

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. However, 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 $[t_0, \infty[$. Define η_c by $\eta_c(t) = \eta(t)$ for $t \le t_0$ and $\eta_c(t) = \eta(t_0)$ for $t \ge t_0$. 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

(a)
$$\ell_c^*(x) = \ell^*(x)$$
 for $x \le x_0 = 1 - t_0^{-1}$ and $\ell_c^*(x) = -\eta(t_0) - \log(1 - x)$ for $x \ge x_0$.

The numerical value of $\eta(t_0)$ is $\eta(t_0) \simeq 0.17924$ and $x_0 \simeq 0.3725$. Furthermore

(b)
$$\ell_c(t) \le t^2/8$$
 for any $t > 0$ and $\ell_c^*(x) \ge 2x^2$ for any $x > 0$.

We now state our second result

Theorem 3.2. For any positive t,

(a)
$$n^{-1}\log \mathbb{E}\left(\exp(tZ - t\mathbb{E}(Z))\right) \le \log(A_n/G_n) + \ell_c(A_nt).$$

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

(b)
$$\mathbb{P}(Z - \mathbb{E}(Z) \ge D_n x) \le \exp(n \log(A_n/G_n) - n\ell_c^*(x)).$$

Remark 3.1. Since the maximum value of η 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],

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

It follows that, for any positive y,

(3.3)
$$\mathbb{P}(Z - \mathbb{E}(Z) \ge D_n - (56/67)ny) \le y^n / (\Delta_1 \Delta_2 \dots \Delta_n).$$

The factor $1/(\Delta_1 \Delta_2 \dots \Delta_n)$ appearing in (3.3) cannot be removed, as shown by (1.7), For sake of completeness, we give here the proof of (1.7). let $\Delta_1 \geq \Delta_2 \geq \dots \geq \Delta_n$ be positive reals and y be any positive real in $[0, \Delta_n]$. Let b_1, b_2, \dots, b_n be independent random variables such that b_k has the Bernoulli law $b(y/\Delta_k)$. Set $T_n = \Delta_1 b_1 + \Delta_2 b_2 + \dots + \Delta_n b_n$. Then $\mathbb{P}(T_n - \mathbb{E}(T_n) \geq D_n - ny) = y^n/(\Delta_1 \Delta_2 \dots \Delta_n)$.

Example 3.1. Take n = 100, $\Delta_1 = 49$ and $\Delta_k = 1$ for $k \ge 2$. Then $\sigma_n = 50$, $D_n = 148$ and $A_n = 1.48$. Let $p = \mathbb{P}(Z - \mathbb{E}(Z) \ge 75)$. The McDiarmid inequality (1.2) applied with x = 3/2 yields $p \le e^{-9/2} \simeq 1.1 \ 10^{-2}$ and (1.3) yields $p \le 7.7 \ 10^{-3}$. Theorem 3.1(b) yields $p \le 8.6 \ 10^{-3}$ and Theorem 3.2(b) yields $p \le 2.7 \ 10^{-8}$.

To conclude this section, we give an inequality, which is a byproduct of the proofs of Theorems 3.1 and 3.2. In some cases this inequality provides better estimates.

Theorem 3.3. For k in [1, n], let the quantities σ_k , D_k , A_k and G_k be defined as in (1.1), (1.4) and (3.1). Set $D_0 = 0$ and $\sigma_0 = 0$. Then, for any k in [0, n] and any positive u,

$$\mathbb{P}\Big(Z - \mathbb{E}(Z) \ge (D_n - D_k)\ell^{*-1}\Big(\frac{(\sigma_n^2 - \sigma_k^2)u}{(D_n - D_k)^2}\Big) + D_k\,\ell_c^{*-1}\Big(\log\Big(\frac{A_k}{G_k}\Big) + \frac{u}{k}\Big)\Big) \le e^{-u},$$

where ℓ^{*-1} and ℓ_c^{*-1} denote the inverse functions of ℓ^* and ℓ_c^* . Furthemore, for k = 1, the function ℓ_c^{*-1} may be replaced by ℓ^{*-1} in the above inequality.

Remark 3.2. For k = n, Theorem 3.3 is just an other formulation of Theorem 3.2. For k = 0, Theorem 3.3 is an other formulation of Theorem 3.1.

We now give a "ready to be used" inequality". From (2.2), $\ell^{*-1}(y) \leq \psi_1^{-1}(y)$. Moreover

(3.4)
$$\psi_1^{-1}(y) = \sqrt{y} \left(1 + \sqrt{1 + (4/9)y} \right)^{-1/2}.$$

Now, from (3.2), $\ell_c^{*-1}(y) \le 1 - \frac{56}{67}e^{-y}$. Hence Theorem 3.3 yields

(3.5)
$$\mathbb{P}\Big(Z - \mathbb{E}(Z) \ge (D_n - D_k)\psi_1^{-1}\Big(\frac{(\sigma_n^2 - \sigma_k^2)u}{(D_n - D_k)^2}\Big) + D_k - \frac{56}{67}kG_ke^{-u/k}\Big) \le e^{-u}.$$

Example 3.1 (continued). Let $p = e^{-9/2}$. Denote by Q the quantile function of $Z - \mathbb{E}(Z)$, which is the inverse of the tail function. The McDiarmid inequality yields $Q(p) \leq 75$. By Theorem 3.3 applied with k = 1 (the optimal choice), $Q(p) \leq 63.35$. Next $Q(p) \leq 73.05$ by Theorem 3.1 and $Q(p) \leq 64.93$ by Theorem 3.2.

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 $\mathbb{E}\left(\exp(tZ - t\mathbb{E}(Z))\right) \leq \ell(\Delta_1 t) + \ell(\Delta_2 t) + \cdots + \ell(\Delta_n t) := L(t)$.

Proof of Lemma 4.1. Let us briefly recall the martingale decomposition of Z. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_k = \sigma(X_1, \ldots, X_n)$. 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

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

By (1.1),

$$(4.2) W_{k-1} \le Y_k \le W_{k-1} + \Delta_k.$$

From this inequality and the convexity of the exponential function,

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

Hence, using the martingale property

(4.3)
$$\Delta_k \mathbb{E} \left(e^{tY_k} \mid \mathcal{F}_{k-1} \right) \le -W_{k-1} e^{t(W_{k-1} + \Delta_k)} + (\Delta_k + W_{k-1}) e^{tW_{k-1}}$$

Set then

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

Since (Z_k) is a martingale sequence, $\mathbb{E}(Y_k | \mathcal{F}_{k-1}) = 0$. 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]. Furthermore, by (4.3),

$$\log \mathbb{E}\left(e^{tY_{k}} \mid \mathcal{F}_{k-1}\right) \leq \gamma(r_{k-1}, \Delta_{k}t).$$

Define now

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

From the above inequality

(4.6)
$$\log \mathbb{E}\left(e^{tY_k} \mid \mathcal{F}_{k-1}\right) \le \ell(\Delta_k t) \text{ almost surely },$$

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

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

We now prove (2.2) , and therefore Theorem 2.1(b). The first step is to compare the functions ψ_1 and ψ_2 .

Lemma 4.2. There exists an unique real x_0 in [0.6670, 0.6675] such that $\psi_1(x) \ge \psi(x)$ for any $x \le x_0$ and $\psi_1(x) < \psi_2(x)$ for $x > x_0$.

Proof of Lemma 4.2. For any x < 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 $f(x) = x^{-4}(\psi(x) - \varphi(x))$. Then

$$f(x) := x^{-4}(\psi(x) - \varphi(x)) = (-5/18) + \sum_{k>4} a_k x^{k-4},$$

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

Lemma 4.3. $\ell^*(x) \ge \psi_1(x)$ for any x in [0, 1].

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

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

Let the function h be defined by $h(x) = (1+x)\log(1+x) - x$ for x > -1, h(-1) = 1 and $h(x) = +\infty$ for x < -1. Since h(0) = h'(0) = 0 and h''(x) = 1/(1+x), the function h is convex and nonnegative. Next

(4.7)
$$\gamma_r^*(x) = rh(x/r) + (1-r)h(-x/(1-r)) := h_r(x).$$

Consequently $\gamma_r^*(x) = +\infty$ for r > 1 - x. Furthermore the above function is continuous with respect to r for r in [0, 1 - x], so that we may take the infimum over]0, 1 - x[in (4.6). Our way to prove Lemma 4.3 is consider h_r as a function of $u_x = 2(r + x) - 1$. Clearly

$$h_r''(x) = (x+r)^{-1} + (1-x-r)^{-1} = 4/(1-u_x^2) \ge 4(1+u_x^2).$$

Now $h_r(0) = h'_r(0) = 0$. Consequently, by the Taylor integral formula,

(4.8)
$$h_r(x) = \int_0^x (x-y)h''_r(y)dy \ge 4\int_0^x (1+u_y^2)(x-y)dy.$$

Some elementary calculations show that

(4.9)
$$4\int_0^x (1+4u_y^2)(x-y)dy = 2x^2 + \frac{4}{9}x^4 + 2x^2(2r-1+(2x/3))^2.$$

Both (4.6), (4.7), (4.8) and (4.9) then imply Lemma 4.3. \blacksquare

To complete the proof of Theorem 2.1(b), it remains to prove that $\ell^*(x) \ge \psi_1(x)$ for $x \ge x_0$.

Lemma 4.4. $\ell^*(x) \ge \psi_1(x)$ for any $x \ge 2/3$.

Proof of Lemma 4.4. Let $t_x = 1/(1-x)$. By definition, $\ell^*(x) \ge xt_x - \ell(t_x)$. Define the function η by

(4.10)
$$\eta(t) = \ell(t) - (t - \log t - 1) = t(e^t - 1)^{-1} + \log(1 - e^{-t}).$$

With this definition, $xt_x - \ell(t_x) = \log t_x - \eta(t_x)$. Now, $t_x \ge 3$ for any $x \ge (2/3)$. Since $\psi_2(x) = \log t_x - t_x^{-2} \log t_x$, the proof of Lemma 4.3 will be complete if we prove that

(4.11)
$$t^2 \eta(t) \le \log t \text{ for any } t \ge 3.$$

Since $\log 3 \ge 1$, it is enough to prove that $t^2 \eta(t) \le 1$. Now, by concavity of the logarithm,

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

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 an unique infimum 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 $t^2\eta(t) \leq 1$ for $t \geq 3$, which implies (4.11). Consequently Lemma 4.3 holds true. Inequality (2.2) follows then from both Lemmas 4.2, 4.3 and 4.4. Theorem 2.1(b) Follows.

Proof of Theorem 3.1. The proof of Theorem 3.1 is based on the convexity 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),

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

and

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

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

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

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

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

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

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

(4.12)
$$L(t) := \ell(\Delta_1 t) + \ell(\Delta_2 t) + \dots + \ell(\Delta_n t) \le (D_n/\sigma_n)^2 \ell(\sigma_n^2 t/D_n).$$

Now

$$L(t) = \int_0^t L'(u) du = \int_0^t \left(\Delta_1 \ell'(\Delta_1 u) \dots + \Delta_n \ell'(\Delta_n u) \right) du.$$

Next, by Lemma 4.5,

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

Hence

$$L(t) \le \Delta \int_0^t \ell'(\sigma^2 u/\Delta) du = (D_n/\sigma_n)^2 \,\ell(\sigma_n^2 t/D_n)$$

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

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

$$\eta' = v(2 - (1 + v)t)$$
 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 if and only $t < t_0$. Now $\eta''(t) < 0$ if and only if $t(e^t + 1) < 3(e^t - 1)$. This condition holds if and only if $t \le t_1$, where t_1 is the unique positive solution of the equation $t = 3 \tanh(t/2)$. Since $t_1 \simeq 2.5757 > 2 > t_0$, the first part of Proposition 3.1 holds true. Now, by definition ℓ_c is continuous and convex and continuously differentiable on the two intervals $[0, t_0]$ and $[t_0, \infty]$. Since $\eta'(t_0) = 0$, the left derivative and the right derivative of ℓ_c at point t_0 are equal. Hence ℓ_c is convex and continuously differentiable on \mathbb{R}^+ .

The proof of (a), being immediate, is omitted. To prove (b), we note that, for any $t \leq t_0$, $\ell_c(t) = \ell(t) \leq t^2/8$, since $\ell^*(x) \geq 2x^2$ for any positive x. Now $\ell_c(t) = t - \log t - 1 + \eta(t_0)$ for $t \geq t_0$, and consequently $(t/4) - \ell'_c(t) = (t-2)^2/(4t)$. Hence $t^2/8 - \ell_c$ is nondecreasing on $[t_0, \infty[$, whence $t^2/8 - \ell_c(t) \geq (t_0^2/8) - \ell_c(t_0) > 0$. Hence (b) holds, wich completes the proof of Proposition 3.1.

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

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

Since $\ell \leq \ell_c$,

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

It follows that

(4.13)
$$\ell(\Delta_1 t) + \ell(\Delta_2 t) + \dots + \ell(\Delta_n t) \le n \log(A_n/G_n) + n\ell_c(A_n t).$$

which, together with Lemma 4.1, implies Theorem 3.2(a). Theorem 3.2(b) follows from the usual Chernoff calculation. \blacksquare

Proof of Theorem 3.3. Here we use the martingale decomposition. With the same notations as in Lemma 4.1, let $S = Z_k - Z_0$ and $T = Z_n - Z_k$. Then

$$Z - \operatorname{I\!E}(Z) = Z_n - Z_0 = S + T.$$

The random variables Z and T have a finite Laplace transform, and, from their martingale decomposition together with (4.6),

(4.14)
$$L_S(t) := \log \mathbb{E}(\exp(tS)) \le \ell(\Delta_1 t) + \ell(\Delta_2 t) + \dots + \ell(\Delta_k t),$$

(4.15)
$$L_T(t) := \log \mathbb{E}(\exp(tT)) \le \ell(\Delta_{k+1}t) + \ell(\Delta_{k+2}t) + \dots + \ell(\Delta_n t).$$

Let L_S^* and L_T^* denote the Young transforms of L_S and L_T respectively. By Lemma 2.1 in Rio (1994), $L_{S+T}^{*-1} \leq L_S^{*-1} + L_T^{*-1}$. Hence

(4.16)
$$\mathbb{P}(Z - \mathbb{E}(Z) > L_T^{*-1}(u) + L_S^{*-1}(u)) \le \exp(-u).$$

Next, by (4.15) together with (4.12),

$$L_T(t) \le (\sigma_n^2 - \sigma_k^2)^{-1} (D_n - D_k)^2 \ell ((\sigma_n^2 - \sigma_k^2) t / (D_n - D_k)),$$

which ensures that

(4.17)
$$L_T^{*-1}(u) \le (D_n - D_k)\ell^{*-1} \left((\sigma_n^2 - \sigma_k^2) u / (D_n - D_k)^2 \right).$$

Now, for k = 1, $L_S(t) \leq \ell(\Delta_1 t)$, whence $L_S^{*-1}(u) \leq \Delta_1 \ell^{*-1}(u)$. Theorem 3.3 in the case k = 1 follows from both this inequality, (4.16), (4.17) and the strict monotonicity of ℓ^{*-1} on \mathbb{R}^+ (which allows to replace > by \geq in (4.16)). For $k \geq 2$, from (4.14) and (4.13), $L_S(t) \leq k \log(A_k/G_k) + k\ell_c(A_k t)$. Therefore

(4.18)
$$L_S^{*-1}(u) \le D_k \, \ell_c^{*-1} \big(\log(A_k/G_k) + (u/k) \big).$$

Theorem 3.3 follows then from both (4.16), (4.17), (4.18) and the strict monotonicity of ℓ^{*-1} and ℓ_c^{*-1} on \mathbb{R}^+ .

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 McDiameter diameter 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 family \mathcal{P} of subsets of $\{1, 2, ..., n\}$ is called partition of $\{1, 2, ..., n\}$ iff: (i) for any I in \mathcal{P} , $I \neq \emptyset$; (ii) for any I and any J in \mathcal{P} , either $I \cap J = \emptyset$ or I = J.

We now define the McDiarmid diameter $\sigma^2(\mathcal{P})$ and the entropy $H(\mathcal{P})$ of a partition \mathcal{P} as follows. Let |J| denote the cardinality of the finite set J. We set

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

Let the geometric means G_J and the entropy be defined by

(5.2)
$$G_J = \left(\prod_{J \in J} A_j\right)^{1/|J|} \text{ and } H(\mathcal{P}) = \sum_{J \in \mathcal{P}} |J| \log(A_J/G_J).$$

The so defined quantities satisfy $\sigma^2(\mathcal{P}) \leq \sigma_n^2$ and $H(\mathcal{P}) \geq 0$. 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

$$\ell_0(t) = t^2/8$$
 for $t \in [0,2]$ and $\ell_0(t) = t - \log t - (3/2) + \log 2$ for $t \ge 4$

For any positive t and any partition \mathcal{P} of $\{1, 2, \ldots, n\}$,

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

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

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

and, for any positive y,

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

Remark 5.1. In Theorem 3.1(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

$$\ell_0^*(x) = 2x^2$$
 for $x \in [0, 1/2]$ and $\ell_0^*(x) = -\log(1-x) + (1/2) - \log 2$ for $x \in [1/2, 1]$,
 $\ell_0^{*-1}(y) = \sqrt{y/2}$ for $y \in [0, 1/2]$ and $\ell_0^{*-1}(y) = 1 - (\sqrt{e}/2)e^{-y}$ 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 3.1(c) applied with $\mathcal{P} = \{[1, 13], [14, 100]\}$ (the optimal partition) yields $Q(p) \leq 62.18$, which improves the results of Section 3. 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 $q = \mathbb{P}(Z - \mathbb{E}(Z) \geq 75)$. Theorem 5.1(b) yields $q \leq 3.0 \, 10^{-7}$ instead of $q \leq 2.7 \, 10^{-8}$ with Theorem 3.2. Recall that the McDiarmid inequality yields $q \leq 1.1 \, 10^{-2}$.

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

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

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

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

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

(5.4)
$$\sum_{J\in\mathcal{P}} |J|\,\ell_0(A_J t) = \int_0^t \Bigl(\sum_{J\in\mathcal{P}} D_J \ell'_0(A_J u)\Bigr) du.$$

Now $\ell'_0(t) = t/4$ for $t \leq 2$ and $\ell'_0(t) = 1 - (1/t)$ for $t \geq 2$, 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

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

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 $5.1(b) \blacksquare$

References

Bentkus, Vidmantas (2007). On measure concentration for separately Lipschitz functions in product spaces. Israel J. Math. 158, 1-17.

Devroye, L., Lugosi, G. (2001). Combinatorial methods in density estimation. Springer Series in Statistics. Springer-Verlag, New York.

Hoeffding, Wassily (1963). Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58, 13-30.

McDiarmid, Colin (1989). On the method of bounded differences, dans Surveys of combinatorics, ed J. Siemons, London Mathematical Society lecture notes series 141.

McDiarmid, Colin (1998). "Concentration," Probabilistic methods for algorithmic discrete mathematics. pp. 195-248. M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed (Editors), Springer, New York.

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

Pinelis, Iosif (2006). On normal domination of (super)martingales. Electron. J. Probab. 11, 1049-1070.

Rio, Emmanuel (1994). Local invariance principles and their application to density estimation. Probab. Theory Related Fields 98, 21-45.

Rio, Emmanuel (2001). Inégalités de concentration pour les processus empiriques de classes de parties. Probab. theory relat. fields 119, 163-175.