The Annals of Probability
2002, Vol. 30, No. 4, 1576–1604

CONCENTRATION INEQUALITIES, LARGE AND MODERATE DEVIATIONS FOR SELF-NORMALIZED EMPIRICAL PROCESSES

BY BERNARD BERCU, ELISABETH GASSIAT AND EMMANUEL RIO

Université Paris-Sud, Université Paris-Sud and Université de Versailles Saint-Quentin

We consider the supremum $\mathcal{W}_n$ of self-normalized empirical processes indexed by unbounded classes of functions $\mathcal{F}$. Such variables are of interest in various statistical applications, for example, the likelihood ratio tests of contamination. Using the Herbst method, we prove an exponential concentration inequality for $\mathcal{W}_n$ under a second moment assumption on the envelope function of $\mathcal{F}$. This inequality is applied to obtain moderate deviations for $\mathcal{W}_n$. We also provide large deviations results for some unbounded parametric classes $\mathcal{F}$.

1. Introduction and main results. Let $(X, X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with values in some measurable space $(\mathbb{X}, \mathcal{X})$. Let $\mathcal{F}$ be a permissible class of real measurable functions on $(\mathbb{X}, \mathcal{X})$. We consider centered and normalized functions, that is, all functions f in $\mathcal{F}$ satisfy

$$\mathbb{E}[f(X)] = 0 \quad \text{and} \quad \mathbb{E}[f^2(X)] = 1. \tag{1.1}$$

Define

$$W_n(f) = \frac{P_n(f)}{\sqrt{P_n(f^2)}}, \tag{1.2}$$

where P_n denotes the empirical measure

$$P_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k}$$

with the convention $0/0 = 0$ if $P_n(f^2) = 0$. For each fixed function f in $\mathcal{F}$, $W_n(f)$ is a self-normalized sum. A striking result about self-normalized sums was obtained by Shao [23] and Dembo and Shao [6]: a large deviations principle holds for $W_n(f)$ without any integrability assumption on f and a moderate deviations principle holds for $W_n(f)$ as soon as $f(X)$ is centered and has a finite second moment. One can note that the precise result requires slightly weaker assumptions. These remarkable properties have to be compared with classical large deviations

Received March 2001; revised September 2001.

AMS 2000 subject classifications. Primary 60E15, 60F10; secondary 62E20, 62F05.

Key words and phrases. Maximal inequalities, self-normalized sums, empirical processes, concentration inequalities, large deviations, moderate deviations, logarithmic Sobolev inequalities.

results on empirical sums $P_n(f)$, where, roughly speaking, one requires the moment generating function of $f(X)$ to be finite in a neighborhood of zero. In this paper, we are interested in the self-normalized empirical process $(W_n(f))_{f\in\mathcal{F}}$. More specifically, we investigate various exponential bounds for the deviation of

$$\mathcal{W}_n = \sup_{f\in\mathcal{F}} W_n(f), \tag{1.3}$$

namely concentration inequalities and asymptotic exponential bounds given by large and moderate deviations principles. Such theoretical results are of practical interest for statistical applications as nonstandard likelihood ratio testing problems [4, 5, 13, 14]. For example, self-normalized score tests may be proposed to obtain a consistent test with exponentially decreasing level in the contamination testing problem [10].

Concentration inequalities have been investigated in depth in the last few years thanks to the important contribution of M. Talagrand [24, 25]. We refer the reader to McDiarmid [18] for an introduction to concentration inequalities with applications. These concentration inequalities apply to the supremum of empirical processes on classes of bounded functions. A different approach to derive concentration inequalities for empirical processes was proposed by Ledoux [15] and later developed by Massart [17] and Rio [21, 22]. All these results assume that the functions are uniformly bounded or at least uniformly bounded on the right-hand side. On the other hand, known results for the deviation of the supremum of empirical processes on classes of unbounded functions lead to upper bounds with nonexponential tails, except for classes of functions having an envelope function for which the moment generating function is finite in a right neighborhood of zero. As a matter of fact, in all other cases, the upper bounds mainly depend on the tail of the envelope function. We refer the reader to Pollard [19], Van der Vaart and Wellner [26] and Giné [11] for theoretical results on empirical processes.

In order to obtain exponential bounds in the case of unbounded functions, different ratios of empirical processes are studied by Pollard [20], Haussler [12] and Bartlett and Lugosi [3]. However, the ratios are not self-normalized, the upper bounds depend on the envelope function and they are not necessarily exponentially decreasing.

Our first result, established in Section 2, is a concentration inequality for $\mathcal{W}_n$.

THEOREM 1.1. *Assume that $\mathcal{F}$ is a countable class of centered and normalized functions with finite bracketing numbers in $L^2(P)$, such that*

$$E = \sup_{n>0} \mathbb{E}\Big[\sup_{f\in\mathcal{F}} \max\big(\sqrt{n}P_n(f), 0\big)\Big] < +\infty. \tag{1.4}$$

Then, for any $\delta > 0$ *and* $\alpha > \sqrt{2}$, *one can find some positive* θ *and* n_0 *depending on* $\mathcal{F}$, α *and* δ *such that, for* $n \geq n_0$ *and for any* x *in* $[0, \theta\sqrt{n}]$,

$$\mathbb{P}(\sqrt{n}\mathcal{W}_n \geq x + \alpha E) \leq 2\exp\left(-\frac{x^2}{4\alpha^2(1+\delta)}\right). \tag{1.5}$$

Assumption (1.4) is satisfied by most of the P-Donsker classes, such as classes with finite bracketing integral (see [27], page 270) or classes of functions introduced in [1]. Sections 3 and 4 are devoted to moderate and large deviations results for $\mathcal{W}_n$. As for concentration inequalities, the earlier moderate and large deviations results on empirical processes $(P_n(f))_{f\in\mathcal{F}}$ require strong assumptions on the class $\mathcal{F}$. In particular, Wu [28] established a functional large deviations principle for $P_n(f)$ under the assumption that an envelope function exists with moment generating function finite in a right neighborhood of zero. Moreover, Ledoux [16] proved moderate deviations results in Banach spaces under subexponential moment conditions on the norms of the random vectors. However, for self normalized empirical processes, this appears to be far too restrictive comparing to the results of Shao [23] and Dembo and Shao [6] and the purpose of Section 3 is to show that we can obtain moderate deviations results for $\mathcal{W}_n$ under quite similar assumptions as in Theorem 1.1.

In order to establish the moderate deviations principle, we need an additional condition on the brackets. We shall say that $\mathcal{F}$ has a finite covering with brackets in $L^2(P)$ satisfying concordance of signs if, for any $\delta > 0$, one can find a finite family $\mathcal{C}$ of pairs of measurable functions in $L^2(P)$ such that, for any f in $\mathcal{F}$, there exists (g, h) in $\mathcal{C}$ with

$$|g| \leq |f| \leq |h|, \qquad gf \geq 0, \qquad gh \geq 0 \quad \text{and} \quad \mathbb{E}[(h-g)^2(X)] \leq \delta. \tag{1.6}$$

THEOREM 1.2. *Use the same assumptions as in Theorem* 1.1. *Moreover, assume that* $\mathcal{F}$ *has a finite covering with brackets in* $L^2(P)$ *satisfying concordance of signs. Then, for any sequence* (x_n) *tending to infinity such that* $x_n = o(\sqrt{n})$, *we have*

$$\lim_{n\to+\infty} \frac{1}{x_n^2} \log \mathbb{P}\left(\mathcal{W}_n \geq \frac{x_n}{\sqrt{n}}\right) = -\frac{1}{2}. \tag{1.7}$$

Recently, Shao [23] has established a large deviations principle for $W_n(f)$ for each fixed function f in $\mathcal{F}$ such that $f(X)$ has continuous distribution. In particular, he has shown that for any $x > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_n(f) \geq x) = -I_f(x)$$

where the rate function I_f is explicitly given. In contrast with moderate deviations results, the denominator $P_n(f^2)$ plays a fundamental role. This introduces further difficulties for proving a large deviations principle for $\mathcal{W}_n$. In Section 4, we shall restrict ourselves to parametric classes associated with exponential models. More precisely, we consider $\mathcal{F} = \{f_\gamma \text{ with } \gamma \in [m, 0^-] \cup [0^+, M]\}$ where m is negative, M is positive,

$$f_\gamma(x) = \begin{cases} \exp(\gamma t(x) - l(\gamma)) - 1, & \text{if } \gamma \neq 0, \\ t(x), & \text{if } \gamma = 0^+, \\ -t(x), & \text{if } \gamma = 0^-. \end{cases} \tag{1.8}$$

The function t is continuous real measurable with moment generating function $l(\gamma) = \log \mathbb{E}[\exp(\gamma t(X))]$. We assume that $[2m, 2M]$ is included in the domain of l so that the functions of $\mathcal{F}$ are square integrable. Define the rate function $\mathcal{I}$ by

$$\mathcal{I}(x) = \begin{cases} \inf_{f \in \mathcal{F}} I_f(x), & \text{if } x \geq 0, \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.9}$$

THEOREM 1.3. *Let $\mathcal{F}$ be the parametric class of centered and square integrable functions given by* (1.8). *Moreover, assume that $t(X)$ has a continuous distribution function and $\mathbb{E}[t(X)] = 0$. Then, $(\mathcal{W}_n)$ satisfies a large deviations principle with continuous rate function $\mathcal{I}$. In particular, for any $x \geq 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{W}_n \geq x) = -\mathcal{I}(x). \tag{1.10}$$

2. Concentration inequalities. The purpose of this section is to establish a concentration inequality for $\mathcal{W}_n$. The main tool for proving this inequality is the entropy method proposed by Ledoux [15], which provides differential inequalities for various functionals of random measures. Due to the structure of self-normalized empirical processes, we avoid symmetrization. We start by recalling a differential inequality for Laplace transforms of functionals of independent random variables which was stated in Massart [17] in a slightly different version. The present form can be found in Rio [21].

THEOREM 2.1. *Let $\mathcal{F}_n$ be the σ-algebra generated by $(X_1, X_2, \dots, X_n)$ and denote by $\mathcal{F}_n^k$ the σ-algebra generated by $(X_1, X_2, \dots, X_{k-1}, X_{k+1}, \dots, X_n)$. Set*

$$\psi(x) = \exp(-x) + x - 1, \qquad \phi(x) = 1 - (1+x)\exp(-x).$$

Let $Z = Z(X_1, \dots, X_n)$ be a given bounded functional and denote by F_n its Laplace transform. Then, for any sequence $(Z_k)_{k\in[1,n]}$ of bounded functionals respectively $\mathcal{F}_n^k$-measurable and for any nonnegative λ, we have

$$\lambda F_n'(\lambda) - F_n(\lambda)\log F_n(\lambda) \leq \sum_{k=1}^n \mathbb{E}\big[\psi\big(\lambda(Z - Z_k)_+\big)e^{\lambda Z} + \phi\big(\lambda(Z - Z_k)_-\big)e^{\lambda Z_k}\big].$$

For all f in $\mathcal{F}$, let $S_n(f) = nP_n(f)$. We shall apply the above differential inequality to the process

$$Z^{(N)}(n) = \sup_{f\in\mathcal{F}} \frac{\max(S_n(f), 0)}{\sqrt{N + S_n(f_+^2)}}$$

with n in $[1, N]$. We set $Z^{(N)} = Z^{(N)}(N)$.

THEOREM 2.2. *Assume that* $\mathcal{F}$ *is a countable class of centered and normalized functions*, *totally bounded in* $L^2(P)$. *Let*

$$E_N = \sup_{n\in[1,N]} \mathbb{E}[Z^{(N)}(n)]$$

and assume that $\zeta = \sup_{N>0} E_N < \infty$. *Set*

$$\eta(t) = t \sup_{f\in\mathcal{F}} \mathbb{P}\big(f^2(X) > t\big). \tag{2.1}$$

Then, *for any* $\delta > 0$, *one can find some positive* ε *and* N_0 *only depending on* η *and* ζ *such that*, *for* $N \ge N_0$ *and for any* λ *in* $[0, \varepsilon\sqrt{N}]$,

$$\mathbb{E}\big[\exp\big(\lambda Z^{(N)}\big)\big] \le \exp\big(\lambda E_N + (1+\delta)\lambda^2\big). \tag{2.2}$$

Consequently, *for any* x *in* $[0, 2(1+\delta)\varepsilon\sqrt{N}]$,

$$\mathbb{P}\big(Z^{(N)} \ge x + E_N\big) \le \exp\bigg(-\frac{x^2}{4(1+\delta)}\bigg). \tag{2.3}$$

PROOF. First of all, relation (2.3) immediately follows from the standard Cramér–Chernoff calculation. Next, in order to avoid heaviness in the notation, we rewrite $Z^{(N)}(n)$ as Z. Our goal is to bound up the Laplace transform F_n of Z via Theorem 2.1. We shall obtain this upper bound by induction on n. Define

$$Z_k = \sup_{f\in\mathcal{F}} \frac{\max(S_n^k(f), 0)}{\sqrt{N + S_n^k(f_+^2)}} \qquad \text{with } S_n^k(f) = \sum_{i=1}^n f(X_i)\mathbb{1}_{i\ne k}.$$

In order to use Theorem 2.1, it is necessary to bound up the random variable $Z - Z_k$. We may assume without loss of generality that $\mathcal{F}$ is finite, that is, $\mathcal{F} = \{f_1, \ldots, f_m\}$ with $m \ge 1$. One the one hand, for the positive part of $Z - Z_k$, let τ be the infimum of integers i such that

$$Z = \frac{\max(S_n(f_i), 0)}{\sqrt{N + S_n(f_{i+}^2)}}.$$

Clearly, we may assume $Z > 0$, which ensures that $S_n(f_\tau) > 0$. If $S_n^k(f_\tau) \le 0$, then $f_\tau(X_k) > 0$ and

$$Z - Z_k \le \frac{S_n(f_\tau)}{\sqrt{N + S_n(f_{\tau+}^2)}} \le \frac{f_{\tau+}(X_k)}{\sqrt{N + S_n(f_{\tau+}^2)}}.$$

Otherwise,

$$Z_k \ge \frac{S_n^k(f_\tau)}{\sqrt{N + S_n^k(f_{\tau+}^2)}} > 0$$

and therefore

$$Z - Z_k \le \frac{S_n(f_\tau)}{\sqrt{N + S_n(f_{\tau+}^2)}} - \frac{S_n^k(f_\tau)}{\sqrt{N + S_n^k(f_{\tau+}^2)}} \le \frac{f_{\tau+}(X_k)}{\sqrt{N + S_n(f_{\tau+}^2)}}.$$

Noting that $\psi(x) \le x^2/2$ for any positive x, we find that for all $\lambda \ge 0$,

$$\psi\big(\lambda(Z - Z_k)_+\big) \le \frac{\lambda^2 f_{\tau+}^2(X_k)}{2N + 2S_n(f_{\tau+}^2)},$$

which immediately implies that

$$\sum_{k=1}^{n} \mathbb{E}\big[\psi\big(\lambda(Z - Z_k)_+\big)e^{\lambda Z}\big] \le \frac{\lambda^2}{2} F_n(\lambda). \tag{2.4}$$

On the other hand, for the negative part of $Z - Z_k$, let τ_k be the infimum of integers i such that

$$Z_k = \frac{\max(S_n^k(f_i), 0)}{\sqrt{N + S_n^k(f_{i+}^2)}}.$$

As before, we may assume $Z_k > 0$, which ensures that $S_n^k(f_{\tau_k}) > 0$. If $S_n(f_{\tau_k}) \le 0$, then $f_{\tau_k}(X_k) < 0$ and

$$Z_k - Z \le \frac{S_n^k(f_{\tau_k})}{\sqrt{N + S_n^k(f_{\tau_k+}^2)}} \le \frac{f_{\tau_k-}(X_k)}{\sqrt{N + S_n^k(f_{\tau_k+}^2)}}.$$

Otherwise,

$$Z \ge \frac{S_n(f_{\tau_k})}{\sqrt{N + S_n(f_{\tau_k+}^2)}} > 0$$

and therefore

$$Z_k - Z \le \frac{S_n^k(f_{\tau_k})}{\sqrt{N + S_n^k(f_{\tau_k+}^2)}} - \frac{S_n(f_{\tau_k})}{\sqrt{N + S_n(f_{\tau_k+}^2)}}$$

$$\le \frac{f_{\tau_k-}(X_k)}{\sqrt{N + S_n(f_{\tau_k+}^2)}} + Z_k\left(1 - \sqrt{\frac{N + S_n^k(f_{\tau_k+}^2)}{N + S_n(f_{\tau_k+}^2)}}\right).$$

Consequently, it follows that

$$Z_k - Z \le \frac{f_{\tau_k-}(X_k)}{\sqrt{N + S_n^k(f_{\tau_k+}^2)}} + \frac{Z_k f_{\tau_k+}^2(X_k)}{2N + S_n(f_{\tau_k+}^2)}. \tag{2.5}$$

Note that either the first term or the second term in this upper bound are null. Since $\phi(x) \le x^2/2$ for any positive x, setting $Y_k = f_{\tau_k+}^2(X_k)$, we deduce that

$$\phi\big(\lambda(Z - Z_k)_-\big) \le \frac{\lambda^2}{2N} f_{\tau_k-}^2(X_k) + \phi\left(\frac{\lambda Y_k Z_k}{2N + Y_k}\right).$$

Let $\mathbb{E}_n^k$ denote the conditional expectation with respect to $\mathcal{F}_n^k$ and $\mathbb{P}_n^k$ be the corresponding conditional probability. Integrating with respect to X_k the above inequality and noting that $\phi'(x) = x\exp(-x)$, we find that

$$\mathbb{E}_n^k[\phi\big(\lambda(Z - Z_k)_-\big)] \le \frac{\lambda^2}{2N}\mathbb{E}_n^k[f_{\tau_k-}^2(X_k)] + \int_0^\infty \mathbb{P}_n^k(V_k \ge t) t e^{-t}\, dt$$

with $V_k = \lambda(2N + Y_k)^{-1} Y_k Z_k$. On the one hand, $\mathbb{E}[f^2(X_k)] = 1$ and the stopping time τ_k is $\mathcal{F}_n^k$-measurable so that $\mathbb{E}_n^k[f_{\tau_k-}^2(X_k)] \le 1$. On the other hand, set

$$\Lambda(t) = \sup_{f \in \mathcal{F}} \mathbb{P}\big(f^2(X) > t\big) = \frac{\eta(t)}{t},$$

where the function η is given by (2.1). Since Z_k is $\mathcal{F}_n^k$-measurable, we obtain that

$$\int_0^\infty \mathbb{P}_n^k(V_k \ge t) t e^{-t}\, dt \le \int_0^{\lambda Z_k} t\Lambda\left(\frac{2Nt}{\lambda Z_k - t}\right) e^{-t}\, dt \le \frac{\lambda Z_k}{2N} I(\lambda Z_k)$$

with

$$I(z) = \int_0^z \eta(2Ntz^{-1}) \exp(-t)\, dt.$$

Therefrom, taking expectations, we deduce that

$$\sum_{k=1}^n \mathbb{E}[\phi\big(\lambda(Z - Z_k)_-\big) e^{\lambda Z_k}] \le \frac{\lambda^2}{2} F_{n-1}(\lambda) + \frac{\lambda}{2}\mathbb{E}[Z_k \exp(\lambda Z_k) I(\lambda Z_k)]. \tag{2.6}$$

We now control the integral $I(z)$ by the straightforward Toeplitz-like lemma below.

LEMMA 2.3. *Assume that $\mathcal{F}$ is totally bounded in $L^2(P)$. Then,* $\lim_{t\to\infty}\eta(t)=0$,

$$I(z)=\big(1-\exp(-z)\big)\varepsilon\left(\frac{\max(1,z)}{2N}\right)$$

for some function ε with values in $[0,1]$ *satisfying* $\lim_{t\to 0}\varepsilon(t)=0$.

PROOF. As the family of functions $\{f^2, f\in\mathcal{F}\}$ is relatively compact in $L^1(P)$, we classically have the uniform integrability of $\{f^2(X), f\in\mathcal{F}\}$ (see [9], page 294), which implies the convergence of η to 0. Next, if $z\le 1$, then

$$I(z)\le\frac{z}{2N}\int_0^{2N}\eta(u)\,du,$$

which implies Lemma 2.3. Otherwise, let $a=(z/2N)^{1/2}$. Since $\eta(t)\le 1$, we get

$$I(z)\le a+\int_a^z\eta(2Ntz^{-1})e^{-t}\,dt\le(1-e^{-z})\left(\frac{ea}{e-1}+\sup_{x\ge 1/a}\eta(x)\right)$$

which completes the proof of Lemma 2.3. □

Now, from (2.6) and Lemma 2.3 together with the elementary fact that for all $x\ge 0$, $x(\exp(x)-1)\le 2\phi(-x)$, we find that

$$\sum_{k=1}^n\mathbb{E}\big[\phi\big(\lambda(Z-Z_k)_-\big)e^{\lambda Z_k}\big]\le\frac{\lambda^2}{2}F_{n-1}(\lambda)+\mathbb{E}\left[\phi(-\lambda Z_k)\varepsilon\left(\frac{\max(1,\lambda Z_k)}{2N}\right)\right].$$

By the Cauchy–Schwarz inequality, $Z_k\le\sqrt{N}$. Consequently, if

$$\varepsilon^*(x)=\sup\{\varepsilon(t):t\in]0,x]\},$$

we can deduce that for $N\ge 1/(2\alpha)$ and for any $\lambda\le\alpha\sqrt{N}$,

$$\sum_{k=1}^n\mathbb{E}\big[\phi\big(\lambda(Z-Z_k)_-\big)e^{\lambda Z_k}\big]\le\frac{\lambda^2}{2}F_{n-1}(\lambda)+\varepsilon^*(\alpha)\mathbb{E}[\phi(-\lambda Z_k)].\tag{2.7}$$

In addition, one can observe that

$$\mathbb{E}[\phi(-\lambda Z_k)]=\lambda F'_{n-1}(\lambda)-F_{n-1}(\lambda)+1.$$

Piecing together the contributions of the positive part (2.4) and of the negative part (2.7), we obtain by Theorem 2.1 that for $N\ge 1/(2\alpha)$ and for any $\lambda\le\alpha\sqrt{N}$,

$$\lambda F'_n-F_n\log F_n\le\frac{\lambda^2}{2}(F_{n-1}+F_n)+\varepsilon^*(\alpha)(\lambda F'_{n-1}-F_{n-1}+1).\tag{2.8}$$

We are now in position to state our induction hypothesis. The induction hypothesis $\mathcal{H}(n)$ at range n is that for any λ in $]0, \alpha\sqrt{N}]$,

$$\begin{cases} F_n(\lambda) < \exp(\lambda E_N + (1+\delta)\lambda^2), \\ \lambda F_n'(\lambda) - F_n(\lambda) + 1 \le C\lambda^2 \exp(\lambda E_N + (1+\delta)\lambda^2). \end{cases}$$

At range 0, we assume that $F_0(\lambda) = 1$. Hence the induction hypothesis holds true at range 0. Let n be some integer in $[1, N]$. Suppose that the induction hypothesis holds at range $n-1$. Set

$$H(\lambda) = \exp(\lambda E_N + (1+\delta)\lambda^2).$$

Then, we find via (2.8) that

$$\lambda F_n'(\lambda) - F_n(\lambda)\log F_n(\lambda) < \frac{\lambda^2}{2} F_n(\lambda) + \frac{\lambda^2}{2} H(\lambda) + \varepsilon^*(\alpha) C\lambda^2 H(\lambda). \tag{2.9}$$

Consequently, F_n is a subsolution of the Differential Equation (DE) corresponding to the equality in (2.9) with $F_n(0) = 1$ and $F_n'(0) = \mathbb{E}[Z] \le E_N$. Moreover, we have

$$\lambda H'(\lambda) - H(\lambda)\log H(\lambda) = \frac{\lambda^2}{2} H(\lambda) + \frac{\lambda^2}{2}(1+2\delta) H(\lambda).$$

Hence, if α and C are such that $\delta = \varepsilon^*(\alpha)C$, H is the solution of (DE) such that $H(0) = 1$ and $H'(0) = E_N$. Therefrom, by the comparison lemma (see [2], page 26), we obtain that for all n in $[1, N]$ and for any λ in $]0, \alpha\sqrt{N}]$,

$$F_n(\lambda) < H(\lambda).$$

It remains to choose α in such a way that the second part of the induction hypothesis holds true. From (2.8) and the above inequality, we derive that

$$\lambda F_n'(\lambda) - F_n(\lambda) + 1 \le F_n(\lambda)\log F_n(\lambda) - F_n(\lambda) + 1 + \lambda^2(1+\delta)H(\lambda).$$

Now, recall that $F_n(\lambda) \ge 1$ and observe that the function $x\log x - x + 1$ is nondecreasing on $[1, +\infty[$. Since $F_n \le H$, we infer that

$$\lambda F_n'(\lambda) - F_n(\lambda) + 1 \le H(\lambda)\log H(\lambda) - H(\lambda) + 1 + \lambda^2(1+\delta)H(\lambda).$$

Set $G(\lambda) = \log H(\lambda) = \lambda E_N + (1+\delta)\lambda^2$. Then, we obtain that

$$\lambda F_n'(\lambda) - F_n(\lambda) + 1 \le \big(\psi\big(G(\lambda)\big) + \lambda^2(1+\delta)\big) H(\lambda). \tag{2.10}$$

However, ψ is a convex function so that

$$\psi\big(G(\lambda)\big) \le \tfrac{1}{2}\big(\psi(2\lambda E_N) + \psi\big(2(1+\delta)\lambda^2\big)\big) \le (E_N^2 + 1 + \delta)\lambda^2. \tag{2.11}$$

Finally, if we choose α in such a way that $\varepsilon^*(\alpha) \le 1/3$ and if we take $C = 3(\zeta^2 + 2)$, we deduce from (2.10) together with (2.11) that

$$\lambda F_n'(\lambda) - F_n(\lambda) + 1 \le \big(E_N^2 + 2(1+\delta)\big)\lambda^2 H(\lambda) \le C\lambda^2 H(\lambda)$$

which immediately implies $\mathcal{H}(n)$. Hence, by induction, $\mathcal{H}(N)$ also holds, completing the proof of Theorem 2.2. □

Now, we deduce the concentration inequality for $\mathcal{W}_n$ from Theorem 2.2.

THEOREM 2.4. *Assume that $\mathcal{F}$ is a countable class of centered and normalized functions, totally bounded in $L^2(P)$. Moreover, assume that for any ζ in* $]0,1[$, *one can find a finite family $\mathcal{G}$ of real measurable functions satisfying the two constraints: for all f in $\mathcal{F}$, there exists g in $\mathcal{G}$ such that*

$$f^2 \geq g^2 \quad \textit{and} \quad \mathbb{E}[f^2(X) - g^2(X)] \leq \zeta. \tag{2.12}$$

Then, for any $\delta > 0$ and $\alpha > \sqrt{2}$, one can find some positive ξ and n_0 depending on $\mathcal{F}$, E given by (1.4) *and α such that, for $n \geq n_0$ and for any λ in* $[0, \xi\sqrt{n}]$,

$$\mathbb{E}\big[\exp(\lambda\sqrt{n}\mathcal{W}_n)\big] \leq 2\exp\big(\alpha\lambda E + (1+\delta)\alpha^2\lambda^2\big). \tag{2.13}$$

PROOF. It follows from (1.2) together with (1.3) that

$$\sqrt{n}\mathcal{W}_n \leq \sup_{f\in\mathcal{F}} \frac{S_n(f)}{\sqrt{n + S_n(f_+^2)}}\left(\frac{n}{S_n(f^2)} + 1\right)^{1/2}.$$

For $\alpha > \sqrt{2}$, set $a_\alpha = (\alpha^2 - 1)^{-1}$ and

$$\Gamma_\alpha = \Big\{\inf_{f\in\mathcal{F}} S_n(f^2) \geq a_\alpha n\Big\}.$$

On Γ_α, we clearly have $\sqrt{n}\mathcal{W}_n \leq \alpha Z$ with $Z = Z^{(n)}(n)$. Hence

$$\mathbb{E}\big[\exp(\lambda\sqrt{n}\mathcal{W}_n)\big] \leq 2\max\big(F_n(\alpha\lambda), \exp(\lambda\sqrt{n})P(\Gamma_\alpha^c)\big), \tag{2.14}$$

where F_n denotes the Laplace transform of Z. On the one hand, we obtain from Theorem 2.2 that for $\lambda < \varepsilon\sqrt{n}/\alpha$,

$$F_n(\alpha\lambda) \leq \exp\big(\alpha\lambda E + (1+\delta)\alpha^2\lambda^2\big). \tag{2.15}$$

On the other hand, let $\mathcal{G}_\zeta$ be a finite family with minimal cardinality among the families satisfying (2.12). First, from (2.12),

$$\min_{g\in\mathcal{G}_\zeta} \mathbb{E}[g^2(X)] \geq 1 - \zeta. \tag{2.16}$$

In addition,

$$\Gamma_\alpha^c \subset \Big\{\inf_{g\in\mathcal{G}_\zeta} S_n(g^2) \leq n a_\alpha\Big\}.$$

Then, applying Proposition A.1 to $\mathcal{G}_\zeta$, we deduce that

$$\mathbb{P}(\Gamma_\alpha^c) \leq |\mathcal{G}_\zeta| \exp(-n\theta_\alpha)$$

with $\theta_\alpha > 0$ provided that $a_\alpha < 1 - \zeta$, that is, $\zeta < (\alpha^2 - 1)^{-1}(\alpha^2 - 2)$. Therefore, we find from (2.14) together with (2.15) that for all $\lambda < \varepsilon\sqrt{n}/\alpha$,

$$E\big[\exp(\lambda\sqrt{n}\mathcal{W}_n)\big] \le 2\exp\big(\alpha\lambda E + (1+\delta)\alpha^2\lambda^2\big) \tag{2.17}$$

as soon as $|\mathcal{G}_\zeta|\exp(\lambda\sqrt{n} - n\theta_\alpha) \le 1$ which can be rewritten as $n\theta_\alpha - c_\zeta \ge \lambda\sqrt{n}$ with $c_\zeta = \log|\mathcal{G}_\zeta|$, completing the proof of Theorem 2.4. One can note that Theorem 1.1 immediately follows from Theorem 2.4. $\square$

3. Moderate deviations. Theorem 1.2 provides the moderate deviations principle for $\mathcal{W}_n$. It is derived from concentration inequalities for the fluctuations process as in Ledoux [16].

PROOF OF THEOREM 1.2. First of all, let $\mathcal{C} = \{(g_i, h_i) \text{ with } i \in I\}$ be a family of brackets satisfying concordance of signs, as defined in Section 1. Next, let $(B_i)_{i\in I}$ be a partition of $\mathcal{F}$ such that, for any f in B_i,

$$|g_i| \le |f| \le |h_i| \quad \text{with} \quad \mathbb{E}[(h_i - g_i)^2(X)] \le \delta \text{ and } g_i f \ge 0,\ g_i h_i \ge 0. \tag{3.1}$$

For any ε in $]0, 1[$, we have the decomposition

$$\mathbb{P}\left(\mathcal{W}_n \ge \frac{x_n}{\sqrt{n}}\right) \le A_n(\delta, \varepsilon) + \sum_{i\in I} B_n^i(\delta, \varepsilon), \tag{3.2}$$

where

$$A_n(\delta, \varepsilon) = \mathbb{P}\left(\sup_{i\in I} \frac{\max(S_n(g_i - \mathbb{E}[g_i]), 0)}{\sqrt{S_n(g_i^2)}} \ge (1-\varepsilon)x_n\right),$$

$$B_n^i(\delta, \varepsilon) = \mathbb{P}\left(\sup_{f\in B_i}\left(\sqrt{n}\max\big(W_n(f), 0\big) - \frac{\max(S_n(g_i - \mathbb{E}[g_i]), 0)}{\sqrt{S_n(g_i^2)}}\right) \ge \varepsilon x_n\right).$$

On the one hand, we apply Theorem 3.1 of Shao [23] to obtain moderate deviations for $A_n(\delta, \varepsilon)$. On the other hand, we bound up the remainder terms $B_n^i(\delta, \varepsilon)$ via the following concentration inequality, which is proven in Appendix B. Let

$$V_n^i(\delta) = \sup_{f\in B_i} \frac{\max(S_n(f - g_i + \mathbb{E}[g_i]), 0)}{\sqrt{S_n(f^2)}}.$$

THEOREM 3.1. *Under the assumptions of Theorem* 1.2, *for any* $\delta > 0$ *and* $\alpha > \sqrt{2}$, *one can find some positive* ξ *and* n_0 *depending on* $\mathcal{F}$, E *given by* (1.4) *and* α *such that*, *for* $n \ge n_0$ *and for any* λ *in* $[0, \xi\sqrt{n}]$,

$$\mathbb{E}\big[\exp\big(\lambda V_n^i(\delta)\big)\big] \le 2\exp\big(\alpha\lambda(1+E) + 16|\log\delta|^{-1}\alpha^2\lambda^2\big). \tag{3.3}$$

Now, by use of Theorem 3.1, we prove that there exists $\delta > 0$ such that $B_n^i(\delta, \varepsilon)$ is bounded by $\exp(-x_n^2)$. For any f in B_i, since $S_n(f^2) \geq S_n(g_i^2)$, we have

$$\frac{\max(S_n(g_i - \mathbb{E}[g_i]), 0)}{\sqrt{S_n(g_i^2)}} \geq \frac{\max(S_n(g_i - \mathbb{E}[g_i]), 0)}{\sqrt{S_n(f^2)}}.$$

Hence, by (1.2),

$$\begin{aligned} &\sqrt{n}\max\bigl(W_n(f), 0\bigr) - \frac{\max(S_n(g_i - \mathbb{E}[g_i]), 0)}{\sqrt{S_n(g_i^2)}} \\ &\qquad \leq \frac{\max(S_n(f - g_i + \mathbb{E}[g_i]), 0)}{\sqrt{S_n(f^2)}}. \end{aligned} \tag{3.4}$$

Consequently, applying (3.3) with $\alpha = 2$, $\delta = \exp(-512/\varepsilon^2)$ and $\lambda = |\log\delta|\frac{\varepsilon x_n}{128}$ together with Markov's inequality, we have for n large enough

$$B_n^i(\delta, \varepsilon) \leq \mathbb{P}\bigl(V_n^i(\delta) \geq \varepsilon x_n\bigr) \leq \exp(-x_n^2). \tag{3.5}$$

In addition, for any i in I, $|\mathbb{E}[g_i]| \leq \sqrt{\delta}$. Hence,

$$\begin{aligned} \sup_{i\in I}\frac{S_n(g_i - \mathbb{E}[g_i])}{\sqrt{S_n(g_i^2)}} &= \sup_{i\in I}\frac{S_n(g_i - \mathbb{E}[g_i])}{\sqrt{S_n((g_i - \mathbb{E}[g_i])^2)}}\sqrt{\frac{S_n((g_i - \mathbb{E}[g_i])^2)}{S_n(g_i^2)}} \\ &\leq \sup_{i\in I}\sqrt{n}W_n(g_i - \mathbb{E}[g_i])\left(1 + \frac{\sqrt{n}|\mathbb{E}[g_i]|}{\sqrt{S_n(g_i^2)}}\right) \\ &\leq \sup_{i\in I}\sqrt{n}W_n(g_i - \mathbb{E}[g_i])\left(1 + \sqrt{\frac{n\delta}{S_n(g_i^2)}}\right). \end{aligned}$$

Therefore,

$$A_n(\delta, \varepsilon) \leq \mathbb{P}\left(\sup_{i\in I} W_n(g_i - \mathbb{E}[g_i]) \geq \frac{(1-\varepsilon)x_n}{(1+\varepsilon)\sqrt{n}}\right) + \mathbb{P}\left(\inf_{i\in I}\frac{S_n(g_i^2)}{n} \leq \frac{\delta}{\varepsilon^2}\right). \tag{3.6}$$

By Theorem 3.1 of Shao [23], we have

$$\lim_{n\to+\infty}\frac{1}{x_n^2}\log\mathbb{P}\left(\sup_{i\in I} W_n(g_i - \mathbb{E}[g_i]) \geq \frac{(1-\varepsilon)x_n}{(1+\varepsilon)\sqrt{n}}\right) = -\frac{1}{2}\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^2. \tag{3.7}$$

Furthermore, one can notice that for any i in I, $\mathbb{E}[g_i^2] \geq 1 - 2\sqrt{\delta}$. Then, since $1 - 2\sqrt{\delta} > \delta/\varepsilon^2$, we deduce from Proposition A.1 that we can find $\theta > 0$ such that

$$\limsup_{n\to+\infty}\frac{1}{n}\log\mathbb{P}\left(\inf_{i\in I}\frac{S_n(g_i^2)}{n} \leq \frac{\delta}{\varepsilon^2}\right) \leq -\theta. \tag{3.8}$$

Then, we deduce from (3.5) to (3.8) that

$$\limsup_{n\to+\infty} \frac{1}{x_n^2} \log \mathbb{P}\Big(\mathcal{W}_n \geq \frac{x_n}{\sqrt{n}}\Big) \leq -\frac{1}{2}\Big(\frac{1-\varepsilon}{1+\varepsilon}\Big)^2$$

which, by the arbitrariness of ε, immediately implies

$$\limsup_{n\to+\infty} \frac{1}{x_n^2} \log \mathbb{P}\Big(\mathcal{W}_n \geq \frac{x_n}{\sqrt{n}}\Big) \leq -\frac{1}{2}.$$

Finally, by Theorem 3.1 of Shao [23], we have for any fixed function f in $\mathcal{F}$,

$$\limsup_{n\to+\infty} \frac{1}{x_n^2} \log \mathbb{P}\Big(\mathcal{W}_n \geq \frac{x_n}{\sqrt{n}}\Big) \geq \limsup_{n\to+\infty} \frac{1}{x_n^2} \log \mathbb{P}\Big(W_n(f) \geq \frac{x_n}{\sqrt{n}}\Big) = -\frac{1}{2},$$

which completes the proof of Theorem 1.2. □

4. Large deviations. Recently, Shao [23] has established a large deviations principle for $W_n(f)$ for each fixed function f in $\mathcal{F}$ such that $f(X)$ has continuous distribution function. In particular, he has shown that for any $x > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(W_n(f) \geq x\big) = -I_f(x)$$

where

$$I_f(x) = -\log \sup_{a\geq 0} \inf_{t\geq 0} \mathbb{E}\big[\exp\big\{t\big(af(X) - x\big(f^2(X) + a^2\big)/2\big)\big\}\big].$$

In this section, we consider the parametric class $\mathcal{F}$ given by (1.8). Due to the structure of the class, it is sufficient to prove a large deviations principle for nonnegative values of the parameter γ. As a matter of fact, one can change $t(x)$ into $-t(x)$ to obtain Theorem 1.3. Consequently, we consider the class $\mathcal{F} = \{f_\gamma \text{ with } \gamma \in [0, M]\}$.

PROOF OF THEOREM 1.3. By rescaling the class, we may without loss of generality assume that $M = 1$. Let δ be a positive real in $]0, 1]$ and set $N = n^2$. For any integer j in $[1, N]$, define $I_j = [(1+\delta)^{j-1-N}, (1+\delta)^{j-N}]$ and set $I_0 = [0, (1+\delta)^{-N}]$. For j in $[1, N]$, let $a_j = \inf I_j$ and define

$$g_j(x) = \operatorname{Sign}\big(f_{a_j}(x)\big) \inf_{\gamma\in I_j} |f_\gamma(x)|. \tag{4.1}$$

In addition, define the function h by $h(0) = 0$ and $h(\gamma) = l(\gamma)/\gamma$ and let

$$g_0(x) = \big(t(x) - h(a_1)\big)_+ - a_1^{-1}\big(\exp\big(a_1 t(x)\big) - 1\big)_-. \tag{4.2}$$

For any positive x and α with $\alpha < x$, we have the decomposition

$$\mathbb{P}(\mathcal{W}_n \geq x) \leq A_n(\delta, \alpha) + B_n(\delta, x - \alpha), \tag{4.3}$$

where

$$A_n(\delta, \alpha) = \mathbb{P}\big(\mathcal{W}_n - \max\big(W_n(g_0), \ldots, W_n(g_N), 0\big) \geq \alpha\big),$$
$$B_n(\delta, \beta) = \mathbb{P}\big(\max\big(W_n(g_0), \ldots, W_n(g_N)\big) \geq \beta\big).$$

Theorem 1.3 follows from both the continuity of $\mathcal{I}$ together with the exponential negligibility of A_n and some large deviations bound for B_n, which are stated below and proved in Appendix C. □

LEMMA 4.1. *Under the assumptions of Theorem* 1.3, *for any positive* α,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log A_n(\delta, \alpha) = -\infty.$$

LEMMA 4.2. *Under the assumptions of Theorem* 1.3, *for any positive* β,

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log B_n(\delta, \beta) \leq -\mathcal{I}(\beta).$$

APPENDIX A

In this Appendix we establish an exponential inequality for empirical processes indexed by square functions similar to the well-known Bennett inequality. However, since our result deals with the left-hand side deviation, we only require a second moment assumption.

PROPOSITION A.1. *Let* $\mathcal{G}$ *be a finite family in* $L^2(P)$ *such that*

$$v = \inf_{g \in \mathcal{G}} \mathbb{E}[g^2] > 0.$$

Moreover, *for any* g *in* $\mathcal{G}$, *let* ϕ_g *be the function defined by* $\phi_g(x) = \mathbb{P}(g^2(X) \geq x)$. *Then*, *for all negative* λ, *we have*

$$\text{(A.1)} \qquad \log \mathbb{E}\Big[\exp\big(\lambda \inf_{g \in \mathcal{G}} S_n(g^2)\big)\Big] \leq \log |\mathcal{G}| + n\lambda\big(H(\lambda) + v\big)$$

where

$$H(\lambda) = \inf_{g \in \mathcal{G}} \int_0^\infty \phi_g(x)(e^{\lambda x} - 1)\, dx.$$

Consequently, *for any* $a > 0$ *with* $a < v$, *one can find* $\theta > 0$ *such that*

$$\text{(A.2)} \qquad \mathbb{P}\bigg(\inf_{g \in \mathcal{G}} \frac{S_n(g^2)}{n} \leq a\bigg) \leq |\mathcal{G}| \exp(-n\theta).$$

PROOF. We proceed as in the proof of Bennett's inequality. For any g in $\mathcal{G}$,

$$\log \mathbb{E}\Big[\exp\big(\lambda \inf_{g\in\mathcal{G}} S_n(g^2)\big)\Big] \le \log|\mathcal{G}| + n \sup_{g\in\mathcal{G}} \log \mathbb{E}\big[\exp\big(\lambda g^2(X)\big)\big].$$

Since $\log x \le x-1$, we obtain

$$\log \mathbb{E}\big[\exp\big(\lambda g^2(X)\big)\big] \le \mathbb{E}\big[\exp\big(\lambda g^2(X)\big) - 1\big],$$

so that

$$\log \mathbb{E}\big[\exp\big(\lambda g^2(X)\big)\big] \le \lambda v + \mathbb{E}\big[\exp\big(\lambda g^2(X)\big) - \lambda g^2(X) - 1\big]. \tag{A.3}$$

Integrating by parts, we have

$$\mathbb{E}\big[\exp\big(\lambda g^2(X)\big) - \lambda g^2(X) - 1\big] = \lambda \int_0^\infty \phi_g(x)(e^{\lambda x} - 1)\, dx,$$

which immediately implies (A.1). Then, by (A.1) and Markov's inequality, we have for all $\lambda < 0$ and for any $a < v$

$$\mathbb{P}\Big(\inf_{g\in\mathcal{G}} S_n(g^2) \le na\Big) \le |\mathcal{G}| \exp\big(n\lambda\big(H(\lambda) + v - a\big)\big).$$

Moreover, by the Lebesgue dominated convergence theorem, $H(\lambda)$ converges to 0 as λ tends to 0. Hence, (A.1) implies (A.2). □

APPENDIX B

The goal of this Appendix is to obtain concentration inequalities for the processes

$$\Delta^{(N)}(n) = \sup_{f\in B_i} \frac{\max(S_n(f - g_i + \mathbb{E}[g_i]), 0)}{\sqrt{N + S_n(f_+^2)}}$$

with n in $[1, N]$. This will be achieved via Theorem 2.1, exactly as in Section 2. We set $\Delta^{(N)} = \Delta^{(N)}(N)$.

THEOREM B.1. *Assume that $\mathcal{F}$ is a countable class of centered and normalized functions, totally bounded in $L^2(P)$. In addition, assume that $\mathcal{F}$ satisfies the bracketing condition* (3.1). *Let*

$$E_N^\delta = \sup_{n\in[1,N]} \mathbb{E}[\Delta^{(N)}(n)]$$

and assume that $\sup_{N>0} E_N^\delta < \infty$. *Then, one can find some positive ε and N_0 only depending on η given by* (2.1) *and E_N^δ such that, for $N \ge N_0$ and for any λ in* $[0, \varepsilon\sqrt{N}]$,

$$\mathbb{E}\big[\exp\big(\lambda\Delta^{(N)}\big)\big] \le \exp(\lambda E_N^\delta + 16|\log\delta|^{-1}\lambda^2). \tag{B.1}$$

PROOF. The proof follows essentially the same line as the one of Theorem 2.2. For the sake of brevity, rewrite $\Delta^{(N)}(n)$ as Δ and denote by F_n the Laplace transform of Δ. Define

$$\Delta_k = \sup_{f \in B_i} \frac{\max(S_n^k(f - g_i) + n\mathbb{E}[g_i], 0)}{\sqrt{N + S_n^k(f_+^2)}}.$$

We may assume without loss of generality that $\mathcal{F}$ is finite, that is, $\mathcal{F} = \{f_1, \dots, f_m\}$ with $m \geq 1$. One the one hand, for the positive part of $\Delta - \Delta_k$, let τ denote the infimum of integers j such that

$$\Delta = \frac{\max(S_n(f_j - g_i) + n\mathbb{E}[g_i], 0)}{\sqrt{N + S_n(f_{j+}^2)}}.$$

Clearly, we may assume $\Delta > 0$, which ensures that $S_n(f_\tau - g_i) + n\mathbb{E}[g_i] > 0$. If $S_n^k(f_\tau - g_i) + n\mathbb{E}[g_i] \leq 0$, then $(f_\tau - g_i)(X_k) > 0$ and

$$\Delta - \Delta_k \leq \frac{S_n(f_\tau - g_i) + n\mathbb{E}[g_i]}{\sqrt{N + S_n(f_{\tau+}^2)}} \leq \frac{(f_\tau - g_i)_+(X_k)}{\sqrt{N + S_n(f_{\tau+}^2)}}.$$

Otherwise,

$$\Delta_k \geq \frac{S_n^k(f_\tau - g_i) + n\mathbb{E}[g_i]}{\sqrt{N + S_n^k(f_{\tau+}^2)}} \geq \frac{S_n^k(f_\tau - g_i) + n\mathbb{E}[g_i]}{\sqrt{N + S_n(f_{\tau+}^2)}} > 0,$$

and consequently

$$\Delta - \Delta_k \leq \frac{(f_\tau - g_i)_+(X_k)}{\sqrt{N + S_n(f_{\tau+}^2)}}.$$

Noting that for any positive x,

$$\psi(x) \leq \frac{x^2}{2}\mathbb{1}_{x \leq 1} + x\mathbb{1}_{x > 1},$$

we deduce that for all $\lambda \geq 0$,

$$\psi\big(\lambda(\Delta - \Delta_k)_+\big) \leq I_k(\lambda) + J_k(\lambda) \tag{B.2}$$

with

$$I_k(\lambda) = \frac{\lambda^2 (f_\tau - g_i)_+^2(X_k)}{2N}\mathbb{1}_{\lambda(\Delta - \Delta_k)_+ \leq 1},$$
$$J_k(\lambda) = \frac{\lambda (f_\tau - g_i)_+(X_k)}{\sqrt{S_n(f_{\tau+}^2)}}\mathbb{1}_{\lambda(\Delta - \Delta_k)_+ > 1}.$$

Furthermore, from condition (3.1), if $g_i \neq 0$ then g_i, f_τ and h_i are simultaneously positive or negative so that $(f_\tau - g_i)_+ \leq (h_i - g_i)_+$. Otherwise, $g_i = 0$ so that $(f_\tau - g_i)_+ \leq |h_i|$. In addition, if $f_\tau > g_i$, one can easily see that necessarily $g_i \geq 0$. Therefrom

$$I_k(\lambda) \leq \frac{\lambda^2 (h_i - g_i)^2 (X_k)}{2N} \mathbb{1}_{\lambda\Delta \leq 1 + \lambda\Delta_k}.$$

Consequently,

$$\sum_{k=1}^n \mathbb{E}[I_k(\lambda) e^{\lambda\Delta}] \leq \frac{e\lambda^2}{2N} \sum_{k=1}^n \mathbb{E}[(h_i - g_i)^2 (X_k) e^{\lambda\Delta_k}] \leq \frac{\delta e \lambda^2}{2} \mathbb{E}[\exp(\lambda\Delta_n)]$$

since $\mathbb{E}[(h_i - g_i)^2(X)] \leq \delta$, Δ_k is $\mathcal{F}_n^k$-measurable and

$$\mathbb{E}[\exp(\lambda\Delta_k)] = \mathbb{E}[\exp(\lambda\Delta_n)].$$

Next, the Laplace transform of Δ_n differs from F_{n-1}. However,

$$\Delta_n \leq \Delta^{(N)}(n-1) + N^{-1/2} |\mathbb{E}[g_i]|. \tag{B.3}$$

Then, as $|\mathbb{E}[g_i]| \leq \sqrt{\delta}$, it follows from (B.3) that

$$\mathbb{E}[\exp(\lambda\Delta_n)] \leq \exp\biggl(\frac{\lambda\sqrt{\delta}}{\sqrt{N}}\biggr) F_{n-1}(\lambda).$$

Hence, for any λ in $]0, \sqrt{N}]$, we obtain that

$$\sum_{k=1}^n \mathbb{E}[I_k(\lambda) e^{\lambda\Delta}] \leq \frac{\delta e^2 \lambda^2}{2} F_{n-1}(\lambda).$$

Next, for the second term $J_k(\lambda)$, we have

$$J_k(\lambda) \leq \frac{\lambda f_{\tau+}(X_k)}{\sqrt{S_n(f_{\tau+}^2)}} \mathbb{1}_{\lambda(h_i - g_i)(X_k) > \sqrt{N}}.$$

Applying the Cauchy–Schwarz inequality, we find

$$\sum_{k=1}^n J_k(\lambda) \leq \lambda M_n(\lambda) \qquad \text{with } M_n(\lambda) = \sqrt{S_n\bigl(\mathbb{1}_{\lambda(h_i - g_i) > \sqrt{N}}\bigr)},$$

which leads to

$$\sum_{k=1}^n \mathbb{E}[J_k(\lambda) e^{\lambda\Delta}] \leq \lambda \mathbb{E}[M_n(\lambda) \exp(\lambda\Delta)].$$

Piecing together these two contributions, we obtain from (B.2) that

$$\sum_{k=1}^n \mathbb{E}\bigl[\psi\bigl(\lambda(\Delta - \Delta_k)_+\bigr) e^{\lambda\Delta}\bigr] \leq \frac{\delta e^2 \lambda^2}{2} F_{n-1}(\lambda) + \lambda \mathbb{E}[M_n(\lambda) \exp(\lambda\Delta)]. \tag{B.4}$$

By the duality variational formula for the entropy (see, e.g., [22], inequality (6) with $y = 1/2$), we have

$$2\mathbb{E}[\lambda M_n(\lambda)\exp(\lambda\Delta)] \le \lambda F_n' - F_n \log F_n + F_n \log \mathbb{E}\big[\exp\big(2\lambda M_n(\lambda)\big)\big]. \tag{B.5}$$

Now, for any $\beta > 0$,

$$2\lambda M_n(\lambda) \le \beta^{-1}\lambda^2 + \beta M_n^2(\lambda).$$

Moreover $M_n^2(\lambda)$ has a binomial $B(n, p)$ distribution with $p \le \delta\lambda^2/N$. Consequently,

$$\log \mathbb{E}\big[\exp\big(2\lambda M_n(\lambda)\big)\big] \le \beta^{-1}\lambda^2 + np(e^\beta - 1) \le \lambda^2\big(\beta^{-1} + \delta(e^\beta - 1)\big).$$

Choosing $\beta = \log(1 + \delta^{-1/2})$ yields

$$\log \mathbb{E}\big[\exp\big(2\lambda M_n(\lambda)\big)\big] \le \frac{2\lambda^2}{\log(1 + \delta^{-1/2})} \le \frac{4\lambda^2}{|\log\delta|}.$$

Finally, it follows from (B.4) and (B.5) that for any λ in $]0, \sqrt{N}]$,

$$\sum_{k=1}^{n} \mathbb{E}\big[\psi\big(\lambda(\Delta - \Delta_k)_+\big)e^{\lambda\Delta}\big] \le \frac{\delta e^2\lambda^2 F_{n-1}}{2} + \frac{2\lambda^2 F_n}{|\log\delta|} + \frac{1}{2}(\lambda F_n' - F_n \log F_n). \tag{B.6}$$

On the other hand, for the negative part of $\Delta_k - \Delta$, let τ_k be the infimum of integers j such that

$$\Delta_k = \frac{\max(S_n^k(f_j - g_i) + n\mathbb{E}[g_i], 0)}{\sqrt{N + S_n^k(f_{j+}^2)}}.$$

Proceeding exactly as in Section 2 with $Y_k = f_{\tau_k+}^2(X_k)$, we find that

$$\begin{aligned}\phi\big(\lambda(\Delta - \Delta_k)_-\big) &\le \frac{\lambda^2}{2N}(f_{\tau_k} - g_i)_-^2(X_k) + \phi\left(\frac{\lambda\Delta_k Y_k}{2N + Y_k}\right),\\ &\le \frac{\lambda^2}{2N}(h_i - g_i)^2(X_k) + \phi\left(\frac{\lambda\Delta_k Y_k}{2N + Y_k}\right).\end{aligned}$$

In addition, (B.3) immediately implies that $\Delta_k \le 2\sqrt{N}$. Similarly to (2.7), we deduce that for α in $]0, 1[$, $N \ge 1/(2\alpha)$ and λ in $]0, \alpha\sqrt{N}]$,

$$\sum_{k=1}^{n} \mathbb{E}\big[\phi\big(\lambda(\Delta - \Delta_k)_-\big)e^{\lambda\Delta_k}\big] \le \frac{\delta e\lambda^2}{2} F_{n-1}(\lambda) + \varepsilon^*(\alpha)\mathbb{E}\left[\phi(-\lambda\Delta_n)\right]. \tag{B.7}$$

The function ϕ is nonincreasing on $\mathbb{R}^-$. Thus, if $\Delta_n \le \Delta^{(n-1)}$,

$$\phi(-\lambda\Delta_n) \le \phi(-\lambda\Delta^{(n-1)}).$$

Otherwise, ϕ is convex on $\mathbb{R}^-$ and $\phi'(x) = x\exp(-x)$. Hence, for any x, y in $\mathbb{R}^+$

$$\phi(-x-y) \le \phi(-x) + y(x+y)\exp(x+y).$$

Applying this elementary inequality to $x = \lambda\Delta^N(n-1)$ and $y = \lambda(\Delta_n - \Delta^N(n-1))$, we get from (B.3) that

$$\phi(-\lambda\Delta_n) \le \phi\big(-\lambda\Delta^N(n-1)\big) + \lambda^2 \frac{\sqrt{\delta}}{\sqrt{N}}\Delta_n \exp(\lambda\Delta_n).$$

Consequently, for any λ in $]0, \alpha\sqrt{N}]$,

$$\phi(-\lambda\Delta_n) \le \phi\big(-\lambda\Delta^N(n-1)\big) + 2e\sqrt{\delta}\lambda^2 \exp\big(\lambda\Delta^N(n-1)\big).$$

Hence

$$\mathbb{E}[\phi(-\lambda\Delta_n)] \le \mathbb{E}\big[\phi\big(-\lambda\Delta^N(n-1)\big)\big] + 2e\sqrt{\delta}\lambda^2 F_{n-1}(\lambda). \tag{B.8}$$

In addition, one can observe that

$$\mathbb{E}\big[\phi\big(-\lambda\Delta^N(n-1)\big)\big] = \lambda F'_{n-1}(\lambda) - F_{n-1}(\lambda) + 1.$$

From the conjunction of (B.6), (B.7) and (B.8), we obtain via Theorem 2.1 that for $N \ge 1/(2\alpha)$ and λ in $]0, \alpha\sqrt{N}]$,

$$\lambda F'_n - F_n \log F_n \le e(2+e)\sqrt{\delta}\lambda^2 F_{n-1} + \frac{4\lambda^2 F_n}{|\log\delta|} + 2\varepsilon^*(\alpha)(\lambda F'_{n-1} - F_{n-1} + 1),$$

provided that $\varepsilon^*(\alpha) \le 1/4$. Finally, as $-x\log x \le e^{-1}$ for x in $]0, 1[$, we find that

$$\lambda F'_n - F_n \log F_n \le \frac{5\lambda^2}{|\log\delta|}(F_n + 2F_{n-1}) + 2\varepsilon^*(\alpha)(\lambda F'_{n-1} - F_{n-1} + 1). \tag{B.9}$$

We are now in position to state our induction hypothesis. Hereafter, assume that $\delta \le 1/e$. The induction hypothesis $\mathcal{H}(n)$ at range n is that for any λ in $]0, \alpha\sqrt{N}]$,

$$\begin{cases} F_n(\lambda) < \exp(\lambda E_N + 16\,|\log\delta|^{-1}\lambda^2), \\ \lambda F'_n(\lambda) - F_n(\lambda) + 1 \le C\lambda^2 \exp(\lambda E_N + 16\,|\log\delta|^{-1}\lambda^2). \end{cases}$$

At range 0, we assume that $F_0(\lambda) = 1$. Hence the induction hypothesis holds true at range 0. Let n be some integer in $[1, N]$. Suppose that the induction hypothesis holds at range $n-1$. Set

$$H(\lambda) = \exp(\lambda E_N + 16|\log\delta|^{-1}\lambda^2).$$

Then, we find via (B.9) that

$$\lambda F'_n - F_n \log F_n < \frac{5\lambda^2}{|\log\delta|}(F_n + 2H) + 2\varepsilon^*(\alpha)C\lambda^2 H. \tag{B.10}$$

Consequently, F_n is a subsolution of the Differential Equation (DE) corresponding to the equality in (B.10) with $F_n(0) = 1$ and $F_n'(0) = \mathbb{E}[Z] \le E_N$. Now, we have

$$\lambda H'(\lambda) - H(\lambda)\log H(\lambda) = 16|\log\delta|^{-1}\lambda^2 H(\lambda).$$

Therefrom, by the comparison lemma in Arnold [2], we obtain that for all n in $[1, N]$ and for any λ in $]0, \alpha\sqrt{N}]$, $F_n(\lambda) < H(\lambda)$ provided that $2\varepsilon^*(\alpha)C \le |\log\delta|^{-1}$. To bound up $\lambda F_n' - F_n + 1$, one has to use exactly the same arguments as in Section 2. Consequently, the end of the proof will be omitted. Finally, by induction, $\mathcal{H}(N)$ holds and the result follows. $\square$

PROOF OF THEOREM 3.1. The proof is similar to the one of Theorem 2.4. First of all, we clearly have

$$V_n^i(\delta) \le \sup_{f\in B_i} \frac{\max(S_n(f - g_i + \mathbb{E}[g_i]), 0)}{\sqrt{n + S_n(f_+^2)}}\left(\frac{n}{S_n(f^2)} + 1\right)^{1/2}.$$

Next, for $\alpha > \sqrt{2}$, set $a_\alpha = (\alpha^2 - 1)^{-1}$ and

$$\Gamma_{\alpha,i} = \left\{\inf_{f\in B_i} S_n(f^2) \ge a_\alpha n\right\}.$$

On $\Gamma_{\alpha,i}$, as $V_n^i(\delta) \le \alpha\Delta$ with $\Delta = \Delta^{(n)}(n)$, we derive that

$$\mathbb{E}\big[\exp(\lambda V_n^i(\delta))\big] \le 2\max\big(F_n(\alpha\lambda), \exp\big(\lambda(1 + \sqrt{\delta})\sqrt{n}\big) P(\Gamma_{\alpha,i}^c)\big) \tag{B.11}$$

where F_n denotes the Laplace transform of Δ. We have already seen from Theorem B.1 that for $\lambda < \varepsilon\sqrt{n}/\alpha$,

$$F_n(\alpha\lambda) \le \exp(\alpha\lambda E_n^\delta + 16|\log\delta|^{-1}\alpha^2\lambda^2). \tag{B.12}$$

In addition, $E_n^\delta \le 1 + E$. Moreover,

$$\Gamma_{\alpha,i}^c \subset \{S_n(g_i^2) \le n a_\alpha\}.$$

Next, the bracketing condition (3.1) implies that $\mathbb{E}[g_i^2] \ge 1 - 2\sqrt{\delta}$ for any $i \in I$. Then, applying Proposition A.1, we deduce that

$$\mathbb{P}(\Gamma_{\alpha,i}^c) \le \exp(-n\theta_\alpha)$$

with $\theta_\alpha > 0$ provided that $a_\alpha < 1 - 2\sqrt{\delta}$, that is, $\delta < (2\alpha^2 - 2)^{-2}(\alpha^2 - 2)^2$. Consequently, we find from (B.11) together with (B.12) that for all $\lambda < \varepsilon\sqrt{n}/\alpha$,

$$E\big[\exp(\lambda V_n^i(\delta))\big] \le 2\exp\big(\alpha\lambda(1 + E) + 16|\log\delta|^{-1}\alpha^2\lambda^2\big) \tag{B.13}$$

as soon as $\exp(\lambda(1 + \sqrt{\delta})\sqrt{n} - n\theta_\alpha) \le 1$, that is, $\lambda < (1 + \sqrt{\delta})^{-1}\sqrt{n}\theta_\alpha$, which completes the proof of Theorem 3.1. $\square$

APPENDIX C

PROOF OF LEMMA 4.1. Let j be some integer in $[0, N]$. If $f_\gamma = 0$ for some γ in I_j, then $g_j = 0$. Otherwise, from the continuity of $\gamma \to f_\gamma$, the sign of f_γ is constant for γ in I_j and $f_\gamma g_j > 0$. Let τ be the infimum of integers j in $[0, N]$ such that the supremum $\mathcal{W}_n$ is realized on I_j. On the one hand, suppose that $\tau \neq 0$.

Then, it follows from (4.1) that

$$|g_\tau(x)| = \inf_{\gamma \in I_\tau} |f_\gamma(x)| \tag{C.1}$$

and, similarly to (3.4), for any γ in I_τ,

$$W_n(f_\gamma) - \max\bigl(W_n(g_\tau), 0\bigr) \leq \frac{P_n(f_\gamma - g_\tau)}{\sqrt{P_n(f_\gamma^2)}}. \tag{C.2}$$

On the other hand, suppose $\tau = 0$. Let us introduce a new parametrization of the class on I_0. In the sequel, for γ in I_0 with $\gamma \neq 0$, we denote by f_γ the function

$$f_\gamma(x) = \gamma^{-1}\bigl(\exp\bigl(\gamma t(x) - l(\gamma)\bigr) - 1\bigr).$$

For the sake of brevity, write $\varepsilon = a_1$. If $t(x)$ belongs to $[0, h(\varepsilon)]$, then $g_0(x) = 0$. Otherwise, either $t(x) > h(\varepsilon)$ or $t(x) < 0$. In the first case, $f_\gamma(x)$ is positive and by convexity of the exponential function,

$$f_\gamma(x) \geq \gamma^{-1}\bigl(\exp\bigl(\gamma\bigl(t(x) - h(\varepsilon)\bigr)\bigr) - 1\bigr) \geq t(x) - h(\varepsilon).$$

In the second case, $f_\gamma(x) < 0$ and

$$-f_\gamma(x) \geq \gamma^{-1}\bigl(1 - \exp\bigl(\gamma t(x)\bigr)\bigr) \geq \varepsilon^{-1}\bigl(1 - \exp\bigl(\varepsilon t(x)\bigr)\bigr).$$

Therefrom, if $g_0(x) \neq 0$, then $g_0(x)$ has the sign of $f_\gamma(x)$ and

$$\inf_{\gamma \in I_0} |f_\gamma(x)| \geq |g_0(x)|. \tag{C.3}$$

With this new parametrization on I_0, we still have (C.2) for $\tau = 0$. Consequently,

$$\mathcal{W}_n - \max\bigl(W_n(g_0), \dots, W_n(g_N), 0\bigr) \leq \max(\Delta_0, \Delta_1, \dots, \Delta_N) \tag{C.4}$$

where

$$\Delta_j = \sup_{\gamma \in I_j} \frac{P_n(|f_\gamma - g_j|)}{\sqrt{P_n(f_\gamma^2)}}. \tag{C.5}$$

We now prove that the upper bound in (C.4) is exponentially negligible. By the Cauchy–Schwarz inequality,

$$\Delta_j \leq \sup_{\gamma \in I_j} \sqrt{P_n\bigl((1 - g_j f_\gamma^{-1})^2 \mathbb{1}_{f_\gamma \neq 0}\bigr)},$$

which implies that

$$\Delta_j \le \sqrt{P_n(m_j)} \qquad \text{with } m_j = 1 - \frac{|g_j|}{\sup_{\gamma \in I_j} |f_\gamma|}. \tag{C.6}$$

In order to control $P_n(m_j)$, it will be convenient to bound up the expectation of $m_j(X)$. Suppose first that $j \neq 0$. Let γ_0 and γ_1 be the elements of I_j such that $|g_j(x)| = |f_{\gamma_0}(x)|$ and $\sup_{\gamma \in I_j} |f_\gamma(x)| = |f_{\gamma_1}(x)|$. Then

$$m_j(x) = 1 - \frac{|f_{\gamma_0}(x)|}{|f_{\gamma_1}(x)|} \le \int_{I_j} \left| \frac{d}{d\gamma} \log |f_\gamma(x)| \right| d\gamma.$$

Let A_j^ξ be the set of reals x such that $|g_j(x)| < \xi a_j$ where $\xi = \sqrt{\delta}$. As

$$\frac{d}{d\gamma} \log |f_\gamma(x)| = \bigl(t(x) - l'(\gamma)\bigr)\left(1 + \frac{1}{f_\gamma(x)}\right),$$

we obtain that for any $x \notin A_j^\xi$,

$$\left| \frac{d}{d\gamma} \log |f_\gamma(x)| \right| \le 2(\xi a_j)^{-1} |t(x) - l'(\gamma)|.$$

Consequently, since $|I_j| = \delta a_j$,

$$\mathbb{E}[m_j(X)] \le \mathbb{P}(X \in A_j^\xi) + 2\sqrt{\delta}\bigl(\mathbb{E}[|t(X)|] + l'(1)\bigr). \tag{C.7}$$

We now study the set A_j^ξ. Since $|\gamma| \ge a_j$ for γ in I_j, we have, by (4.1),

$$|g_j(x)| \ge 1 - \exp\Bigl(-a_j \inf_{\gamma \in I_j} |h(\gamma) - t(x)|\Bigr).$$

Therefore, A_j^ξ is included in the set of reals x such that

$$\inf_{\gamma \in I_j} |h(\gamma) - t(x)| \le a_j^{-1} |\log(1 - a_j \xi)|. \tag{C.8}$$

Let Q denote the maximal concentration function of the real random variable $t(X)$, defined by

$$Q(x) = \sup_{y \in \mathbb{R}} \mathbb{P}\bigl(y \le t(X) \le y + x\bigr).$$

We infer from (C.8) and the concavity of the logarithm function that

$$\mathbb{P}(X \in A_j^\xi) \le \kappa \qquad \text{with } \kappa = Q\bigl(2|\log(1 - \xi)| + \delta l'(1)\bigr).$$

Hence from (C.7), we get that

$$\mathbb{E}[m_j(X)] \le \kappa + 2\sqrt{\delta}\bigl(\mathbb{E}[|t(X)|] + l'(1)\bigr) = \zeta. \tag{C.9}$$

We now bound the Laplace transform of Δ_j. By (C.6), the convexity of the exponential function and the fact that $0 \le m_j(X) \le 1$,

$$\mathbb{E}[\exp(t\Delta_j^2)] \le \exp(n\zeta(e^t - 1)).$$

Using the Chernoff calculation and recalling that $N = n^2$, we infer that

$$\text{(C.10)} \qquad \frac{1}{n}\log \mathbb{P}(\max(\Delta_1, \dots, \Delta_N) \ge \alpha) \le \frac{2}{n}\log n + \alpha^2 - \alpha^2 \log\left(\frac{\alpha^2}{\zeta}\right).$$

It remains to prove a similar bound for $\mathbb{P}(\Delta_0 \ge \alpha)$. From (C.6), we have

$$\frac{1}{n}\log \mathbb{P}(\Delta_0 \ge \alpha) \le \alpha^2 - \alpha^2 \log\left(\frac{\alpha^2}{\mathbb{E}[m_0(X)]}\right).$$

Furthermore, the function m_0 takes its values in $[0, 1]$ and converges pointwise to 0 as n tends to ∞, provided that $t(x) \ne 0$. Hence, by the Lebesgue dominated convergence theorem,

$$\text{(C.11)} \qquad \lim_{n\to\infty} \mathbb{E}[m_0(X)] = 0 \quad \text{and} \quad \limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\Delta_0 \ge \alpha) = -\infty$$

for any positive δ. Finally, Lemma 4.1 follows from (C.10) and (C.11) together with the fact that $\log(\alpha^2/\zeta)$ tends to ∞ as δ tends to 0. □

CONTINUITY OF THE RATE FUNCTION $\mathcal{I}$. In order to perform a continuous parametrization of the class $\mathcal{F}$, we set $r_\gamma = \gamma^{-1} f_\gamma$ for $\gamma > 0$. Let

$$Z(a, \gamma, x) = ar_\gamma(X) - x(r_\gamma^2(X) + a^2)/2, \qquad F(t, a, \gamma, x) = \mathbb{E}[\exp(tZ(a, \gamma, x))]$$

and

$$R(a, \gamma, x) = \inf_{t\ge 0} \mathbb{E}[\exp\{t(ar_\gamma(X) - x(r_\gamma^2(X) + a^2)/2)\}].$$

Following the approach of Shao [23], Lemma 8.1, we obtain that for any $d > 0$, one can find $A > 0$ such that, for any x in $[d, 1]$,

$$\text{(C.12)} \qquad \mathcal{I}(x) = -\log \sup_{\gamma\in[0,1]} \sup_{a\in[0,A]} R(a, \gamma, x).$$

The continuity of $\mathcal{I}$ immediately follows from the lemma below. □

LEMMA C.1. *Assume that $t(X)$ has a continuous distribution function. Then, for any $A > 0$ and any $d > 0$, R is continuous on* $[0, A] \times [0, 1] \times [d, 1]$.

PROOF. Clearly, R is upper semicontinuous. Hence we only have to prove that for any nonnegative y, $(R > y)$ is an open set. Let $b_0 = (a_0, \gamma_0, x_0)$ in $[0, A] \times [0, 1] \times [d, 1]$ and assume that $R(b_0) > y$. Using the fact that $Z(b_0)$ is an analytic function of $t(X)$, we obtain that $\mathbb{P}(Z(b_0) = 0) = 0$. Hence there exists

$s > 0$ such that $\mathbb{P}(Z(b_0) \geq 2s) > 0$. Thus, one can find a neighborhood V of b_0 and some positive δ such that $\mathbb{P}(Z(b) \geq s) > \delta$ for any b in V. Consequently, choosing $T = |\log\delta|/s$, we obtain that $F(T, b) > 1$ for any b in V. As $F(0, b) = 1$, we deduce that $R(b) = \inf\{F(t, b) : t \in [0, T]\}$ for any b in V, which implies that $(R > y)$ is an open set. □

PROOF OF LEMMA 4.2. We first bound up the random variables $W_n(g_j)$ for $j \neq 0$. Let $K = [0, l'(1)]$. If $t(x) > l'(1)$, then

$$\frac{d}{d\gamma} f_\gamma(x) = \bigl(t(x) - l'(\gamma)\bigr)\exp\bigl(\gamma\bigl(t(x) - h(\gamma)\bigr)\bigr) > 0$$

for any γ in $]0, 1]$. Hence $f_\gamma(x) > 0$. In the same way, if $t(x) < 0$, then $\frac{d}{d\gamma} f_\gamma(x) < 0$ and $f_\gamma(x) < 0$. Hence $g_j(x) = f_{a_j}(x)$ for $t(x)$ not in K. For the sake of brevity, rewrite f_{a_j} as f_j. Noting that $\sqrt{P_n(f^2)}$ is an L^2 pseudonorm, we get

$$\begin{aligned}\sqrt{P_n(f_j^2)}\bigl(W_n(g_j) - W_n(f_j)\bigr) &\leq W_n(g_j)\bigl(\sqrt{P_n(f_j^2)} - \sqrt{P_n(g_j^2)}\bigr) + P_n(|g_j - f_j|)\\ &\leq 2\sqrt{P_n\bigl((f_j - g_j)^2\bigr)},\end{aligned}$$

so that

$$W_n(g_j) - W_n(f_j) \leq 2\sqrt{\frac{P_n((f_j - g_j)^2 \mathbb{1}_{t(x)\in K})}{P_n(f_j^2)}}. \tag{C.13}$$

If $t(x)$ belongs to K, then

$$|f_j(x) - g_j(x)| \leq \int_{I_j} |t(x) - l'(\gamma)| \exp\bigl(\gamma\bigl(t(x) - h(\gamma)\bigr)\bigr)\,\mathrm{d}\gamma \leq \delta a_j l'(1) e^{l'(1)}.$$

Let $C = 2l'(1)\exp(l'(1))$ and set $\xi = \sqrt{\delta}$. It follows from (C.13) that

$$\mathbb{P}\bigl(W_n(g_j) \geq \beta\bigr) \leq \mathbb{P}\bigl(W_n(f_j) \geq \beta - \xi\bigr) + \mathbb{P}\bigl(P_n(|r_{a_j}|) \leq C\xi\bigr). \tag{C.14}$$

In order to bound from below the random variables $P_n(|r_\gamma|)$, we bound up the concentration function of r_γ around 0. Noting that $|r_\gamma(x)| \leq y$ if and only if

$$h(\gamma) + \gamma^{-1}\log(1 - \gamma y) \leq t(x) \leq h(\gamma) + \gamma^{-1}\log(1 + \gamma y)$$

and recalling that $\varphi(x) = \log(1 + x) - \log(1 - x)$ is a convex function, we have

$$\mathbb{P}(|r_\gamma(X)| \leq y) \leq Q\bigl(\gamma^{-1}\varphi(\gamma y)\bigr) \leq Q\bigl(\varphi(y)\bigr) = G(y). \tag{C.15}$$

Now G is the distribution function of a random variable V with values in $[0, 1]$. Denote by L^* the Legendre transform of the log-Laplace of V. Since G is continuous at point 0, $L^*(x)$ goes to infinity as x tends to 0. Furthermore, by (C.15), for any γ in $]0, 1]$,

$$\mathbb{P}\big(P_n(|r_\gamma|) \le x\big) \le \exp\big(-nL^*(x)\big). \tag{C.16}$$

Using the same arguments, we bound up $W_n(g_0)$. First,

$$W_n(g_0) - W_n(t) \le 2\sqrt{\frac{P_n((t-g_0)^2)}{P_n(t^2)}}. \tag{C.17}$$

Next, observe that if $t(x) \ge 0$, then $|t(x) - g_0(x)| \le h(a_1)$. In addition, if $-a_1^{-1/2} \le t(x) < 0$,

$$|t(x) - g_0(x)| \le \frac{1}{a_1}\psi\big(-a_1 t(x)\big) \le \frac{a_1 t^2(x)}{2} \le \frac{\sqrt{a_1}}{2}|t(x)|.$$

Consequently,

$$\sqrt{\frac{P_n((t-g_0)^2)}{P_n(t^2)}} \le \frac{h(a_1)}{P_n(|t|)} + \frac{\sqrt{a_1}}{2}$$

as soon as $t(X_i) \ge -a_1^{-1/2}$ for any i in $[1, n]$. Now it is easy to check that $t(X)$ satisfies the concentration bound (C.15) so that

$$\mathbb{P}\big(P_n(|t|) \le \sqrt{a_1}\big) \le \exp\bigg(-nL^*\bigg(\frac{2h(a_1)}{\sqrt{a_1}}\bigg)\bigg). \tag{C.18}$$

Hence

$$\mathbb{P}\big(W_n(g_0) \ge \beta\big) \le \mathbb{P}\big(W_n(t) \ge \beta - 2\sqrt{a_1}\big) + \exp\bigg(-nL^*\bigg(\frac{2h(a_1)}{\sqrt{a_1}}\bigg)\bigg) + n\mathbb{E}[t^2(X)]a_1$$

where we recall that $a_1 = (1+\xi^2)^{-n^2}$. Consequently, as the rate function I_t is continuous (see Lemma 8.1 of [23]), it implies that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(W_n(g_0) \ge \beta\big) \le -I_t(\beta). \tag{C.19}$$

Finally, Lemma 4.2 follows from the conjunction of (C.14), (C.16), (C.19), Lemma below and the continuity of the rate function. □

APPENDIX D

LEMMA D.1. *For any positive x*,

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\gamma\in]0,1]} \mathbb{P}\big(W_n(f_\gamma) \ge x\big) \le -\mathcal{I}(x).$$

PROOF. Let A be some real satisfying $A > l(1)$. Define the functions $\overline{t}(x) = \inf\{t(x), A\}$, $\overline{f}_\gamma(x) = \exp(\gamma\overline{t}(x) - l(\gamma)) - 1$ and $\overline{r}_\gamma = \gamma^{-1}\overline{f}_\gamma$. For the sake of brevity, we rewrite f_γ and r_γ as f and r, respectively. First,

$$W_n(f) \le \frac{\max(P_n(\overline{f}), 0)}{\sqrt{P_n(f^2)}} + \frac{P_n(f - \overline{f})}{\sqrt{P_n(f^2)}}.$$

If $t(x) \le A$, then $f - \overline{f} = 0$. Otherwise $0 \le f - \overline{f} \le f$ as $A > l(1)$. Hence, by the Cauchy–Schwarz inequality,

$$\frac{P_n(f - \overline{f})}{\sqrt{P_n(f^2)}} \le \sqrt{P_n(\mathbb{1}_{t>A})}.$$

Next, recalling that $P_n(\overline{r}) \le P_n(r)$, we get for any positive real δ

$$\frac{\max(P_n(\overline{r}), 0)}{\sqrt{P_n(r^2)}} - \frac{\max(P_n(\overline{r}), 0)}{\sqrt{\delta^2 + P_n(r^2)}} \le \frac{\delta^2}{2P_n(r^2)}.$$

Consequently, if $\overline{W}_n(r, \delta) = \frac{P_n(\overline{r})}{\sqrt{\delta^2 + P_n(r^2)}}$, we deduce that for any $0 \le y \le x/2$,

$$\text{(D.1)} \quad \mathbb{P}\big(W_n(f) \ge x\big) \le D_n(x - 2y) + \mathbb{P}\left(P_n(r^2) \le \frac{\delta^2}{2y}\right) + \mathbb{P}\big(P_n(\mathbb{1}_{t>A}) \ge y\big)$$

with $D_n(z) = \mathbb{P}(\overline{W}_n(r, \delta) \ge z)$. In order to bound up $D_n(z)$, we need the following lemma.

LEMMA D.2. *For any ε and x in* $]0, 1]$,

$$\mathbb{P}\left(\frac{P_n(\overline{r}_\gamma)}{\sqrt{P_n(r_\gamma^2) + \varepsilon}} \ge x\right) \le \frac{e^{2A}}{x^2\varepsilon}\exp\big(-nI_{f_\gamma}(x)\big).$$

PROOF. We shall follow the same approach as that of Shao [23]. Exactly as in [23], equation (2.3), page 288, we have

$$\text{(D.2)} \quad \mathbb{P}\left(\frac{P_n(\overline{r})}{\sqrt{P_n(r^2) + \varepsilon}} \ge x\right) = \mathbb{P}\left(\sup_{b\ge 0} 2bP_n(\overline{r}) - x\big(b^2 + \varepsilon + P_n(r^2)\big) \ge 0\right).$$

Now $\overline{r} \le e^A - 1$ so that $P_n(\overline{r}) \le e^A - 1$. Hence the supremum in (D.2) is realized in the interval $J = [0, C]$ with $C = x^{-1}(e^A - 1)$. For k positive integer, set

$$J_k = \big[\sqrt{(k-1)\varepsilon}, \sqrt{k\varepsilon}\,\big], \qquad K = \big[\varepsilon^{-1}x^{-2}e^{2A}\big]$$

and $b_k = \sup J_k$. Since the union of J_k for k in $[1, K]$ covers J, we get from (D.2)

$$\mathbb{P}\left(\frac{P_n(\overline{r})}{\sqrt{P_n(r^2)+\varepsilon}} \geq x\right) \leq \sum_{k=1}^{K} \mathbb{P}\big(2b_k P_n(\overline{r}) - x\big(b_k^2 + P_n(r^2)\big) \geq 0\big)$$
$$\leq \sum_{k=1}^{K} \inf_{t\geq 0} \mathbb{E}\big[\exp\big(2tb_k P_n(\overline{r}) - tx\big(b_k^2 + P_n(r^2)\big)\big)\big]$$
$$\leq K \sup_{b\geq 0} \inf_{t\geq 0} \mathbb{E}\big[\exp\big(2tbP_n(r) - tx\big(b^2 + P_n(r^2)\big)\big)\big],$$

as $P_n(\overline{r}) \leq P_n(r)$, which implies Lemma D.2. □

Starting from (D.1) and Lemma D.2, we now complete the proof of Lemma D.1. By use of (C.15), we have

$$\frac{1}{n}\log \mathbb{P}\left(P_n(r^2) \leq \frac{\delta^2}{2y}\right) \leq -L^*\left(\frac{\delta}{\sqrt{2y}}\right).$$

In addition, if we take $A = l(1) + \delta^{-1}$, from Markov's inequality applied to $\exp(t(X))$, we obtain that $\mathbb{P}(t(X) > A) \leq \exp(-1/\delta)$, whence

$$\frac{1}{n}\log \mathbb{P}\big(P_n(\mathbb{1}_{t>A}) \geq y\big) \leq -\frac{y}{\delta} - y\log\left(\frac{y}{e}\right).$$

Furthermore, from Lemma D.2 with $\varepsilon = \delta^2$,

$$\frac{1}{n}\log D_n(x-2y) \leq \frac{2}{n}\big(l(1) + \delta^{-1} - \log\big(\delta(x-2y)\big)\big) - \mathcal{I}(x-2y).$$

Hence for any positive δ,

$$\limsup_{n\to\infty} \frac{1}{n}\log \sup_{\gamma\in]0,1]} \mathbb{P}\big(W_n(f_\gamma) \geq x\big)$$
$$\leq -\min\left(L^*\left(\frac{\delta}{\sqrt{2y}}\right),\ \frac{y}{\delta} + y\log\left(\frac{y}{e}\right), \mathcal{I}(x-2y)\right)$$

which, by the arbitrariness of δ, ensures that

$$\limsup_{n\to\infty} \frac{1}{n}\log \sup_{\gamma\in]0,1]} \mathbb{P}\big(W_n(f_\gamma) \geq x\big) \leq -\mathcal{I}(x-2y).$$

Finally, Lemma D.1 follows from the continuity of $\mathcal{I}$. □

REFERENCES

[1] ANDERSEN, N. T., GINÉ, E. OSSIANDER, M. and ZINN, J. (1988). The central limit theorem and the law of iterated logarithm for empirical processes under local conditions. *Probab. Theory Related Fields* **77** 271–305.

[2] ARNOLD, V. (1974). *Equations différentielles ordinaires.* Mir, Moscow.

[3] BARTLETT, P. and LUGOSI, G. (1999). An inequality for uniform deviations of sample averages from their means. *Statist. Probab. Lett.* **4** 55–62.

[4] DACUNHA-CASTELLE, D. and GASSIAT, E. (1997). Testing in locally conic models, and application to mixture models. *ESAIM Probab. Statist.* **1** 285–317.

[5] DACUNHA-CASTELLE, D. and GASSIAT, E. (1999). Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *Ann. Statist.* **27** 1178–1209.

[6] DEMBO, A. and SHAO, Q. M. (1998). Self-normalized large deviations in vector spaces. In *Proceedings of the Oberwolfach Meeting on High Dimensional Probability* (E. Eberlein, M. Hahn and M. Talagrand, eds.) 28–32. Birkhäuser, Boston.

[7] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. Springer, New York.

[8] DUDLEY, R. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929.

[9] DUNFORD, N. and SCHWARTZ, J. T. (1953). *Linear Operators. Part I: General Theory.* Interscience, New York.

[10] GASSIAT, E. (2002). Likelihood ratio inequalities with applications to various mixtures. *Ann. Inst. H. Poincaré Probab. Statist.* To appear.

[11] GINÉ, E. (1996). Empirical processes and applications: an overview. *Bernoulli* **2** 1–28.

[12] HAUSSLER, D. (1992). Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Inform. Comput.* **100** 78–150.

[13] KERIBIN, C. (1999). *Tests de modèles par maximum de vraisemblance*. Thèse de l'Université d'Evry-Val d'Essonne.

[14] KERIBIN, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā Ser. A* **62** 49–66.

[15] LEDOUX, M. (1996). On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.* **1** 63–87.

[16] LEDOUX, M. (1992). Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi. *Ann. Inst. H. Poincaré Probab. Statist.* **28** 267–280.

[17] MASSART, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863–884.

[18] MCDIARMID, C. (1998). Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics* 195–248. Springer, Berlin.

[19] POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

[20] POLLARD, D. (1995). Uniform ratio limit theorems for empirical processes. *Scand. J. Statist.* **22** 271–278.

[21] RIO, E. (2001). Inégalités de concentration pour les processus empiriques: Classes de parties. *Probab. Theory Related Fields* **119** 163–175.

[22] RIO, E. (2000). Inégalités exponentielles pour les processus empiriques. *C. R. Acad. Sci. Paris Sér. I* **330** 597–600.

[23] SHAO, Q. M. (1997). Self-normalized large deviations. *Ann. Probab.* **25** 285–328.

[24] TALAGRAND, M. (1996). New concentration inequalities for product spaces. *Invent. Math.* **126** 505–563.

[25] TALAGRAND, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. IHES* **81** 73–205.

[26] VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

[27] VAN DER VAART, A. (1998). *Asymptotic Statistics.* Cambridge Univ. Press.

[28] WU, L. M. (1994). Large deviations, moderate deviations and LIL for empirical processes. *Ann. Probab.* **22** 17–27.

B. BERCU
E. GASSIAT
LABORATOIRE DE MATHÉMATIQUES
UMR 8628 CNRS
BÂTIMENT 425
UNIVERSITÉ PARIS-SUD
91405 ORSAY CEDEX
FRANCE
E-MAIL: Bernard.Bercu@math.u-psud.fr
Elisabeth.Gassiat@math.u-psud.fr

E. RIO
LABORATOIRE DE MATHÉMATIQUES
UMR 8100 CNRS
BÂTIMENT FERMAT
45 AVENUE DES ETATS-UNIS
78035 VERSAILLES CEDEX
FRANCE
E-MAIL: Emmanuel.Rio@math.uvsq.fr