The Annals of Probability
1994, Vol. 22, No. 1, 28–76

SHARPER BOUNDS FOR GAUSSIAN AND EMPIRICAL PROCESSES[1]

BY M. TALAGRAND

University of Paris VI and Ohio State University

Under natural conditions on a class $\mathcal{F}$ of functions on a probability space, near optimal bounds are given for the probabilities

$$P\left(\sup_{f\in\mathcal{F}}\left|\sum_{i\le n} f(X_i) - nE(f)\right| \ge M\sqrt{n}\right).$$

The method is a variation of this author's method to study the tail probability of the supremum of a Gaussian process.

1. Introduction. Consider a probability space (Ω, Σ, P), and consider n independent identically distributed (i.i.d.) random variables $X_1, \dots, X_n$, valued in Ω, of law P. Consider a function f on Ω. (We make the convention that by "function" we mean "measurable function" and by "set" we mean "measurable set.") For n large, the quantity $\sum_{i\le n} f(X_i) - nEf$ is approximately $N(0, \sqrt{n}(Ef^2)^{1/2})$. One could say that one of the objectives of empirical process theory is to understand how well this approximation holds uniformly over a class of functions. Consider such a class of functions $\mathcal{F}$. We are interested in this paper in the quantity

$$\sup_{f\in\mathcal{F}}\left|\sum_{i\le n} f(X_i) - nEf\right|,$$

which for simplicity will be denoted by

$$\left\|\sum_{i\le n} f(X_i) - nEf\right\|_{\mathcal{F}}.$$

Observe that this need not be a r.v. (i.e., it might fail to be measurable). Measurability questions for empirical processes are, however, well understood, and in order not to waste space on these, we will assume once and for all that $\mathcal{F}$ is countable, so that no measurability problem will arise. We are interested in bounds for

$$(1.1)\qquad \tau_{\mathcal{F}}(M) = \tau_{n,\mathcal{F}}(M) = P\left(\left\|\sum_{i\le n} f(X_i) - nEf\right\|_{\mathcal{F}} \ge M\sqrt{n}\right).$$

This question has been studied in particular by Massart [10] and Alexander [1], following classical work by Kiefer [8] and Dvoretzky, Kiefer and Wolfowitz

Received March 1992; revised January 1993.
[1]Supported in part by an NSF grant.
AMS 1991 *subject classifications*. 60G50, 60E99, 62E99.
Key words and phrases. Uniform approximation, isoperimetric inequalities, tail probabilities.

[3].

Throughout the paper we will assume that $\mathscr{F}$ is uniformly bounded. One reason for this hypothesis is that it is often desirable, from the point of view of statistics, that $\mathscr{F}$ has good properties independently of the underlying probability, in which case $\mathscr{F}$ has to be uniformly bounded [2]. Another reason is that we need to have very sharp bounds when $\mathscr{F}$ consists of just one function f, and that, when f is bounded, we can appeal to the work of [7]. Assuming $\mathscr{F}$ uniformly bounded, there is then no loss of generality to assume that $\mathscr{F}$ consists only of functions f such that $0 \le f \le 1$. This allows us to control the contributions of each individual function f to (1.1). It then remains to have hypotheses that ensure that $\mathscr{F}$ is not too large. The hypotheses we will use are classical. We first recall two standard notions.

Consider a metric space (T, d). We denote by $N(T, d, \varepsilon)$ the smallest number of (open) balls of radius ε needed to cover T.

Consider now two functions f_1, f_2 on Ω. We define the bracket $[f_1, f_2]$ as

$$[f_1, f_2] = \{f;\ f_1 \le f \le f_2\}.$$

For two sets C_1, C_2, we define similarly

$$[C_1, C_2] = \{C;\ C_1 \subseteq C \subseteq C_2\}.$$

We can now state one typical result.

THEOREM 1.1. *Consider a (countable) class $\mathscr{C}$ of (measurable) subsets of Ω. Assume that there exists a number $V \ge 1$ and a number $v \ge 1$ such that either of the following holds.*

(i) *Given $\varepsilon > 0$ and any probability Q on Ω that is supported by a finite set, we have*

$$N(\mathscr{C}, d_Q, \varepsilon) \le \left(\frac{V}{\varepsilon}\right)^v, \tag{1.2}$$

where

$$d_Q(C_1, C_2) = Q(C_1 \Delta C_2). \tag{1.3}$$

(ii) *Given $\varepsilon > 0$, $\mathscr{C}$ can be covered by at most $(V/\varepsilon)^v$ brackets $[C_1, C_2]$ for which $P(C_2 \setminus C_1) \le \varepsilon$.*

Then for all $M > 0$ we have

$$\tau_{\mathscr{C}}(M) \le \frac{K(V)}{M}\left(\frac{K(V)M^2}{v}\right)^v e^{-2M^2}, \tag{1.4}$$

where $K(V)$ depends on V only.

COMMENT.

1. An actual (possibly nonoptimal) dependence of $K(V)$ upon V will be carried out in the proof.
2. The condition $v \geq 1$ is assumed for convenience in the computations. If one assumes only $v > 0$, the same proof shows that $\tau_{\mathscr{C}}(M) \leq K(V, v)M^{2v-1}e^{-2M^2}$, where $K(V, v)$ depends on V, v only.
3. Proving that (1.4) holds for all $M > 0$ rather than from M not too large requires a number of unpleasant uninspiring computations. We have, however, decided to perform these to get the seemingly final result.

An important family of classes of sets are the so-called Vapnik–Červonenkis (VC) classes. Their importance stems from the fact that (modulo measurability) these are the classes of sets that behave well independently of the underlying probability [2]. Let us recall that $\mathscr{C}$ is called a VC class of index (= dimension) v if it does not shatter any subset of Ω of cardinality $v + 1$, but does shatter at least a subset of cardinality v. ($\mathscr{C}$ shatters $\{x_1, \dots, x_n\}$ if given a subset I of $\{1, \dots, n\}$, we can find a $C \in \mathscr{C}$ such that $x_i \in C$ if and only if $i \in I$.) Very recently, Haussler (improving a previous result of Dudley) proved that a VC class of index v satisfies (1.2) with $V = 4e$ [6]. Thus, in particular, (1.4) holds for such classes, where $K(4e)$ is a universal constant. [This is why the dependence of the constant of (1.4) upon V is not a critical issue.]

How sharp is (1.4)? In particular, what is the correct power of M in the right-hand side of (1.4)? Historically, the most important VC classes are the classes Q_d of sets of the type $x + (\mathbb{R}^+)^d (x \in \mathbb{R}^d)$ in $\mathbb{R}^d$. It is well known that Q_d has index d. It is known in that case (by looking at the limiting Gaussian process) that if P is uniform on $[0, 1]^d$, then for n large, $\tau_{Q_d}(M)$ is at least $cM^{2d-2}e^{-2M^2}$, where c depends on d only [8]. Thus one certainly could not do better than the exponent $2v - 2$ in (1.4). In the case of Q_q, the correct power of M is M^{2d-2}, not M^{2d-1}. This is due to the fact that the size of the coefficient of e^{-2M^2} in (1.4) is not influenced by the "dimension" of all $\mathscr{C}$, but rather by the "dimension" of the subset of $\mathscr{C}$ consisting of the sets $C \in \mathscr{C}$ for which $P(C) = 1/2$. In the example of Q_d this subset is genuinely smaller than $\mathscr{C}$, and the correct power in (1.4) should be $2v - 2$ rather than $2v - 1$. It is apparently not known whether this phenomenon occurs for all VC classes. But we would like to point out that this is a combinatorial question about VC classes, except when $v = 1$, where it is known that $2v - 1$ is then the optimal power (see [16] and the discussion therein) that is unrelated to the considerations of the present paper. Our methods do allow us to obtain the correct bounds for the usual classes, as is shown by the following result.

THEOREM 1.2. *Consider a class $\mathscr{C}$ of sets that satisfies either hypothesis* (i) *or hypothesis* (ii) *of Theorem* 1.1. *Assume moreover that for some numbers $V' \geq 1$, $v' \geq w > 0$ and all $\delta \geq \varepsilon > 0$, we have*

$$(1.5) \qquad N(\mathscr{C}_\delta, d_P, \varepsilon) \leq V' \delta^w \varepsilon^{-v'},$$

where $\mathscr{C}_\delta = \{C \in \mathscr{C};\ |P(C) - \frac{1}{2}| \leq \delta\}$. *Then, for* $M \geq K\sqrt{w}$, *we have*

$$\tau_{\mathscr{C}}(M) \leq K(v, v', w, V, V')M^{2v'-2w}e^{-2M^2},$$

where $K(v, v', w, V, V')$ *depends only on* v, v', w, V, V'.

COMMENTS.

1. In the case of Q_d, P uniform on $[0, 1]^d$, it is simple to see that (1.5) holds for $w = 1$, $v' = d$.
2. It would not be too hard to carry out an explicit dependence for $K(v, v', w, V, V')$ in v, v', w, V, V' (see Proposition 2.8 below).

Another question of interest is the possibility of obtaining small values of the numerical constants involved in (1.4) in the case of VC classes. As will be explained later, while (1.4) holds for each M, the reasons for which it holds when $M \gg n^{1/4}$ are rather uninteresting. Our proof, as it is written, provides unreasonable values of the constants involved; these large values occur out of the necessity to control the values of M near $\sqrt{n}$. If one restricted attention to the values of $M \leq n^{1/4}$, the values of the constants provided by our proof would already not be outrageous. But we have written the computations in the simplest possible way, without any attempt to get sharp constants, and certainly much improvement is possible in that direction. In particular, it must be pointed out that while our approach is unlikely ever to yield optimal constants, it essentially does not use chaining (that makes the search of sharp constants hopeless). We have, however, felt that the search of sharp numerical constants is better left to others with the talent and the taste for it.

Let us now turn to classes of functions. The following result parallels Theorem 1.1.

THEOREM 1.3. *Consider a (countable) class* $\mathscr{F}$ *of (measurable) functions* f *such that* $0 \leq f \leq 1$. *Assume that either of the following holds, where* $V, v \geq 1$.

(i) *Given* $\varepsilon > 0$ *and any probability* Q *on* Ω *that is supported by a finite set, we have*

$$N(\mathscr{F}, d_Q, \varepsilon) \leq \left(\frac{V}{\varepsilon}\right)^v, \tag{1.6}$$

where

$$d_Q(f, g) = \left(\int (f-g)^2\, dQ\right)^{1/2}. \tag{1.7}$$

(ii) *Given* $\varepsilon > 0$, $\mathscr{F}$ *can be covered by at most* $(V/\varepsilon)^v$ *brackets* $[f_1, f_2]$ *for which* $E(f_2 - f_1)^2 \leq \varepsilon^2$.

Then, for all $M > 0$, we have

$$\tau_{\mathscr{F}}(M) \le \left(K(V)\frac{M}{\sqrt{v}}\right)^{v} e^{-2M^2}. \tag{1.8}$$

COMMENT. Due to a different definition of distances [there is no square root in the right-hand side of (1.3)], one should replace v by $2v$ to compare this statement with Theorem 1.1. One then realizes that exactly one power of M has been lost in (1.8) versus (1.4).

There is a large number of possible variations on the theme of this work. As exemplified by previous work in this area, one of the challenges is to obtain clean results through clean computations. We have given only results where this could be reasonably well achieved. We have, however, tried to make the basic steps of the proof stand out in sufficient generality that many variations could be carried out with limited effort. One obvious such variation would be a result that would be to Theorem 1.3 what Theorem 1.2 is to Theorem 1.1. Other possibilities are suggested in Section 2 or in the course of the paper. One interesting case, in the situation of Theorem 1.3, is when one has more control of

$$\sigma(\mathscr{F}) = \sup_{f \in \mathscr{F}} \left(E(f - Ef)^2\right)^{1/2}. \tag{1.9}$$

While our approach does lead to progress in that case over the previous work of [1] and [10], we have not succeeded to produce there a clean result that could be considered as more or less the final word. Thus we will discuss only two reasonably simple results, with somewhat sketchy proofs.

We now discuss the methods and the organization of this paper. The basic idea was invented in the paper [19], where I study the tails of the supremum of a Gaussian process with unique point of maximal variance. The idea is simply that the main contribution to $P(\sup_{t \in T} X_t > u)$ should come from the variable X_s where variance is maximal. This is expressed by conditioning with respect to X_s and using the "concentration of measure phenomenon" on the conditioned process, in the form of the Gaussian isoperimeteric inequality. As it turns out, this method gives optimal order bounds for the tails of the supremum of a Gaussian process in all known cases. Unfortunately, the subsequent uses of this approach (in particular [15]) leave room for improvements. So, our first task is to spell out the basic principle in the case of Gaussian processes (Theorem 2.3) and to demonstrate how to use it. This is the object of Section 2. The advantage of working in the Gaussian setting is that there are much fewer technical difficulties, so that the ideas stand out more.

The program is then to follow the same overall approach in the case of empirical processes. The techniques to achieve that have been well under control for some time. First, one has to find a substitute for the Gaussian isoperimetric inequality. What one needs is only of fast decay of $\tau_{\mathscr{F}}(M)$ for certain classes $\mathscr{F}$. The inequalities proved in [1] and [10] would be (almost) sufficient for this purpose, but we find it much simpler to use the isoperimetric inequalities of [20], [21], and [23] that apparently take care once and for all of this class of problems, in the most general situation and in an optimal way. This is the purpose of Section 3.

In Section 4 we prove (variations of) classical estimates on the tails of the binomial law. These are obtained by brute force through Stirling's formulas.

In Section 5 we use these ingredients to mimic the proof of Theorem 2.3 and to obtain the basic inequality in the case of classes of sets. The proofs of Theorems 1.1 and 1.2 are then completed in Section 6.

In Section 7 we consider the case of classes of functions. The main difficulty there is that, in contrast with the case of sets, it is a priori not obvious how to work conditionally on the event $\{\sum_{i \le n} f(X_i) - nEf \ge u\}$. But, fortunately, there is an almost miraculous way to go around the problem (Theorem 7.1) using moment generating functions and an extreme point argument. This method, however, gives no hope of reducing the power of M in (1.7) to $v - 1$.

Finally, in Section 8 we discuss the case of classes of functions for which one controls the quantity of $\sigma(\mathscr{F})$ considered in (1.9).

Throughout the paper, K denotes a universal constant that may vary at each occurrence. Specific constants are denoted by K_1, K_2, and so forth.

2. Gaussian processes and partitioning lemma. First, we recall the following result, due essentially to R. M. Dudley, and which we will use many times.

PROPOSITION 2.1 (Metric entropy bound). *Consider a centered process $(Y_t)_{t \in T}$. Suppose that there is a distance d on T such that, for all $u > 0$, we have*

$$\forall\, s, t \in T, \qquad P(|X_s - X_t| \ge u\, d(s,t)) \le 2\exp(-u^2).$$

Then

$$E \sup_{t \in T} Y_t \le K \int_0^\infty \sqrt{\log N(T, d, \varepsilon)}\, d\varepsilon. \tag{2.1}$$

Concentrations of measure properties are essential in the understanding of Gaussian processes. The following convenient inequality is one of the many forms of these properties and is closely related to the Gaussian isoperimetric inequality. It is due to Maurey and Pisier [13]. (For our purposes, it is essentially irrelevant to have the best constant in the exponent; the proof then becomes simpler.)

PROPOSITION 2.2. *Consider a bounded Gaussian process $(X_t)_{t \in T}$. Then, setting $\sigma^2 = \sup_{t \in T} EX_t^2$, we have, for all $u > 0$,*

$$P\Big(\sup_{t \in T} X_t - E \sup_{t \in T} X_t \ge u\Big) \le \exp\Big(-\frac{u^2}{2\sigma^2}\Big). \tag{2.2}$$

We now turn to the basic inequality.

THEOREM 2.3. *Consider a Gaussian process* $(X_t)_{t\in T}$ *and set* $H = E\sup_{t\in T} X_t$. *Consider a r.v.* Y *such that the family* $\{Y, X_t, t\in T\}$ *is jointly Gaussian. Assume that*

$$\forall t\in T,\qquad E(Y(X_t - Y)) \le 0. \tag{2.3}$$

Set $\sigma = (EY^2)^{1/2}$, $a = \sup_{t\in T}(E(X_t - Y)^2)^{1/2}$.
Then, if $a \le \sigma$, $u \ge H$, *we have*

$$\begin{aligned} P\Big(\sup_{t\in T} X_t \ge u\Big) \le{}& \frac{1}{2}\exp\left(-\frac{(u-H)^2}{2a^2}\right) \\ &+ \frac{a}{\sqrt{a^2+\sigma^2}}\exp\left(-\frac{(u-H)^2}{2(\sigma^2+a^2)}\right) + \Phi\left(\frac{u-H}{\sigma}\right), \end{aligned} \tag{2.4}$$

where $\Phi(u) = \int_u^\infty (1/\sqrt{2\pi})e^{-s^2/2}\,ds$.
If, moreover, we have $u \ge 2H$, $u \ge H + \sigma$, *then we have*

$$P\Big(\sup_{t\in T} X_t \ge u\Big) \le \Phi\left(\frac{u}{\sigma}\right)\left(1 + K\frac{au}{\sigma^2}\exp\frac{1}{2}\left(\frac{au}{\sigma^2}\right)^2\right)\exp\frac{2uH}{\sigma^2}, \tag{2.5}$$

where K *is universal.*

COMMENT. Observe that (2.3) holds in particular if $EY^2 \ge \sup_{t\in T} EX_t^2$.

PROOF. For $t \in T$, we set

$$a_t = \frac{E((X_t - Y)Y)}{\sigma^2} = \frac{E(X_tY) - \sigma^2}{\sigma^2} \le 0.$$

We set

$$Z_t = X_t - (1 + a_t)Y,$$

so that $E(Z_tY) = 0$. Simple algebra now shows that

$$\forall s,t\in T,\qquad E(Z_s - Z_t)^2 \le E(X_s - X_t)^2.$$

Thus it follows by the Sudakov–Fernique [4] inequality that

$$E\sup_{t\in T} Z_t \le H. \tag{2.6}$$

(We should also mention here that there is a more elementary argument that yields $E\sup_{t\in T} Z_t \le 2H$.)
Now, we write

$$\begin{aligned} P\Big(\sup_{t\in T} X_t \ge u \mid Y = w\Big) &= P\Big(\sup_{t\in T}(X_t - Y) \ge u - w \mid Y = w\Big) \\ &= P\Big(\sup_{t\in T}(Z_t + a_tY) \ge u - w \mid Y = w\Big). \end{aligned}$$

For $w \geq 0$, since $a_t \leq 0$, we have

$$\sup_{t \in T} (Z_t + a_t w) \leq \sup_{t \in T} Z_t.$$

Since $E(Z_t Y) = 0$ for all $t \in T$, the process $(Z_t)_{t \in T}$ is independent of Y. Also, $EZ_t^2 \leq E(Z_t + a_t Y)^2 = E(X_t - Y)^2 \leq a^2$. Thus, from (2.6) and (2.2), we get, for $w \leq u - H$, $w \geq 0$, that

$$P\Big(\sup_{t \in T} X_t \geq u | Y = w\Big) \leq \exp\left(-\frac{(u - w - H)^2}{2a^2}\right). \tag{2.7}$$

For $w \leq 0$, since

$$|a_t| = \frac{|E((X_t - Y)Y)|}{\sigma^2} \leq \frac{a}{\sigma} \leq 1,$$

we have

$$\sup_{t \in T} (Z_t + a_t w) \leq |w| + \sup_{t \in T} Z_t = -w + \sup_{t \in T} Z_t,$$

and thus, for $u \geq H$, by (2.2) again

$$\begin{aligned} P\Big(\sup_{t \in T} (Z_t + a_t Y) \geq u - w | Y = w\Big) &\leq P\Big(\sup_{t \in T} Z_t \geq u\Big) \\ &\leq \exp\left(-\frac{(u - H)^2}{2a^2}\right). \end{aligned} \tag{2.8}$$

We have

$$\begin{aligned} P\Big(\sup_{t \in T} X_t \geq u\Big) &= \int_{-\infty}^{\infty} P\Big(\sup_{t \in T} X_t \geq u | Y = w\Big) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{w^2}{2\sigma^2}\right) dw \\ &= \int_{-\infty}^{0} + \int_{0}^{u-H} + \int_{u-H}^{\infty} := I_1 + I_2 + I_3. \end{aligned}$$

By (2.8), we have

$$I_1 \leq \frac{1}{2} \exp\left(-\frac{(u - H)^2}{2a^2}\right).$$

By (2.7), we have, setting $u - H = s$, by a routine computation,

$$\begin{aligned} I_2 &\leq \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(s - w)^2}{2a^2} - \frac{w^2}{2a^2}\right) dw \\ &= \frac{a}{\sqrt{a^2 + \sigma^2}} \exp\left(\frac{s^2}{2a^4(a^{-2} + \sigma^{-2})} - \frac{s^2}{2a^2}\right) \\ &= \frac{a}{\sqrt{a^2 + \sigma^2}} \exp\left(-\frac{s^2}{2(\sigma^2 + a^2)}\right). \end{aligned}$$

This proves (2.4).

We now turn to the proof of (2.5). First, we observe that

$$\exp\left(-\frac{(u-H)^2}{2a^2}\right) = \exp\left(-\frac{(u-H)^2}{2(a^2+\sigma^2)}\right)\exp\left(-\frac{(u-H)^2\sigma^2}{2a^2(a^2+\sigma^2)}\right).$$

Now, since $u \geq \sigma + H$, $a \leq \sigma$, we have

$$\exp\left(-\frac{(u-H)^2\sigma^2}{2a^2(a^2+\sigma^2)}\right) \leq \exp\left(-\frac{\sigma^2}{4a^2}\right) \leq \frac{Ka}{\sigma},$$

so that the sum of the first two terms of the right-hand side of (2.4) is at most

$$\frac{Ka}{\sigma}\exp\left(-\frac{(u-H)^2}{2(\sigma^2+a^2)}\right).$$

Since

$$\frac{1}{\sigma^2+a^2} \geq \frac{1}{\sigma^2} - \frac{a^2}{\sigma^4}$$

this is at most, setting $\xi = ua/\sigma^2$,

$$\text{(2.9)}\qquad \frac{Ka}{\sigma}\exp\left(-\frac{(u-H)^2}{2\sigma^2}+\frac{u^2a^2}{2\sigma^4}\right) \leq \frac{K\sigma}{u}\xi\exp\frac{\xi^2}{2}\exp\left(-\frac{(u-H)^2}{2\sigma^2}\right).$$

If we recall the well-known fact that, for $x \geq 1$,

$$\text{(2.10)}\qquad \Phi(x) \geq \frac{1}{2x\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right),$$

we see since $u - H \geq \sigma$ that the right-hand side of (2.9) is at most

$$K\xi\exp\frac{\xi^2}{2}\Phi\left(\frac{u-H}{\sigma}\right).$$

To conclude, we observe that

$$\Phi(x-y) \leq e^{2xy}\Phi(x)$$

if $x \geq 1$, $y > 0$. Indeed the function

$$f(y) = e^{2xy}\Phi(x) - \Phi(x-y)$$

satisfies $f(0) = 0$,

$$f'(y) = 2xe^{2xy}\Phi(x) - \frac{1}{\sqrt{2\pi}}e^{-(x-y)^2/2},$$

so that $f'(y) \geq 0$ by (2.10). □

As it turns out, Theorem 2.3 seems to be a universal tool to get proper bounds on the supremum of a Gaussian process, by breaking the index set into suitable pieces to which one applies the basic inequality. We first consider a simple case that is closely connected to our further results on empiricals. We will then sketch out several applications of Theorem 2.3 that demonstrate its power and that also have versions for empirical processes. The following is (an improvement of) a result of Samarodnitsky [15]. (It should be recalled that the results of [15] were actually obtained after this author pointed out the relevence of the approach of [19]).

THEOREM 2.4. *Consider a Gaussian process $(X_t)_{t\in T}$. Let $\sigma^2 = \sup_{t\in T} EX_t^2$. Consider the canonical distance d on T given by $d(s,t)^2 = E(X_s - X_t)^2$. Assume that for some constant $A > \sigma$, some $v > 0$ and some $0 \le \varepsilon_0 \le \sigma$ we have*

$$\varepsilon < \varepsilon_0 \Rightarrow N(T, d, \varepsilon) \le \left(\frac{A}{\varepsilon}\right)^v.$$

Then for $u \ge \sigma^2[(1+\sqrt{v})/\varepsilon_0]$ we have

$$P\Big(\sup_{t\in T} X_t \ge u\Big) \le \left(\frac{KAu}{\sqrt{v}\,\sigma^2}\right)^v \Phi\left(\frac{u}{\sigma}\right), \tag{2.11}$$

where K is universal.

COMMENTS. If $\varepsilon_0 = \sigma$, the condition on u is $u \ge \sigma(1+\sqrt{v})$. It is not restrictive as we cannot expect an interesting bound unless u is of order at least $E\sup_{t\in T} X_t$, which can be of order as large as $\sigma\sqrt{v}\sqrt{\log(eA/\sigma)}$.

The method of proof is as follows. Consider $a \le \sigma$ and $H > 0$. We partition T in N pieces $(T_i)_{i\le N}$, each of diameter $\le a \le \sigma$, and for each of which $E\sup_{t\in T_i} X_t \le H$. By (2.5) we get, when $a \le \sigma$, $u \ge H + \sigma$, $u \ge 2H$,

$$P\Big(\sup_{t\in T} X_t \ge u\Big) \le N\Phi\left(\frac{u}{\sigma}\right)\left(1 + K\frac{au}{\sigma^2}\exp\frac{1}{2}\left(\frac{au}{\sigma^2}\right)^2\right)\exp\left(\frac{2uH}{\sigma^2}\right). \tag{2.12}$$

As it turns out, the only way the term e^{2uH/σ^2} does not have a catastrophic influence is if uH/σ^2 is bounded independently of u. If we brutally partition T in $N \le (A/a)^v$ sets T_i of diameter $2a$, by Proposition 2.1, we get only

$$H \le \int_0^{2a} \sqrt{\log(A/\varepsilon)^v}\, d\varepsilon \le Ka\sqrt{v}\left(\log\frac{A}{a}\right)^{1/2}.$$

Then one has to take a of order $u^{-1}(\log u)^{-1/2}$, and then N gets too large [an extra term in $(\log u)^{v/2}$ appears]. To get around this difficulty (which will creep up again in the proof of Theorem 1.1), we need an (essentially well known) partitioning lemma (see [12]; considerably more subtle results are obtained in [22]).

LEMMA 2.5. *Consider a metric space (T,d) and $p, q \in \mathbb{Z}$, $p < q$. Consider a partition $\mathscr{P}_q$ of T, such that each set of $\mathscr{P}_q$ has diameter $\le 4^{-q}$. Consider integers k_l, $p \le l \le q$. Then one can find an increasing sequence $(\mathscr{P}_l)_{p\le l\le q}$ of* *partitions of* T *with the following properties*:

(2.13) *Each set of* $\mathscr{P}_l$ *has diameter less than or equal to* 4^{-l+1}.

(2.14) *Each atom of* $\mathscr{P}_l$ *contains at most* k_l *atoms of* $\mathscr{P}_{l+1}$.

$$\forall\, l < q, \qquad \operatorname{card} \mathscr{P}_l \le N(T, d, 4^{-l}) + \frac{\operatorname{card} \mathscr{P}_{l+1}}{k_l}. \tag{2.15}$$

Proof. The partitions $\mathscr{P}_l$ are constructed by decreasing induction over l. We show how to construct $\mathscr{P}_l$ once $\mathscr{P}_{l+1}$ has been obtained. Set $N = N(T, d, 4^{-l})$. Consider points $(t_i)_{i \le N}$ of T such that each point of T is within distance 4^{-l} of at least one point t_i, $i \le N$. For $i \le N$, define A_i as the union of all the sets of $\mathscr{P}_{l+1}$ that intersect the ball centered at t_i of radius 4^{-l}. Thus, since each set of $\mathscr{P}_{l+1}$ has diameter $\le 4^{-l}$, A_i has diameter at most

$$2 \cdot 4^{-l} + 2 \cdot 4^{-l} = 4^{-l+1}.$$

Define now $C_i = A_i \setminus \bigcup_{j<i} A_j$. These form a partition Q for T, that is coarser than $\mathscr{P}_{l+1}$. Certain atoms of Q might contain more than k_l atoms of $\mathscr{P}_{l+1}$. Any such atom can be in turn partitioned in sets, all of which except one are the union of exactly k_l atoms of $\mathscr{P}_{l+1}$, the exceptional set being the union of at most k_l sets of $\mathscr{P}_{l+1}$. This constructs $\mathscr{P}_l$, and (2.15) is obvious. □

Corollary 2.6. *Suppose that in the preceding lemma we have* $N(T, d, 4^{-l}) \le (A4^l)^v$ *for* $l \ge p$ *and* $\operatorname{card} \mathscr{P}_q \le (A4^q)^v$. *Then, if* $k_l \ge 2 \cdot 4^v$ *for all* l, *we have* $\operatorname{card} \mathscr{P}_l \le 2(A4^l)^v$.

Proof. By decreasing induction over l,

$$(A4^l)^v + \frac{2(A4^{l+1})^v}{k_l} \le 2(A4^l)^v. \qquad \square$$

We go back to the proof of Theorem 2.4 and we show how to partition T to deduce (2.11) from (2.12). There is no loss of generality to assume T finite. We consider q large enough that $s, t \in T \Rightarrow d(s,t) \ge 4^{-q}$. We use Corollary 2.6 with $k_l = [2 \cdot 4^v] + 1 \le 3 \cdot 4^v$, and, for each $l \le q$, with $4^{-l} \le \varepsilon_0$, we find a partition of T in $N \le 2(A4^{-l})^v$ sets $(T_i)_{i \le N}$ such that, for $m \ge l$, we have, for all $i \le N$,

$$N(T_i, d, 4^{-m+1}) \le (3 \cdot 4^v)^{m-l}.$$

Using Proposition 2.1, we have by a simple calculation

$$E \sup_{t \in T_i} X_t \le K\sqrt{v}\, 4^{-l}. \tag{2.16}$$

We now take for l the smallest integer such that $4^{-l} \le \sqrt{v}\,\sigma^2/4Ku$, where K is the constant of (2.16). Thus, assuming, as we may, $K \ge 1$, we have $4^{-l} \le \varepsilon_0$ provided $u \ge \sqrt{v}\,\sigma^2/\varepsilon_0$. We have $N \le (16KAu/\sqrt{v}\,\sigma^2)^v$, since

$4^{-l} \geq \sqrt{v}\,\sigma^2/16Ku$. Also we have

$$H \leq v\sigma^2/4u, \qquad a \leq 4^{-l+1} \leq \sqrt{v}\,\sigma^2/u.$$

Observe that, since $u \geq \sqrt{v}\,\sigma$, we have $H \leq \sqrt{v}\,\sigma \leq u/4$, and we have $a \leq \sigma$, so that the result follows by (2.12). □

Our next application is an improvement of another result of Samarodnitsky [15] that is related to Theorem 1.2.

PROPOSITION 2.7. *Consider a Gaussian process $(X_t)_{t \in T}$ and denote by d the distance induced by the process on T. Set $\sigma = \sup_{t \in T}(EX_t^2)^{1/2}$, and, for $\delta > 0$, set*

$$T_\delta = \{t \in T;\ E(X_t^2) \geq \sigma^2 - \delta^2\}.$$

Consider numbers $v \geq w \geq 1$ and assume that for all $\delta > 0$, $\varepsilon > 0$, $\varepsilon \leq \delta(1+\sqrt{v})/\sqrt{w}$, we have

$$N(T_\delta, d, \varepsilon) \leq A\delta^w \varepsilon^{-v}. \tag{2.17}$$

Then, for $u \geq 2\sigma\sqrt{w}$, we have

$$P\Big(\sup_{t \in T} X_t \geq u\Big) \leq \frac{Aw^{w/2}}{v^{v/2}} K^{v+w}\left(\frac{u}{\sigma^2}\right)^{v-w} \Phi\left(\frac{u}{\sigma}\right). \tag{2.18}$$

PROOF. We set $\delta_0 = 0$, $\delta_1 = \sqrt{w}\,\sigma^2/u$, and, for $k \geq 1$, we set $\delta_k = 2^{k-1}\delta_1$. For $k \geq 1$, we set $U_k = T_{\delta_k} \setminus T_{\delta_{k-1}}$. Setting $\varepsilon_0 = \delta_1(1+\sqrt{v})/\sqrt{w}$, we have

$$\frac{\sigma^2(1+\sqrt{v})}{\varepsilon_0} \leq \frac{\sigma^2\sqrt{w}}{\delta_1} = u,$$

and thus, setting $\sigma_k^2 = \sigma^2 - \delta_{k-1}^2$ and applying Theorem 2.4 to U_k, we get

$$P\Big(\sup_{t \in U_k} X_t \geq u\Big) \leq A\delta_k^w \left(\frac{Ku}{\sqrt{v}\,\sigma_k^2}\right)^v \Phi\left(\frac{u}{\sigma_k}\right).$$

Denote by k_0 the largest integer such that $\delta_{k_0-1} \leq \sigma/2$. Since $u \geq 2\sqrt{w}\,\sigma$, we have $k_0 \geq 2$ and $\delta_{k_0-1} \geq \sigma/4$. For $k \leq k_0$, we have $\sigma_k^2 \geq 3\sigma^2/4$, so that

$$P\Big(\sup_{t \in U_k} X_t \geq u\Big) \leq A\delta_k^w \left(\frac{Ku}{\sqrt{v}\,\sigma^2}\right)^v \frac{\sigma}{u} \exp\left(-\frac{u^2}{2(\sigma^2 - \delta_{k-1}^2)}\right).$$

Since we have

$$\frac{1}{\sigma^2 - \delta_{k-1}^2} \geq \frac{1}{\sigma^2} + \frac{\delta_{k-1}^2}{\sigma^4},$$

we see that

$$P\Big(\sup_{t \in U_k} X_t \geq u\Big) \leq A\delta_k^w \left(\frac{Ku}{\sqrt{v}\,\sigma^2}\right)^v \exp\left(-\frac{u^2\delta_{k-1}^2}{2\sigma^4}\right) \Phi\left(\frac{u}{\sigma}\right).$$

Now, we have

$$\sum_{k\geq 1}\delta_k^w \exp\left(-\frac{u^2\delta_{k-1}^2}{2\sigma^4}\right) = \delta_1^w + \sum_{k\geq 2}\delta_k^w \exp\left(-\frac{u^2\delta_{k-1}^2}{2\sigma^4}\right)$$

$$\leq \delta_1^w\left(1+\sum_{k\geq 2}2^{wk}\exp(-w2^{2k-3})\right)$$

$$\leq (K\delta_1)^w.$$

Thus, if we set $T' = \bigcup_{k\leq k_0} U_k$, we have, recalling the value of δ_1, that

$$P\left(\sup_{t\in T'} X_t \geq u\right) \leq A\frac{w^{w/2}}{v^{v/2}}K^{w+v}\left(\frac{u}{\sigma^2}\right)^{v-w}\Phi\left(\frac{u}{\sigma}\right). \tag{2.19}$$

For $t \notin T'$, we have

$$E(X_t^2) \leq \sigma^2 - \delta_{k_0-1}^2 \leq \sigma^2 - \sigma^2/16 = 15\sigma^2/16.$$

Thus, if we use (2.17) for $\delta = \sigma$ and we apply Theorem 2.4, we see that for $u \geq \sigma\sqrt{w}$ we have

$$P\left(\sup_{t\in T\setminus T'} X_t \geq u\right) \leq A\sigma^w\left(\frac{Ku}{\sqrt{v}\,\sigma^2}\right)^v\Phi\left(\frac{4u}{\sigma\sqrt{15}}\right). \tag{2.20}$$

To conclude, it suffices to check that for $x \geq \sqrt{w}$ we have $\Phi(4x/\sqrt{15}) \leq (Kw)^{w/2}x^{-w}\Phi(x)$, so that the left-hand side of (2.20) is dominated by the left-hand side of (2.18). □

COMMENT. If we suppose $w > 0$ rather that $w > 1$, the only difference lies in the dependence in w of the right-hand side of (2.18).

We finish this section with two more applications of Theorem 2.3. The following was proved in [17] using also the concentration of measure phenomenon, but in a different way.

PROPOSITION 2.8. *Consider a Gaussian process $(X_t)_{t\in T}$. We assume that T is compact for the natural distance d and that the process $(X_t)_{t\in T}$ is continuous for d (see, e.g., [19] for complete definitions). Let $\sigma = \sup_{t\in T}(EX_t^2)^{1/2}$. Then, given $\varepsilon > 0$, we can find $u(\varepsilon)$ such that*

$$u \geq u(\varepsilon) \Rightarrow P\left(\sup_{t\in T} X_t \geq u\right) \leq \exp\left(u\varepsilon - \frac{u^2}{2\sigma^2}\right).$$

SKETCH OF PROOF. Since the process is continuous, we have

$$\lim_{\delta\to 0} E \sup_{d(s,t)\leq\delta} |X_s - X_t| = 0.$$

Fix $\varepsilon > 0$ and pick $\delta > 0$ such that

$$E \sup_{d(s,t)\le\delta} |X_s - X_t| \le \varepsilon\sigma^2/8.$$

Set $a = \sigma^2\sqrt{\varepsilon/4u}$. By (2.5) we get, for u large enough that $a \le \delta$, that

$$\begin{aligned} P\Big(\sup_{t\in T} X_t \ge u\Big) &\le N(T,d,a)\Phi\Big(\frac{u}{\sigma}\Big)\Big(1 + K\exp\frac{\varepsilon u}{4}\Big)\exp\Big(\frac{\varepsilon u}{4}\Big) \\ &\le KN(T,d,a)\Phi\Big(\frac{u}{\sigma}\Big)\exp\Big(\frac{\varepsilon u}{2}\Big). \end{aligned}$$

Now, by Sudakov minorization, for all $\eta > 0$, we have

$$N(T,d,a) \le \exp\frac{\eta}{a^2}$$

for a small enough. Thus, taking $\eta = \sigma^4\varepsilon^2/8$, we have $N(T,d,a) \le \exp \varepsilon u/2$ for a small enough (i.e., u large enough). This completes the proof. □

COMMENT.

1. In [17] a result of the same nature is proved when the process $(X_t)_{t\in T}$ is only assumed to be bounded; but this does not seem to follow from Theorem 2.3.
2. It should be pointed out that, instead of (2.5), one could use the cruder inequality

$$P\Big(\sup_{t\in T} X_t \ge u\Big) \le 2\exp\bigg(-\frac{(u-H)^2}{2\sigma^2}\bigg),$$

which follows immediately from (2.2).

We now turn to the main result of [19].

PROPOSITION 2.9. *Consider a Gaussian process $(X_t)_{t\in T}$; suppose that T is compact for the canonical distance d. Suppose that there is a unique point $s \in T$ for which*

$$EX_s^2 = \sup_{t\in T} EX_t^2.$$

For $h > 0$, set $T_h = \{t \in T;\ EX_tX_s \ge \sigma^2 - h^2\}$.
Assume that $E\sup_{t\in T} X_t < \infty$ and that

$$\lim_{h\to\infty} \frac{E\sup_{t\in T_h} X_t}{h} = 0.$$

Then

$$\lim_{u\to 0} \frac{P(\sup_{t\in T} X_t \ge u)}{\Phi(u/\sigma)} = 1.$$

PROOF. Consider $1 > \eta > 0$. Consider h_0 such that

$$h \le 2h_0 \Rightarrow E \sup_{t \in T_h} X_t \le \eta^2 h. \tag{2.20}$$

We can and do assume $h_0 \le \sigma\eta^2$. Since we have assumed that T is compact, we have

$$\sup_{t \notin T_{h_0}} EX_t^2 < \sigma^2,$$

since otherwise there would be $s' \notin T_{h_0}$ (so that $s' \neq s$) for which $EX_{s'}^2 = \sigma^2$, and this contradicts the hypothesis that there is a unique point of maximal variance.

By Proposition 2.8 we see that it suffices to prove that for some universal constant K we have

$$P\Big(\sup_{t \in T_{h_0}} X_t \ge u\Big) \le \Phi\Big(\frac{u}{\sigma}\Big)(1 + K\eta)$$

for u large enough. We fix u and we set $\alpha = \sigma^2/\eta u$. We set $V_{-1} = \varnothing$. For $k \ge 0$, we set

$$V_k = T_{2^k \alpha}, \qquad U_k = V_k \setminus V_{k-1}.$$

Consider the smallest integer p such that $2^p \alpha \ge h_0$. We have

$$T_{h_0} \subset \bigcup_{0 \le k \le p} U_k.$$

Setting $H_k = E \sup_{t \in V_k} X_t$, we see by (2.20) that for $k \le p$ we have $H_k \le 4\alpha\eta^2 2^k$. Setting

$$b_k = \sup_{t \in V_k} \big(E|X_t - X_s|^2\big)^{1/2},$$

we see (since $s \in V_k$) that

$$b_k \le KH_k \le K\alpha\eta^2 2^k. \tag{2.21}$$

By (2.5) we have

$$P\Big(\sup_{t \in U_0} X_t \ge u\Big) \le \Phi\Big(\frac{u}{\sigma}\Big)(1 + K\eta \exp K\eta^2)\exp 2\eta$$

$$\le \Phi\Big(\frac{u}{\sigma}\Big)(1 + K\eta).$$

To estimate $P(\sup_{t \in U_k} X_t \ge u)$ for $k \ge 1$, we appeal again to (2.5). We now take the r.v. Y of Theorem 2.3 to be $Y_k = (1 - (\alpha 2^{k-1})^2/\sigma^2)X_s$. [It is then a simple matter to see that (2.3) holds by definition of T_h.] Then we have, since $k \le p$ and $h_0 \le \sigma\eta^2$,

$$\big(E(Y_k - X_s)^2\big)^{1/2} = (\alpha 2^{k-1})^2/\sigma \le \frac{h_0}{\sigma}\alpha 2^{k-1} \le \eta^2 \alpha 2^k.$$

It follows from (2.21) and the triangle inequality that if we set

$$a_k = \sup_{t \in V_k} \left(E|X_t - Y_k|^2 \right)^{1/2},$$

then $a_k \le K\eta^2 \alpha 2^k$. Thus by (2.5) we get (since $\eta \le 1$)

$$\begin{aligned} P\Big(\sup_{t \in U_k} X_t \ge u \Big) &\le \Phi\left(\frac{u}{\sigma\left(1 - (\alpha 2^{k-1})^2/\sigma^2\right)} \right)\left(1 + K\eta 2^k \exp(K\eta^2 2^{2k})\right)\exp(K\eta 2^k) \\ &\le \Phi\left(\frac{u}{\sigma\left(1 - (\alpha 2^{k-1})^2/\sigma^2\right)} \right)\exp(K\eta 2^{2k}). \end{aligned}$$

We observe that, for $x \le 1$, we have $(1 - x)^{-1} \ge 1 + x$, so that

$$\frac{u}{\sigma\left(1 - (\alpha 2^{k-1})^2/\sigma^2\right)} \ge \frac{u}{\sigma} + \frac{(\alpha 2^{k-1})^2 u}{2\sigma^3}.$$

Also, it is immediate to see that, by a change of variable,

$$\Phi(x + y) \le e^{-xy}\Phi(x)$$

for $y > 0$, so that, recalling the value of α,

$$\begin{aligned} \Phi\left(\frac{u}{\sigma\left(1 - (\alpha 2^{k-1})^2/\sigma^2\right)} \right) &\le \Phi\left(\frac{u}{\sigma}\right)\exp\left(-\frac{(\alpha 2^{k-1})^2 u^2}{\sigma^4} \right) \\ &\le \Phi\left(\frac{u}{\sigma}\right)\exp\left(-\frac{2^{2k}}{4\eta} \right). \end{aligned}$$

Thus we have

$$\sum_{1 \le k \le p} P\Big(\sup_{t \in U_k} X_t \ge u \Big) \le \Phi\left(\frac{u}{\sigma}\right)\left(\sum_{1 \le k \le p} \exp\left(-2^{2k}\left(\frac{1}{4\eta} - K\eta \right) \right) \right).$$

For η sufficiently small, this latter sum is $\le \eta$. This completes the proof. □

3. Isoperimetric bounds. We first recall some general tools. Throughout the paper, we denote by $(\varepsilon_i)_{i \le n}$ an independent Bernoulli sequence [i.e., $P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = 1/2$] that is independent of the sequence $(X_i)_{i \le n}$. We denote by P_ε and E_ε, respectively, the conditional probability and the conditional expectation given $(X_i)_{i \le n}$.

LEMMA 3.1 (Giné and Zinn [5]). *Consider any class $\mathscr{F}$ of functions on a probability space. Then we have*

$$E\left\|\sum_{i\le n} f(X_i) - nE(f)\right\|_{\mathscr{F}} \le 2E\left\|\sum_{i\le n}\varepsilon_i f(X_i)\right\|_{\mathscr{F}}, \tag{3.1}$$

$$\forall\, t>0,\quad P\left(\left\|\sum_{i\le n} f(X_i) - nE(f)\right\|_{\mathscr{F}} \ge 4t\right) \le 4P\left(\left\|\sum_{i\le n}\varepsilon_i f(X_i)\right\|_{\mathscr{F}} \ge t\right). \tag{3.2}$$

The following is also an idea from [5].

COROLLARY 3.2. *Consider a class $\mathscr{F}$ of nonnegative functions on a probability space. Then*

$$E\left\|\sum_{i\le n} f(X_i)\right\|_{\mathscr{F}} \le n\sup_{f\in\mathscr{F}} E(f) + 2E\left\|\sum_{i\le n}\varepsilon_i f(X_i)\right\|_{\mathscr{F}}.$$

PROOF. We observe that, for all $f\in\mathscr{F}$, we have

$$\begin{aligned}\sum_{i\le n} f(X_i) &\le \left|\sum_{i\le n} f(X_i) - nE(f)\right| + nE(f)\\ &\le \left\|\sum_{i\le n} f(X_i) - nE(f)\right\|_{\mathscr{F}} + n\sup_{f\in\mathscr{F}} E(f).\end{aligned}$$

To get the result, we take the supremum over f on the left-hand side, we take expectations, and we use (3.1). □

LEMMA 3.3 (Ledoux and Talagrand [9], Theorem 4.12). *Consider a class $\mathscr{F}$ of functions such that $-1\le f\le 1$ for $f\in\mathscr{F}$. Then*

$$E\left\|\sum_{i\le n}\varepsilon_i f^2(X_i)\right\|_{\mathscr{F}} \le 4E\left\|\sum_{i\le n}\varepsilon_i f(X_i)\right\|_{\mathscr{F}}.$$

Combining Lemma 3.3 and Corollary 3.2, we get the following result.

COROLLARY 3.4. *Consider a class $\mathscr{F}$ of functions on a probability space and assume that $-1\le f\le 1$ for all $f\in\mathscr{F}$. Set $\sigma^2=\sup_{f\in\mathscr{F}} E(f^2)$. Then*

$$E\left\|\sum_{i\le n} f^2(X_i)\right\|_{\mathscr{F}} \le n\sigma^2 + 8E\left\|\sum_{i\le n}\varepsilon_i f(X_i)\right\|_{\mathscr{F}}. \tag{3.3}$$

Throughout the paper we will use the following notation: We consider for $L,S > 0$, the function $\varphi_{L,S}(t)$ defined for $t \geq 0$ by

$$\varphi_{L,S}(t) = \frac{t^2}{L^2 S} \qquad \text{if } t \leq LS,$$

$$\varphi_{L,S}(t) = \frac{t}{L}\left(\log \frac{et}{LS}\right)^{1/2} \qquad \text{if } t \geq LS.$$

We observe that $\varphi_{L,S}(t)/t$ increases.

We now come to the following absolutely general principle.

THEOREM 3.5. *Consider a class $\mathscr{F}$ of functions on a probability space. Assume that $0 \leq f \leq 1$ for all $f \in \mathscr{F}$. Set*

$$H = E\left\|\sum_{i \leq n} \varepsilon_i f(X_i)\right\|_{\mathscr{F}}, \qquad \sigma = \sup_{f \in \mathscr{F}} \left(E(f - Ef)^2\right)^{1/2}.$$

Then, setting $S = n\sigma^2 + H$, for some universal constant K_1, we have

$$t \geq K_1 H \Rightarrow P\left(\left\|\sum_{i \leq n} f(X_i) - nE(f)\right\|_{\mathscr{F}} \geq t\right) \leq \exp\left(-\varphi_{K_1, S}(t)\right).$$

REMARK. The author hopes that the relative case with which Theorem 3.5 is disposed of will convince the reader to learn how to manipulate the isoperimetric inequalities of [20], [21] and [23].

PROOF. There is no loss of generality to assume that the probability P has no atoms. Consider a function θ which to each finite subset F of Ω associates a number $\theta(F)$. Assume the following:

$$F \subset G \Rightarrow \theta(F) \leq \theta(G), \tag{3.4}$$

$$\theta(F \cup G) \leq \theta(F) + \theta(G), \tag{3.5}$$

$$\theta(F) \leq \operatorname{card} F. \tag{3.6}$$

Since we assume that P has no atoms, the points $X_1, \dots, X_n$ are almost surely distinct. We can consider the function $Z = \theta(\{X_1, \dots, X_n\})$. When this function is measurable, it follows from the isoperimetric inequality of [21] that, for some universal constant K_2, we have, for $t \geq K_2 EZ$,

$$P(Z \geq t) \leq \exp\left(-\frac{t}{K_2} \log \frac{et}{K_2 EZ}\right). \tag{3.7}$$

This inequality can also be derived from the newer and simple inequality of [23] (but the derivation is less immediate); but the inequality of [21] itself has now received a very simple and elementary proof [24].

Consider the class $\mathscr{G}$ of functions $f - Ef$ for $f \in \mathscr{F}$. Thus

$$\Big\| \sum_{i \le n} f(X_i) - nE(f) \Big\|_{\mathscr{F}} = \Big\| \sum_{i \le n} g(X_i) \Big\|_{\mathscr{G}}.$$

Also, we observe that (averaging in $X_1, \dots, X_n$ inside the supremum rather than outside)

$$E \Big\| \sum_{i \le n} \varepsilon_i Ef \Big\|_{\mathscr{G}} \le H,$$

so that, by the triangle inequality,

$$E \Big\| \sum_{i \le n} \varepsilon_i g(X_i) \Big\|_{\mathscr{G}} \le 2H.$$

We apply (3.7) to the functions

$$\theta(\{x_1, \dots, x_n\}) = E_\varepsilon \Big\| \sum_{i \le n} \varepsilon_i g(x_i) \Big\|_{\mathscr{G}}$$

and

$$\theta(\{x_1, \dots, x_n\}) = \Big\| \sum_{i \le n} g^2(x_i) \Big\|_{\mathscr{G}},$$

respectively, to get that, for $u \ge 2K_2 H$, $v \ge 8K_2 S$, we have

$$P\bigg(E_\varepsilon \Big\| \sum_{i \le n} \varepsilon_i g(X_i) \Big\|_{\mathscr{G}} \ge u \bigg) \le \exp\bigg(-\frac{u}{K_2} \log \frac{eu}{2K_2 H} \bigg), \tag{3.8}$$

$$P\bigg(\Big\| \sum_{i \le n} g^2(X_i) \Big\|_{\mathscr{G}} \ge v \bigg) \le \exp\bigg(-\frac{v}{K_2} \log \frac{ev}{8K_2 S} \bigg). \tag{3.9}$$

[We have used that, by Corollary 3.4, we have $E\|\Sigma_{i \le n} g^2(X_i)\|_{\mathscr{G}} \le 8S$.]

We now appeal to the isoperimetric inequality of [20] to get that, for $t \ge 4E_\varepsilon\|\Sigma_{i \le n}\varepsilon_i g(X_i)\|_{\mathscr{G}}$, we have

$$P_\varepsilon\bigg(\Big\| \sum_{i \le n} \varepsilon_i g(X_i) \Big\|_{\mathscr{G}} \ge t \bigg) \le 2\exp\bigg(-\frac{t^2}{32\|\Sigma_{i \le n} g^2(X_i)\|_{\mathscr{G}}} \bigg).$$

Thus we get

$$P\bigg(\Big\| \sum_{i \le n} \varepsilon_i g(X_i) \Big\|_{\mathscr{G}} \ge t \bigg) \le 2\exp\bigg(-\frac{t^2}{32v} \bigg) + P\bigg(E_\varepsilon \Big\| \sum_{i \le n} \varepsilon_i g(X_i) \Big\|_{\mathscr{G}} \ge \frac{t}{4} \bigg)$$
$$+ P\bigg(\Big\| \sum_{i \le n} g^2(X_i) \Big\|_{\mathscr{G}} \ge v \bigg).$$

We use (3.8) with $u = t/4$. Thus, for $t \geq 8K_2H$, $v \geq 8K_2S$, we have

$$P\left(\left\|\sum_{i\leq n}\varepsilon_i g(X_i)\right\|_{\mathscr{G}} \geq t\right) \leq 2\exp\left(-\frac{t^2}{32v}\right) + \exp\left(-\frac{t}{8K_2}\log\frac{et}{8K_2H}\right) + \exp\left(-\frac{v}{K_2}\log\frac{ev}{8K_2S}\right).$$

If $t \leq 8K_2S$, we take $v = 8K_2S$, and we observe that $v \geq t^2/v$ to obtain the result. If $t \geq 8K_2S$, we take

$$v = t\left(\log\frac{et}{K_2S}\right)^{-1/2}$$

and the result follows by simple calculations. □

Let us observe a simple property of the function $\varphi_{L,S}(t)$.

LEMMA 3.6. *One can find a number $K(L)$ depending on L only such that*

$$\forall\, t \leq K(L)\sqrt{S}, \qquad \varphi_{L,S}\big(K(L)t\sqrt{S}\big) \geq 11t^2. \tag{3.10}$$

COMMENT. The number 11 could be replaced by any other.

PROOF. If $K(L)t\sqrt{S} \leq LS$, then

$$\varphi_{L,S}\big(K(L)t\sqrt{S}\big) = \frac{K(L)^2}{L^2}t^2.$$

If $K(L)t\sqrt{S} \geq LS$, then we have

$$\varphi_{L,S}\big(K(L)t\sqrt{S}\big) = \frac{K(L)t\sqrt{S}}{L}\left(\log\frac{etK(L)}{L\sqrt{S}}\right)^{1/2}.$$

Since the log is at least 1, if $K(L)t\sqrt{S} \geq 11t^2L$, the result holds. Otherwise, we have $t \geq K(L)\sqrt{S}/11L$, so that since $t \leq K(L)\sqrt{S}$, we have

$$\varphi_{L,S}\big(K(L)t\sqrt{S}\big) \geq \frac{t^2}{L}\left(\log\left(\frac{eK(L)^2}{11L^2}\right)\right)^{1/2}.$$

Thus it suffices to take $K(L)^2 = 11L^2\exp(11L)^2$. □

Here is a simple corollary of Theorem 3.5.

COROLLARY 3.7. *There exist two universal constants $K_3, a_0 > 0$, with the following property. Consider a class $\mathscr{F}$ of functions on a probability space and assume that $0 \leq f \leq 1$ for all $f \in \mathscr{F}$. Assume that*

$$\sup_{f\in\mathscr{F}} E(f - Ef)^2 \leq a_0.$$

Let

$$H = E\left\| \sum_{i \le n} \varepsilon_i f(X_i) \right\|_{\mathscr{F}}.$$

Then, if

$$n \ge K_3 H, \qquad M\sqrt{n} \ge K_3 H, \tag{3.11}$$

we have

$$\tau(M) \le K_3 \exp(-11M^2).$$

COMMENT. The number a_0 will be used throughout the paper.

PROOF. We set $S = na_0 + 8H$. Thus, by Theorem 3.5, we have

$$M\sqrt{n} \ge K_1 H \Rightarrow \tau(M) \le \exp\left(-\varphi_{K_1, S}(M\sqrt{n})\right).$$

We observe (and this will be used many times) that, since $|f(X_i) - Ef| \le 1$, we have $\tau(M) = 0$ unless $M\sqrt{n} \le n$, so that one can always assume $M \le \sqrt{n}$.

We now use Lemma 3.6 with $t = M\sqrt{n}$, $S' = n/K(K_1)^2$. We observe that, since $M \le \sqrt{n}$, we have $t \le K(K_1)\sqrt{S'}$, Thus, by (3.10), we have

$$\varphi_{K_1, S'}(M\sqrt{n}) \ge 11M^2.$$

Since $\varphi_{K_1, x}(t)$ is an increasing function of x, we are done if $S' \ge S$. But this occurs if $a_0 = 1/2K(K_1)^2$, $n \ge 16K(K_1)^2 H$. □

4. Binomial coefficients. Certainly it is hard to say something new about binomial coefficients. However, we have not found in the literature the exact property we need here. In any case, the reader might appreciate that we give a computationally very simple derivation, of the bounds we need. Consider $1 \le k \le n - 1$. Using Stirling's formulas as, for example, in Robbins [14], we get

$$\binom{n}{k} \le \frac{K\sqrt{n}}{\sqrt{k(n-k)}} \frac{n^n}{k^k (n-k)^{n-k}}.$$

Setting $t = k/n$ yields

$$\binom{n}{k} \le \frac{K\sqrt{n}}{\sqrt{k(n-k)}} \left(\frac{1}{t^t (1-t)^{1-t}} \right)^n.$$

Thus, given $\alpha > 0$, we have

$$\alpha^k (1-\alpha)^{n-k} \binom{n}{k} \le \frac{K\sqrt{n}}{\sqrt{k(n-k)}} \left(\left(\frac{\alpha}{t}\right)^t \left(\frac{1-\alpha}{1-t}\right)^{1-t} \right)^n.$$

Setting $u = t - \alpha = k/n - \alpha$, we get

$$\alpha^k(1-\alpha)^{n-k}\binom{n}{k} \le \frac{K\sqrt{n}}{\sqrt{k(n-k)}}\exp(-n\Psi(u,\alpha)). \tag{4.1}$$

where

$$\begin{aligned}\Psi(u,a) = & -(u+\alpha)\log\alpha + (u+\alpha)\log(u+\alpha)\\ & +(1-(u+\alpha))\log(1-(u+\alpha))\\ & -(1-(u+\alpha))\log(1-\alpha).\end{aligned}$$

To understand better the function Ψ, one checks by direct computation that

$$\Psi(0,\alpha) = 0, \qquad \frac{\partial\Psi}{\partial u}(0,\alpha) = 0$$

and

$$\frac{\partial^2\Psi}{\partial u^2} = \frac{1}{(u+\alpha)(1-\alpha-u)} = \frac{4}{1-4\left(u-\left(\frac{1}{2}-\alpha\right)\right)^2}.$$

In particular, we have

$$\frac{\partial^2\Psi}{\partial u^2} \ge 4\left(1+4\left(u-\left(\frac{1}{2}-\alpha\right)\right)^2\right).$$

Thus

$$\frac{\partial\Psi}{\partial u}(u,\alpha) \ge 4u + \frac{16}{3}\left(\left(u-\left(\frac{1}{2}-\alpha\right)\right)^3 + \left(\frac{1}{2}-\alpha\right)^3\right). \tag{4.2}$$

The function $h(\beta) = (u-\beta)^3 + \beta^3$ is minimum at $\beta = u/2$, so that

$$\frac{\partial\Psi}{\partial u}(u,\alpha) \ge 4u + \frac{4}{3}u^3$$

and thus

$$\Psi(u,\alpha) \ge 2u^2 + \frac{u^4}{3}. \tag{4.3}$$

LEMMA 4.1. *Assume that $a_0 \le \alpha \le 1 - a_0$ (where a_0 has been introduced in Corollary 3.7). Then, for all k such that $\alpha n \le k \le n$, we have*

$$\alpha^k(1-\alpha)^{n-k}\binom{n}{k} \le \frac{K}{\sqrt{n}}\exp\left(-n\left(2u^2+\frac{u^4}{4}\right)\right),$$

where we have set $u = k/n - \alpha$.

PROOF. If $k \le n(1 - a_0/2)$, we have $n - k \ge na_0/2$, so that

$$\frac{\sqrt{n}}{\sqrt{k(n-k)}} \le \frac{1}{\sqrt{n}} \frac{1}{\sqrt{a_0^2/2}}$$

and the result follows from (4.3) in that case.

If $k \ge n(1 - a_0/2)$, $k < n$, we have $u \ge a_0/2$, so that by (4.1) and (4.3) (and since $n - k \ge 1$)

$$\alpha^k(1-\alpha)^{n-k}\binom{n}{k}$$
$$\le \frac{K}{\sqrt{n}\sqrt{(1-a_0/2)}} \exp\left(-n\left(2u^2 + \frac{u^4}{4}\right)\right)\left[\sqrt{n}\exp(-na_0^4/12)\right]$$

and the result follows since the last term is bounded independently of n. Only the case $k = n$ remains. It is left to the reader. □

COROLLARY 4.2. *Assume that* $a_0 \le \alpha \le 1 - a_0$. *Consider* $0 \le w \le u \le 1$. *Then*

$$\sum_{l \ge n(w+\alpha)} \alpha^l(1-\alpha)^{n-l}\binom{n}{l} \le \frac{K}{u\sqrt{n}} \exp\left(-n\left(2u^2 + \frac{u^4}{4}\right) + 5nu(u-w)\right).$$

PROOF. Consider the function $h(x) = 2x^2 + x^4/4$. It is convex, so that for all x we have

$$h(x) \ge h(u) + (x-u)h'(u).$$

Thus, by Lemma 4.1, we have

$$\sum_{l \ge n(w+\alpha)} a^l(1-a)^{n-l}\binom{n}{l}$$
$$\le \frac{K}{\sqrt{n}} \exp(-nh(u)) \sum_{l \ge n(w+\alpha)} \exp((nu - (l - n\alpha))h'(u)).$$

We observe that $4u \le h'(u) \le 5u$. If we denote by l_0 the smallest integer with $l_0 \ge n(w + \alpha)$, we have

$$\sum_{l \ge n(w+\alpha)} \exp((nu - l + n\alpha)h'(u))$$
$$= (1 - \exp(-h'(u)))^{-1}\exp((nu - l_0 + n\alpha)h'(u))$$
$$\le \frac{K}{h'(u)} \exp((nu - l_0 + n\alpha)h'(u))$$
$$\le \frac{K}{u} \exp(5n(u-w)u).$$

□

5. Basic inequality: the case of sets.

THEOREM 5.1. *Consider a class $\mathscr{C}$ of subsets of Ω. Consider a certain set $C_0 \in \mathscr{C}$ and assume that $a_0 \le P(C_0) \le 1 - a_0$ (where a_0 has been determined in Corollary 3.7). We set*

$$H = E\left\| \sum_{i \le n} \varepsilon_i 1_{C \triangle C_0}(X_i) \right\|_{\mathscr{C}}, \qquad a = \sup_{C \in \mathscr{C}} P(C \triangle C_0).$$

Then, if $M \ge 4/a_0$, we have

$$\tau_{\mathscr{C}}(M) \le \frac{K}{M} e^{-2M^2} \exp\left(KaM^2 + \frac{KMH}{\sqrt{n}} - \frac{M^4}{4n} \right).$$

We first present a lemma that will allow us to bring in classes consisting only of small sets.

LEMMA 5.2. *In the situation of Theorem 5.1, consider the classes of sets*

$$\mathscr{C}_1 = \{C_0 \setminus C; C \in \mathscr{C}\}, \qquad \mathscr{C}_2 = \{C \setminus C_0; C \in \mathscr{C}\}.$$

Consider a subset I of $\{1, \dots, n\}$ and the event

$$\Omega_I = \{i \in I \Leftrightarrow X_i \in C_0\}.$$

Set $k = \operatorname{card} I$. Consider the random variables defined on Ω_I by

$$F_1 = \left\| \sum_{i \in I} 1_C(X_i) - k \frac{P(C)}{P(C_0)} \right\|_{\mathscr{C}_1}, \tag{5.1}$$

$$F_2 = \left\| \sum_{i \notin I} 1_C(X_i) - (n-k) \frac{P(C)}{P(C_0^c)} \right\|_{\mathscr{C}_2}. \tag{5.2}$$

Then on Ω_I we have

$$\left\| \sum_{i \le n} 1_C(X_i) - nP(C) \right\|_{\mathscr{C}} \le |nP(C_0) - k| \left(1 + \frac{2a}{a_0} \right) + F_1 + F_2. \tag{5.3}$$

PROOF. Consider a set $C \in \mathscr{C}$. We observe that

$$C = (C \setminus C_0) \cup (C_0 \setminus (C_0 \setminus C)).$$

Thus

$$\begin{aligned} \operatorname{card}\{i \le n: X_i \in C\} = \operatorname{card}\{i \le n; X_i \in C_0\} \\ + \operatorname{card}\{i \le n; X_i \in C \setminus C_0\} \\ - \operatorname{card}\{i \le n; X_i \in C_0 \setminus C\}. \end{aligned}$$

Now we observe that

$$\left|nP(C \setminus C_0) - (n-k)\frac{P(C \setminus C_0)}{P(\Omega \setminus C_0)}\right| = \frac{P(C \setminus C_0)}{P(\Omega \setminus C_0)}|nP(\Omega \setminus C_0) - (n-k)|$$
$$\leq \frac{a}{a_0}|nP(C_0) - k|.$$

A similar inequality for

$$\left|nP(C_0 \setminus C) - \frac{kP(C_0 \setminus C)}{P(C_0)}\right|$$

completes the proof. □

Consider the probability P_1 on Ω given by $P_1(A) = P(A \cap C_0)/P(A)$. Consider i.i.d. r.v. $Y_1, \dots, Y_k$ distributed according to P_1. Then, conditionally on Ω_I, F_1 is distributed like

$$V(k) = \left\| \sum_{i \leq k} 1_C(Y_i) - kP_1(C) \right\|_{\mathscr{C}_1}.$$

Thus we will be able to bound the tails of $V(k)$ using Theorem 3.5.

LEMMA 5.3. *Consider a class $\mathscr{F}$ of functions and set*

$$H = E\left\| \sum_{i \leq n} \varepsilon_i f(X_i) \right\|_{\mathscr{F}}.$$

Consider a subset C_0 of Ω, with $a_0 \leq \alpha = P(C_0) \leq 1 - a_0$. Consider the variables $(Y_i)_{i \leq k}$ as above. Then

$$E\left\| \sum_{i \leq k} \varepsilon_i f(Y_i) \right\|_{\mathscr{F}} \leq KH.$$

PROOF. We observe first that

$$E_\varepsilon \left\| \sum_{i \leq n} \varepsilon_i f(X_i) \right\|_{\mathscr{C}} \geq E_\varepsilon \left\| \sum_{X_i \in C_0} \varepsilon_i f(X_i) \right\|_{\mathscr{C}}.$$

Thus, if we set

$$G_l = E\left\| \sum_{i \leq l} \varepsilon_i f(Y_i) \right\|_{\mathscr{F}},$$

taking expectations we have

$$H \geq \sum_{0 \leq l \leq n} \alpha^l (1-\alpha)^{n-l} \binom{n}{l} G_l.$$

Thus there exists $l \geq \alpha n$ such that $G_l \leq 2H$. Since $k \leq n \leq l/a_0$, we have $E(G_k) \leq E(G_l)/a_0$. □

We now observe that

$$\sup_{C \in \mathscr{C}} P_1(C) \leq \frac{a}{a_0}.$$

Thus we see from Theorem 3.5 that

$$P(F_1 \geq t|\Omega_I) \leq K_1 \exp(-\varphi_{K_1, S'}(t)) \quad \text{if } t \geq K_1 H,$$

where $S' \leq K(na + H)$. Certainly the same bounds do hold for $P(F_2 \geq t|\Omega_I)$, so similar bounds (with different constants) hold for $P(F_1 + F_2 \geq t|\Omega_I)$. We observe that these bounds are independent of k.

PROPOSITION 5.4. *We have*

$$\left\| \sum_{i \leq n} 1_C(X_i) - nP(C) \right\|_{\mathscr{C}} \leq nU\left(1 + \frac{2a}{a_0}\right) + W,$$

where the random variables U and W have the following properties: $W \geq 0$, *and for all* $t \geq 0$, $u > w > 0$, *we have*

$$(5.4) \quad P(U \geq w, W \geq t) \leq \frac{K}{\sqrt{n}\,u} \exp(-nh(u) - \varphi(t) + 5nu(u - w)),$$

where $h(u) = 2u^2 + u^4/4$, *and where*

$$\varphi(t) = 0 \qquad \text{if } t < K_4 H,$$

$$\varphi(t) = \varphi_{K_4, S}(t) \quad \text{if } t \geq K_4 H,$$

where we have $S = an + H$.

PROOF. We set $U = (1/n)|na - \sum_{i \leq n} 1_{C_0}(X_i)|$. On each set Ω_I, we define $W = F_1 + F_2$, where F_1 and F_2 are given by (5.1) and (5.2). Thus we have

$$\{U \geq w\} = \bigcup \{\Omega_I; |\text{card } I - n\alpha| \geq w\}.$$

Thus

$$P(\{U \geq w\}) \leq \sum_{l \geq n(w+\alpha)} \alpha^l (1-\alpha)^{n-l} \binom{n}{l} + \sum_{l \leq n(\alpha - w)} \alpha^l (1-\alpha)^{n-l} \binom{n}{l}$$

$$= \sum_{l \geq n(w+\alpha)} \alpha^l (1-\alpha)^{n-l} \binom{n}{l} + \sum_{l \geq n(w+1-\alpha)} (1-\alpha)^l \alpha^{n-l} \binom{n}{l},$$

so that, by Corollary 4.2, we have

$$P(\{U \geq w\}) \leq \frac{K}{\sqrt{n}\,u} \exp(-nh(u) + 5nu(u - w)).$$

The result then follows from the fact that

$$P(W \geq t | \Omega_I) \leq K \exp(-\varphi(t))$$

whenever $I \subset \{1, \dots, n\}$. □

We now go back to the proof of Theorem 5.1. We keep the notation of Proposition 5.4.

Consider the function φ defined in the proof of that proposition. The function $\varphi(t)/t$ increases. We define u by $u(1 + 2a/a_0) = M/\sqrt{n}$, and we note that $u \leq 1$. We consider the smallest $d \geq 1/u$ such that $\varphi(d)/d \geq 11u$.

Suppose that $nU(1 + 2a/a_0) + W \geq M\sqrt{n}$.

Consider the smallest $l \geq 0$ such that

$$nU\left(1 + \frac{2a}{a_0}\right) \geq M\sqrt{n} - (l+1)d.$$

Then, if $l > 0$, we have

$$nU\left(1 + \frac{2a}{a_0}\right) \leq M\sqrt{n} - ld,$$

so that $W \geq ld$. Thus, since $W \geq 0$, we can find $l \geq 0$ such that

$$W \geq ld, \qquad U \geq u - \frac{(l+1)d}{n}.$$

By Proposition 5.4 this implies

$$\begin{aligned}\tau(M) &:= P\left(\left\| \sum_{i \leq n} 1_C(X_i) - nP(C) \right\|_{\mathscr{C}} \geq M\sqrt{n}\right) \\ &\leq \frac{K}{\sqrt{n}\,u} \exp(-nh(u)) \sum_{l \geq 0} \exp(5u(l+1)d - \varphi(ld)).\end{aligned}$$

For $l \geq 1$, we have

$$\varphi(ld) \geq l\varphi(d) \geq 11uld \geq 5u(l+1)d + lud$$

so that

$$\sum_{l \geq 1} \exp(5u(l+1)d - \varphi(ld)) \leq \sum_{l \geq 1} \exp(-lud) \leq K$$

as $ud \geq 1$. Thus we have

$$\tau(M) \leq \frac{K}{\sqrt{n}\,u} \exp(-nh(u)) \exp(5ud).$$

To estimate d, we use (3.10) with $t = uK(K_4)\sqrt{S}$ to see that $\varphi_{K_4, S}(d_0) \geq 11d_0u$, where $d_0 = KuS$, so that we have $d \leq \max(1/u, KH, KuS)$, and thus

$$ud \leq K\left(1 + \frac{HM}{\sqrt{n}} + aM^2\right).$$

To conclude the proof, it remains to evaluate $\exp(-nh(u))$. We write

$$h(u) \geq h\left(\frac{M}{\sqrt{n}}\right) + \left(u - \frac{M}{\sqrt{n}}\right)h'\left(\frac{M}{\sqrt{n}}\right)$$
$$\geq h\left(\frac{M}{\sqrt{n}}\right) - 5\left(\frac{M}{\sqrt{n}} - u\right)\frac{M}{\sqrt{n}}$$
$$\geq h\left(\frac{M}{\sqrt{n}}\right) - Ka\frac{M^2}{n}.$$

Theorem 5.1 is proved. □

6. Classes of sets. Our aim is to prove Theorem 1.1. The hardest case is (i), where condition (1.2) holds. Until further notice, we assume that $\mathscr{C}$ is a class of sets that satisfies (1.2), where $V \geq e$. The necessary modifications to cover case (ii) will be indicated later on.

LEMMA 6.1. *Consider a class $\mathscr{C}$ of sets that satisfies* (1.2) *and points $x_1, \dots, x_n$ of Ω. Set*

$$b = \left\| \sum_{i \leq n} 1_C(x_i) \right\|_{\mathscr{C}}.$$

Then we have

$$E\left\| \sum_{i \leq n} \varepsilon_i 1_C(x_i) \right\|_{\mathscr{C}} \leq K\sqrt{bv \log \frac{Vn}{b}}.$$

PROOF. Consider the distance δ on C given by

$$\delta(A, B) = \left(\operatorname{card}\{i \leq n; x_i \in A \triangle B\}\right)^{1/2}.$$

Thus we have $\delta(A, B) = \sqrt{n d_Q(A, B)}$, where d_Q is the distance on $\mathscr{C}$ given by (1.3), with $Q = (1/n)\sum_{i \leq n}\delta_{X_i}$. By (1.2), we have

$$N(\mathscr{C}, \delta, \varepsilon) = N\left(\mathscr{C}, d_Q, \varepsilon^2/n\right) \leq \left(\frac{Vn}{\varepsilon^2}\right)^v. \tag{6.1}$$

The diameter of $\mathscr{C}$, for the distance δ, is at most $\sqrt{2b}$. By Proposition 2.1, we have

$$E\left\| \sum_{i \leq n} \varepsilon_i 1_C(x_i) \right\|_{\mathscr{C}} \leq \int_0^{\sqrt{2b}} \sqrt{\log N(\mathscr{C}, \delta, \varepsilon)}\, d\varepsilon.$$

Using (6.1), the result follows by a routine computation. □

PROPOSITION 6.2. *Set* $a = \sup_{C \in \mathscr{C}} P(C)$. *Then we have*

$$E\Big\| \sum_{i \le n} 1_C(X_i) \Big\|_{\mathscr{C}} \le 2na + Kv \log \frac{V}{a}, \tag{6.2}$$

$$E\Big\| \sum_{i \le n} \varepsilon_i 1_C(X_i) \Big\|_{\mathscr{C}} \le K\sqrt{vn} \left(\left(a + \frac{v}{n} \log \frac{V}{a} \right) \log \frac{V}{a} \right)^{1/2}. \tag{6.3}$$

PROOF. *Step 1*. Consider the r.v. $b = \|\sum_{i \le n} 1_C(X_i)\|_{\mathscr{C}}$.
From Lemma 6.1, we have

$$E_\varepsilon \Big\| \sum_{i \le n} \varepsilon_i 1_C(X_i) \Big\|_{\mathscr{C}} \le K \sqrt{bv \log \frac{Vn}{b}}. \tag{6.4}$$

The function $x \to x \log(Vn/x)$ increases for $x \le n$ (provided $V \ge e$). Thus, if we set $b' = \max(b, Eb)$, we have

$$\begin{aligned} E_\varepsilon \Big\| \sum_{i \le n} \varepsilon_i 1_C(X_i) \Big\|_{\mathscr{C}} &\le K \sqrt{b'v \log \frac{Vn}{b'}} \\ &\le K \sqrt{b'v \log \frac{Vn}{Eb}}. \end{aligned}$$

Since $Eb' \le 2Eb$ and since $E\sqrt{b'} \le \sqrt{Eb'}$, *we get*

$$E \Big\| \sum_{i \le n} \varepsilon_i 1_C(X_i) \Big\|_{\mathscr{C}} \le K \sqrt{Ebv \log \frac{Vn}{Eb}}. \tag{6.5}$$

Step 2. By Corollary 3.2, we have

$$Eb \le na + K \sqrt{Ebv \log \frac{Vn}{Eb}}.$$

If $Eb \ge 2na$, we have $Eb - na \ge Eb/2$, so that since $Vn/Eb \le V/a$,

$$Eb \le K \sqrt{Ebv \log \frac{V}{a}}$$

and thus $Eb \le Kv \log(V/a)$. Thus, in any case, we have (6.2). And (6.3) then follows from (6.5). □

We now show that we need only be concerned with sets C such that $a_0 \le P(C) \le 1 - a_0$.

PROPOSITION 6.3. *Consider a class* $\mathscr{C}$ *of sets that satisfies* (1.2). *Assume that* $\sup_{C \in \mathscr{C}} P(C) \le a_0$. *For* $M \ge 0$, *we set*

$$\tau(M) = P\left(\Big\| \sum_{i \le n} 1_C(X_i) - nP(C) \Big\|_{\mathscr{C}} > M\sqrt{n} \right).$$

Then we have

$$(6.6) \qquad \forall\, n \geq 1, \forall\, M > K_4\sqrt{v \log V}, \qquad \tau(M) \leq \frac{K}{M}\exp(-2M^2).$$

PROOF. Set

$$H = E\left\| \sum_{i \leq n} \varepsilon_i 1_C(X_i) \right\|_{\mathscr{C}}.$$

It follows from Corollary 3.7 that (if K is large enough) (6.6) holds whenever

$$(6.7) \qquad n \geq K_3 H, \qquad M\sqrt{n} \geq K_3 H.$$

Now, it follows from (6.3) that

$$H \leq K\sqrt{vn}\left(\left(a_0 + \frac{v}{n}\log\frac{V}{a_0}\right)\log\frac{V}{a_0}\right)^{1/2},$$

so that (6.7) holds as soon as $n \geq K_4 v \log V$, $M \geq K_4\sqrt{v \log V}$. Since, as already noted, we have $n \geq M\sqrt{n}$, we have $n \geq K_4 v \log V$ as soon as $M \geq K_4\sqrt{v \log V}$. □

After these preliminaries, we go back to the proof of Theorem 1.1. We assume $M \geq K_4\sqrt{v \log V}$.

By Corollary 4.2, we can assume $a_0 \leq P(C) \leq 1 - a_0$ for $C \in \mathscr{C}$. The method of proof is similar to that of Theorem 2.4. We will split $\mathscr{C}$ into sets to which Theorem 5.1 will apply. We will take $a = v/M^2$.

Consider the smallest p such that $4^{-p+1} \leq a$. Consider $q \geq p$, which will be determined later. We construct the partitions of $T = \mathscr{C}$ given by Corollary 2.6, with $k_l = [3 \cdot 4^v] \geq 2 \cdot 4^v$. Thus $N = \operatorname{card} \mathscr{P}_p \leq (KVM^2/v)^v$. Consider the atoms $(\mathscr{C}_j)_{j \leq N}$ of $\mathscr{P}_p$. Set

$$H_j = E\left\| \sum_{i \leq n} \varepsilon_i 1_{C \,\triangle\, C_j}(X_i) \right\|_{\mathscr{C}_j},$$

where C_j is an arbitrary element of $\mathscr{C}_j$. Set $H = \sup_{j \leq n} H_j$. We will apply Theorem 5.1 to each class $\mathscr{C}_j$. Thus we need to control

$$(6.8) \qquad \frac{K_4 M H}{\sqrt{n}} - \frac{M^4}{4n}.$$

For this, we evaluate H.

PROPOSITION 6.4. *Assume* $3(q-p)v \leq n4^{-q}$. *Then*

$$(6.9) \qquad H \leq K\left(\sqrt{nv}\left[\sqrt{4^{-p}} + \sqrt{q4^{-q}} + \sqrt{4^{-q}\log V}\right] + qv + v\log V\right).$$

COMMENT. This proposition replaces (2.16) in the Gaussian case, in the proof of Theorem 2.4. If we had $H \le K\sqrt{n\sigma 4^{-p}}$, the proof would be much simpler. The larger M, the larger are the extra terms of (6.9). It will require rather significant work to show that the term $-M^4/4n$ in (6.8) absorbs these extra terms for all values of $M \le \sqrt{n}$. (This would be easier to do if one restricted to the case $M \le n^{1/4}$.)

PROOF. Let us fix $j \le n$. Then, by the usual chaining argument, we have

$$H_j \le \sum_{p<l\le q} E\left\| \sum_{i\le n} \varepsilon_i f(X_i) \right\|_{\mathscr{D}_l} + E\sup_{r\le R}\left\| \sum_{i\le n} \varepsilon_i f(X_i) \right\|_{\mathscr{G}_r}.$$

There, card $\mathscr{D}_l \le (3\cdot 4^v)^{l-p}$, and each function f in $\mathscr{D}_l$ is of the type $1_C - 1_{C'}$, where $P(C \vartriangle C') \le 4^{-l+1}$. Also, $R \le (3\cdot 4^v)^{q-p}$, and, for a certain set $C_r \in \mathscr{C}$, $\mathscr{G}_r$ consists of the functions $1_C - 1_{C_r}$ for $C \in \mathscr{C}_r$, where

$$\mathscr{C}_r = \{C \in \mathscr{C};\, P(C \vartriangle C_r) \le 4^{-q+1}\}.$$

Let us observe that

$$\left\| \sum_{i\le n} \varepsilon_i (1_C - 1_{C_r})(X_i) \right\|_{\mathscr{C}_r} \stackrel{\mathscr{D}}{=} \left\| \sum_{i\le n} \varepsilon_i 1_{C\vartriangle C_r}(X_i) \right\|_{\mathscr{C}_r}, \tag{6.10}$$

where $\stackrel{\mathscr{D}}{=}$ means equality in distribution. Let us also observe that the class $\{C \vartriangle C_r,\ C \in \mathscr{C}_r\}$ consists only of sets of probability $\le 4^{-q+1}$ and satisfies (2.1).

Now we have to evaluate the expectation of the supremum of a family of r.v.'s when we control their tails. This is standard.

PROPOSITION 6.5. *Consider (not necessarily independent) r.v.* $(Z_r)_{r\le R}$. *Assume that for numbers A, L and B we have*

$$t \ge A \Rightarrow P(|Z_r| \ge t) \le \exp(-\varphi_{L,B}(t)). \tag{6.11}$$

Then, if $\log R \le B$, *we have*

$$E\max_{r\le R}|Z_r| \le K\left(A + L\sqrt{B\log R}\right).$$

PROOF. We write, for any number $W \ge A$,

$$\begin{aligned} E\max_{r\le R}|Z_r| &= \int_0^\infty P\left(\max_{r\le R}|Z_r| \ge t\right) dt \\ &\le \int_0^\infty \min\left(1, \sum_{r\le R} P(|Z_r| \ge t)\right) dt \\ &\le W + \int_W^\infty R\exp(-\varphi_{L,B}(t))\, dt. \end{aligned}$$

Since $\varphi_{L,B}(t)/t$ increases, we have

$$\int_W^\infty \exp(-\varphi_{L,B}(t))\,dt \le \int_W^\infty \exp\left(-t\frac{\varphi_{L,B}(W)}{W}\right)dt$$
$$= \frac{W}{\varphi_{L,B}(W)}\exp(-\varphi_{L,B}(W)).$$

We take now $W = \max(A, L\sqrt{B\log R})$, so that if $\log R \le B$, we have

$$\varphi_{L,B}(W) \ge \varphi_{L,B}\left(L\sqrt{B\log R}\right) = \log R \ge \log 2. \qquad \square$$

We return to the proof of Proposition 6.4.

In the case of $\mathscr{D}_l$, it follows from Theorem 3.5 (or, if one prefers, from Bernstein's inequality!) that, for $f \in \mathscr{D}_l$, if we set $a = 4^{-l+1}$, $Z = |\sum_{i\le n}\varepsilon_i f(X_i)|$ satisfies (6.11) with

$$L \le K, \qquad A \le K\sqrt{na}, \qquad B \le na + \sqrt{na}.$$

Since $v \ge 1$, we have $\log\operatorname{card}\mathscr{D}_l \le 3(l-p)v$. Thus, if we have $3(l-p)v \le na$ and $na \ge 1$, we have

$$E\left\|\sum_{i\le n}\varepsilon_i f(X_i)\right\|_{\mathscr{D}_l} \le K\left(\sqrt{na} + \sqrt{na}\sqrt{\log\operatorname{card}\mathscr{D}_l}\right)$$
$$\le K\sqrt{n4^{-l}}\sqrt{(l-p)v}.$$

By (6.10), Proposition 6.2 and Theorem 3.5, the variables

$$Z_r = \left\|\sum_{i\le n}\varepsilon_i(1_C - 1_{C_r})(X_i)\right\|_{\mathscr{C}_r}$$

satisfy (6.11) for $L \le K$, $B = na + A$, where $a = 4^{-q+1}$ and

$$A = K\sqrt{nv}\left(\left(a + \frac{v}{n}\log\frac{V}{a}\right)\log\frac{V}{a}\right)^{1/2}. \tag{6.12}$$

Thus, by Proposition 6.5, provided that

$$3(q-p)v \le B,$$

we have

$$E\sup_{r\le R}\left\|\sum_{i\le n}\varepsilon_i f(X_i)\right\|_{\mathscr{G}_r} = K\left(A + \sqrt{v(q-p)(A+na)}\right)$$
$$\le K\left(A + \sqrt{v(q-p)A} + \sqrt{v(q-p)na}\right).$$

Since $a = 4^{-q+1} \le 4\cdot 4^{p-q}$, we have

$$\log\frac{V}{a} \ge \frac{1}{K}(q-p),$$

so that $KA \geq \sqrt{vna(q-p)}$, $KA \geq v(q-p)$, and thus

$$E \sup_{r \leq R} \left\| \sum_{i \leq n} \varepsilon_i f(X_i) \right\|_{\mathscr{G}_r} \leq KA \leq K\left(\sqrt{vn4^{-q}(q+\log V)} + v(q+\log V)\right)$$

$$\leq K\left(\sqrt{vn4^{-q}q} + \sqrt{vn4^{-q}\log V} + vq + v\log V\right).$$

To complete the proof, it suffices to observe that

$$\sum_{l \geq p} 4^{-l/2} \leq K\sqrt{4^{-p}}.$$

Proposition 6.4 is proved. □

Since $4^{-p} \leq v/M^2$, we are left with the task to show that we can select $q \geq p$ such that

$$3(q-p)v \leq n4^{-q} \tag{6.13}$$

and that we can control

$$R = \frac{KM}{\sqrt{n}}\left(\sqrt{vqn4^{-q}} + vq + \sqrt{vn4^{-q}\log V} + v\log V\right) - \frac{M^4}{4n}.$$

For that purpose, we simply pick the largest q for which (6.13) holds. We observe that

$$3(q-p)4^{q-p} \leq \frac{n}{v}4^{-p} \leq \frac{n}{M^2},$$

so that $q - p \leq K\log(Kn/M^2)$. Also, by the definition of q, we have

$$n4^{-q-1} \leq 3(q+1-p)v \leq 3qv, \tag{6.14}$$

so that, using that $\sqrt{q\log V} \leq 2(q + \log V)$, we have

$$R \leq \frac{KMv}{\sqrt{n}}(q + \log V) - \frac{M^4}{4n}.$$

We first bound

$$R_1 = \frac{KMv}{\sqrt{n}}q - \frac{M^4}{8n}.$$

Since $4^{-p+1} \geq v/M^2$, we have $p \leq K\log(4M^2/v)$, so that, writing $q = p + (q-p)$, we have $q \leq K\log(Kn/v)$ and

$$R_1 \leq K\frac{Mv}{\sqrt{n}}\log\frac{Kn}{v} - \frac{M^4}{8n}.$$

Taking the supremum over M of the right-hand side yields

$$R_1 \leq Kv\left(\frac{v}{n}\right)^{1/3}\left(\log K\frac{n}{v}\right)^{1/3} \leq Kv,$$

since $n \geq v$, as follows from the inequalities $n \geq \sqrt{M}$, $M \geq K_4\sqrt{\sigma\log V}$.

Thus we have

$$R \leq Kv + \frac{KMv}{\sqrt{n}} \log V - \frac{M^4}{8n}.$$

Observe that, if $M \leq \sqrt{n}/\log V$, we have $R \leq Kv$. On the other hand, taking the supremum over n of the right-hand side gives

$$R \leq Kv\left[1 + \frac{v}{M^2}(\log V)^2\right].$$

Thus we have proved the following.

THEOREM 6.6. *Under condition* (i) *of Theorem* 1.1, *if* $M \geq K_4\sqrt{v \log V}$ *and if either* $n \geq M^2(\log V)^2$ *or* $M \geq \sqrt{v}\log V$, *we have*

$$\tau_{\mathscr{C}}(M) \leq \frac{K}{M}\left(\frac{KVM^2}{v}\right)^v e^{-2M^2},$$

where K *is universal. To deduce* (1.4), *it then suffices to take* $K(V)$ *large enough that the right-hand side of* (1.4) *is greater or equal to one when* $M \leq \sqrt{v}\log V$.

We now turn to the proof of Theorem 1.1 under condition (ii). For a class $\mathscr{F}$ of functions and $\varepsilon > 0$, we denote by $N_{[\,]}(\mathscr{F}, \varepsilon)$ the smallest number of brackets $[f_1, f_2]$, such that $E(f_2 - f_1)^2 \leq \varepsilon^2$, needed to cover $\mathscr{F}$. An essential ingredient of the proof is as follows.

PROPOSITION 6.7 (Ossiander's bracketing theorem [11]).

$$E\left\|\sum_{i \leq n} f(X_i) - nEf\right\|_{\mathscr{F}} \leq K\sqrt{n}\int_0^\infty \sqrt{\log N_{[\,]}(\mathscr{F}, \varepsilon)}\, d\varepsilon.$$

Suppose now that $\mathscr{C}'$ is a class of sets that satisfies condition (ii) of Theorem 1.1, and that moreover all the sets C of $\mathscr{C}'$ are contained in a certain set C_0 with $P(C_0) = b$. Then, by Proposition 6.7, using a computation similar to that of Lemma 6.1, we get

$$E\left\|\sum_{i \leq n} f(X_i) - nEf\right\|_{\mathscr{C}'} \leq K\sqrt{vnb \log \frac{V}{b}},$$

so that, by Lemma 2.7 of [5], we have

$$E\left\|\sum_{i \leq n} \varepsilon_i f(X_i)\right\|_{\mathscr{C}'} \leq K\sqrt{vnb \log \frac{V}{b}}. \tag{6.15}$$

This in particular applies to $\mathscr{C}' = \mathscr{C}$ when $b = 1$, so that

$$E\left\|\sum_{i \leq n} \varepsilon_i f(X_i)\right\|_{\mathscr{C}} \leq K\sqrt{vn \log V}.$$

The estimate (6.15), used instead of (6.3), is sufficient to make Proposition 6.3 work [by first breaking $\mathscr{C}$ into pieces to which (6.15) applies]. The rest of the proof of Theorem 1.1 is nearly identical to what it was in case (i), with the major difference that each class $\mathscr{C}_r$ is now contained in brackets $[C_1, C_2]$ with $P(C_2 \setminus C_1) \le 4^{-q}$ and satisfies $N_{[\,]}(\mathscr{C}_r, \sqrt{\varepsilon}) \le (V/\varepsilon)^v$. Thus, in the proof of Proposition 6.4, the quantity A for (6.12) can now be replaced by

$$K\sqrt{nv4^{-q}\log(4^q V)}\,,$$

and the conclusion of Proposition 6.4 can be reinforced to

$$H \le K\Big(\sqrt{nv4^{-q}q} + K\sqrt{nv4^{-q}\log V}\Big).$$

We then take q as before, so that by (6.14), we get

$$n4^{-q} \le Kv\log\frac{n}{M^2},$$

so that

$$R \le Kv + K\frac{Mv}{\sqrt{n}}\sqrt{\log\frac{n}{M^2}\log V} - \frac{M^4}{4n}.$$

It is simple to see that if $n \ge M^2\log V\log\log V$, then $R \le Kv$.

Also, it is simple to see that

$$R \le Kv\left(1 + \frac{v}{M^2}\log\frac{M^2}{v}\log V\right),$$

so that $R \le Kv$ for $M \ge \sqrt{v\log V\log\log V}$. Thus we have shown the following result.

THEOREM 6.8. *In case* (ii) *of Theorem* 1.1, *if* $M \ge K\sqrt{v\log V}$ *and if either* $n \ge M^2\log V\log\log V$ *or* $M \ge \sqrt{v\log V\log\log V}$, *we have*

$$\tau_{\mathscr{C}}(M) \le \frac{K}{M}\left(\frac{KVM^2}{v}\right)^v e^{-2M^2},$$

where K is universal.

We now explain how to prove Theorem 1.2. We observe that, for all β and u,

$$\begin{aligned}(u-\beta)^3 + \beta^3 &= u^3 - 3\beta u^2 + 3\beta^2 u\\ &= u(u^2 - 3\beta u + 3\beta^2) \ge \frac{u^3}{8} + \frac{\beta^2 u}{4},\end{aligned}$$

since

$$u^2 - 3\beta u + 3\beta^2 \ge \frac{u^2}{8} + \frac{\beta^2}{4} \quad \text{for all } \beta.$$

Thus, by (4.2), we get

$$\frac{\partial \Psi}{\partial u}(u,\alpha) \geq 4u + \frac{u^3}{8} + \frac{u}{4}\left(\frac{1}{2} - \alpha\right)^2$$

and thus

$$\Psi(u,\alpha) \geq 2u^2\left[1 + \frac{1}{16}\left(\frac{1}{2} - \alpha\right)^2\right] + \frac{u^4}{32}.$$

Only the obvious changes are needed to the proof of Theorem 5.1 to obtain that if moreover all the sets in $\mathscr{C}$ satisfy $|P(C) - \frac{1}{2}| \geq \beta$, then the conclusion still holds with the term $e^{-2M^2(1+\beta^2/16)}$ instead of e^{-2M^2}. Again, only obvious changes are needed to show that if $\mathscr{C}$ satisfies either condition of Theorem 1.1 and moreover all the sets in $\mathscr{C}$ satisfy $|P(C) - \frac{1}{2}| \geq \beta$, the conclusion holds with the term $e^{-2M^2(1+\beta^2/16)}$ instead of e^{-2M^2}. The dependence in β is certainly not sharp, but this statement suffices to derive Theorem 1.2 the way Proposition 2.8 follows from Theorem 2.4.

We now discuss some possible variations on Theorem 1.1. One such variation in case 2 is to consider more general ways to control the bracketing entropy than just assuming polynomial decay. One can, for example, assume that

$$N_{[\,]}(\mathscr{C},\varepsilon) \leq \varphi(\varepsilon),$$

where φ satisfies $\varphi(\varepsilon/2) \leq A\varphi(\varepsilon)$ (polynomial-type control) or $\log\varphi(\varepsilon/2) \leq \beta\log\varphi(\varepsilon)$ for some $\beta < 4$ (exponential-type control). The main difference from Theorem 1.1 is that the statements are not so clean, since it is harder to optimize the size of the pieces in which $\mathscr{C}$ will be broken for applications of Theorem 5.1. Let us also note that in the case where we assume $\log\varphi(\varepsilon/2) \leq \beta\log\varphi(\varepsilon)$, Lemma 2.5 is not needed, and the proof greatly simplifies.

Another variation would be to assume that all the sets in $\mathscr{C}$ satisfy $P(C) \leq a < \frac{1}{2}$ and to obtain a bound for $\tau_{\mathscr{C}}(M)$ of the type

$$\text{Polynomial term in } M \times \frac{K}{M}\exp\left(-n\Psi\left(\frac{M}{\sqrt{n}}, a\right) + \text{perturbation term}\right),$$

with the smallest possible perturbation term. In order to get a clean result, when $a = \frac{1}{2}$, we have replaced $-n\Psi(M/\sqrt{n}, \frac{1}{2})$ by $-2M^2 - M^4/4n$ and we have arranged that this little room given by the extra term $M^4/4n$ kills all the perturbations. It seems, however, that one can go through the same proof keeping the term $-n\Psi(M/\sqrt{n}, a)$ and tracking the perturbations (although the detailed careful computations remain to be done).

7. Moment generating functions. When the function f is not the indicator of a set, it is not obvious how to obtain results conditioned on the event $\{\sum_{i\leq n} f(X_i) - nE(f) \geq M\sqrt{n}\}$. A substitute for this is obtained through the next result.

THEOREM 7.1. *Consider a function f on a probability space (Ω, P) and assume that $0 \le f \le 1$. Set $a = Ef$ and assume that $a_0 \le a \le 1 - a_0$.*

Consider a class $\mathscr{G}$ of functions on Ω. Assume that for all g in $\mathscr{G}$, we have $Eg = Efg = 0$. Set

$$H = E\left\| \sum_{i \le n} \varepsilon_i g(X_i) \right\|_{\mathscr{G}}, \qquad \sigma = \sup_{g \in \mathscr{G}} \left(Eg^2\right)^{1/2}, \qquad b = \sup_{g \in \mathscr{G}} \|g\|_\infty.$$

Let $S = a\sigma^2 + bH$.

Then, for $t \ge 0$, $u \ge 0$, we have

$$(7.1) \qquad \begin{aligned} &P\left(\sum_{i \le n} f(X_i) - nEf \ge t; \left\| \sum_{i \le n} g(X_i) \right\|_{\mathscr{G}} \ge u \right) \\ &\qquad \le K \exp\left(-n\Psi\left(a, \frac{t}{n}\right) - \varphi(u) \right), \end{aligned}$$

where $\varphi(u) = 0$ if $u \le K_4 H$, and $\varphi(u) = \varphi_{K_4, S}(u)$ for $u \ge K_4 H$.

COMMENT. It should be said that moment generating functions are not a sharp tool. For example, they do not allow us to capture the correct factor in front of the exponential in (4.1) (see [7]), thereby creating an irretrievable loss of one power of M in (1.8). The use of this technique is, however, motivated by the success of Theorem 7.1.

PROOF. *Step* 1. By approximation, we can assume that $\mathscr{G}$ is finite. Consider the set

$$\mathscr{H} = \{h : 0 \le h \le 1, Eh = a, \forall g \in \mathscr{G}, Ehg = 0\}.$$

This is a convex subset of the unit ball of $L^\infty(P)$ that contains f. Provided with the weak* topology $\sigma(L^\infty(P), L^1(P))$, this is a compact set. Let us fix λ and μ and consider the function θ on $\mathscr{H}$ given by

$$(7.2) \qquad \theta(h) = E \exp\left(\lambda\left(\sum_{i \le n} h(X_i) - na \right) + \mu \left\| \sum_{i \le n} g(X_i) \right\|_{\mathscr{G}} \right).$$

Since the exponential is convex, θ attains its maximum at an extreme point of $\mathscr{H}$.

Step 2. We show that an extreme point of $\mathscr{H}$ is of the type 1_A. Indeed, consider $h \in \mathscr{H}$ and assume that $P(B) > 0$, where $B = \{0 < h < 1\}$. Then, for some $\varepsilon > 0$, we have $P(B_\varepsilon) > 0$, where $B_\varepsilon = \{\varepsilon < h < 1 - \varepsilon\}$. Since we assume that $\mathscr{G}$ is finite and since there is no loss of generality to assume that P has no atoms, we can find a function w that is 0 outside B_ε, such that $\|w\|_\infty \le \varepsilon$, $Ew = 0$ and $E(wg) = 0$ for all g in $\mathscr{G}$. Then $h + w$ and $h - w$ both belong to $\mathscr{H}$, so that h is not an extreme point.

Step 3. Thus we have shown that for some set A with $1_A \in \mathscr{H}$, we have $\theta(f) \le \theta(1_A)$. Since $1_A \in \mathscr{H}$, we have $P(A) = a$, and $E(g1_A) = 0$ for all g in $\mathscr{G}$.

Thus $E(g1_{\Omega\setminus A}) = 0$ for all g in $\mathscr{G}$. Consider now, for a subset I of $\{1,\dots,n\}$, the event

$$\Omega_I = \{i \in I \Leftrightarrow X_i \in A\}.$$

To compute $\theta(1_A)$, we write

$$\begin{aligned}\theta(1_A) &= \sum_{I\subset\{1,\dots,n\}} P(\Omega_I) E\left(\exp\lambda\left(\sum_{i\le n} 1_A(X_i) - na\right)\right.\\ &\qquad\qquad \left. + \mu\left\|\sum_{i\le n} g(X_i)\right\|_{\mathscr{G}}\middle|\Omega_I\right)\\ &\le \sum_{I\subset\{1,\dots,n\}} P(\Omega_I)\exp\lambda(\operatorname{card} I - na)\\ &\qquad \times E\left(\exp\left(\mu\left\|\sum_{i\in I} g(X_i)\right\|_{\mathscr{G}}\right)\exp\left(\mu\left\|\sum_{i\notin I} g(X_i)\right\|_{\mathscr{G}}\right)\middle|\Omega_I\right).\end{aligned}\tag{7.3}$$

Consider now the probabilities on Ω:

$$P_1(C) = P(C\cap A)/P(A),$$
$$P_2(C) = P(C\setminus A)/(1-P(A)).$$

Consider $Y_1,\dots,Y_n; Z_1,\dots,Z_n$ that are i.i.d. distributed like P_1 (resp. P_2).

From (7.3) we get

$$\begin{aligned}\theta(1_A) &\le \sum_{k\le n} a^k(1-a)^{n-k}\binom{n}{k}\exp(\lambda(k-na))E\exp\mu\left\|\sum_{i\le k} g(Y_i)\right\|_{\mathscr{G}}\\ &\qquad\times E\exp\mu\left\|\sum_{i\le n-k} g(Z_i)\right\|_{\mathscr{G}}.\end{aligned}\tag{7.4}$$

Step 4. Consider now the function

$$\xi(\mu) = \xi_K(\mu) = \int_{-\infty}^{KH}\mu e^{\mu x}\,dx + \int_{KH}^{\infty}\mu e^{\mu x}\exp(-\varphi_{K,S}(x))\,dx.\tag{7.5}$$

If we use Lemma 5.3 as well as Theorem 3.5 (after rescaling), we see that for K sufficiently large, the last two terms on the right of (7.4) are dominated by $\xi(\mu)$. Thus we have

$$\begin{aligned}\theta(f) \le \theta(1_A) &\le \xi(\mu)^2\sum_{k\le n} a^k(1-a)^{n-k}e^{\lambda(k-na)}\\ &= \xi(\mu)^2\left((1-a)e^{-a\lambda} + ae^{\lambda(1-a)}\right)^n.\end{aligned}$$

Step 5. By Chebyshev's inequality we have

$$\begin{aligned}P&\left(\sum_{i\le n} f(X_i) - nEf \ge t;\ \left\|\sum_{i\le n} g(X_i)\right\|_{\mathscr{G}} \ge u\right)\\ &\qquad\le e^{-(\lambda t+\mu u)}\xi(\mu)^2\left((1-a)e^{-a\lambda} + ae^{\lambda(1-a)}\right)^n.\end{aligned}$$

We observe (this is actually the derivation of the Chernoff bounds for the binomial law) that

$$\inf_{\lambda} e^{-\lambda t}\big((1-a)e^{-a\lambda}+ae^{\lambda(1-a)}\big)^{n}=\exp\left(-n\Psi\left(\frac{t}{n},a\right)\right).$$

If we observe that $\xi(\mu)=E\exp\mu Y$ for a certain r.v. Y, we see by Cauchy-Schwarz that $\xi(\mu)^2\le\xi(2\mu)$. Consider $u\ge 4KH$ [where K is the constant of (7.5)]. Then set

$$\mu=\frac{\varphi_{K,S}(u/4)}{u},$$

so that $\varphi_{K,S}(x)\ge 4\mu x$ for $x\ge u/4$. Thus

$$\int_{u/4}^{\infty}2\mu e^{2\mu x}\exp(-\varphi_{K,S}(x))\,dx\le\int_{u/4}^{\infty}2\mu e^{-2\mu x}\,dx\le e^{-\mu u/2}.$$

Thus

$$\xi(2\mu)\le\int_{-\infty}^{u/4}2\mu e^{2\mu x}\,dx+e^{-\mu u/2}\le 2e^{\mu u/2}$$

and

$$e^{-\mu u}\xi(2\mu)\le 2e^{-\mu u/2}=2\exp(-(1/2)\varphi_{K,S}(u/4)).$$

This completes the proof. □

We can now state and prove the basic inequality for classes of functions.

THEOREM 7.2. *Consider a class $\mathscr{F}$ of functions on Ω and assume that $0\le f\le 1$ for each f in F. Assume that $\sigma^2=\sup_{f\in\mathscr{F}}E(f-Ef)^2\ge a_0$. We set*

$$H=E\sup_{f,f'\in\mathscr{F}}\left|\sum_{i\le n}\varepsilon_i(f-f')(X_i)\right|,\qquad \rho=\sup_{f,f'\in\mathscr{F}}\big(E(f-f')^2\big)^{1/2}.$$

Then, provided that $\rho\le a_0$, we have

$$\tau_{\mathscr{F}}(M)\le Ke^{-2M^2}\exp\left(K\rho^2M+\frac{KMH}{\sqrt{n}}-\frac{M^4}{4n}\right).$$

PROOF. By approximation, we can assume that $\mathscr{F}$ is finite, so that there exists $f_1\in\mathscr{F}$ for which $\sigma^2=E(f_1-Ef_1)^2$.

For a function $f\in\mathscr{F}$, we set

$$\theta(f)=\frac{1}{\sigma^2}E\big((f-f_1-E(f-f_1))(f_1-Ef_1)\big).$$

Thus $|\theta(f)|\le\rho/\sigma\le\rho/a_0\le 1$ and $\theta(f)\le 0$ since $E(f-Ef)^2\le E(f_1-Ef_1)^2$.

For a function $f\in\mathscr{F}$, we write

$$g'(f)=f-f_1(1+\theta(f)),\qquad g(f)=g'(f)-E(g'(f)).$$

Consider the class $\mathscr{G}$ of all functions of the type $g(f)$ for $f \in \mathscr{F}$. We observe that for $g \in \mathscr{G}$ we have $Eg = 0$, $E(f_1 g) = 0$ (which follows from a straightforward computation).

Since $|\theta(f)| \le 1$, we have $\|g\|_\infty \le 4$ for $g \in \mathscr{G}$.

Also,

$$\left(Eg(f)^2\right)^{1/2} \le \left(Eg'(f)^2\right)^{1/2} \le \rho + \sigma(\rho/\sigma) \le 2\rho$$

and

$$E\left\| \sum_{i \le n} \varepsilon_i g(X_i) \right\|_{\mathscr{G}} \le E\left\| \sum_{i \le n} \varepsilon_i((f - f_1)(X_i) - E(f - f_1)) \right\|_{\mathscr{F}} + \sup_{f \in \mathscr{F}} \theta(f) E\left| \sum_{i \le n} \varepsilon_i(f_1(X_i) - Ef_1) \right|.$$

By an argument used in the proof of Theorem 3.5, the first term is less than $4H$. The second is less than or equal to $(\rho/\sigma)\sigma = \rho$. Since we certainly have $\rho \le KH$ (by Khintchine's inequality), we get

$$E\left\| \sum_{i \le n} \varepsilon_i g(X_i) \right\|_{\mathscr{G}} \le KH. \tag{7.6}$$

For $f \in \mathscr{F}$, we have

$$f = f_1(1 + \theta(f)) + g(f) + E(g'(f)),$$

so that, since $-1 \le \theta(f) \le 0$, $Eg(f) = 0$, we have

$$\begin{aligned} \sum_{i \le n} f(X_i) - nEf &= (1 + \theta(f))\Big(\sum_{i \le n} f_1(X_i) - nEf_1\Big) + \sum_{i \le n} g(f)(X_i) \\ &\le (1 + \theta(f))\Big(\sum_{i \le n} f_1(X_i) - nEf_1\Big) + \Big\| \sum_{i \le n} g(X_i) \Big\|_{\mathscr{G}} \qquad (7.7) \\ &\le \max\Big(0, \sum_{i \le n} f_1(X_i) - nEf_1\Big) + \Big\| \sum_{i \le n} g(X_i) \Big\|_{\mathscr{G}}. \end{aligned}$$

Consider a number $d > 0$ and suppose that $\sum_{i \le n} f_1(X_i) - nEf_1 \ge 0$. If l is the largest integer such that $\sum_{i \le n} f_1(X_i) - nEf_1 \le M\sqrt{n} - ld$, then by (7.7) we have

$$\sup_{f \in \mathscr{F}} \Big(\sum_{i \le n} f(X_i) - nEf \Big) \ge M\sqrt{n} \Rightarrow \Big\| \sum_{i \le n} g(X_i) \Big\|_{\mathscr{G}} \ge ld,$$

so that

$$\begin{aligned} &P\Big(\sup_{f \in \mathscr{F}} \Big(\sum_{i \le n} f(X_i) - nEf \Big) \ge M\sqrt{n} \Big) \\ &\quad \le \sum_{l \ge 0} P\Big(\sum_{i \le n} f(X_i) - nEf \ge M\sqrt{n} - (l+1)d; \Big\| \sum_{i \le n} g(X_i) \Big\|_{\mathscr{G}} \ge ld \Big) \\ &\qquad + P\Big(\Big\| \sum_{i \le n} g(X_i) \Big\|_{\mathscr{G}} \ge M\sqrt{n} \Big), \end{aligned}$$

where the last term occurs because of the case $\sum_{i \le n} f(X_i) - nEf \le 0$. To each term of the sum we apply Theorem 7 [using (7.6)], and we apply Theorem 3.5 to the last term. The computation then parallels that of the end of the proof of Proposition 5.4. [To control $\tau(M)$, we then apply the same argument to the class $\mathscr{F}' = \{1 - f;\ f \in \mathscr{F}\}$.] □

8. Control of the variance. Consider a class $\mathscr{F}$ of functions on Ω, such that $0 \le f \le 1$ for each $f \in \mathscr{F}$. Assume that

$$\forall f \in \mathscr{F}, \qquad E(f - Ef)^2 \le \sigma^2 < 1/4.$$

We would like, using this information, to improve upon Theorem 1.3. In the case of one single function, sharp bounds for

$$P\Big(\sum_{i \le n} f(X_i) - nEf \ge nt\Big) \tag{8.1}$$

are better expressed when $Ef = 0$. In that case, setting $\sigma^2 = Ef^2$, $b = \sup f$, Hoeffding [7] shows that

$$P\Big(\sum_{i \le n} f(X_i) \ge nt\Big) \le \exp(-n\theta(t, \sigma^2, b)), \tag{8.2}$$

where the function $\theta(t) = \theta(t, \sigma^2, b)$ is best understood by the relations $\theta(0) = \theta'(0) = 0$,

$$\theta''(t) = \frac{1}{\sigma^2 + t(b - \sigma^2/b) - t^2}.$$

It can be shown that $\theta(t, \sigma^2, b)$ is a decreasing function of σ, a fact that we will use many times.

Let us observe that

$$\theta(t) = \frac{t^2}{2\sigma^2} - \frac{t^3}{6\sigma^4}\left(b - \frac{\sigma^2}{b}\right) + O(t^4), \tag{8.3}$$

so that the influence of b on the right-hand side of (8.3) starts to be felt (unless $b - \sigma^2/b \le 0$) for $nt^3 \ge \sigma^4$, that is, when $t = M/\sqrt{n}$, for $M^3 \ge \sigma^4\sqrt{n}$. For values of M just slightly larger ($M^3 \ge \sigma^4\sqrt{n}\log n$), the influence of the second term on the right of (8.3) is more important than any polynomial term in M in front of the exponential. In other words, when proving bounds of the type

$$\tau_{\mathscr{G}}(M) \le (LM)^a \exp\left(-n\varphi\left(\frac{M}{\sqrt{n}}, \sigma^2\right)\right), \tag{8.4}$$

the number a becomes unimportant for $M^3 \ge \sigma^4\sqrt{n}\log n$, if $\theta(t, \sigma, 1) - \varphi(t, \sigma^2) \ge ct^3/\sigma^4 (c > 0)$. This is in particular the case if one uses for φ the function derived from the use of Bernstein's or Bennett's inequality for which one has (when $b = 1$)

$$\varphi(t, \sigma^2) = \frac{t^2}{2\sigma^2} - \frac{t^3}{6\sigma^4} + O(t^4).$$

This is an interesting contrast with the situation of Theorems 3.1 and 3.2, which corresponds to the use of (8.2) for $\sigma = b$, so that the approximation $\theta(t) = t^2/2\sigma^2$ of (8.2) remains good until $nt^4 \sim 1$, which corresponds to values of M of order $n^{1/4}$ rather than $n^{1/6}$ (a point at which, as should have been apparent from the proof, the sharpness of Theorems 1.1 and 1.3 becomes an illusion).

It must also be mentioned [again because of the contribution of the second term on the right of (8.3)] that for these large values of M, no argument using (8.2) will be reasonably sharp unless it always uses (8.2) with $b = 1 - Ef$ rather than $b = 1$. This is a level of sophistication the need for which has yet to be demonstrated. For these reasons, we will concentrate our effort on the values of M with $M^3 \le \sigma^4\sqrt{n}\,(nt^3 \le \sigma^4)$ and only briefly indicate what could be done for other values. For these values of M, one sees that changing b by a factor 2 will not matter much. Thus we will replace the class $\mathscr{F}$ by the class of functions $\{f - Ef,\ f \text{ in } \mathscr{F}\}$. In other words, we will assume that $\mathscr{F}$ consists of functions f for which $-1 \le f \le 1$ and $Ef = 0$, $Ef^2 \le \sigma^2$.

Let us observe that

$$\frac{1}{\sigma^2 + t(b - \sigma^2/b) - t^2} \ge \frac{1}{\sigma^2 + tb} \ge \frac{1}{\sigma^2} - \frac{tb}{\sigma^4},$$

so that

$$\theta(t, \sigma^2, b) \ge \frac{b^2}{2\sigma^2} - \frac{t^3 b}{6\sigma^3}.$$

In particular, if a function h satisfies $Eh = 0$, $Eh^2 \le \sigma^2$, $h \le b$, we have

$$P\Big(\sum_{i \le n} h(X_i) \ge nt\Big) \le \exp\Big(-\frac{nt^2}{2\sigma^2} + \frac{nt^3 b}{6\sigma^3}\Big). \tag{8.5}$$

We now come to a basic observation.

Lemma 8.1. *Consider a function f, $\|f\|_\infty \le 1$, $Ef^2 \le \sigma^2$, $Ef = 0$. Consider a function g, $\|g\|_\infty \le 2$, $Eg = 0$, $Eg^2 = \rho^2$, $E(fg) \le 0$. Consider $t, u \ge 0$. Suppose that*

$$nt^3 \le \sigma^4, \qquad 2^4 t^2/\sigma^2 \le u \le t\rho/\sigma. \tag{8.6}$$

Then we have

$$P\Big(\sum_{i \le n} f(X_i) \ge nt;\ \sum_{i \le n} g(X_i) \ge nu\Big) \le K \exp\Big(-\frac{nt^2}{2\sigma^2} - \frac{nu^2}{4\rho^2}\Big). \tag{8.7}$$

PROOF. Consider $\alpha > 0$. The left-hand side of (8.7) is at most

$$P\Big(\sum_{i \le n} (f + \alpha g)(X_i) \ge n(t + \alpha u)\Big).$$

Since $\alpha > 0$, $E(fg) \le 0$, we have $E(f + \alpha g)^2 \le \sigma^2 + \alpha^2\rho^2$. By (8.5) we have (since $|f + \alpha g| \le 1 + 2\alpha$)

$$P\Big(\sum_{i \le n} (f + \alpha g)(X_i) \ge n(t + \alpha u)\Big)$$
$$\le \exp\left(-\frac{n(t + \alpha u)^2}{2(\sigma^2 + \alpha^2\rho^2)} + \frac{n(t + \alpha u)^3(1 + 2\alpha)}{6\sigma^4}\right).$$

We take $\alpha = u\sigma^2/t\rho^2$. The first term inside the exponential becomes $-n(t^2/2\sigma^2 + u^2/2\rho^2)$. For the second term, since $u \le t\rho/\sigma$, we have $\alpha u \le t$, so that $n(t + \alpha u)^3/\sigma^4 \le 8nt^3/\sigma^4 \le 8$. Also, since $u \ge 2^4 t^2/\sigma^2$, we have $16t^3\alpha/\sigma^4 \le nu^2/\rho^2$, so that the second term inside the exponential is less than or equal to $8/6 + nu^2/6\rho^2$. □

It is well known that bounds of the type of Theorem 3.5 can be recovered by working through the usual chaining argument (see, e.g., [9], Chapter 11). The importance of Lemma 8.1 is that it allows us to mimic these arguments "conditionally on $\sum_{i \le n} f(X_i) \ge nt$."

Our objective now is to (indicate how to) prove that, for $M \le n^{1/6}\sigma^{4/3}$, under hypothesis (i) or (ii) of Theorem 1.1, we have

$$\tau_{\mathscr{F}}(M) \le \left(\frac{KVM\sigma^2}{\sqrt{v}}\right)^v \exp\left(-\frac{M^2}{2\sigma^2}\right) \tag{8.8}$$

for $M \ge K(V, v, \sigma)$. Here, and in the rest of this section, $K(V, v, \sigma)$ denotes a number, depending only on V, v, σ, which may vary at each occurrence. (Figuring out the best possible dependence given by this approach requires checking many computational details and more energy than the author has left at this point.) Let us fix M. Consider the largest p for which

$$\rho = 4^{-p+1} \ge \frac{\sigma^2\sqrt{v}}{M}.$$

The approach is (as usual) to cut $\mathscr{F}$ into $(KV/\rho)^v$ pieces $\mathscr{G}$ for which

$$\tau_{\mathscr{G}}(M) \le K^v \exp\left(-\frac{M^2}{2\sigma^2}\right). \tag{8.9}$$

Proceeding as in Section 6, $\mathscr{G}$ will have the following property: there is an increasing sequence of partitions $(\mathscr{P}_l)_{l \ge p}$ of $\mathscr{G}$ into less than $(2 \cdot 4^{l-p})^v$ atoms $\mathscr{G}_i$ for which

$$g, g' \in \mathscr{G}_i \Rightarrow \big(E(g - g')^2\big)^{1/2} \le 4^{-l+1}.$$

Moreover,

$$g, g' \in \mathscr{G} \Rightarrow \left(E(g - g')^2\right)^{1/2} \le \rho.$$

To prove (8.9), there is no loss of generality to assume that $\mathscr{G}$ is finite. Let us consider f in $\mathscr{G}$, such that

$$\forall\, g \in \mathscr{G}, \qquad Ef^2 \ge Eg^2.$$

Set $\mathscr{G}' = \{g - f;\ g \in \mathscr{G}\}$. As usual, we write

$$\begin{aligned} P\Big(\sup_{\mathscr{G}} \sum_{i \le n} g(X_i) \ge M\sqrt{n}\Big) &\le \sum_{1 \le l \le l_0} P\Big(\sum_{i \le n} f(X_i) \ge (M - l\rho\sqrt{v})\sqrt{n}, \\ &\qquad \sup_{\mathscr{G}'} \sum_{i \le n} g(X_i) \ge (l-1)\rho\sqrt{vn}\Big) \\ &\quad + P\Big(\sup_{\mathscr{G}'} \sum_{i \le n} g(X_i) \ge l_0\rho\sqrt{vn}\Big). \end{aligned} \tag{8.10}$$

The last term will be evaluated through Theorem 3.5. The inequality $l_0\sqrt{v} > KM/\sigma$ will suffice to make this term of smaller order. Thus we can assume $l_0\sqrt{v} \le KM/\sigma$. We fix an l such that $1 \le l \le l_0$, and set

$$t = (M - l\rho\sqrt{v})/\sqrt{n}\,, \ u = (l-1)\rho\sqrt{v/n}\,.$$

We have to get bounds for

$$P\Big(\sum_{i \le n} f(X_i) \ge nt;\ \sup_{\mathscr{G}'} \sum_{i \le n} g(X_i) \ge nu\Big). \tag{8.11}$$

We will assume $l > 1$ and leave the easy case $l = 1$ to the reader. The definition of ρ shows that if M is larger than a suitable constant $K(V, v, \sigma)$, we have $l_0\rho\sqrt{v} \le KM\rho/\sigma \le M/2$, so that we have $t \ge M/2\sqrt{n}$. We also observe that

$$\rho\sqrt{v/n} \le u \le KM\rho/\sigma\sqrt{n}\,. \tag{8.12}$$

We now use chaining. Consider $r \ge p$, and, for $p \le s \le r$ and each atom R of $\mathscr{P}_s$, let us select $h_R \in R$ such that $E(fh_R)$ is as large as possible. For $R \in \mathscr{P}_r$, set $\mathscr{G}(R) = \{g - h_R;\ g \in R\}$. For $R \in \mathscr{P}_{p+1}$, set $g_R = h_R - f$. For $R \in \mathscr{P}_s$, $p+1 < s < r$, set $g_R = h_R - h_{R'}$, where R' is the atom of $\mathscr{P}_{s-1}$ that contains R. Observe that by construction we have $E(fg_R) \le 0$. Consider a sequence $(u_s)_{p < s \le r}$ such that $\sum_{p < s \le r} u_s \le u/2$.

Thus, if $\sup_{\mathscr{G}'} \sum_{i \le n} g(X_i) \ge nu$, then we must have either $\sum_{i \le n} g_R(X_i) \ge nu_s$ for some $p + 1 \le s \le r$ and some atom R of $\mathscr{P}_s$, or $\sup_{\mathscr{G}(R)} \sum_{i \le n} g(X_i) \ge nu/2$ for some atom R of $\mathscr{P}_r$.

Setting $\rho_s = 4^{-s+1}$, we see that if

$$2^4 t^2/\sigma^2 \le u_s \le t\rho_{s-1}/\sigma, \tag{8.13}$$

then by (8.7) we have that the quantity (8.11) is bounded by

$$\sum_{p<s\le r} K(2\cdot 4^{s-p})^{v}\exp\left(-\frac{nt^2}{2\sigma^2}-\frac{nu_s^2}{4\rho_s^2}\right) + \sum_{R \text{ atom of } \mathscr{P}_r} P\left(\left\|\sum_{i\le n} g(X_i)\right\|_{\mathscr{G}(R)} \ge \frac{nu}{2}\right). \tag{8.14}$$

This last term is evaluated through Theorem 3.5. Consider $q \ge r$, such that $3(q-r)v \le n4^{-2q}$. The version of Proposition 6.4 for functions shows that

$$\begin{aligned} H(R) &= E\left\|\sum_{i\le n}\varepsilon_i g(X_i)\right\|_{\mathscr{G}'(R)} \\ &\le K\left(\sqrt{nv}\left(4^{-r}+4^{-q}\left(\sqrt{q+\log V}\right)\right)+qv+v\log V\right). \end{aligned} \tag{8.15}$$

We select r such that $\rho_r \sim u\sigma/K_5 t$, where K_5 is universal and will be determined later. Since we assume $M^4 \le \sqrt{n}\,\sigma^4$, we leave the reader to check that, for $M \ge K(V, v, \sigma)$, by taking q as large as possible, we get $H(R) \le K\sqrt{nv}\,\rho_r$ (recall that $\sqrt{n} \ge M$). [This observation also applies to the computation of the last term of (8.9).] Also, we see that taking $M \ge K(V, v, \sigma)$ yields $H(R) \le Kn\rho_r^2$. Thus, provided that

$$K\sqrt{nv}\,\rho_r \le nu \le Kn\rho_r^2, \tag{8.16}$$

we have by Theorem 3.5 that the last term of (8.14) is bounded by

$$(2\cdot 4^{r-p})^{v}\exp\left(-\frac{nu^2}{K\rho_r^2}\right) \le (2\cdot 4^{r-p})^{v}\exp\left(-\frac{2nt^2}{\sigma^2}\right),$$

by a suitable choice of K_5. We observe that

$$4^{r-p} = \frac{\rho}{\rho_r} \le K\frac{t\rho}{\sigma u} \le \frac{Kt}{\sigma}\frac{\sqrt{n}}{\sqrt{v}}.$$

Thus the last term of (8.14) is bounded by

$$\left(\frac{Kt}{\sigma}\sqrt{n}\right)^{v}\exp\left(-\frac{2nt^2}{\sigma^2}\right) \le \exp\left(-\frac{nt^2}{\sigma^2}\right)$$

for $t \ge K\sigma\sqrt{v/n}\,\sqrt{\log(1+v)}$, which we may assume since $M \ge K(V, v, \sigma)$. Finally, by (8.12), we have

$$\exp\left(-\frac{nt^2}{\sigma^2}\right) \le \exp\left(-\frac{nt^2}{2\sigma^2}-\frac{nu^2}{K\rho^2}\right).$$

Observe also that (8.16) reduces to

$$K\sqrt{nv}\,\frac{u\sigma}{t} \le nu \le n\frac{u^2\sigma^2}{Kt^2},$$

that is, $t \ge K\sqrt{v}\,\sigma/\sqrt{n}$, $Kt^2 \le u\sigma^2$. The first condition is automatic since

$t \geq M/2\sqrt{n}$; for the second, it suffices to see that $Kt^2 \leq \sigma^2\rho\sqrt{v/n} = \sigma^4 v/M\sqrt{n}$, that is, $Km^3 \leq \sigma^4\sqrt{n}$, which holds.

We now take $u_s = u4^{-(s-\rho)}/K$, where K is large enough that (8.12) implies $u_p \leq t\rho_{p-1}/\sigma$, so that $u_s \leq t\rho_{s-1}/\sigma$ for $s \geq p$. To check (8.13), it suffices to show that $u_r \geq 2^4 t^2/\sigma^2$. But $u_r = u\rho_r/4\rho$ is of order $u\sigma/K\rho t$ and is bigger than $2^4t^2/\sigma^2$, provided $u\sigma^3 \geq Kt^3\rho$. According to (8.12), it suffices that $Kt^3\sqrt{n} \leq \sigma^3\sqrt{v}$, that is $KM^3 \leq n\sigma^3\sqrt{v}$. But $M^3 \leq \sqrt{n}\,\sigma^4$ and we can assume $n \geq Kv/\sigma^2$.

Now, we have proved that the quantity (8.14) is bounded by

$$K \exp\left(-\frac{nt^2}{2\sigma^2} - \frac{nu^2}{K\rho^2}\right).$$

We then leave the reader to combine this with (8.10) to yield (8.9), and hence (8.8).

What can be done when $M^3 \geq \sqrt{n}\,\sigma^4$? The main obstacle in the above approach using large values of M is in the proof of Lemma 8.1, namely in the inequality

$$\begin{aligned} (8.17) \qquad & P\left(\sum_{i \leq n} (f + \alpha g)(X_i) \geq n(t + \alpha u)\right) \\ & \qquad \leq \inf_{\alpha \geq 0} \exp\bigl(-n\theta(t + \alpha u, \sigma^2 + \alpha^2\rho^2, 1 + 2\alpha)\bigr). \end{aligned}$$

The problem is that when $nt^3 \geq \sigma^4$, the value of $\theta(t, \sigma^2, b)$ depends a lot on b, so that replacing 1 by $1 + 2\alpha$ creates a big decrease of θ. One situation where this difficulty is diminished is when one has a control over $\|g\|_\infty$, as the term $1 + 2\alpha$ can be replaced by $1 + 2\alpha\|g\|_\infty$. This is, for example, the case when one considers hypothesis by (ii) of Theorem 1.3 [and one mimics the proof of Ossiander's theorem "conditionally on $\Sigma_{i \leq n} f(X_i) \geq nt$"] and in particular when one controls the covering numbers of $\mathscr{F}$ for the L^∞ norm.

Another rather fascinating twist is as follows. As pointed out in Hoeffding's paper, the inequality

$$P\left(\sum_{i \leq n} f(X_i) \geq nt\right) \leq \exp\bigl(-n\theta(t, Ef^2, \sup f)\bigr)$$

is rather sharp for functions f that take the value $b = \sup f$ on a set of probability $\sigma^2/(b^2 + \sigma^2)$ (where $\sigma^2 = Ef^2$). Now, if $\alpha^2 Eg^2 \ll Ef^2$, the function $f + \alpha g$ cannot take the value $b = 1 + \alpha \sup g$ on a set of probability $\sigma^2/(b^2 + \sigma^2)$ [where $\sigma^2 = E(f + \alpha g)^2$], nor can it be reasonably close to any such function. Thus one can expect that in such a case (8.17) is not sharp. It can actually be shown (by adapting suitably a lemma of Bennett [7], Lemma 2) that (8.17) can be much improved in that case. The improvments that we developed were apparently optimal. They did allow us to prove (8.8) for the values of M^3 much beyond $\sqrt{n}\,\sigma^3$ [the exponent being of course replaced by $-n\theta(M/\sqrt{n}, \sigma^2, 1)$] although they were not sufficient to get a clean result for all values of M (or even of $M \leq \sigma^2\sqrt{n}$).

Should anyone be really interested in the situation studied in this section for large M, we would like to point out a less accurate, but much simpler method. It is simply to write that, for a class of functions $\mathscr{G}$ and $f \in \mathscr{G}$,

$$(8.18)\quad \begin{aligned} \tau_{\mathscr{G}}(M) &\le P\left(\left|\sum_{i\le n} f(X_i)\right| \ge \sqrt{n}\,(M-w)\right) + P\left(\left\|\sum_{i\le n} g(X_i)\right\|_{\mathscr{G}'} \ge w\sqrt{n}\right) \\ &\le 2\exp\left(-n\theta\left(\frac{M-w}{\sqrt{n}}, \sigma^2, 1\right)\right) + P\left(\left\|\sum_{i\le n} g(X_i)\right\|_{\mathscr{G}'} \ge w\sqrt{n}\right), \end{aligned}$$

where $\mathscr{G}' = \{g - f;\ g \in \mathscr{G}\}$.

One then evaluates the last term through Theorem 3.5 and one optimizes over w. Finally, one breaks a general class $\mathscr{F}$ in pieces to which (8.18) can be applied efficiently. It should be pointed out that this rather straightforward approach can be used in the situations considered in Theorems 1.1 and 1.3, and that the power of M one obtains in front of the exponential is only twice the optimal, a result that already improves considerably on the previous work in this area. For simplicity, we will discuss this approach in the present case only for $M \le \sigma^2\sqrt{n}$. The point of this condition is that it is simple to see that the function $\theta(t) = \theta(t, \sigma^2, 1)$ satisfies $\theta'(t) \le Kt/\sigma^2$ for $t \le \sigma^2$. If we set

$$H = E\left\|\sum \varepsilon_i g(X_i)\right\|_{\mathscr{G}'}, \qquad \rho = \sup_{g\in\mathscr{G}} \left(E(g-f)^2\right)^{1/2}$$

and $S = n\rho^2 + H$, it then follows from Theorem 3.5 and the fact that

$$\theta\left(\frac{M-w}{\sqrt{n}}\right) \ge \theta\left(\frac{M}{\sqrt{n}}\right) - \frac{KwM}{\sigma^2 n}$$

that (8.18) implies

$$\tau_{\mathscr{G}}(M) \le 2\exp\left(-n\theta\left(\frac{M}{\sqrt{n}}\right) + \frac{KwM}{\sigma^2}\right) + \exp\left(-\varphi_{K,S}(w\sqrt{n})\right),$$

provided $w \ge KH/\sqrt{n}$. Consider then the smallest number w_0 such that $\varphi_{K,S}(w_0\sqrt{n}) \ge n\theta(M/\sqrt{n})$. Then

$$\tau_{\mathscr{G}}(M) \le 3\exp\left(-n\theta\left(\frac{M}{\sqrt{n}}\right) + \frac{Kw_0 M}{\sigma^2} + \frac{KHM}{\sqrt{n}\,\sigma^2}\right).$$

If we recall that $n\theta(M/\sqrt{n}) \le KM^2/\sigma^2$, we see from the definition of $\varphi_{K,S}$ that if $S \ge M^2/\sigma^2$, we have $w_0 \le K\sqrt{S/n}\,M/\sigma$, so that

$$\tau_{\mathscr{G}}(M) \le 3\exp\left(-n\theta\left(\frac{M}{\sqrt{n}}\right) + K\sqrt{\frac{S}{n}}\frac{M^2}{\sigma^3} + \frac{KHM}{\sqrt{n}\,\sigma^2}\right).$$

Let us take $\rho = v\sigma^3/M^2$, so that $S \ge M^2/\sigma^2$ whenever $M^6 \le nv^2\sigma^8$ (the reader should observe that this is essentially the case $M^3 \le \sqrt{n}\,\sigma^4/K$ considered before). Thus, in that case, splitting $\mathscr{F}$ into $K(V/\rho)^v$ pieces $\mathscr{G}'$ of diameter $\le \rho$, we get

$$\tau_{\mathscr{G}}(M) \le \left(\frac{KVM^2}{v\sigma^3}\right)^v \exp\left(-n\theta\left(\frac{M}{\sqrt{n}}\right) + K\sqrt{\frac{H}{n}}\,\frac{M^2}{\sigma^2} + \frac{KHM}{\sqrt{n}\,\sigma^2}\right),$$

where H is controlled by (8.15). At least for $M \ge K(V, v, \sigma)$, the last two terms in the exponent will disappear by previous arguments, yielding a reasonable bound, although not as good as the bound obtained previously in (8.8).

If $S \le M^2/\sigma^2$, then the definition of $\varphi_{K,S}$ shows that

$$w_0 \le \frac{KM^2}{\sqrt{n}\,\sigma^2}\log\left(\frac{eM^2}{\sigma^2 S}\right)^{-1/2},$$

yielding

$$\tau_{\mathscr{G}}(M) \le 2\exp\left(-n\theta\left(\frac{M}{\sqrt{n}}\right) + \frac{KM^2}{\sqrt{n}\,\sigma^4}\log\left(\frac{eM^2}{\sigma^2 S}\right)^{-1/2} + \frac{KHM}{\sqrt{n}\,\sigma^2}\right). \tag{8.19}$$

If it were true that $S = n\rho^2$, knowing that $\mathscr{F}$ can be split into $K(V/\rho)^v$ pieces for which (8.18) holds, optimization over ρ would mean taking

$$\rho = \frac{\sqrt{n}\,\sigma}{M}\exp\left(-\frac{M^{2/3}}{n^{1/6}\sigma^{3/4}v^{2/3}}\right).$$

But, on the other hand, we do not know how to do better than (8.15) to control H (and, in particular, we need $n\rho \ge 1$!). This makes optimization of (8.18) unwieldy. A simple choice is, however, $\rho = v\sigma^3/M_0^2$, where $M_0^6 = nv^2\sigma^8$, so that $\rho = n^{-1/3}v^{1/3}\sigma^{5/3}$. This can be shown to work in the usual way, at least for $M \ge K(V, v, \sigma)$. For $M \ge M_0$, this yields

$$\tau_{\mathscr{F}}(M) \le \left(\frac{KVn^{1/3}}{v^{1/3}\sigma^{5/3}}\right)^v \exp\left(-n\theta\left(\frac{M}{\sqrt{n}}\right) + \frac{KM^3}{\sqrt{n}\,\sigma^4}\log\left(\frac{eM^2}{M_0^2}\right)^{-1/2}\right).$$

Acknowledgment. I am indebted to Professor Dudley, who brought these questions to my attention and who usefully commented on the results. Part of this research was done while the author was visiting the M.S.R.I.

REFERENCES

[1] ALEXANDER, K. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067.

[2] DUDLEY, R. (1984). A course on empirical processes. *Ecole d'Été de Probabilités de St. Flour 1982. Lecture Notes in Math.* **1097** 2–142. Springer, New York.

[3] DVORETZKY, A., KIEFER, J. C. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

[4] FERNIQUE, X. (1975). Regularité des trajectoires des fonctions aléatoires gaussiennes. *Ecole d'Été de Probabilités de St. Flour. Lecture Notes in Math.* **480** 1–96. Springer, New York.

[5] GINÉ, E. and ZINN, J. (1984). On the central limit theorem for empirical processes. *Ann. Probab.* **12** 929–989.
[6] HAUSSLER, D. (1991). Sphere packing numbers for subsets of the n-cube with bounded Vapnik Červonenkis dimension. To appear.
[7] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.
[8] KIEFER, J. C. (1961). On large deviations of the empiric D. F. of vector chance variables and a law of the iterated logarithm. *Pacific J. Math.* **2** 649–660.
[9] LEDOUX, M. and TALAGRAND, M. (1992). *Probability in Banach Spaces*. Springer, New York.
[10] MASSART, P. (1986). Rates of convergence in the central limit theorem for empirical processes. *Ann. Inst. H. Poincaré* **22** 381–423.
[11] OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2-bracketing. *Ann. Probab.* **15** 897–1919.
[12] PISIER, G. (1983). Some applications of the metric entropy condition to harmonic analysis. *Banach Spaces, Harmonic Analysis and Probability. Lecture Notes in Math.* **995** 123–154. Springer, New York.
[13] PISIER, G. (1986). Probabilistic methods in the geometry of Banach spaces. *Probability and Analysis. Lecture Notes in Math.* **1206** 167–241. Springer, New York.
[14] ROBBINS, H. (1955). A remark on Stirling's formula. *Amer. Math. Monthly* **62** 26–29.
[15] SAMARODNITSKY, G. (1991). Probability tails of Gaussian extrema. *Stochastic Process. Appl.* **38** 55–84.
[16] SMITH, D. L. and DUDLEY, R. M. (1992). Exponential bounds in Vapnik–Červonenkis classes of index 1. To appear.
[17] TALAGRAND, M. (1984). Sur l'intégrabilité des vecteurs Gaussiens. *Z. Wahrsch. Verw. Gebiete* **68** 1–8.
[18] TALAGRAND, M. (1987). Regularity of Gaussian processes. *Acta. Math.* **159** 99–149.
[19] TALAGRAND, M. (1988). Small tails for the supremum of a Gaussian process. *Ann. Inst. H. Poincaré* **24** 307–315.
[20] TALAGRAND, M. (1989). An isoperimetric theorem on the cube and the Khinchine–Kahane inequalities. *Proc. Amer. Math. Soc.*. **104** 905–909.
[21] TALAGRAND, M. (1989). Isoperimetry and the integrability of sums of independent Banach-space valued random variables. *Ann. Probab.* **17** 1546–1570.
[22] TALAGRAND, M. (1990). Sample boundedness of stochastic processes under increment conditions. *Ann. Probab.* **18** 1–49.
[23] TALAGRAND, M. (1991). A new isoperimetric inequality for product measure, and the tails of sums of independent random variables. *Geom. Funct. Anal.* 211–223.
[24] TALAGRAND, M. (1993). Isoperimetric inequalities and concentration of measure in product spaces. To appear.
[25] VAPNIK, V. N. and ČERVONENKIS, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 264–321.

EQUIPE D'ANALYSE-TOUR 46
U.A. AU CNRS NO. 754
4 PL. JUSSIEU
75230 PARIS CEDEX 05
FRANCE

DEPARTMENT OF MATHEMATICS
OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210-1174