The Annals of Probability
1988, Vol. 16, No. 3, 1281–1290

CHARACTERISTICS OF NORMAL SAMPLES

By Victor Goodman

Indiana University

A "law of large numbers" for the maximum of i.i.d. univariate normal random variables is extended to a general multivariate case. Let $\mathbf{Z}_i$ denote i.i.d. Banach space valued random variables with a centered Gaussian distribution. Let $\mathbf{K}$ denote the unit ball of the reproducing kernel Hilbert space. Then with probability 1, the maximum distance from the sample points $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$ to the set $\sqrt{2 \log n}\ \mathbf{K}$ approaches zero. In addition, the sample forms epsilon nets for this set as n tends to infinity.

0. Introduction. Gnedenko (1943) proved that the maximum ξ_n of the first n random variables in an i.i.d. standard normal sequence satisfies

$$P\left\{ \left| \xi_n - \sqrt{2 \log n} \right| < \varepsilon \right\} \to 1$$

for all $\varepsilon > 0$ as $n \to \infty$. His general result regarding degeneracy of an arbitrary maximum ξ_n was termed the "law of large numbers for the maximum of a random sequence." We formulate a similar law for multivariate samples. For a simple multivariate normal case one takes $\mathbf{Z}_1, \mathbf{Z}_2, \dots$ to be isonormally distributed on R^d. A straightforward calculation shows that the maximum Euclidean length of the first n sample points also obeys Gnedenko's law of large numbers. In addition, there are sufficiently many sample points of nearly this maximum length so that such points form an epsilon net for the expanding sphere

$$\left\{ \mathbf{x} \in R^d \colon |\mathbf{x}| = \sqrt{2 \log n} \right\}.$$

One may say that the law of large numbers appears simultaneously for all directions.

An interesting consequence of this result is that the sample $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$ appears as a surprisingly regular set. With probability 1, the sample is eventually contained in the ball of radius $\varepsilon + \sqrt{2 \log n}$. However, the sample is also eventually dense in the ball of radius $\sqrt{2 \log n}$. As a result, the unnormalized sample data approximates the Euclidean ball of radius $\sqrt{2 \log n}$ quite well. Do other multivariate samples approximate properly scaled deterministic sets in a similar manner? We show that this phenomenon appears in a very general context for normal samples, and we apply it to describe samples of some Gaussian processes.

1. Gaussian measures. Let ν denote a centered Gaussian measure on the Borel σ-algebra of a real separable Banach space B with norm $\| \ \|$. Then for each

Received September 1986; revised March 1987.
AMS 1980 *subject classifications.* Primary 60B11, 60D05, 60G15; secondary 60B12, 60F10, 60F20.
Key words and phrases. Gaussian processes, i.i.d. samples, reproducing kernels, cluster set.

fixed element $\mathbf{y}$ in the topological dual B^*, the random variable

$$\mathbf{x} \to \langle \mathbf{y}, \mathbf{x} \rangle$$

has a mean zero normal distribution. It is well known that the reproducing kernel map S: $B^* \to B$ exists as a compact linear operator and is characterized by the identity

$$\operatorname{Cov}(\langle \mathbf{y}, \mathbf{x} \rangle \langle \mathbf{z}, \mathbf{x} \rangle) = \langle \mathbf{y}, S(\mathbf{z}) \rangle$$

which holds for all $\mathbf{y}, \mathbf{z} \in B^*$. In addition, an inner product $(\ ,\ )_\nu$ may be defined on the range of S by the covariance

$$(S(\mathbf{y}), S(\mathbf{z}))_\nu \equiv \langle \mathbf{y}, S(\mathbf{z}) \rangle.$$

The associated seminorm $|\ |_\nu$ on the range of S is a norm and the completion of the range in this norm is the reproducing kernel Hilbert space $\mathbf{H}_\nu$. The unit ball of $\mathbf{H}_\nu$ will be denoted by $\mathbf{K}$. It is known that $\mathbf{K}$ is a compact subset of B. Details of the preceding may be found in Lemma 2.1 of Goodman, Kuelbs and Zinn (1981).

If $\mathbf{C}$ is a subset of a Banach space and $\varepsilon > 0$, the metric entropy of $\mathbf{C}$ is denoted by $H(\varepsilon, \mathbf{C})$. This quantity is the logarithm of the minimal cardinality of coverings of the set $\mathbf{C}$ by sets of diameter not exceeding 2ε. A formal definition may be found in Section 1 of Dudley (1973). A Banach ball with radius r, centered at $\mathbf{x} \in B$, will be denoted by $B_r(\mathbf{x})$.

2. Normal samples.

THEOREM 2.1. *Suppose that* $\mathbf{Z}_1, \mathbf{Z}_2, \dots$ *are i.i.d. Banach space valued random vectors with a centered Gaussian distribution* ν *on a real separable Banach space* B. *If* $\mathbf{K}$ *denotes the unit ball of the reproducing kernel Hilbert space for* ν, *then with probability* 1,

$$\max_{i \le n} d\big(\mathbf{Z}_i, \sqrt{2 \log n}\, \mathbf{K}\big) \to 0, \tag{2.1}$$

$$\max_{\mathbf{y} \in \sqrt{2 \log n}\, \mathbf{K}} d\big(\mathbf{y}, \{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n\}\big) \to 0 \tag{2.2}$$

as $n \to \infty$. *Here* $d(\ ,\ )$ *denotes the Banach norm distance from a point to a set.*

EXAMPLE 2.1. Consider a d-dimensional standard Brownian motion process $\mathbf{B}_t$, $0 \le t \le 1$. Then $\mathbf{B}_t$ has a Gaussian distribution on the Banach space consisting of all continuous R^d valued functions with the supremum norm, and it is well known that

$$\mathbf{K} = \left\{ \int_0^t \mathbf{f}(s)\, ds \colon \mathbf{f} \in L^2([0,1], R^d) \text{ and } \|\mathbf{f}\|_2 \le 1 \right\}$$

[see Strassen (1964)]. Strassen observed that for all $\mathbf{k} \in \mathbf{K}$,

$$|\mathbf{k}(t)| \le \sqrt{t} \quad \text{and} \quad |\mathbf{k}(t)| = o(\sqrt{t}) \quad \text{as } t \downarrow 0.$$

Theorem 2.1 implies the following result for i.i.d. d-dimensional Brownian motions.

Almost surely, the initial n sample paths of every sample sequence are uniformly close to the family

$$\mathbf{S}_n \equiv \left\{\mathbf{g}\colon |\mathbf{g}(t)| \le \sqrt{2t\log n},\, 0 \le t \le 1\right\}.$$

This strengthens the main result of LePage and Schreiber (1985) where the preceding approximation was obtained with a uniform error of $o(\sqrt{\log n})$. In addition, one may show that if $|\mathbf{y}| \le \sqrt{t}$, there is a function in $\mathbf{K}$ whose graph contains the point $(t, \mathbf{y})$. It follows from Theorem 2.1 that the union of graphs for the first n sample paths is asymptotically dense in the union of graphs of functions in $\mathbf{S}_n$.

Therefore, we may paraphrase the summary in LePage and Schreiber (1985). Eventually, an *unnormalized* plot of n independent R^d valued Brownian paths over $[0,1]$ is almost certain to have the appearance of a shaded region with the boundary $|\mathbf{y}| = \sqrt{2t\log n}$.

EXAMPLE 2.2. Theorem 2.1 should be compared with Theorem 9 of de Acosta and Kuelbs (1983), where the special case $a_n = 1$ applies to the normalized random sample

$$\left\{1/\sqrt{2\log n}\,\mathbf{Z}_i\colon i \le n\right\}.$$

Theorem 9 implies that the maximum distance from this sample to the set $\mathbf{K}$ converges to zero and that the sample is asymptotically dense for $\mathbf{K}$. Theorem 2.1 may be viewed as an improvement in the rate for convergence of this set to $\mathbf{K}$ and the rate for its denseness from $o(1)$ to $o(1/\sqrt{\log n})$.

3. Proving Theorem 2.1. The theorem is a consequence of three lemmas for a centered Gaussian measure ν.

LEMMA 3.1 [Talagrand (1984)]. *For each $\varepsilon > 0$, there is a random variable ψ_ε such that*

$$E\left[\exp\left(\tfrac{1}{2}\psi_\varepsilon\right)\right] < \infty \quad \textit{and for all } \lambda > 0, \quad \nu(\lambda\mathbf{K} + \mathbf{B}_\varepsilon(\mathbf{0})) = P\left\{\psi_\varepsilon < \lambda^2\right\}.$$

LEMMA 3.2. *For any nonnull element $\mathbf{h} \in \mathbf{H}_\nu$ and any $\varepsilon > 0$,*

$$\nu(\mathbf{B}_\varepsilon(\mathbf{h})) \ge \nu\left(\mathbf{B}_{\varepsilon/2}(\mathbf{0})\right)\left\{\Phi(|\mathbf{h}|_\nu[1 + \varepsilon/2\|\mathbf{h}\|]) - \Phi(|\mathbf{h}|_\nu[1 - \varepsilon/2\|\mathbf{h}\|])\right\}.$$

Here Φ denotes the standard normal distribution function.

LEMMA 3.3. *Let $H(\varepsilon, \mathbf{K})$ denote the metric entropy of* $\mathbf{K}$. *Then*

$$\lim_{\varepsilon\downarrow 0} \varepsilon^2 H(\varepsilon, \mathbf{K}) = 0.$$

PROOF OF THEOREM 2.1. We apply Lemma 3.1, setting $\lambda = \sqrt{2\log n}$. Then

$$\begin{aligned}\nu\left(\sqrt{2\log n}\,\mathbf{K} + \mathbf{B}_\varepsilon(\mathbf{0})\right) &= P\{\psi_\varepsilon < 2\log n\}\\ &= P\left\{\exp\left(\tfrac{1}{2}\psi_\varepsilon\right) < n\right\}.\end{aligned}$$

Since each $\mathbf{Z}_n$ has distribution equal to ν, the preceding equality gives

$$\sum_n P\{d(\mathbf{Z}_n, \sqrt{2\log n}\ \mathbf{K}) > \varepsilon\} \leq \sum_n P\{\exp(\tfrac{1}{2}\psi_\varepsilon) \geq n\} < \infty.$$

Then the Borel–Cantelli lemma implies that for each $\varepsilon > 0$, the event

$$\{d(\mathbf{Z}_n, \sqrt{2\log n}\ \mathbf{K}) > \varepsilon, \text{i.o.}\}$$

has zero probability. This proves (2.1) of Theorem 2.1.

Next, apply Lemma 3.2 for the choice $\mathbf{h} = \sqrt{2\log n}\ \mathbf{k}$, where $\mathbf{k} \in \mathbf{H}_\nu$ satisfies $|\mathbf{k}| \leq 1$. As shown in Section 2 of Goodman, Kuelbs and Zinn (1981), there is a constant c such that $\|\mathbf{k}\| \leq c|\mathbf{k}|_\nu$ for any $\mathbf{k} \in H_\nu$. Hence, $\|\mathbf{k}\| \leq c$ and Lemma 3.2 gives

$$\nu(\mathbf{B}_\varepsilon(\mathbf{h})) \geq d_\varepsilon\left[\Phi\left(\sqrt{2\log n}\ |\mathbf{k}|_\nu + \varepsilon/2c\right) - \Phi\left(\sqrt{2\log n}\ |\mathbf{k}|_\nu - \varepsilon/2c\right)\right].$$

Here d_ε denotes the quantity $\nu(\mathbf{B}_{\varepsilon/2}(\mathbf{0}))$. Since the preceding difference of normal distribution functions is an integral of the form

$$c'\int_{a-\delta}^{a+\delta}\exp(-x^2/2)\,dx, \tag{3.1}$$

the estimate is minimal for $|\mathbf{k}|_\nu = 1$. A standard estimation for this case is made by multiplying the integrand by $x/(a+\delta)$. The resulting lower bound for (3.1) is

$$c'/(a+\delta)\exp\left(-\tfrac{1}{2}(a^2+\delta^2)\right)[\exp(a\delta) - \exp(-a\delta)],$$

where $a = \sqrt{2\log n}$ and $\delta = \varepsilon/2c$. We then obtain

$$\nu(\mathbf{B}_\varepsilon(\mathbf{h})) \geq c''/n\left(\sqrt{2\log n} + \delta\right)\left[\exp\left(\delta\sqrt{2\log n}\right) - \exp\left(-\delta\sqrt{2\log n}\right)\right]. \tag{3.2}$$

Lemma 3.3 allows us to estimate $H(\varepsilon, \sqrt{2\log n}\ \mathbf{K})$, the logarithm of the minimal cardinality of coverings of the set $\sqrt{2\log n}\ \mathbf{K}$ by balls of radius ε. A simple scaling argument shows that

$$H(\varepsilon, a\mathbf{K}) = H(\varepsilon/a, \mathbf{K}).$$

Since $\varepsilon/a \to 0$ as $n \to \infty$, Lemma 3.3 implies that

$$H(\varepsilon, a\mathbf{K}) = o(a^2/\varepsilon^2) = o(\log n).$$

Next, we consider the probability that a sample of size n misses some ε neighborhood of a point in the set $a\mathbf{K}$:

$$\begin{aligned}
&\log P\{\mathbf{Z}_i \notin \mathbf{B}_\varepsilon(\mathbf{h}) \text{ for } 1 \leq i \leq n\}\\
&\quad = n\log(1 - \nu(\mathbf{B}_\varepsilon(\mathbf{h})))\\
&\quad \leq -c''/\left(\sqrt{2\log n} + \delta\right)\left[\exp\left(\delta\sqrt{2\log n}\right) - \exp\left(-\delta\sqrt{2\log n}\right)\right]\\
&\quad \approx -c''\left(\sqrt{2\log n} + \delta\right)\exp\left(\delta\sqrt{2\log n}\right),
\end{aligned}$$

using the estimate (3.2). Since we also have an estimate for the minimal number of balls $\mathbf{B}_\varepsilon(\mathbf{h})$ needed to cover $a\mathbf{K}$, we obtain a bound for the probability that the sample misses at least one ball in a minimal covering of $a\mathbf{K}$. It follows that

$$\log P\Big\{\max_{\mathbf{y}\in\sqrt{2\log n}\,\mathbf{K}} d(\mathbf{y},\{\mathbf{Z}_1,\mathbf{Z}_2,\dots,\mathbf{Z}_n\}) > 2\varepsilon\Big\} \tag{3.3}$$
$$\le o(\log n) - 1/\big(\sqrt{2\log n} + \delta\big)\exp\big(\delta\sqrt{2\log n}\,\big),$$

which shows that the series

$$\sum_n P\Big\{\max_{\mathbf{y}\in\sqrt{2\log n}\,\mathbf{K}} d(\mathbf{y},\{\mathbf{Z}_1,\mathbf{Z}_2,\dots,\mathbf{Z}_n\}) > 2\varepsilon\Big\}$$

is summable. We apply the Borel–Cantelli lemma and since ε is arbitrary, condition (2.2) of the theorem is proved. □

4. Proofs of Lemmas 3.1–3.3. Lemma 3 in Talagrand (1984) contains an inequality which is essentially Lemma 3.1 here. We present an alternate proof which contains more information concerning the exponential moment of ψ_ε.

DEFINITION 4.1. Let $|\mathbf{x}|$ denote the function on the Banach space given by

$$|\mathbf{x}| = \begin{cases} |\mathbf{x}|_\nu, & \text{if } \mathbf{x}\in H_\nu,\\ +\infty, & \text{otherwise.}\end{cases}$$

For fixed $\varepsilon > 0$, let $\psi_\varepsilon(\mathbf{y})$ denote the function on the Banach space given by

$$\psi_\varepsilon(\mathbf{y}) = \inf_{\mathbf{x}\in\mathbf{B}_\varepsilon(\mathbf{y})} |\mathbf{x}|^2.$$

Note that by the definition of ψ_ε, for any $\lambda > 0$, the set $\lambda\mathbf{K} + \mathbf{B}_\varepsilon(\mathbf{0})$ is equal to the set $\{\psi_\varepsilon < \lambda^2\}$.

For any bounded operator T: $B \to B$, let

$$\psi_{T,\varepsilon}(\mathbf{y}) = \inf_{x\in\{y+z:\ \|\mathbf{Tz}\|<\varepsilon\}} |\mathbf{x}|^2.$$

Then by definition, $\psi_\varepsilon = \psi_{I,\varepsilon}$.

PROPOSITION 4.1. *Let P: $B\to B$ be a continuous projection with finite-dimensional range $\mathscr{R}(P)\subset H_\nu$ and suppose that P is self-adjoint on H_ν. Then for all $\mathbf{y}\in B$,*

$$\psi_{P,\varepsilon}(\mathbf{y}) = \inf_{\mathbf{x}\in\mathbf{B}_\varepsilon(P\mathbf{y})\cap\mathscr{R}(P)} |\mathbf{x}|^2.$$

In addition, a similar identity holds for P replaced by the projection, $Q \equiv I - P$.

PROOF. By definition,

$$\psi_{P,\varepsilon}(\mathbf{y}) = \inf_{\mathbf{x}\in\{\mathbf{y}+\mathbf{z}:\ \|P\mathbf{z}\|<\varepsilon\}} |\mathbf{x}|^2$$
$$= \inf_{\{\mathbf{x}:\ \|P(\mathbf{x}-\mathbf{y})\|<\varepsilon\}} |\mathbf{x}|^2.$$

Then if $\mathbf{x}\in\mathscr{R}(P)\cap\mathbf{B}_\varepsilon(P\mathbf{y})$, $\mathbf{x}$ satisfies the condition $\|P(\mathbf{x}-\mathbf{y})\| < \varepsilon$. Hence,

$$\psi_{P,\varepsilon}(\mathbf{y}) \le \inf_{\mathbf{x}\in\mathbf{B}_\varepsilon(P\mathbf{y})\cap\mathscr{R}(P)} |\mathbf{x}|^2.$$

On the other hand, by the orthogonality of P and Q,

$$\begin{aligned}\psi_{P,\varepsilon}(\mathbf{y}) &= \inf_{\{\mathbf{x}:\, \|P(\mathbf{x}-\mathbf{y})\|<\varepsilon\}} |P\mathbf{x} + Q\mathbf{x}|^2 \\ &= \inf_{\{\mathbf{x}:\, \|P(\mathbf{x}-\mathbf{y})\|<\varepsilon\}} \left(|P\mathbf{x}|^2 + |Q\mathbf{x}|^2\right) \\ &\geq \inf_{\{\mathbf{x}:\, \|P(\mathbf{x}-\mathbf{y})\|<\varepsilon\}} |P\mathbf{x}|^2 \\ &= \inf_{\{\mathbf{x}\in\mathscr{R}(P):\, \|P(\mathbf{x}-\mathbf{y})\|<\varepsilon\}} |\mathbf{x}|^2.\end{aligned}$$

This shows the reverse inequality for the two expressions. Similar arguments prove the equality for the case of P replaced by Q. □

COROLLARY 4.1. *If P: $B \to B$ is an operator as in the statement of Proposition* 4.1, *then for all* $\mathbf{y} \in B$

$$\psi_{P,\varepsilon}(\mathbf{y}) = \psi_{P,\varepsilon}(P\mathbf{y}).$$

Moreover, the random functions $\psi_{P,\varepsilon}(\mathbf{y})$ and $\psi_{Q,\varepsilon}(\mathbf{y})$ are independent with respect to the probability measure ν.

PROOF. The preceding equality is immediate from Proposition 4.1 and therefore $\psi_{P,\varepsilon}(\mathbf{y})$ is a function of $P\mathbf{y}$. Similarly, from Proposition 4.1, one concludes that $\psi_{Q,\varepsilon}(\mathbf{y})$ is a function of $Q\mathbf{y}$, but the random vectors $P\mathbf{y}$ and $Q\mathbf{y}$ are stochastically independent. □

COROLLARY 4.2. *If P: $B \to B$ is an operator as in the statement of Proposition* 4.1, *then for any* $\varepsilon > 0$,

$$E\left[\exp\left(\tfrac{1}{2}\psi_{P,\varepsilon}\right)\right] < \infty.$$

PROOF. Let ν' denote the projection of the measure ν on the finite-dimensional space $\mathscr{R}(P)$ under P. Let ψ'_ε denote the corresponding function of Definition 4.1 on the space $\mathscr{R}(P)$ where the infimum is again taken with respect to the Banach space norm. It follows from Corollary 4.1 and Proposition 4.1 that

$$\psi_{P,\varepsilon}(\mathbf{y}) = \psi'_\varepsilon(P\mathbf{y}).$$

It suffices to show that

$$E'\left[\exp\left(\tfrac{1}{2}\psi'_\varepsilon\right)\right] < \infty.$$

But, since the norms $|\ |$ and $\|\ \|$ are equivalent on $\mathscr{R}(P)$, there is a $\delta > 0$ such that for all $\mathbf{y} \in \mathscr{R}(P)$,

$$\psi'_\varepsilon(\mathbf{y}) \leq (|\mathbf{y}| - \delta)^2.$$

Furthermore, since the space $\mathscr{R}(P)$ is finite dimensional, it is easy to see that the expectation

$$E'\left[\exp\left(\tfrac{1}{2}(|\mathbf{y}| - \delta)^2\right)\right],$$

taken with respect to ν', is finite. □

PROPOSITION 4.2. *Let P: $B \to B$ be a continuous projection with finite-dimensional range $\mathscr{R}(P) \subset H_\nu$ and suppose that P is self-adjoint on H_ν. Then for fixed $\varepsilon > 0$, $s > 0$ and $t > 0$ such that $\varepsilon = s + t$,*

$$\psi_\varepsilon(\mathbf{y}) \le \psi_{P,s}(\mathbf{y}) + \psi_{Q,t}(\mathbf{y})$$

for all $\mathbf{y} \in B$. Here, Q denotes the projection $I - P$.

PROOF. Suppose that $\mathbf{p} \in \mathscr{R}(P)$ satisfies

$$\|\mathbf{p} - P\mathbf{y}\| < s$$

and that $\mathbf{q} \in \mathscr{R}(Q)$ satisfies

$$\|\mathbf{q} - Q\mathbf{y}\| < t.$$

Then

$$\|\mathbf{p} + \mathbf{q} - \mathbf{y}\| < \varepsilon$$

and by the definition of $\psi_\varepsilon(\mathbf{y})$,

$$\psi_\varepsilon(\mathbf{y}) \le |\mathbf{p} + \mathbf{q}|^2$$
$$= |\mathbf{p}|^2 + |\mathbf{q}|^2.$$

However, by Proposition 4.1, the infimum of the right-hand expression over such $\mathbf{p}$ is equal to $\psi_{P,s}(\mathbf{y}) + |\mathbf{q}|^2$. Again, by Proposition 4.1, the infimum over such $\mathbf{q}$ gives the desired inequality. □

PROOF OF LEMMA 3.1. Given $\varepsilon > 0$, let $s = t = \varepsilon/2$. Then for any projection P: $B \to B$ as in Proposition 4.1, Proposition 4.2 and Corollary 4.1 imply that the inequality

$$E\left[\exp\left(\tfrac{1}{2}\psi_\varepsilon\right)\right] \le E\left[\exp\left(\tfrac{1}{2}\psi_{P,\varepsilon}\right)\right] E\left[\exp\left(\tfrac{1}{2}\psi_{Q,\varepsilon}\right)\right] \tag{4.1}$$

holds. Furthermore, the first expectation on the right-hand side of (4.1) is finite by Corollary 4.2 for any choice of P. Let P be chosen so that $Q = I - P$ satisfies

$$\nu(S_Q) > \tfrac{1}{2},$$

where

$$S_Q \equiv \{\mathbf{x}: \|Q\mathbf{x}\| < t\}.$$

This is possible by Theorem 4.6 of Dudley (1967). We apply the main result of Borell (1975), which gives the bound

$$\nu(\lambda \mathbf{K} + S_Q) \ge \Phi(\lambda + \alpha). \tag{4.2}$$

Here, Φ denotes the standard normal distribution function and α is defined by the equation

$$\Phi(\alpha) = \nu(S_Q). \tag{4.3}$$

By the choice of Q, $\alpha > 0$. Now, by the definition of $\psi_{Q,t}(\mathbf{y})$,

$$\lambda\mathbf{K} + S_Q = \left\{\mathbf{y}: \psi_{Q,t}(\mathbf{y}) < \lambda^2\right\}.$$

Therefore, $\psi_{Q,t}$ is dominated in tail distribution by $(Z - \alpha)^2$ where Z denotes a standard normal random variable. This shows that the latter expectation in (4.1) is finite and proves Lemma 3.1. □

REMARK. One may optimize the bound for the exponential moment of $\frac{1}{2}\psi_\varepsilon$ by taking the infimum of the right-hand expression in (4.1) over all finite-dimensional projections P as in Proposition 4.1 with the constraints

$$\nu\{\mathbf{x}: \|(I-P)\mathbf{x}\| \le t\} > \tfrac{1}{2}$$

and

$$s + t = \varepsilon.$$

PROOF OF LEMMA 3.2. Suppose $h \in H_\nu \cap \mathscr{R}(S)$. Let $h = ak$ where $|k|_\nu = 1$. The Cameron–Martin formula gives

$$\nu(\mathbf{B}_\varepsilon(\mathbf{h})) = \int_{\mathbf{B}_\varepsilon(\mathbf{0})} \exp\big(a\langle \mathbf{k},\mathbf{x}\rangle - \tfrac{1}{2}a^2\big)\, d\nu(\mathbf{x}).$$

Let P be any projection as in the statement of Proposition 4.1, satisfying $\mathbf{h} \in \mathscr{R}(P)$. Then

$$\{\mathbf{x}: \|P\mathbf{x}\| < \varepsilon/2\} \cap \{\mathbf{x}: \|Q\mathbf{x}\| < \varepsilon/2\} \subset \mathbf{B}_\varepsilon(\mathbf{0}),$$

and we obtain the inequality

$$\begin{aligned}\nu(\mathbf{B}_\varepsilon(\mathbf{h})) &\ge \int_{\{\mathbf{x}:\, \|P\mathbf{x}\| < \varepsilon/2\} \cap \{\mathbf{x}:\, \|Q\mathbf{x}\| < \varepsilon/2\}} \exp\big(a\langle \mathbf{k},\mathbf{x}\rangle - \tfrac{1}{2}a^2\big)\, d\nu(\mathbf{x}) \\ &= \nu(\{\mathbf{x}: \|Q\mathbf{x}\| < \varepsilon/2\}) \int_{\{\mathbf{x}:\, \|P\mathbf{x}\| < \varepsilon/2\}} \exp\big(a\langle \mathbf{k},\mathbf{x}\rangle - \tfrac{1}{2}a^2\big)\, d\nu(\mathbf{x}).\end{aligned} \tag{4.4}$$

The last equality follows from the independence of $P\mathbf{x}$ and $Q\mathbf{x}$. Now take P to be equal to the one-dimensional projection $\langle \mathbf{k},\mathbf{x}\rangle\mathbf{k}$. The region of integration in (4.4) becomes

$$\{\mathbf{x}: |\langle \mathbf{k},\mathbf{x}\rangle| < \varepsilon/2\|\mathbf{k}\|\}.$$

Finally, since $\langle \mathbf{k},\mathbf{x}\rangle$ is a standard normal random variable and the integral in (4.4) is the expression for

$$\nu\{a - \varepsilon/2\|\mathbf{k}\| < \langle \mathbf{k},\mathbf{x}\rangle < a + \varepsilon/2\|\mathbf{k}\|\},$$

we obtain the difference of the standard normal distribution functions in the estimate (4.4). In addition, the inequality

$$\nu(\{\mathbf{x}: \|Q\mathbf{x}\| < \varepsilon/2\}) \ge \nu(\{\mathbf{x}: \|\mathbf{x}\| < \varepsilon/2\})$$

follows from Theorem 5 of Gross (1962) since the operator norm of the projection Q on H_ν is equal to 1. This gives the inequality of Lemma 3.2. □

PROOF OF LEMMA 3.3. For fixed $\varepsilon > 0$, let $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ be any subset of $\lambda\mathbf{K}$ such that the sets

$$\mathbf{B}_\varepsilon(\mathbf{w}_1), \mathbf{B}_\varepsilon(\mathbf{w}_2), \dots, \mathbf{B}_\varepsilon(\mathbf{w}_m)$$

are disjoint.

Here, $\lambda > 0$ is arbitrary and $\mathbf{K}$ denotes the unit ball of the reproducing kernel Hilbert space for ν. An inequality in the derivation of (3.2) gives

$$\nu(\mathbf{B}_\varepsilon(\mathbf{w}_i)) \geq d_\varepsilon[\Phi(\lambda + \varepsilon/2c) - \Phi(\lambda - \varepsilon/2c)],$$

where c and d_ε do not depend on λ.

Since these sets are disjoint, the sum of the probabilities does not exceed 1 and we obtain an upper bound on m,

$$m \leq \{d_\varepsilon[\Phi(\lambda + \varepsilon/2c) - \Phi(\lambda - \varepsilon/2c)]\}^{-1}.$$

Let m_λ denote the cardinality of the largest subset of $\lambda\mathbf{K}$ which is 2ε-discrete. The preceding inequality shows that

$$\lim_{\lambda \to \infty} \lambda^{-2} \log(m_\lambda) \leq \tfrac{1}{2}.$$

The quantities m_λ provide bounds for the metric entropy of the set $\mathbf{K}$. If the points $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ form a maximal 2ε-discrete subset, then any other point $\mathbf{k} \in \lambda\mathbf{K}$ has the property that the set $\mathbf{B}_\varepsilon(\mathbf{k})$ intersects at least one of the sets $\mathbf{B}_\varepsilon(\mathbf{w}_i)$. This shows that the sets $\mathbf{B}_{2\varepsilon}(\mathbf{w}_i)$ form a covering for $\lambda\mathbf{K}$. We obtain the bound

$$\limsup_{\lambda \to \infty} \lambda^{-2} H(2\varepsilon, \lambda\mathbf{K}) \leq \tfrac{1}{2}.$$

Now let $\varepsilon = \frac{1}{2}$. It follows from the identity $H(1, \lambda\mathbf{K}) = H(1/\lambda, \mathbf{K})$ that

$$\limsup_{\varepsilon \downarrow 0} \varepsilon^2 H(\varepsilon, \mathbf{K}) \leq \tfrac{1}{2}.$$

This inequality is valid for the unit ball of the reproducing kernel Hilbert space of an arbitrary Gaussian measure ν. However, one may consider the scaled measure $\nu(\lambda^{-1})$ whose unit ball is the set $\lambda\mathbf{K}$. It follows from the preceding inequality that

$$\limsup_{\varepsilon \downarrow 0} \varepsilon^2 H(\varepsilon, \lambda\mathbf{K}) \leq \tfrac{1}{2}.$$

The change of variables $\varepsilon = \varepsilon\lambda$ in the preceding limit gives

$$\limsup_{\varepsilon \downarrow 0} \varepsilon^2 H(\varepsilon, \mathbf{K}) \leq 1/2\lambda^2.$$

Since λ is arbitrary, the lemma follows. □

Acknowledgments. The author wishes to thank the Editor for pointing out an error in the original manuscript and M. Talagrand for supplying a crucial argument in the proof of Lemma 3.3.

REFERENCES

Borell, C. (1975). The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* **30** 207–216.

de Acosta, A. and Kuelbs, J. (1983). Limit theorems for moving averages. *Z. Wahrsch. verw. Gebiete* **64** 67–123.

Dudley, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.* **1** 290–330.

DUDLEY, R. M. (1973). Sample functions of the Gaussian process. *Ann. Probab.* **1** 66–103.

GNEDENKO, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math.* **44** 423–453.

GOODMAN, V., KUELBS, J. and ZINN, J. (1981). Some results on the LIL in Banach space with applications to weighted empirical processes. *Ann. Probab.* **9** 713–752.

GROSS, L. (1962). Measurable functions on Hilbert space. *Trans. Amer. Math. Soc.* **105** 372–390.

LEPAGE, R. and SCHREIBER, B. (1985). An iterated logarithm law for families of Brownian paths. *Z. Wahrsch. verw. Gebiete* **70** 341–344.

STRASSEN, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. verw. Gebiete* **3** 211–226.

TALAGRAND, M. (1984). Sur l'intégrabilité des vecteurs gaussiens. *Z. Wahrsch. verw. Gebiete* **68** 1–8.

DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47405