The Annals of Probability
1987, Vol. 15, No. 4, 1339–1351

TAIL BEHAVIOUR FOR THE SUPREMA OF GAUSSIAN PROCESSES WITH APPLICATIONS TO EMPIRICAL PROCESSES

BY ROBERT J. ADLER[1] AND GENNADY SAMORODNITSKY[2]

Technion-Israel Institute of Technology and University of Washington, and Technion-Israel Institute of Technology

Initially we consider "the" standard isonormal linear process L on a Hilbert space H, and applying metric entropy methods obtain bounds for the probability that $\sup_C Lx > \lambda$, $C \subset H$ and λ large. Under the assumption that the entropy function of C grows polynomially, we find bounds of the form $c\lambda^\alpha \exp(-\frac{1}{2}\lambda^2/\sigma^2)$, where σ^2 is the maximal variance of L. We use a notion of entropy finer than that usually employed and specifically suited to the nonstationary situation. As a result we obtain, in the nonstationary setting, more precise bounds than any in the literature.

We then treat a number of examples in which the power α is identified. These include the distributions of the maxima of the rectangle indexed, pinned Brownian sheet on $\mathbb{R}^k$ for which $\alpha = 2(2k-1)$, and the half plane indexed pinned sheet on $\mathbb{R}^2$ for which $\alpha = 2$.

1. Introduction. Our motivation comes from the theory of empirical processes, where $X_1, \ldots, X_n$ represent i.i.d. observations from some k-dimensional distribution, and our aim is to test the hypothesis that the parent distribution is given by a measure ν: $\nu(A) = P\{X_i \in A\}$ on the unit cube. A natural test procedure is to form the empirical measure ν_n: $\nu_n(A) = (1/n)\sum_{i=1}^n I_A(X_i)$ (I_A is the indicator function of A) and compare ν_n to ν via a Kolmogorov–Smirnov type statistic of the form

$$\sup_{\mathscr{A}} \left\{ \sqrt{n}\,|\nu_n(A) - \nu(A)| \right\}, \tag{1.1}$$

for some family $\mathscr{A}$ of Borel subsets of $[0,1]^k$. It is well known [Dudley (1978, 1984)] that $\sqrt{n}\,(\nu_n - \nu)$ converges weakly to a Gaussian process on $\mathscr{A}$, under conditions related to the size of $\mathscr{A}$. Consequently, the study of (1.1) reduces, in the limit, to the study of the supremum of a particular Gaussian process over a class of sets.

Unfortunately, however, it is also well known that it borders on the impossible to obtain the exact distributions of Gaussian maxima and that, except for certain special cases, the only results available relate to the tail of the distribution. In

Received December 1984; revised January 1987.

[1]Research supported in part by AFOSR Contracts 84-0104, 85-0384 and F49620-85-C-0114 while visiting the Center for Stochastic Processes, Chapel Hill, North Carolina.

[2]Research supported in part by the Wolf Foundation.

AMS 1980 *subject classifications*. Primary 60G15, 60G57; secondary 60F10, 62G30.

Key words and phrases. Gaussian processes, isonormal process, supremum, metric entropy, Brownian sheet, empirical processes.

this paper we shall be concerned with proving results of the form

$$(1.2)\qquad P\Big\{\sup_{x\in S} X(x) > \lambda\Big\} \le c\lambda^{\alpha}\exp\big(-\tfrac{1}{2}\lambda^2/\sigma^2\big), \quad \lambda \text{ large},$$

where S is a parameter set for a zero mean Gaussian process X, c is a (generally unknown) constant, $\sigma^2 = \sup_{x\in S} E\{X^2(x)\}$ and α is a parameter depending on certain relationships between the size of S and the covariance structure of X. We shall be interested in identifying the smallest possible α. Bounds on the sup in (1.1) (for large n) then follow from (1.2) by taking $S = \mathscr{A}$ and X = Brownian sheet (defined below).

In order to handle as many choices of S as possible, we shall concentrate initially on the most general of Gaussian processes, the so-called *isonormal* process. This process, together with the requisite notion of entropy, is defined in the following section. There we also formulate and prove tail bounds akin to (1.2), and in Section 3 we apply the general results to specific problems related to set-indexed Brownian sheets.

2. The isonormal process and a Fernique inequality. The central idea is to study one, canonical, Gaussian process, and then relate any particular process to this one. It is defined as a linear map L from a real, infinite-dimensional Hilbert space H into real Gaussian variables with $EL(x) = 0$ and $EL(x)L(y) = (x, y)$ for all $x, y \in H$ and is called the *isonormal* Gaussian process on H [cf. Segal (1954) and Dudley (1967, 1973)].

Since Gaussian distributions are uniquely determined by their means and covariances, the isonormal process L can be regarded as the only real Gaussian process for, if $\{x_t, t \in T\}$ is any real Gaussian process with mean $Ex_t = m_t$, then $L(x_t - m_t) + m_t$ is another version of the process, where we take $L^2(\Omega, P)$ for H. On H, L "remembers" the covariance structure of x_t, and, by its linearity, also keeps track of all joint distributions. Thus, we can in general neglect the specific joint distributions of x_t on (Ω, P) and work only with the abstract geometric structure of the function $t \to x_t - m_t \in H$. To see precisely how this works in practice, see the examples in Section 3.

In order to study the structure of H, we shall require the notion of *metric entropy*. Let $\|\cdot\|$ be the induced norm on H, and for $\varepsilon > 0$ let $N(C, \varepsilon) \equiv N_C(\varepsilon)$ be the minimal number of points $x_1, \dots, x_n$ from C such that, for all $y \in C$, $\min_i\{\|x_i - y\|\} \le \varepsilon$. We assume N finite for each $\varepsilon > 0$. Consequently, there exist sets $A_1, \dots, A_{N(C,\varepsilon)}$ covering C such that, for all n, $\|x - y\| \le 2\varepsilon$ for all $x, y \in A_n$. Set $H_C(\varepsilon) = \log N_C(\varepsilon)$. Then $H_C(\varepsilon)$ is the *metric entropy* function of C. Metric entropy is known to play an important role in continuity problems for Gaussian processes and can also be used to study suprema problems. In particular, let us now and hereafter assume that $H_C(\varepsilon)$ exhibits at most a logarithmic growth, or, more conveniently, that there exist positive constants a and κ such that

$$(2.1)\qquad N_C(\varepsilon) \le a\varepsilon^{-\kappa},$$

for small enough ε. Then for large enough $p \geq 2$ and all $\lambda > (1 + 4\kappa \ln p)^{1/2}$,

$$P\Big\{ \sup_{x \in C} |Lx| > \lambda\big(\sigma + 2p^{-2}\big)\Big\} \leq \tfrac{5}{2}(\pi/2)^{1/2} a p^{2\kappa} \psi(\lambda), \tag{2.2}$$

where

$$\sigma := \sup_{x \in C} \|x\| = \sup_{x \in C} \big[E\big(L^2(x)\big)\big]^{1/2}, \qquad \psi(\lambda) = \sqrt{2/\pi} \int_\lambda^\infty e^{-u^2/2}\, du. \tag{2.3}$$

[Inequality (2.2), when specialized to Gaussian processes defined on a Euclidean space, is due to Fernique (1975). The proof for the isonormal process requires only minimal, essentially notational, changes. Details can be found in Adler and Samorodnitsky (1985). The case $\|x\| \equiv \sigma$ for all $x \in C$ is treated in Weber (1980).]

Whereas (2.2) provides a route to essentially the best bounds for $P\{\sup_{x \in C} |Lx| > \lambda\}$ when L has constant variance this is not necessarily the case if $\|x\|$ is nonconstant on C (c.f. comments at the end of this section), and it will be the case of nonconstant variance that will interest us in the following section. To obtain sharper results in the nonstationary situation, we need to impose an additional restriction on C and shall assume that it possesses some sort of scaling property. In particular, for each $\theta > 0$ let $\mathscr{G}_\theta$ be a partition of C satisfying

$$\sup_{x, y \in A} \|x - y\| \leq \theta, \quad \text{for all } A \in \mathscr{G}_\theta. \tag{2.4}$$

Define $N_C^{\mathscr{G}}(\theta) := \#\mathscr{G}_\theta$. Clearly, $N_C^{\mathscr{G}}(\theta) \geq N_C(\theta)$, since the latter entropy is related to a $\mathscr{G}_\theta$ of minimal cardinality. In general, however, we shall want to choose $\mathscr{G}_\theta$ so that both entropies are effectively the same. Now we introduce the "scaling hypothesis," by assuming the existence of a function f such that in addition to (2.1) we have

$$N_A\big(f(\theta)\varepsilon\big) \leq a\varepsilon^{-\kappa}, \quad \text{for all } A \in \mathscr{G}_\theta, \tag{2.5}$$

and small enough $\varepsilon, \theta > 0$. Such an f always exists. [Take $f \equiv 1$! Clearly, however, for this partitioning procedure to have any value, we shall want to choose $\mathscr{G}_\theta$ and f such that as well as (2.4) and (2.5) holding we have that $f(\theta) \searrow 0$ as $\theta \searrow 0$. Nevertheless, it is not necessary to assume this now, and the bound in Theorem 2.1 is correct for any f. If f does not decrease to zero with θ, however, it is uninteresting.]

Note that it would be nice to replace (2.5) with the more pleasing condition $N_A(f(\theta)\varepsilon) \leq N_C(\varepsilon)$ comparing entropies. However, such a condition turns out to be impractical in examples, since we generally do not have the precise form of $N_C(\varepsilon)$, but only its growth rate.

Note, also, that we can always take $N_C^{\mathscr{G}}(\theta)$ to be nonincreasing and f left continuous. Consequently, fixing some $p \geq 2$, the function

$$g(\theta) := \theta + 2f(\theta)/p^2$$

can also be taken to be left continuous, so that its inverse

$$g^{-1}(\eta) := \sup\{\theta : g(\theta) \leq \eta\}$$

is well defined.

THEOREM 2.1. *Suppose for* $\varepsilon \in (0, \varepsilon_0]$, $N_C(\varepsilon) \le a\varepsilon^{-\kappa}$, *and, furthermore, that for all* $\theta \in (0, \theta_0]$ *we have that* f *satisfies* (2.5). *Then for every* $p \ge \max(2, \varepsilon_0^{-1/2})$, $\sigma_A := \sup_{x \in A}\|x\|$, *any* $A \in \mathscr{G}_\theta$, *and all* $\lambda > g(\theta)(1 + 4\kappa \ln p)^{1/2}$,

$$
\begin{aligned}
P\Big\{\sup_{x\in A}|Lx| > \lambda\Big\} &\le 2\psi\Big(\big[\lambda - g(\theta)(1 + 4\kappa\ln p)^{1/2}\big]/\sigma_A\Big) \\
&\quad + 8ap^{2\kappa}\psi(\lambda/g(\theta)) \\
&\quad + 8ap^{2\kappa}\sigma_A\lambda^{-1}\exp\big(-\lambda^2/2\sigma_A^2\big)\exp\big(\lambda^2 g^2(\theta)/2\sigma_A^4\big).
\end{aligned}
\tag{2.6}
$$

An easy consequence of (2.6), that we shall require later, goes as follows. Set $\theta_\lambda := g^{-1}([\lambda^2(1 + 4\kappa \ln p)]^{-1/2})$, and take $A \in \mathscr{G}_{\theta_\lambda}$. Then for appropriate $c_1 = c_1(p, \kappa, \sigma_A)$, $c_2 = c_2(p, \kappa)$, bounded for σ_A bounded away from zero,

$$
P\Big\{\sup_{x\in A}|Lx| > \lambda\Big\} \le c_1\lambda^{-1}\exp\big(-\tfrac{1}{2}\lambda^2/\sigma_A^2\big) + c_2\lambda^{-2}\exp\big[-\tfrac{1}{2}\lambda^4(1 + 4\kappa\ln p)\big].
\tag{2.7}
$$

[Substitute θ_λ into (2.6), and apply the inequality $\psi(u) < \sqrt{2/\pi}\, u^{-1}e^{-u^2/2}$.]

PROOF. The idea of the proof is simple. If θ is small, then so are the sets in $\mathscr{G}_\theta$. For $A \in \mathscr{G}_\theta$, choose some $x^* \in A$. For each $x \in A$, write $Lx = Lx^* + L(x - x^*)$. Since $\|x - x^*\|$ must be small, $L(x - x^*)$ should be also small (stochastically). To show this we consider $L(x - x^*)$ conditional on Lx^*, using an idea used previously in Adler and Brown (1986) and Berman (1985) for certain Gaussian processes on R^k. Consequently, $Lx = Lx^* +$ a smaller-order term. The various conditions on Lx^* and the smaller-order term that make $|Lx| > \lambda$ are what lead to the various terms in (2.6). Details are as follows.

Take $A \in \mathscr{G}_\theta$ and let x^* be a point in A satisfying $\|x^*\| = \sup_A\|x\|$. Consider the process $L^*x := L(x - x^*) = Lx - Lx^*$, and let A^* be its image in $\mathscr{L}^2(\Omega, P)$. Let I be the (identity) operator on A^* that simply identifies each element of A^* as a Gaussian variable. The inner product (u, v) of $u = L^*x$ and $v = L^*y$ in A^* is given by $E(L^*x, L^*y)$, I is isonormal on A^* and $\sup_{A^*}|Iu| = \sup_A|L^*x|$. Furthermore, it is trivial to check that

$$
\sup_{A^*}\|u\|_* \le \theta^2 \quad \text{and} \quad \|u - v\|_* = \|x - y\|.
$$

Thus, the entropy function for I is identical to that for L on the original space. Consequently, comparing (2.5) and (2.1) it follows from (2.2) that

$$
P\Big\{\sup_A|L^*x| > \lambda\big[\theta + 2f(\theta)p^{-2}\big]\Big\} \le \tfrac{5}{2}ap^{2\kappa}\int_\lambda^\infty e^{-u^2/2}\,du.
\tag{2.8}
$$

Furthermore, precisely the same bound holds if we replace L^*x by $L^{**}(x) := Lx - E(Lx|Lx^*)$. This follows as for L^*, on noting that $\|u - v\|_{**} \le \|u - v\|_*$, which follows from an easy calculation on conditional variances.

Now note that the event that interests us, $\sup_A|Lx| > \lambda$, is included in the union of the four events:

$$|Lx^*| > \lambda - g(\theta)(1 + 4\kappa \ln p)^{1/2}, \tag{2.9}$$

$$\sup_A |L^*x| > \lambda, \tag{2.10}$$

$$\sup_A Lx > \lambda \quad \text{and} \quad 0 \le Lx^* \le \lambda - g(\theta)(1 + 4\kappa \ln p)^{1/2}, \tag{2.11}$$

$$\inf_A Lx < -\lambda \quad \text{and} \quad -\lambda + g(\theta)(1 + 4\kappa \ln p)^{1/2} \le Lx^* \le 0. \tag{2.12}$$

The probability of (2.9) is bounded by the first term in (2.6), while the second term there bounds the probability of (2.10) by (2.8). The probabilities of (2.11) and (2.12), which are clearly identical, are a little more involved to derive.

Note first that if $\eta > 0$, then $E(Lx|Lx^* = \eta) \le \eta$, since x^* is a point of maximal norm. Consequently, $E(Lx|Lx^*) \le Lx^*$ on the set where $Lx^* \ge 0$, and so (2.11) is contained in the event

$$\sup_A L^{**}x > \lambda - Lx^* \quad \text{and} \quad 0 \le Lx^* \le \lambda - g(\theta)(1 + 4\kappa \ln p)^{1/2}.$$

But $L^{**}x$ and Lx^* are independent, so the probability of this event can be bounded by

$$\int_0^\gamma P\Big(\sup_A L^{**}x > \lambda - u\Big) p(u/\sigma_A)\sigma_A^{-1}\, du,$$

with $\gamma = \lambda - g(\theta)(1 + 4\kappa \ln p)^{1/2}$. Applying (2.8) for L^{**}, we can bound this by

$$2ap^{2\kappa} \int_0^\lambda \psi\left(\frac{\lambda - u}{g(\theta)}\right) p(u/\sigma_A)\sigma_A^{-1}\, du.$$

Set $z = \lambda(\lambda - u)$ and note $p(x + y) \le p(x)e^{-xy}$ to further bound this by

$$\frac{2ap^{2\kappa}}{\lambda\sigma_A} p\left(\frac{\lambda}{\sigma_A}\right) \int_0^\infty \psi\left(\frac{z}{\lambda g(\theta)}\right) e^{z/\sigma_A^2}\, dz.$$

This is now a standard integral, and turns out to be no more than half the last factor in (2.6). This completes the proof of the theorem. □

In what follows we shall require the following easy consequence of Theorem 2.1, necessary because the behaviour of the right-hand sides of (2.6) and (2.7) is not good for very small σ_A.

COROLLARY 2.1. *Theorem* 2.1 *and inequality* (2.7) *continue to hold if we replace σ_A in the right-hand side of* (2.6) *and* (2.7) *by any $\sigma > \sigma_A$, as long as we double all constants.*

PROOF. Note firstly that if Z_t, $t \in T$, is any a.s. bounded collection of variables, and Y independent with $P\{Y > 0\} = P\{Y < 0\} = \frac{1}{2}$, then

$$P\Big\{\sup_T |Z_t| > \lambda\Big\} \le 2P\Big\{\sup_T |Z_t + Y| > \lambda\Big\}. \tag{2.13}$$

Now take $\sigma \ge \sigma_A$ and Y zero mean Gaussian with variance $\sigma^2 - \sigma_A^2$, independent of Lx for all $x \in A$, and define a new process L^* by $L^*x = Lx + Y$. Consider the image of A under L^*, call it A^*, as part of an $\mathscr{L}^2$ space of Gaussian variables, where for any two points, u, v in the image such that $u = L^*x$, $v = L^*y$, $x, y \in A$ their inner product $(u, v)_*$ is given by $E(L^*x, L^*y)$. Then clearly

$$\|u\|_*^2 = \|x\|^2 + \sigma^2 - \sigma_A^2, \qquad \|u - v\|_* = \|x - y\|.$$

Consequently, $\sup_{A^*}\|u\|_* = \sigma$ and A^* has the same entropy function as A. Let I be the identity map on this set. Then I is clearly isonormal on A^*, and $\sup_{A^*}|Iu| = \sup_A|L^*x|$. Thus, we can apply Theorem 2.1 and (2.7) to I and then apply (2.13) with $Z = L$ to prove the corollary. □

It is clear from (2.7) that for large λ we find that the dominant term in the bound is $O(\lambda^{-1}\exp(-\frac{1}{2}\lambda^2/\sigma_A^2))$. *But this is of the order of the probability that a single zero mean Gaussian variable with variance* σ_A^2 *is greater than* λ. That is, we have replaced the supremum of L over A by its value at one point only. Essentially, this has been done by making A small as λ becomes large, since $A \in \mathscr{G}_{\theta_\lambda}$ and θ_λ will be small for λ large. That is, we have achieved at this stage a *discretization* of the supremum problem. This is actually the heart of the solution of the general problem, for all we need do now is sum the bounds of Theorem 2.1 and its corollaries over the various sets in $\mathscr{G}_\theta$ to bound the supremum over the whole of C.

To sum these bounds efficiently, we require further assumptions on the structure of C, as in the following, final result, for which we set

$$C_\delta^+ := \{x \in C: \|x\| > \delta\}, \qquad C_\delta^- := \{x \in C: \|x\| \le \delta\},$$

and recall that θ_λ was defined as $g^{-1}([\lambda^2(1 + 4\kappa \ln p)]^{-1/2})$.

THEOREM 2.2. *Suppose* $N_C(\varepsilon) \le a\varepsilon^{-\kappa}$ *for* $\varepsilon \in (0, \varepsilon_0]$, *and that there exist constants* c, β, γ *and* δ_0 *such that for each* $\theta \in (0, \theta_0]$ *there exists a partition* $\mathscr{G}_\theta$ *of* C *and a constant* n_θ *so that*

$$n(\delta_1, \delta_2, \theta) \le c(\delta_2 - \delta_1)^\beta N_C^{\mathscr{G}}(\theta) + n_\theta(\delta_2 - \delta_1)^\gamma, \qquad 0 \le \delta_2 - \delta_1 \le \delta_0, \tag{2.14}$$

where

$$n(\delta_1, \delta_2, \theta) := \#\left\{A \in \mathscr{G}_\theta: A \cap \left[C_{\delta_1}^+ \cap C_{\delta_2}^-\right] \ne \varnothing\right\}. \tag{2.15}$$

Then there exist constants c_1 *and* c_2 *such that for sufficiently large* λ

$$P\Big(\sup_{x\in C}|Lx| > \lambda\Big) \le c_1 N_C^{\mathscr{G}}(\theta_\lambda)\lambda^{-1-2\beta}e^{-\lambda^2/2\sigma^2} + c_2 n_{\theta_\lambda}\lambda^{-1-2\gamma}e^{-\lambda^2/2\sigma^2}, \tag{2.16}$$

where $\sigma := \sup_{x \in C}\|Lx\|$.

PROOF. For fixed λ define the sequence $\{\delta_i\}$ given by

$$\delta_0^2 = \tfrac{1}{2}\sigma^2, \qquad \delta_i^2 = \sigma^2 - (m-i)\lambda^{-2}, \qquad i = 1,\dots,m,$$

where $m := 1 + [\frac{1}{2}\sigma^2\lambda^2]$. Clearly, it will suffice for us to bound $P(\sup_{C_{\delta_0}^+}|Lx| > \lambda)$. Apply Corollary 2.1 to obtain

$$P\Big\{\sup_{C_{\delta_0}^+}|Lx| > \lambda\Big\} \le c\sum_{i=1}^{m} n(\delta_{i-1}, \delta_i, \theta_\lambda)\lambda^{-1}\exp\big(-\tfrac{1}{2}\lambda^2/\delta_i^2\big).$$

Note that $\delta_i - \delta_{i-1} \le 1/(\sigma\lambda^2)$. Take λ large enough for (2.14) to hold, and substitute to bound the above sum by

$$c\big[\sigma^{-\beta}\lambda^{-1-2\beta}N^{\mathscr{G}}(\theta_\lambda) + n_{\theta_\lambda}\sigma^{-\gamma}\lambda^{-1-2\gamma}\big]\sum_{i=1}^{m}\exp\big(-\tfrac{1}{2}\lambda^2/\delta_i^2\big). \tag{2.17}$$

Thus, to complete the proof we need only bound the last summation by $c\exp(-\frac{1}{2}\lambda^2/\sigma^2)$. This can be done as follows. Set $\alpha_i = \exp\{-\frac{1}{2}\lambda^2/(\sigma^2 - (m-i)\lambda^{-2})\}$. It is easy to check that $\alpha_{i-1} \le \alpha_i e^{-1/2\sigma^4} < \alpha_i$, so that the sum in (2.17) is bounded by

$$\alpha_m \sum_{k=1}^{m}\big(e^{-1/2\sigma^4}\big)^k \le \frac{\alpha_m}{1 - e^{-1/2\sigma^4}} = c\exp\big(-\tfrac{1}{2}\lambda^2/\sigma^2\big).$$

This completes the proof. □

COMMENT. Theorems 2.1 and 2.2 have much in common with Theorémè 2.1.1 of Weber (1978) and the results in Weber (1980). In both cases Weber treats only spaces C for which $\|x\|$ is constant over C (i.e., processes of constant variance), and has results dependent only on the behaviour of the entropy function. That his results cannot be expected to always yield minimal powers of λ in a bound like (2.16) follows from the fact that it is easy to construct examples of processes with the same κ in (2.1) but with suprema tail distributions behaving like $\lambda^\alpha e^{-c\lambda^2}$ for quite different α. [An example is given in Adler and Samorodnitsky (1985), where the reader can also find more results akin to those of this section.] In fact, for none of the examples of the following section, do either Weber's general results, or those of Berman (1985) (in the case of Example 3.1), yield the best powers of λ.

3. Examples. Our examples deal not with the isonormal process on Hilbert space H but with processes whose parameter space is somewhat less abstract. Thus, we shall have to translate these processes to the isonormal case. But this is easy, for if X_t is a zero mean Gaussian process on, say, a metric space (S, d) with continuous covariance function $R(s, t)$, then we simply identify H with the $\mathscr{L}^2$ space of X, and $C \subset H$ with the set $\{x \in H\colon x = X_t \text{ for some } t \in S\}$. For $x = X_t$, $y = X_s$ in C we have $(x, y)_H = R(t, s)$. Clearly, L is now the identity operator, so that Lx is simply x identified as a Gaussian variable rather than an element of H. Furthermore, $\sup_{x\in C}|Lx| \equiv \sup_{t\in S}|X_t|$.

Entropy calculations are only slightly more involved, for we shall generally partition C by first partitioning S (this is usually geometrically simpler) and then letting the above identification induce a corresponding partition on C. We shall work the first example carefully to explain what is happening. In the later two, we shall skimp on detail.

All the examples are connected with Brownian sheets. Let λ_k be Lebesgue measure on $[0,1]^k$. The zero mean Gaussian process W defined on Borel sets in $[0,1]^k$ with covariance

$$E[W(A)W(B)] = \lambda_k(A \cap B), \tag{3.1}$$

is called the *set indexed Brownian sheet*. The pinned version of W, denoted by $\mathring{W}(A) := W(A) - \lambda_k(A)W([0,1]^k)$ has covariance

$$E(\mathring{W}(A)\mathring{W}(B)) = \lambda_k(A \cap B) - \lambda_k(A)\lambda_k(B). \tag{3.2}$$

For the special case of W indexed only by k-intervals of the form $A_{\mathbf{t}} = \prod_{i=1}^k [0, t_i]$, we write $W(\mathbf{t}) := W(A_{\mathbf{t}})$ and $\mathring{W}(\mathbf{t}) = \mathring{W}(A_{\mathbf{t}})$, and call $W(\mathbf{t})$ and $\mathring{W}(\mathbf{t})$ the point indexed sheet and pinned sheet, respectively. $W(\mathbf{t})$ is of particular interest as a natural k-dimensional generalisation of Brownian motion while $\mathring{W}(A)$ arises as a weak limit in the empirical measure setting of the introduction [cf. Dudley (1978)]. We start with the point indexed pinned sheet.

EXAMPLE 3.1. There exists a finite c such that

$$P\left\{ \sup_{[0,1]^k} |\mathring{W}(\mathbf{t})| > \lambda \right\} \le c\lambda^{2(k-1)} e^{-2\lambda^2}. \tag{3.3}$$

[This result was originally established in somewhat greater generality in Adler and Brown (1986), where it was also shown that this bound serves, for different c, as a lower bound as well. It is not, however, obtainable from any other general Gaussian bound. Using Berman's (1985) result, for example, the best bound possible is only $O(\lambda^{2k-1} e^{-2\lambda^2})$.]

We rederive the result here to show how it can be obtained from the general theory. We shall apply Theorem 2.2, so we are basically concerned with finding a good bound for $n(\delta_1, \delta_2, \theta)$, and the other factors in (2.14).

We commence by noting

$$\begin{aligned} \|\mathring{W}(\mathbf{t}) - \mathring{W}(\mathbf{s})\|^2 &= E\left[(\mathring{W}(\mathbf{t}) - \mathring{W}(\mathbf{s}))^2 \right] \\ &\le \lambda(A_{\mathbf{t}} \triangle A_{\mathbf{s}}) \le \sum_{i=1}^k |t_i - s_i|, \end{aligned} \tag{3.4}$$

for all $\mathbf{s}, \mathbf{t} \in [0,1]^k$. Now, for each $\theta > 0$ set $m_\theta := 1 + [k\theta^{-2}]$ ($[x] :=$ integer part of x) and define a partition I_θ of $(0,1]^k$ by

$$I_\theta = \left\{ A \subset [0,1]^k : A = \prod_{i=1}^k \left(\frac{n_i}{m_\theta}, \frac{n_i + 1}{m_\theta} \right], n_i = 0, 1, \dots, m_\theta - 1 \right\}.$$

Furthermore, let $\mathscr{G}_\theta$ be the partition I_θ induces in H, the $\mathscr{L}^2$ space of $\mathring{W}$. By (3.4), if $x, y \in A \in \mathscr{G}_\theta$, then $\|x - y\| \le \theta$, so that $\mathscr{G}_\theta$ is a partition of the type required for Theorem 2.2, and

$$N_C^{\mathscr{G}}(\theta) = 1 + [k/\theta^2]^k \le 3k^k \theta^{-2k}, \tag{3.5}$$

the inequality following by simple algebra. By (3.5), C has polynomial entropy with $\kappa \le 2k$. We now check the scaling property.

Fix $\varepsilon > 0$, set $p_\varepsilon := 1 + [\varepsilon^{-2}]$, divide each $A \in I_\theta$ into p_ε^k equal k-intervals and map these into the corresponding $A \in \mathscr{G}_\theta$. Applying (3.4) once again, it is easy to check that

$$N_A(\theta\varepsilon) \le 3\varepsilon^{-2k}, \quad \text{for all } \varepsilon < (2k)^{-1/2} \text{ and } A \in \mathscr{G}_\theta.$$

Thus, we can take $f(\theta) = \theta$ in (2.5) and, for some $p \ge 2$,

$$\theta_\lambda = \lambda^{-1}\left[(1 + 2p^{-2})(1 + 8k \ln p)^{1/2}\right]^{-1}. \tag{3.6}$$

All that remains is to investigate $n(\delta_1, \delta_2, \theta)$. Firstly note that it suffices to consider $\delta_1 > \frac{1}{4}$, for we can break C into two parts, over which $\|x\| \le \frac{1}{4}$ and $\|x\| > \frac{1}{4}$. Over the first part the inequality (2.2) gives us an upper bound of $O(e^{-8\lambda^2})$ for the tail of the supremum, which is clearly of smaller order than the desired (3.3). Thus, the case $\delta_i \le \frac{1}{4}$ can be neglected. Now note that $C_{\delta_1}^+ \cap C_{\delta_2}^-$ is the image of the following set, in which we write $|\mathbf{t}|$ for $t_1 \times \cdots \times t_k$:

$$\begin{aligned} I(\delta_1, \delta_2) &= \left\{\mathbf{t}: \delta_1^2 < |\mathbf{t}|(1 - |\mathbf{t}|) \le \delta_2^2\right\} \\ &= \left\{\mathbf{t}: \tfrac{1}{2} - \left(\tfrac{1}{4} - \delta_1^2\right)^{1/2} < |\mathbf{t}| \le \tfrac{1}{2} - \left(\tfrac{1}{4} - \delta_2^2\right)^{1/2}\right\} \\ &\quad \cup \left\{\mathbf{t}: \tfrac{1}{2} + \left(\tfrac{1}{4} - \delta_2^2\right)^{1/2} \le |\mathbf{t}| < \tfrac{1}{2} + \left(\tfrac{1}{4} - \delta_1^2\right)^{1/2}\right\}. \end{aligned} \tag{3.7}$$

The second line follows via a little elementary algebra. To count the number of A from I_θ that intersect $I(\delta_1, \delta_2)$ it suffices to count the number of lattice points of the form $(n_1/m_\theta, \ldots, n_k/m_\theta)$ falling in $I(\delta_1, \delta_2)$. But this is relatively easy, for if we fix $n_1, \ldots, n_{k-1}$, then some more algebra applied to (3.7) shows that no more than $32\sqrt{2}\,(\delta_2 - \delta_1)^{1/2} m_\theta$ values of n_k are permissible. Allowing $n_1, \ldots, n_{k-1}$ to vary, we thus obtain

$$\begin{aligned} n(\delta_1, \delta_2, \theta) &\le c(m_\theta)^{k-1}(\delta_2 - \delta_1)^{1/2} m_\theta \\ &\le c(k)\theta^{-2k}(\delta_2 - \delta_1)^{1/2} \\ &\le c(\delta_2 - \delta_1)^{1/2} N_C^{\mathscr{G}}(\theta). \end{aligned}$$

But this is all we need, for substitution into (2.16), on noting that $\sigma^2 = \frac{1}{4}$ for this problem, immediately establishes the required (3.3).

EXAMPLE 3.2. Let $\mathscr{R}_k$ be the set of all k-intervals of the form $[\mathbf{s}, \mathbf{t}] = \prod_{i=1}^k [s_i, t_i]$ contained in $[0,1]^k$. Then there exists a constant c such that

$$P\left\{\sup_{\mathscr{R}_k} |\mathring{W}(A)| > \lambda\right\} \le c\lambda^{2(2k-1)} e^{-2\lambda^2}. \tag{3.8}$$

Before we prove this result, we shall establish its sharpness by showing that there exists a c' such that

$$c'\lambda^{2(2k-1)} e^{-2\lambda^2} \le P\left\{\sup_{\mathscr{R}_k} \mathring{W}(A) > \lambda\right\}. \tag{3.9}$$

We shall prove this for $k = 2$. For $k > 2$ the proof is basically the same, the notation is just a little longer. Let $A = [\mathbf{s}, \mathbf{t}]$ be a rectangle in $[0,1]^2$, and define a mapping T: $\mathscr{R}_2 \to [0,1]^4$ by

$$T([\mathbf{s},\mathbf{t}]) := \left(\frac{t_1 - s_1}{t_1}, t_1, \frac{t_2 - s_2}{t_2}, t_2 \right).$$

Clearly, we must have $0 \le s_i \le t_i \le 1$, $i = 1,2$, for $[\mathbf{s},\mathbf{t}]$ to be in $\mathscr{R}_2$, and so it is easy to see that T is one–one and onto. The inverse mapping is defined by

$$T^{-1}(z_1, z_2, z_3, z_4) = \left[(z_2(1 - z_1), z_4(1 - z_3)), (z_2, z_4) \right]. \tag{3.10}$$

Now define a process $X(\mathbf{z})$ on $[0,1]^4$ by $X(\mathbf{z}) = \mathring{W}(T^{-1}(\mathbf{z}))$. This process is clearly Gaussian with zero mean, and it follows from (3.10) and (3.2) that

$$E\left[X^2(\mathbf{z})\right] = \lambda(T^{-1}(\mathbf{z})) - \left(\lambda(T^{-1}(\mathbf{z}))\right)^2 = |\mathbf{z}| - |\mathbf{z}|^2. \tag{3.11}$$

This is the variance of the point indexed sheet on $[0,1]^4$. After a page or so of elementary algebra, one can derive the rather useful inequality that for any $A, B \in \mathscr{R}_2$,

$$\lambda(A \cap B) \le \prod_{i=1}^{4} \left[T_i(A) \wedge T_i(B) \right],$$

where $T_i(A)$ is the ith coordinate of $T(A)$. An immediate consequence of this is that

$$E\left[X(\mathbf{u})X(\mathbf{v})\right] = E\left[\mathring{W}(T^{-1}\mathbf{u})\mathring{W}(T^{-1}\mathbf{v})\right] \le \prod_{i=1}^{4} u_i \wedge v_i - |\mathbf{u}| \cdot |\mathbf{v}|.$$

That is, the covariance function of X is dominated by that of the point indexed sheet on $[0,1]^4$. Consequently, by (3.11) and Slepian's inequality [Slepian (1962)], the tail of $\sup X$ dominates that of the sheet. Theorem 2.2 of Adler and Brown (1986) states that this, in turn, dominates $c'\lambda^6 e^{-2\lambda^2}$ for some c' [or $c'\lambda^{2(2k-1)}e^{-2\lambda^2}$ for general k], so that (3.9) is proven.

Now to the upper bound. We shall give the main steps of the derivation and skip all the algebra, most of which is similar to that in the previous example. To define $\mathscr{G}_\theta$, set $m_\theta = 1 + [2k/\theta^2]$, and let $\mathscr{G}_\theta$ be the image in H of the partition of $\mathscr{R}_k$ given by $\bigcup_{\mathbf{J} \in L_k(\theta)} A(\mathbf{J})$, where $L_k(\theta)$ is the set of all integer $2k$-tuples of the form $(j_1^{(1)}, j_1^{(2)}, \dots, j_k^{(1)}, j_k^{(2)})$ with $j_i^{(1)} \le j_i^{(2)}, \dots, i = 1, \dots, k$, $j_i^{(l)} = 0, 1, \dots, m_\theta - 1$, $i = 1, \dots, k$, $l = 1,2$, and $A(\mathbf{J})$ is the collection of all k-intervals $[\mathbf{x}, \mathbf{y}]$ satisfying $|x_i - j_i^{(1)}/m_\theta| \le \theta^2/2k$, $|y_i - j_i^{(2)}/m_\theta| \le \theta^2/2k$, $i = 1, \dots, k$. It is easy to see that $\mathscr{G}_\theta$ is a partition of the required form, and that

$$N_C^{\mathscr{G}}(\theta) \le 3.4^k k^{2k} \theta^{-4k} = c\theta^{-4k}.$$

Consequently, we have polynomial entropy with parameter $\kappa = 4k$. Continuing the same procedure, it is easy to see that, for each $A \in \mathscr{G}_\theta$, $N_A(\theta\varepsilon) \le c\varepsilon^{-4k}$, so that as in the previous case we have $f(\theta) = \theta$ and $\theta_\lambda = c\lambda^{-1}$.

Now consider $C_{\delta_1}^+ \cap C_{\delta_2}^-$, which we can write as

$$\left\{ B = \prod_{i=1}^{k} [x_i, y_i] \colon \delta_1^2 < \prod_i^k (y_i - x_i) - \left[\prod_i^k (y_i - x_i) \right]^2 \le \delta_2^2 \right\}.$$

Again we can assume $\delta_1 > \frac{1}{4}$, and follow the procedure of the previous example to eventually obtain

$$n(\delta_1, \delta_2, \theta) \le c N_C^{\mathscr{G}}(\theta)(\delta_2 - \delta_1)^{1/2}, \quad \text{for } \delta_2 - \delta_1 \le c\theta^2.$$

If $\delta_2 - \delta_1 > c\theta^2$, then dividing the interval (δ_1, δ_2) into subintervals of length at most $c\theta^2$ we obtain, from the above bound,

$$n(\delta_1, \delta_2, \theta) \le c(\delta_2 - \delta_1)\theta^{-1} N_C^{\mathscr{G}}(\theta).$$

Overall we obtain

$$n(\delta_1, \delta_2, \theta) \le c N_C^{\mathscr{G}}(\theta)(\delta_2 - \delta_1)^{1/2} + c\theta^{-1} N_C^{\mathscr{G}}(\theta)(\delta_2 - \delta_1).$$

Substituting all the above into Theorem 2.2, together with the fact that $\sigma = \frac{1}{2}$, we prove (3.8).

In the final example we treat $\mathring{W}$ indexed by all half squares in R^2, i.e., the space

$$\mathscr{D}_2 := \{ A \subset [0,1]^2 \colon A = [0,1]^2 \\ \cap \{(x, y) \colon \alpha x + \beta y + \gamma \le 0 \text{ some } \alpha, \beta, \gamma \in [-\infty, \infty]\} \}.$$

EXAMPLE 3.3. For the Brownian sheet indexed by half squares, we have

$$P\left\{ \sup_{\mathscr{D}_2} |\mathring{W}(A)| > \lambda \right\} \le c\lambda^2 e^{-2\lambda^2}, \tag{3.12}$$

for some finite, positive c.

Note firstly that if $A \in \mathscr{D}_2$, then $\mathring{W}(A) = -\mathring{W}(A^c)$. Consequently, we need only consider half of $\mathscr{D}_2$, say those half squares that contain at least one of the points $(1,0)$ or $(1,1)$. We write this as $\mathscr{D}_2^+$.

Let $S_1, \dots, S_4$ denote the four sides of the unit square, $\{(x, y) \colon 0 \le x,\ y \le 1\}$ on which, respectively, $x = 0,\ x = 1,\ y = 0,\ y = 1$. To define $\mathscr{G}_\theta$ set $m_\theta = [\theta^{-2}]$ and $x_i^{(k)}(\theta)$ the point on S_k at a distance i/m_θ from its start. Now let $A(\theta, k, l, i, j)$ be the collection of all half planes in $\mathscr{D}_2^+$ with boundary intersecting S_k between $x_i^{(k)}$ and $x_{i+1}^{(k)}$, and S_l between $x_j^{(l)}$ and $x_{j+1}^{(l)}$, $k, l = 1, \dots, 4$, $k \ne l$, $i, j = 0, 1, \dots, m_\theta - 1$. These A provide a partition of $\mathscr{D}_2^+$, and we take the induced partition in the $\mathscr{L}^2$ space of $\mathring{W}$ as $\mathscr{G}_\theta$. Clearly, $\mathscr{G}_\theta$ has the properties we generally require and, furthermore,

$$N_C^{\mathscr{G}}(\theta) = \binom{4}{2}(m_\theta + 1)^2 \le 24\theta^{-4}. \tag{3.13}$$

Consequently, we have polynomial entropy with $\kappa = 4$. To further subdivide these sets, simply subdivide each interval $[x_i^{(k)}, x_{i+1}^{(k)}]$ more finely, so that simple calculations yield that $N_A(\varepsilon\theta) \le 4\varepsilon^{-4}$ for each such A. Consequently, $f(\theta) = \theta$ and for $p \ge 2$,

$$\theta_\lambda = \lambda^{-1}\left[(1 + 2p^{-2})(1 + 8 \ln p)^{1/2}\right]^{-1}.$$

It remains to estimate $n(\delta_1, \delta_2, \theta)$, for which we must describe $C^+_{\delta_1} \cap C^-_{\delta_2}$. As before, this is made up of the image of all half squares whose intersections with $[0,1]^2$ have area S satisfying either

$$a_1 = \tfrac{1}{2} + \left(\tfrac{1}{4} - \delta_2^2\right)^{1/2} \le S < \tfrac{1}{2} + \left(\tfrac{1}{4} - \delta_1^2\right)^{1/2} = b_1 \tag{3.14}$$

or

$$a_2 = \tfrac{1}{2} - \left(\tfrac{1}{4} - \delta_1^2\right)^{1/2} < S \le \tfrac{1}{2} - \left(\tfrac{1}{4} - \delta_2^2\right)^{1/2} = b_2. \tag{3.15}$$

We further divide $C^+_{\delta_1} \cap C^-_{\delta_2}$, into the image of half squares whose intersection with $[0,1]^2$ is a proper quadrilateral, and those that yield a triangle. We shall count only the first case. The second can be treated similarly and yields same order of magnitude bounds on $n(\delta_1, \delta_2, \theta)$. Because of symmetry, we need only treat quadrilaterals including all of the side S_2, for we then simply add a factor of two to our counting to account for the side S_3. Such quadrilaterals can be parametrized by two points u and v representing, respectively, the points of intersection of the boundary of the half plane with the sides S_3 and S_4 of $[0,1]^2$. Then the area of the quadrilateral is given by $1 - \frac{1}{2}(u+v)$. For such a quadrilateral to be in the preimage of $C^+_{\delta_1} \cap C^-_{\delta_2}$, it thus follows from (3.14) and (3.15) that

$$2(1-b_i) \le u + v \le 2(1-a_i), \quad \text{for } i = 1 \text{ or } 2. \tag{3.16}$$

For any $A(\theta, 3, 4, j_1, j_2) \in \mathscr{G}_\theta$ that contains a set from $C^+_{\delta_1} \cap C^-_{\delta_2}$, we have that

$$0 \le u - j_1/m_\theta \le \theta^2, \qquad 0 \le v - j_2/m_\theta \le \theta^2,$$

for some u and v satisfying (3.16). Thus,

$$2(1-b_i)m_\theta - 4 \le j_1 + j_2 \le 2(1-a_i)m_\theta, \quad \text{for } i = 1 \text{ or } 2. \tag{3.17}$$

For fixed a_i, b_i the number of pairs (j_1, j_2) satisfying (3.17) is no more than $4m_\theta[1 + (b_i - a_i)m_\theta]$. Note that via a little algebra

$$16(b_i - a_i) = \left(1 - 4\delta_1^2\right)^{1/2} - \left(1 - 4\delta_2^2\right)^{1/2} \le c(\delta_2 - \delta_1)^{1/2}.$$

Thus,

$$\begin{aligned} n(\delta_1, \delta_2, \theta) &\le 4m_\theta + c(\delta_2 - \delta_1)^{1/2} m_\theta^2 \\ &\le c\theta^{-2} + c(\delta_2 - \delta_1)^{1/2} N_C^{\mathscr{G}}(\theta). \end{aligned}$$

Since these calculations are good for (say) $\delta_2 - \delta_1 < \frac{1}{2}$ we can now apply Theorem 2.2 and the fact that $\sigma = \frac{1}{2}$ to obtain (3.12) and so complete the proof.

Acknowledgments. Some of the results presented here, when restricted to the class of homogeneous Gaussian fields on $\mathbb{R}^k$, have a significant overlap with the "extended Fernique inequality" in Berman (1985). We had already obtained these results independently before hearing, from Professor Berman, of this work. However, when he very kindly sent us a preliminary (still untyped) version of his results we took advantage of the opportunity to combine what was best in both proofs, and so the statements and proofs of Theorem 2.1, when restricted to simple random fields, have much in common with his results. We are grateful to Larry Brown, who did most of the hard work in Adler and Brown (1986). It was his insight on the problems tackled there that set us off on the current work. Both a referee, and Professor Weber himself, drew our attention to the results of Weber (1978, 1980). We are grateful to Professor Weber for correspondence helping to clarify the relationships between his work and an earlier version of this paper. Further details on these relationships can be found in Adler and Samorodnitsky (1985). We would also like to thank another referee for pointing out an error in an earlier version of the paper.

REFERENCES

ADLER, R. J. and BROWN, L. D. (1986). Tail behaviour for suprema of empirical processes. *Ann. Probab.* **14** 1–30.

ADLER, R. J. and SAMORODNITSKY, G. (1985). Tail behaviour for the suprema of Gaussian processes with a view towards empirical processes. Technical Report 127, Center for Stochastic Processes, Univ. of North Carolina, Chapel Hill, N.C.

BERMAN, S. M. (1985). An asymptotic bound for the tail of the distribution of the maximum of a Gaussian process. *Ann. Inst. H. Poincaré Sect. B* (*N.S.*) **25** 45–57.

DUDLEY, R. M. (1967). The sizes of compact sets 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.

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

DUDLEY, R. M. (1984). A course on empirical processes. *École d'Été de Probabilités de Saint Flour XII–1982. Lecture Notes in Math.* **1097** 1–142. Springer, New York.

FERNIQUE, X. (1975). Regularité des trajectories des fonctions aleatoires gaussienns. *École d'Été de Probabilités de Saint Flour IV–1974. Lecture Notes in Math.* **480** 1–96. Springer, New York.

SEGAL, I. E. (1954). Distributions in Hilbert space and canonical systems of operators. *Trans. Amer. Math. Soc.* **88** 12–41.

SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.* **41** 463–501.

WEBER, M. (1978). Classes supérieures de processus gaussiens. *Z. Wahrsch. verw. Gebiete* **42** 113–128.

WEBER, M. (1980). Analyse asymptotique des processus gaussiens stationnaires. *Ann. Inst. H. Poincaré Sect. B* (*N.S.*) **16** 117–176.

ISRAEL INSTITUTE OF TECHNOLOGY
TECHNION
FACULTY OF INDUSTRIAL ENGINEERING
AND MANAGEMENT
TECHNION CITY
HAIFA 32 000
ISRAEL

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
AT CHAPEL HILL
CHAPEL HILL, NORTH CAROLINA 27514