The Annals of Probability
1986, Vol. 14, No. 1, 1–30

TAIL BEHAVIOUR FOR SUPREMA OF EMPIRICAL PROCESSES

BY ROBERT J. ADLER[1] and LAWRENCE D. BROWN[2]

Technion and Cornell University

We consider multivariate empirical processes $X_n(t) := \sqrt{n}(F_n(t) - F(t))$, where F_n is an empirical distribution function based on i.i.d. variables with distribution function F and $t \in \mathbb{R}^k$. For X_F the weak limit of X_n, it is shown that

$$c(F, k)\lambda^{2(k-1)}e^{-2\lambda^2} \leq P\left\{\sup_t X_F(t) > \lambda\right\} \leq C(k)\lambda^{2(k-1)}e^{-2\lambda^2}$$

for large λ and appropriate constants c, C. When $k = 2$ these constants can be identified, thus permitting the development of Kolmogorov—Smirnov tests for bivariate problems. For general k the bound can be used to obtain sharp upper–lower class results for the growth of $\sup_t X_n(t)$ with n.

1. Introduction. It is well known that the Kolmogorov–Smirnov (KS) statistic, based on a sample from any univariate random variable with continuous distribution function (d.f.), is distribution free. It is also well known that in the multivariate situation this is not the case, and it is to this situation that we shall soon direct our efforts. In the beginning, however, Kolmogorov (1933) showed that the one-sided statistic, $T_n = \sup\{\sqrt{n}(F_n(x) - F(x)): x \in \mathbb{R}^1\}$, where F denotes the underlying d.f. and F_n the empirical d.f., satisfies

$$P\{T_n > \lambda\} \to e^{-2\lambda^2} \quad \forall\, \lambda \quad \text{as } n \to \infty. \tag{1.1}$$

Smirnov (1944) extended this result to the two-sample problem, Feller (1948) gave it a neater proof, and Doob (1949) followed by Donsker (1951, 1952) and the theory of weak convergence explained it in terms of the convergence of $\sqrt{n}(F_n - F)$ to a limiting Gaussian process whose maximum had the tail distribution $\exp(-2\lambda^2)$.

In the multivariate case, there is no simple analogue to (1.1), and the best one can hope to obtain is either a limiting distribution for some specific F, or bounds that may be valid for a family of F sharing, perhaps, some regularity properties. The first attack on this problem was made by Kiefer and Wolfowitz (1958), who showed that if $T_n^{(k)}$ is the one-sided KS statistic in k dimensions, then for some $\alpha = \alpha(k) > 0$ and $c = c(k) < \infty$,

$$P\left\{T_n^{(k)} > \lambda\right\} \leq ce^{-\alpha\lambda^2} \quad \forall\, n, \lambda, F. \tag{1.2}$$

Despite the fact that this bound is obviously very crude, it did at least suffice to

Received April 1984; revised May 1985.
[1]Supported in part by AFOSR 83-0068.
[2]Supported in part by NSF MCS 8200031.
AMS 1980 *subject classifications.* Primary; 62G30; secondary; 60F10, 60F15, 62E15, 62E20.
Key words and phrases. Tail behaviour of suprema, empirical processes, Kolmogorov–Smirnov tests, Gaussian random fields.

prove the existence of a limiting distribution for $T_n^{(k)}$ as $n \to \infty$. [The full weak convergence of the empirical d.f. to an appropriate limiting Gaussian random field was later established by Dudley (1966, 1967).] However, although Kiefer and Wolfowitz established the existence of this limiting distribution, no explicit form for it is known. Indeed, there is only one nontrivial case where reasonably accurate bounds are known, this being the case where F is uniform on the unit square. Here the limiting distribution of $\sqrt{n}(F_n - F)$ is that of a pinned Brownian sheet, and fairly close lower and upper bounds on the distribution of its maximum appear in Goodman (1976) and Cabaña and Wschebor (1982), respectively. We shall have more to say on this later, when Goodman's lower bound is extended to arbitrary dimensions.

In a classic paper, Kiefer (1961) greatly improved on (1.2) and showed that for all $\varepsilon > 0$ there is a $c = c(k, \varepsilon)$ such that

$$P\{T_n^{(k)} > \lambda\} \leq ce^{-2(1-\varepsilon)\lambda^2} \quad \forall\, n, \lambda, F. \tag{1.3}$$

This is a particularly interesting bound since, viewed as a result on the maximum of the limiting Gaussian field, rather than as a result on $T_n^{(k)}$ itself, it is one of the few forerunners of the general inequality for continuous Gaussian processes, $X(t)$, that states that for all $a < (2\sigma^2)^{-1}$ and $\lambda > \lambda(a)$ large enough

$$P\Big\{\sup_t X(t) > \lambda\Big\} \leq e^{-a\lambda^2}, \tag{1.4}$$

where $\sigma^2 = \sup_t\{\operatorname{var} X(t)\}$ (Fernique, 1970, 1971, Landau and Shepp, 1971, and Marcus and Shepp, 1971). Note that since $\sup_x\{\operatorname{var}[\sqrt{n}(F_n(x) - F(x))]\} = \frac{1}{4}$, Kiefer's bound, for *fixed* λ, is obtainable from (1.4) and weak convergence. Kiefer's result, as a bound for *all* λ, is not, however, thus obtainable. In any case, historically, (1.3) is particularly interesting, since its consequences for the limiting Gaussian field represent, to the best of our knowledge, the first time that a uniform bound was placed on the maxima of a large family of Gaussian processes. (The *statistical* significance of such a lower bound is that it permits construction of "confidence intervals" for an unknown F.) Furthermore, Kiefer exploited (1.3) to prove a law of the iterated logarithm (LIL) for the multivariate KS statistic.

The main thrust of the current work will be to further refine (1.3), in two directions, and then to investigate the consequences of the refinement. For a start, it follows easily from the results of Section 4 and an embedding theorem (Section 5) that (1.3) can be replaced by: There is a $c = c(F)$ such that for $n > n(\lambda)$

$$P\{T_n^{(k)} > \lambda\} \leq c\lambda^{2(k-1)}e^{-2\lambda^2}. \tag{1.5}$$

This, as with Kiefer's result, is of interest beyond the KS situation, since, in the Gaussian process setting, it provides a family of processes for which (1.4) can be improved upon. However, we can do better than just (1.5), and we shall also show that as long as F satisfies mild regularity conditions, there is a $c = c(F)$ such that for $n > n(\lambda)$

$$P\{T_n^{(k)} > \lambda\} \geq c\lambda^{2(k-1)}e^{-2\lambda^2}. \tag{1.6}$$

The upper and lower bound together enable us to improve on Kiefer's LIL and to obtain an exact upper–lower class result in its place (Section 5).

An upper bound similar in spirit to (1.5) has recently been obtained by Alexander (1984). Treating a more general situation of empirical measures, ν_n, indexed by a Vapnik–Červonenkis class of functions, $\mathscr{F}$ say, he showed that

$$P\left\{\sup_{f\in\mathscr{F}}|\nu_n(f)|>\lambda\right\}\leq 16\lambda^{v2^{12}}e^{-2\lambda^2}\quad\forall\,\lambda\geq 8,$$

where v is a strictly positive integer describing the "size" of $\mathscr{F}$. Alexander's result, while clearly being an improvement on (1.3), also gives, for the cases we consider, an enormous overestimate of the power of λ in the upper bound. (In fairness to Alexander, we should point out that he was not primarily interested in sharp powers of λ, so it is not surprising that his power is so large.)

Unlike Alexander, however, we shall have little to say about the sizes of the constants in our bounds, other than to guarantee their finiteness. Thus, from the point of view of actually applying the KS statistic in a statistical setting, these results are of limited interest. We shall remedy this situation in Section 3, where, for the two-dimensional case, we shall develop an explicit, sharp upper bound, and a reasonable lower bound. The various applications of these results are spelled out in detail in Brown and Adler (1985). The argument leading to the upper bound is rather interesting, since it is based on finding the worst possible F (a task actually performed by Kiefer) and comparing it, via Slepian's (1962) inequality, to all other cases. The distribution of the maximum in the worst possible case is what then provides the bound. In fact, this methodology of "comparison" will also be used to obtain the lower bound (1.6), and may, in a certain sense, be considered the main methodological theme of this paper.

The following section is devoted to peripheral and support material. There we obtain lower bounds for the distribution of the maximum of the pinned Brownian sheet in k dimensions and some related distributions. While these do have some intrinsic interest, our main interest in them will arise from their usefulness as "comparison distributions." We close this section with notation and some background results.

Let $X_1, X_2, \ldots,$ be independent random variables with d.f. $F(x)$, which we assume to be continuous, and which can therefore, without loss of generality, be taken to be concentrated on the unit cube $I^k=[0,1]^k$ of $\mathbb{R}^k$ with univariate marginals uniform on $[0,1]$. We denote a point in I^k by either x or $(x_1,\ldots,x_k)$ and introduce the usual partial order.

$$x<y\Leftrightarrow x_i<y_i,\qquad i=1,\ldots,k,\qquad x,\,y\in I^k.$$

For $x<y$ we write $[x,y]$ for the set $\Pi_{i=1}^k[x_i,y_i]$, and use $I_A(\cdot)$ for the indicator function of the set $A\subset I^k$. Thus we can formally introduce the empirical d.f. F_n as

$$F_n(x):=n^{-1}\sum_{i=1}^n I_{[0,x]}(X_i). \tag{1.7}$$

Let W_F be the pinned Brownian sheet based on F; i.e., the zero mean Gaussian process with covariance function

$$R_F(x, y) = E\{W_F(x)W_F(y)\} = F(x \wedge y) - F(x)F(y), \qquad x, y \in I^k, \tag{1.8}$$

where $x \wedge y$ is the coordinatewise minimum $(x_1 \wedge y_1, \ldots, x_k \wedge y_k)$.

Then, as is well known [$k = 1$, Donsker (1952); $k > 1$, Dudley (1966, 1967)], $\sqrt{n}(F_n - F)$ converges weakly to W_F in the space of all bounded functions on I^k under the topology of uniform convergence. Thus, in particular, if

$$T_n^k = T_n^k(F) := \sup\{\sqrt{n}(F_n(x) - F(x)): x \in I^k\} \tag{1.9}$$

is the one-sided KS statistic, then

$$T_n^k(F) \to_{\mathscr{L}} M_F := \sup\{W_F(x): x \in I^k\} \quad \text{as } n \to \infty. \tag{1.10}$$

This last result provides the obvious motivation for the next two sections, both of which are concerned with the distribution of M_F. In fact, one can go beyond the central limit result (1.9) to a much stronger embedding type result. However, since we shall not need this result until Section 5, we shall introduce it only then.

2. Two special cases. We consider firstly the distribution of M_F when F is the uniform distribution, U say, on I^k. We shall, however, require a slightly more general result later, and to this end let $W^{(k)}$ denote the (unpinned) Brownian sheet on I^k, i.e., the zero mean Gaussian process with covariance function

$$E\{W^{(k)}(x)W^{(k)}(y)\} = \prod_{i=1}^{k}(x_i \wedge y_i), \qquad x, y \in I^k,$$

and write $\mathring{W}^{(k)}$ for the pinned version of $W^{(k)}$ on I^k. Then a version of $\mathring{W}$ can be obtained from W by the correspondence

$$\mathring{W}^{(k)}(x) = W^{(k)}(x) - |x|W^{(k)}(1), \qquad x \in I^k, \tag{2.1}$$

where $|x| = \prod_{i=1}^{k} x_i$. In the general notation of the previous section $\mathring{W}^{(k)} \equiv W_U$. The result we shall need is

THEOREM 2.1.

$$P\Big\{\sup_{I^k} W^{(k)}(x) > \lambda \,\Big|\, W^{(k)}(1) = w\Big\} \geq e^{-2\lambda(\lambda - w)} \sum_{n=0}^{k-1} \frac{[2\lambda(\lambda - w)]^n}{n!}, \tag{2.2}$$

a.s. for all $\lambda > w$. *Furthermore, the case* $w = 0$ *yields*

$$P\Big\{\sup_{I^k} \mathring{W}^{(k)}(x) > \lambda\Big\} \geq e^{-2\lambda^2} \sum_{n=0}^{k-1} (2\lambda^2)^n/(n!). \tag{2.3}$$

When $k = 1$, (2.2) follows from the Brownian line crossing probabilities (Feller, 1971, pages 340–341). When $k = 2$ (2.3) is given explicitly in Goodman (1976),

and (2.2) is also there implicitly. Cabaña and Wschebor (1982) and Park and Skoug (1978) also have (2.2) in the two-dimensional case, and shortly after obtaining the above result for general k, we received a copy of Cabaña (1984) which states the same result with a virtually identical, albeit more detailed, proof. However, for the sake of completeness, we shall give a brief proof of the theorem.

PROOF. The proof proceeds by induction. As noted above, (2.2) is known to be true when $k = 1$ and $k = 2$. Now write $a_k(\lambda, w)$ for the conditional probability on the left in (2.2) and define

$$h_k(\lambda, w) = P\left\{ \sup_{I^k} \left[\frac{\mathring{W}^{(k)}(x) - \lambda}{|x|} \right] > w \right\}. \tag{2.4}$$

Then, after some calculation, it readily follows from (2.1) that $a_k(\lambda, w) = h_k(\lambda, -w)$. If we now follow the formulation of Goodman (1976) of treating the $(k + 1)$ parameter, real-valued, $W(x)$ as a $C_0[0,1]^k$-valued, single parameter process, then by applying Goodman's Theorem 2 and mimicking his manipulations on page 980, it is straightforward to establish the relation

$$a_{k+1}(\lambda, w) \geqq 1 - \int_{-\lambda}^{-w} (1 - e^{-2\lambda(w+u)}) h_k(\lambda, du). \tag{2.5}$$

Exploiting the above relationship between a_k and h_k thus yields a recurrence formula for a_{k+1}, and it is now a matter of elementary calculus to check the induction hypothesis and so complete the proof.

There are two points worth noting here. The proof uses the fact that $W^{(k+1)}$ can be treated as a $C_0[0,1]^k$-valued, *Markov* process in real time. The Markovianess is extremely important and is also a property of W_F for any F. However, also used is the fact that $W^{(k)}(x) - |x|W^{(k)}(x)$ and $W^{(k)}(1)$ are independent. (Here $W^{(k)}$ is the $C_0[0,1]^k$-valued process.) This is not true for general W_F and so the above proof only works in the uniform case.

The second result which, unlike Theorem 2.1, is of little independent interest, will be extremely useful for us later. To state it, we introduce, for each $\varepsilon \in (0, \frac{1}{2})$, the d.f. $F_\varepsilon(x) = F_\varepsilon(x_1, \ldots, x_k)$ which distributes probability 2ε uniformly on the cube $A_\varepsilon := [\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon]^k$, with the remaining probability of $1 - 2\varepsilon$ distributed uniformly on that part of the main diagonal of I^k disjoint from A_ε.

THEOREM 2.2. *There is a finite $c = c(\varepsilon, k)$ such that for all $\lambda > 0$*

$$P\left\{ \sup_{I_k} W_{F_\varepsilon}(x) > \lambda \right\} \geq c\lambda^{2(k-1)} e^{-2\lambda^2}. \tag{2.6}$$

PROOF. We commence with some necessary notation. Let $\psi_\varepsilon(u, v)$ denote the two-dimensional normal density with zero means and covariance matrix Σ_ε

defined by

$$\Sigma_\varepsilon := \begin{bmatrix} \left(\tfrac{1}{2}-\varepsilon\right)\left(\tfrac{1}{2}+\varepsilon\right) & \left(\tfrac{1}{2}-\varepsilon\right)^2 \\ \left(\tfrac{1}{2}-\varepsilon\right)^2 & \left(\tfrac{1}{2}-\varepsilon\right)\left(\tfrac{1}{2}+\varepsilon\right) \end{bmatrix}. \tag{2.7}$$

Furthermore, let $\hat{\Sigma}_\varepsilon$ be the matrix identical to Σ_ε, but with the sign of the off-diagonal entries reversed, and let $\hat{\psi}_\varepsilon$ be the two-dimensional normal density with zero mean and covariance matrix $\hat{\Sigma}_\varepsilon$. For each positive ε and λ, and integral k, set

$$\psi_{k,\varepsilon}(\lambda) := \int_{-\infty}^{2\varepsilon\lambda}\int_{-\infty}^{2\varepsilon\lambda}\left\{\sum_{n=0}^{k-1}\frac{[(u-2\varepsilon\lambda)(v-2\varepsilon\lambda)/\varepsilon]^n}{n!}\right\}\hat{\psi}_\varepsilon(u,v)\,du\,dv, \tag{2.8}$$

and write $Q_\varepsilon(\lambda)$ for the normal quadrant integral

$$Q_\varepsilon(\lambda) := 1 - \int_{-\infty}^{\lambda}\int_{-\infty}^{\lambda}\psi_\varepsilon(u,v)\,du\,dv. \tag{2.9}$$

We can now turn to the main part of the proof.

We shall obtain only a lower bound for $P\{\sup(W_{F_\varepsilon}(x)\colon x \in A_\varepsilon) > \lambda\}$, which, *a fortiori*, will provide the lower bound required. Let a_ε and b_ε be the two extreme corners of A_ε, i.e., $a_\varepsilon = (\tfrac{1}{2}-\varepsilon,\ldots,\tfrac{1}{2}-\varepsilon)$, $b_\varepsilon = (\tfrac{1}{2}+\varepsilon,\ldots,\tfrac{1}{2}+\varepsilon)$. Then define the process $Z(x)$ on I^k by

$$Z(x) := \left\{W_{F_\varepsilon}(a_\varepsilon + 2\varepsilon x) - (1-|x|)W_{F_\varepsilon}(a_\varepsilon) - |x|W_{F_\varepsilon}(b_\varepsilon)\right\}/\sqrt{2\varepsilon}\,. \tag{2.10}$$

Then it is straightforward to check that $Z(x)$ is a standard pinned sheet on I^k as in (2.1). Consequently, for $u, v < \lambda$, it follows that

$$\begin{aligned} &P\left\{\sup_{A_\varepsilon} W_{F_\varepsilon}(x) > \lambda \,|\, W_{F_\varepsilon}(a_\varepsilon) = u,\ W_{F_\varepsilon}(b_\varepsilon) = v\right\} \\ &\quad = P\left\{\sup_{I^k}\left(\frac{\mathring{W}^{(k)}(x) - (\lambda-u)/\sqrt{2\varepsilon}}{|x|}\right) \geq \frac{u-v}{\sqrt{2\varepsilon}}\right\}. \end{aligned} \tag{2.11}$$

But this is precisely the probability defined at (2.4). Thus, using the equivalence noted there between this probability and a_k, we can bound it by Theorem 2.1. Using this bound, (2.11), and the fact that the joint density of $\langle W_{F_\varepsilon}(a_\varepsilon), W_{F_\varepsilon}(b_\varepsilon)\rangle$ is given by ψ_ε, we obtain

$$\begin{aligned} P\left\{\sup_{I^k} W_{F_\varepsilon}(x) > \lambda\right\} &\geq P\left\{W_{F_\varepsilon}(a_\varepsilon) > \lambda \text{ or } W_{F_\varepsilon}(b_\varepsilon) > \lambda\right\} \\ &\quad + \int_{-\infty}^{\lambda}\int_{-\infty}^{\lambda} e^{-(\lambda-u)(\lambda-v)/\varepsilon} \\ &\qquad \cdot\left\{\sum_{n=0}^{k-1}\frac{[(\lambda-u)(\lambda-v)/\varepsilon]^n}{n!}\right\}\psi_\varepsilon(u,v)\,du\,dv. \end{aligned} \tag{2.12}$$

Consider the integrand and make the transformations $x = u - \lambda(1-2\varepsilon)$, $y =$

$v - \lambda(1 - 2\varepsilon)$. Tedious but straightforward algebra yields that it is equivalent to

$$e^{-2\lambda^2}\hat{\psi}_\varepsilon(x, y) \sum_{n=0}^{k-1} [(x - 2\varepsilon\lambda)(y - 2\varepsilon\lambda)/\varepsilon]^n/n!$$

Substituting this into (2.12), changing the bounds on the integral, and replacing the rightmost probability by $Q_\varepsilon(\lambda)$ now yields that for every $\lambda > 0$

$$P\left\{\sup_{I_k} W_{F_\varepsilon}(x) > \lambda\right\} \geq Q_\varepsilon(\lambda) + e^{-2\lambda^2}\psi_{k,\varepsilon}(\lambda). \tag{2.13}$$

To obtain (2.6) now simply take λ large enough so that the dominant term in the sum in $\psi_{k,\varepsilon}$ is $O(\lambda^{2(k-1)})$. Then choose an appropriate c. This completes the proof.

In general, we shall use Theorem 2.2 to form a basis for comparison between the maxima of pinned sheets based on different d.f.s. The crucial result that underlies all these comparisons is a basic result of Slepian (1962), which we record here as

LEMMA 2.1 (*Slepian's inequality*). *Let X and Y be two zero mean a.s. continuous Gaussian processes defined over some set T. If* $\operatorname{var} X(t) = \operatorname{var} Y(t)$, $\forall$ $t \in T$, *and*

$$\operatorname{cov}(X(t), X(s)) \leq \operatorname{cov}(Y(t), Y(s)) \quad \forall s, t \in T, \tag{2.14}$$

then

$$P\left\{\sup_T X(t) > \lambda\right\} \geq P\left\{\sup_T Y(t) > \lambda\right\} \quad \forall\, \lambda. \tag{2.15}$$

Note that Slepian's inequality does not extend to comparisons of $|\sup X|$ and $|\sup Y|$, and so the sharp results of the following section are not easily extendable to the two-sided KS statistic. Nevertheless, we can always use the fact that for symmetric processes

$$P\{\sup X > \lambda\} \leq P\{\sup|X| > \lambda\} \leq 2P\{\sup X > \lambda\} \tag{2.16}$$

to obtain bounds for the two-sided case. For the bounds of Section 4, in which constants are not identified, this is clearly sufficient.

We now consider, as an example of our "comparison methodology" the two-dimensional case. As already noted, the bounds we obtain here are somewhat more precise than those we succeeded in obtaining in higher dimensions.

3. The two-dimensional case. Throughout this section, we shall denote points in I^2 by $x = (x_1, x_2)$ and F will denote a continuous d.f. on I_2 possessing uniform marginals. The degenerate distribution, uniform on the negative slope diagonal $x_1 + x_2 = 1$, will be denoted by $G(x)$; i.e.,

$$G(x) = (x_1 + x_2 - 1)^+, \qquad x \in I^2. \tag{3.1}$$

Our aim in this section will be to devise good (nonasymptotic) bounds for $P\{\sup W_F > \lambda\}$. We start with

THEOREM 3.1. *For any two-dimensional d.f. F satisfying the above conditions and for any* $\lambda > 0$,

$$P\Big\{\sup_{I^2} W_F(x) > \lambda\Big\} \le P\Big\{\sup_{I^2} W_G(x) > \lambda\Big\}. \tag{3.2}$$

Furthermore,

$$P\Big\{\sup_{I^2} W_G(x) > \lambda\Big\} \le \sum_{n=1}^{\infty} (8n^2\lambda^2 - 2)e^{-2n^2\lambda^2}. \tag{3.3}$$

PROOF. Let m be the mapping I^2 onto $\{(x_1, x_2) \in I^2: x_1 + x_2 \ge 1\}$ defined by

$$G(m(x)) = G(m_1(x), m_2(x)) = F(x), \qquad \forall\, x \in I^2, \tag{3.4}$$

$$m_2(x) - m_1(x) = x_2 - x_1 \qquad \forall\, x \in I^2. \tag{3.5}$$

We must check that m is well defined. But this is easy, for by (3.5) we see that we are mapping lines of slope one onto themselves. At the beginning of each such line $F = G = 0$ and, as each line leaves I^2, the marginal uniformity of F and G ensures that they are again equal. Since both F and G are nondecreasing along such lines, and G is continuous and strictly increasingly when it is not zero, it is easy to arrange (3.4) in a unique fashion.

Now consider the processes W_F and W_G. We shall compare $\sup\{W_F(x): x \in I^2\}$ to $\sup\{W_G(x): x \in m(I^2)\} = \sup\{W_G(x): x \in I^2\}$. Note firstly that for $x \in I^2$

$$\operatorname{var} W_F(x) = \operatorname{var} W_G(m(x)), \tag{3.6}$$

a simple consequence of (3.4) and (1.8). We want to show that

$$F(x \wedge y) \ge G(m(x) \wedge m(y)), \qquad x, y \in I^2, \tag{3.7}$$

from which it would follow

$$\begin{aligned} \operatorname{cov}(W_F(x), W_F(y)) &= F(x \wedge y) - F(x)F(y) \\ &\ge G(m(x) \wedge m(y)) - G(m(x))G(m(y)) \\ &= \operatorname{cov}(W_G(m(x)), W_G(m(y))). \end{aligned} \tag{3.8}$$

But then we shall have completed the proof of the first part of the theorem, viz. (3.2), since (3.6) and (3.8) are precisely the ingredients for Slepian's inequality.

Returning to (3.7), consider firstly the case $x < y$. Then $F(x \wedge y) = F(x) = G(m(x)) \ge G(m(x) \wedge m(y))$ so (3.7) holds. The case $y < x$ is obviously identical, so now consider (3.7) for x, y with $x_1 > y_1$ and $x_2 < y_2$. (The remaining case is handled analogously.) Then $x \wedge y = (y_1, x_2)$. Write $w = (m_1(y), m_2(x))$. There are three possible cases to consider: $m(x) \ge w \ge m(y)$, $m(y) \ge w \ge m(x)$, and $w = m(x) \wedge m(y)$. We shall consider only the third case explicitly, but the reasoning is valid for all the cases. Note (drawing a picture helps to see the

inequalities) that

$$
\begin{aligned}
F(x \wedge y) &= F(y_1, x_2) \\
&\geq [F(x) - (x_1 - y_1)] \vee [F(y) - (y_2 - x_2)] \quad \text{by marginal uniformity} \\
&\geq \tfrac{1}{2}\{F(x) + F(y) - [(x_1 - x_2) - (y_1 - y_2)]\} \\
&= \tfrac{1}{2}\{G(m(x)) + G(m(y)) \\
&\qquad - [(m_1(x) - m_2(x)) - (m_1(y) - m_2(y))]\} \quad \text{by (3.4), (3.5)} \\
&\geq m_2(x) + m_1(y) - 1 \quad \text{by (3.1).}
\end{aligned} \tag{3.9}
$$

Hence, if $m_2(x) + m_1(y) - 1 \geq 0$ then the above yields

$$F(x \wedge y) \geq G(m_1(y), m_2(x)). \tag{3.10}$$

On the other hand, if $m_2(x) + m_1(y) - 1 < 0$, then $G(m_1(y), m_2(x)) = 0$ and so (3.10) is trivially true. Thus, in general,

$$F(x \wedge y) \geq G(m_1(y), m_2(x)) = G(w) = G(m(x) \wedge m(y)).$$

From this we immediately obtain (3.7) and the proof of (3.2).

It remains to establish the inequality (3.3). To this end, let $\mathring{W}(t)$, $t \in [0,1]$ be a standard Brownian bridge with covariance function

$$E\{\mathring{W}(t)\mathring{W}(s)\} = (s \wedge t) - st. \tag{3.11}$$

Define the two-parameter field X on I^2 by

$$X(x_1, x_2) = \begin{cases} \mathring{W}(x_1) - \mathring{W}(1 - x_2) & x_1 + x_2 - 1 \geq 0, \\ 0 & x_1 + x_2 - 1 \leq 0. \end{cases}$$

Then comparison of covariance functions shows that X is a version of W_G. Thus

$$
\begin{aligned}
P\{\sup \mathring{W}_G(x_1, x_2) > \lambda\} &= P\{\sup[\mathring{W}(x_1) - \mathring{W}(1 - x_2)\colon x_1 + x_2 - 1 \geq 0] > \lambda\} \\
&= P\{\sup[\mathring{W}(s) - \mathring{W}(t)\colon s > t] > \lambda\} \\
&\leq P\{\sup[\mathring{W}(s) - \mathring{W}(t)\colon s, t \in [0,1]] > \lambda\} \\
&= P\left\{\left[\sup_{[0,1]} (\mathring{W}(s))^+ + \sup_{[0,1]} (\mathring{W}(s))^-\right] > \lambda\right\}.
\end{aligned} \tag{3.12}
$$

But the last probability is known exactly, having been determined in Kac, Kiefer, and Wolfowitz [(1955), Equation (4.6)], and is precisely the sum given on the right of (3.3), and so we are finished. [Note: the original KKW formula is for $\frac{1}{2}[\sup(\mathring{W})^+ + \sup(\mathring{W})^-]$ and so needs to be slightly corrected to obtain our (3.3).]

REMARK. Note that the inequality in (3.12) is far from sharp, and a little reflection shows that this inequality, while retaining a bound of the right order of magnitude, "costs," roughly, a factor of two, i.e., we expect that the final upper bound in (3.3) is too large by a factor of at most two, while the inequality in (3.2) contributes further imprecision. Indeed, comparison of the general upper bound (3.3) with the specific lower bound in the uniform case, (2.3) with $k = 2$, shows,

for large λ, a difference between the bounds of precisely a factor of four. The precise value of the probability in (3.3) is given by $P\{\sup[\mathring{W}(s) - \mathring{W}(t)\colon 0 \le t < s \le 1] \ge \lambda\}$ [cf. (3.12)], but this seems hard to calculate. However, numerical estimates of this probability are easy to obtain via simulation and some are listed in Brown and Adler (1985). Furthermore, calculation of (3.3) and comparison with (2.3) for moderate λ, say $\lambda \in [1, 3]$, yields that for the uniform distribution (3.3) overestimates the true probability by less than a factor of four and that the KS test statistics derivable from (3.3) are in fact quite useful. For details see Brown and Adler (1985).

We now turn to the more difficult problem of finding a uniform lower bound for the two-dimensional case. Here we shall need to impose assumptions on F in order to avoid degeneracies (e.g., F concentrated on the diagonal $x_1 = x_2$, which reduces to the one-dimensional case). Let $\|x\| = |x_1| + |x_2|$ denote the "city block" norm of x. Then we shall prove

THEOREM 3.2. *Let F be a d.f. on I^2, with uniform marginals, such that there exists an $x_0 \in I^2$, a neighbourhood N of x_0, and a constant $\beta \in (0, 1]$ satisfying*

$$F(x_0) = \tfrac{1}{2} \tag{3.13}$$

and

$$\text{if } x, y \in N \text{ and either } x_1 = y_1 \text{ or } x_2 = y_2 \text{ then } |F(x) - F(y)| > \beta\|x - y\|. \tag{3.14}$$

Then there exists a finite $c = c(F) > 0$ such that

$$P\{\sup W_F(x) > \lambda\} \ge c\lambda^2 e^{-2\lambda^2}. \tag{3.15}$$

REMARK 1. The impact of (3.14) is that it guarantees that F is strictly increasing along horizontal and vertical lines in N (and so on all increasing paths) and thus "spreads mass."

REMARK 2. Theorem 3.2, as it stands, is a special case of the more general result Theorem 4.2. What makes it of special interest, however, is the fact that in two dimensions it is possible to obtain estimates for c. We shall discuss these at the end of the proof. Furthermore, the two-dimensional case turns out to be somewhat simpler than its higher-dimensional analogue, thereby making its proof more transparent and interesting.

It is clear that the conditions of Theorem 3.2 hold if F has a density bounded away from zero. However, absolute continuity is not a requisite of the theorem, and it is easy to build examples of non-absolutely continuous F satisfying (3.13) and (3.14). A trivial example is the extremal case (3.1).

PROOF OF THEOREM 3.2. The aim of the proof will be to compare W_F with W_{F_ε}, where F_ε is the distribution function of the preceding section, and then use Slepian's inequality and Theorem 2.2 to complete the argument. The comparison will only be possible over a region in the neighbourhood of $(\frac{1}{2}, \frac{1}{2})$ in the domain of

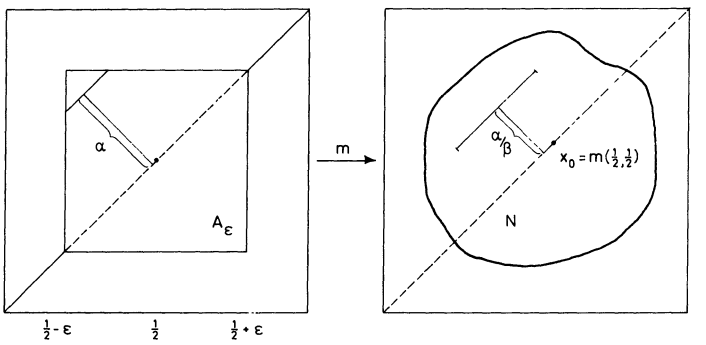

FIG. 1. *The map m takes lines of slope one in A_ε to lines of slope one in N.*

W_{F_ε}, together with a subset of N in the domain of W_F, but it will turn out that such a comparison will suffice for our purposes. We start by building a mapping between the above two neighbourhoods, and by noting that the reader's path through the forthcoming algebra will be considerably simpler if the argument is followed graphically with pen and paper. Figure 1 provides a starting point for such a venture.

For $x \in R^2$, let $\hat{x}$ be the projection of x on the diagonal $\{x\colon x_1 = x_2\}$, i.e., $\hat{x}$ has both coordinates equal to $\frac{1}{2}(x_1 + x_2)$. Define

$$d = d(N) := \inf\{\|x - x_0\|, \qquad x \notin N.\} \tag{3.16}$$

Let $\varepsilon = d\beta/3$ and define a map m from A_ε into N, sending lines of slope one to lines of slope one, satisfying

$$m(y) - \widehat{m(y)} = x_0 - \hat{x}_0 + (y - \hat{y})/\beta \tag{3.17}$$

and

$$F_\varepsilon(y) = F(m(y)), \tag{3.18}$$

with F_ε as in Section 2. It is necessary to demonstrate that this map is well defined and one–one.

To this end, fix $y \in A_\varepsilon$ and let $m_0 = x_0 + (y - \hat{y})/\beta$. Since $\|m_0 - x_0\| = \|y - \hat{y}\|/\beta < 2\varepsilon/\beta < d$, we have that $m_0 \in N$. Also, let $m_\mu = m_0 + \mu(1,1)$ for real μ. Now note that by (3.14) $F(m_\mu)$ is *strictly* increasing in μ as long as $m_\mu \in N$. Furthermore, if $\mu = \varepsilon/\beta$, then $m_\mu \geq x_0$ since $(y - \hat{y})_i \leq \varepsilon$ for $y \in A_\varepsilon$. Similarly, $\mu = -\varepsilon/\beta$ implies $m_\mu < x_0$. Consequently,

$$F(m_{-\varepsilon/\beta}) \leq \tfrac{1}{2} \leq F(m_{\varepsilon/\beta}). \tag{3.19}$$

Now consider for what values of μ we shall have $m_\mu \in N$. Clearly, $(m_\mu)_i - (x_0)_i < \varepsilon/\beta + \mu$, $i = 1,2$, since $y \in A_\varepsilon$. Hence $m_\mu \in N$ for $|\mu| < 2\varepsilon/\beta$. Now take $\mu > \mu'$, with $|\mu| \vee |\mu'| < 2\varepsilon/\beta$. Then by (3.14)

$$F(m_\mu) - F(m_{\mu'}) > \beta(\mu - \mu').$$

But since $F(m_{2\varepsilon/\beta}) \geq F(m_{\varepsilon/\beta}) + \beta\varepsilon/\beta \geq \frac{1}{2} + \varepsilon$ and $F(m_{-2\varepsilon/\beta}) \leq \frac{1}{2} - \varepsilon$, it fol-

lows that there is a unique $\mu \in [-2\varepsilon/\beta, 2\varepsilon/\beta]$ such that $F(m_\mu) = F_\varepsilon(y)$ since $y \in A_\varepsilon$ implies $\frac{1}{2} - \varepsilon \le F_\varepsilon(y) \le \frac{1}{2} + \varepsilon$. Let $m(y) = m_\mu$ for this μ. Then clearly (3.18) is satisfied as is (3.17), so that m is well defined for each $y \in A_\varepsilon$. Furthermore, the above argument also establishes that m is one–one. This completes the first part of the argument.

Let $m(A_\varepsilon)$ be the image of A_ε under the mapping m and consider $W_F(x)$ for $x \in m(A_\varepsilon)$. Clearly, for $y \in A_\varepsilon$ (3.18) and (1.8) imply

$$E\{W_F^2(m(y))\} = E\{W_{F_\varepsilon}^2(y)\}. \tag{3.20}$$

Now take $x < y$, $x, y \in A_\varepsilon$. Then $F_\varepsilon(x) < F_\varepsilon(y)$ and so $F(m(x)) < F(m(y))$. Thus (1.8) immediately yields

$$\begin{aligned} E\{W_F(m(x))W_F(m(y))\} &= F(m(x) \wedge m(y)) - F(m(x))F(m(y)) \\ &\le F(m(x)) \wedge F(m(y)) - F(m(x))F(m(y)) \\ &= F_\varepsilon(x) - F_\varepsilon(x)F_\varepsilon(y) \\ &= E\{W_{F_\varepsilon}(x)W_{F_\varepsilon}(y)\}. \end{aligned} \tag{3.21}$$

By symmetry, (3.21) also holds for $x > y$. Now suppose $z = x \wedge y$ is distinct from both x and y. Set $u = m(x)$, $v = m(y)$, and $w = u \wedge v$. Observe, either geometrically or algebraically, that

$$\|x - z\| + \|y - z\| = \|(x - \hat{x}) - (y - \hat{y})\|. \tag{3.22}$$

Now take any x, y in the interior of A_ε. Then since F_ε has uniform marginals it is easy to see that

$$\begin{aligned} F_\varepsilon(x) + F_\varepsilon(y) - 2F_\varepsilon(z) &< (\|x - z\| + \|y - z\|) \\ &= \|(x - \hat{x}) - (y - \hat{y})\|. \end{aligned} \tag{3.23}$$

Now suppose that $v \ge u$. We shall show that this is impossible. Write $t = (u_1, v_2)$. Then by geometry and assumption (3.14)

$$\begin{aligned} F(v) - F(u) &= F(v) - F(t) + F(t) - F(u) \\ &\ge \beta\{\|v - t\| + \|t - u\|\} \\ &\ge \beta\|(v - u) - (\hat{v} - \hat{u})\| \\ &= \|(y - x) - (\hat{y} - \hat{x})\|, \end{aligned}$$

the last line following from (3.17). The above and (3.23) now yield

$$\begin{aligned} 0 < F(v) - F(u) &= F_\varepsilon(y) - F_\varepsilon(x) \\ &< F_\varepsilon(y) + F_\varepsilon(x) - 2F_\varepsilon(z) \\ &< F(v) - F(u), \end{aligned}$$

which is clearly untenable. Thus we cannot have $v \ge u$ nor, by symmetry, $u \ge v$. Consequently, $w = u \wedge v$ is distinct from both u and v. Then, again by geome-

try, assumption, and (3.22), we have

$$\begin{aligned} F(u) + F(v) - 2F(w) &\geq \beta\{\|u - w\| + \|v - w\|\} \\ &= \beta\{\|(u - \hat{u}) - (v - \hat{v})\|\} \\ &= \|(x - \hat{x}) - (y - \hat{y})\| \\ &= \|x - \hat{x}\| + \|y - \hat{y}\| \\ &> F_\varepsilon(x) + F_\varepsilon(y) - 2F_\varepsilon(z). \end{aligned}$$

Thus, since $F(u) = F_\varepsilon(x)$, $F(v) = F_\varepsilon(y)$

$$F_\varepsilon(z) = F(m(z)) > F(w).$$

From this it immediately follows that for *all* $x, y \in A_\varepsilon$,

$$E\{W_{F_\varepsilon}(x)W_{F_\varepsilon}(y)\} \geq E\{W_F(m(x))W_F(m(y))\},$$

with strict inequality if $u \wedge v \neq u$ or v. But this is all we need, for by Slepian's inequality,

$$\begin{aligned} P\Big\{\sup_{I^2} W_F(x) > \lambda\Big\} &\geq P\Big\{\sup_{m(A_\varepsilon)} W_F(x) > \lambda\Big\} \\ &\geq P\Big\{\sup_{A_\varepsilon} W_{F_\varepsilon}(x) > \lambda\Big\}. \end{aligned}$$

The last probability is precisely that given by the RHS of (2.13), which, with $k = 2$, is asymptotically of the form $c\lambda^2 e^{-2\lambda^2}$. This completes the proof of the theorem.

It is interesting to note that there are "$1\frac{1}{2}$-dimensional" d.f.s that yield supremum tail probabilities strictly between the one-dimensional $O(e^{-2\lambda^2})$ and two-dimensional $O(\lambda^2 e^{-2\lambda^2})$. As an example, take H to be the d.f. on I^2 with density

$$h(x, y) = \begin{cases} 2 & (x, y) < \left(\frac{1}{2}, \frac{1}{2}\right) \quad \text{or} \quad (x, y) > \left(\frac{1}{2}, \frac{1}{2}\right), \\ 0 & \text{otherwise.} \end{cases} \tag{3.24}$$

Clearly, H fails to satisfy the conditions of Theorem 3.2. However, it is a relatively easy exercise to estimate the exceedence probabilities of W_H, using the fact that the two processes

$$\begin{aligned} W_1(x, y) &:= \sqrt{2}\left\{W_H(x/2, y/2) - W_H\left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right\}, \\ W_2(x, y) &:= \sqrt{2}\left\{W_H(1 - x/2, 1 - x/2) - W_H\left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right\}, \end{aligned}$$

$(x, y) \in I^2$, are both versions of the pinned Brownian sheet $\mathring{W}^{(2)}$. This fact, together with Theorems 2.1 and 3.2, conditioning on and then integrating out $W_H(\frac{1}{2}, \frac{1}{2})$, readily yields

$$P\Big\{\sup_{I^2} W_H(x) > \lambda\Big\} = O(\lambda e^{-2\lambda^2}), \tag{3.25}$$

thus indicating that noneven powers of λ in tail bounds cannot be excluded. [Indeed, there is no good reason even to exclude noninteger powers, as these do

occur as tail bounds for other classes of Gaussian processes; see, for example, Section 12.2 of Leadbetter, Lindgren, and Rootzen (1983).]

4. Bounds for the general case. Our aim in this section will be to obtain, in $k > 2$ dimensions, bounds of the same general form as those we have just obtained for two dimensions. In particular, if F is a continuous d.f. on I^k with uniform (one-dimensional) marginals, then the two central results are as follows:

THEOREM 4.1. *There exist constants c_k, $k \geq 1$, independent of F and λ, such that for F as above*

$$P\{\sup W_F(x) > \lambda\} \leq c_k \lambda^{2(k-1)} e^{-2\lambda^2}. \tag{4.1}$$

THEOREM 4.2. *Suppose, in addition to the above, there exists an $x^* \in I^k$, a neighbourhood N of x^* and a constant $\underline{\psi} > 0$ satisfying*

$$F(x^*) = \tfrac{1}{2}, \tag{4.2}$$

and throughout N, F possesses continuous first-order partial derivatives $\psi_i := \partial F/\partial x_i$ satisfying

$$\inf_i \inf_N \psi_i(x) = \underline{\psi} > 0. \tag{4.3}$$

Then for each such F there exists a constant $c = c(F)$, independent of λ, such that

$$P\left\{\sup_{I^k} W_F(x) > \lambda\right\} \geq c\lambda^{2(k-1)} e^{-2\lambda^2}. \tag{4.4}$$

Both of these results, while clearly indicating the correct order of magnitude behaviour of the tail of $\sup W_F$, are considerably weaker than their two-dimensional counterparts, since the style of their proofs is such that it is impossible to closely monitor inequalities so as to estimate the constants of the bounds. Consequently, the statistical value of Theorems 4.1 and 4.2 is somewhat limited. Nevertheless, they have interesting probabilistic consequences, as we shall see in Section 5, as well as being of intrinsic interest for the reasons mentioned in the introduction.

We shall prove Theorem 4.1 first, by a method totally different from that used for the two-dimensional upper bound. There, recall, the argument was based on finding a "worst possible F." In dimensions three and above there seems to be no analogous unique worst F, and the proof is forced to take a different route. We start with some necessary lemmas, for which we define the following event for $x_1, x_2 \in I^k$, $x_1 \leq x_2$, and $\lambda > 0$

$$A = A(x_1, x_2, \lambda) := \{\sup(W_F(x): x_1 \leq x \leq x_2) > \lambda\}. \tag{4.5}$$

Also, write

$$\sigma^2(x) := \operatorname{var}(W_F(x)) = F(x)[1 - F(x)]. \tag{4.6}$$

LEMMA 4.1. *Take* $\frac{1}{4} < \alpha < \frac{1}{2}$, $x_1, x_2 \in I^k$, $x_1 \le x_2$, *and* $\lambda > 1$. *If*

$$\alpha \le F(x_1) \le F(x_2) \le 1 - \alpha \tag{4.7}$$

and

$$F(x_2) - F(x_1) \le \tfrac{1}{2}\alpha^2\lambda^{-2}, \tag{4.8}$$

then

$$P\{A\} \le O\big(\lambda^{-1}\exp(-\lambda^2/2\sigma^2(x_1))\big), \tag{4.9}$$

where, for any function $f\colon \mathbb{R} \to \mathbb{R}$,

$$f(a) \le O(a) \Leftrightarrow \limsup_{a\to\infty}(f(a)/a) \le K < \infty.$$

PROOF. Since it is generally difficult to work with the maxima of the pinned sheet W_F, the main idea of the proof is to relate W_F to its unpinned version Z_F, where Z_F is the zero mean Gaussian field on I^k satisfying

$$E\{Z_F(x)\cdot Z_F(y)\} = F(x \wedge y).$$

Then $Z_F(x) - F(x)Z_F(1)$ is a version of W_F, so that using this version in all that follows, we can write

$$W_F(x) = V(x) - [F(x) - F(x_1)]Z(1), \qquad x \in [x_1, x_2], \tag{4.10}$$

where

$$V(x) := W_F(x_1) + [Z_F(x) - Z_F(x_1)], \qquad x \in [x_1, x_2]. \tag{4.11}$$

The idea of the proof is that for λ large (4.8) implies the second term in (4.10) will be small, while $V(x)$ will be close to $W_F(x_1)$.

Note firstly, by direct calculation of covariances, that $W_F(x)$ and $Z_F(1)$ are independent, so that with A at (4.5)

$$P\{A\} = 2P\{A \text{ and } Z_F(1) \ge 0\}. \tag{4.12}$$

Thus, by (4.10)

$$P\{A\} \le 2P\Big\{Z_F(1) \ge 0 \text{ and } \sup_{[x_1,x_2]} V(x) > \lambda\Big\} \tag{4.13}$$

$$\le 2P\Big\{\sup_{[x_1,x_2]} V(x) > \lambda\Big\}.$$

To bound the last probability, write $V(x) = W_F(x_1) + U(x)$, where

$$U(x) := Z_F(x) - Z_F(x_1)$$

is independent of $W_F(x_1)$. Suppose we can show the existence of a finite $c > 0$ such that for all $\eta > 0$

$$P\Big\{\sup_{[x_1,x_2]} U(x) > \eta\Big\} \le c(1 + \eta\lambda/\alpha)^{-1}e^{-\eta^2\lambda^2/\alpha^2}. \tag{4.14}$$

Then, allowing c to vary from line to line and setting $\sigma^2 = \sigma^2(x_1)$ for notational

convenience, we have

$$P\{A\} \le 2P\{W(x_1) \ge \lambda\} + \int_{-\infty}^{\lambda} P\{\sup U(x) \ge \lambda - w\}\, dP\{W(x_1) \le w\}$$

$$\le c\lambda^{-1}\exp(-\lambda^2/2\sigma^2)$$

$$+ c\int_{-\infty}^{\lambda} (1 + \lambda - w)^{-1}\exp\left\{-\frac{\lambda^2(\lambda - w)^2}{\alpha^2} - \frac{w^2}{2\sigma^2}\right\} dw,$$

on using standard inequalities for the first probability and (4.14) for the integrand, after noting $\lambda > 1$ and $\alpha < \frac{1}{2}$. Consider the integral, which we bound by

$$\int_0^{\infty} \exp\left\{-\frac{\lambda^2 y^2}{\alpha^2} - \frac{(\lambda - y)^2}{2\sigma^2}\right\} dy.$$

Set $\beta = (\lambda^2/\alpha^2) + (1/2\sigma^2)$ and rewrite the exponent here as

$$-\frac{\lambda^2}{2\sigma^2} - \beta y^2 + \frac{2y\lambda}{2\sigma^2} - \frac{(\lambda/2\sigma^2)^2}{\beta} + \frac{(\lambda/2\sigma^2)^2}{\beta}$$

$$\le \frac{-\lambda^2}{2\sigma^2} + k(\alpha, \sigma^2) - \beta\left[y - \frac{\lambda}{2\sigma^2\beta}\right]^2,$$

where $k(\alpha, \sigma^2) = \sup_\lambda[(\lambda/2\sigma^2)^2/\beta] < \infty$ depends on α and σ^2 but not on λ. Consequently, the above integral is bounded by

$$e^{-\lambda^2/2\sigma^2} \cdot e^k \cdot \int_0^{\infty} e^{-\beta[y - \lambda/2\sigma^2\beta]}\, dy \le c \cdot e^{-\lambda^2/2\sigma^2}\beta^{-1/2}$$

$$\le c\lambda^{-1}e^{-\lambda^2/2\sigma^2}.$$

Combining this with the above, we find

$$P\{A\} \le c\lambda^{-1}e^{-\lambda^2/2\sigma^2},$$

which thus establishes the lemma, subject to (4.14). It remains to establish (4.14).

A straightforward application of the multivariate "reflection principle" yields

$$P\left\{\sup_{[x_1, x_2]} U(x) > \lambda\right\} \le 2^k P\{U(x_2) > \lambda\}.$$

By (4.8) $\operatorname{var} U(x_2) \le \frac{1}{2}\alpha^2\lambda^{-2}$, so that (4.14) now follows by standard inequalities.

Without much extra work we can also prove a stronger version of the preceding lemma. Under the conditions of the lemma, we have, for $x_1 \le x \le x_2$ that

$$\sigma^2(x) \ge \sigma^2 - \alpha^2/(4\lambda^2).$$

Consequently,

$$\frac{\lambda^2}{2\sigma^2(x)} - \frac{\lambda^2}{2\sigma^2} \le \frac{\lambda^2}{2[\sigma^2 - \alpha^2/4\lambda^2]} - \frac{\lambda^2}{2\sigma^2}$$

$$= \frac{\alpha^2/4}{2\sigma^2[\sigma^2 - \alpha^2/4\lambda^2]} = O(1).$$

Thus Lemma 4.1 immediately yields

LEMMA 4.2. *Under the conditions and notation of Lemma* 4.1

$$P\{A\} \le O\big(\lambda^{-1}\exp(-\lambda^2/2\hat{\sigma}^2)\big), \tag{4.15}$$

where

$$\hat{\sigma}^2 = \inf\{\sigma^2(x)\colon x_1 \le x \le x_2\}.$$

To state the next lemma define the event

$$B = \{\sup[W_F(x)\colon F(x) \le \alpha \text{ or } F(x) \ge (1-\alpha) > \lambda]\}.$$

LEMMA 4.3. *Let* $\alpha \in (0, \frac{1}{2})$ *and* $\beta \in (1, (4\alpha(1-\alpha))^{-1})$. *Then*

$$P\{B\} \le O\big(\exp(-2\beta\lambda^2)\big). \tag{4.16}$$

PROOF. This is a straightforward application of (1.4), on noting that $F(x) < \alpha$ and $F(x) > 1 - \alpha$ both imply $(2\sigma^2(x))^{-1} \le 2[4\alpha(1-\alpha)]^{-1}$. We now turn to the

PROOF OF THEOREM 4.1. The idea of the proof is as follows. Divide I^k into a large number of small cubes and separate these cubes into two groups. In the first group put those cubes over which W_F has small variance and use Lemma 4.3 to show that the maximum of W_F over this group is asymptotically unimportant. For the second group use Lemma 4.2 to bound the (distribution of the) maximum of W_F over each cube, then sum these bounds, and thus obtain a final bound.

We now spell out the proof in detail and note that the only real difficulty lies in finding a convenient labelling system for the various cubes. We commence with cubes over which W_F has large variance (i.e., close to $\frac{1}{2}$). Fix the dimension k, choose $\alpha \in (\frac{1}{4}, \frac{1}{2})$, $\lambda > 1$, and set $\gamma = \alpha^2/(2k\lambda^2)$. Let $\mathbf{1} = (1, \dots, 1)$, and let $x \in I^k$ be such that also $x + \gamma\mathbf{1} \in I^k$. Then the uniformity of the marginals of F implies

$$F(x + \gamma\mathbf{1}) \le F(x) + \gamma k = F(x) + \alpha^2/(2\lambda^2). \tag{4.17}$$

Now consider the lattice of points of the form $\gamma(n_1, \dots, n_k)$, where $n_i = 0, 1, \dots, [\gamma^{-1}]$. Then each of these points has a unique expression as $p + j\gamma\mathbf{1}$, where $p \in \pi$ and π is the set of $\gamma(n_1, \dots, n_k)$ with $\min\{n_i\colon 1 \le i \le k\} = 0$. For each $p \in \pi$ define, inductively,

$$j_1 = j_1(p) = \max\{j\colon F(p + j\gamma\mathbf{1}) < \alpha\},$$

$$j_i = j_i(p) = \max\{j\colon F(p + j\gamma\mathbf{1}) - F(p + j_{i-1}\gamma\mathbf{1}) \le \alpha^2/(2\lambda^2)\}.$$

Furthermore, define

$$J = J(p) = \min\{i\colon F(p + j_i\gamma\mathbf{1}) > 1 - \alpha\}.$$

Note that (4.17) implies $j_i - j_{i-1} \ge 1$ for all i and p. Also, for $1 \le j_i \le J$,

$$0 < \alpha - \alpha/(2\lambda^2) \le F(p + j_i\gamma\mathbf{1}) \le 1 - \alpha + \alpha^2/(2\lambda^2) < 1.$$

Now set $j^*(p) = j_{J(p)} - 1$, and define

$$S(p) = \bigcup_{k=0}^{j^*(p)} \{x\colon p + (j_1 + k)\gamma\mathbf{1} \le x \le p + (j_1 + k + 1)\gamma\mathbf{1}\},$$

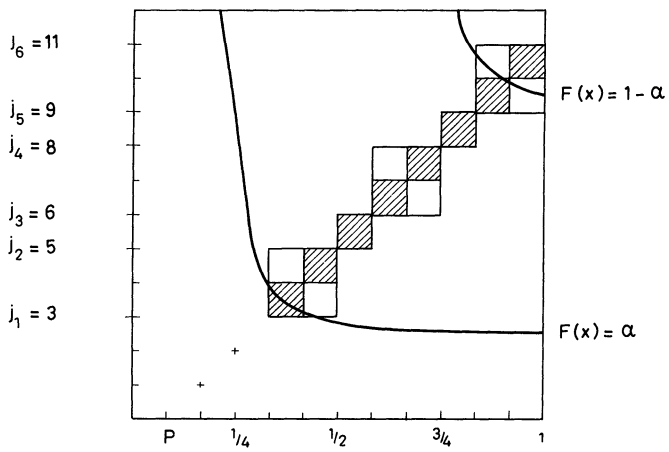

FIG. 2. *An example of the sets S(p) and S'(p) with* $\gamma = 12$, $p = (\frac{1}{12}, 0)$. *The set* π *is given by the marked points on the axes. S(p) = shaded squares, S'(p) = all squares, J(p) = 6.*

and

$$S^*(p) = \bigcup_{i=1}^{J(p)-1} \{x\colon p + j_i\gamma\mathbf{1} \le x \le p + j_{i+1}\gamma\mathbf{1}\}.$$

(An example for $k = 2$ is depicted in Figure 2.)

From the definitions of S and S^* it is clear that

$$\{x\colon \alpha \le F(x) \le 1 - \alpha\} \subseteq \bigcup_{p\in\pi} S(p) \subseteq \bigcup_{p\in\pi} S^*(p).$$

Thus, with A as at (4.5)

$$\begin{aligned} &P\{\sup(W_F(x)\colon \alpha \le F(x) \le 1 - \alpha) > \lambda\} \\ &\quad \le P\{\sup(W_F(x)\colon x \in S^*(p) \text{ for some } p \in \pi) > \lambda\} \\ &\quad \le \sum_{p\in\pi} \sum_{i=1}^{J(p)-1} P\{A(p + j_i\gamma\mathbf{1}, p + j_{i+1}\gamma\mathbf{1}, \lambda)\} \qquad (4.18) \\ &\quad \le \sum_{p\in\pi} O\left(\sum_{i=1}^{J(p)-1} \lambda^{-1}\exp[-\lambda^2/2\sigma^2(p + j_i\gamma\mathbf{1})] \right), \end{aligned}$$

the last inequality following from Lemma 4.2. Now note that for all i, p,

$$F(p + j_{i+2}\gamma\mathbf{1}) - F(p + j_i\gamma\mathbf{1}) > \alpha^2/2\lambda^2,$$

and set $I = \min\{i\colon \frac{1}{2} - i\alpha^2/2\lambda^2 < \alpha\} = \min\{i\colon \frac{1}{2} + i\alpha^2/2\lambda^2 > 1 - \alpha\}$. Then it follows that the sequence

$$\{\sigma^2(p + j_i\gamma\mathbf{1})\colon i = 1, \ldots, J(p)\} \qquad (4.19)$$

is dominated by the sequence

$$\{a_0, a_0, a_0, a_0, a_1, \ldots, a_i, a_i, a_i, a_i, \ldots, a_I, a_I, a_I, a_I\}, \tag{4.20}$$

in which $a_i = \frac{1}{4} - (i\alpha^2/2\lambda^2)^2$, where by "domination" we mean that the elements of (4.20) may be rearranged so that, termwise, they dominate corresponding elements of (4.19). Furthermore, there may also be more terms in (4.20) than in (4.19). As a consequence of this we have that

$$\begin{aligned}
\sum_{i=1}^{J(p)-1} \lambda^{-1}\exp\left[-\lambda^2/2\sigma^2(p + j_i\gamma\mathbf{1})\right] &\le 4\sum_{i=1}^{I} \lambda^{-1}\exp\left[-\lambda^2/2\left(\tfrac{1}{4} - \left(i\alpha^2/2\lambda^2\right)^2\right)\right] \\
&\le 4\sum_{i=1}^{I} \lambda^{-1}\exp\left[-2\lambda^2\left(1 + \left(i\alpha^2/\lambda^2\right)^2\right)\right] \\
&= 4e^{-2\lambda^2}\sum_{i=1}^{I} \lambda^{-1}\exp\left[-2\alpha^4 i^2/\lambda^2\right] \\
&= e^{-2\lambda^2}O\left(\int_0^{\infty} e^{-2\alpha^4 y}\,dy\right) \\
&= O(e^{-2\lambda^2}).
\end{aligned}$$

Note that π has at most $(2 + 2k\lambda^2/\alpha^2)^{k-1} = O(\lambda^{2(k-1)})$ points. Combining this fact, the above, (4.18), and Lemma 4.3 yields

$$\begin{aligned}
P\left\{\sup_{I^k} W_F(x) > \lambda\right\} &\le P\{\sup(W_F(x)\colon \alpha \le F(x) \le 1 - \alpha) > \lambda\} \\
&\quad + P\{\sup(W_F(x)\colon F(x) < \alpha \text{ or } F(x) > 1 - \alpha) > \lambda\} \\
&= O(\lambda^{2(k-1)}e^{-2\lambda^2}) + O(e^{-2\beta\lambda^2}) \\
&= O(\lambda^{2(k-1)}e^{-2\lambda^2}).
\end{aligned}$$

This completes the proof of Theorem 4.1.

PROOF OF THEOREM 4.2. Our aim here will be to attempt to mimick the proof of the two-dimensional case (Theorem 3.2), by comparing W_F to W_{F_ε}. However, for $k > 2$ dimensions, the mapping on which the comparison is based is definable as a linear mapping only in an arbitrarily small neighbourhood of the point x^* of the theorem, and we shall not be able to say anything concrete about the size of the neighbourhood, and thus, *a fortiori*, anything nonasymptotic about the lower bound that we shall obtain.

The first part of the proof carefully sets up some geometrical structures and is totally nonprobabilistic. Probability will enter only when the groundwork is ready.

Let G be the uniform distribution on I^k and y^* the point $(\frac{1}{2})^{1/k}\mathbf{1}$. Then $G(y^*) = \frac{1}{2}$ and

$$\gamma := \frac{\partial G}{\partial y_i}(y^*) = \left(\frac{1}{2}\right)^{(k-1)/k}. \tag{4.21}$$

Note that γ is independent of i. In order to compare F and G, it is convenient to consider new coordinate systems for F and G, obtained by rotation and translation. To this end, let $\psi = \langle \psi_1, \ldots, \psi_k \rangle$ and write γ interchangeably for the constant (4.21) and the constant vector $\gamma \mathbf{1}$. Define unit vectors, with $\| \cdot \|$ now denoting the usual Euclidean norm,

$$V_1 = \psi / \|\psi\|, \qquad W_1 = \gamma / \|\gamma\|,$$

and extend to two orthonormal bases $\mathbf{V} := \{V_1, \ldots, V_k\}$ and $\mathbf{W} := \{W_1, \ldots, W_k\}$ for $\mathbb{R}^k$. Choose the origins of the new spaces to be x^* and y^*, respectively. Then if $v(x)$ and $w(y)$ are, respectively, the representations of x and y in the new coordinate systems, we have

$$v(x^*) = 0, \qquad v_1(x) = \psi'(x - x^*)/\|\psi\|,$$
$$w(y^*) = 0, \qquad w_1(y) = \mathbf{1}'(y - y^*)/\sqrt{k} = \gamma'(y - y^*)/\|\gamma\|.$$

The d.f.s F and G can be transferred in a natural fashion to $\mathbf{V}$ and $\mathbf{W}$ space, respectively. Let $\tilde{F}$ and $\tilde{G}$ be the corresponding functions defined by

$$\tilde{F}(v(x)) = F(x), \qquad \tilde{G}(w(y)) = G(y).$$

(Note that $\tilde{F}$ and $\tilde{G}$ are not necessarily d.f.s on $\mathbf{V}$ and $\mathbf{W}$ space.) Now define maps π^v and π^w from $\mathbf{V}$ and $\mathbf{W}$ space, respectively, to the original domains of F and G by

$$\pi^v(v(x)) = x - x^*, \qquad \pi^w(w(y)) = y - y^*.$$

Thus π^v and π^w transform from the coordinate systems of the $\mathbf{V}$ and $\mathbf{W}$ spaces to systems centered at x^* and y^* but oriented like the original Cartesian system.

We shall need to impose on the $\mathbf{V}$ and $\mathbf{W}$ spaces concepts of ordering inherited from the original spaces. To this end, write

$$v^{(1)} \ll v^{(2)} \Leftrightarrow \pi_i^v(v^{(1)}) \leq \pi_i^v(v^{(2)}), \qquad i = 1, \ldots, k,$$
$$w^{(1)} \ll w^{(2)} \Leftrightarrow \pi_i^w(w^{(1)}) \leq \pi_i^w(w^{(2)}), \qquad i = 1, \ldots, k,$$

and define $v^{(1)} \wedge v^{(2)}$ and $w^{(1)} \wedge w^{(2)}$ accordingly.

This completes the necessary geometrical groundwork. We now build the mapping upon which the comparison between F and G will be based. Let

$$\rho = \frac{\|\gamma\|}{\|\psi\|} = \frac{\sqrt{k}\,\gamma}{\|\psi\|}, \qquad \alpha = \frac{k \max\{\psi_i : i = 1, \ldots, k\}}{\min\{\psi_i : i = 1, \ldots, k\}}, \qquad \beta = \rho + 2\alpha.$$

Define the mapping $m = (m_1, \ldots, m_k)$ from a neighbourhood of zero in $\mathbf{W}$ space to a neighbourhood of zero in $\mathbf{V}$ space, via its coordinate mappings, by first setting

$$m_i(w) = \beta w_i, \qquad i = 2, \ldots, k \tag{4.22}$$

and then choosing $m_1(w)$ such that

$$\tilde{F}(m(w)) = \tilde{G}(w). \tag{4.23}$$

We need to check that m_1 is, in fact, well defined. For $w_i = 0$, $i = 2, \ldots, k$, and general w_1, $\tilde{G}(w)$ is clearly strictly increasing as a function of w_1. Further-

more, since the unit vector V_1 of the V space has, as a vector in the original space, strictly positive coordinates, it follows that $\tilde{F}(v_1, 0, \ldots, 0)$ is strictly increasing as a function of v_1. Since $\tilde{F}(0) = \tilde{G}(0) = \frac{1}{2}$, it follows that m is well defined for w of the form $(w_1, 0, \ldots, 0)$. The implicit function theorem now defines m uniquely for sufficiently small neighbourhoods.

Having defined our mapping, let us consider some of its properties. Note that for small neighbourhoods of the origin

$$\tilde{G}(w) = \tfrac{1}{2} + \|\gamma\| w_1 + o(\|w\|),$$
$$\tilde{F}(v) = \tfrac{1}{2} + \|\psi\| v_1 + o(\|v\|).$$

Combining these facts with (4.22), (4.23), and the definition of β we obtain that for small w

$$m_1(w) = \rho w_1 + o(\|w\|). \tag{4.24}$$

Consequently, for small w^i

$$\left|\left[m_1(w^1) - m_1(w^2)\right] - \left[\rho w_1^1 - \rho w_1^2\right]\right| \leq \|w^1 - w^2\|. \tag{4.25}$$

Now let q be the linear map approximating m, i.e., set

$$q_1(w) = \rho w_1, \qquad q_i(w) = \beta w_i, \qquad i = 2, \ldots, k. \tag{4.26}$$

Then by (4.25) for small w^i

$$\left|\left[m(w^1) - m(w^2)\right] - \left[q(w^1) - q(w^2)\right]\right| \leq \|w^1 - w^2\|. \tag{4.27}$$

Finally, note that as a consequence of (4.27) we also have

$$\pi_i^v\big(m(w^1) - m(w^2) - \big(q(w^1) - q(w^2)\big)\big) \leq \|w^1 - w^2\|. \tag{4.28}$$

This completes our listing of properties of m and its linear approximation. We can now turn to the final part of the proof, the comparison of W_F and W_G, which we commence by comparing $\tilde{F}$ and $\tilde{G}$.

Firstly, let N be a small enough neighbourhood of zero in W space so that (4.24)–(4.28) are true for $w^i \in N$. Take $w^1, w^2 \in N$ with $w^1 \wedge\!\!\wedge w^2 \in N$. Suppose

$$w^1 \wedge\!\!\wedge w^2 = w^p \quad \text{for } p = 1 \text{ or } 2. \tag{4.29}$$

Then

$$\begin{aligned} \tilde{F}\big(m(w^1) \wedge\!\!\wedge m(w^2)\big) &= \tilde{F}\big(m(w^1)\big) \wedge \tilde{F}\big(m(w^2)\big) \\ &= \tilde{G}(w^1) \wedge\!\!\wedge \tilde{G}(w^2) \\ &= \tilde{G}(w^p) \\ &= \tilde{G}(w^1 \wedge\!\!\wedge w^2). \end{aligned} \tag{4.30}$$

Now consider the case $w^3 = w^1 \wedge\!\!\wedge w^2 \neq w^p$ for either $p = 1$ or 2. We shall obtain (4.30) also for this case, but with inequality replacing the equality. For each coordinate $j = 1, \ldots, k$, $w^3 = w^1 \wedge\!\!\wedge w^2$ implies that

$$\pi_j^w(w^1) - \pi_j^w(w^3) = 0 \quad \text{or} \quad \pi_j^w(w^2) - \pi_j^w(w^3) = 0.$$

Fix j and let $p := p(j) = 1$ or 2 be such that

$$\pi_j^w(w^p) - \pi_j^w(w^3) = 0. \tag{4.31}$$

Thus $w^3 = \langle w_1^{p(1)}, \ldots, w_k^{p(k)} \rangle$. Rewrite q as

$$q(w) = \beta w + (\rho - \beta) w_1 V_1,$$

and apply this to (4.31) with $w = w^p - w^3$ to obtain, via the linearity of all the operators,

$$\pi_j^v\big(q(w^p - w^3)\big) = -2\alpha\big(w_1^p - w_1^3\big)\pi_j^v(V_1). \tag{4.32}$$

Now note, from the definition of w_1 and since $w^p \gg w^3$,

$$w_1^p - w_1^3 = \sum_{i=1}^{k} \pi_i^v\big(w_1^p - w_1^3\big)/\sqrt{k} \geq \|w^p - w^3\|/\sqrt{k}\,.$$

Furthermore, it follows from the definition of V_1 and α that

$$\pi_j^v(V_1) \geq (\sqrt{k}\,\alpha)^{-1}.$$

Substituting the above two inequalities into (4.32) yields

$$\pi_j^v\big(q(w^p - w^3)\big) \leq -2\|w^p - w^3\|.$$

Combining this with (4.28) thus yields

$$\pi_j^v\big(m(w^p) - m(w^3)\big) \leq -\|w^p - w^3\| < 0. \tag{4.33}$$

However, what we have just shown is that for *every* $j = 1, \ldots, k$ there is a $p = p(j)$ satisfying (4.33). Consequently, for every $w^1, w^2 \in N$ with $w^1 \wedge w^2 \in N$, it follows that

$$m(w^1 \wedge w^2) \gg \big(m(w^1) \wedge m(w^2)\big),$$

from which it follows that

$$\begin{aligned} \tilde{F}\big(m(w') \wedge m(w^2)\big) &\leq \tilde{F}\big(m(w^1 \wedge w^2)\big) \\ &= \tilde{G}(w^1 \wedge w^2). \end{aligned} \tag{4.34}$$

Combining this with (4.30), we find that the above inequality holds for *all* $w^1, w^2 \in N'$, where $N' \subset N$ is a neighbourhood of zero such that $w^1, w^2 \in N'$ implies $w^1 \wedge w^2 \in N$ and, consequently, that (4.34) holds.

To obtain the final comparison between F and G, we need to return to the original coordinate system. However, this is now easy. Since the "minimum" relationship in (4.34) is really that of the original coordinate system, it trivially follows that via m we have constructed a map, say m^*, from some neighbourhood N^* of y^* to a neighbourhood $m^*(N^*)$ of x^* satisfying

$$F(m^*(y)) = G(y), \qquad y \in N^*,$$

$$F\big(m^*(y^1) \wedge m^*(y^2)\big) \leq G(y^1 \wedge y^2), \qquad y^1, y^2 \in N^*.$$

Slepian's inequality now yields

$$P\Big\{\sup_{I^k} W_F(x) > \lambda\Big\} \geq P\Big\{\sup_{m(N^*)} W_F(x) > \lambda\Big\}$$

$$\geq P\Big\{\sup_{N^*} W_G(y) > \lambda\Big\} \tag{4.35}$$

$$\geq P\Big\{\sup_{B_\delta} W_G(y) > \lambda\Big\},$$

where δ is chosen small enough so that $B_\delta := [(\frac{1}{2})^{1/k} - \delta, (\frac{1}{2})^{1/k} + \delta]^k \subset N^*$. (Note that δ depends on N^* and so on F.) Thus, to complete the proof, we need only find a lower bound for the last probability.

To do this, we shall once again compare the process of interest to W_{F_ε} on the small square A_ε (c.f. Section 2). Choosing δ small enough so that

$$\delta < [1/k] \wedge \Big[(\tfrac{1}{2})^{1/k} - (\tfrac{1}{3})^{1/k}\Big], \tag{4.36}$$

set $\varepsilon = \delta/9$ and define a mapping m from A_ε into I^k by

$$m(x) = 3\big(x - (\tfrac{1}{2})\mathbf{1}\big) + (\tfrac{1}{2})^{1/k}\mathbf{1} + \rho(x)\mathbf{1}, \tag{4.37}$$

where $\rho(x)$ is the unique value for which $G(m(x)) = F_\varepsilon(x)$.

As was the case for our previous mappings, m maps lines of slope one into lines of slope one, and it is not hard to check that (4.36) ensures that $m(x) \in B_\delta$ if $x \in A_\varepsilon$. Furthermore, one can check that for any $x, y \in A_\varepsilon$

$$F_\varepsilon(x \wedge y) \geq G(m(x) \wedge m(y)). \tag{4.38}$$

(We leave the details of the proof to the reader, who by now should either be adept enough at these special case arguments to prove (4.38) by himself, or bored enough that he would not read the proof anyway.) Consequently, Slepian's inequality can, once again, be applied and we find

$$P\Big\{\sup_{B_\delta} W_G(x) > \lambda\Big\} > P\Big\{\sup_{A_\varepsilon} W_{F_\varepsilon}(x) > \lambda\Big\}.$$

But this last probability is known and is bounded from below in Theorem 2.2 by $c\lambda^{2(k-1)}e^{-2\lambda^2}$. This completes the proof of the theorem.

5. An upper–lower class theorem. We now return to the one-sided KS statistic $T_n^{(k)}$ of the introduction and study the way it grows with n. In a fundamental paper treating the one-dimensional case, Chung (1949) proved the following result for a sequence $\lambda(n) \uparrow \infty$:

$$P\{T_n^{(1)} > \lambda(n) \text{ infinitely often}\} = 0 \text{ (or } = 1) \tag{5.1}$$

if

$$\sum_n \frac{\lambda^2(n)}{n} e^{2\lambda^2(n)} < \infty \text{ (or } = \infty). \tag{5.2}$$

Kiefer (1961) obtained a weaker version of Chung's result for the multivariate

case and proved the following LIL for every k and continuous F:

$$
\begin{aligned}
1 &= P\left\{ \limsup_{n\to\infty} \frac{T_n^{(k)}}{\left(\frac{1}{2}\log\log n\right)^{1/2}} = 1 \right\} \\
&= P\left\{ \liminf_{n\to\infty} \frac{T_n^{(k)}}{\left(\frac{1}{2}\log\log n\right)^{1/2}} = -1 \right\}.
\end{aligned} \tag{5.3}
$$

Kiefer's proof of (5.3) was based on inequality (1.3), which is not fine enough to pick up the higher iterated logarithm terms that (5.2) yields. Having improved on Kiefer's inequality in the previous sections (at least insofar as the limit process W_F is concerned), we can now complete the task Kiefer began and obtain a multidimensional analogue of (5.1).

Unlike Chung and Kiefer's basic inequalities for $P\{T_n > \lambda\}$, we have only inequalities for $P\{\sup_x W_F(x) > \lambda\}$, and so we shall need to proceed via an embedding into the Kiefer process. To define this process write $B \equiv C[0,1]^k$ for the Banach space of continuous functions on $[0,1]^k$. Let B^* be its topological dual, $\mathscr{B}$ its Borel σ algebra, and μ the measure induced on $\mathscr{B}$ by the process W_F. It is well known that B^* can be embedded in $\mathscr{L}^2(B, \mathscr{B}, \mu)$ by a one–one continuous and weakly continuous mapping. Let H denote the closure of the range of this mapping. Then H is a Hilbert space with inner product $(\cdot,\cdot)_H$.

The Kiefer process is then defined as the B-valued, real parameter Gaussian process K_t, $t \geq 0$, with covariance kernel

$$E(\langle f, K_s\rangle\langle g, K_t\rangle) = (s \wedge t)\cdot(f,g)_H, \qquad f, g \in B^*.$$

As a consequence of this definition, we have the following properties:

$$P\{K_1 \in A\} = P\{W_F \in A\}, \tag{5.4}$$

$$
\begin{aligned}
P\{(K_t - K_s) \in A\} &= P\{K_{t-s} \in A\} \\
&= P\left\{\sqrt{t-s}\,K_1 \in A\right\} \quad \text{for } t > s,
\end{aligned} \tag{5.5}
$$

$$(K_t - K_s) \quad \text{and} \quad K_u \text{ are independent for all } t > s \geq u. \tag{5.6}$$

Here A is any Borel subset of $C[0,1]^k$ with topology generated by the sup norm $\|k\| = \sup\{|k(t)|\colon\ t \in I^k\}$. Then Theorem 7.1 of Dudley and Philipp (1983) implies the following embedding theorem, which is a strengthening of an earlier result of Kiefer (1972).

THEOREM 5.1 (Dudley–Phillip). *Let $X_1, X_2, \ldots$ be an infinite sequence of i.i.d. r.v.s, with common d.f. F. Let F_n be the empirical d.f. based on $X_1, \ldots, X_n$. Then, enlarging the probability space if necessary, for every $\theta > 0$ there exists a Kiefer process K_t, $t \geq 0$, such that*

$$\sup_x \left|n[F_n(x) - F(x)] - K_n(x)\right| \leq o\left(n^{1/2}(\log n)^{-\theta}\right) \tag{5.7}$$

with probability one. Consequently, we also have with probability one

$$\left|T_n - n^{-1/2}\sup_x K_n(x)\right| \leq o\left((\log n)^{-\theta}\right).$$

As an immediate consequence of this result, along with a LIL for sums of Banach space random variables, it is now easy to obtain Kiefer's LIL (5.3) [c.f. Kuelbs and Philipp (1980) and Goodman, Kuelbs, and Zinn (1981), especially Theorem 6.1]. Indeed, the Banach space results yield much more than (5.3), for they also identify the cluster points of F_n in $C[0,1]^k$ in terms of the unit ball of a certain Hilbert space. It is not possible, however, to follow this path to obtain a multivariate version of Chung's upper–lower class theorem, the problem being that no appropriate upper–lower class theorem is known for K_t. [Note that whereas Kuelbs (1975) does have a result of this type for K_t, it is not applicable here, since it gives results not for the growth of $\|K_t\|$ but the growth of $\|K_t\|_*$, where $\|\cdot\|_*$ is another unspecified norm (albeit equivalent to the sup norm).] Consequently, we shall have, in essence, to prove such a result in order to obtain a generalization of Chung's theorem.

To state our result, we shall say a nonnegative, nondecreasing, continuous function $\psi(t)$ defined for large values of t is a *lower function* for $\{K_t,\ t \geq 0\}$ if

$$P\{\|K_n\| > \sqrt{n}\,\psi(n) \text{ for an unbounded set of } n\} = 1, \tag{5.8}$$

and an *upper function* for $\{K_t\colon\ t \geq 0\}$ if

$$P\{\|K_t\| > t^{1/2}\psi(t) \text{ for only a bounded set of } t\} = 1. \tag{5.9}$$

Since the definition of K_t is dependent on F, whether or not any given ψ is a lower or upper function depends on F as well as ψ. Thus we write $\psi \in \mathsf{L}(F)$ and $\psi \in \mathsf{U}(F)$, respectively, to denote this dependency. Note that (5.8) implies the weaker condition,

$$P\{\|K_t\| > t^{1/2}\psi(t) \text{ for an unbounded set of } t\} = 1,$$

which is usually taken as the definition of a lower function. However, the stronger result (5.8) is what is needed to apply Theorem (5.1), and since our proof will be strong enough to prove (5.8), we use it to define the notion of lower class. We can now state

THEOREM 5.2. *Let F be continuous on I^k with uniform marginals. For ψ as above, set*

$$I_k(\psi) = \int^{\infty} \frac{\psi^{2k}(t)}{t} e^{-2\psi^2(t)}\,dt. \tag{5.10}$$

If $I_k(\psi) < \infty$, then $\psi \in \mathsf{U}(F)$. Furthermore, if F satisfies the conditions of Theorem 4.2, and $I_k(\psi) = \infty$, then $\psi \in \mathsf{L}(F)$.

A simple argument, dating back at least to Erdös (1942) and spelled out in detail in Sirao (1959), shows that there is no loss of generality in Theorem 5.2 in assuming that for large t

$$\left(\left(\tfrac{1}{4}\right)\log\log t\right)^{1/2} \leq \psi(t) \leq (\log\log t)^{1/2}. \tag{5.11}$$

Furthermore, a straightforward application of the Abel–Dini theorem easily yields the following corollary.

COROLLARY 5.1. *Let $p \geq 3$ be integral and define*

$$\psi_{k,\delta}(t) := 2^{-1/2}\left[\log_2 t + (k+1)\log_3 t + \log_4 t + \cdots + (1+\delta)\log_{p+1}(t)\right]^{1/2}.$$

Then $\delta > 0$ implies $I_k(\psi_{k,\delta}) < \infty$ and $\delta < 0$ implies $I_k(\psi_{k,\delta}) = \infty$, so that $\psi_{k,\delta} \in \mathsf{U}(F)$ if $\delta > 0$ and $\psi_{k,\delta} \in \mathsf{L}(F)$ if $\delta < 0$ and F satisfies the conditions of Theorem 4.2.

As a further consequence of (5.11) and Theorem 5.1 we can also derive the following corollary of Theorem 5.2, which generalises Chung's univariate test:

COROLLARY 5.2. *For all F*

$$P\{T_n^{(k)} > \psi(n) \text{ i.o.}\} = 0 \quad \text{if } \sum_n \frac{\psi^{2k}(n)}{n} e^{-2\psi^2(n)} < \infty.$$

For F satisfying the conditions of Theorem 4.2

$$P\{T_n^{(k)} > \psi(n) \text{ i.o.}\} = 1 \quad \text{if } \sum_n \frac{\psi^{2k}(n)}{n} e^{-2\psi^2(n)} = \infty.$$

This result, of course, implies Kiefer's LIL (5.3). All that now remains is the

PROOF OF THEOREM 5.2. Clearly, in view of Theorem 5.1 and (5.11), it is sufficient to obtain an upper-lower class result for $\|K_t\|$. The result for T_n^k then follows easily. We consider the convergent case first, i.e., $I_k(\psi) < \infty$. Define a sequence t_n satisfying

$$t_{n+1} = t_n\left(1 + \psi^{-2}(t_n)\right), \tag{5.12}$$

where $t_1 > 3$ is sufficiently large so that (5.11) holds for $t > t_1$, and so $\lim_n t_n = \infty$. Set $I_n = [t_n, t_{n+1}]$ and

$$A_n = \left\{ \sup_{t \in I_n} \frac{\|K_t\|}{t^{1/2}\psi(t)} > 1 \right\}.$$

Then, applying the Banach space version of Lévy's inequality, we have

$$\begin{aligned} P\{A_n\} &\leq P\left\{ \sup_{t \in I_n} \|K_t\| > t_n^{1/2}\psi(t_n) \right\} \\ &\leq 2P\left\{ \|K_{t_{n+1}}\| > t_n^{1/2}\psi(t_n) \right\} \\ &= 2P\left\{ t_{n+1}^{-1/2}\|K_{t_{n+1}}\| > (t_n/t_{n+1})^{1/2}\psi(t_n) \right\}. \end{aligned}$$

Now apply the scaling law (5.5) and Theorem 4.1, letting C be a constant that may vary from line to line, to obtain

$$\begin{aligned} P\{A_n\} &\leq C(t_n/t_{n+1})^{k-1}\left[\psi(t_n)\right]^{2(k-1)}\exp\left[-2\psi^2(t_n)t_n/t_{n+1}\right] \\ &\leq C\left[\psi(t_n)\right]^{2(k-1)}\exp\left[-2\psi^2(t_n)\right], \end{aligned}$$

since $t_n < t_{n+1}$ and $(t_n/t_{n+1}) = (1 + \psi^{-2}(t_n))^{-1} > 1 - \psi^{-2}(t_n) > \frac{1}{2}$ for large enough n.

To complete the proof, it is clearly sufficient to show $\Sigma_1^\infty P\{A_n\}$ converges. But

$$\begin{aligned}\sum_{n=1}^{\infty} P\{A_n\} &\le C\sum_n \int_{t_{n-1}}^{t_n} [\psi(t_n)]^{2(k-1)} e^{-2\psi^2(t_n)} \frac{t_n}{s(t_n - t_{n-1})}\, ds \\ &= C\sum_n \int_{t_{n-1}}^{t_n} [\psi(t_n)]^{2(k-1)} e^{-2\psi^2(t_n)} \frac{t_n \psi^2(t_{n-1})}{s t_{n-1}}\, ds \\ &\le C\int_0^\infty \frac{\psi^{2k}(s)}{s} e^{-2\psi^2(s)}\, ds \\ &= CI_k(\psi),\end{aligned}$$

the last inequality from the definition of t_k and the ultimate monotonicity of $\psi^{2k}(s)e^{-2\psi^2(s)}$. By assumption $I_k(\psi) < \infty$ and so the proof of the convergence part of the theorem is complete.

Now assume $I_k(\psi) = \infty$. Let $\alpha_n = (\log n)^2$,

$$\beta_n = \prod_{i=2}^{n} \alpha_i \quad \text{and} \quad t_n = [\beta_n], \qquad n \ge 2.$$

Also set

$$\varepsilon_n = -1 + \frac{(1 + t_n/t_{n+1})}{(1 - t_n/t_{n+1})^{1/2}} > 0, \qquad n \ge 2.$$

Following Chung's (1949) argument, set

$$H_n = \{\|K_{t_n}\| < t_n^{1/2}\psi(t_n)\},$$

$$H_{n,n+1} = \{\|K_{t_{n+1}} - K_{t_n}\| > (1 + \varepsilon_n)(t_{n+1} - t_n)^{1/2}\psi(t_{n+1})\}.$$

Then, if both H_n and $H_{n,n+1}$ occur, we have

$$\begin{aligned}\|K_{t_{n+1}}\| &\ge (1 + \varepsilon_n)(t_{n+1} - t_n)^{1/2}\psi(t_{n+1}) - t_n^{1/2}\psi(t_n) \\ &= t_{n+1}^{1/2}\psi(t_{n+1})\{(1 - t_n/t_{n+1})^{1/2}(1 + \varepsilon_n) - t_n^{1/2}\psi(t_n)/[t_{n+1}^{1/2}\psi(t_{n+1})]\} \\ &\ge t_{n+1}^{1/2}\psi(t_{n+1}).\end{aligned}$$

That is, $H_n \cdot H_{n,n+1} \Rightarrow \overline{H}_{n+1}$. Thus, noting from (5.6) that $H_{n,n+1}$ and H_m are independent for $m \le n$, it follows that

$$P\left\{\prod_{m=2}^{n} H_m\right\} \cdot P\{H_{n,n+1}\} \le P\left\{\prod_{m=2}^{n} H_m \cdot \overline{H}_{n+1}\right\}.$$

That is

$$\begin{aligned}P\left\{\prod_{m=2}^{n+1} H_m\right\} &\le P\left\{\prod_{m=2}^{n} H_m\right\} \cdot (1 - P\{H_{n,n+1}\}) \\ &\le P\{H_2\} \cdot \prod_{m=2}^{n} (1 - P\{H_{m,m+1}\}).\end{aligned}$$

Clearly, then, if we can show $\sum_{n=2}^{\infty} P\{H_{n,n+1}\} = \infty$, we will have proven the theorem. Applying Theorem 4.2 and (5.5) we have

(5.13)

$$P\{H_{n,n+1}\} \geq C(1+\varepsilon_n)^{2(k-1)}[\psi(t_{n+1})]^{2(k-1)}\exp\left\{-2\psi^2(t_{n+1})(1+\varepsilon_n)^2\right\}.$$

Consider the exponent and note that

$$\frac{1}{\alpha_{n+1}} - \frac{1}{\beta_{n+1}} \leq \frac{t_n}{t_{n+1}} \leq \frac{1}{\alpha_{n+1} - \beta_n^{-1}} \leq \frac{1}{\alpha_{n+1}} + O(1/\beta_n\alpha_{n+1}).$$

Consequently,

$$(1+\varepsilon_n)^2 = 1 + 3/\alpha_{n+1} + O(\alpha_{n+1}^{-2}),$$

so that by (5.11)

$$\begin{aligned}\psi^2(t_{n+1})(1+\varepsilon_n)^2 &\leq \psi^2(t_{n+1}) + (\log\log\beta_{n+1})\cdot 4/\alpha_{n+1} \\ &= \psi^2(t_{n+1}) + \frac{\log\sum_{2}^{n+1}\log^2 m}{4\log^2(n+1)} \\ &\leq \psi^2(t_{n+1}) + 1\end{aligned}$$

for large enough n. Substituting into (5.13) and setting $\gamma_n = (\alpha_{n+1})^{1+1/n}$ yields

$$\begin{aligned}\sum_n P\{H_{n,n+1}\} &\geq C\sum_{n=3}^{\infty}[\psi(t_n)]^{2(k-1)}e^{-2\psi^2(t_n)} \\ &\geq C\sum_n\int_{\alpha_n}^{\gamma_n}[\psi(t_n)]^{2(k-1)}e^{-2\psi^2(t_n)}\frac{\alpha_n}{s(\gamma_n-\alpha_n)}\,ds \\ &\geq C\sum_n\int_{\alpha_n}^{\alpha_{n+1}}\frac{[\psi(s^n)]^{2k}e^{-2\psi^2(s^n)}}{s}\cdot\frac{\alpha_n}{\psi^2(\beta_n)(\gamma_n-\alpha_n)}\,ds\end{aligned}$$

since $\psi^{2k}(t)e^{-2\psi^2(t)}$ is eventually decreasing in t. A change of variables leads to

$$\sum_n P\{H_{n,n+1}\} \geq C\sum_n a_n\int_{(\alpha_n)^n}^{(\alpha_{n+1})^{n+1}}\frac{[\psi(t)]^{2k}e^{-2\psi^2(t)}}{t}\,dt,$$

where

(5.14)
$$a_n = \frac{\alpha_n}{n\psi^2(\beta_n)(\gamma_n-\alpha_n)}.$$

If we can now show that a_n is bounded away from zero for large enough n, we shall have $\sum P\{H_{n,n+1}\} \geq CI_k(\psi) = \infty$ and the proof will be complete. Firstly, note that by (5.11)

(5.15)
$$\begin{aligned}\psi^2(\beta_n) &\leq \log\log\left(\prod_{m=2}^{n}\log^2 m\right) = \log\left(\sum_{m=2}^{n}\log(\log^2 m)\right) \\ &\leq \log((n-2)2\log\log n) \leq 2\log n.\end{aligned}$$

Furthermore,

$$\alpha_n/(\gamma_n - \alpha_n) > \alpha_n/(\alpha_{n+1} - \alpha_n) = \left[\frac{(\log n + 1)^2}{(\log n)^2} - 1\right]^{-1} \tag{5.16}$$

$$\geq \frac{1}{2}\left[\frac{\log n + 1}{\log n} - 1\right]^{-1}.$$

Substituting (5.15) and (5.16) into (5.14) yields

$$\begin{aligned} a_n &\geq \tfrac{1}{4}[n(\log(n+1) - \log n)]^{-1} \\ &= \tfrac{1}{4}[n \log(1 + 1/n)]^{-1} \\ &> \tfrac{1}{8}, \end{aligned}$$

which completes the proof of the theorem.

REFERENCES

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

BROWN, L. D. and ADLER, R. J. (1985). Multivariate Kolmogorov–Smirnov tests. Unpublished manuscript.

CABAÑA, E. M. (1984). On the transition density of multidimensional parameter Wiener process with one barrier. *J. Appl. Probab.* **21** 197–200.

CABAÑA, E. M. and WSCHEBOR, M. (1982). The two parameter Brownian bridge: Kolmogorov inequalities and upper and lower bounds for the distribution of the maximum. *Ann. Probab.* **10** 289–302.

CHUNG, K. L. (1949). An estimate concerning the Kolmogoroff limit distribution. *Trans. Amer. Math. Soc.* **67** 36–50.

DONSKER, M. (1951). An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* **6** 1–12.

DONSKER, M. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **23** 277–281.

DOOB, J. L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Statist.* **20** 393–403.

DUDLEY, R. M. (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126.

DUDLEY, R. M. (1967). Measures on non-separable metric spaces. *Illinois J. Math.* **11** 449–453.

DUDLEY, R. M. and PHILIPP, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. verw. Gebiete.* **62** 509–552.

ERDŐS, P. (1942). On the law of the iterated logarithm. *Ann. of Math.* **43** 419–436.

FELLER, W. (1948). On the Kolmogorov–Smirnov limit theorems for empirical distributions. *Ann. Math. Statist.* **19** 177–189.

FELLER, W. (1971). *An Introduction to Probability Theory and its Applications* **2**. Wiley, New York.

FERNIQUE, X. (1970). Intégrabilité des vecteurs gaussiens. *C. R. Acad. Sci. Paris Ser. A.* **270** 1698–1699.

FERNIQUE, X. (1971). Régularité des processus gaussiens. *Invent. Math.* **12** 304–320.

GOODMAN, V. (1976). Distribution estimates for functionals of the two parameter Wiener process. *Ann. Probab.* **4** 977–982.

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.

KAC, M., KIEFER, J. and WOLFOWITZ, J. (1955). On tests of normality and other tests of goodness of fit based on distance methods. *Ann. Math. Statist.* **26** 189–211.

KIEFER, J. (1961). On large deviations of the empiric distribution function of vector chance variables and a law of the iterated logarithm. *Pacific J. Math.* **11** 649–660.

KIEFER, J. (1972). Skorohod embedding of multivariate R.V.'s and the sample DF. *Z. Wahrsch. verw. Gebiete.* **24** 1–35.

KIEFER, J. and WOLFOWITZ, J. (1958). On the deviations of the empiric distribution function of vector chance variables. *Trans. Amer. Math. Soc.* **87** 173–186.

KOLMOGOROV, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Inst. Ital. Atti. Giorn.* **4** 83–91.

KUELBS, J. (1975). Sample path behaviour for Brownian motion in Banach spaces. *Ann. Probab.* **3** 247–261.

KUELBS, J. and PHILIPP, W. (1980). Almost sure invariance principles for partial sums of mixing B-valued random variables. *Ann. Probab.* **8** 1003–1036.

LANDAU, H. J. and SHEPP, L. A. (1971). On the supremum of a Gaussian process. *Sankhyā Ser. A* **32** 369–378.

LEADBETTER, M. R., LINDGREN, G., and ROOTZEN, H. (1983). Extremes and Related Properties of Random Sequences and Processes. *Springer, New York.*

MARCUS, M.B. and SHEPP, L. A. (1971). Sample behaviour of Gaussian processes. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2** 423–441. Univ. California Press.

PARK, C. and SKOUG, D. L. (1978). Distribution estimates of barrier-crossings probabilities of the Yeh–Wiener process. *Pacific J. Math.* **78** 455–466.

SIRAO, T. (1959). On the continuity of Brownian motion with a multidimensional parameter. *Nagoya Math. J.* **16** 135–156.

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

SMIRNOV, N. V. (1944). Approximation of the distribution law of a random variable by empirical data. *Uspehi Mat. Nauk.* **10** 179–206. In Russian.

FACULTY OF INDUSTRIAL ENGINEERING
AND MANAGEMENT
TECHNION
HAIFA, ISRAEL 32000

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853