Exponential Bounds in Vapnik-Červonenkis Classes of Index 1

By

Daphne L. Smith
University of California, Berkeley

Technical Report No. 129
December 1987

Research supported by the University of California Presidential Fellowship.

Department of Statistics
University of California
Berkeley, California

Exponential Bounds in Vapnik-Červonenkis Classes of Index 1

Daphne L. Smith
University of California, Berkeley

ABSTRACT

The smallest classes $\mathscr{C}$ which satisfy the Vapnik-Červonenkis combinatorial condition are assigned an index of 1. We show that over all classes $\mathscr{C}$ of index 1, the classical exponential inequalities for empirical processes are optimal.

Keywords and phrases. Empirical process, exponential bound, Vapnik-Červonenkis class, Brownian bridge, set-indexed gaussian process, partial order.

AMS 1980 subject classifications. Primary 60F05; secondary 60G15, 60B10, 60F10.

Abbreviated Title: Vapnik - Červonenkis classes of index 1

1. Introduction

Let $X_1, X_2, \ldots$ be i.i.d. random variables with common distribution P on a measurable space $(\Omega, \mathscr{A})$. We define the empirical measure

$$P_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i}$$

and the normalized empirical process on $\mathscr{A}$ defined by

$$\nu_n = n^{1/2}(P_n - P), \; n \geq 1.$$

For a given class of events $\mathscr{C} \subset \mathscr{A}$, let $D_n(\mathscr{C}) = \sup_{A \in C} |\nu_n(A)|$. Under certain conditions on the class $\mathscr{C}$, exponential bounds for the probability that $D_n(\mathscr{C})$ exceeds M, $M \geq 1$ have been obtained. The inequalities which arise have the general form

$$\Pr\{D_n(\mathscr{C}) > M\} < ae^{-\gamma M^2} \tag{1.1}$$

where a, γ are positive constants that depend on $\mathscr{C}$ but not on M. The results of Dvoretzky (1956), Kiefer (1961), Devroye (1982) and Alexander (1984) are all of this form, for various $\mathscr{C}$.

The classical example comes from considering the class of intervals on the real line. Here, we take $\mathscr{C}_1 = \{[0,t] : t \in [0,1]\}$. The fundamental result (Kolmogorov, 1932) gives

$$\Pr\{\sup_{A \in C_1} \nu_n(A) > M\} \sim e^{-2M^2}$$

as $n \to \infty$, for all M.

In Euclidean space we may consider another well known example. Let $\mathscr{C} = E_d : \{\{x : x_j \leq t_j, j = 1, \ldots, d\}, \mathbf{t} \in \mathbf{R}^d\}$. When $d > 1$, Kiefer (1961) established (1.1) with $\gamma = 2(1-\varepsilon)$, for all $\varepsilon > 0$ and for some $a = a(\varepsilon, d)$. Unlike the case $d = 1$, the form of the limiting distribution of $D_n(\mathscr{C})$ depends on P.

A new direction was set when Vapnik and Červonenkis (1968, 1971) introduced combinatorial ideas that lead to results for general families $\mathscr{C}$ of geometric regions, defined as follows. Let X be a set and $\mathscr{C}$ a class of subsets of X. A finite set $A \subset X$ is *shattered* by $\mathscr{C}$ if every subset $E \subset A$ is of the form $A \cap C$ for some $C \in \mathscr{C}$. Now $\mathscr{C}$ is called a *Vapnik-Červonenkis class* (or VC class) if for some $n \geq 1$, no n-element subset of X is shattered by $\mathscr{C}$. Let $S(\mathscr{C})$ equal the cardinality of the largest subset $A \subset X$ which is shattered by $\mathscr{C}$. We will call $S(\mathscr{C})$ the index of $\mathscr{C}$.

Several familiar classes of geometric regions are VC classes. These include the classes of all rectangles, all closed balls, all polyhedra with at most m faces, and quadrants of the form $(-\infty, \mathbf{t}]$, in $\mathbf{R}^d$. If $\mathscr{C}$ is a VC class, then $\{C \cap D : C, D \in \mathscr{C}\}$,

$\{C \cup D : C, D \in \mathscr{C}\}$ and $\{A^c : A \in \mathscr{C}\}$ are also VC classes. The class of all closed convex sets in $\mathbb{R}^2$ is not a VC class. Additional facts about VC classes are presented in Dudley (1978, 1984, etc.).

Assuming suitable measurability conditions Alexander (1984) in the more general setting of VC classes of functions shows that for VC classes with index d and $M \geq 8$, a bound in (1.1) of the form $a \exp(-(2-\varepsilon) M^2)$ results. The exponent $(2-\varepsilon) M^2$ cannot be improved; a bound of the form $a \exp(-2M^2)$ is the best possible. (Even a simple case where $\mathscr{C} = \{A\}$ with $P(A) = 1/2$ has this exact bound, Hoeffding (1963)). However, Alexander's work (intended for asymptotic use) yields a constant $a = a(\varepsilon, d)$ which is impractically large. Is further refinement possible?

Bounds such as those desired for (1.1) yield faster rates of convergence when ν_n is indexed by some small family of sets, such as the VC classes with index 1. But we shall give an example on $[0, 1]$ that shows:

(1.2) *there is no constant* $\beta < \infty$ *which will satisfy the inequality*

$$\Pr\{D_n(\mathscr{C}) > M\} < \beta e^{-2M^2} \text{ for } \textit{all classes with } S(\mathscr{C}) = 1.$$

2. An Asymptotic Distribution when $S(\mathscr{C}) = 1$

We introduce a collection of subsets on $[0,1]$. Let $X = [0,1]$ and define k disjoint subintervals

$$A_1 = [0, t_1]$$
$$A_{i+1} = (t_i, t_{i+1}],\ i = 1, 2, \ldots, k-1$$

with parameters $t_1 < \frac{1}{2}$; $0 < t_1 < t_2 < \cdots < t_k = 1$ such that $X = \bigcup_{i=1}^{k} A_i$. Let $\mathscr{C}_k$ be a collection of the k sets, $\mathscr{C}_i = A_1 \cup A_{i+1}$, $0 \le i \le k-1$. It is easy to check that $S(\mathscr{C}_k) = 1$ since no 2-point set $\{a, b\} \subset [0,1]$ is shattered by $\mathscr{C}_k$. In this section we show that as $M \to \infty$

(2.1) $$\Pr\{\sup_{A \in C_k} \nu_n(A) > M\} \sim (k-1)e^{-2M^2} \text{ and the}$$

and the result (1.2) follows.

In order to prove (2.1), we now consider appropriate Gaussian limit processes. Let $(\Omega, \mathscr{A}, P)$ be a probability space and let W_P denote the *isonormal* Gaussian process with $EW_P(A) = 0$ and $EW_P(A)W_P(B) = P(A \cap B)$ for *all* $A, B \in \mathscr{A}$. We define $G_P(A) := W_P(A) - P(A)W_P(\Omega)$ so that G_P is a Gaussian process indexed by a class of measurable sets $\mathscr{C} \subset \mathscr{A}$ with mean 0 and covariance

$$EG_P(A)G_P(B) = P(A \cap B) - P(A)P(B).$$

Now Dudley (1978) [under suitable measurability conditions] gives the weak convergence result

$$\underset{A \in \mathscr{C}}{\nu}(A) \to G_P(A) \quad \text{as } n \to \infty$$

which holds under different conditions on $\mathscr{C}$, and in particular for Vapnik-Červonenkis classes.

Remark 2.1. When P is uniform we have $G_P([0,t]) = Y_t$, $0 \le t \le 1$, where Y_t is the Brownian bridge. We will use $G_P(t)$ to denote $G_P([0,t])$.

Remark 2.2. Instead of the usual linear ordering of the parameter sets A the collection $\mathscr{C}_K$ is constructed by a treelike partial ordering. See Dudley (1984): Let $\mathscr{C} \subset 2^X$ satisfy $S(\mathscr{C}) = 1$ and $\phi \in \mathscr{C}$. Then the partial ordering of $\mathscr{C}$ by inclusion is treelike.

We note that a partial order $(\mathscr{C}, \subseteq)$ is called linear if for all $A, B \in \mathscr{C}$, either $A \subseteq B$ or $B \subseteq A$. A partial order will be called treelike iff for all $B \subseteq \mathscr{C}$ and $L(B) := \{A : A \subseteq B\}$, the restriction of $\subseteq$ to $L(B)$ is linear.

In the treelike p.o. $(\mathscr{C}_k, \subseteq)$ we may successfully compare events for which $t \in A_1$ and $t_i \in A_{i+1}$, $i = 1, \ldots, k-1$ but events for which $t_i \in A_i$ and $t_j \in A_j$ for $i \neq j \neq 1$ are not comparable for inclusion.

Remark 2.3. If A and B are disjoint measurable sets then $G_P(A \cup B) = G_P(A) + G_P(B)$ a.s. since the variance of $G_P(A \cup B) - G_P(A) - G_P(B) = 0$. Thus, if $X = \bigcup_{i \le k} A_i$ for disjoint A_i we have the linear relationship $G_P(X) = \sum_{i \le k} G_P(A_i) = 0$.

In the subsequent analysis, we will require several well-known [Doob (1965, 1949)] facts about the Brownian bridge. One such result is given here. We omit the straightforward proof.

Lemma 2.4. Let Y_t denote the Brownian bridge, $0 \le s \le t \le 1$.
Then $\Pr(\sup_{0 \le s \le t} Y_s > M \mid Y_t = y) = \exp(-2M(M-y)/t)$ where $M \ge \max(0, y)$.

We now state our first theorem.

Theorem 2.5. Let $(X, \mathscr{A}, P)$ be a probability space, P Lebesgue measure, $X = [0,1]$ and $\mathscr{C}_k$ the collection $\{\mathscr{C}_i, i = 0, \ldots, k-1\}$. Then as $M \to \infty$,

$$\Pr\{\sup_{A \in \bigcup \mathscr{C}_i} G_P(A) > M\} \text{ is less than or asymptotic to}$$

$$(k-1)\, e^{-2M^2}.$$

Proof. The proof of theorem 2.5 is done in two stages. First we determine an upper bound for $\Pr\{\sup G_P(A) > M\}$ when $A \in \mathscr{C}_k$; then we determine the form of the asymptotic distribution. For the class $\mathscr{C}_k$ we have $X = \bigcup_{i \le k} A_i$ and assign probabilities to the disjoint A_i as follows:

$$P(A_j) = p_j,\ j = 1, \ldots, k \text{ where } p_1 < \tfrac{1}{2} \tag{2.2}$$

and assume $P(A_1 \cup A_j) > \frac{1}{2}$ for each $j > 1$. In terms of the parameters t we have

$$p_1 = t_1,\ p_j = t_j - t_{j-1},\ j > 1.$$

Define events

$$B_i := \{\sup_{t \in A_i} G_P([0,t]) > M\}, \quad 1 \le i \le k.$$

We will indicate a method for obtaining bounds for $\Pr(B_i)$ over the collection $\mathscr{C}_k$ by considering the following inequality.

(2.3) $$\Pr(\sup_{t \in \bigcup A_i} G_P([0,t]) > M)$$

(2.4) $$= \Pr(\bigcup_i B_i) \le \Pr(B_1) + \sum_{j>1} \Pr(B_j \backslash B_1).$$

Now let $Z_t := G_P([0,t])$. When $t \in A_1$, we have $Z_t \equiv Y_t$ when Y_t denotes the Brownian bridge.

We wish to find the distribution of the supremum of the Brownian bridge indexed by the parameter sets A_i. We begin by computing the distribution on the set A_1. Later we will consider the general case.

Now

(2.5) $$\Pr(B_1) = \Pr(Y_{t_1} > M) + \Pr(\sup Y_t > M \text{ for } t \in [0,t_1) \text{ and } Y_{t_1} < M).$$

The probability given in the first term is simply the tail distribution of the Brownian bridge with $E\,Y_{t_1} = 0$ and variance $\sigma_1^2 = t_1(1-t_1)$. To evaluate the second term we integrate the conditional probability $\Pr(\sup_{0 \le t < t_1} Y_t > M \mid Y_{t_1} = y)$ with respect to the distribution of Y_{t_1}. Using Lemma 2.4 and after some calculation the expression in (2.5) reduces to

(2.6) $$\Pr(B_1) = K(M/\sigma_1) + \exp(-2M^2)\,K[M(1-2t_1)/\sigma_1].$$

where $K(\lambda) = 1 - \Phi(\lambda)$, and Φ is the standard normal distribution function.

Next, to evaluate the second term in the inequality given by (2.4) we start with

(2.7) $$\Pr(B_j \backslash B_1) = \Pr[\sup_{t \in C_j} Z_t > M] - \Pr[\sup_{t \in A_1} Z_t > M].$$

Define $\nu_j = t_1 + t_j - t_{j-1}$. Note that $\nu_j > ½$. We consider the first probability in (2.7)

(2.8) Write $$\Pr[\sup_{t \in C_j} Z_t > M] = \Pr[\sup_{t \in [0,\nu_j]} Y_t > M]$$

$$= \Pr[Y_{\nu_j} \ge M] + \Pr[\sup_{t \in [0,\nu_j]} Y_t > M \text{ and } Y_{\nu_j} < M].$$

$$= \Pr_1 + \Pr_2$$

The distribution for Y_{ν_j} has the same form as in (2.6) with t_1 replaced by ν_j and variance $\sigma_j^2 := \nu_j(1-\nu_j)$. The case for the parameter ν_j can be found using analysis similar to that used in computing (2.6). Therefore, we find that

$$\Pr_1 = K(M/\sigma_j), \text{ and}$$

$$\Pr_2 = \int_{-\infty}^{M} e^{-2M(M-y)/v_j} \frac{1}{\sqrt{2\pi}\,\sigma_j} e^{-y^2/(2\sigma_j^2)}\,dy$$

and after combining the exponential terms and completing the square this integral reduces to

$$= e^{-2M^2}[1 - K[M(2v_j - 1)/\sigma_j]]$$

where $v_j > ½$.

The inequality (2.5) may be evaluated by substituting (2.5), (2.8) and rearranging the terms to yield the upper bound:

$$\text{(2.9)} \qquad \Pr(\bigcup_i B_i) \le \sum_{j=2}^{k} \exp(-2M^2) + K[M/\sigma_1] + \sum_{j>1} [K[M/\sigma_j] - K[M/\sigma_1]]$$

$$+ \exp(-2M^2)\,K[M(1 - 2t_1)/\sigma_1]$$

$$- \sum_{j>1} \exp(-2M^2)\,[K[M(2v_j - 1)/\sigma_j] + K[M(1 - 2t_1)/\sigma_1]]$$

$$= (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}) + (\mathrm{IV}) - (\mathrm{V}).$$

We complete the proof by finding the dominant terms in the asymptotic expression for (2.9) as $M \to \infty$.

In (II) + (III) let

$$\gamma_j := (2\sigma_j^2)^{-1} - 2, \quad j = 1,\ldots,k$$

Choosing $\gamma = \min(\gamma_1, \ldots, \gamma_k) > 0$ will condense the notation.

The Mills' ratio expansion for the tail-end area $K(\cdot)$ gives (II) + (III) $= O(\frac{1}{M}\exp(-(2 + \gamma)M^2))$. Computing along the same lines we obtain for

$$\delta_m := \begin{cases} (1 - 2t_1)^2/\sigma_1^2 & \text{if } m = 1 \\ (2v_j - 1)^2/\sigma_j^2 & \text{if } m = j > 1 \end{cases}$$

that

$$(\mathrm{IV}) = O[\frac{1}{M}\exp(-(2 + \delta_1)M^2)]$$

$$\text{and } (\mathrm{V}) = O[\sum_{j=1}^{k} \frac{1}{M}\exp(-(2 + \delta_j)M^2)].$$

The appropriate substitutions allow us to conclude that $\Pr[\cup B_i] \le (k-1)\exp(-2M^2) + O[\frac{1}{M}\exp(-(2+\delta)M^2)]$ where $\delta := \min(\gamma, \delta_1, \ldots, \delta_k) > 0$. This inequality yields the desired result. □

We consider now the remaining result of this paper.

Theorem 2.6. Let $(X, \mathscr{A}, P)$ be a probability space, P Lebesgue measure, $X = [0,1]$ and $\mathscr{C}_k = \{\mathscr{C}_i; i = 0, \ldots, k-1\}$. Then as $M \to \infty$,

$$\Pr[\sup_{A \in \bigcup C_i} G_P(A) > M] \sim (k-1)e^{-2M^2}.$$

Proof. To prove theorem 2.6 we make the assumption used in the computation of the upperbound namely, $P(A_1) < ½$ and $P(A_1 \cup A_j) > ½$ for each $j > 1$. We will determine a lower bound for (2.3) by using the inequality

(2.10) $$\Pr[\bigcup_i B_i] \ge \Sigma \Pr(B_i) - \sum_{i<j} \Pr(B_i \cap B_j).$$

For $j > 1$, we begin our evaluation with

(2.11) $$\Pr(B_j) \ge \Pr(B_1 \cup B_j) - \Pr(B_1).$$

In view of (2.8) and (2.6) it is easy to verify that

$$\begin{aligned} \Pr(B_j) \ \ge\ & K[M/\sigma_j] + \exp(-2M^2)[1 - K(M(2v_j - 1)/\sigma_j)] \\ & - K[M/\sigma_1] - \exp(-2M^2)K[M(1-2t_1)/\sigma_1],\ j > 1. \end{aligned}$$

To bound $\Pr(B_i \cap B_j)$ we refer to the simple case $B_2 \cap B_3$ and then generalize the argument to account for arbitrary indices. But first we will require a calculation.

Remark 2.7. In $\mathscr{C} = \bigcup_{i \le 3} \mathscr{C}_i$ we assume $X = \bigcup_{i=1}^{3} A_i$ with $P(A_1) = p_1$, where $p_1 < ½$ and $P(A_j) = (1-p_1)/2$, $j = 2, 3$.

If we put $G_P(A_1) = x$, $G_P(A_2) = y$ we have the constraint $x + y + z = 0$ so that $z \equiv -x-y$. Now the joint distribution function of (x, y) is bivariate normal with density function $f(x,y)$ and covariance $\Sigma_{xy} = P(A_1 \cap A_2) - P(A_1)P(A_2)$. One easily calculates $\Sigma_{xy} = -p_1(1-p_1)/2$ for $x \ne y$, $\Sigma_{xx} = p_1(1-p_1)$, and $\Sigma_{yy} = (1-p_1)(1+p_1)/4$.

To bound $\Pr(B_2 \cap B_3)$ we will consider

$$\Pr[\sup_{t \in A_2} G_P(t) > M, \sup_{t \in A_3} G_P(t) > M].$$

We will make use of the conditional distributions to obtain a bound for this expression. Now from Lemma 2.4 we have

$$\Pr[\sup_{t\in A_2} G_P(t) > M \mid G_P(A_1) = x,\ G_P(A_2) = y]$$

$$= \begin{cases} \exp(-2(M-x)(M-x-y)/(t_2-t_1)) & \text{if } x+y<M \text{ and } x<M. \\ 1 & \text{if } x+y \ge M \text{ or } x \ge M. \end{cases}$$

The formula for A_3 can be obtained similarly. Thus,

$$\Pr[\sup_{t\in A_2} G_P(t) > M,\ \sup_{t\in A_3} G_P(t) > M] \tag{2.12}$$

$$= \iint \Pr[\sup_{t\in A_2} G_P(t) > M \mid G_P(A_1) = x,\ G_P(A_2) = y]$$

$$\cdot \Pr[\sup_{t\in A_3} G_P(t) > M \mid G_P(A_1) = x,\ G_P(A_3) = z]\ f(x,y)\,dx\,dy$$

which in accordance with Lemma 2.4 becomes

$$= \iint \Big(1_{\{x+y\ge M \text{ or } x\ge M\}}$$

$$+ 1_{\{x+y<M \text{ and } x<M\}} \exp(-2(M-x)(M-x-y)/(t_2-t_1))\Big)$$

$$\times \Big(1_{\{-y\ge M \text{ or } x\ge M\}}$$

$$+ 1_{\{-y<M \text{ and } x<M\}} \exp(-2(M-x)(M+y)/(t_3-t_2))\Big) f(x,y)\,dx\,dy.$$

To obtain an explicit formula for (2.12) we expand the product in the integrand and transform the integrals to standard bivariate normal form. It is a simple but tedious calculation (which will be omitted here) to then show that computation of the asymptotic bounds may be condensed and each exponential term is of the form $\exp(-(2+\delta_i)M^2)$, $\delta_i = \delta_i(p_1) > 0$ with $p_1 < ½$. Therefore,

$$\Pr[\sup_{t\in A_2} G_P(t) > M,\ \sup_{t\in A_3} G_P(t) > M]$$

$$= F_1 + F_2 + F_3 + F_4 + F_5 - F_6,$$

where

$$F_i = \exp(-(2+\delta_i)M^2) \quad i = 1, \ldots, 6$$

with
$$\delta_1 = (2p_1-1)^2[2p_1(1-p_1)]^{-1},$$

$$\delta_2 = 2(2-3p_1)^2[(1+9p_1)(1-p_1)]^{-1},$$

$$\delta_3 = 2(9p_1^2-8p_1+3)[(1+9p_1)(1-p_1)]^{-1},$$

$$\delta_4 = \begin{cases} [2p_1(1-p_1)]^{-1} - 2 & \text{when } p_1 \in (¼, ½) \\ 2(3-8p_1)/(1+8p_1) & \text{when } p_1 \in (0, ¼] \end{cases}$$

$$\delta_5 = 2(1 - p_1)/(1 + p_1)$$

$$\text{and } \delta_6 = \delta_5.$$

Let $\delta := \min(\delta_1, \delta_2, \ldots, \delta_6)$. then (2.12) $\leq O\exp(-(2 + \delta)M^2)$, $\delta > 0$. This inequality provides an upper bound for $\Pr(B_2 \cap B_3)$.

To compute a bound for $\Pr(B_i \cap B_j)$ we will use a new description of the intervals in $\mathscr{C}_k$. Let $\mathscr{C}_k'$ be the class obtained by reordering the subintervals of X as follows. W.l.o.g. assume k is odd. Reparametrize the sets $A_4, A_5, \ldots, A_k$, for k odd, so that $0 < t_1 < t_2 < t_4 < t_{2m-2} < t_{2m} < t_{k-1}$ and $0 < t_1 < t_3 < t_5 < t_{2m-1} < t_{2m+1} < t_k$. Then the sets $A_2, A_4, \ldots$ form one branch of the treelike ordering and the sets $A_3, A_5, \ldots$ form the other branch.

Insert Diagram 1

Thus the event $B_2 \cap B_3$ occurs on $\mathscr{C}_k$ and $\mathscr{C}_k'$ with equal probability.
Hence (2.13)

$$\Pr(B_2 \cap B_3) \leq \Pr((B_2 \cup \cdots \cup B_{k-1}) \cap (B_3 \cup \cdots \cup B_k))$$

for which this intersection probability can be written

$$\Pr(B_2 \cup \cdots \cup B_{k-1}) + \Pr(B_3 \cup \cdots \cup B_k)$$
$$-\Pr(B_2 \cup B_3 \cup \cdots \cup B_{k-1} \cup B_k).$$

To generalize our argument to include the case of pairwise intersections of the events B_i and B_j, $i \neq j$ we note the restrictions on i and j occur in $\binom{k-1}{2}$ ways. Therefore, $\Pr(\cup B_i) \geq (k-1)e^{-2M^2} - \binom{k-1}{2} O[\exp(-(2+\delta)M^2)]\ \delta > 0,$
or passing to the limit as $M \to \infty$

$$\Pr(\cup B_i) \sim (k-1)e^{-2M^2}$$

which completes the proof of this theorem. □

Acknowledgement. I would like to thank my thesis advisor Professor Richard M. Dudley for guidance and helpful discussion during the course of my thesis work.

REFERENCES

ALEXANDER, K. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.*. **12** 1041-1067. Correction (1987); **15** 428-430.

BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading.

DEVROYE, L. (1982). Bounds for the uniform deviation of empirical measures. *J. Multivar. Anal.* **12** 72-79.

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

DOOB, J.L. (1965). *Stochastic Processes*. Wiley, New York.

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

DUDLEY, R.M. (1978). Central limit theorems for empirical measures. *Ann. Prob.* **6** 899-929. Correction (1979) **7** 909-911.

DUDLEY, R.M. (1984). A course on empirical processes. *Ecole d'ete de probabilities de St. Four, 1982. Lecture notes in Math.* **1097** 1-142. Springer, New York.

DUDLEY, R.M. (1985). Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, 1983 (L.M. Le Cam and R.A. Olshen, eds.) Wadsworth, Monterey, California.

DURBIN, J. (1973). Distribution theory for tests based on the sample distribution function. *SIAM*, Philadelphia.

DVORETZKY, A., KIEFER, J., and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Stat.* **27** 642-669.

ERDELYI, A. (1956). *Asymptotic expansions*. Dover publications, New York.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13-30.

JOHNSON, N.L. and KOTZ, S. (1972). *Distribution in statistics: continuous multivariate distributions*. Wiley, New York.

KARLIN, S. and TAYLOR, H. (1975). *A First Course in Stochastic Processes* (2nd ed.). Academic Press, New York.

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

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

MCKEAN, H.P. (1963). Brownian motion with a several-dimensional time. *Theory of Prob. and its Appl.* **8** 335-354.

PITT, C.D. (1971). A Markov property for Gaussian processes in a multidimensional parameter. *Arch. Rat. Mech.* **43** 368-391.

POLLARD, D. (1984). *Convergence of stochastic processes.* Springer-Verlag, New York.

RUBEN, H. (1961). Probability content of regions under spherical normal distirbutions, III: The bivariate normal integral. *Ann. Math. Statist.* **32** 171-186.

SEGAL, I.E. (1954). Abstract probability spaces and a theorem of Kolmogoroff. *Amer. J. Math.* **76** 721-732.

SMITH, D.L. (1985). Vapnik-Červonenkis classes and the supremum distribution of a Gaussian process. Ph.D. dissertation, Massachusetts Institute of Technology.

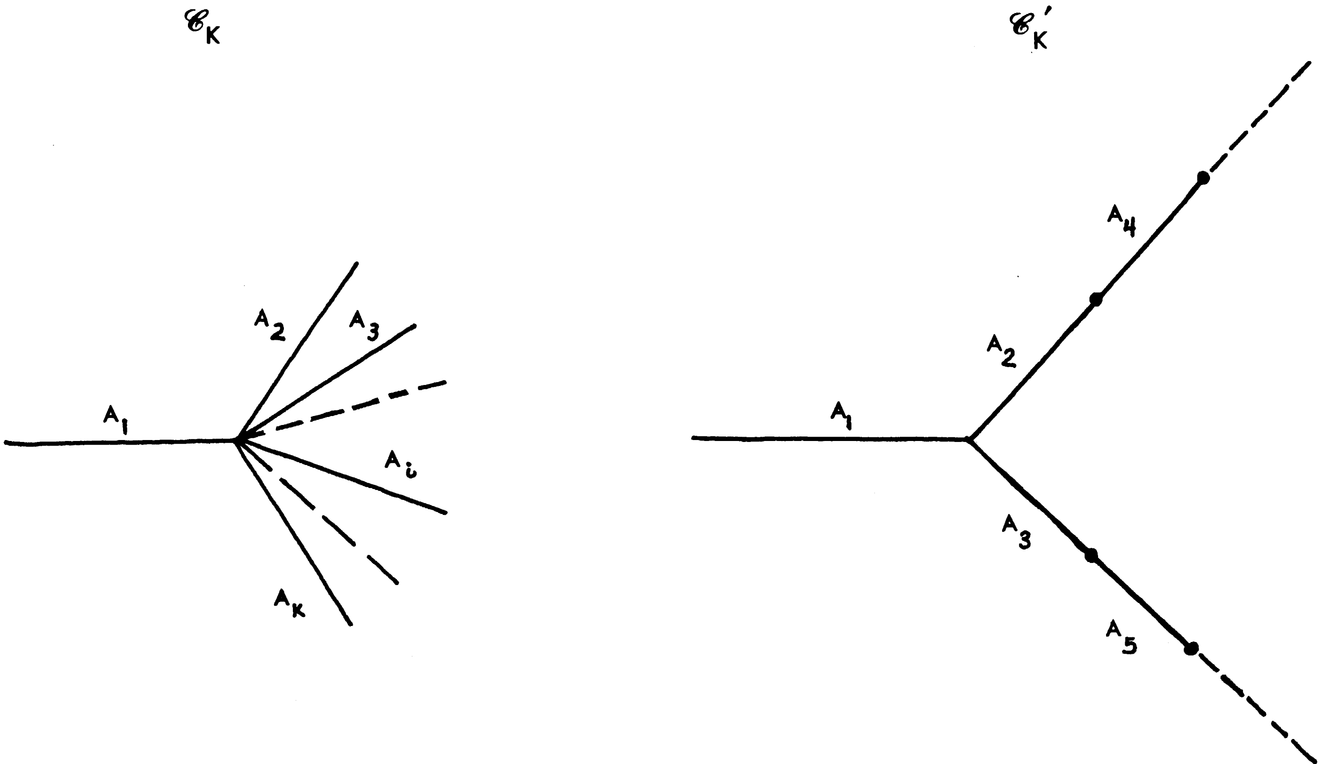
VAPNIK, V.N. and ČERVONENKIS, A. YA. (1968). Uniform convergence of frequencies of occurrence of events to their probabilities. *Sov. Math Doklady* **9** 915-918.

VAPNIK, V.N. and ČERVONENKIS, A. YA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Prob. Appl.* **16** 264-280.

WENOCUR, R. and DUDLEY, R. (1981). Some special Vapnik-Červonenkis classes. *Discrete Math.* **33** 313-318.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

Diagram 1



TECHNICAL REPORTS

Statistics Department

University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.

2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosciences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.

3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.

4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11. No. 1, 185-214.

5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.

6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.

7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.

8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.

9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.

10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.

11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.

12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.

14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.

15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.

16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.

17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review.

18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.

19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).

20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting. 1985, Vol. 4, 251-262.

21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.

22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.

23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics, (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984.

25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.

26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.

27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.

28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimate Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.

29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.

30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES, 1985 Vol 3 pp. 1-13.

31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.

32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.

33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.

34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.

35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science, Feb 1986, Vol. 1, No. 1, 3-39.

36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.

37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly, Feb 1986, Vol. 93, No. 2, 123-125.

38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.

39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.

40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.

41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.

42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.

43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?

44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.

45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.

46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.

47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data.

48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.

49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift, 1986. D. Reidel.

50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada, January, 1986.

51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.

53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.

54. BLACKWELL, D. (November 1985). Approximate normality of large products.

55. FREEDMAN, D. A. (December 1985, revised Dec. 1986). As others see us: A case study in path analysis. Prepared for the Journal of Educational Statistics.

56. LE CAM, L. and YANG, G. L. (January 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies.

57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.

58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.

59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.

60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.

61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.

62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.

63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.

64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.

65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.

66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.

67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data.

68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).

69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.

71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.

72. LEHMANN, E.L. (July 1986). Statistics - an overview.

73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.

74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial, and some continued fractions of Gauss.

75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.

76. O'SULLIVAN, F. (September 1986). Relative risk estimation.

77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.

78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.

79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks.

80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.

81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.

82. DOKSUM, K.J. and LO, A.Y. (November 1986). Consistent and robust Bayes Procedures for Location based on Partial Information.

83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.

85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory.

86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.

87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.

88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.

89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised January 1988). Estimating and testing in a two sample generalized odds rate model.

90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.

91. DIACONIS, P and FREEDMAN, D.A. (March 1987). A finite version of de Finetti's theorem for exponential families, with uniform asymptotic estimates.

92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.

92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.

93. DABROWSKA, D.M. (June 1987, revised Feb 1988). Nonparametric quantile regression with censored data.

94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.

95. RIZZARDI, F. (Aug 1987). Two-Sample t-tests where one population SD is known.

96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in *Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley.*

97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for cancer.

98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.

99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.

100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.

101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators.

102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.

103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.

104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer.

105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.

106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.

107. CHENG, C-S. (August 1987). Some orthogonal main-effect plans for asymmetrical factorials.

108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.

109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \le k \le n} S_k^+ / ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.

110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.

113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.

114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.

115. BICKEL, P.J. and RITOV, Y. (September 1987). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (September 1987). Unachievable information bounds in non and semiparametric models.

117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.

118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.

119. HESSE, C.H. (October 1987). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.

120. DONOHO, D. and LIU, R. (October 1987, revised March 1988). Geometrizing rates of convergence, II.

121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. To appear in *Statistics a Guide to the Unknown.*

122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.

123. DONOHO, D.L. and MACGIBBON, B. (November 1987). Minimax risk for hyperrectangles.

124. ALDOUS, D. (November 1987). Stopping times and tightness II.

125. HESSE, C.H. (November 1987). The present state of a stochastic model for sedimentation.

126. DALANG, R.C. (December 1987). Optimal stopping of two-parameter processes on hyperfinite probability spaces.

127. Same as No. 133.

128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.

129. SMITH, D.L. (December 1987). Exponential bounds in Vapnik-Červonenkis classes of index 1.

130. STONE, C.J. (November 1987). Uniform error bounds involving logspline models.

131. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.

132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes

133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.

134. DUBINS, L.E. and SCHWARZ, G. (December 1987). A sharp inequality for martingales and stopping-times.

135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). On the risk of lung cancer for ex-smokers.

136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.

137. DIACONIS, P. and FREEDMAN, D.A. (January 1988). On the consistency of Bayes estimates.

138. DONOHO, D.L. and LIU, R.C. (January 1988). Geometrizing rates of convergence, III.

139. BERAN, R. (January 1988). Refining simultaneous confidence sets.

140. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.

141. BRILLINGER, D.R. (January 1988). Two reports on trend analysis: a) An Elementary Trend Analysis of Rio Negro Levels at Manaus, 1903-1985 b) Consistent Detection of a Monotonic Trend Superposed on a Stationary Time Series

142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density.

143. DALANG, R.C. (February 1988). Randomization in the two-armed bandit problem.

144. DABROWSKA, D.M. and DOKSUM, K.A. (February 1988). Graphical comparisons of cumulative hazards for two populations.

145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.

146. BICKEL, P.J. and RITOV, Y. (February 1988). Estimating integrated squared density derivatives.

147. STARK, P.B. (March 1988). Strict bounds and applications.

148. DONOHO, D.L. and STARK, P.B. (March 1988). Rearrangements and smoothing.

149. NOLAN, D. (March 1988). Asymptotics for a multivariate location estimator.

150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.

151. NOLAN, D. (March 1988). Limit theorems for a random convex set.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.