

Calibrating Prediction Regions

Rudolf Beran*

University of California, Berkeley

Technical Report No. 183

November 1988

***This research was supported in part by NSF Grant DMS 87-01426.**

**Part of the work was done while the author was a guest of Sonderforschungsbereich 123
at Universität Heidelberg.**

The author thanks G. Sawitzki and F. Seillier for helpful comments.

**Department of Statistics
University of California
Berkeley, California**

Calibrating Prediction Regions

Rudolf Beran*

University of California, Berkeley

ABSTRACT

Suppose the variable X to be predicted and the learning sample Y_n that was observed have a joint distribution, which depends on an unknown parameter θ . The parameter θ can be finite or infinite dimensional. A prediction region D_n for X is a random set, depending on Y_n , that contains X with prescribed probability α . This paper studies methods for controlling simultaneously the conditional coverage probability of D_n , given Y_n , and the overall (unconditional) coverage probability of D_n . The basic construction yields a prediction region D_n which has the following properties in regular models: Both the conditional and overall coverage probabilities of D_n converge to α as the size n of the learning sample increases. The convergence of the former is in probability. Moreover, the asymptotic distribution of the conditional coverage probability about α is typically normal; and the overall coverage probability tends to α at rate n^{-1} . Can one reduce the dispersion of the conditional coverage probability about α and increase the rate at which overall coverage probability converges to α ? Both issues are addressed. The paper establishes a lower bound for the asymptotic dispersion of conditional coverage probability. The paper also shows how to calibrate D_n so as to make its overall coverage probability converge to α at the faster rate n^{-2} . This calibration adjustment does not affect the asymptotic distribution or dispersion of the conditional coverage probability, in a first-order analysis. In general, a bootstrap Monte Carlo algorithm accomplishes the calibration of D_n . In special cases, analytical calibration is possible.

* This research was supported in part by NSF Grant DMS 87-01426. Part of the work was done while the author was a guest of Sonderforschungsbereich 123 at Universität Heidelberg. The author thanks G. Sawitzki and F. Seillier for helpful comments.

1. INTRODUCTION

Prediction regions are of interest in the following context. The variable X to be predicted and the learning sample Y_n that was observed have a joint distribution $P_{\theta,n}$. The parameter θ is unknown but is restricted to a finite or infinite dimensional parameter space. A prediction region for X is a random set $D_n = D_n(\alpha, Y_n)$, depending on the learning sample, that contains X with prescribed probability α .

Let $P_\theta(\cdot | Y_n)$ denote the conditional distribution of X given Y_n . The *conditional coverage probability of D_n given Y_n* is

$$CP(D_n | Y_n, \theta) = P_\theta[X \in D_n | Y_n]. \quad (1.1)$$

The *overall coverage probability of D_n* is

$$\begin{aligned} CP(D_n | \theta) &= E_\theta CP(D_n | Y_n, \theta) \\ &= P_{\theta,n}[X \in D_n], \end{aligned} \quad (1.2)$$

where the expectation is with respect to the distribution $Q_{\theta,n}$ of Y_n . If $CP(D_n | Y_n, \theta)$ equals α , exactly or asymptotically, then so does $CP(D_n | \theta)$. The converse may not be true, as shown by the following example.

Example 1. Consider the stationary first-order autoregressive model

$$X_i = \theta X_{i-1} + E_i \quad (1.3)$$

where the $\{E_i\}$ are i.i.d. standard normal random variables and $|\theta|$ is no larger than $1 - \epsilon$, for some small positive ϵ . Suppose the learning sample is $Y_n = (X_1, \dots, X_n)$ and the goal is to predict $X = X_{n+1}$. Let $\hat{\theta}_n$ denote the least squares estimate of θ based on Y_n , clipped so that $|\hat{\theta}_n|$ does not exceed $1 - \epsilon$:

$$\hat{\theta}_n = \max \left\{ \min \left\{ \sum_{i=1}^{n-1} X_i X_{i+1} / \sum_{i=1}^{n-1} X_i^2, 1 - \epsilon \right\}, -1 + \epsilon \right\}. \quad (1.4)$$

Let Φ denote the standard normal cdf and let $z_\alpha = \Phi^{-1}(\alpha)$. For the one-sided prediction interval

$$D_n = (-\infty, z_\alpha + \hat{\theta}_n X_n], \quad (1.5)$$

the conditional coverage probability is

$$CP(D_n | Y_n, \theta) = \Phi[z_\alpha + (\hat{\theta}_n - \theta) X_n] \quad (1.6)$$

because the conditional distribution of X given Y_n is $N(\theta X_n, 1)$ here. Since $\hat{\theta}_n$ is a consistent estimate of θ , both $CP(D_n | Y_n, \theta)$ and $CP(D_n | \theta)$ converge to α as n increases, the former in probability.

On the other hand, for the alternative one-sided prediction interval

$$D_n = (-\infty, z_\alpha(1 - \hat{\theta}_n^2)^{-1/2}], \quad (1.7)$$

the overall coverage probability tends to α as n increases, because X has a $N(0, (1 - \theta^2)^{-1})$ distribution. However

$$CP(D_n|Y_n, \theta) = \Phi[z_\alpha(1 - \hat{\theta}_n^2)^{-1/2} - \theta X_n] \quad (1.8)$$

does not converge to α , because in the limit it is distributed as $\Phi[(z_\alpha - \theta Z)(1 - \theta^2)^{-1/2}]$, where Z is a standard normal random variable.

This paper presents methods for controlling simultaneously the overall coverage probability and the conditional coverage probability of a prediction region. Both goals are important. Several recent authors, including Cox and Hinkley (1974), Cox (1975, 1986), Atwood (1984), Bai and Olshen (1988) have emphasized overall coverage probability. Other recent authors, including Guttman (1970), Butler and Rothman (1980), Butler (1982), Stine (1985), have discussed conditional coverage probability as well. Terminology varies greatly in the works just cited.

The main findings of this paper are as follows. For a natural construction of the prediction region D_n , the overall coverage probability $CP(D_n|\theta)$ converges to α at rate n^{-1} while $n^{1/2}[CP(D_n|Y_n, \theta) - \alpha]$ has a normal limit law with mean zero and variance $\sigma^2(\theta)$. Appropriate calibration of D_n makes the overall coverage probability tend to α at rate n^{-2} , without affecting the normal limit law for the centered conditional coverage probability. At the same time, the asymptotic variance $\sigma^2(\theta)$ can be minimized by the construction, without affecting the n^{-1} or n^{-2} rate of convergence for overall coverage probability. Sections 3 and 4 give details.

The calibration operation mentioned above is introduced in section 2. In the simplest examples, the calibration of D_n can be done exactly. In a larger class of examples, an asymptotic approximation is available for the calibration operation. This approximation is linked to ideas in Cox (1975). In general, a bootstrap Monte Carlo algorithm accomplishes the calibration, without bogging down in complicated algebra.

2. CONSTRUCTIONS

An extension of the classical pivotal method generates a prediction region D_n whose conditional coverage probability and overall coverage probability both converge to α as the size of the learning sample increases. Calibration of the critical value of D_n seeks to reduce error in overall coverage probability without increasing the dispersion of the conditional coverage probability. Monte Carlo approximations are available for the critical values and calibrated critical values of D_n .

2.1 Constructing D_n

Let $R_n = R_n(X, Y_n)$ be a root for the prediction region — a function of X and Y_n which will be referred to a critical value in order to generate the desired prediction region D_n for X . Let $A_n(\cdot, \theta, Y_n)$ be the conditional cdf of R_n given Y_n . Assume this cdf to be continuous. Suppose $\hat{\theta}_n = \hat{\theta}_n(Y_n)$ is a consistent estimate of θ based on the learning sample. The plug-in estimate of $A_n(\cdot, \theta, Y_n)$ is then $A_n(\cdot, \hat{\theta}_n, Y_n)$. Define the prediction region D_n by referring R_n to the *largest* α th quantile of $A_n(\cdot, \hat{\theta}_n, Y_n)$:

$$\begin{aligned} D_n &= \{x : R_n(x, Y_n) \leq A_n^{-1}(\alpha, \hat{\theta}_n, Y_n)\} \\ &= \{x : A_n[R_n(x, Y_n), \hat{\theta}_n, Y_n] \leq \alpha\}. \end{aligned} \quad (2.1)$$

This construction is motivated by prediction interval (1.5) for Example 1.

The conditional coverage probability of D_n for X is

$$CP(D_n | Y_n, \theta) = A_n[A_n^{-1}(\alpha, \hat{\theta}_n, Y_n), \theta, Y_n]. \quad (2.2)$$

Under regularity conditions, such as those to be discussed in section 3, both $CP(D_n | Y_n, \theta)$ and $CP(D_n | \theta)$ converge to α as n increases, the former converging in probability.

Example 2. Suppose X and the elements of $Y_n = (X_1, \dots, X_n)$ are i.i.d. $N(\mu, \sigma^2)$ random variables, the parameter $\theta = (\mu, \sigma^2)$ being unknown. Let $\hat{\theta}_n = (\bar{X}_n, s_n^2)$ denote the usual unbiased estimate of θ based on Y_n . A classical root for this problem,

$$R_n(X, Y_n) = (X - \bar{X}_n) / [s_n(1 + n^{-1})^{1/2}], \quad (2.3)$$

generates, through (2.1), the one-sided prediction interval

$$D_n = (-\infty, \bar{X}_n + s_n z_\alpha]. \quad (2.4)$$

It is easily seen, in this case, that both $CP(D_n | Y_n, \theta)$ and $CP(D_n | \theta)$ tend to α as n increases, the first convergence being in probability. Nevertheless, prediction interval (2.4) is disappointing in two ways. First, the simpler root $R_n(X, Y_n) = X$ yields the same prediction interval by construction (2.1). Secondly, the classical interval for this problem replaces z_α in (2.4) with $(1 + n^{-1})^{1/2}$ times the α th quantile of the t -distribution with $n - 1$ degrees of freedom. This substitution produces a prediction interval whose overall coverage probability is exactly α and whose conditional coverage probability still converges to α as n increases. The possibility of reducing the coverage probability error of D_n in general models is the subject of the next section.

2.2 Calibrating D_n

Let $H_n(\cdot, \theta)$ denote the cdf of the transformed root $A_n(R_n, \hat{\theta}_n, Y_n)$. By (1.2) and the second line of (2.1) respectively,

$$\begin{aligned} \text{CP}(D_n|\theta) &= E_\theta \text{CP}(D_n|Y_n, \theta) \\ &= H_n(\alpha, \theta) \end{aligned} \quad (2.5)$$

Both expressions for overall coverage probability of $D_n = D_n(\alpha)$ will be used in the sequel. Suppose $H_n(\cdot, \theta)$ is continuous at its α th quantiles and $H_n^{-1}(\alpha, \theta)$ denotes the *largest* α th quantile. If θ were known, replacing prediction region $D_n(\alpha)$ with $D_n[H_n^{-1}(\alpha, \theta)]$ would yield a prediction region whose coverage probability is exactly α . Since θ is unknown but has an estimate $\hat{\theta}_n$, it is natural to consider the calibrated prediction set

$$D_{n,1}(\alpha) = D_n[H_n^{-1}(\alpha, \hat{\theta}_n)]. \quad (2.6)$$

It will be shown in section 3 that $\text{CP}(D_{n,1}|\theta)$ typically converges to α at a faster rate than $\text{CP}(D_n|\theta)$. On the other hand, $\text{CP}(D_{n,1}|Y_n, \theta)$ converges in probability to α at the same rate as $\text{CP}(D_n|Y_n, \theta)$.

Example 2 (continued). Both the root $R_n = X$ and the root (2.3) yield the one-sided prediction interval (2.4). By (2.2),

$$\begin{aligned} \text{CP}(D_n|Y_n, \theta) &= \Phi[z_\alpha + \sigma^{-1}\{(s_n - \sigma)z_\alpha + (\bar{X}_n - \mu)\}] \\ &= \alpha + O_p(n^{-1/2}). \end{aligned} \quad (2.7)$$

Let J_r denote the cdf of the t-distribution with r degrees of freedom. Since

$$H_n(x, \theta) = J_{n-1}[z_x(1 + n^{-1})^{-1/2}], \quad (2.8)$$

it follows by the second line of (2.5) that

$$\begin{aligned} \text{CP}(D_n|\theta) &= J_{n-1}[z_\alpha(1 + n^{-1})^{-1/2}] \\ &= \alpha - (4n)^{-1}(z_\alpha^3 + 3z_\alpha)\phi(z_\alpha) + O(n^{-2}). \end{aligned} \quad (2.9)$$

From (2.8),

$$H_n^{-1}(\alpha, \theta) = \Phi[J_{n-1}^{-1}(\alpha)(1 + n^{-1})^{1/2}]. \quad (2.10)$$

Consequently, the calibrated prediction interval is

$$D_{n,1}(\alpha) = (-\infty, \bar{X}_n + s_n(1 + n^{-1})^{1/2}J_{n-1}^{-1}(\alpha)], \quad (2.11)$$

the classical answer in this situation. Clearly $\text{CP}(D_{n,1}|\theta) = \alpha$ and

$$\text{CP}(D_{n,1}|Y_n, \theta) = \alpha + O_p(n^{-1/2}) \quad (2.12)$$

by extending the reasoning for (2.7). In fact, both $n^{1/2}[\text{CP}(D_n|Y_n, \theta) - \alpha]$ and $n^{1/2}[\text{CP}(D_{n,1}|Y_n, \theta) - \alpha]$ converge weakly to the same normal limit distribution. This phenomenon will be explained more generally in section 4.

2.3 Monte Carlo Approximations

In a few cases, such as Example 2, $D_{n,1}$ can be constructed analytically. In a larger class of problems, it is possible to approximate $D_{n,1}$ by using asymptotic expansions for $H_n(\cdot, \hat{\theta}_n)$ and, if necessary, for $A_n(\cdot, \hat{\theta}_n, Y_n)$. The most general approach approximates the first or both of these cdf's by Monte Carlo methods. Algorithm 1 below assumes that $A_n(\cdot, \hat{\theta}_n, Y_n)$ can be found analytically and gives an approximation for $H_n(\cdot, \hat{\theta}_n)$. Algorithm 2 provides approximations for both cdf's. Both algorithms are bootstrap algorithms in the sense that they rely on sampling from fitted models.

The following representations for $H_n(\cdot, \hat{\theta}_n)$ and $A_n(\cdot, \hat{\theta}_n, Y_n)$ underlie the two algorithms. Let \tilde{X} be a random variable whose conditional distribution, given Y_n , is $P_{\hat{\theta}_n}(\cdot | Y_n)$. Then

$$A_n(x, \hat{\theta}_n, Y_n) = \Pr[R_n(\tilde{X}, Y_n) \leq x | Y_n]. \quad (2.13)$$

Let (X^*, Y_n^*) be random variables whose conditional joint distribution, given Y_n , is $P_{\hat{\theta}_n}$. Let $\theta_n^* = \hat{\theta}_n(Y_n^*)$. Then

$$H_n(x, \hat{\theta}_n) = \Pr[A_n\{R_n(X^*, Y_n^*), \theta_n^*, Y_n^*\} \leq x | Y_n]. \quad (2.14)$$

Let \tilde{X}^* be a random variable whose conditional distribution, given X^*, Y_n^* and Y_n , is $P_{\theta_n^*}(\cdot | Y_n^*)$. In view of (2.13),

$$A_n[R_n(X^*, Y_n^*), \theta_n^*, Y_n^*] = \Pr[R_n(\tilde{X}^*, Y_n^*) \leq R_n(X^*, Y_n^*) | X^*, Y_n^*, Y_n]. \quad (2.15)$$

Representation (2.14) is the basis for the bootstrap Algorithm 1 below. Representations (2.13), (2.14) and (2.15) are the foundation for the double bootstrap Algorithm 2.

Algorithm 1. (A_n is known, H_n to be found). Draw J bootstrap samples $\{(X_j^*, Y_{n,j}^*) : 1 \leq j \leq J\}$ from the fitted distribution $P_{\hat{\theta}_n}$ for (X, Y_n) . These bootstrap samples are conditionally independent, given Y_n . For each j , calculate $\theta_{n,j}^* = \hat{\theta}_n(Y_{n,j}^*)$ and $R_{n,j}^* = R_n(X_j^*, Y_{n,j}^*)$. The empirical cdf of the values $\{A_n(R_{n,j}^*, \theta_{n,j}^*, Y_{n,j}^*) : 1 \leq j \leq J\}$ approximates $H_n(\cdot, \hat{\theta}_n)$ for sufficiently large J .

Algorithm 2. (Both A_n and H_n to be found). Draw K bootstrap variables $\{\tilde{X}_k : 1 \leq k \leq K\}$ from the fitted conditional distribution $P_{\hat{\theta}_n}(\cdot | Y_n)$ for X . These bootstrap variables are conditionally independent, given Y_n . The empirical cdf of the values

$\{R_n(\tilde{X}_k, Y_n): 1 \leq k \leq K\}$ approximates $A_n(\cdot, \hat{\theta}_n, Y_n)$ for sufficiently large K .

Draw J bootstrap samples $\{(X_j^*, Y_{n,j}^*): 1 \leq j \leq J\}$ as in Algorithm 1 and calculate $\theta_{n,j}^* = \hat{\theta}_n(Y_{n,j}^*)$ for each j . Then, for each j , draw K bootstrap variables $\{\tilde{X}_{k,j}^*: 1 \leq k \leq K\}$ from the fitted conditional distribution $P_{\theta_{n,j}^*}(\cdot | Y_{n,j}^*)$. These bootstrap variables are conditionally independent given Y_n and the $\{(X_j^*, Y_{n,j}^*)\}$. Let Z_j be the proportion of the values $\{R_n(\tilde{X}_{k,j}^*, Y_{n,j}^*): 1 \leq k \leq K\}$ which are less than or equal to $R_n(X_j^*, Y_{n,j}^*)$. The empirical cdf of the $\{Z_j: 1 \leq j \leq J\}$ approximates $H_n(\cdot, \hat{\theta}_n)$ for sufficiently large J and K .

A SUN 3/140 workstation suffices to carry out Algorithms 1 and 2 for J and K near 1000 and modest sample size n . Both algorithms rely on simple random sampling with replacement. More efficient algorithm based on importance sampling are likely future developments.

Example 1 (continued). Table 1 reports some results from a simulation study of three prediction intervals in the gaussian autoregressive model (1.3). For the root $R_n = X$, the prediction region D_n defined by (2.1) is simply the one-sided prediction interval (1.5). The corresponding conditional cdf $A_n(x, \hat{\theta}_n, Y_n)$ has the analytical expression (3.5). Algorithm 1 produced the bootstrap approximation to $H_n^{-1}(\alpha, \hat{\theta}_n)$ that is needed to construct the calibrated prediction interval $D_{n,1}$. The third prediction interval considered was

$$D_n'(\alpha) = D_n[\alpha + (2n)^{-1} z_\alpha \phi(z_\alpha)]. \quad (2.15)$$

This interval is an analytical approximation to $D_{n,1}$ for this example, as will be explained in section 4.

In the study comparing D_n , $D_{n,1}$ and D_n' , the value of α was .90 and $\hat{\theta}_n$ was defined by (1.4) with ϵ being 10^{-6} . The numerical entries in Table 1 support several conclusions:

(a) For each θ , the overall coverage probability of $D_{n,1}$ or D_n' is closer to α than is the overall coverage probability of D_n . All three overall coverage probabilities converge rapidly to α as the size n of the learning sample increases.

(b) For each θ , the standard deviation of the conditional coverage probability is nearly the same for D_n , $D_{n,1}$ and D_n' . All three standard deviations tend to zero as n increases.

Asymptotics supporting these conclusions are developed in the next two sections. Of course, the asymptotics do not fully explain the remarkably good small sample performance of D_n' and $D_{n,1}$ in Example 1.

Table 1. Overall Coverage Probabilities and Standard Deviations of Conditional Coverage Probabilities for D_n , $D_{n,1}$ and D_n' in Example 1.

n	θ	CP (\cdot θ) for			S.D. of CP (\cdot Y_n, θ) for		
		D_n	$D_{n,1}$	D_n'	D_n	$D_{n,1}$	D_n'
3	.9	.876	.902	.916	.110	.102	.097
	.5	.858	.889	.899	.149	.142	.130
	.1	.851	.882	.892	.167	.162	.147
	0.	.850	.881	.891	.169	.164	.149
	-.1	.850	.881	.891	.169	.164	.149
	-.5	.856	.887	.897	.150	.144	.131
	-.9	.878	.904	.918	.102	.092	.088
5	.9	.884	.897	.907	.081	.076	.074
	.5	.875	.892	.898	.116	.113	.106
	.1	.874	.894	.897	.122	.118	.113
	0.	.874	.893	.897	.122	.119	.112
	-.1	.874	.894	.897	.121	.118	.112
	-.5	.875	.893	.899	.113	.110	.103
	-.9	.885	.898	.908	.081	.076	.072

NOTE: The intended coverage probability is .90. $D_{n,1}$ uses 999 bootstrap samples. The table entries are calculated from 2500 Monte Carlo trials.

3. CONVERGENCE OF COVERAGE PROBABILITIES

This section establishes conditions under which the conditional and overall coverage probabilities of D_n and $D_{n,1}$ converge to α , as the size of the learning sample increases. The first restriction is to assume that

$$CP(D_n | Y_n, \theta) = C(\alpha, \theta, \hat{\theta}_n, U_n), \quad (3.1)$$

where the function C does not depend on n and $U_n = U_n(Y_n)$ is a statistic. This situation arises frequently in examples, as will be shown later in the section. Suppose the parameter space is metric. Let $C(\theta)$ denote the set of sequences $\{\theta_n: n \geq 1\}$ in the parameter space such that θ_n converges to θ . Let $L(U_n | \theta)$ stand for the distribution of U_n under Q_θ .

Proposition 1. Suppose (3.1) holds, $C(\alpha, \theta, t, u)$ is continuous in (θ, t, u) at points where $t = \theta$, and $C(\alpha, \theta, \theta, u) = \alpha$ for all possible values of (θ, u) . Suppose that for every sequence $\{\theta_n\}$ in $C(\theta)$, $\hat{\theta}_n \rightarrow \theta$ in $Q_{\theta, n}$ probability and $\{L(U_n | \theta_n)\}$ is tight. Then, for every sequence $\{\theta_n\}$ in $C(\theta)$,

$$CP(D_n | Y_n, \theta_n) \rightarrow \alpha \text{ in } Q_{\theta, n} \text{ probability} \quad (3.2)$$

and

$$CP(D_n | \theta_n) \rightarrow \alpha. \quad (3.3)$$

These conclusions also extend to $D_{n,1}$ if $C(\alpha, \theta, t, u)$ is continuous in (α, θ, t, u) at points where $t = \theta$.

The proof of this result is in section 6. The uniformity of convergences (3.2) and (3.3) over compact subsets of the parameter space enhances their trustworthiness. The following examples illustrate the scope of Proposition 1.

Example 1 (continued). Suppose the root for prediction in this autoregressive model is

$$R_n(X, Y_n) = X. \quad (3.4)$$

The conditional cdf of R_n given X_n is

$$A_n(x, \theta, Y_n) = \Phi(x - \theta X_n). \quad (3.5)$$

Let $\hat{\theta}_n$ be the consistent estimate of θ given in (1.4). The prediction region D_n defined by (2.1) coincides with the one-sided prediction interval already described in (1.5). Proposition 1 is applicable with $U_n = X_n$ and

$$C(x, \theta, t, u) = \Phi[\Phi^{-1}(x) + (t - \theta)u]. \quad (3.6)$$

A two-sided prediction interval for X can be obtained in two ways: Combine the one-sided interval (1.5) with the analogous one-sided interval based on the root $-X$, each of nominal coverage probability $(1 + \alpha)/2$; or use in (2.1) the root

$$R_n(X, Y_n) = |X - \hat{\theta}_n X_n| \quad (3.7)$$

where $\hat{\theta}_n X_n$ is the usual point predictor of X . Let $\Psi(x) = \Phi(x) - \Phi(-x)$ be the cdf of the folded-over standard normal distribution. Because the model is gaussian, both methods just described yield the prediction interval

$$D_n = [\hat{\theta}_n X_n - \Psi^{-1}(\alpha), \hat{\theta}_n X_n + \Psi^{-1}(\alpha)] \quad (3.8)$$

Proposition 1 is again applicable, this time with $U_n = X_n$ and

$$C(x, \theta, t, u) = \Phi[\Psi^{-1}(x) + (t - \theta)u] - \Phi[-\Psi^{-1}(x) + (t - \theta)u]. \quad (3.9)$$

Example 3. Suppose the learning sample $Y_n = (X_1, \dots, X_n)$ follows the linear model

$$X_i = \beta c_i + E_i, \quad (3.10)$$

where the $\{c_i\}$ are known constants. The goal is to predict

$$X = \beta c + E, \quad (3.11)$$

given Y_n and the assumption that the errors E , $\{E_i\}$ are i.i.d. with continuous cdf F , which has mean zero and finite variance. The unknown parameter $\theta = (\beta, F)$ is estimated by $\hat{\theta}_n = (\hat{\beta}_n, \hat{F}_n)$, where $\hat{\beta}_n$ is the least squares estimate of β and \hat{F}_n is the empirical cdf of the residuals $\{X_i - \hat{\beta}_n c_i : 1 \leq i \leq n\}$

The root

$$R_n(X, Y_n) = X \quad (3.12)$$

in (2.1) generates the one-sided prediction interval

$$D_n = (-\infty, \hat{F}_n^{-1}(\alpha) + \hat{\beta}_n c] \quad (3.13)$$

for X . Define the distance between $\theta = (\beta, F)$ and the element $t = (b, G)$ of the parameter space by

$$d(t, \theta) = |b - \beta| + m(G, F), \quad (3.14)$$

where m is bounded Lipschitz distance. Assume $\max_i c_i^2 / \sum_1^n c_i^2$ tends to zero and $\sum_1^n c_i^2$ tends to infinity as n increases. Then the estimate $\hat{\theta}_n$ described above is consistent for θ . Proposition 1 is applicable with

$$C(x, \theta, t) = F[G^{-1}(x) + (b - \beta)c]. \quad (3.15)$$

The continuity required of the function C is immediate here because weak convergence of cdf's to a continuous cdf implies uniform convergence.

The root

$$R_n(X, Y_n) = |X - \hat{\beta}_n c| \quad (3.16)$$

in (2.1) yields the two-sided prediction interval

$$D_n = [\hat{\beta}_n c - \hat{G}_n^{-1}(\alpha), \hat{\beta}_n c + \hat{G}_n^{-1}(\alpha)], \quad (3.17)$$

where $\hat{G}_n(x) = \hat{F}_n(x) - \hat{F}_n(-x)$ is the estimated cdf of $|E|$. The application of Proposition 1 to this case is similar to that of the previous paragraph.

4. FURTHER ASYMPTOTICS

The asymptotic distributions of $CP(D_n|Y_n, \theta)$ or $CP(D_{n,1}|Y_n, \theta)$ and the rates of convergence to α of $CP(D_n|\theta)$ or $CP(D_{n,1}|\theta)$ are the topics of this section. Attention is focused on the simplest case: the learning sample Y_n and the variable X to be predicted are independent. The distribution of Y_n is $Q_{\theta,n}$ and the distribution of X is P_θ . The prediction root has the form $R(X, \hat{\theta}_n)$, where $\hat{\theta}_n = \hat{\theta}_n(Y_n)$ is a consistent estimate of θ . The parameter space in the exposition will be an open subset of the real line. The extension to Euclidean parameter spaces is straightforward, but requires heavier notation.

In this setting, the conditional cdf of the root R given Y_n is

$$A(x, \theta, \hat{\theta}_n) = P_\theta[R(X, \hat{\theta}_n) \leq x], \quad (4.1)$$

where $\hat{\theta}_n$ is held fixed on the right side. The conditional coverage probability, given Y_n , of the prediction region D_n generated by (2.1) is thus a function $C(\alpha, \theta, \hat{\theta}_n)$, which is explicitly $A[A^{-1}(\alpha, \hat{\theta}_n, \hat{\theta}_n), \theta, \hat{\theta}_n]$. Note that $C(\alpha, \theta, t)$ is a cdf in its first argument. It will be assumed throughout that

$$C(\alpha, \theta, \theta) = \alpha \quad (4.2)$$

for every possible θ . This occurs, in particular, when $A(x, \theta, t)$ is continuous in x .

4.1 Coverage Probabilities of D_n

The asymptotic behavior of $CP(D_n|Y_n, \theta)$ and $CP(D_n|\theta)$ can be studied by developing expansions for these two quantities. In the discussion, notation like $f^{(i,j,k)}(x, \theta, t)$ represents the partial derivative $\partial^{i+j+k} f(x, \theta, t) / \partial x^i \partial \theta^j \partial t^k$. Let $C_1(\theta)$ denote the class of all sequences $\{\theta_n\}$ in the parameter space such that $\{n^{1/2}(\theta_n - \theta)\}$ converges to a finite limit. Let $\langle x \rangle$ denote the integer part of x . The following

assumptions describe the leading case.

Assumption A(r). For $1 \leq j \leq r$ and r an even integer, there exist functions $\{a_{j,k}(\theta)\}$ such that, as n increases,

$$n^{r/2} [E_{\theta_n}(\hat{\theta}_n - \theta_n)^j - \sum_{k=\langle (j+1)/2 \rangle}^{r/2} n^{-k} a_{j,k}(\theta_n)] \rightarrow 0 \quad (4.3)$$

whenever $\{\theta_n\}$ belongs to $C_1(\theta)$.

Assumption B. If $\{\theta_n\}$ belongs to $C_1(\theta)$, then the distribution of $n^{1/2}(\hat{\theta}_n - \theta_n)$ under Q_{θ_n} converges weakly to the $N(0, a_{2,1}(\theta))$ distribution.

These assumptions are satisfied, for example, when $\{Q_{\theta,n}\}$ is a smoothly parametrized exponential family and $\hat{\theta}_n$ is the maximum likelihood estimate of θ . In this situation, $\hat{\theta}_n$ is a smooth function of a sample mean and (4.3) follows by a Taylor expansion argument (compare Theorem 5.1 on p. 101 of Lehmann, 1983). The asymptotic normality of Assumption B is a consequence of the central limit theorem.

Proposition 2A. Suppose Assumption B and (4.2) hold. Suppose $C^{(0,0,1)}(\alpha, \theta, t)$ exists and is continuous in (θ, t) at points where $t = \theta$. Then, for every sequence $\{\theta_n\}$ in $C_1(\theta)$,

$$L[n^{1/2}\{CP(D_n|Y_n, \theta_n) - \alpha\}|\theta_n] \Rightarrow N(0, \sigma^2(\theta)) \quad (4.4)$$

where

$$\sigma^2(\theta) = [C^{(0,0,1)}(\alpha, \theta, \theta)]^2 a_{2,1}(\theta). \quad (4.5)$$

In the extension of Proposition 2A to vector parameter θ , $C^{(0,0,1)}(\alpha, \theta, t)$ is a column vector, $a_{2,1}(\theta)$ is a matrix, and

$$\sigma^2(\theta) = [C^{(0,0,1)}(\alpha, \theta, \theta)]' a_{2,1}(\theta) [C^{(0,0,1)}(\alpha, \theta, \theta)] \quad (4.6)$$

Proposition 2B. Suppose Assumption A(2) and (4.2) hold, and $a_{2,1}(\theta)$ is continuous in θ . Suppose $C^{(0,0,2)}(\alpha, \theta, t)$ exists, is bounded in (θ, t) , and is continuous in (θ, t) at points where $t = \theta$. Then, for every sequence $\{\theta_n\}$ in $C_1(\theta)$,

$$n[CP(D_n|\theta_n) - \alpha - n^{-1}b_1(\alpha, \theta_n)] \rightarrow 0 \quad (4.7)$$

where

$$b_1(\alpha, \theta) = \sum_{j=1}^2 a_{j,1}(\theta) C^{(0,0,j)}(\alpha, \theta, \theta) / j! \quad (4.8)$$

Propositions 2A and 2B are proved in Section 6. The following example illustrates their content.

Example 2 (continued). In this example, the parameter $\theta = (\mu, \sigma^2)$ is two-dimensional, $t = (m, s^2)$ and $C(\alpha, \theta, t)$ is

$$\begin{aligned} C(\alpha, \theta, t) &= \Phi[z_\alpha + \sigma^{-1} \{(s - \sigma) z_\alpha + (m - \mu)\}] \\ &= \alpha + \delta \phi(z_\alpha) + 2^{-1} \delta^2 \phi'(\bar{z}), \end{aligned} \quad (4.9)$$

where

$$\delta = \sigma^{-1} [(s - \sigma) z_\alpha + (m - \mu)] \quad (4.10)$$

and \bar{z} lies between z_α and $z_\alpha + \delta$. The reasoning for Proposition 2A and 2B applies, with $\hat{\theta}_n = (\bar{X}_n, s_n^2)$. Thus, $CP(D_n | Y_n, \theta_n) = C(\alpha, \theta_n, \hat{\theta}_n)$ is asymptotically normal as in (4.4) with

$$\sigma^2(\theta) = (2^{-1} z_\alpha^2 + 1) \phi^2(z_\alpha). \quad (4.11)$$

Moreover, taking expectations through the expansion for $CP(D_n | Y_n, \theta_n)$ implied by (4.9) establishes conclusion (4.7) with

$$\begin{aligned} b_1(\alpha, \theta) &= -4^{-1} z_\alpha \phi(z_\alpha) + 2^{-1} (2^{-1} z_\alpha^2 + 1) \phi'(z_\alpha) \\ &= -4^{-1} (z_\alpha^3 + 3z_\alpha) \phi(z_\alpha). \end{aligned} \quad (4.12)$$

This argument is an alternative derivation for (2.9).

The next two examples illustrate what can happen in situations where the reasoning for Proposition 2A and 2B does not apply, in some respect.

Example 1 (continued). For root (3.4) in this AR(1) model, the expression (3.6) entails

$$\begin{aligned} CP(D_n | Y_n, \theta_n) &= \Phi[z_\alpha + (\hat{\theta}_n - \theta_n) X_n] \\ &= \alpha + \delta_n \phi(z_\alpha) + 2^{-1} \delta_n^2 \phi'(z_n) \end{aligned} \quad (4.13)$$

where δ_n represents $(\hat{\theta}_n - \theta_n) X_n$ and z_n lies between z_α and $z_\alpha + \delta_n$. Proposition 2A does not apply to this example. Indeed, if $\{\theta_n\}$ lies in $C_1(\theta)$, then $n^{1/2} [CP(D_n | Y_n, \theta_n) - \alpha]$ converges weakly to the product of two independent $N(0, 1)$ random variables and $\phi(z_\alpha)$.

On the other hand, taking expectations through (4.13) establishes

$$CP(D_n|\theta_n) = \alpha - (2n)^{-1} z_\alpha \phi(z_\alpha) + o(n^{-1}) \quad (4.14)$$

By symmetry, the expectation of δ_n is zero in this gaussian model. While Proposition 2B is not applicable to this example, conclusion (4.14) is analogous to (4.7).

Example 4. Suppose X and the elements of $Y = (X_1, \dots, X_n)$ are iid random variables with unknown continuous cdf F . This cdf, which is the parameter θ here, is estimated by the empirical cdf. Let $X_{(1)} \leq \dots \leq X_{(n)}$ denote the order statistics of the learning sample. Define $X_{(n+1)}$ to be ∞ . From the root $R_n = X$, definition (2.1) generates the one-sided prediction interval

$$D_n = (-\infty, X_{(<\alpha n + 1>)}], \quad (4.15)$$

where $<\cdot>$ is the integer part function. Evidently, $CP(D_n|Y_n, \theta)$ equals $F(X_{(<\alpha n + 1>)})$ for every continuous F . Since the i th uniform order statistic has a $Beta(i, n - i + 1)$ distribution, it follows that

$$n^{1/2} [CP(D_n|Y_n, \theta) - \alpha] \Rightarrow N(0, \alpha(1 - \alpha)) \quad (4.16)$$

and

$$\begin{aligned} CP(D_n|\theta) &= (n + 1)^{-1} <\alpha n + 1> \\ &= \alpha + O(n^{-1}) \end{aligned} \quad (4.17)$$

Both convergences are uniform over all continuous F . These conclusions parallel Propositions 2A and 2B, though for different reasons.

Proposition 2B has several implications and extensions:

(a) Two plausible estimates for $CP(D_n|\theta)$ are the plug-in estimate $CP(D_n|\hat{\theta}_n)$ and the naive estimate $\hat{CP} = \alpha$. Under the assumptions of Proposition 2B,

$$n[CP(D_n|\hat{\theta}_n) - \alpha - n^{-1} b_1(\alpha, \hat{\theta}_n)] \rightarrow 0 \quad (4.18)$$

in $Q_{\theta_n, n}$ probability. Consequently,

$$n[CP(D_n|\hat{\theta}_n) - CP(D_n|\theta_n)] \rightarrow 0 \quad (4.19)$$

in $Q_{\theta_n, n}$ probability, provided $b_1(\alpha, \theta)$ is continuous in θ . By contrast, the naive estimate satisfies

$$n[\hat{CP} - CP(D_n|\theta_n)] \rightarrow -b_1(\alpha, \theta), \quad (4.20)$$

and so is less efficient than the plug-in estimate. This fact and (2.5) are the bases for the calibration procedure which generates $D_{n,1}$ from D_n .

(b) Suppose $C^{(0,0,2)}(\alpha, \theta, t)$ is bounded in all three arguments and continuous at points where $t = \theta$, in addition to the assumptions of Proposition 2B. Suppose also that $b_1(\alpha, \theta)$ is continuous in both arguments. Then

$$n [H_n^{-1}(\alpha, \hat{\theta}_n) - \alpha + n^{-1} b_1(\alpha, \hat{\theta}_n)] \rightarrow 0 \quad (4.21)$$

in $Q_{\theta, n}$ probability, for every sequence $\{\theta_n\}$ in $C_1(\theta)$.

(c) Higher order expansions for $H_n(\alpha, \theta)$ and $H_n^{-1}(\alpha, \theta)$ exist under stronger regularity conditions. For example, suppose Assumption A(4) and (4.2) hold, $a_{4,2}(\theta)$ is continuous, and $C^{(0,0,4)}(\alpha, \theta, t)$ is bounded in all three arguments and is continuous at points where $t = \theta$. Then, for every convergent sequence $\{\alpha_n\}$ in the unit interval and every sequence $\{\theta_n\}$ in $C_1(\theta)$,

$$n^2 [H_n(\alpha_n, \theta_n) - \alpha_n - \sum_{i=1}^2 n^{-i} b_i(\alpha_n, \theta_n)] \rightarrow 0, \quad (4.22)$$

where b_1 is given by (4.8) and

$$b_2(\alpha, \theta) = \sum_{j=1}^4 a_{j,2}(\theta) C^{(0,0,j)}(\alpha, \theta, \theta) / j! \quad (4.23)$$

Moreover, if $b_2(\alpha, \theta)$ and the derivative $b_1^{(1,0)}(\alpha, \theta)$ are continuous in both arguments, then

$$n^2 [H_n^{-1}(\alpha, \hat{\theta}_n) - \alpha - \sum_{i=1}^2 n^{-i} c_i(\alpha, \hat{\theta}_n)] \rightarrow 0 \quad (4.24)$$

in $Q_{\theta, n}$ probability, where

$$\begin{aligned} c_1(\alpha, \theta) &= -b_1(\alpha, \theta) \\ c_2(\alpha, \theta) &= b_1(\alpha, \theta) b_1^{(1,0)}(\alpha, \theta) - b_2(\alpha, \theta). \end{aligned} \quad (4.25)$$

4.2 Coverage Probabilities of $D_{n,1}$

From the definition (2.6) of $D_{n,1}$, it is immediate that

$$CP(D_{n,1} | Y_n, \theta) = C[H_n^{-1}(\alpha, \hat{\theta}_n), \theta, \hat{\theta}_n]. \quad (4.26)$$

Consequently, the behavior of $H_n^{-1}(\alpha, \hat{\theta}_n)$ strongly influences the conditional and overall coverage probabilities of $D_{n,1}$. The discussion surrounding (4.21) and (4.24) motivates the assumptions made about $H_n^{-1}(\alpha, \hat{\theta}_n)$ in this section.

Proposition 3A. Suppose Assumption B and (4.2) hold. Suppose $C^{(0,0,1)}(\alpha, \theta, t)$ exists and is continuous in all three arguments at points where $t = \theta$. Suppose that

$$n [H_n^{-1}(\alpha, \hat{\theta}_n) - \alpha] \rightarrow 0 \quad (4.27)$$

in Q_{θ_n} probability for every sequence $\{\theta_n\}$ in $C_1(\theta)$. Then

$$L[n^{1/2}\{CP(D_{n,1}|Y_n, \theta_n) - \alpha\}|\theta_n] \Rightarrow N(0, \sigma^2(\theta)) \quad (4.28)$$

where $\sigma^2(\theta)$ is defined by (4.5).

Comparing Propositions 2A and 3A reveals that both $CP(D_n|Y_n, \theta_n)$ and $CP(D_{n,1}|Y_n, \theta_n)$ have the *same* normal limiting distribution whose asymptotic variance is proportional to the asymptotic variance of $\hat{\theta}_n$. Thus, the estimate $\hat{\theta}_n$ should be asymptotically efficient to minimize the asymptotic dispersion about α of $CP(D_n|Y_n, \theta)$ or $CP(D_{n,1}|Y_n, \theta)$. This statement can be formalized in a lower asymptotic minimax bound and in a convolution theorem, which will be presented elsewhere.

As the next result shows, calibration of D_n achieves a bias correction of order n^{-1} in the conditional coverage probability of D_n . Let

$$r_n(\alpha, \theta) = n^2[H_n^{-1}(\alpha, \theta) - \alpha + n^{-1}b_1(\alpha, \theta)] \quad (4.29)$$

for b_1 defined as in (4.8).

Proposition 3B. Suppose Assumption A(2) and (4.2) hold. Suppose $C^{(1,0,0)}(\alpha, \theta, t)$, $C^{(1,0,1)}(\alpha, \theta, t)$, $C^{(2,0,0)}(\alpha, \theta, t)$, $b_1^{(0,1)}(\alpha, \theta)$, $b_1^{(0,2)}(\alpha, \theta)$ exist and are bounded in all of their arguments at points where $t = \theta$. Suppose that for every sequence $\{\theta_n\}$ in $C_1(\theta)$,

$$\limsup_{n \rightarrow \infty} |r_n(\alpha, \theta_n)| < \infty \quad (4.30)$$

and

$$\limsup_{n \rightarrow \infty} E_{\theta_n} |r_n(\alpha, \hat{\theta}_n)| < \infty. \quad (4.31)$$

Then

$$CP(D_{n,1}|\theta_n) = \alpha + O(n^{-2}). \quad (4.32)$$

Propositions 3A and 3B are proved in Section 6. Some extensions are possible:

(a) A plausible analytical approximation to $D_{n,1}$ is the prediction region

$$D_n'(\alpha) = D_n[\alpha - n^{-1}b_1(\alpha, \hat{\theta}_n)]. \quad (4.33)$$

If $\{\theta_n\}$ is any sequence in $C_1(\theta)$, it follows from the proofs for Propositions 3B and 3A that

$$CP(D_n'|\theta_n) = \alpha + O(n^{-2}) \quad (4.34)$$

and

$$L [n^{1/2} \{ CP(D_n' | Y_n, \theta_n) - \alpha \} | \theta_n] \Rightarrow N(0, \sigma^2(\theta)), \quad (4.35)$$

like prediction region $D_{n,1}$. Nevertheless, as shown below in the discussion of example 2, prediction regions $D_{n,1}$ and D_n' can still differ significantly. The analytical calibration idea that underlies D_n' is related to Cox (1975).

(b) Calibration of D_n can be iterated. Let $H_{n,1}(\alpha, \theta) = CP[D_{n,1}(\alpha) | \theta]$ and define

$$\begin{aligned} D_{n,2}(\alpha) &= D_{n,1}[H_{n,1}^{-1}(\alpha, \hat{\theta}_n)] \\ &= D_n\{H_n^{-1}[H_{n,1}^{-1}(\alpha, \hat{\theta}_n), \hat{\theta}_n]\}. \end{aligned} \quad (4.36)$$

Under stronger regularity conditions, the conclusion of Proposition 3A extends to $CP(D_{n,2} | Y_n, \theta_n)$ and

$$CP(D_{n,2} | \theta_n) = \alpha + O(n^{-3}). \quad (4.37)$$

In other words, $D_{n,2}$ achieves a second-order bias correction to $CP(D_n | Y_n, \theta_n)$. A nested double bootstrap Monte Carlo algorithm can be used to approximate $H_{n,1}(\alpha, \hat{\theta}_n)$ and so the critical value of $D_{n,2}$.

Thus, for all sufficiently large n , the coverage probability of $D_{n,2}$ is more accurate than that of $D_{n,1}$, which in turn is more accurate than the coverage probability of D_n . For *fixed* n , calibration and iterated calibration may or may not improve coverage probability. Examples 1 and 2, the numerical study reported in section 2.3, and experience with asymptotic corrections in other problems suggest that at least one round of calibration is often worthwhile. On the other hand, the discussion below of example 4 shows that regularity conditions are essential for the success of calibration.

Example 2 (continued). In this normal location-scale model, the function $b_1(\alpha, \theta)$ is given by (4.12). Let

$$\alpha_n = \alpha + (4n)^{-1} (z_\alpha^3 + 3z_\alpha) \phi(z_\alpha) \quad (4.38)$$

so that D_n' is just $D_n(\alpha_n)$. From (2.9) it is immediate that

$$\begin{aligned} CP(D_n' | \theta) &= \alpha_n - 4^{-1} n^{-1} (z_{\alpha_n}^3 + 3z_{\alpha_n}) \phi(z_{\alpha_n}) + O(n^{-2}) \\ &= \alpha + O(n^{-2}) \end{aligned} \quad (4.39)$$

uniformly in θ , as expected. However, in this example $CP(D_{n,1} | \theta)$ equals α exactly, as discussed after (2.11). Both $n^{1/2} [CP(D_n' | Y_n, \theta) - \alpha]$ and $n^{1/2} [CP(D_{n,1} | Y_n, \theta) - \alpha]$ converge weakly to a $N(0, \sigma^2(\theta))$ distribution, the variance being given by (4.11). This example illustrates the extension of Propositions 3A and 3B to vector θ and to D_n' .

Example 4 (continued). Success of the calibration adjustment requires continuity of $H_n(\alpha, \theta)$ in α . Such continuity is missing in example 4, where by (4.17),

$$H_n(\alpha, \theta) = (n+1)^{-1} \langle \alpha n + 1 \rangle,$$

the discrete uniform distribution supported on the values $\{n^{-1}j: 0 \leq j \leq n\}$. In this instance,

$$\begin{aligned} H_n^{-1}(\alpha, \theta) &= n^{-1} \langle \alpha(n+1) \rangle \quad \text{if } 0 \leq \alpha < 1 \\ &= \infty \quad \text{if } \alpha = 1. \end{aligned} \quad (4.41)$$

Consequently, for α less than 1,

$$D_{n,1} = (-\infty, X_{(\beta)}], \quad (4.42)$$

where β is $\langle \alpha(n+1) \rangle + 1$. It follows that

$$n^{1/2} [\text{CP}(D_{n,1} | Y_n, \theta) - \alpha] \Rightarrow N(0, \alpha(1-\alpha)) \quad (4.43)$$

and

$$\text{CP}(D_{n,1} | \theta) = \alpha + O(n^{-1}) \quad (4.44)$$

uniformly over all continuous F . In this nonparametric case, $D_{n,1}$ is no better than D_n , in either conditional or overall coverage probability.

Example 1 (continued). For the gaussian AR(1) model, with prediction root $R_n = X$, the coverage probability calculation (4.14) and the reasoning behind (4.33) yield the analytically calibrated prediction interval.

$$D_n'(\alpha) = D_n[\alpha + (2n)^{-1} z_\alpha \phi(z_\alpha)], \quad (4.45)$$

where D_n is given by (1.5). By direct arguments based on (4.14), both $\text{CP}(D_{n,1} | \theta_n)$ and $\text{CP}(D_n' | \theta_n)$ equal $\alpha + o(n^{-1})$ while $\text{CP}(D_n | \theta_n)$ is $\alpha + O(n^{-1})$. Moreover, the asymptotic variance of $n^{1/2} [\text{CP}(\cdot | Y_n, \theta_n) - \alpha]$ is $\phi^2(z_\alpha)$ for each of the prediction intervals D_n , $D_{n,1}$ and D_n' . These asymptotics are consistent with the simulation results reported in section 2.2. They indicate that calibration works beyond the setting of Propositions 3B and 3A.

5. PROOFS

This section proves the Propositions stated earlier in the paper.

Proof of Proposition 1. Suppose (3.2) does not hold. By going to a subsequence assume without loss of generality that

$$Q_{\theta_{m,n}}[|\text{CP}(D_n | Y_n, \theta_n) - \alpha| > \varepsilon] > \delta \quad (5.1)$$

for some sequence $\{\theta_n\} \in C(\theta)$ and some positive (ϵ, δ) . By going to a further subsequence, assume without loss of generality that $\hat{\theta}_n \rightarrow \theta$ in $Q_{\theta_n, n}$ probability and that U_n converges weakly under $Q_{\theta_n, n}$ to a random variable U . For Skorokhod versions of the random variables involved, $(\hat{\theta}_n, U_n)$ converges almost surely to (θ, U) , by Wichura (1970). From this and the assumed properties of C , it follows that $C(\alpha, \theta_n, \hat{\theta}_n, U_n) \rightarrow \alpha$ with probability one. This contradicts (5.1) because of (3.1). Hence (3.2) and (3.3) hold.

The argument for the coverage probabilities of $D_{n,1}$ is analogous and uses two additional facts: In view of (3.3), which holds for every possible α , and (2.5), it follows that $H_n^{-1}(\alpha, \theta_n)$ converges to α whenever $\{\theta_n\}$ belongs to $C(\theta)$. Moreover,

$$CP(D_{n,1} | Y_n, \theta_n) = C[H_n^{-1}(\alpha, \hat{\theta}_n), \theta_n, \hat{\theta}_n, U_n]. \quad (5.2)$$

Proof of Proposition 2A. It follows from (4.2) that

$$\begin{aligned} CP(D_n | Y_n, \theta_n) &= C(\alpha, \theta_n, \hat{\theta}_n) \\ &= \alpha + (\hat{\theta}_n - \theta_n) C^{(0,0,1)}(\alpha, \theta_n, \bar{\theta}_n) \end{aligned} \quad (5.3)$$

where $\bar{\theta}_n$ lies between θ_n and $\hat{\theta}_n$. Assumption B implies that $(\theta_n, \bar{\theta}_n)$ converges to (θ, θ) in $Q_{\theta_n, n}$ probability. The asymptotic normality (4.4) thus follows from (5.3), Assumption B, and the continuity assumption on $C^{(0,0,1)}(\alpha, \theta, t)$.

Proof of Proposition 2B. In this case, it follows from (4.2) that

$$\begin{aligned} CP(D_n | Y_n, \theta_n) &= \alpha + (\hat{\theta}_n - \theta_n) C^{(0,0,1)}(\alpha, \theta_n, \theta_n) \\ &\quad + 2^{-1} (\hat{\theta}_n - \theta_n)^2 C^{(0,0,2)}(\alpha, \theta_n, \bar{\theta}_n) \end{aligned} \quad (5.4)$$

where $\bar{\theta}_n$ lies between θ and $\hat{\theta}_n$. By Assumption A(2),

$$\begin{aligned} E_{\theta_n}(\hat{\theta}_n - \theta_n) &= n^{-1} a_{1,1}(\theta_n) + o(n^{-1}) \\ E_{\theta_n}(\hat{\theta}_n - \theta_n)^2 &= n^{-1} a_{2,1}(\theta_n) + o(n^{-1}). \end{aligned} \quad (5.5)$$

The second line in (5.5) and the continuity assumptions on $C^{(0,0,2)}(\alpha, \theta, t)$ and $a_{2,1}(\theta)$ entail

$$n(\hat{\theta}_n - \theta_n)^2 [C^{(0,0,2)}(\alpha, \theta_n, \bar{\theta}_n) - C^{(0,0,2)}(\alpha, \theta_n, \theta_n)] \rightarrow 0 \quad (5.6)$$

in $Q_{\theta_n, n}$ probability. A uniform integrability argument using the boundedness of $C^{(0,0,2)}$ and the continuity of $a_{2,1}(\theta)$ shows that convergence (5.6) also occurs in expectation. The conclusion (4.7) follows from the facts just established, by taking expectations through (5.4).

Proof of Proposition 3A. It follows from (4.2) and (4.26) that

$$CP(D_{n,1}|Y_n, \theta_n) = H_n^{-1}(\alpha, \hat{\theta}_n) + (\hat{\theta}_n - \theta_n) C^{(0,0,1)}[H_n^{-1}(\alpha, \hat{\theta}_n), \theta_n, \bar{\theta}_n], \quad (5.7)$$

where $\bar{\theta}_n$ lies between θ_n and $\hat{\theta}_n$. In view of (4.27), the result follows by the reasoning for Proposition 2A.

Proof of Proposition 3B. Let $\{\theta_n\}$ be any sequence in $C_1(\theta)$. Write K_n for the quantile function H_n^{-1} and define

$$K_{n,o}(\alpha, \theta) = \alpha - n^{-1} b_1(\alpha, \theta). \quad (5.8)$$

By Taylor expansion,

$$C[K_n(\alpha, \hat{\theta}_n), \theta_n, \hat{\theta}_n] = C[K_{n,o}(\alpha, \hat{\theta}_n), \theta_n, \hat{\theta}_n] + n^{-2} r_n(\alpha, \hat{\theta}_n) C^{(1,0,0)}(k_n, \theta_n, \hat{\theta}_n) \quad (5.9)$$

where k_n lies between $K_n(\alpha, \hat{\theta}_n)$ and $K_{n,o}(\alpha, \hat{\theta}_n)$. Since $C^{(1,0,0)}$ is bounded, it follows from (4.26) and (4.31) that

$$CP(D_{n,1}|\theta_n) = E_{\theta_n} C[K_{n,o}(\alpha, \hat{\theta}_n), \theta_n, \hat{\theta}_n] + O(n^{-2}). \quad (5.10)$$

The following three Taylor expansions hold:

$$\begin{aligned} \delta_n &\equiv K_{n,o}(\alpha, \hat{\theta}_n) - K_{n,o}(\alpha, \theta_n) \\ &= -n^{-1} (\hat{\theta}_n - \theta_n) b_1^{(0,1)}(\alpha, \bar{\theta}_{n,1}) \\ &= -n^{-1} [(\hat{\theta}_n - \theta_n) b_1^{(0,1)}(\alpha, \theta_n) + 2^{-1} (\hat{\theta}_n - \theta_n)^2 b_1^{(0,2)}(\alpha, \bar{\theta}_{n,2})] \end{aligned} \quad (5.11)$$

where $\bar{\theta}_{n,1}$ and $\bar{\theta}_{n,2}$ lie between $\hat{\theta}_n$ and θ_n . Also

$$\begin{aligned} C[K_{n,o}(\alpha, \hat{\theta}_n), \theta_n, \hat{\theta}_n] &= C[K_{n,o}(\alpha, \theta_n), \theta_n, \hat{\theta}_n] \\ &+ \delta_n \{ C^{(1,0,0)}[K_{n,o}(\alpha, \theta_n), \theta_n, \theta_n] + (\hat{\theta}_n - \theta_n) C^{(1,0,1)}[K_{n,o}(\alpha, \theta_n), \theta_n, \bar{\theta}_{n,3}] \} \\ &+ 2^{-1} \delta_n^2 C^{(2,0,0)}(\bar{k}_n, \theta_n, \hat{\theta}_n) \end{aligned} \quad (5.12)$$

where \bar{k}_n lies between $K_{n,o}(\alpha, \hat{\theta}_n)$ and $K_{n,o}(\alpha, \theta_n)$ while $\bar{\theta}_{n,3}$ lies between $\hat{\theta}_n$ and θ_n .

From Assumption A2, (5.10), and the boundedness of the derivatives in (5.11), (5.12), it is apparent that

$$CP(D_{n,1}|\theta_n) = E_{\theta_n} C[K_{n,o}(\alpha, \theta_n), \theta_n, \hat{\theta}_n] + O(n^{-2}). \quad (5.13)$$

On the other hand, (4.30) and an expansion analogous to (5.9) establish

$$\begin{aligned} E_{\theta_n} C[K_{n,o}(\alpha, \theta_n), \theta_n, \hat{\theta}_n] &= E_{\theta_n} C[K_n(\alpha, \theta_n), \theta_n, \hat{\theta}_n] + O(n^{-2}) \\ &= \alpha + O(n^{-2}). \end{aligned} \quad (5.14)$$

The second line in (5.14) relies on the continuity of $H_n(\alpha, \theta)$ in α , itself a consequence of (2.5), the first line in (5.3), and the boundedness of $C^{(1,0,0)}$. Combining

(5.13) with (5.14) completes the proof.

REFERENCES

- Atwood, C.L., "Approximate tolerance intervals based on maximum likelihood estimates," *Journal of the American Statistical Association*, 79 (1984), 459-465.
- Bai, C. and Olshen, R.A., "Discussion of a paper by P. Hall," *The Annals of Statistics* 16 (1988), 953-956.
- Butler, R.W., "Nonparametric interval and point prediction using data trimmed by a Grubbs type outlier rule," *The Annals of Statistics*, 10 (1982), 197-204.
- Butler, R. and Rothman, E.D., "Predictive intervals based on reuse of the sample," *Journal of the American Statistical Association*, 75 (1980), 881-889.
- Cox, D.R., "Prediction intervals and empirical Bayes confidence intervals," in *Perspectives in Probability and Statistics*, ed. J. Gani, London: Academic Press, 1975.
- Cox, D.R., "Discussion of a paper by R.W. Butler," *Journal of the Royal Statistical Society, Ser. B*, 48 (1986), 27.
- Cox, D.R. and Hinkley, D.V., *Theoretical Statistics*, London: Chapman and Hall, 1974.
- Guttman, I., *Statistical Tolerance Regions: Classical and Bayesian*, London: Griffin, 1970.
- Lehmann, E.L., *Theory of Point Estimation*. New York: Wiley, 1983.
- Stine, R.A., "Bootstrap prediction intervals for regression," *Journal of the American Statistical Association*, 80 (1984), 1026-1031.
- Wichura, M.J., "On the construction of almost uniformly convergent random variables with given weakly convergent image laws," *The Annals of Mathematical Statistics*, 41 (1970), 284-291.

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 Hydrosiences, (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. Statistical Science, 1988, Vol.3 No. 2 239-271.
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, 579-611.
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. Scandinavian J. Statist., 1988, 15, 1-23.
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. (June 1987). As others see us: A case study in path analysis. Journal of Educational Statistics, 12, 101-128.
56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.
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. Biometrika, 1987, 74, 799-808.
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. To appear in Statistical Science.
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. (Nov 1986, revised Aug 1988). 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. Ann. Inst. Henri Poincaré, 1987, 23, 397-423.
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. J. Amer. Statist. Assoc., 1988, 83, 744-749.
90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.
91. DIACONIS, P and FREEDMAN, D.A. (April 1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. To appear in the Journal of Applied Probability.
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. CANCELLED
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 carcinogenesis. To appear in Environmental Health Perspectives.
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. Annals of Statistics, June, 1988.
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. IEEE Computer Graphics and applications, June, 1988.
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. (Aug 1987, revised Oct 1988). 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 $E \max_{1 \leq k \leq n} S_k^+ / ES_n^+$: 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. Annals of Statistics, June, 1988.
111. BICKEL, P.J. and GHOSH, J.K. (August 1987, revised June 1988). 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. (Sept. 1987, revised Aug 1988). Large sample theory of estimation in biased sampling regression models I.
116. RITOV, Y. and BICKEL, P.J. (Sept.1987, revised Aug. 1988). Achieving 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. (Oct. 1987, revised Mar. 1988, Oct. 1988). Geometrizing rates of convergence, II.
121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. Statistics a Guide to the Unknown, pp. 249-260 (Eds. J.M. Tanur et al.) Wadsworth, Pacific Grove.
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., MACGIBBON, B. and LIU, R.C. (Nov.1987, revised July 1988). 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, revised June 1988). Optimal stopping of two-parameter processes on nonstandard 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. (Nov.1987, revised Sept. 1988). Uniform error bounds involving logspline models.
131. Same as No. 140
132. HESSE, C.H. (Dec. 1987, revised June 1989). 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. CANCELLED
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. (April 1988). On the uniform consistency of Bayes estimates for multinomial probabilities.
- 137a. DONOHO, D.L. and LIU, R.C. (1987). Geometrizing rates of convergence, I.
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. (Jan. 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. (Feb. 1988, revised Nov. 1988). Randomization in the two-armed bandit problem.
144. DABROWSKA, D.M., DOKSUM, K.A. and SONG, J.K. (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. (Feb.1988, revised August 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. (Mar. 1988, revised May 1989). Asymptotics for multivariate trimming.
152. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On a theorem of Kuchler and Lauritzen.
153. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the problem of types.
154. DOKSUM, K.A. (May 1988). On the correspondence between models in binary regression analysis and survival analysis.
155. LEHMANN, E.L. (May 1988). Jerzy Neyman, 1894-1981.
156. ALDOUS, D.J. (May 1988). Stein's method in a two-dimensional coverage problem.
157. FAN, J. (June 1988). On the optimal rates of convergence for nonparametric deconvolution problem.
158. DABROWSKA, D. (June 1988). Signed-rank tests for censored matched pairs.
159. BERAN, R.J. and MILLAR, P.W. (June 1988). Multivariate symmetry models.
160. BERAN, R.J. and MILLAR, P.W. (June 1988). Tests of fit for logistic models.
161. BREIMAN, L. and PETERS, S. (June 1988). Comparing automatic bivariate smoothers (A public service enterprise).
162. FAN, J. (June 1988). Optimal global rates of convergence for nonparametric deconvolution problem.
163. DIACONIS, P. and FREEDMAN, D.A. (June 1988). A singular measure which is locally uniform. (Revised by Tech Report No. 180).
164. BICKEL, P.J. and KRIEGER, A.M. (July 1988). Confidence bands for a distribution function using the bootstrap.
165. HESSE, C.H. (July 1988). New methods in the analysis of economic time series I.
166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.
167. BREIMAN, L., STONE, C.J. and KOOPERBERG, C. (August 1988). Confidence bounds for extreme quantiles.
168. LE CAM, L. (August 1988). Maximum likelihood an introduction.
169. BREIMAN, L. (Aug.1988, revised Feb. 1989). Submodel selection and evaluation in regression I. The X-fixed case and little bootstrap.
170. LE CAM, L. (September 1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge.
171. STONE, C.J. (September 1988). Large-sample inference for logspline models.
172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.
173. MILLAR, P.W. (October 1988). Optimal estimation in the non-parametric multiplicative intensity model.
174. YOR, M. (October 1988). Interwinings of Bessel processes.
175. ROJO, J. (October 1988). On the concept of tail-heaviness.
176. ABRAHAMS, D.M. and RIZZARDI, F. (September 1988). BLSS - The Berkeley interactive statistical system: An overview.

177. MILLAR, P.W. (October 1988). Gamma-funnels in the domain of a probability, with statistical implications.
178. DONOHO, D.L. and LIU, R.C. (October 1988). Hardest one-dimensional subproblems.
179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signal when the low frequency information is missing.
180. FREEDMAN, D.A. and PITMAN, J.A. (Nov. 1988). A measure which is singular and uniformly locally uniform. (Revision of Tech Report No. 163).
181. DOKSUM, K.A. and HOYLAND, ARNLJOT (Nov. 1988, revised Jan. 1989, Aug. '89). Models for variable stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution.
182. DALANG, R.C., MORTON, A. and WILLINGER, W. (November 1988). Equivalent martingale measures and no-arbitrage in stochastic securities market models.
183. BERAN, R. (November 1988). Calibrating prediction regions.
184. BARLOW, M.T., PITMAN, J. and YOR, M. (Feb. 1989). On Walsh's Brownian Motions.
185. DALANG, R.C. and WALSH, J.B. (Dec. 1988). Almost-equivalence of the germ-field Markov property and the sharp Markov property of the Brownian sheet.
186. HESSE, C.H. (Dec. 1988). Level-Crossing of integrated Ornstein-Uhlenbeck processes
187. NEVEU, J. and PITMAN, J.W. (Feb. 1989). Renewal property of the extrema and tree property of the excursion of a one-dimensional brownian motion.
188. NEVEU, J. and PITMAN, J.W. (Feb. 1989). The branching process in a brownian excursion.
189. PITMAN, J.W. and YOR, M. (Mar. 1989). Some extensions of the arcsine law.
190. STARK, P.B. (Dec. 1988). Duality and discretization in linear inverse problems.
191. LEHMANN, E.L. and SCHOLZ, F.W. (Jan. 1989). Ancillarity.
192. PEMANTLE, R. (Feb. 1989). A time-dependent version of Pólya's urn.
193. PEMANTLE, R. (Feb. 1989). Nonconvergence to unstable points in urn models and stochastic approximations.
194. PEMANTLE, R. (Feb. 1989, revised May 1989). When are touchpoints limits for generalized Pólya urns.
195. PEMANTLE, R. (Feb. 1989). Random walk in a random environment and first-passage percolation on trees.
196. BARLOW, M., PITMAN, J. and YOR, M. (Feb. 1989). Une extension multidimensionnelle de la loi de l'arc sinus.
197. BREIMAN, L. and SPECTOR, P. (Mar. 1989). Submodel selection and evaluation in regression — the X-random case.
198. BREIMAN, L., TSUR, Y. and ZEMEL, A. (Mar. 1989). A simple estimation procedure for censored regression models with known error distribution.
199. BRILLINGER, D.R. (Mar. 1989). Two papers on bilinear systems: a) A study of second- and third-order spectral procedures and maximum likelihood identification of a bilinear system. b) Some statistical aspects of NMR spectroscopy, *Actas del 2° congreso latinoamericano de probabilidad y estadística matemática*, Caracas, 1985.
200. BRILLINGER, D.R. (Mar. 1989). Two papers on higher-order spectra: a) Parameter estimation for nonGaussian processes via second and third order spectra with an application to some endocrine data. b) Some history of the study of higher-order moments and spectra.
201. DE LA PENA, V. and KLASS, M.J. (April 1989). L bounds for quadratic forms of independent random variables.
202. FREEDMAN, D.A. and NAVIDI, W.C. (April 1989). Testing the independence of competing risks.
203. TERDIK, G. (May 1989). Bilinear state space realization for polynomial stochastic systems.
204. DONOHO, D.L. and JOHNSTONE, I.M. (May 1989). Minimax risk over l_p -Balls.
205. PEMANTLE, R., PROPP, J. and ULLMAN, D. (May 1989). On tensor powers of integer programs.

- 206. MILASEVIC, P. and NOLAN, D. (May 1989). Estimation on the sphere: A geometric approach.
- 207. SPEED, T.P. and YU, B. (July 1989). Stochastic complexity and model selection: normal regression.
- 208. DUBINS, L.E. (June 1989). A group decision device: Its pareto-like optimality.
- 209. BREIMAN, L. (July 1989). Fitting additive models to regression data.
- 210. PEMANTLE, R. (July 1989). Vertex-reinforced random walk.
- 211. LE CAM, L. (August 1989). On measurability and convergence in distribution.
- 212. FELDMAN, R.E. (July 1989). Autoregressive processes and first-hit probabilities for randomized random walks.
- 213. DONOHO, D.L., JOHNSTONE, I.M., HOCH, J.C. and STERN, A.S. (August 1989). Maximum entropy and the nearly black object.

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.