Probab. Theory Relat. Fields 95, 125–140 (1993)

Probability Theory and Related Fields

On nondifferentiable functions and the bootstrap*

Lutz Dümbgen

Department of Statistics, University of California, Berkeley, CA 94720, USA

Received January 2, 1992; in revised form June 5, 1992

Summary. We investigate a class of statistical problems, where usual bootstrap methods fail, and discuss two alternative solutions. In particular, a stochastic procedure for constructing confidence sets is proposed. Special applications are the eigenvalues of a covariance matrix and minimum distance functionals.

Mathematics Subject Classifications: 62E20, 62G07, 62G15

1 Introduction

We investigate a class of statistical problems, where usual bootstrap methods fail, and discuss two alternative solutions. One special case are the eigenvalues of a covariance matrix, which is a well-known counterexample of Beran and Srivastava (1985, 1987); see also Eaton and Tyler (1991). Other examples are given in connection with minimum distance procedures.

In general we are dealing with a sequence of statistical models that can be partially described by an unknown parameter t_n in a real Banach space $(\mathbf{X}, \|\cdot\|)$. For each n let $\hat{t}_n$ be an estimator for t_n defined on a probability space (Ω_n, P_n). For the random elements $B_n := \sqrt{n}(\hat{t}_n - t_n)$ we assume that

A1 B_n converges in distribution to a random variable B, where $L := \mathscr{L}(B)$ is a tight Borel distribution on $\mathbf{X}$

(the factors $\sqrt{n}$ could be replaced with any other constants $r_n > 0$ tending to ∞). The second basic assumption is that there are consistent estimators for $\mathscr{L}(B_n)$. That means, there are random distributions $\hat{L}_n = \hat{L}_n(\cdot \mid \omega_n)$ on $\mathbf{X}$, defined for $\omega_n \in \Omega_n$, such that

A2 $\quad\quad$ $\hat{L}_n$ converges weakly to L in probability

* Work supported by the Miller Institute for Basic Research in Science, Berkeley

Present address: Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, W-6900 Heidelberg, Germany

(we refer to Hoffmann-Jørgensen's, 1984, concept of weak convergence; see the appendix for more details). Two examples for A1–2 are given at the end of this section.

It follows from A1–2 that one can construct confidence sets $C_n \subset \mathbf{X}$ for t_n. But suppose that one is merely interested in $\phi(t_n)$, where ϕ is any function from $\mathbf{X}$ into another Banach space $(\mathbf{Y}, \|\cdot\|)$. One might use $\phi(C_n)$ as a confidence set for $\phi(t_n)$. However, these sets are too conservative in general. Another problem is that it is not always clear how to compute or represent $\phi(C_n)$.

An alternative is the naive bootstrap. It estimates the distribution of $\sqrt{n}(\phi(\hat{t}_n) - \phi(t_n))$ by

$$\hat{M}_n := \mathscr{L}_{X \sim \hat{L}_n}(\sqrt{n}(\phi(\hat{t}_n + \sqrt{n^{-1}}\, X) - \phi(\hat{t}_n)))$$

($\hat{t}_n$, $\hat{L}_n$ regarded as fixed). Under some regularity conditions this method works well, if ϕ is compactly differentiable in the sense of Reeds (1976); see Gill (1989). However the functions ϕ we are interested in are not necessarily differentiable, although they have other regularity properties. One consequence is that the naive bootstrap fails in general.

There is a possible modification of the bootstrap, called the 'rescaled bootstrap' hereafter. In an i.i.d. setting it is equivalent to the sample size modification described by Bretagnolle (1983); see also Beran and Srivastava (1987) and Eaton and Tyler (1991). The naive bootstrap estimator $\hat{M}_n$ is replaced by

$$\tilde{M}_n := \mathscr{L}_{X \sim \hat{L}_n}(\sqrt{m}(\phi(\hat{t}_n + \sqrt{m^{-1}}\, X) - \phi(\hat{t}_n))),$$

where $m = m(n)$ are numbers such that $m \to \infty$ and $m/n \to 0$. In the present framework it typically works in case of constant sequences of parameters, $t_n = t$ for all n. But if this condition is violated, the rescaled bootstrap might fail either. In that sense it is very nonrobust and shouldn't be used. Both the naive and the rescaled bootstrap are treated in Sect. 2.

In Sect. 3 we discuss an ad hoc method, which is more reliable but computer-intensive. The basic idea is to use the classical recipe of inverting tests: Each hypothetical parameter in $\mathbf{X}$ is tested for being a plausible candidate for t_n, and then the set of plausible parameters is used to define a confidence set for $\phi(t_n)$. In many classical problems this approach reduces to looking only at one or a few possible parameters, but in the present situation no such reduction seems possible. Therefore it is proposed to replace the whole parameter space $\mathbf{X}$ by a random finite subset, which leads to a stochastic procedure as in Beran and Millar (1987).

Some technical details and most proofs are deferred to the appendix.

Example 1.1: sample covariance matrices. Let $\mathbf{X}$ be the space of real, symmetric $d \times d$ matrices $x = (x_{i,j} \colon 1 \leq i, j \leq d)$, and for a probability measure Q on $\mathbf{R}^d$ with finite second moments let $\Sigma(Q) \in \mathbf{X}$ be its covariance matrix. Given a sequence $(Q_n)_n$ of such distributions, let $\hat{Q}_n$ be the empirical distribution of n independent random variables with distribution Q_n. Under mild regularity conditions the estimators $\hat{t}_n := \Sigma(\hat{Q}_n)$ for $t_n := \Sigma(Q_n)$ meet A1, see Beran and Srivastava (1985): Suppose that

$$(1) \qquad Q_n \xrightarrow[\text{weakly}]{} Q \quad \text{and} \quad E_{Y \sim Q_n}(\|Y\|^4) \to E_{Y \sim Q}(\|Y\|^4) < \infty$$

for a fixed Q. Then $t_n \to t$, and $\mathscr{L}(B_n)$ converges weakly to a centered Gaussian distribution $L = L(Q)$ on $\mathbf{X}$. Furthermore (1) holds in probability, if Q_n is replaced with $\hat{Q}_n$. Thus by resampling from $\hat{Q}_n$ one gets a consistent estimator $\hat{L}_n$ for the distribution of B_n. The distribution L is typically nonsingular; for instance, if Q is absolutely continuous with respect to Lebesgue measure. This fact is needed in Sect. 3.

Example 1.2: empirical discrete measures. Now let Q_n, $\hat{Q}_n$ be probability measures on the set $\{1, 2, \ldots\}$ of positive integers. Any such distribution Q is identified with $(Q\{1\}, Q\{2\}, \ldots) \in \ell_1$ or $(\sqrt{Q\{1\}}, \sqrt{Q\{2\}}, \ldots) \in \ell_2$, where ℓ_p is the Banach space of real sequences $r = (r_1, r_2 \ldots)$ with finite norm $\|r\|_p := (\sum_i |r_i|^p)^{1/p}$. In both cases, if

$$(2) \qquad Q_n\{i\} \to Q\{i\}\ \forall i \quad \text{and} \quad \sum_i \sqrt{Q_n\{i\}} \to \sum_i \sqrt{Q\{i\}} < \infty$$

for a fixed Q, then the $B_n := \sqrt{n}(\hat{Q}_n - Q_n)$ satisfy A1 with a centered Gaussian distribution $L = L(Q)$. Again one can bootstrap $\mathscr{L}(B_n)$ in the usual way, because (2) still holds in probability, if Q_n is replaced with $\hat{Q}_n$.

2 Naive and rescaled bootstrap

In this section we restrict our attention to the following asymptotic framework:

B1 $\qquad \sqrt{n}(t_n - t) \to \Delta \qquad$ for fixed points $t, \Delta \in \mathbf{X}$.

The case $\Delta = 0$ corresponds to the standard asymptotics '$Q_n = Q$ for all n' in examples 1.1 and 1.2. Further suppose that the function $\phi: \mathbf{X} \to \mathbf{Y}$ is compactly differentiable at t in a weak sense:

B2 There are a linear space $\mathbf{X}_o \subset \mathbf{X}$ containing $\operatorname{supp}(L)$ and Δ and a function $\Phi: \mathbf{X}_o \to \mathbf{Y}$ such that

$$\lim_{r \downarrow 0,\, x' \to x} r^{-1}(\phi(t + r x') - \phi(t)) = \Phi(x) \qquad \forall x \in \mathbf{X}_o$$

($\operatorname{supp}(L)$ is the smallest closed set $A \subset \mathbf{X}$ with $LA = 1$). Note that the function Φ in B2 is automatically continuous and positive homogeneous on $\mathbf{X}_o$, i.e. $\Phi(rx) = r\Phi(x)$ for all $x \in \mathbf{X}_o$ and $r \geqq 0$. If Φ is linear, then φ is compactly differentiable tangentially at $\mathbf{X}_o$ in the sense of Gill (1989). At the end of this section we give examples for ϕ with a detailed discussion of its analytic properties. The important thing is that B2 is satisfied, but Φ is nonlinear in general.

It is notationally convenient to reformulate the problem in terms of

$$\begin{aligned}
\Phi_n(x) &:= \sqrt{n}(\phi(t + \sqrt{n^{-1}}\, x) - \phi(t)), \\
\Delta_n &:= \sqrt{n}(t_n - t), \\
\hat{\Delta}_n &:= \sqrt{n}(\hat{t}_n - t) = B_n + \Delta_n.
\end{aligned}$$

We also write

$$f(x, z) := f(z+x) - f(z) \qquad \text{for functions } f\colon \mathbf{X} \to \mathbf{Y}.$$

Then $\Phi_n(\hat{\Delta}_n) - \Phi_n(\Delta_n) = \Phi_n(B_n, \Delta_n)$, and

$$\hat{M}_n = \mathscr{L}_{X \sim \hat{L}_n}(\Phi_n(X, \hat{\Delta}_n)),$$
$$\tilde{M}_n = \mathscr{L}_{X \sim \hat{L}_n}(\Phi_m(X, \sqrt{m/n}\hat{\Delta}_n)).$$

The next result gives the asymptotic joint distribution of $\Phi_n(\hat{\Delta}_n) - \Phi_n(\Delta_n)$ and $\hat{M}_n$. While $\Phi_n(\hat{\Delta}_n) - \Phi_n(\Delta_n)$ has a definite limiting distribution, the bootstrap estimator $\hat{M}_n$ converges in distribution to a random measure:

Proposition 1 *Suppose that* A1–2 *and* B1–2 *hold. Then the random elements* $\Phi_n(\hat{\Delta}_n) - \Phi_n(\Delta_n)$ *converge in distribution to* $\Phi(B, \Delta)$. *Further, for any fixed* $f \in \mathscr{C}(\mathbf{Y}, [0, 1])$, *the random elements* $\hat{M}_n f$ *converge in distribution to the random variable* $M_{(\Delta+B)} f$, *where* $M_{(z)} := \mathscr{L}(\Phi(B, z))$.

Generally, for any $f \in \mathscr{C}(\mathbf{Y} \times \mathbf{Y}, [0, 1])$ *the random elements* $\hat{M}_n f(\Phi_n(B_n, \Delta_n), \cdot)$ *converge in distribution to the random variable* $M_{(\Delta+B)} f(\Phi(B, \Delta), \cdot)$.

An exceptional case is Φ being linear. For then $\Phi(\cdot, z) = \Phi$ for all $z \in \mathbf{X}_o$, and $\hat{M}_n$ converges weakly to the correct, nonrandom limit $M_{(0)}$ in probability.

The rescaled bootstrap $\tilde{M}_n$ has a different behavior:

Proposition 2 *Under* A1–2 *and* B1–2, *the rescaled bootstrap estimator* $\tilde{M}_n$ *converges weakly to* $M_{(0)}$ *in probability.*

For $\Delta = 0$ this is the correct limit, but otherwise the distributions $M_{(0)}$ and $M_{(\Delta)}$ can be different, unless Φ is linear.

Both Propositions 1 and 2 can be derived from an extended Continuous Mapping Theorem given by van der Vaart and Wellner (1989); see the appendix.

Example 2.1: eigenvalues of a symmetric matrix. Let $\mathbf{X}$ be as in example 1.1, and for any real, symmetric $\tilde{d} \times \tilde{d}$ matrix $\tilde{x}$ let $\phi(\tilde{x})$ be the vector of its $\tilde{d}$ eigenvalues in decreasing order. This function ϕ is positive homogeneous and Lipschitz continuous. In order to formulate other regularity properties we have to introduce some notation: The set of all orthogonal $d \times d$ matrices is denoted by $\mathbf{T}$. For a (possibly void) subset $E = \{e_1, \ldots, e_{k-1}\}$ of $\{1, \ldots, d-1\}$ with $0 =: e_0 < e_1 < e_2 < \ldots < e_k := d$ define

$$\Phi(x \mid E) := (\phi(x_{i,j} : e_{s-1} < i, j \leqq e_s) : 1 \leqq s \leqq k).$$

Here is a basic result, which is given in Eaton and Tyler (1991):

Lemma 1 *Let* $(\xi_n)_n$ *be a sequence in* $\mathbf{X}$. *Further let* $\tau_n, \tau \in \mathbf{T}$ *and* $E = \{e_1, \ldots, e_{k-1}\} \subset \{1, \ldots, d-1\}$ *such that*

$$\begin{aligned}
&\tau_n \to \tau, \\
&\tau_n^* \xi_n \tau_n = \operatorname{diag}(\phi(\xi_n)) && \textit{for all } n, \\
&\phi_i(\xi_n) - \phi_{e(s)}(\xi_n) = 0 && \textit{for } e_{s-1} < i \leqq e_s \textit{ and } 1 \leqq s \leqq k, \\
&\phi_{e(s)}(\xi_n) - \phi_{e(s)+1}(\xi_n) \to \infty && \textit{for } 1 \leqq s < k.
\end{aligned}$$

Then

$$\phi(x, \xi_n) - \Phi(\tau^* x \tau \mid E) \to 0$$

uniformly in x on bounded subsets of $\mathbf{X}$ *($[\cdot]^*$ denotes transposition).*

A particular consequence of Lemma 1 is condition B2: For any fixed $t \in \mathbf{X}$ let $E(t)$ be the set of all indices $i \in \{1, \ldots, d-1\}$ such that $\phi_i(t) > \phi_{i+1}(t)$, and let $\mathbf{T}(t)$ be the set of all $\tau \in \mathbf{T}$ such that $\tau^* t \tau = \operatorname{diag} \phi(t)$). Then

$$(3) \qquad \lim_{r \downarrow 0, x' \to x} r^{-1}(\phi(t + r x') - \phi(t)) = \Phi(\tau^* x \tau \mid E(t)) \qquad \text{for } x \in \mathbf{X} \text{ and } \tau \in \mathbf{T}(t).$$

This follows by letting $\xi_n = r_n^{-1} t$, where $r_n \downarrow 0$. Formula (3) shows that ϕ is differentiable at t if and only if all eigenvalues of t are different.

Example 2.2: minimum distance functionals. Let $(q_\theta : \theta \in \Theta)$ be a parametric family in $\mathbf{X}$, and define

$$\phi(x) := \inf_{\theta \in \Theta} \|q_\theta - x\|.$$

This functional ϕ is obviously Lipschitz continuous with constant 1 for any parametric family. Minimum distance tests based on functionals of this type are discussed by Pollard (1980). It would be interesting to have upper confidence bounds for the distance $\phi(t_n)$ rather than just a test, whether the parametric family is correct or not. Suppose that Θ is an open subset of $\mathbf{R}^p$ such that

$$\inf_{\tilde\theta \in \Theta, |\tilde\theta - \theta| \geqq \varepsilon} \|q_{\tilde\theta} - q_\theta\| > 0 \qquad \forall \varepsilon > 0 \quad \forall \theta \in \Theta,$$

and $\Theta \ni \theta \mapsto q_\theta$ is continuously differentiable with nonsingular derivatives $D_\theta : \mathbf{R}^p \to \mathbf{X}$. Then one can prove

Lemma 2 *Let $(s_n)_n$ be a sequence in $\mathbf{X}$ such that $\phi(s_n) = O(\sqrt{n^{-1}})$ and $s_n \to q_\theta$ for some $\theta \in \Theta$. Let $(\theta_n)_n$ be any sequence in Θ such that $\|q_{\theta_n} - s_n\| \leqq \phi(s_n) + n^{-1}$, and define $\pi_n := \sqrt{n}(s_n - q_{\theta_n})$. Then*

$$\sqrt{n}\,\phi(s_n + \sqrt{n^{-1}}\,x) - \min_{h \in \mathbf{R}^p} \|D_\theta h - \pi_n - x\| \to 0$$

uniformly in x on bounded subsets of $\mathbf{X}$.

In particular, Lemma 2 implies that B2 is satisfied for $t = q_\theta$, $\theta \in \Theta$, with limit

$$\Phi(x) := \min_{h \in \mathbf{R}^p} \|D_\theta h - x\|.$$

Moreover, there is an expansion for the corresponding minimum distance approximations: For $r > 0$ let $\psi_r : \mathbf{X} \to \Theta$ be any function such that $\|Q_{\psi_r(x)} - x\| \leqq \phi(x) + r^2$ for all $x \in \mathbf{X}$. Then

$$\lim_{r \downarrow 0, x' \to x} r^{-1} \psi_r(q_\theta + r x') = \arg\min_{h \in \mathbf{R}^p} \|D_\theta h - x\|$$

whenever this argmin is uniquely defined. Even if uniqueness is guaranteed for sufficiently many $x \in \mathbf{X}$, the argmin is nonlinear in general.

3 Robust confidence sets

Now the points $\Delta_n = \sqrt{n}(t_n - t)$ are no longer assumed to converge to a fixed point in $\mathbf{X}$. The only assumption (if any) is that

B3 $$t_n \to t.$$

Moreover, $\Phi_n: \mathbf{X} \to \mathbf{Y}$ could be any function; only in the examples we refer to $\Phi_n(x) = \sqrt{n}(\phi(t + \sqrt{n^{-1}}x) - \phi(t))$.

Generally a confidence set for $\Phi_n(\Delta_n)$ is a random subset C_n of $\mathbf{Y}$ defined on (Ω_n, P_n). For a fixed level $\alpha \in (0, 1/2)$ the confidence sets C_n are said to be asymptotically valid, if

$$\limsup_{n \to \infty} P_n\{\Phi_n(\Delta_n) \notin C_n\} \leqq \alpha.$$

In this paper we restrict our attention to confidence sets, which are determined by one shape function $S: \mathbf{Y} \to \mathbf{R}$ and one confidence bound $R_n: \Omega_n \to \mathbf{R}$ as follows:

$$C_n = \{y \in \mathbf{Y}: S(\Phi(\hat{\Delta}_n) - y) \leqq R_n\}.$$

For instance, if $S(y) = \|y\|$, then C_n is the closed ball with center $\Phi_n(\hat{\Delta}_n)$ and radius R_n. If $\mathbf{Y} = \mathbf{R}$ and $S(y) = -y$, then C_n is the halfline $(-\infty, \Phi_n(\hat{\Delta}_n) + R_n]$. Generally it is assumed that S is Lipschitz continuous.

The naive bootstrap leads to the sets $C_n^{(b)} := \{S(\Phi_n(\hat{\Delta}_n) - \cdot) \leqq \hat{q}_n(\hat{\Delta}_n)\}$, where

$$\hat{q}_n(\xi) := \sup\{r \in \mathbf{R}: \Pr_{X \sim \hat{L}_n}\{S \circ \Phi_n(X, \xi) \geqq r\} \geqq \alpha\}.$$

Alternatively, we test the hypothesis "$\Delta_n = \xi$" for all points ξ in a random finite subset $\hat{\Xi}_n$ of $\mathbf{X}$. The tests are based on the teststatistic

$$T_n(x, \xi) := S(\Phi_n(x) - \Phi_n(\xi)).$$

Thus $I\{T_n(\hat{\Delta}_n, \xi) > \hat{q}_n(\xi)\}$ is a test of "$\Delta_n = \xi$" with nominal level α. Now the confidence bound $\hat{R}_n$ is defined as

$$\hat{R}_n := \max\{T_n(\hat{\Delta}_n, \xi): \xi \in \hat{\Xi}_n, T_n(\hat{\Delta}_n, \xi) \leqq \hat{q}_n(\xi)\},$$

and the corresponding confidence set is denoted by $\hat{C}_n$.

This description is convenient for theoretical considerations. But one can avoid the computation of all quantiles $\hat{q}_n(\xi)$, $\xi \in \hat{\Xi}_n$, by computing $\hat{R}_n$ inductively after arranging the values $T_n(\hat{\Delta}_n, \xi)$ in decreasing order.

The particular choice of T_n seems intuitive and often leads to reasonable confidence sets. However, one can also find examples, where this particular teststatistic does not work well. One possible improvement is to replace $T_n(x, \xi)$ with $w_n(x) T_n(x, \xi)$, where $w_n > 0$ is a suitable weight function (studentization). The choice of an appropriate teststatistic as well as more flexible shapes for $\hat{C}_n$ will be the subject of future research.

In what follows the asymptotic behavior of $\hat{C}_n$ is studied under conditions A1–2, where we also assume that

A3 $$\operatorname{supp}(L) \text{ is a linear subspace of } \mathbf{X}.$$

It is well-known that this requirement is always satisfied for a centered Gaussian distribution L. In addition one needs conditions on the random sets $\hat{\Xi}_n$ and the functions Φ_n.

On the one hand it is assumed that $\hat{\Xi}_n$ is asymptotically dense in $\Delta_n + \operatorname{supp}(L)$ in a weak sense:

C1 $$\min_{\xi \in \hat{\Xi}_n} \|\xi - \Delta_n - z\| \underset{p}{\to} 0 \qquad \forall z \in \operatorname{supp}(L).$$

On the other hand the sets $\hat{\Xi}_n$ must not be too large:

C2 $$\max_{\xi \in \hat{\Xi}_n} \|\xi - \hat{\Delta}_n\| = o_p(\sqrt{n}).$$

An explicit construction of $\hat{\Xi}_n$ might be as follows: Let $\hat{L}_n$ be a Markov kernel in the classical sense, and let $B_{1,n}, B_{2,n}, \ldots$ be random variables, which are conditionally independent with distribution $\hat{L}_n$ given $\hat{\Delta}_n$, $\hat{L}_n$. Then the random sets

$$\hat{\Xi}_n := \{\hat{\Delta}_n - B_{i,n} : 1 \leqq i \leqq k_n\}$$

meet assumption C1, provided that k_n tends to infinity. This follows from the fact that the joint distribution of $B_n, B_{1,n}, \ldots, B_{k,n}$ converges weakly to L^{k+1} for any fixed integer $k > 0$. For

$$\begin{aligned} \limsup_{n\to\infty} P_n\{\min_{\xi \in \hat{\Xi}_n} \|\xi - \Delta_n - z\| \geqq \varepsilon\} &\leqq \inf_k \limsup_{n\to\infty} P_n\{\min_{i \leqq k} \|B_n - z - B_{i,n}\| \geqq \varepsilon\} \\ &\leqq \inf_k \mathrm{E}(L\{\|B - z - \cdot\| \geqq \varepsilon\}^k) \\ &= \Pr\{L\{\|B - z - \cdot\| < \varepsilon\} = 0\} \\ &= 0 \qquad \forall z \in \operatorname{supp}(L). \end{aligned}$$

The second requirement C2 can always be achieved by a proper truncation. Note also that $\max_{i \leqq k_n} \|B_{i,n}\| = o_p(\sqrt{k_n})$, provided that $\mathrm{E}(\|B_{1,n}\|^2 \mid \hat{\Delta}_n, \hat{L}_n) \underset{p}{\to} \mathrm{E}(\|B\|^2)$.

The following assumption for Φ_n is used in order to establish a lower bound for $\hat{R}_n$ and the asymptotic validity of $\hat{C}_n$:

D1 There is a function Φ: $\operatorname{supp}(L) \to \mathbf{Y}$ such that

$$\Phi_n(x_n, \Delta_n) \to \Phi(x) \qquad \text{whenever } x_n \to x \in \operatorname{supp}(L).$$

Typically one can verify this condition only along certain subsequences $(N_k)_k$ of (n). Then the results below have to be modified in an obvious way. On the other hand one can often prove the following stronger compactness condition, which is used in order to obtain an upper bound for $\hat{R}_n$:

D2 For arbitrary points $\xi_n \in \mathbf{X}$ such that $\|\xi_n - \Delta_n\| = o(\sqrt{n})$ and for any subsequence of (n) there are a subsubsequence $(n_k)_k$ and a function $\tilde{\Phi}$: $\operatorname{supp}(L) \to \mathbf{Y}$ such that

$$\Phi_{n_k}(x_k, \xi_{n_k}) \to \tilde{\Phi}(x) \qquad \text{whenever } x_k \to x \in \operatorname{supp}(L).$$

For instance, if $\mathbf{Y}$ has finite dimension and the functions Φ_n are uniformly Lipschitz continuous, then D2 is satisfied. A special case of D1–2 is:

D3 There is a continuous, linear function Φ_o: $\operatorname{supp}(L) \to \mathbf{Y}$ such that

$$\Phi_n(x_n, \xi_n) \to \Phi_o(x) \quad \text{whenever } x_n \to x \in \operatorname{supp}(L) \quad \text{and} \quad \|\xi_n - \Delta_n\| = o(\sqrt{n}).$$

For $\Phi_n = \sqrt{n}(\phi(t + \sqrt{n^{-1}}\cdot) - \phi(t))$, the latter assumption is satisfied, if B3 holds and ϕ is continuously Frechet differentiable in a neighborhood of t.

Before stating the results let us introduce some further notation: In case of D1 the function $S(\Phi(x) - \Phi(z))$ is denoted by $T(x, z)$, and $q(z)$ stands for the $(1-\alpha)$-quantile of $\mathscr{L}(S \circ \Phi(B, z))$. If D1–2 hold, let $\mathscr{Z}$ be the family of all sequences $Z = (n_k, \xi_k)_k$, where $(n_k)_k \subset (n)$, $\xi_k \in \mathbf{X}$, $\|\xi_k - \Delta_{n_k}\| = o(\sqrt{n_k})$, and $\Phi_{n_k}(\cdot, \xi_k)$ converges to some function $\Phi(\cdot, Z)$ as in D2. Then let $q(Z)$ be the $(1-\alpha)$-quantile of $\mathscr{L}(S \circ \Phi(B, Z))$, and define

$$T(x, Z) := \limsup_{k \to \infty} T_{n_k}(\Delta_{n_k} + x, \xi_k).$$

This slight abuse of notation can be justified, because $\Phi(x, z) = \Phi(x, Z)$, $q(z) = q(Z)$, and $T(x, z) = T(x, Z)$ for $x, z \in \operatorname{supp}(L)$ and $Z := (n, \Delta_n + z)_n$.

Theorem 1 *Suppose that* A1–3, C1 *and* D1 *are satified, and define*

$$f^*(x) := \sup\{T(x, z): z \in \operatorname{supp}(L),\ T(x, z) < q(z)\} \qquad (x \in \operatorname{supp}(L)).$$

If f: $\mathbf{X} \to [-\infty, \infty)$ is upper semicontinuous such that $f \leqq f^*$ on supp (L), *then*

$$\exp(\hat{R}_n) \geqq \exp(f(B_n)) + o_p(1).$$

Moreover, let Φ satisfy the following assumption:

(4) supp $(L) \ni z \mapsto T(B, z)$ *has no local maximum at* 0 *almost surely.*

Then

$$\limsup_{n \to \infty} P_n\{\Phi_n(\Delta_n) \notin \hat{C}_n\} \leqq \Pr\{S \circ \Phi(B) \geqq q(0)\}.$$

In particular, the confidence sets $\hat{C}_n$ are asymptotically valid, if the distribution of $S \circ \Phi(B)$ is continuous.

Theorem 2 *Suppose that* A1–3, C2, *and* D1–2 *are satisfied. Let*

$$F^*(x) := \sup\{T(x, Z): Z \in \mathscr{Z},\ T(x, Z) \leqq q(Z)\} \qquad (x \in \operatorname{supp}(L)).$$

Then F^* is bounded from above, and

$$\hat{R}_n \leqq F(B_n) + o_p(1)$$

for any lower semicontinuous F: $\mathbf{X} \to \mathbf{R}$ such that $F \geqq F^*$ on supp (L).

These lower and upper bounds for $\hat{R}_n$ are not very handy, but in some special cases they lead to good and simpler bounds. Especially in case of D3 they imply that the confidence sets $\hat{C}_n$ and $C_n^{(b)}$ are asymptotically equivalent:

Proposition 3 *Suppose that* A1–3, C1–2, *and* D3 *are satisfied, where* $\mathscr{L}(S \circ \Phi_o(B))$ *is continuous. Then the confidence bounds* $\hat{q}_n(\hat{\Delta}_n)$ *and* $\hat{R}_n$ *converge to* $q(0)$ *in probability. In particular,*

$$\lim_{n \to \infty} P_n\{\Phi(\Delta_n) \notin \hat{C}_n\} = 1 - \alpha.$$

For under the conditions of Proposition 3,

$$f^* = \sup\{S \circ \Phi_o(z) \colon z \in \operatorname{supp}(L),\ S \circ \Phi_o(z) < q(0)\} = q(0)$$

on supp (L). Further, $q(Z) = q(0)$ for all $Z \in \mathscr{Z}$, so that $F^* \leq q(0)$.

In the remainder of this section we apply Theorems 1 and 2 to Examples 2.1 and 2.2.

Example 2.1 (continued). Suppose that B3 holds, where $\phi_d(t) > 0$ and $t = \operatorname{diag}(\phi(t))$. Further suppose that L is absolutely continuous with respect to Lebesgue measure on $\mathbf{X}$, which is denoted by Leb. Now consider the functions

$$\psi(x) := (\log \phi_1(x), \ldots, \log \phi_d(x)), \qquad S(y) := \max_{1 \leq i \leq d} |y_i|,$$

where $\log r$ may be defined arbitrarily for $r \leq 0$ (the use of ψ instead of ϕ is motivated by the asymptotic theory for normal distributions Q in Example 1.1). Under these assumptions the confidence sets $\hat{C}_n$ are asymptotically valid, and the confidence bounds $\hat{R}_n$ are stochastically bounded.

Limit functions corresponding to $\Psi_n := \sqrt{n}(\psi(t + \sqrt{n^{-1}}\cdot) - \psi(t))$ are denoted with the letter Ψ in place of Φ. Condition D2 is satisfied, and all possible limits $\Psi(\cdot, Z)$ have the form

$$(5) \qquad t^{-1}\Phi(\tau^* x \tau, \operatorname{diag}(y) \mid E) = t^{-1}(\Phi(\operatorname{diag}(y) + \tau^* x \tau \mid E) - \Phi(\operatorname{diag}(y) \mid E))$$

for some $E = E(Z) = \{e_1, \ldots, e_{k-1}\} \subset \{1, \ldots, d-1\}$ with $E \supset E(t)$, $\tau = \tau(Z) \in \mathbf{T}(t)$ and $y = y(Z) \in \mathbf{R}^d$. This follows from Lemma 1 together with a simple compactness argument. In particular, if t has d different eigenvalues, then D3 holds with limit

$$\Psi_o(x) := t^{-1}(x_{1,1}, \ldots, x_{d,d}).$$

For proving the asymptotic validity of the confidence sets $\hat{C}_n$, suppose (without loss of generality) that D1 holds with limit $\Psi(x) = t^{-1}\Phi(\tau_o^* x \tau_o, \operatorname{diag}(y_o) \mid E_o)$. Since $\Psi(z + r\,\mathrm{Id})$ equals $\Psi(z) + t^{-1}(r, \ldots, r)$ for any $r \in \mathbf{R}$ and $z \in \mathbf{X}$, condition (4) in Theorem 1 is satisfied. Moreover,

$$(6) \qquad \operatorname{Leb}\{x \in \mathbf{X} \colon \text{some component of } \Psi(x) \text{ equals } \lambda\} = 0 \qquad \forall \lambda \in \mathbf{R}.$$

Hence $\mathscr{L}(S \circ \Psi(B))$ is continuous by our assumption on L.

The function F^* is given by

$$(7) \qquad F^*(x) = \sup\{T(x, z) \colon z \in \mathbf{X},\ T(x, z) \leq \tilde{q}(z)\},$$

where $\tilde{q}(z)$ is the maximum of all $(1-\alpha)$-quantiles of $\mathscr{L}(S \circ \Psi(\tau^* B \tau, z))$, $\tau \in E(t)$. This is very similar to $f^*(x)$ except that '$<$' and $q(\cdot)$ are replaced with '$\leq$' and $\tilde{q}(\cdot)$ respectively. In fact, if $E_o = E(t)$, then $q(\cdot) = \tilde{q}(\cdot)$, because $\Phi(\cdot \mid E(t))$

$=\Phi(\tau^*(\cdot)\tau \mid E(t))$ for all $\tau \in \mathbf{T}(t)$. If $E_o \neq E(t)$, this does not hold in general. However, if L is the distribution $L(Q)$ in Example 1.1, where Q is a nonsingular Gaussian measure on $\mathbf{R}^d$, then $\tau^* B\tau =_{\mathscr{L}} B$ for all $\tau \in \mathbf{T}(t)$ and again $q(\cdot)=\tilde{q}(\cdot)$.

Example 2.2 (continued). One can easily apply Theorems 1 and 2, if $\mathbf{X}$ is finite dimensional with $\operatorname{supp}(L)=\mathbf{X}$. Alternatively we consider a Hilbert space $\mathbf{X}$ with inner product $\langle \cdot, \cdot \rangle$. Let B3 be satisfied with $t=q_\theta$, $\theta \in \Theta$, let $\phi(t_n)=O(\sqrt{n^{-1}})$, and let L be centered Gaussian with $\operatorname{supp}(L) \not\subset D_\theta \mathbf{R}^p$. If one is interested in upper confidence bounds for $\phi(t_n)$, one has to consider $S(y)=-y$. Under the present assumptions the confidence sets $\hat{C}_n$ are asymptotically valid. Moreover, the corresponding confidence bounds $\phi(\hat{t}_n)+\sqrt{n^{-1}}\,\hat{R}_n$ can be bounded explicitly: There are a function $G^*\colon [0, \infty) \to [-\infty, \infty)$ and constants $c^*, d^* \in (0, \infty)$ to be defined later with the following properties:

$$\sqrt{n}\,\phi(\hat{t}_n)+\hat{R}_n \leqq G^*(\sqrt{n}\,\phi(\hat{t}_n))+o_p(1), \tag{8}$$

and

$$\begin{aligned} &G^* = -\infty \text{ on } [0, c^*),\\ &G^* \text{ is continuous on } [c^*, \infty) \text{ with } G^*(c^*)=0,\\ &G^*(r)-r \text{ is nondecreasing in } r \in [c^*, \infty) \text{ with limit } d^* \text{ as } r \to \infty. \end{aligned} \tag{9}$$

Note that $G^*(r)<r$ for some $r>0$. This indicates that the confidence sets $\hat{C}_n$ capture the positive bias of the estimator $\phi(\hat{t}_n)$.

Precisely, let Π be the orthogonal projection onto $(D_\theta \mathbf{R}^p)^\perp$, and let $\mathbf{N}$ be the set of all $v \in \Pi\mathbf{X}$ with $\|v\| \leqq 1$. For $v \in \mathbf{N}$ and $c \geqq 0$ let

$$\begin{aligned} \Psi(x \mid c, v) &:= (c^2+2c\langle v, x\rangle + \|\Pi x\|^2)^{1/2} = (c^2 - c^2\|v\|^2 + \|cv+\Pi x\|^2)^{1/2},\\ \tilde{q}(c, v) &:= \alpha\text{-quantile of } \mathscr{L}(\Psi(B \mid c, v)). \end{aligned}$$

Then

$$\begin{aligned} G^*(r) &:= \sup\{c \geqq 0\colon \tilde{q}(c, v) \leqq r \text{ for some } v \in \mathbf{N}\},\\ c^* &:= \tilde{q}(0, 0),\\ d^* &:= \sup\{(1-\alpha)\text{-quantile of } \mathscr{L}(\langle v, B\rangle)\colon v \in \mathbf{N}\}. \end{aligned}$$

For proving the preceding claims note first that condition D2 is satisfied by the Lipschitz continuity of ϕ. In particular, let $\Psi_n(x) := \sqrt{n}\,\phi(s_n+\sqrt{n^{-1}}\,x)$, where $s_n \to t$ and $\phi(s_n)=O(\sqrt{n^{-1}})$. It follows from Lemma 2 that there are $\pi_n \in \Pi\mathbf{X}$ such that $\Psi_n - \|\pi_n + \Pi(\cdot)\|$ converges to 0 uniformly on bounded sets. On the other hand, for each subsequence of (n) there exist a subsubsequence $(n_k)_k$, a number $c \geqq 0$ and a point $v \in \mathbf{N}$ such that $\sqrt{n_k}\,\phi(t_{n_k}) \to c$ and $\pi_{n_k} \xrightarrow[\text{weakly}]{} cv$. But then Ψ_{n_k} converges to $\Psi(\cdot \mid c, v)$ uniformly on compact sets. These functions have obviously no local maximum on $\operatorname{supp}(L)$, and

$$\mathscr{L}(\Psi(B \mid c, v)) \quad \text{is continuous.} \tag{10}$$

Hence Theorem 1 yields the asymptotic validity of $\hat{C}_n$.

As for the upper bound (8), one can argue similarly as in the proof of (7). Without loss of generality suppose that D1 holds with limit $\Phi = \Psi(\cdot \mid c_o, v_o) - c_o$ for some $v_o \in \mathbf{N}$ and $c_o = \lim_{n\to\infty} \sqrt{n}\,\phi(t_n)$. If $Z = (n_k, \sqrt{n_k}(s_k - t))_k$ is a point in $\mathscr{X}$ such that $T(x, Z) < \infty$ for some $x \in \mathbf{X}$, then $\phi(s_k) = O(\sqrt{n_k^{-1}})$. Hence $\Phi(\cdot, Z) = \Psi(\cdot \mid c, v) - c$ with $c = c(Z)$, $v = v(Z)$. In particular,

$$T(x, Z) = c - \Psi(x \mid c_o, v_o), \qquad q(Z) = c - \tilde{q}(c, v),$$

and a little algebra yields

$$F^*(x) \leqq G^*(\Psi(x \mid c_o, v_o)) - \Psi(x \mid c_o, v_o).$$

Now let $G(r) := G^*(r)$ for $r \geqq \tilde{q}(0, 0)$ and $G(r) := r - \tilde{q}(0, 0)$ for $0 \leqq r \leqq \tilde{q}(0, 0)$. Then G is continuous on $[0, \infty)$, according to (9), and Theorem 2 implies that $\sqrt{n}\,\phi(\hat{t}_n) + \hat{R}_n$ is not greater than $G(\sqrt{n}\,\phi(\hat{t}_n)) + o_p(1)$. But (10) implies that the asymptotic probability of $\sqrt{n}\,\phi(\hat{t}_n)$ being in $(\tilde{q}(0, 0) - \varepsilon, \tilde{q}(0, 0))$ is arbitrarily small for suitable $\varepsilon > 0$, and $\sqrt{n}\,\phi(\hat{t}_n) + \hat{R}_n < 0$ means that $\sqrt{n}\,\phi(\hat{t}_n) + \hat{R}_n = -\infty$. This yields (8).

Claims (9–10) are proved in the appendix.

Appendix

If P is an outer probability measure on a set Ω, we call (Ω, P) a probability space and P a distribution. A mapping X from Ω into some other set $\mathbf{X}$ is called a random element, and the distribution of X is defined to be the outer measure $\mathscr{L}(X) := P \circ X^{-1}$. If $(\Omega, \mathscr{A}, P)$ is a probability space in the classical sense, P is identified with the corresponding outer measure. In particular, if Ω is a topological space and $\mathscr{A}$ is the Borel σ-field, then P is called a Borel distribution. For a bounded, nonnegative function f on Ω the integral of f with respect to P is defined as

$$Pf := \int_0^\infty P\{f \geqq r\}\, \mathrm{d}r.$$

Here are some properties of this integral: For any bounded $f, g: \Omega \to [0, \infty)$ and $r \geqq 0$,

$$P(rf) = rPf, \qquad P(r+f) = r + Pf,$$
$$P(f \vee g) \leqq Pf + Pg, \qquad |Pf - Pg| \leqq \sup_{\omega \in \Omega} |(f-g)(\omega)|.$$

Following Hoffmann-Jørgensen (1984) (see also van der Vaart and Wellner, 1989), a sequence of distributions L_n on a metric space $(\mathbf{X}, d)$ converges weakly to a Borel distribution L on $\mathbf{X}$, if

$$L_n f \to Lf \qquad \forall f \in \mathscr{C}(\mathbf{X}, [0, 1]);$$

$\mathscr{C}(\mathbf{X},[0,1])$ denotes the space of continuous functions from $\mathbf{X}$ into $[0,1]$. A sequence of random elements $X_n:(\Omega_n, P_n)\to\mathbf{X}$ converges in distribution to an $\mathbf{X}$-random variable X, if $\mathscr{L}(X_n)$ converges weakly to $\mathscr{L}(X)$.

Let $\hat{L}_n$ be random distributions on $\mathbf{X}$ defined on (Ω_n, P_n); that means, $\hat{L}_n = \hat{L}_n(\cdot|\omega_n)$ is a distribution on $\mathbf{X}$ for all $\omega_n\in\Omega_n$. Then $\hat{L}_n$ converges weakly to L in probability, if

$$\hat{L}_n f \xrightarrow[p]{} Lf \quad \forall f\in\mathscr{C}(\mathbf{X},[0,1])$$

('$\xrightarrow[p]{}$' denotes convergence in outer probability).

A basic tool is the following result, see van der Vaart and Wellner (1989):

Theorem 3 *For $n=1,2,\ldots$ let L_n be a distribution on a metric space $\mathbf{X}$ and g_n a function from $\mathbf{X}$ into another metric space $\mathbf{Y}$. Suppose that L_n converges weakly to a tight Borel distribution L on $\mathbf{X}$ and*

$$g_n(x_n)\to g(x) \quad \textit{whenever} \quad x_n\to x\in\mathbf{X}_o\subset\mathbf{X},$$

where $\mathbf{X}_o$ is such that $L(\mathbf{X}\setminus\mathbf{X}_o)=0$ and g is a function from $\mathbf{X}_o$ into $\mathbf{Y}$. Then g is continuous, and $L_n\circ g_n^{-1}$ converges weakly to $L\circ g^{-1}$.

Proof of Propositions 1 and 2 For proving Proposition 1 it suffices to verify the general part, where f maps $\mathbf{Y}\times\mathbf{Y}$ into $[0,1]$. With

$$G_n(x):=\hat{L}_n f(\Phi_n(x,\Delta_n),\Phi_n(\cdot,\Delta_n+x)) \quad\text{and}\quad g(x):=Lf(\Phi(x,\Delta),\Phi(\cdot,\Delta+x))$$

one has to show that $G_n(B_n)$ converges in distribution to $g(B)$. Using some standard approximation arguments one can deduce that there are events $A_n\subset\Omega_n$ such that $P_n(\Omega_n\setminus A_n)\to 0$ and $\hat{L}_n\xrightarrow[\text{weakly}]{} L$ along $(A_n)_n$. Now one can utilize Theorem 3 iteratively as follows: By B1–2,

$$\Phi_n(x_n,\Delta_n+z_n)\to\Phi(x,\Delta+z) \quad\text{whenever}\quad x_n\to x\in\mathbf{X}_o \text{ and } z_n\to z\in\mathbf{X}_o.$$

Therefore,

$$G_n(x_n)\to g(x) \quad\text{whenever}\quad x_n\to x\in\mathbf{X}_o \text{ along } (A_n)_n.$$

With the nonrandom functions $g_n(x):=\inf_{\omega_n\in A_n} G_n(x|\omega_n)$ and $d_n(x):=\sup_{\omega_n\in A_n} G_n(x|\omega_n) - g_n(x)$ one can deduce from Theorem 3 that $g_n(B_n)$ converges in distribution to $g(B)$ and $d_n(B)\xrightarrow[p]{}0$. Thus the assertion follows from

$$|G_n(B_n)-g_n(B_n)|\leqq I(\Omega_n\setminus A_n)+d_n(B_n)=o_p(1).$$

As for Proposition 2, note that

$$\Phi_m(x_n,z_n)\to\Phi(x) \quad\text{whenever}\quad x_n\to x\in\mathbf{X}_o \text{ and } z_n\to 0.$$

Since $\sqrt{m/n}\,\hat{\Delta}_n\xrightarrow[p]{}0$, one may assume that $\sqrt{m/n}\,\hat{\Delta}_n\to 0$ along $(A_n)_n$. Then Theorem 3 shows that $\hat{L}_n f\circ\Phi_m(\cdot,\sqrt{m/n}\hat{\Delta}_n)\to Lf\circ\Phi$ along $(A_n)_n$ for any function $f\in\mathscr{C}(\mathbf{Y},[0,1])$ □

Proof of Theorem 1 By the continuity of S and Φ, the limit teststatistic T is continuous on $\operatorname{supp}(L) \times \operatorname{supp}(L)$. This implies that f^* is lower semicontinuous from $\operatorname{supp}(L)$ into $[-\infty, \infty]$. For that reason one needs an auxiliary function f rather than just an extension of f^*. Together with Theorem 3, the continuity of T also implies that the quantiles $q(z)$ are continuous in $z \in \operatorname{supp}(L)$; note that they are unique, because L has connected support.

Since $\operatorname{supp}(L)$ is separable, there are events $A_n \subset \Omega_n$ with $P_n(\Omega_n \setminus A_n) \to 0$ such that

$$\hat{L}_n \xrightarrow[\text{weakly}]{} L \quad \text{and} \quad \min_{\xi \in \hat{\Xi}_n} \|\xi - \Delta_n - z\| \to 0 \quad \forall z \in \operatorname{supp}(L) \quad \text{along}\,(A_n)_n.$$

In what follows we consider an arbitrary fixed number $\varepsilon > 0$ and fixed points $\omega_n \in A_n^*$, where $A_n^* \subset A_n$ such that

$$P_n(\Omega_n \setminus A_n^*) \leqq L(\mathbf{X} \setminus C) + o(1) \quad \text{and} \quad B_n \to C \ \text{along}\,(A_n^*)_n$$

for a compact set $C \subset \operatorname{supp}(L)$ to be specified later. That means, we treat all random elements as fixed points.

For proving the lower bound let C be any compact subset of $\operatorname{supp}(L)$ such that $L(\mathbf{X} \setminus C) \leqq \varepsilon$. It suffices to show that

$$\exp(\hat{R}_n) \geqq \exp(f(B_n)) + o(1).$$

For any subsequence of (n) there are a subsubsequence $(n_k)_k$ and a point $c \in C$ such that $B_{n_k} \to c$. Then $\limsup_{k \to \infty} f(B_{n_k}) \leqq f(c)$, and it suffices to consider the case $f^*(c) > -\infty$. We show that

$$\hat{R}_{n_k} \geqq f^*(c) + o(1).$$

For that purpose let $z = z(c)$ be an arbitrary point in $\operatorname{supp}(L)$ such that $T(c, z) < q(z)$. Then there are points $\xi_n = \xi_n(z) \in \hat{\Xi}_n$ such that $\|\xi_n - \Delta_n - z\| \to 0$. But then

$$T_{n_k}(\hat{\Delta}_{n_k}, \xi_{n_k}) \to T(c, z) \quad \text{and} \quad \hat{q}_{n_k}(\xi_{n_k}) \to q(z),$$

according to D1 and Theorem 3. Hence $T_{n_k}(\hat{\Delta}_{n_k}, \xi_{n_k}) \leqq \hat{q}_{n_k}(\xi_{n_k})$ for sufficiently large k. In particular,

$$\hat{R}_{n_k} \geqq T(c, z) + o(1).$$

The validity of $\hat{C}_n$ can be proved with similar arguments. Let $\mathbf{X}_o$ be the set of all $x \in \operatorname{supp}(L)$ such that $\operatorname{supp}(L) \ni z \mapsto T(x, z)$ has no local maximum at 0. By (4), $\mathbf{X}_o$ is a Borel set with $L\mathbf{X}_o = 1$. Now one can find a compact subset C of $\mathbf{X}_o \cap \{S \circ \Phi(\cdot) < q(0)\}$ such that $L(\mathbf{X} \setminus C) \leqq L\{S \circ \Phi(\cdot) \geqq q(0)\} + \varepsilon$. It suffices to show that

$$T(\hat{\Delta}_n, \Delta_n) \leqq \hat{R}_n$$

eventually as $n \to \infty$. We can restrict our attention to a subsequence $(n_k)_k$ of (n) such that B_{n_k} converges to some $c \in C$. By the continuity of T and q, there is a point $z = z(c) \in \operatorname{supp}(L)$ such that

$$T(c, 0) < T(c, z) < q(z).$$

Given points $\xi_n = \xi_n(z) \in \hat{\Xi}_n$ with $\|\xi_n - \Delta_n - z\| \to 0$,

$$\begin{aligned} T_{n_k}(\hat{\Delta}_{n_k}, \Delta_{n_k}) &\to T(c, 0), \\ T_{n_k}(\hat{\Delta}_{n_k}, \xi_{n_k}) &\to T(c, z), \\ \hat{q}_{n_k}(\xi_{n_k}) &\to q(z). \end{aligned}$$

Hence

$$T_{n_k}(\hat{\Delta}_{n_k}, \Delta_{n_k}) \leqq T_{n_k}(\hat{\Delta}_{n_k}, \xi_{n_k}) \leqq \hat{q}_{n_k}(\xi_{n_k})$$

for sufficiently large k □

Proof of Theorem 2 Condition D2 implies that the functions Φ_n are asymptotically equicontinuous in the following sense: For each $\eta > 0$ there are an $\varepsilon(\eta) > 0$ and an integer $N(\eta)$ such that $\|\Phi_n(\xi') - \Phi_n(\xi)\| \leqq \eta$ whenever $n \geqq N(\eta)$, $\|\xi' - \xi\| \leqq \varepsilon(\eta)$, $\|\xi - \Delta_n\| \leqq \varepsilon(\eta)\sqrt{n}$. In particular,

$$\|\Phi(x_1, Z) - \Phi(x_2, Z)\| \leqq \eta \quad \text{and} \quad T(x_1, Z) \leqq T(x_2, Z) + \ell\eta$$

for all $x_1, x_2 \in \operatorname{supp}(L)$ with $\|x_1 - x_2\| \leqq \varepsilon(\eta)$, and for any $Z \in \mathscr{Z}$, where ℓ denotes the Lipschitz constant of S. Hence $T(\cdot, Z)$ is either real-valued or constant (∞ or $-\infty$), and the family of all real-valued $T(\cdot, Z)$ is equicontinuous.

Another important consequence of D2 is a compactness property of the set $\mathscr{Z}$: For each sequence $(Z_l)_l$ in $\mathscr{Z}$ and for any $x \in \operatorname{supp}(L)$ there are a subsequence $(Z_{l_s})_s$ and a $Z_o \in \mathscr{Z}$ such that $q(Z_{l_s}) \to q(Z_o)$ and $T(x, Z_{l_s}) \to T(x, Z_o)$. The proof is omitted, because it is straightforward (though notationally somewhat tedious). In addition to D2 one has to utilize the fact that

$$q(Z) = \lim_{k \to \infty} q_{n_k}(\xi_k) \qquad \forall Z \in \mathscr{Z},$$

where $q_n(\xi)$ is defined as $\hat{q}_n(\xi)$ with L in place of $\hat{L}_n$. This property of $\mathscr{Z}$ certainly implies that $q(\cdot)$ is bounded on $\mathscr{Z}$. Together with the equicontinuity of the functions $T(\cdot, Z)$, one can also deduce that F^* is upper semicontinuous from $\operatorname{supp}(L)$ into $[-\infty, \max_{Z \in \mathscr{Z}} q(Z)]$.

Since $\|B_n\|$ is stochastically bounded, condition C2 is equivalent to $\max_{\xi \in \hat{\Xi}_n} \|\xi - \Delta_n\| = o_p(\sqrt{n})$. With (A_n) as in the proof of Theorem 1, one may also assume that $\max_{\xi \in \hat{\Xi}_n} \|\xi - \Delta_n\| = o(\sqrt{n})$ along $(A_n)_n$. For a fixed $\varepsilon > 0$ let A_n^*, C be as in the proof of Theorem 1, where $L(\mathbf{X} \setminus C) \leqq \varepsilon$. It suffices to show that

$$\hat{R}_n \leqq F(B_n) + o(1)$$

for any fixed sequence of points $\omega_n \in A_n$. Let $\xi_n \in \hat{\Xi}_n$ be such that

$$\hat{R}_n = T_n(\hat{\Delta}_n, \xi_n) \leqq \hat{q}_n(\xi_n),$$

provided that $\hat{R}_n > -\infty$. One only needs to consider a subsequence $(n_k)_k$ of (n) such that

$$\hat{R}_{n_k} > -\infty \ \forall k, \quad B_{n_k} \to c \in C, \quad Z := (n_k, \xi_{n_k})_k \in \mathscr{Z}.$$

But then,

$$\begin{aligned} T(c, Z) &= \limsup_{k \to \infty} T_{n_k}(\Delta_{n_k} + B_{n_k}, \xi_{n_k}) \\ &= \limsup_{k \to \infty} \hat{R}_{n_k} \\ &\leqq \lim_{k \to \infty} \hat{q}_{n_k}(\xi_{n_k}) = q(Z). \end{aligned}$$

Hence $\limsup_{k \to \infty} \hat{R}_{n_k} \leqq F^*(c)$, whereas $\liminf_{k \to \infty} F(B_{n_k}) \geqq F(c) \geqq F^*(c)$ □

Proof of (6) By the special form of Ψ it suffices to consider the function ϕ instead of Ψ. Note that $\mathbf{T}$ is a compact, continuously differentiable manifold with dimension $(d-1)d/2$. The set $\{x \in \mathbf{X} : \phi_i(x) = \lambda$ for some $i \in \{1, \ldots, d\}\}$ may be written as $\{\tau \operatorname{diag}(\lambda, y)\tau^* : \tau \in \mathbf{T},\ y \in \mathbf{R}^{d-1}\}$. Thus it is the image of $\mathbf{T} \times \mathbf{R}^{d-1}$ under a continuously differentiable mapping. Therefore it has Lebesgue measure 0, because $\mathbf{T} \times \mathbf{R}^{d-1}$ has dimension $(d-1)d/2 + (d-1) = \dim(\mathbf{X}) - 1$ □

Proof of (7) Without loss of generality let $\tau_o = \mathrm{Id}$. Fix any point $x \in \mathbf{X}$. For $Z \in \mathscr{Z}$ the limit $\Psi(\cdot, Z)$ has the form (5) with $E = E(Z)$, $\tau = \tau(Z)$, and $y = y(Z)$. In general, E and E_o are different. However, in order to compute $F^*(x)$ we only have to consider such points $Z \in \mathscr{Z}$, where $T(x, Z) < \infty$. This implies that $\sqrt{n_k}\, \|\phi(t_{n_k}) - \phi(t + \sqrt{n_k^{-1}}\, \xi_k)\|$ is bounded in k. Hence one can deduce from Lemma 1 that $E = E_o$, and one may even assume that

$$\sqrt{n_k}(\phi(t_{n_k}) - \phi(t + \sqrt{n_k^{-1}}\, \xi_k)) \to y_o - y$$

without changing $T(x, Z)$. Then $T(x, Z) = S(\Psi(x) - \Psi(z))$, where $z := \operatorname{diag}(y - y_o)$, and $q(Z)$ is the $(1-\alpha)$-quantile of $\mathscr{L}(S \circ \Psi(\tau^* B \tau, z))$. Hence $F^*(x)$ is not greater than the right hand side of (7).

On the other hand, for any $\tau \in \mathbf{T}(t)$ and for any $z \in \mathbf{X}$ let $s_n := \tau(\operatorname{diag}(\phi(t_n)) + \sqrt{n^{-1}}\, z)\tau^*$. Then one can deduce from Lemma 1 that $Z := (n, \sqrt{n}(s_n - t))_n$ belongs to Z with $\Psi(x, Z) = \Psi(\tau^* x \tau, z)$ and $T(x, Z) = T(x, z)$ □

Proof of (9–10) Standard theory for Gaussian measures on a Hilbert space shows that there is a complete orthonormal system $(x_i : i \in I)$ in $\operatorname{supp}(\mathscr{L}(\Pi B))$ such that

$$\Pi B =_{\mathscr{L}} \sum_{i \in I} Z_i \sigma_i x_i,$$

with independent standard normal random variables Z_i and constants $\sigma_i > 0$ such that $\sum_{i \in I} \sigma_i^2 < \infty$. But then

$$\Psi(B|c, v)^2 =_{\mathscr{L}} c^2(1 - \sum_{i \in I} \langle v, x_i \rangle^2) + \sum_{i \in I} (Z_i \sigma_i + c\langle v, x_i \rangle)^2.$$

One can deduce (10) from this representation by conditioning on all but one Z_i, $i \in I$.

Furthermore, $\tilde{q}(c, v)$ is strictly increasing in $c \in [0, \infty)$ for any fixed $v \in \mathbf{N}$. This implies that

$$\tilde{q}(c) := \min\{\tilde{q}(c, v) : v \in \mathbf{N}\}$$

is continuous and strictly increasing in $c \geqq 0$. Note that $\tilde{q}(\cdot)$ is welldefined and continuous. For $\mathbf{N}$ is sequentially compact with respect to the weak topology, and $\tilde{q}(c_n, v_n) \to \tilde{q}(c, v)$ whenever $c_n \to c$ and $v_n \xrightarrow[\text{weakly}]{} v$. On the other hand, for all $v \in \mathbf{N}$, $\delta > 0$ and $c \geqq 0$,

$$\{x \in \mathbf{X} : \Psi(x|c, v) \leqq \tilde{q}(c, v)\} \subset \{x \in \mathbf{X} : \Psi(x|c + \delta, v) \leqq \tilde{q}(c, v) + \delta\},$$

because $\Psi(x|c, v)$ is Lipschitz continuous in c with constant 1. Hence, $\Pr\{\Psi(B|c + \delta, v) \leqq \tilde{q}(c, v) + \delta\} \geqq \alpha$, which implies that $\tilde{q}(c) - c$ is nonincreasing in $c \geqq 0$. Finally, since

$$\Psi(x|c, v) - c = \langle v, x \rangle + O(c^{-1} \|x\|^2) \quad \text{as} \quad c \to \infty \qquad \text{uniformly in } v \in \mathbf{N},$$

one can deduce that $\tilde{q}(c) - c \to -d^*$ as $c \to \infty$. All these properties of $\tilde{q}(\cdot)$ imply (9) □

Acknowledgements. I am grateful to Rudy Beran, Warry Millar and Peter Bickel for their interest in this subject and some stimulating discussions.

References

Beran, R., Millar, P.W.: Stochastic estimation and testing. Ann. Stat. **15**, 1131–1154 (1987)

Beran, R., Srivastava, M.S.: Bootstrap tests and confidence regions for functions of a covariance matrix. Ann. Stat. **13**, 95–115 (1985) Correction: Ann. Stat. **15**, 470–471 (1987)

Bretagnolle, J.: Lois limites du bootstrap de certaines fonctionelles. Ann. Inst. Henri Poincaré **19**, 281–296 (1983)

Eaton, M.L., Tyler, D.E.: On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. Ann. Stat. **19**, 260–271 (1991)

Gill, R.D.: Non- and semi-parametric maximum likelihood estimators and the von Mises method. Scand. J. Stat. **16**, 97–128 (1989)

Hoffmann-Jørgensen, J.: Stochastic processes on Polish spaces. (unpublished, 1984)

Pollard, D.: The minimum distance method of testing. Metrika **27**, 43–70 (1980)

Reeds, J.A.: On the definition of von Mises functionals. Ph.D. dissertation, Harvard University, 1976

Vaart, A.W. van der, Wellner, J.A.: Prohorov and continuous mapping theorems in the Hoffmann-Jørgensen weak convergence theory, with applications to convolution and asymptotic minimax theorems. Preprint, 1989 Lecture Notes-Monograph Series, IMS, Hayward