A remark on consistent estimation

Document status and date:
Published: 01/01/1999

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Report 99-004
A Remark on Consistent Estimation
Erik W. van Zwet
Willem R. van Zwet

A REMARK ON CONSISTENT ESTIMATION

Erik W. van Zwet and Willem R. van Zwet

University of Utrecht, University of Leiden and EURANDOM at Eindhoven

Abstract

In this paper we re-examine two auxiliary results in Putter and van Zwet (1996). Viewed in a new light these results provide some insight in two related phenomena, to wit consistency of estimators and local asymptotic equivariance. Though technically quite different, our conclusions will be similar to those in Beran (1997) and LeCam and Yang (1990).

Keywords and phrases: Non-parametric models, consistency, local asymptotic equivariance, convolution theorem, sets of the first category

AMS subject classifications: 62G05, 62G20

1   Introduction

Let $(\mathcal{X}, \mathcal{A})$ be a measurable space and $\mathcal{P}$ a collection of probability measures on $(\mathcal{X}, \mathcal{A})$. Let Π be a topology on $\mathcal{P}$, so that $(\mathcal{P}, \Pi)$ is a topological space. Finally, let $X_1, X_2, \ldots$ denote a sequence of i.i.d. random variables with values in $\mathcal{X}$ and (unknown) common distribution $P \in \mathcal{P}$.

For $N = 1, 2, \ldots$, we consider a map $\tau_N : (\mathcal{P}, \Pi) \to (\mathcal{R}, \rho)$, where $(\mathcal{R}, \rho)$ is a metric space. Both spaces $(\mathcal{P}, \Pi)$ and $(\mathcal{R}, \rho)$ are equipped with the σ-algebra of Borel sets $\mathcal{B}(\mathcal{P}, \Pi)$ and $\mathcal{B}(\mathcal{R}, \rho)$, which are generated by the open sets in $(\mathcal{P}, \Pi)$ and $(\mathcal{R}, \rho)$ respectively. Probability distributions on these spaces are probability measures on the Borel sets, and are induced by measurable maps from $(\mathcal{X}^\infty, \mathcal{A}^\infty, P^\infty)$ to $(\mathcal{P}, \mathcal{B}(\mathcal{P}, \Pi))$ or $(\mathcal{R}, \mathcal{B}(\mathcal{R}, \rho))$. We assume throughout that each τ_N is measurable.

Having observed the i.i.d. sample $X_1, \ldots, X_N$ with common distribution $P \in \mathcal{P}$, our aim is to estimate the somewhat abstract $\mathcal{R}$-valued "parameter" $\tau_N(P)$. For a measurable map $t_N : \mathcal{X}^N \to \mathcal{R}$, let $T_N = t_N(X_1, \ldots, X_N)$ be an estimator of $\tau_N(P)$ based on $X_1, \ldots, X_N$. We shall say that T_N is a *consistent estimator* of $\tau_N(P)$ for $P \in \mathcal{P}$ if

$$(1.1) \qquad \rho(T_N, \tau_N(P)) \to^P 0, \qquad \text{for every } P \in \mathcal{P},$$

where $\to^P$ indicates convergence in probability under P as $N \to \infty$. The more formally inclined reader should view this expression as shorthand for the correct but laborious statement that the sequence $\{T_N\}_{N=1}^\infty$ is a consistent sequence of estimators of the sequence $\{\tau_N(P)\}_{N=1}^\infty$. If we wish to stress the role of the metric ρ in (1.1) we call T_N ρ-consistent.

In many applications the topology Π on $\mathcal{P}$ will be metrized by the Hellinger metric H so that $(\mathcal{P}, H)$ is a metric space. Recall that for $P, Q \in \mathcal{P}$ with densities f and g with respect to a common σ-finite measure μ on $(\mathcal{X}, \mathcal{A})$, the Hellinger distance H of P and Q is defined as

$$(1.2) \qquad H(P, Q) = \left\{ \int_{\mathcal{X}} (f^{1/2} - g^{1/2})^2 d\mu \right\}^{1/2}.$$

Many results in asymptotic statistics do not hold for all underlying distributions $P \in \mathcal{P}$, but only for $P \in \mathcal{P} \backslash D$ where the exceptional set D is in some sense small compared to $\mathcal{P}$. For a finite dimensional parametric family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^k$, we may identify $\mathcal{P}$ with Θ, and the exceptional subset of Θ will typically be small in the sense that it has Lebesgue measure zero. On the more general topological spaces of distributions $(\mathcal{P}, \Pi)$ that we consider in this paper, there is no obvious analogue for Lebesgue measure for which "small" sets can naturally be described as sets of measure zero. It is therefore hardly surprising that in the present set-up, the exceptional set D will be small in a topological sense: D will be a set of the first category in $(\mathcal{P}, \Pi)$. We recall that a *set of the first category* is a countable union of nowhere dense sets, and that a set is *nowhere dense* in $(\mathcal{P}, \Pi)$ if its closure does not contain an open set in $(\mathcal{P}, \Pi)$.

When thinking of exceptional sets $D \subset \mathcal{P}$ of the first category as "small", a word of caution may not be amiss. Even for parametric families $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^k$, such sets D may correspond to subsets of the parameter space of positive Lebesgue measure

2

and one has to impose regularity conditions to make sure that this phenomenon does not occur. By allowing exceptional sets of the first category, one is – in a sense – sweeping some technical difficulties under the rug in exchange for cleaner statements of results. Of course one also has the added generality of dealing with arbitrary $\mathcal{P}$ rather than parametric families, but again this comes at a price. With increasing complexity of the space $\mathcal{P}$, the concept of a set of the first category gradually looses its significance. For instance, as long as $(\mathcal{P}, \Pi)$ is topologically complete, we know that the complement of a set of the first category is everywhere dense, but in incomplete spaces a set of the first category may equal the entire space. This gradual loss of significance corresponds precisely to what one would expect: as the complexity of the model $\mathcal{P}$ increases, many results will hold in decreasing generality. All in all, we feel that exceptional sets of the first category may well be the proper concept in much of asymptotic statistical theory.

The framework introduced so far was used to discuss consistency of bootstrap estimators in Putter and van Zwet (1996). Two of the auxiliary results obtained in that paper a re-examined in the present note from a somewhat different point of view, to wit consistency of estimators in general and the related subject of local asymptotic equivariance. Though technically quite different, our conclusions will be similar in spirit to those in Beran (1997) and LeCam and Yang (1990).

2 Consistency

The parameter sequence $\{\tau_N(P)\}$ will be called *locally asymptotically constant* (LAC) at $P \in \mathcal{P}$ if for every $C > 0$,

$$(2.1) \qquad \lim_{N \to \infty} \sup_{\{Q \in \mathcal{P}: \ H(P,Q) \leq CN^{-1/2}\}} \rho(\tau_N(Q), \tau_N(P)) = 0.$$

If τ_N converges to a limit τ pointwise in $(\mathcal{R}, \rho)$, then the LAC property is obviously equivalent to the statement that the limit of $\tau_N(P_N)$ is the same for any sequence P_N contained in shrinking Hellinger balls of radii of order $N^{-1/2}$ and centered at P. The following proposition is simply a re-statement of lemma 5.3 in Putter and van Zwet (1996).

Proposition 2.1 *Let the topology Π on $\mathcal{P}$ be metrized by Hellinger metric H and suppose that*

(i) For every $N = 1, 2, \ldots$, the map $\tau_N : (\mathcal{P}, \Pi) \to (\mathcal{R}, \rho)$ is continuous;

(ii) There exists a ρ-consistent estimator $T_N = t_N(X_1, \ldots, X_N)$ of $\tau_N(P)$.

Then there exists a set D of the first category in $(\mathcal{P}, \Pi)$ such that $\{\tau_N(P)\}$ is LAC at every $P \in \mathcal{P} \backslash D$.

Let P^N and Q^N denote the joint distribution of $X_1, \ldots, X_N$ under the models P and Q respectively. We have

$$(2.2) \qquad 1 - \frac{1}{2} H^2(Q^N, P^N) = (1 - \frac{1}{2} H^2(Q, P))^N$$

and hence, if P and Q are at Hellinger distance of order $N^{-1/2}$, the N-dimensional distributions P^N and Q^N are at Hellinger distance of order 1. This is the case where the two models are contiguous and, for large N, the joint distributions of $X_1, \ldots, X_N$ are essentially different under P and Q. Yet, in order to be continuous and estimable, the parameter $\tau_N(P)$ has to be essentially the same under both models for large N and "most" P. Hence, if e.g. $\tau_N(P) = \theta(P^N)$ is a parameter of the distribution of P^N which varies non-trivially with its argument P^N, then one must expect that $\tau_N(P)$ will not be estimable.

A case of particular interest is as follows. Consider a sequence of random variables $Y_N = y_N(X_1, \ldots, X_N; P)$, where y_N is a measurable map from $\mathcal{X}^N \times \mathcal{P}$ to a *separable* metric space $(\mathcal{Y}, \eta)$. Let $\mathcal{R}$ be the space of all probability distributions on $(\mathcal{Y}, \eta)$ equipped with Prohorov's metric ρ which is defined for distributions $R_1, R_2 \in \mathcal{R}$ by

(2.3)
$$\rho(R_1, R_2) = \inf\{\varepsilon > 0 : R_1(A) \le R_2(A^\varepsilon) + \varepsilon, \text{for all } A \in \mathcal{B}(\mathcal{Y}, \eta)\},$$

where A^ε is an ε-neighborhood of A. Since $(\mathcal{Y}, \eta)$ is separable, ρ metrizes weak convergence of probability measures in $\mathcal{R}$ (c.f. Dudley (1989), sect. 11.3). Note that the separability of $(\mathcal{Y}, \eta)$ also implies that $(\mathcal{R}, \rho)$ is separable (Billingsley (1968), p. 239). The parameter to be estimated is

$$(2.4) \qquad\qquad \tau_N(P) = \mathcal{L}_P(Y_N),$$

the probability distribution of Y_N under P. Obviously τ_N maps $(\mathcal{P}, H)$ into $(\mathcal{R}, \rho)$. This is the estimation problem that the bootstrap is designed to solve.

We shall say that Y_N is *locally asymptotically distributionfree* (LAD) if its distribution τ_N satisfies (2.1) with ρ taken to be Prohorov metric. Obviously, the conclusion of Proposition 2.1 may now be interpreted to assert that Y_N is LAD.

At first sight the dependence of $Y_N = y_N(X_1, \ldots, X_N; P)$ on P may seem somewhat unusual. It is introduced to allow us to standardize the statistic of interest and this standardization may depend on P. If $(\mathcal{Y}, \eta)$ is a Euclidean space $\mathbb{R}^k$, one may for instance wish to study an estimator $Z_N = z_N(X_1, \ldots, X_N)$ of a k-dimensional parameter $\zeta(P)$ and in this case the distribution $\tau_N(P)$ of $Y_N = a_N(Z_N - \zeta(P))$ is the distribution of interest for some normalizing sequence of constants a_N. Perhaps an even more compelling reason for allowing Y_N to depend on P is that otherwise its distribution will typically not be estimable in cases of interest.

To see this, suppose that $Y_N = y_N(X_1, \ldots, X_N)$ does not depend on P. In the first place this ensures continuity of each τ_N in view of (2.2) and because $\rho(\tau_N(Q), \tau_N(P))$ will tend to zero if $H(Q^N, P^N)$ does. Hence assumption (i) of Proposition 2.1 is automatically satisfied and we obtain

Corollary 2.1 *Let $\tau_N(P)$ be the law of $Y_N = y_N(X_1, \ldots, X_N)$, where Y_N does not depend on P. If there exists a Prohorov-consistent estimator of $\tau_N(P)$, then there exists a set D of the first category in $(\mathcal{P}, H)$ such that Y_N is LAD at every $P \in \mathcal{P} \backslash D$.*

In other words, the distribution $\tau_N(P)$ of $Y_N = y_N(X_1, \ldots, X_N)$ can only be estimated Prohorov-consistently on the basis of $X_1, \ldots, X_N$ if Y_N is LAD at "most" P. But we already noted that if P and Q are at Hellinger distance of order $N^{-1/2}$, then P^N and Q^N will be essentially different N-dimensional distributions. Hence the LAD property of Y_N at a given P_0 indicates that the distribution of $Y_N = y_N(X_1, \ldots, X_N)$ is insensitive to significant changes of the distribution P^N of the random vector $(X_1, \ldots, X_N)$ in all possible directions around P_0^N. Such a statistic is not of much interest since no statistical procedure based on it will enable us to distinguish between significantly different contiguous models.

Let us therefore return to the case where $Y_N = y_N(X_1, \ldots, X_N; P)$ does indeed depend on P. If $\tau_N(P)$ is to be continuous as well as estimable, then clearly the role of this dependence of Y_N on P must be to offset the change in the distribution of the statistic that would otherwise occur for changes in the underlying distribution P

of Hellinger distance of order $N^{-1/2}$. To arrive at the proper dependence of Y_N on P that will ensure the LAD property which is necessary for the estimability of $\tau_N(P)$, one starts with the statistic of interest $Z_N = z_N(X_1, \ldots, X_N)$ and asks what the influence of a contiguous change in the distribution of the X_i will be on the distribution of Z_N. This is often relatively easy as the statistic is supposedly constructed for some purpose related to distinguishing between the different models $P \in \mathcal{P}$ and the effect of contiguous changes of P on its distribution in usually known to first order. The second step is to construct a dependence on P that will destroy this first order change. This step is basically intelligent guesswork.

Another approach to this problem is to propose a particular dependence of Y_N on P and then try to find out in which cases this produces the desired result of making the distribution $\tau_N(P)$ of Y_N continuous and LAC by checking (2.1). The prime example of this was alluded to above: if $(\mathcal{Y}, \eta)$ is a Euclidean space $\mathbb{R}^k$ and the k-vector Z_N is a consistent estimator of a k-dimensional parameter $\zeta(P)$, then we may choose

$$(2.5) \qquad Y_N = N^{1/2}(Z_N - \zeta(P)), \quad Z_N = z_N(X_1, \ldots, X_N)$$

This transformation will certainly remove at least part of the first order effect of the change in P on the distribution of the statistic Z_N. Moreover, the distribution $\tau_N(P)$ of Y_N is clearly of much interest for statistical purposes. To make sure that the normalization in (2.5) makes sense we assume that

$$(2.6) \qquad \text{the sequence } \{\tau_N(P)\} \text{ is tight for every } P \in \mathcal{P}.$$

The problem of estimating this distribution $\tau_N(P)$ was studied by Efron (1979) who proposed the so-called naïve bootstrap as an estimator.

Let us assume that ζ is continuous. Then the continuity of each individual τ_N is also guaranteed. We saw earlier that for every fixed N the distribution of Z_N is continuous and this continues to hold for the distribution of $N^{1/2}Z_N$. Similarly the continuity of ζ implies the continuity of $N^{1/2}\zeta$ for fixed N. Hence assumption (i) of Proposition 2.1 may be replaced by the requirement that $\zeta : (\mathcal{P}, H) \rightarrow \mathbb{R}^k$ is continuous.

In the present case where Y_N is given by (2.5), the LAC property of $\tau_N(P)$ – or equivalently the LAD property of Y_N – is usually expressed

by saying that Z_N is a *regular estimator* of $\zeta(P)$ or that Z_N is *locally asymptotically equivariant* (LAE). This terminology is usually reserved for the case where the distribution $\tau_N(P)$ of Y_N converges weakly to a limit distribution, but we shall adopt this terminology also in the slightly weaker case where (2.6) holds. Proposition 2.1 now reduces to

Corollary 2.2 *Let* $(\mathcal{Y}, \eta) = \mathbb{R}^k$, *let* $\zeta \; : \; (\mathcal{P}, H) \to \mathbb{R}^k$, *let* Z_N *and* Y_N *as defined in (2.5) be random variables with values in* $\mathbb{R}^k$ *and suppose that (2.3), (2.4) and (2.6) hold. If* ζ *is continuous and there exists a Prohorov-consistent estimator of* $\tau_N(P)$, *then there exists a set* D *of the first category in* $(\mathcal{P}, H)$ *such that* Z_N *is LAE at every* $P \in \mathcal{P} \backslash D$.

Interest in LAE estimators is motivated by the convolution theorem. In the present setting we shall use slight modification of the version of this theorem given in van der Vaart (1989). First of all a path in $\mathcal{P}$ through a fixed point $P \in \mathcal{P}$ is a map $t \to P_t$ from some interval $(0, \varepsilon)$ into $\mathcal{P}$ such that

$$(2.7) \qquad \int [t^{-1}(dP_t^{1/2} - dP^{1/2}) - \frac{1}{2}gdP^{1/2}]^2 \to 0 \;\; \text{as } t \downarrow 0$$

for some $g \in L_2(P)$. The collection of all such functions g constitutes the *tangent cone* $T(P)$ at the point P. Second, we shall assume that the functional $\zeta \; : \; (\mathcal{P}, H) \to \mathbb{R}^k$ is *differentiable* in the sense that there exists a vector-valued function $\dot{\zeta}_P \in L_2(P)^k$ such that for every path P_t through P

$$(2.8) \qquad\qquad t^{-1}(\zeta(P_t) - \zeta(P)) \to \int \dot{\zeta}_P g dP.$$

With these definitions we have a version of the

Convolution Theorem. *In the set-up of Corollary 2.2 with (2.3), (2.4) and (2.6) being satisfied, suppose that* ζ *is differentiable and* Z_N *is LAE. Then for every* P *for which the tangent cone* $T(P)$ *contains a point in the interior of its linear span,*

(2.9)
$$\lim_{N \to \infty} \sup_{\{Q \in \mathcal{P}: \; H(P,Q) \leq CN^{-1/2}\}} \rho(\tau_N(Q), M_N(P) * N(0, \Sigma(\zeta, P))) = 0.$$

Here $\{M_N(P)\}$ denotes a tight and LAE sequence of probability distributions on $\mathbb{R}^k$ which only depends on P, * denotes convolution and

$N(0, \Sigma(\zeta, P))$ denotes the k-variate normal distribution with mean 0 and as a covariance matrix the Cramér-Rao bound $\Sigma(\zeta, P)$ for estimating ζ.

Combining Corollary 2.2 and the convolution theorem we find

Corollary 2.3 *Consider the set-up of Corollary 2.2 with (2.3), (2.4) and (2.6) being satisfied. Suppose that ζ is differentiable and that there exists a Prohorov-consistent estimator of $\tau_N(P)$. Then there exists a set D of the first category in $(\mathcal{P}, H)$ such that (2.9) holds for all $P \in \mathcal{P} \backslash D$ for which the tangent cone $T(P)$ contains an interior point of its linear span.*

Now let us step back for a moment and examine what we have shown. In a large non-parametric model $\mathcal{P}$, the tangent cone $T(P)$ will contain an interior point of its linear span for many points $P \in \mathcal{P}$. In fact, $T(P)$ may well be a linear space for all interior points $P \in \mathcal{P}$. In such cases we find roughly speaking that if we estimate a differentiable functional $\zeta(P)$ at rate $N^{-1/2}$ by a statistic $Z_N = z_N(X_1, \ldots, X_N)$, then at "most" points $P \in \mathcal{P}$ it is not possible to estimate the distribution $\tau_N(P)$ of $Y_N = N^{1/2}(Z_N - \zeta(P))$ consistently, unless this distribution behaves asymptotically like $M_N(P) * N(0, \Sigma(\zeta, P))$ under all sequences of underlying distributions inside Hellinger balls of radius $N^{-1/2}$ around P.

Let us compare Corollary 2.3 with the results obtained in Beran (1997). Theorem 2.1 in Beran (1997) deals with the parametric case where $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^k$ under the assumption that $\tau_N(P)$ converges to a limit distribution, the support of which satisfies a technical condition. Otherwise, the conditions are more or less similar to ours: a LAN assumption and Fréchet-differentiability of ζ. Among other things this theorem asserts that if the parametric bootstrap estimates $\tau_N(P)$ Prohorov-consistently, then (2.9) will hold for all $\theta \in \Theta$. There is no exceptional set where the conclusion does not hold. The important differences with Corollary 2.3 in the present paper are that we do not assume convergence but only tightness of $\tau_N(P)$, but that we do have an exceptional set of the first category. Incidentally this set of the first category corresponds to a set of Lebesgue measure zero in the parameter space (cf. Putter (1994, Theorem 2.6.1). The fact that we discuss estimability of $\tau_N(P)$ as opposed to consistency of a particular estimator – the parametric bootstrap – is a minor matter.

Theorem 2.2 in Beran (1997) deals with general non-parametric families $\mathcal{P}$. This is achieved by replacing the class $\mathcal{P}$ of all possible

distributions on $(\mathcal{X}, \mathcal{A})$ by the class $\mathcal{P}_0$ of all discrete distributions with finite support, which is dense in $\mathcal{P}$. This reduces the problem to a finite dimensional one which can be treated as before. Instead of an exceptional set of the first category, the exceptional set is now $\mathcal{P} \backslash \mathcal{P}_0$. Again the main difference with Corollary 2.3 above is the requirement of weak convergence of $\tau_N(P)$ instead of tightness. We shall return to this matter in Section 3.

There is also an interesting connection with van Zwet (1995). One of the conclusions of that paper is that, roughly speaking, the naïve bootstrap can only be consistent if Y_N is either asymptotically normal or asymptotically distributionfree (or of course a sum of variables of both types). This seems to come very close to Corollary 2.3.

3 Local Asymptotic Constancy

In this section we do not assume throughout that the topological space $(\mathcal{P}, \Pi)$ is metrizable. We shall say that the sequence $\tau_N : (\mathcal{P}, \Pi) \rightarrow (\mathcal{R}, \rho)$ is *equicontinuous* at $P \in \mathcal{P}$ if for every $\varepsilon > 0$ there exists a neighborhood U_ε of P such that $Q \in U_\varepsilon$ implies that

$$(3.1) \qquad \qquad \sup_N \rho(\tau_N(Q), \tau_N(P)) \le \varepsilon.$$

The following proposition is a re-statement of Lemma 4.2 in Putter and van Zwet (1996)

Proposition 3.1 *Suppose that*

(i) For every $N = 1, 2, \ldots$, the map $\tau_N : (\mathcal{P}, \Pi) \rightarrow (\mathcal{R}, \rho)$ is continuous;

(ii) τ_N converges pointwise to a limit τ.

Then there exists a set D of the first category in $(\mathcal{P}, \Pi)$ such that the sequence τ_N is equicontinuous at every $P \in \mathcal{P} \backslash D$.

A comparison with the results in Section 2 shows that the convergence $\tau_N(P) \rightarrow \tau(P)$ for all $P \in \mathcal{P}$ makes a great deal of difference for the estimability of $\tau_N(P)$. Obviously, equicontinuity of τ_N on $\mathcal{P} \backslash D$ implies that if we can find a consistent estimator $\hat{P}_N$ of P, then $\tau_N(P)$ can be estimated consistently for all $P \in \mathcal{P} \backslash D$. But a comparison with Proposition 2.1 also shows that if $\tau_N \rightarrow \tau$, we no longer need the estimability of $\tau_N(P)$ to conclude that $\{\tau_N(P)\}$ is LAC at every

$P \in \mathcal{P} \backslash D$. In fact, we now have more because if Π is metrized by the Hellinger metric, (3.1) ensures that for any $\delta_N \downarrow 0$,

$$(3.2) \qquad \lim_{N \to \infty} \sup_{\{Q \in \mathcal{P} \ : \ H(P,Q) \leq \delta_N\}} \rho(\tau_N(Q), \tau_N(P)) = 0,$$

which is much stronger than (2.1).

References

Beran, R.J. (1997), Diagnosing bootstrap success, *Ann. Inst. Statist. Math.* **49**, 1-24.

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

Dudley, R.M. (1989), *Real Analysis and Probability*, Wadsworth, Belmont.

Efron, B. (1979), The bootstrap: Another look at the jackknife, *Ann. Statist.*, **7**, 1-26

Le Cam, L.M. and Yang, G.L. (1990), *Asymptotics in Statistics: Some Basic Concepts*, Springer-Verlag, New York.

Putter, H (1994), *Consistency of Resampling Methods*, PhD thesis, University of Leiden.

Putter, H and van Zwet, W. R. (1996), Resampling: Consistency of substitution estimators, *Ann. Statist.* **24**, 2297-2318

van der Vaart, A.W. (1989), On the asymptotic information bound, *Ann. Statist.* **17**, 1487-1500

van Zwet, W.R. (1995), Resampling: The jackknife and the naïve bootstrap. Unpublished manuscript.