Chapter 10
Finite samples and asymptotics

Chris A.J. Klaassen

Abstract Willem van Zwet is a scientist and a scholar with a broad spectrum of research interests. This is reflected by the five papers in this section, which study very different fundamental problems and which have four of his PhD students and his youngest son as coauthor.

10.1 Introduction

Willem van Zwet is a scientist and a scholar with a broad spectrum of research interests within statistics and probability theory with the stress on statistics. This is clear from his list of publications. This breadth is obvious also from the selection of papers we discuss in this section: asymptotic normality of rank test statistics, Hellinger distance and contiguity, estimation of parameters and score functions, and consistency and asymptotic equivariance. Willem has tried to instill this attitude of broad interest towards research also into his PhD students.

As a thesis advisor Willem van Zwet has been unparalleled. He didn't suggest standard problems to his students, but he has boosted their careers by tackling interesting, fundamental, relevant problems. Most of his PhD students have become full professor in statistics (12 out of 16, so far). Actually the five papers in this section all have been written with scientific children, i.e. PhD students, as coauthors, except for the most recent one, which has been written in collaboration with his youngest biological son Erik.

One of the distinctions between these five papers is that three of them belong to the realm of asymptotic statistics, whereas the other two are of the finite sample type. Research in mathematical statistics is or should be motivated mainly by

Chris A.J. Klaassen
Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
e-mail: C.A.J.Klaassen@uva.nl

real life problems where data have to be interpreted. Since data always come in finite amounts, the core business of mathematical statistics should be finite sample statistics. Typically however, finite sample problems are difficult to handle mathematically. That is why one resorts to approximations, which are obtained typically by letting the sample size tend to infinity. This results in asymptotic statistics. This approach is applied quite often; actually, to such an extent that asymptotic statistics has become the norm, or at least the benchmark at which different techniques are compared.

10.2 Asymptotic Normality of Nonparametric Tests for Independence

David van Dantzig (1900–1959) is one of the initiators of the Mathematical Centre in Amsterdam. It was founded in February 1947 as a non-profit institute aiming at the promotion of pure mathematics and its applications. The topologist Van Dantzig intended to contribute to the reconstruction of Dutch society after the second world war by stimulating the study and application of mathematical statistics. In his philosophy the statistician should choose for his data a statistical model with the weakest of possible assumptions. Therefore, a main theme at the Mathematical Centre, the cradle of much of Dutch mathematical statistics, has been the study of nonparametric and rank procedures during the first decades of its existence. These methods had developed strongly, also internationally, with the monograph of Hájek and Šidák (1967) as a milestone. The present paper fits perfectly well in this tradition. It presents conditions for asymptotic normality for a class of rank statistics used for testing independence, and it seems to be one of the last ones in a series of papers that present weaker and weaker conditions for asymptotic normality to hold. We mention Wald and Wolfowitz (1944), Noether (1949), Hoeffding (1951), Hájek (1961), and Bhuchongkul (1964).

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent and identically distributed random vectors with continuous distribution function $H(\cdot, \cdot)$ on $\mathbb{R}^2$ and marginal distribution functions $F(\cdot)$ and $G(\cdot)$. Locally most powerful rank tests of the null hypothesis of independence $H(\cdot, \cdot) = F(\cdot)G(\cdot)$ are of the type

$$T_n = \frac{1}{n} \sum_{i=1}^{n} J_n\left(\frac{R_i}{n}\right) K_n\left(\frac{Q_i}{n}\right) = \frac{1}{n} \sum_{i=1}^{n} J_n\left(F_n(X_i)\right) K_n\left(G_n(Y_i)\right)$$

$$= \int \int J_n\left(F_n(x)\right) K_n\left(G_n(y)\right) dH_n(x, y), \qquad (10.1)$$

where R_i is the rank of X_i among $X_1, \ldots, X_n$, where Q_i is the rank of Y_i among $Y_1, \ldots, Y_n$, and where $F_n(\cdot), G_n(\cdot)$, and $H_n(\cdot, \cdot)$ are the empirical versions of $F(\cdot), G(\cdot)$, and $H(\cdot, \cdot)$, respectively. Assuming existence of functions $J(\cdot)$ and $K(\cdot)$ on the unit interval such that

$$\int \int \left[J_n\left(F_n(x)\right) K_n\left(G_n(y)\right) - J\left(F_n(x)\right) K\left(G_n(y)\right) \right] dH_n(x,y) = o_P\left(\frac{1}{\sqrt{n}}\right) \quad (10.2)$$

holds and with the notation

$$\mu = \int \int J\left(F(x)\right) K\left(G(y)\right) dH(x,y) \qquad (10.3)$$

we note heuristically that

$$
\begin{aligned}
T_n - \mu &= \int \int J\left(F_n(x)\right) K\left(G_n(y)\right) dH_n(x,y) \\
&\quad - \int \int J\left(F(x)\right) K\left(G(y)\right) dH(x,y) + o_P\left(\frac{1}{\sqrt{n}}\right) \\
&= \int \int \left\{ J\left(F_n(x)\right) - J\left(F(x)\right) \right\} K\left(G_n(y)\right) dH_n(x,y) \\
&\quad + \int \int J\left(F(x)\right) \left\{ K\left(G_n(y)\right) - K\left(G(y)\right) \right\} dH_n(x,y) \qquad (10.4) \\
&\quad + \int \int J\left(F(x)\right) K\left(G(y)\right) d\left\{ H_n(x,y) - H(x,y) \right\} + o_P\left(\frac{1}{\sqrt{n}}\right) \\
&= \int \int \left\{ F_n(x) - F(x) \right\} J'\left(F(x)\right) K\left(G(y)\right) dH(x,y) \\
&\quad + \int \int \left\{ G_n(y) - G(y) \right\} J\left(F(x)\right) K'\left(G(y)\right) dH(x,y) \\
&\quad + \int \int J\left(F(x)\right) K\left(G(y)\right) d\left\{ H_n(x,y) - H(x,y) \right\} + o_P\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}
$$

holds. Consequently, it is intuitively clear that asymptotic normality of $\sqrt{n}(T_n - \mu)$ holds with limit variance as in (3.10) of the paper, and that this asymptotic normality will hold uniformly over appropriate classes of distribution functions $H(\cdot,\cdot)$. However, the technical difficulties are many, especially since the authors have been aiming at minimal conditions. Actually they have replaced (10.4) above by an expression with 13 terms, 10 of which had to be shown to be asymptotically negligible.

Since $H(\cdot,\cdot)$ is not necessarily equal to $F(\cdot)G(\cdot)$, the asymptotic normality is shown under so-called fixed alternatives to the hypothesis of independence. A well-known example is the Van der Waerden normal scores rank correlation coefficient with $J(u) = K(u) = \Phi^{-1}(u) \approx \sqrt{-2\log[u(1-u)]}$ and $J'(u) = K'(u) = 1/\phi(\Phi^{-1}(u)) \approx \sqrt{2\pi}\,[u(1-u)]^{-1}, 0 < u < 1$. Clearly conditions are needed on the behavior of $J(u)$ and $K(u)$ as u comes close to 0 or 1, in order for (10.4) to be valid and for the asymptotic normality to hold. Ruymgaart, Shorack, and Van Zwet improve on the conditions of Bhuchongkul (1964), and for the case $J(\cdot) = K(\cdot)$ they need that $|J(u)|\,[u(1-u)]^{1/4-\delta}$ and $|J'(u)|\,[u(1-u)]^{5/4-\delta}$ are bounded for some positive δ, thus incorporating the Van der Waerden normal scores rank correlation coefficient.

Frits Ruymgaart, Willem's third PhD student, has generalized these results in Ruymgaart (1974) and in his PhD thesis, Ruymgaart (1973), to the case where the

score functions $J(\cdot)$ and $K(\cdot)$ may have discontinuities. It is very likely that this is the reason why Ruymgaart (1974) has gotten many more citations in the literature than the article under discussion. In any case, these papers are still being cited because of their relevance to the study of semiparametric copula models. For example the Van der Waerden normal scores rank correlation coefficient is semiparametrically efficient in the normal copula model; see Klaassen and Wellner (1997). In the normal copula model one assumes that if all components of a random vector are transformed into normal random variables, then the resulting random vector has a multivariate normal distribution. It was noted by Li (2000) that this normal copula model was in use for the pricing of credit default swaps. This practice has been blamed for the global financial crisis of 2007–2009. Motivated by risk management problems, the study of copula models has led to research on a generalization of T_n from (10.1), namely

$$T_n = \frac{1}{n} \sum_{i=1}^{n} J\left(\frac{R_i}{n}, \frac{Q_i}{n}\right) = \frac{1}{n} \sum_{i=1}^{n} J\left(F_n(X_i), G_n(Y_i)\right); \qquad (10.5)$$

see e.g. Fermanian, Radulovic, and Wegkamp (2004) and Schmidt and Stadtmüller (2006). We may conclude that Ruymgaart, Shorack, and Van Zwet have added a technically complicated, but thorough and useful result to the statistical literature, which is a milestone in a long development.

10.3 A Note on Contiguity and Hellinger Distance

Consider two sequences of probability measures (P_n) and (Q_n) defined on a common sequence of measurable spaces $(\mathscr{X}_n, \mathscr{A}_n)$. The sequence of probability measures (Q_n) is called *contiguous* with respect to the sequence (P_n), if for every sequence $(A_n), A_n \in \mathscr{A}_n$, the convergence $P_n(A_n) \to 0$ implies $Q_n(A_n) \to 0$, notation $(Q_n) \lhd (P_n)$. The two-sided version of this fundamental concept has been introduced by Le Cam (1960), and advertised and applied by Roussas (1972). Jaroslav Hájek visited Lucien Le Cam in Berkeley and popularized the concept in Hájek and Šidák (1967) calling the most important results Le Cam's first, second, and third lemma.

Willem van Zwet and his first PhD student, his contemporary Kobus Oosterhoff, were the first to geometrize contiguity for product measures, and they chose the Hellinger distance as a natural metric for this. Their results have been published as Oosterhoff and Van Zwet (1979) in the Hájek Memorial Volume, which is just the proper place for this paper, given Hájek's interest in the topic.

Let $P_n = P_n^{(n)}$ be the product measure $P_n^{(n)} = \prod_{i=1}^{n} P_{ni}$ and $Q_n = Q_n^{(n)} = \prod_{i=1}^{n} Q_{ni}$ with $(\mathscr{X}_n, \mathscr{A}_n)$ the product space of $(\mathscr{X}_{ni}, \mathscr{A}_{ni}), i = 1, \ldots, n$. Let μ_{ni} be a σ-finite measure on $(\mathscr{X}_{ni}, \mathscr{A}_{ni})$ dominating both P_{ni} and Q_{ni}. Denote the densities of P_{ni} and Q_{ni} by p_{ni} and q_{ni}, respectively, and write $H(P_{ni}, Q_{ni})$ for the Hellinger distance of P_{ni} and Q_{ni},

$$H(P_{ni}, Q_{ni}) = \sqrt{\int \left(p_{ni}^{1/2} - q_{ni}^{1/2} \right)^2 d\mu_{ni}}. \tag{10.6}$$

It is easy to see, as Oosterhoff and Van Zwet (1979) show, that

$$\lim_{n \to \infty} \sum_{i=1}^{n} H^2(P_{ni}, Q_{ni}) = 0 \implies (Q_n^{(n)}) \lhd (P_n^{(n)})$$

$$\implies \limsup_{n \to \infty} \sum_{i=1}^{n} H^2(P_{ni}, Q_{ni}) < \infty \tag{10.7}$$

holds. As the one-sided contiguity is an 'asymmetric' property and the Hellinger distance is symmetric, the gap between the contiguity in the middle term and the left hand or the right hand side of (10.7) can be closed only by an asymmetric condition. Indeed, with the additional notation

$$H_c(P_{ni}, Q_{ni}) = \sqrt{\int_{|q_{ni} - p_{ni}| \geq c p_{ni}} \left(p_{ni}^{1/2} - q_{ni}^{1/2} \right)^2 d\mu_{ni}} \tag{10.8}$$

the main result of Oosterhoff and Van Zwet (1979) may be formulated as follows,

$$(Q_n^{(n)}) \lhd (P_n^{(n)}) \iff \begin{cases} \limsup_{n \to \infty} \sum_{i=1}^{n} H^2(P_{ni}, Q_{ni}) < \infty, \\[2mm] \lim_{c \to \infty} \limsup_{n \to \infty} \sum_{i=1}^{n} H_c^2(P_{ni}, Q_{ni}) = 0. \end{cases} \tag{10.9}$$

The second result from this paper is related to the First Lemma of Le Cam. Let Λ_n be the loglikelihood ratio

$$\Lambda_n = \sum_{i=1}^{n} \log \left(q_{ni}(X_{ni}) / p_{ni}(X_{ni}) \right). \tag{10.10}$$

The First Lemma of Le Cam implies

$$\Lambda_n \xrightarrow{P_n^{(n)}}_w \mathcal{N}(-\frac{1}{2}\sigma^2, \sigma^2) \implies (Q_n^{(n)}) \lhd (P_n^{(n)}). \tag{10.11}$$

Oosterhoff and Van Zwet succeeded in formulating necessary and sufficient conditions for this convergence to normality of the loglikelihood ratio in terms of Hellinger distances as follows. For any $\sigma \geq 0$ we have

$$\left. \begin{array}{l} \Lambda_n \xrightarrow{P_n^{(n)}}_w \mathcal{N}(-\frac{1}{2}\sigma^2, \sigma^2), \\[3mm] \lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \max_{1 \leq i \leq n} H_\varepsilon^2(P_{ni}, Q_{ni}) = 0 \end{array} \right\} \iff$$

$$\begin{cases} \lim_{n \to \infty} \sum_{i=1}^{n} H^2(P_{ni}, Q_{ni}) = \frac{1}{4}\sigma^2, \\[2mm] \lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \sum_{i=1}^{n} H_\varepsilon^2(P_{ni}, Q_{ni}) = 0. \end{cases} \tag{10.12}$$

In order to stress the relation to the Hellinger distance we have chosen to formulate
(2.3) of Theorem 1 and the uniform asymptotic negligibility condition (3.3) of The-
orem 2 of Oosterhoff and Van Zwet (1979) in terms of $H_c(P_{ni}, Q_{ni})$ from (10.8); cf.
(3.7) of ibid. To see the equivalence of (2.3) of Theorem 1 with the second condition
at the right hand side of (10.9), we note that for $c > 2$

$$(1 - c^{-1/2})^{-2} H_c^2(P_{ni}, Q_{ni}) \leq Q_{ni}(q_{ni}(X_{ni}) \geq c\, p_{ni}(X_{ni}))$$
$$\leq (1 - c^{-1})^{-2} H_c^2(P_{ni}, Q_{ni}) \qquad (10.13)$$

holds. To derive the equivalence of the uniform asymptotic negligibility condition
(3.3) of Theorem 2 with the second condition at the left hand side of (10.12) we note
that for $\varepsilon > 0$

$$(\sqrt{1+\varepsilon} - 1)^2 P_{ni}(|q_{ni}(X_{ni}) - p_{ni}(X_{ni})| \geq \varepsilon\, p_{ni}(X_{ni}))$$
$$\leq H_\varepsilon^2(P_{ni}, Q_{ni}) \qquad (10.14)$$
$$\leq P_{ni}(|q_{ni}(X_{ni}) - p_{ni}(X_{ni})| \geq \varepsilon\, p_{ni}(X_{ni}))$$
$$+ Q_{ni}(|q_{ni}(X_{ni}) - p_{ni}(X_{ni})| \geq \varepsilon\, p_{ni}(X_{ni}))$$

holds, and we use (10.11) to obtain that the uniform asymptotic negligibility condi-
tion

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \max_{1 \leq i \leq n} P_{ni}(|q_{ni}(X_{ni}) - p_{ni}(X_{ni})| \geq \varepsilon\, p_{ni}(X_{ni})) = 0 \qquad (10.15)$$

implies

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \max_{1 \leq i \leq n} Q_{ni}(|q_{ni}(X_{ni}) - p_{ni}(X_{ni})| \geq \varepsilon\, p_{ni}(X_{ni})) = 0. \qquad (10.16)$$

It is well known that contiguity holds in regular parametric models for i.i.d. ran-
dom variables. Let $\{P(\theta) : \theta \in \Theta\}, \Theta \subset \mathbb{R}^k$, be a collection of distributions that
have densities $p(\theta)$ with respect to some σ-finite measure μ. The most important
condition for regularity is the existence of a score function $\dot{\ell}(\theta) \in L_2^k(P(\theta))$ such
that the map $\theta \mapsto p^{1/2}(\theta)$ is continuously Fréchet differentiable in $L_2(\mu)$ as follows,

$$|h|^{-2} \int \left(p^{1/2}(\theta + h) - p^{1/2}(\theta) - \frac{1}{2} h^T \dot{\ell}(\theta) p^{1/2}(\theta) \right)^2 d\mu \to 0, h \to 0, \quad (10.17)$$

with the map $\theta \mapsto \dot{\ell}(\theta) p^{1/2}(\theta)$ from Θ to $L_2^k(P(\theta))$ continuous. In these regular
models Local Asymptotic Normality holds and via (10.11) this yields the contiguity

$$\{P^n(\theta_n + n^{-1/2} t_n)\} \lhd \{P^n(\theta_n)\}, \qquad (10.18)$$

as $n \to \infty$, $\theta_n \to \theta$, and $t_n \to t$ for fixed θ and t, and hence the corresponding mutual
contiguity; see e.g. Section 2.1 of Bickel et al. (1993). To circumvent the cum-
bersome proof of Local Asymptotic Normality, one might use the characterization
(10.9) of contiguity in order to prove (10.18) as follows. By (a) from the proof of

Lemma A.9.5 of Bickel et al. (1993) we see that (10.17) holds uniformly for θ in compacts. Consequently,

$$n \int \left(p^{1/2}(\theta_n + n^{-1/2} t_n) - p^{1/2}(\theta_n) - \frac{1}{2} n^{-1/2} t_n^T \ell(\theta_n) p^{1/2}(\theta_n) \right)^2 d\mu \to 0 \quad (10.19)$$

holds as $n \to \infty$, $\theta_n \to \theta$, and $t_n \to t$ for fixed θ and t. It follows that the first condition at the right hand side of (10.9) is satisfied. To prove that the second condition at the right hand side of (10.9) is satisfied as well, we first note that for $c \geq 1$

$$\int_{p(\theta+h) \geq c\, p(\theta)} \left(p^{1/2}(\theta + h) - p^{1/2}(\theta) \right)^2 d\mu$$

$$\leq 3 \int_{p(\theta+h) \geq c\, p(\theta)} \left(p^{1/2}(\theta + h) - p^{1/2}(\theta) - \frac{1}{2} h^T \ell(\theta) p^{1/2}(\theta) \right)^2 d\mu$$

$$+ 3 \int_{p(\theta+h) \geq c\, p(\theta)} \left(p^{1/2}(\theta - h) - p^{1/2}(\theta) + \frac{1}{2} h^T \ell(\theta) p^{1/2}(\theta) \right)^2 d\mu \quad (10.21)$$

$$+ 3 \int_{p(\theta+h) \geq c\, p(\theta)} p(\theta) d\mu$$

and

$$\int_{p(\theta+h) \geq c\, p(\theta)} \left\{ \left(p^{1/2}(\theta + h) - p^{1/2}(\theta) \right)^2 - \left(c^{1/2} - 1 \right)^2 p(\theta) \right\} d\mu \geq 0 \quad (10.21)$$

hold, and hence for sufficiently large values of c

$$\int_{p(\theta+h) \geq c\, p(\theta)} \left(p^{1/2}(\theta + h) - p^{1/2}(\theta) \right)^2 d\mu$$

$$\leq 3 \left[1 - 3(\sqrt{c} - 1)^{-2} \right]^{-1} \times \qquad\qquad (10.22)$$

$$\int_{p(\theta+h) \geq c\, p(\theta)} \left\{ \left(p^{1/2}(\theta + h) - p^{1/2}(\theta) - \frac{1}{2} h^T \ell(\theta) p^{1/2}(\theta) \right)^2 \right.$$

$$\left. + \left(p^{1/2}(\theta - h) - p^{1/2}(\theta) + \frac{1}{2} h^T \ell(\theta) p^{1/2}(\theta) \right)^2 \right\} d\mu. \quad (10.23)$$

To complete the proof of the contiguity in (10.18), we note that the second condition at the right hand side of (10.9) follows from (10.19) and (10.22) by the substitutions $\theta = \theta_n$ and $h = n^{-1/2} t_n$.

One of the simplest examples of nonregular parametric models for i.i.d. random variables where contiguity may be determined easily via (10.9), is the location family for the exponential distribution. With μ_{ni} Lebesgue measure on $(\mathbb{R}, \mathscr{B})$ and

$$p_{ni}(x) = e^{-x} \mathbf{1}_{(0,\infty)}(x), \quad q_{ni}(x) = p_{ni}(x - \frac{t}{n}), \quad x \in \mathbb{R}, \quad (10.24)$$

some computation shows

$$\limsup_{n\to\infty} \sum_{i=1}^{n} H^2(P_{ni}, Q_{ni}) = |t| \tag{10.25}$$

and

$$\lim_{c\to\infty} \limsup_{n\to\infty} \sum_{i=1}^{n} H_c^2(P_{ni}, Q_{ni}) = |t| \mathbf{1}_{[t<0]}, \tag{10.26}$$

which by (10.9), i.e. Theorem 1 of Oosterhoff and Van Zwet (1979), implies the one-sided contiguity from the left hand side of (10.9), but only for t positive.

More tedious computations are necessary in case of triangular densities

$$p_{ni}(x) = (1 - |x|) \vee 0, \quad q_{ni}(x) = (1 - |x - \frac{\sigma}{\sqrt{n\log n}}|) \vee 0, \quad x \in \mathbb{R}. \tag{10.27}$$

In my master thesis, Klaassen (1974), written under supervision of Kobus Oosterhoff, the conditions at the right hand side of (10.12) have been checked, and hence Theorem 2 of Oosterhoff and Van Zwet (1979) is applicable here.

The main result, Theorem 1 of ibid., is called by Jacod and Shiryaev (1987, p. 576) 'the first general contiguity result'. It has been generalized to nonindependent and continuous time cases, and it has been one of the roots of Jacod and Shiryaev (1987).

Furthermore, the paper is cited in several publications as a reference for contiguity. We mention Strasser (1985), Bickel, Klaassen, Ritov, and Wellner (1993), Cabaña and Cabaña (1997) and Bose, Gangopadhyay, and Goswami (2007). The results are explicitly used in e.g. Khmaladze (1988), Eubank (2000), Pfanzagl (2000), Putter and Young (2001), and Ferger (2001).

We conclude that Oosterhoff and Van Zwet (1979) is a fundamental paper, which presents useful tools for verifying contiguity.

10.4 On Estimating a Parameter and its Score Function
On Estimating a Parameter and its Score Function, II

In the early seventies several semiparametrically efficient estimators for the symmetric location model have been constructed. These estimators were called adaptive because they adapt to the unknown underlying symmetric density $f(\cdot)$ of the errors in the observations in such a way that they attain the asymptotic variance bound of $1/I(f)$ with $I(f)$ the Fisher information for location. After a chat with Peter Huber about these estimators, Willem van Zwet suggested me, his PhD student, to study these estimators and to show that there is some loss somewhere. This has been a very stimulating research program and it resulted in my thesis Klaassen (1981). The articles under discussion are generalizations to much more general models of the inequality in Theorem 3.2.1 of ibid. for the symmetric location case. These inequal-

ities improve the Cramér-Rao lower bound for unbiased estimators of the parameter of interest by adding a multiple of the integrated mean squared error of an estimator, given the parameter of interest, of the efficient influence function. In this way they state that the parameter of interest can be estimated accurately only if, given the parameter of interest, the efficient influence function can be estimated accurately as well. So, if the parameter of interest is the only unknown parameter, this additional term vanishes and the inequality reduces to the ordinary Cramér-Rao inequality.

The charm of these inequalities is that they are finite sample results. They have been proved for i.i.d. situations where, given the parameter of interest, there exists a sufficient statistic with respect to the nuisance parameter. All these inequalities are based on a conditional version of Projection Lemma 4.1 of Hájek (1968). An asymptotic version of these results does not need these sufficient statistics and it states that asymptotically efficient estimation of the parameter of interest is possible if and only if the efficient influence function can be estimated consistently; for a generalization of this, see Klaassen (1987).

The first paper is applicable in situations where adaptive estimation should be possible, i.e. where the semiparametrically efficient influence function is the same as the efficient influence function for the case that the nuisance parameter is known. The second paper studies the general semiparametric situation. Still another PhD student of Willem is a coauthor here, namely Aad van der Vaart. In chapter 5 of his PhD thesis Van der Vaart (1988), he continues research on models of the above type. There he constructs asymptotically efficient estimators for semiparametric models with a sufficient statistic with respect to the nuisance parameter.

10.5 A Remark on Consistent Estimation

A fundamental rule of thumb in statistics states that 'substituting unknown parameters in statistical procedures by estimators of them, yields appropriate procedures.' Consequently, if one is simulating the distribution of a statistic and the distribution of the underlying random variables is unknown, one may replace the latter distribution by an estimator of it, like the empirical. The resulting bootstrap was introduced by Efron (1979). Clearly, an important question is :'When does the bootstrap work and when it doesn't?' For some important classes of situations the validity of the bootstrap was proved by Bickel and Freedman (1981), who also presented some counter-examples.

Hein Putter, writing his doctoral thesis Putter (1994) under supervision of Willem van Zwet, has studied the question in the setting of a general substitution estimator. In Putter and Van Zwet (1996) they write 'This is commonly called a "plug-in esti-mator," but this expression is of the same sad grammatical level as "see-through clothes."' However, more importantly in the context of the paper under discussion they prove that substitution estimators work under all underlying distributions, except for a 'small' subset within the set of underlying distributions metrized by Hellinger distance, namely for a subset of the first category.

Under local asymptotic normality Beran (1997) has proved that the bootstrap fails in estimating the distribution of an estimator consistently, precisely at those parameter values at which the bootstrapped estimator is not locally asymptotically equivariant. He also has shown that the set of these parameter values has Lebesgue measure 0. At the points where an estimator is locally asymptotically equivariant or regular, the Hájek-LeCam convolution theorem holds; Hájek (1970).

Willem van Zwet and his second son Erik have used results from Putter and Van Zwet (1996) in order to prove within a very general framework that, if the distribution of an estimator can be estimated consistently in the Prohorov metric, then there exists a subset of the first category within the set of underlying distributions metrized by Hellinger distance, such that the estimator is locally asymptotically equivariant outside this subset of the first category. In Van Zwet and Van Zwet (1999) they prove under somewhat stricter conditions, but still within the same very general framework, that for the same subset of the first category as above the distribution of the estimator has a locally asymptotically uniform convolution structure. We conclude that Willem and Erik van Zwet have generalized the connection between validness of the bootstrap and local asymptotic equivariance and convolution structure as noted by Beran (1997), to the general i.i.d. case in an elegant way.

References

1. Beran, R.J. (1997), Diagnosing bootstrap success, *Ann. Inst. Statist. Math.* **49**, 1–14.
2. Bhuchongkul, S. (1964), A class of nonparametric tests for independence in bivariate populations, *Ann. Math. Statist.* **35**, 138–149.
3. Bickel, P.J. and Freedman, D.A. (1981), Some Asymptotic Theory for the Bootstrap, *Ann. Statist.* **9**, 1196-1217.
4. Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. (1993), *Efficient and Adaptive Estimation in Semiparametric Models*, Johns Hopkins Univ. Press, Baltimore; reprint (1998), Springer, New York.
5. Bose, A., Gangopadhyay, S., and Goswami, A. (2007), A note on random coin tossing, *Indag. Math. (N.S.)* **18**, 405–416.
6. Cabaña, A. and Cabaña, E.M. (1997), Transformed empirical processes and modified Kolmogorov-Smirnov tests for multivariate distributions, *Ann. Statist.* **25**, 2388–2409.
7. Efron, B. (1979), Bootstrap Methods: Another Look at the Jackknife, *Ann. Statist.* **7**, 1–26.
8. Eubank, R.L. (2000), Testing for no effect by cosine series methods, *Scand. J. Statist.* **27**, 747–763.
9. Ferger, D. (2001), Analysis of change-point estimators under the null hypothesis, *Bernoulli* **7**, 487–506.
10. Fermanian, J.-D., Radulovic, D., and Wegkamp, M. (2004), Weak convergence of empirical copula processes, *Bernoulli* **10**, 847–860.
11. Hájek, J. (1961), Some extensions of the Wald-Wolfowitz-Noether theorem, *Ann. Math. Statist.* **32**, 506–523.
12. Hájek, J. (1968), Asymptotic normality of simple linear rank statistics under alternatives, *Ann. Math. Statist.* **39**, 325–346.
13. Hájek, J. (1970), A characterization of limiting distributions of regular estimates, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14**, 323–330.

14. Hájek, J., and Šidák, Z. (1967), *Theory of rank tests*, Academic Press, New York, Academia Publishing House of the Czechoslovak Academy of Sciences, Prague.
15. Hoeffding, W. (1951), A combinatorial central limit theorem, *Ann. Math. Statist.* **22**, 558-566.
16. Jacod, J. and Shiryaev, A.N. (1987), *Limit Theorems for Stochastic Processes*, Springer, Berlin.
17. Khmaladze, E.V. (1988), An innovation approach to goodness-of-fit tests, *Ann. Statist.* **16**, 1503–1516.
18. Klaassen, C.A.J. (1974), Applications of Hellinger distance in hypothesis testing (in Dutch), Master thesis, University of Nijmegen.
19. Klaassen, C.A.J. (1981). *Statistical Performance of Location Estimators*, Mathematical Centre Tracts 133, Mathematical Centre, Amsterdam.
20. Klaassen, C.A.J. (1987). Consistent estimation of the influence function of locally asymptotically linear estimators. *Ann. Statist.* **15**, 1548–1562.
21. Klaassen, C.A.J. and Van Zwet, W.R. (1985), On estimating a parameter and its score function, *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer (Berkeley, Calif., 1983)*, L.M. Le Cam and R.A. Olshen (eds.), Wadsworth Statist./Probab. Ser., Wadsworth, Belmont, CA, **II**, 827–839.
22. Klaassen, C.A.J., Van der Vaart, A.W., and Van Zwet, W.R. (1988), On estimating a parameter and its score function, II, *Statistical Decision Theory and Related Topics IV, (West Lafayette, Ind., 1986)*, S.S. Gupta and J.O. Berger (eds.), Springer, New York, **2**, 281–288.
23. Klaassen, C.A.J. and Wellner, J.A. (1997), Efficient estimation in the bivariate normal copula model: normal margins are least favourable, *Bernoulli* **3**, 55–77.
24. Le Cam, L.M. (1960), Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses, *Univ. California Publ. Statist.* **3**, 37–98.
25. Li, D.X. (2000), On Default Correlation: A Copula Function Approach, *Journal of Fixed Income* **9**, 43-54.
26. Noether, G.E. (1949), On a theorem of Wald and Wolfowitz, *Ann. Math. Statist.* **20**, 455–458.
27. Oosterhoff, J. and Van Zwet, W.R. (1979), A note on contiguity and Hellinger distance. *Contributions to Statistics, Jaroslav Hájek Memorial Volume*, J. Jurečková (ed.), Reidel, Dordrecht, 157–166.
28. Pfanzagl, J. (2000), On local uniformity for estimators and confidence limits, *J. Statist. Plann. Inf.* **84**, 27–53.
29. Putter, H. (1994), *Consistency of Resampling Methods*, PhD thesis, University of Leiden.
30.
31. Putter, H. and Van Zwet, W.R. (1996), Resampling: consistency of substitution estimators, *Ann. Statist.* **24**, 2297–2318.
32. Putter, H. and Young, G.A. (2001), On the effect of covariance function estimation on the accuracy of Kriging predictors, *Bernoulli* **7**, 421–438.
33. Roussas, G.G. (1972), *Contiguity of Probability Measures: Some Applications in Statistics*, Cambridge Tracts in Mathematics and Mathematical Physics, **63**, Cambridge University Press, London-New York.
34. Ruymgaart, F.H. (1973), *Asymptotic theory of rank tests for independence*, Mathematical Centre Tracts **43**, Mathematisch Centrum, Amsterdam.
35. Ruymgaart, F.H. (1974), Asymptotic normality of nonparametric tests for independence, *Ann. Statist.* **2**, 892–910.
36. Ruymgaart, F.H., Shorack, G.R., and Van Zwet, W.R. (1972), Asymptotic normality of nonparametric tests for independence, *Ann. Math. Statist.* **43**, 1122–1135.
37. Schmidt, R. and Stadtmüller, U. (2006), Non-parametric estimation of tail dependence, *Scand. J. Statist.* **33**, 307–335.
38. Strasser, H. (1985), *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*, de Gruyter Studies in Mathematics **7**, de Gruyter, Berlin.
39. Van der Vaart, A.W. (1988), *Statistical Estimation in Large Parameter Spaces*, CWI Tracts 44, CWI, Amsterdam.

40. Van Zwet, E.W. and Van Zwet, W.R. (1999), A remark on consistent estimation, *Johann Pfanzagl—on the occasion of his 70th birthday, Math. Methods Statist.*, **8**, 277–284.
41. Wald, A. and Wolfowitz, J. (1944), Statistical tests based on permutations of the observations, *Ann. Math. Statist.* **15**, 358–372.