FROM A TO Z: ASYMPTOTIC EXPANSIONS BY VAN ZWET

W. ALBERS

University of Twente

Refinements of first order asymptotic results are reviewed, with a number of Ph.D. projects supervised by van Zwet serving as stepping stones. Berry-Esseen bounds and Edgeworth expansions are discussed for R-, L- and U-statistics. After these special classes, the question about a general second order theory for asymptotically normal statistics is addressed. As a final topic, empirical Edgeworth expansions are considered.

AMS subject classifications: 62E20, 62G20, 60F05.
Keywords and phrases: Second order asymptotics, Berry-Esseen, (empirical) Edgeworth expansion, R-, L-, and U-statistics, symmetric statistics, bootstrap.

1 Introduction

In this paper an attempt is made to sketch van Zwet's contributions to the area of asymptotic expansions. Such a task is not particularly simple, as it concerns an expanding area in more than one sense, which also covers an impressively long period: from the early Seventies till now. (Hence the attempt to capture this comprehensive aspect in a literal manner in the title!) As a consequence, the resulting picture could easily become so loaded with details that the reader will have difficulty to focus, and the remaining impression will be blurred.

To avoid this from happening, we shall impose severe restrictions. In the first place, technical details will be dealt with rather loosely, and references will be given only sparingly. Both are amply available in the papers which we do refer to. Moreover, striving for completeness as far as references are concerned, would simply exhaust the available space and thus replace the intended sketch. A more essential restriction, however, is the fact that we shall not try to cover the whole area, but instead will select a single path through it. Our selection criterion, which seems suitable for an occasion like this, will be van Zwet's joint work on asymptotic expansions with quite a few of his students, during and following their Ph.D. projects under his guidance. Other contributions he made will typically only be included if these provided essential tools in these Ph.D. projects, or answered questions arising from such work.

As will be clear from the above, almost no attention at all will be devoted to areas of and approaches to asymptotic expansions other than those used by van Zwet, and hence the efforts of many important contributors to the field as a whole will remain unmentioned. Moreover, those working on similar topics as van Zwet, or even together with him, may still go unnoticed. Finally, those who do get mentioned, may feel that they are represented only bleakly. So let us hasten to apologize to whom it may concern, once more asking understanding for the consequences of just hacking a rather single-minded path, linking van Zwet's contributions from the point where it more or less began, till today.

The organization of the paper is as follows. In section 2 we briefly consider the classical case and the corresponding standard techniques. The next two sections are devoted to rank tests. In section 3 the one- and two-sample cases figure, which are linked to the Ph.D. thesis of Albers (1974). Section 4 is devoted to the simple linear rank test, concerned with Does' thesis from 1982. In section 5 we move from R-statistics to L-statistics. Such linear combinations of order statistics were studied in the Ph.D. thesis of Helmers (1978). Note that we do not adhere strictly to chronological order: from time to time we backtrack a little, to pick up developments which have been unfolding simultaneously. This is also the case for U-statistics, which we consider in section 6. No Ph.D. project was directly involved here, but, as was joked among his students, it was really unavoidable that Willem would do something about U-statistics: his university, the "Rijksuniversiteit Leiden", is commonly denoted by its abbreviation as "the RUL". Hence U-statistics form the missing link in his roots between R- and L-statistics! (Incidentally, since 1998 it is simply "Universiteit Leiden", so this time the ranks seem to have gone missing.)

Several questions arose from the research till this point. In section 7 we briefly consider the one about "why first order efficiency implies second order efficiency", while section 8 is devoted to the question how things can be generalized, leading to the results for symmetric statistics. This material is used in section 9 for empirical Edgeworth expansions and the bootstrap, which are the topics of the Ph.D. thesis of Putter (1994).

2 The classical case

For several decades now there has been a profound interest in refinements of first order asymptotic results, such as asymptotic normality of test statistics and estimators. A definite impetus in this respect was provided by the special invited paper on Edgeworth expansions in nonparametric statistics by Peter Bickel (1974). He lists the following four reasons for interest in higher order terms:

1) better numerical approximations than with simple normal approxima-

tions,
2) qualitative insight into the regions of unreliability of first order results,
3) discrimination between first order equivalent procedures, for example in terms of Hodges-Lehmann deficiencies,
4) challenging probabilistic problems.

The starting point for both first and second order results has been the classical case of sums of independent identically distributed (i.i.d.) random variables (r.v.'s). Let $X_1, \ldots, X_N$ be i.i.d. r.v.'s with positive and finite variance σ^2 and let F_N denote the distribution function (d.f.) of $S_N = N^{-1/2} \sum_{j=1}^N (X_j - EX_j)/\sigma$. Then by virtue of the central limit theorem $\sup_x |F_N(x) - \Phi(x)| = o(1)$, where Φ is the standard normal d.f. An improvement of this first order result is provided by the Berry-Esseen (BE) bound, which allows replacement of the mere "$= o(1)$" by "$\leq CN^{-1/2}E|X_1|^3/\sigma^3$", for some positive constant C, assuming of course that $E|X_1|^3 < \infty$. Further progress beyond this rate of convergence result requires replacement of Φ by an Edgeworth expansion (EE). A typical result runs like

$$\sup_N |F_N(x) - \tilde{F}_N(x)| = o(N^{-1}), \tag{1}$$

where $\tilde{F}_N(x)$ equals

$$\Phi(x) - \phi(x)\left[\frac{\kappa_3}{6N^{\frac{1}{2}}}(x^2 - 1) + \frac{\kappa_4}{24N}(x^3 - 3x) + \frac{\kappa_3^2}{72N}(x^5 - 10x^3 + 15x)\right], \tag{2}$$

in which κ_3 and κ_4 are the 3^{rd} and 4^{th} cumulant of X_1, respectively, and $\phi = \Phi'$. The choice in (2) is a two-step EE; omitting the terms of order N^{-1} produces the one-step EE, which gives $o(N^{-1/2})$ rather than $o(N^{-1})$ in (1).

These first two improvements beyond the BE bound are of primary interest, for example in second order comparisons. Hodges and Lehmann (1970) focussed attention on this area with admirable clarity in a paper with the concise title "Deficiency". Usually two competing statistical procedures A and B are compared as follows: if B requires $k = k_n$ observations to match the performance of A based on n observations, the ARE $e = lim_{n\to\infty} n/k_n$ of B with respect to A is studied. As Hodges and Lehmann point out, a more natural quantity than this ratio would be the simple difference $d_n = k_n - n$. Especially, whenever $e = 1$, i.e. the procedures are first order equivalent, study of this deficiency d_n is rewarding. For example, obtaining $d = \lim_{n\to\infty} d_n$ (if it exists) allows a perfectly simple comparison: one procedure eventually just requires d more observations than the other one. However, to obtain this type of knowledge, the distributions involved have to be known up to $o(N^{-1})$, rather than merely up to $o(1)$, i.e. a result like the one given by (1) and (2) is required. In their paper, Hodges and

Lehmann demonstrated the use of the deficiency concept through some very convincing and elegant examples. Even more important perhaps, was the fact that at the end of the paper they posed a number of open problems. This stimulated many studies in the Seventies. One of their questions concerned the deficiency of rank tests with respect to their parametric competitors, which inspired the research covered in sections 3 and 4.

To ensure that (1) actually holds, obviously a moment condition like $E|X_1|^r < \infty$ for some $r \geq 4$, is needed. But we also have to assume something like Cramér's condition (C):

$$\limsup_{|t|\to\infty} |\rho^*(t)| < 1, \tag{3}$$

where ρ^* is the characteristic function (ch.f.) of X_1. We shall now take a look at the arguments used in the proof of (1). This will explain why (3) is used, but more importantly, it will be helpful in the sections that follow. According to Esseen's smoothing lemma, the difference between F_N and $\tilde{F}_N$ can be measured by comparing their Fourier transforms ρ_N and $\tilde{\rho}_N$, respectively. In fact, for all $T > 0$, we have that $\sup_x |F_N(x) - \tilde{F}_N(x)|$ is of order

$$\int_{-T}^{T} \left| \frac{\rho_N(t) - \tilde{\rho}_N(t)}{t} \right| dt + \frac{1}{T}. \tag{4}$$

As $\rho_N(t)$, the ch.f. of S_N, satisfies $|\rho_N(t)| = |\rho^*(N^{-1/2}t/\sigma)|^N$, an expansion for ρ_N holds for $|t| < \delta N^{1/2}$, for some $\delta > 0$. Now $\tilde{F}_N$ in (2) has been chosen such that $\tilde{\rho}_N$ precisely equals the expansion for ρ_N truncated after the fourth term, which suffices to make the integral in (4) sufficiently small for $T = \delta N^{1/2}$. But to get $o(N^{-1})$ in (1), we need that $T^{-1} = o(N^{-1})$. On the remaining set I it no longer helps to look at $\rho_N - \tilde{\rho}_N$ and we simply need to show that

$$\int_I \left| \frac{\rho_N(t)}{t} \right| dt = o(N^{-1}). \tag{5}$$

(The accompanying result $\int_I |\tilde{\rho}_N(t)/|t| dt = o(N^{-1})$ is trivial.) If X_1 has a lattice distribution, ρ^* is periodic and $|\rho_N(t)| = |\rho^*(N^{-1/2}t/\sigma)|^N$ will keep returning to 1 and (5) may not hold true. To see that things in fact do go wrong, just consider the binomial case, where (1) clearly is false. Hence the strong non-lattice condition (3), to stay out of this kind of trouble.

3 One- and two-sample rank tests

The basic question is how to extend results for the i.i.d. case to more complicated situations. As concerns first order results, a lot of effort was devoted in the Fifties and Sixties to obtaining asymptotic normality for classes of rank statistics. As we saw above, in the early Seventies, similar questions arose for second order problems. For the easiest case, the one-sample linear rank statistic, this led to the Ph.D. thesis of Albers (1974) and to Albers, Bickel and van Zwet (1976). The idea is that here it is not necessary to expand the statistic: a direct approach will work, using an appropriate conditioning argument.

Let $X_1, \ldots, X_N$ be i.d.d. r.v.'s with common d.f. G. Consider the order statistics $0 < Z_1 < Z_2 < \ldots < Z_N$ of $|X_1|, \ldots, |X_N|$ and the anti-ranks $D_1, \ldots, D_N$ defined through $|X_{D_j}| = Z_j$. Let $V_j = 1$ if $X_{D_j} > 0$, and 0 otherwise, $j = 1, \ldots, N$, then the hypothesis that the distribution determined by G is symmetric about zero is tested on the basis of

$$T_N = \sum_{j=1}^{N} a_j V_j, \tag{6}$$

where the scores a_j are typically generated by some continuous function J on $(0,1)$, e.g. through $a_j = J(j/(N+1))$ (approximate scores). For J equal to 1, t or $\Phi^{-1}([1+t]/2)$, we obtain the sign, the Wilcoxon signed rank or the one-sample van der Waerden test, respectively.

The problem is that the summands in (6) are independent under the hypothesis only. The key step is to note that, conditional on $Z = (Z_1, \ldots, Z_N)$, the V_j are independent under the alternative as well. Hence the classical theory applies after all and an EE like (2) can be given for the conditional d.f. of T_N. A serious obstacle, however, is that the V_j are obviously lattice r.v.'s and (3) will not hold. Fortunately, we are generally saved by the fact that in this respect the i.i.d. case is least favourable. If $|\rho_N(t)| = |\rho^*(N^{-1/2}t/\sigma)|^N$, the only way to keep $|\rho_N|$ away from 1, is to do so for $|\rho^*|$ through (3). But in the case of varying components, for (5) it amply suffices if for each t there is a positive fraction among the ch.f.'s of the summands which are not close to 1 in modulus. This in its turn is easily achieved by letting the a_j vary, i.e. by letting J be non-constant. (On the other hand, a constant J produces the binomially distributed sign statistic, for which the situation is indeed hopeless).

Hereafter it remains to obtain an unconditional expansion for the d.f. of T_N by taking the expectation with respect to Z of the conditional EE. Although attention is restricted to the hypothesis and contiguous location alternatives, there are still a lot of technicalities involved and the resulting paper needs almost 50 Annals pages. The resulting expansions, however, are completely explicit and enable quick and illuminating comparisons to first order equivalent tests, such as parametric counterparts. As an example we mention that the aforementioned Hodges-Lehmann deficiency d_N (the additional number of observations required to match the power) of the normal scores test with respect to the t-test satisfies

$$d_N \sim \tfrac{1}{2} \log \log N. \tag{7}$$

Hence the bad news is that its limit is infinite; the good news is that for all practical purposes a single additional observation suffices. Several extensions of the basic result for the one-sample case were realized; we merely mention adaptive rank tests (Albers (1979)) and two-stage rank tests (Albers (1991)).

Next we turn to the two-sample problem. We modify the situation described at the beginning of this section as follows: $X_1, \ldots, X_N$ are still independent, but now $X_1, \ldots, X_m$ have common d.f. F and $X_{m+1}, \ldots, X_N$ have common d.f. G. The Z_j in this case are the order statistics if $X_1, \ldots, X_N$, the anti-ranks are defined through $X_{D_j} = Z_j$ and $V_j = 1$ if $m+1 \leq D_j \leq N$ and $V_j = 0$ otherwise, $j = 1, \ldots, N$. Then T_N from (6) stands for the general linear rank statistic for testing the hypothesis that $F = G$. An asymptotic expansion to order N^{-1} for the d.f. of this T_N under the hypothesis and contiguous alternatives, was obtained by Bickel and van Zwet (1978). This paper is the natural counterpart of the one-sample paper by Albers, Bickel and van Zwet (1976), but there is also a major difference.

In the one-sample problem we are always dealing with symmetric distributions and therefore the terms of order $N^{-1/2}$ in the expansions vanish. Hence, when comparing first order equivalent tests, deficiencies of order (almost) 1 (cf. (7)) will typically arise. For the two-sample case there is no reason to expect symmetry, and terms of order $N^{-1/2}$ do occur. Consequently, one would expect to find deficiencies of order $N^{1/2}$, but this is not what happens. In fact, the results for the one- and two-sample case are typically qualitatively the same. This quite surprising result is due to the fact that invariably for first order efficient tests all terms of order $N^{-1/2}$ agree, and hence drop out in the deficiency computations. The phenomenon of first order efficiency implying second order efficiency, noted earlier by Pfanzagl (see e.g. Pfanzagl (1979)), was sufficiently intriguing to be studied in its own right and we shall come back to it in section 7.

Although the techniques employed are similar to those of Albers, Bickel and van Zwet (1976), the occurrence of the $N^{-1/2}$-terms makes the two-sample case essentially more complicated to handle. An additional complication is that the distance to the independent case is larger here. For, after conditioning on Z in the one-sample problem, T_N is distributed as $\sum_{j=1}^{N} a_j W_j$, where the W_j are independent Bernoulli r.v.'s. In the two-sample case, however this step produces a T_N which is distributed as $\sum_{j=1}^{N} a_j W_j$, given that $\sum_{j=1}^{N} W_j = N - m$. Hence an additional trick, essentially due to Erdös and Renyi (1959), is required to obtain again an explicit representation for the conditional ch.f. of T_N. The foregoing hopefully demonstrates that it would be a major understatement to call the two-sample case a straightforward generalization of the one-sample problem. In fact, it took almost 70 pages in the Annals to do so!

4 The simple linear rank test

Let $X_1, \ldots, X_N$ be independent r.v.'s with d.f.'s $F_1, \ldots, F_N$, respectively and denote the rank of X_j among $(X_1, \ldots, X_N)$ by $R_j, j = 1, \ldots, N$. In addition to the scores a_j, we have a second vector $(c_1, \ldots, c_N)$, the regression constants. This leads to the general simple linear rank statistic

$$T_N = \sum_{j=1}^{N} c_j a_{R_j}, \tag{8}$$

which can be used to test the hypothesis of randomness $F_1 = \ldots = F_N$. The two-sample case from the previous section is contained in (8) as a special case for the choice $c_j = 0, j = 1, \ldots, m$, $c_j = 1, j = m+1, \ldots, N$.

For this general statistic, a direct approach no longer seems feasible and we resort to the more or less traditional road of attack, which consists of decomposing or expanding the statistic itself. The basic scheme suggests to write

$$T_N = S_N + R_N, \tag{9}$$

where S_N is a sum of independent r.v.'s, thus allowing application of the classical approach from section 2, while the remainder $R_N = T_N - S_N$ is supposed to be negligible in comparison to S_N. For example, under the hypothesis we can compare T_N from (8) to $S_N = \sum_{j=1}^{N} c_j J(F(X_j))$, where J is the score function which generates the a_j and F is the common d.f. of the X_j under H_0. This approach has been used extensively to obtain asymptotic normality results for T_N, under varying sets of conditions. (Note that there is a trade-off: allowing quite general F_j means strong conditions on the c_j and a_j, whereas e.g. under contiguous alternatives the conditions on regression constants and scores can be much milder.)

Typically, the first steps on the road towards second order results are taken by just pressing the above argument a bit harder: S_N, being an i.i.d. sum, also allows a classical BE bound, while generally not merely $R_N = o_P(|T_N|)$, but in fact $R_N = O_P(N^{-1/2}|T_N|)$ will hold. Let G_N be the d.f. of the standardized version $T_N^* = (T_N - ET_N)/\sigma(T_N)$, then we simply use, for some sequence $\epsilon_N > 0$,

$$G_N(x) \le P\left(\frac{S_N - ET_N}{\sigma(T_N)} \le x + \epsilon_N\right) + P\left(\frac{|R_N|}{\sigma(T_N)} > \epsilon_N\right) \tag{10}$$

together with a similar inequality in the opposite direction. The last probability in (10) can be bounded by $E|R_N|^r/(\epsilon_N \sigma(T_N))^r$ for some large r. As $R_N = O_P(N^{-1/2}|T_N|)$, this will typically be of order $N^{-r/2}\epsilon_N^{-r}$. The first probability on the right-hand side of (10) will equal $\Phi(x+\epsilon_N)+O(N^{-1/2})$ by virtue of the classical BE bound. Hence $G_N(x)-\Phi(x)$ will be of order

$$\epsilon_N + N^{-1/2} + \epsilon_N^{-r} N^{-r/2}. \tag{11}$$

This sketch shows that this type of argument is not only simple (apart from the technicalities involved!), but unfortunately also simply not good enough: no matter how we choose the ϵ_N in (11), we will never get the "right" rate $N^{-1/2}$. Something like $N^{-1/2+\delta}$ or maybe $N^{-1/2}\log N$ will be the best (11) gets us. Of course, one could object that for practical purposes it really does not matter that much whether the error behaves like $N^{-1/2}$ or like $N^{-1/2}\log N$. The point is, however, that a method which already falls short of providing the right answer in the first improvement step, will be quite useless to get any further, i.e. to obtain asymptotic expansions.

To get rid of the final δ, we need to replace a crude inequality like (10) by a more delicate analysis using the smoothing lemma from section 2: just use (4) for $\tilde{\rho}_N(t)=\exp(-t^2/2)$ and $T\sim N^{1/2}$. To begin with, replace R_N in (9) by $Q_N+\tilde{R}_N$, with e.g.

$$Q_N = N^{-1}\sum_{j=1}^{N} c_j J'(F(X_j))\,\{R_j - E(R_j|X_j)\} \tag{12}$$

and $\tilde{R}_N$ as the new remainder $T_N-(S_N+Q_N)$. For the standardized version, write $T_N^* = \bar{S}_N + \bar{Q}_N + \bar{R}_N$ and use that its ch.f. satisfies

$$\rho_N(t) = Ee^{it\bar{S}_N} + itEe^{it\bar{S}_N}\bar{Q}_N + O\left(\frac{t^2}{2}E\bar{Q}_N^2 + |t|E|\bar{R}_N|\right). \tag{13}$$

The first term in (13) equals $\exp(-t^2/2)+O(N^{-1/2}\exp(-t^2/4))$ because of the classical theory. As $\bar{S}_N$ and $\bar{Q}_N$ are sums of independent r.v.'s and as such are almost independent, the expectation in the second term can be shown to be $O(N^{-1/2}e^{-t^2/4})$ as well. Finally, since both $E\bar{Q}_N^2$ and $E|\bar{R}_N|$ are $O(N^{-1})$, it follows that the integral in (4) can be made $O(N^{-1/2})$ for $T\sim N^{1/4}$, rather than for $T=\delta^* N^{1/2}$ for some $\delta^*>0$, which is what we really need. To bridge the gap, we expand ρ_N in this interval much farther, producing a remainder term $|t|^m E|\bar{Q}_N|^m/m!$. As $E|\bar{Q}_N|^m$ can be shown to behave like $(cm)^m N^{-m/2}$ for some constant c, it follows that this remainder term leads to a contribution of order $(\delta^* ce)^m$ in (4). Hence for δ^* sufficiently small and e.g. $m=\log N$, this can be made negligible and the desired result follows.

This ingenious argument, which we have discussed in a little bit more detail to convey the flavor of the techniques used, was mentioned by Bickel (1974) in connection with U-statistics, applied by his Ph.D. student Bjerve to L-statistics and used by Hušková on simple linear rank statistics. Nevertheless, it may have brought us at the right BE rate of order $N^{-1/2}$, but there it stops again: the trick with $m = \log N$ works only up to $T \sim N^{1/2}$. As we already mentioned in section 2, what is really needed to obtain asymptotic expansions, is a way to deal with the integral in (5).

This is precisely what we find in van Zwet (1982): he essentially shows that there exist positive β and b such that

$$(14) \qquad |\rho_N(t)| = O(N^{-\beta \log N}) \text{ for } \log N \leq |t| \leq bN^{-3/2}.$$

Clearly (14) amply suffices to show (5). The techniques used to derive this smoothness property depend on the particular structure of T_N. They are related to the arguments used in Albers, Bickel and van Zwet (1976) and Bickel and van Zwet (1978), according to which some variation in the summands already suffices to keep their lattice character from destroying the smoothness.

Using van Zwet (1982) as a starting point, Does could make remarkable progress in his 1982 Ph.D. thesis. To begin with, he obtained a BE bound under weaker conditions, allowing unbounded score functions, such as the important special case Φ^{-1}, which is optimal for normal underlying distributions (see Does (1982a,b)). But he also obtained the desired expansions to $o(N^{-1})$, both under the hypothesis (Does (1983)) and under contiguous alternatives (Does (1984)). In view of (14), the emphasis in this work lies on studying the integral from (4) over the region $|t| \leq \log N$. This is a highly technical matter, using more sophisticated versions of (12) and (13). A large part of the effort required is due to the desire to not merely prove the result, but to do so under mild conditions which allow direct verification in applications.

5 Linear combinations of order statistics

Since we started with R-statistics in section 3 for the one- and two-sample case, it made sense to continue this development in the next section for the case of the simple linear rank statistic. As a consequence, our changing from R-statistics to L-statistics in the present section means going back in time a little: the developments for linear combinations of order statistics were parallel to or even preceded those for T_N from (8). Incidentally, the fact that this is not some matter of chance, is discussed by van Zwet (1983). He observes that there exists a striking similarity between the techniques employed in both areas, and uses the image of two armies marching on parallel roads. In fact, they are basically going the same place, in the sense that he manages to show under very general conditions that asymptotic normality of a two-sample rank statistic under a fixed alternative follows from a similar result for an appropriate L-statistic. Another occasion where the two areas meet, was encountered in Albers, Bickel and van Zwet (1976). Here it was observed that an asymptotic expansion for the d.f. of the one sample rank statistic under fixed alternatives would require such an expansion for the d.f. of an L-statistic. For this reason, attention was restricted to contiguous alternatives.

Let $X_1, \ldots, X_N$ be i.i.d. r.v.'s from some d.f. F, then we replace (8) by

$$T_N = N^{-1} \sum_{i=1}^{N} c_{iN} X_{i:N}, \tag{15}$$

where $X_{i:N}$ is the i^{th} order statistic of $X_1, \ldots, X_N$ and the c_{iN} are weights. Just as was the case with R-statistics, asymptotic results for L-statistics are available under varying sets of conditions. Typically, either the weights are smooth, i.e. $c_{iN} = J(i/(N+1))$ for some smooth function on $(0,1)$, or F is supposed to be smooth. In the latter case, however, attention has to be restricted to trimmed L-statistics, i.e. with $c_{iN} = 0$ for $i < N\alpha$ or $i > N\beta$, for certain $0 < \alpha < \beta < 1$.

Again we begin with the BE case. As we already mentioned in the previous section, Bjerve obtained such a result for L-statistics, using an argument due to Bickel. He considered the trimmed case, while Helmers (1977) applied the same type of approach to smooth weights. The result of Bickel for U-statistics was further improved by Callaert and Janssen (1978). Using this latter paper, Helmers (1981) improved his previous result by proving it under weaker conditions.

Next we move to asymptotic expansions. Here the pattern is again the same as in the previous section: a special argument is required to deal with (5), and here as well this is provided by van Zwet. To be more precise, by van Zwet (1977) it is shown that the ch.f. ρ_N of the standardized $T_N^* = (T_N - ET_N)/\sigma(T_N)$ satisfies, for every positive integer r and for $t \neq 0$,

$$|\rho_N(t)| = O(|t|^{-r} + e^{-\gamma N}), \tag{16}$$

where $\gamma > 0$ depends on r. Using (16), Helmers (1980) obtained an EE to $o(N^{-1})$ for L-statistics with smooth weights; the companion result from the trimmed case is contained in Helmers (1979) (the special case of the trimmed mean was already covered by Bjerve). Just as in the case of R-statistics, it is a highly laborious and technical matter to achieve all this under reasonably mild conditions. Collected together, all this material can be found in Helmers' Ph.D. thesis (1978). An additional remark is that van Zwet (1979) demonstrated that for the special case of uniform underlying distributions stronger results can be obtained than in general.

As was discussed in sections 2 and 3, the interest in second order analysis of R-statistics was stimulated by the desire to obtain deficiencies. Similarly, one can wonder about deficiencies of first order efficient tests based on L-statistics. For results of this nature we refer to Bening (1995).

6 U-statistics

After R- and L-statistics, we shall now consider U-statistics. Let again $X_1, \ldots, X_N$ be i.i.d. r.v.'s, but this time introduce for symmetric h (i.e. $h(x,y) = h(y,x)$) the U-statistic

$$U_N = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} h(X_i, X_j), \tag{17}$$

where we assume $Eh(X_1, X_2) = 0$ and $Eh^2(X_1, X_2) < \infty$. Defining $g(x) = E(h(X_1, X_2)|X_1 = x)$ and $\psi(x,y) = h(x,y) - g(x) - g(y)$, we can write

$$U_N = \hat{U}_N + \Delta_N, \tag{18}$$

with $\hat{U}_N = (N-1)\sum_{i=1}^N g(X_i)$ and $\Delta_N = \sum_{i=1}^{N-1}\sum_{j=i+1}^N \psi(X_i, X_j)$. Provided that $Eg^2(X_1) > 0$, we have that $U_N/\sigma(U_N)$ is asymptotically standard normal.

As was mentioned in section 4, the first BE bound for U-statistics was already obtained by Bickel (1974). Moreover, in the previous section we discussed how this result was used by Helmers to obtain a BE bound for L-statistics, and that he subsequently sharpened his result by using an improved version of the BE bound for U-statistics due to Callaert and Janssen. The final step in this apparent interplay between rate of convergence results for L- and U-statistics was due to Helmers and van Zwet (1982), who obtained the BE-bound for $U_N/\sigma(U_N)$ under the natural condition that $E|g(X_1)|^3 < \infty$.

The situation for asymptotic expansions to $o(N^{-1})$ is as follows: the first result on EE's for U-statistics was obtained by Callaert, Janssen and Veraverbeke (1980) (also see Janssen's Ph.D. thesis from 1978). However, they had to impose a complicated smoothness condition on the distribution of h, which was difficult to verify, and also clearly more strict than necessary. But, just as in section 3, it turns out that problems caused by a possible lattice character, become less, rather than more pronounced as the situation gets more complicated. In the former case, the i.i.d. sum was least favorable and some variation in the summands already sufficed to obtain the required smoothness. Here we observe that in going from single to double sums, like those in (17), the magnitude of the jumps in the d.f. for the lattice case typically goes down from $N^{-1/2}$ to $N^{-3/2}$ (cf. the "bad" sign statistic to the "good" Wilcoxon or signed rank statistic, which falls under (17)).

Consequently, Bickel, Götze and van Zwet (1986) succeeded in establishing the EE to $o(N^{-1})$ under very mild conditions that are easy to verify and do not involve smoothness of the d.f. of $h(X_1, X_2)$, but only of the d.f. of $g(X_1)$. In fact, conditions on g are given such that $\hat{U}_N$ from (18) admits an EE, supplemented with a moment condition on $\psi(X_1, X_2)$ to control the behavior of the remainder Δ_N in (18).

However, one awkward condition remains. Let $\omega_1, \omega_2, \dots$ be some orthonormal sequence of eigenfunctions of the kernel ψ with respect to the d.f. F of the X_i, and let $\lambda_1, \lambda_2, \dots$ be the corresponding eigenvalues, i.e.

$$\int \psi(x, y)\omega_j(x)dF(x) = \lambda_j \omega_j(y). \tag{19}$$

Then it is assumed that a sufficient number of these λ_j are nonzero. The meaning of this condition only becomes clear during the proof. Again, the source of trouble is the behavior of the ch.f. $\rho_N(t)$ for large $|t|$, making it hard to prove (5). In the present case the problem is that for these large $|t|$ this behavior is no longer governed by $\hat{U}_N$, but instead by the remainder Δ_N. To avoid degeneration in the subsequent analysis, a certain number of eigenvalues should be nonzero.

7 Efficiency of first and second order

After completion of sections 3-6, we have reached the level where BE bounds and EE's to $o(N^{-1})$ are available for R-, U- and L-statistics. Before climbing on to the next level, we briefly pause to contemplate the phenomenon of first order efficiency implying second order efficiency, which we encountered in section 3 in connection with two-sample rank tests. In the mean time, several other groups, such as Pfanzagl and his students, had also made significant contributions to higher order theory. Here we merely mention that Pfanzagl (1979) demonstrated that this phenomenon happens in general when first order efficient tests are compared. The powers of such tests typically agree to second, rather than merely to first order. Now it is one thing to observe this state of affairs, but because of the technicalities involved, it is quite something else to understand why it does happen. Fortunately, Bickel, Chibisov and van Zwet (1981) provide a nice intuitive explanation of the phenomenon.

The idea (very roughly!) is as follows. For $N = 1, 2, \dots$, let X_N be the outcome of an experiment and suppose that this X_N has density either $p_{N,0}$ or $p_{N,1}$. (Usually N simply stands for the number of independent r.v.'s in the N^{th} testing problem.) The test function of the most powerful level-α_N test in this case is

$$\phi_N(\Lambda_N) = \begin{cases} 1 & \text{if } \Lambda_N > c_N, \\ 0 & \text{otherwise,} \end{cases} \tag{20}$$

where Λ_N is the log likelihood ratio $\log\{p_{N,1}(X_N)/p_{N,0}(X_N)\}$. Typically, we are interested in the contiguous case, where $\alpha_N = E_{N,0}\phi_N(\Lambda_N)$ remains bounded away from zero, while the power $\pi_N^* = E_{N,1}\phi_N(\Lambda_N)$ remains bounded away from one. Under these circumstances, Λ_N is generally asymptotically normal and moreover usually admits an EE for π_N^* like

$$\pi_N^* = c_0 + c_1 N^{-1/2} + o(N^{-1/2}). \tag{21}$$

Let Z_N be a competing first order efficient test statistic, with level α_N, test function $\psi_N(Z_N) = 1$ for $Z_N > d_N$ and $\psi_N(Z_N) = 0$ otherwise, and power π_N admitting an EE $\pi_N = c_0 + c_1' N^{-1/2} + o(N^{-1/2})$. Note that we use the same c_0 here as in (21) by virtue of the first order efficiency. However, calculation for explicit examples invariably shows that also $c_1' = c_1$, implying that Z_N is in fact second order efficient. To understand why $\pi_N^* - \pi_N = o(N^{-1/2})$, rather than of the exact order $N^{-1/2}$, we observe that this power difference equals

$$E_{N,1}\{\phi_N(\Lambda_N) - \psi_N(Z_N)\} = E_{N,0}\Big\{(e^{\Lambda_N} - e^{d_N})(\phi_N(\Lambda_N) - \psi_N(Z_N))\Big\}. \tag{22}$$

Note that the contribution involving e^{d_N} in (22) can be smuggled in because both tests have level α_N and thus $E_{N,0}\phi(\Lambda_N) = E_{N,0}\psi_N(Z_N)$. Since Z_N is first order efficient, we can write $Z_N = \Lambda_N + \Delta_N$, with $\Delta_N = o_P(|\Lambda_N|)$ (cf.(9)). The factor $(\phi_N(\Lambda_N) - \psi_N(Z_N))$ in (22) will be non-zero only on the set where Λ_N is between c_N and $d_N - \Delta_N$. In view of the first order equivalence of the tests, c_N and d_N are close and therefore Λ_N is with large probability close to d_N on this set. Consequently, when the second factor in (22) is non-zero - which happens with small probability - the first factor will typically be small. This provides the acceleration from precise order $N^{-1/2}$ to $o(N^{-1/2})$.

As a final remark in this section we mention that Bickel, Götze and van Zwet (1983) have extended the approach above to the study of third-order efficiency of maximum likelihood-type estimates.

8 Symmetric statistics

Nowadays, many scientists are thrilled by studies of the expanding universe. Some, however, seem to have reversed preferences and rather pursue universal expansions! As van Zwet (1984) pointed out, the multitude of results obtained till then (and described in the previous sections) may have been extremely useful for statistical applications, but from a probabilistic perspective it still looks rather ad hoc, without much hope for a general theory. Consequently, he started the development of a general second order theory for asymptotically normal statistics. As the statistics involved are functions of i.i.d. r.v.'s $X_1, \ldots, X_N$, it can be assumed without loss of generality that the functions involved are symmetric. But this restriction to symmetric statistics is the only limitation imposed. Nevertheless, even this limitation can be avoided, but for arbitrary functions the conditions involved will be much more complicated and difficult to verify.

Consider $T = t(X_1, \ldots, X_N)$, where the function t is symmetric in its N arguments. As we have seen, a common approach towards second order results involved Taylor expansion of T (cf. e.g. (9) and (12)). But the smoothness of t, which is needed for this method, does not seem to be essential. The proper approach for the general case is Hoeffding's decomposition, which expands T in a series of U-statistics of increasing order. Assume that $ET^2 < \infty$ and write

$$T - ET = \sum_{i=1}^{N} T_i + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} T_{ij} + R, \tag{23}$$

where $T_i = E(T|X_i) - ET$ and $T_{ij} = E(T|X_i, X_j) - E(T|X_i) - E(T|X_j) + ET$.

To illuminate the idea behind (23), let $\hat{T}_m$ be the L_2-projection of T on the linear space spanned by functions of at most m r.v.'s from $X_1, \ldots, X_N$, then

$$\hat{T}_0 = ET, \quad \hat{T}_1 - \hat{T}_0 = \sum_{i=1}^{N} T_i,$$
$$\hat{T}_2 - \hat{T}_1 = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} T_{ij}, \quad R = \sum_{j=3}^{N} (\hat{T}_j - \hat{T}_{j-1}).$$

(The alternative term ANOVA-type decomposition is sometimes used, in view of these repeated orthogonal projections.) Using (23) and properties of L_2-projections, van Zwet (1984) obtained the BE bound for T, assuming that $E|E(T|X_1)|^3 = O(N^{-3/2})$, together with a simple moment condition to control the behavior of $T - \hat{T}_1$. If this result is applied to special cases like U- and L-statistics, it reproduces the optimal results for these situations (e.g. $E|g(X_1)|^3 < \infty$ and $Eh^2(X_1, X_2) < \infty$ for U-statistics, cf. section 6).

For the present general case, the step from the BE bound to an appropriate EE, is essentially more complicated than in the special cases studied before. In view of the similarity between (18) and (23), at first sight one would expect that the approach of Bickel, Götze and van Zwet from section 6 for U-statistics, would lead in a rather straightforward manner to an EE to $o(N^{-1})$ here as well. Unfortunately, the behavior of the "sole" difference R between (23) and (18) turns out to be extremely complex. In a sense, this is not completely surprising: the term preceding R in (23) corresponds to Δ_N from (18), and already Δ_N required a peculiar eigenvalue condition (cf. (19)) to ensure its proper behavior. Hence, for terms of still higher order, things probably get even worse.

The situation at present is as follows: an EE to $o(N^{-1})$ does exist (see Götze and van Zwet (1991)), but is as yet not in a form fit for publication.

Bentkus, Götze and van Zwet (1997) present an EE to $O(N^{-1})$, which thus not includes the terms of order N^{-1}, but does attain the right order, and not something like $O(N^{-1+\delta})$ (cf. the discussion following (11)). Incidentally, they also show that without the eigenvalue condition, the need of which was in some doubt, the EE to $o(N^{-1})$ for U-statistics is not necessarily valid. The result obtained looks quite natural: take the one-step EE (cf. (2)) and use for κ_3 simply the third cumulant of

$$\sum_{i=1}^{N} T_i + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} T_{ij},$$

i.e. neglect R in (23). This leads to an error $O(N^{-1})$ under appropriate moment conditions: a fourth for T_1, a third for T_{12} and a relatively simple one to control the behavior of R. In addition, as expected, a Cramér-type condition on T_1 is needed. Just as in the BE case, the general result obtained here turns out to be comparable to the best available results for special cases. The proof is long and tedious, among others since the traditional smoothing lemma (cf. (4)) does not seem to work anymore; its role is now played by a nonstandard smoothing inequality, on which a technique called data-dependent smoothing is based.

9 Empirical Edgeworth expansions

In this final section we consider the results obtained by Putter in his '94 Ph.D. thesis. He studies substitution estimators (formerly known as plug-in estimators!), with the bootstrap as the most prominent example. Besides results on consistency of such substitution estimators (see Putter and van Zwet (1996)), he also pays ample attention to so-called empirical Edgeworth expansions (EEE's), which provide the link to the present review.

In analogy to our observation in section 5 about R- and L-statistics, the existence of such a link is no coincidence: the closer one looks, the better one sees the relation between bootstrap and EEE. To begin with, practitioners often hope that the bootstrap automatically works, and thus effectively replaces the need for statistical thinking by routine application of simulation, but (un?)fortunately, this is not the case. Van Zwet in particular has shown that typically the bootstrap requires asymptotically linear and asymptotically normal statistics. Moreover, finer properties such as second order correctness, which have made the bootstrap even more popular, typically require the validity of an Edgeworth expansion. Hence it seems that the bootstrap and appropriate expansion techniques work under similar circumstances.

In addition, the use of expansions helps to understand the behavior of the bootstrap. Consider for example the second order correctness property, which means that the error of the bootstrap approximation can actually be of a smaller order of magnitude than the error in the customary normal approximation. Specifically, let $X_1, \ldots, X_N$ be i.i.d. r.v.'s from a d.f. F and let $T_N = t_N(X_1, \ldots, X_N)$ be a symmetric statistic (cf. section 8). Suppose the d.f. G_N of the standardized version

$$T_N^* = (T_N - ET_N)/\sigma(T_N) \tag{24}$$

admits an EE. Then the bootstrap approximation G_N^* for G_N relies on replacing F by some empirical version, like the empirical d.f. $\hat{F}_N$. Consequently, the coefficients in the EE for G_N^* are just the empirical counterparts of the corresponding coefficients in the EE for G_N.

But now a similar argument applies as in section 7: these coefficients are of order $N^{-1/2}$ (or even N^{-1}) to begin with, and estimation errors are $o_P(1)$ (typically even $O_P(N^{-1/2})$), which in combination leads to an approximation error $o_P(N^{-1/2})$ (or even $O_P(N^{-1})$), rather than merely $O_P(N^{-1/2})$, for this EEE, and thus for the bootstrap. Incidentally, do note that we have considered the standardized version T_N^*. For T_N itself, $\sigma(T_N)$ will occur already in the leading term of the EE, leading to an estimation error of at least order $N^{-1/2}$. As $\sigma(T_N)$ is typically unknown in practice, the statistic of real interest is neither T_N^* from (24) nor T_N, but a Studentized version

$$\tilde{T}_N = (T_N - ET_N)/S_N, \tag{25}$$

where S_N^2 is some appropriate estimator of $\sigma^2(T_N)$.

The above immediately prompts the following question: instead of merely using the EEE to explain the bootstrap, can't we use it to replace the bootstrap altogether? In this way, a lot of simulation effort can be avoided. This attractive idea is studied extensively by Putter. Generally speaking, it turns out that both bootstrap and EEE indeed outperform the ordinary normal approximation. In the mutual comparison, the bootstrap seems to be slightly better than the EEE, which agrees with intuition as the bootstrap also estimates higher order coefficients, whereas the EEE stops after one (or two) steps.

Up to now, we have mainly outlined the motivation and the general ideas. At the end of this section, we shall briefly also consider some specific aspects, such as methods applied, types of estimators used, etc. But, as usual, we largely refer to the relevant papers, which in this case are Putter (1994) and Putter and van Zwet (1998). Consider symmetric statistics T_N with $ET_N = 0$, then one-step EE's with error $o(N^{-1/2})$ are established for $T_N/\sigma(T_N)$ and for $\tilde{T}_N = T_N/S_N$ (cf. (25)). For S_N^2 the well-known jackknife estimator of variance is used. Next, the coefficients in these EE's are estimated in a similar fashion, also using jackknife techniques, and it is shown that the resulting one-step EEE's have error $o_P(N^{-1/2})$.

As concerns the methods of proof, for the EE's the key tool again is Hoeffding's decomposition (cf. (18) and (23)). Extensive use is made of the results by Bickel, Götze and van Zwet (1986) on the EE for U-statistics, which were discussed in section 6. For the step from EE's to EEE's, it suffices to show the consistency of the jackknife estimators applied. It is demonstrated that the results obtained are sufficiently general to allow application to U-statistics, L-statistics, smooth functions of the sample mean, as well as smooth functionals of the empirical d.f. Moreover, it is also demonstrated how the results can be used to prove second order correctness of the bootstrap for Studentized U-statistics of degree two, a case which was studied earlier by Helmers (1991) under stronger moment conditions.

Here our sketch comes to an end. Maybe this comes across a little abruptly, leaving the reader out on a limb. But remember that this is the appropriate place to be at the end of a journey through a tree-like structure such as this review!

REFERENCES

Albers, W. (1974). *Asymptotic expansions and the deficiency concept in statistics.* Mathematical Centre Tract 58. Mathematisch Centrum, Amsterdam. (Ph.D. thesis)

Albers, W., Bickel, P.J., van Zwet, W.R. (1976). Asymptotic expansions for the power of distribution free tests in the one-sample problem. *Annals of Statistics* **4** 108–156.

Albers, W. (1979). Asymptotic deficiencies of one-sample rank tests under restricted adaptation. *Annals of Statistics* **7** 944–954.

Albers, W. (1991). Second order analysis of two-stage rank tests for the one-sample problem. *Annals of Statistics* **19** 1042–1052.

Bening, V.E. (1995). A formula for deficiency: one sample L- and R-tests I, II. *Mathematical Methods Statistics* **4** 167–188, 274–293.

Bentkus, V., Götze, F., van Zwet, W.R. (1997). An Edgeworth expansion for symmetric statistics. *Annals of Statistics* **25** 851–896.

Bickel, P.J. (1974). Edgeworth expansions in nonparametric statistics. *Annals of Statistics* **2** 1–20.

Bickel, P.J., van Zwet, W.R. (1978). Asymptotic expansions for the power of distribution free tests in the two-sample problem. *Annals of Statistics* **6** 937–1004.

Bickel, P.J., Chibisov, D.M., van Zwet, W.R. (1981). On efficiency of first and second order. *International Statistical Review* **49** 169–175.

Bickel, P.J., Götze, F., van Zwet, W.R. (1983). A simple analysis of third-order efficiency of estimates. *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II* 749–768.

Bickel, P.J., Götze, F., van Zwet, W.R. (1986). The Edgeworth expansion for U-statistics of degree two. *Annals of Statistics* **14** 1463–1484.

Callaert, H., Janssen, P. (1978). The Berry-Esseen theorem for U-statistics. *Annals of Statistics* **6** 417–421.

Callaert, H., Janssen, P., Veraverbeke, N. (1980). An Edgeworth expansion for U-statistics. *Annals of Statistics* **8** 299–312.

Does, R.J.M.M. (1982a). *Higher order asymptotics for simple linear rank statistics.* Mathematical Centre Tract 151. Mathematisch Centrum, Amsterdam (Ph.D. Thesis).

Does, R.J.M.M. (1982b). Berry-Esseen theorems for simple linear rank statistics under the null-hypothesis. *Annals of Probability* **10** 982–991.

Does, R.J.M.M. (1983). An Edgeworth expansion for simple linear rank statistics under the null-hypothesis. *Annals of Statistics* **11** 607–624.

Does, R.J.M.M. (1984). Asymptotic expansions for simple linear rank statistics under contiguous alternatives. *Asymptotic Statistics* **2** (eds. P. Mandl and M. Hušková), 221-230, North Holland.

Erdös, P., Rényi, A. (1959). On the central limit theorem for samples from a finite population. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **4** 49–61.

Götze, F., van Zwet, W.R. (1991). Edgeworth expansions for asymptotically linear statistics. Preprint 91–034, Universität Bielefeld.

Helmers, R. (1977). The order of the normal approximation for linear combinations of order statistics with smooth weight functions. *Annals of Probability* **5** 940–953.

Helmers, R. (1978). *Edgeworth expansions for linear combinations of order statistics.* Mathematical Centre Tract 105. Mathematisch Centrum, Amsterdam. (1982; Ph.D. thesis 1978)

Helmers, R. (1979). Edgeworth expansions for trimmed linear combinations of order statistics. *Asymptotic Statistics* **1** (eds. P. Mandl and M. Hušková), 221-232.

Helmers, R. (1980). Edgeworth expansions for linear combinations of order statistics with smooth weight functions. *Annals of Statistics* **8** 1361–1374.

Helmers, R. (1981). A Berry-Esseen theorem for linear combinations of order statistics. *Annals of Probability* **9** 342–347.

Helmers, R., van Zwet, W.R. (1982). The Berry-Esseen bound for U-statistics. *Statistical Decision Theory and Related Topics, III, Vol. I* 497–512.

Helmers, R. (1991). On the Edgeworth expansion and the bootstrap approximation for a Studentized U-statistic. *Annals of Statistics* **19** 470–484.

Hodges, J.L., Lehmann, E.L. (1970) Deficiency. *Annals of Mathematical Statistics* **41** 783–801.

Pfanzagl, J. (1979). First order efficiency implies second order efficiency. *Contributions to Statistics*, 167–196, Reidel, Dordrecht.

Putter, H. (1994). *Consistency of resampling methods.* (Ph.D. thesis)

Putter, H., van Zwet, W.R. (1996). Resampling: consistency of substitution estimators. *Annals of Statistics* **24** 2297–2318.

Putter, H., van Zwet, W.R. (1998). Empirical Edgeworth expansions for symmetric statistics. *Annals of Statistics* **26** 1540–1569.

van Zwet, W.R. (1977). Asymptotic expansions for the distribution functions of linear combinations of order statistics. *Statistical Decision Theory and Related Topics, II* 421–437.

van Zwet, W.R. (1979). The Edgeworth expansion for linear combinations of uniform order statistics. *Asymptotic Statistics* **1** (eds. P. Mandl and M. Hušková), 93–101, North Holland.

van Zwet, W.R. (1982). On the Edgeworth expansion for the simple linear rank statistic. *Nonparametric Statistical Inference, Vol. I, II* 889-909.

van Zwet, W.R. (1983). Ranks and order statistics. *Recent Advances in Statistics* 407–422, Academic Press, New York.

van Zwet, W.R. (1984). A Berry-Esseen bound for symmetric statistics. *Z. Wahrsch. Verw. Gebiete* **66** 425–440.

FACULTY OF MATHEMATICAL SCIENCES
P.O. BOX 217
7500 AE ENSCHEDE
NETHERLANDS
albers@math.utwente.nl