AN EDGEWORTH EXPANSION FOR SYMMETRIC STATISTICS

V. Bentkus[1] F. Götze[1] W. van Zwet[1,2]

University of Bielefeld,
University of Leiden and University of North Carolina at Chapel Hill

January 1994

Abstract. We consider asymptotically normal statistics which are symmetric functions of N i.i.d. random variables. For these statistics we prove the validity of an Edgeworth expansion with remainder $O(1/N)$ under Cramér's condition on the linear part of the statistic and moment assumptions for all parts of the statistic. By means of a counter-example we show that it is generally not possible to obtain an Edgeworth expansion with remainder $o(1/N)$ without imposing additional assumptions on the structure of the non-linear part of the statistic

1. Introduction and results

Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed (i.i.d.) random variables taking values in an arbitrary measurable space $(\mathcal{X}, \mathcal{B})$ with common distribution $\mathbf{P}$. Let the measurable function $t : \mathcal{X}^N \to \mathbf{R}$ be symmetric in its N arguments, in the sense that it is invariant under permutations of these arguments. Consider the symmetric statistic

$$T = t(X_1, \ldots, X_N).$$

[1]Research supported by the SFB 343 at Bielefeld.

[2]Research supported by the Netherlands' Organization for Scientific Research.

1991 *Mathematics Subject Classification.* Primary 62E20; secondary 60F05.

Key words and phrases. Asymptotic expansion, Edgeworth expansions, Symmetric statistics, Hoeffding's decomposition, U-statistics, functions of sample means, functionals of empirical distribution functions, linear combinations of order statistics, Student's statistic.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

We shall derive Edgeworth expansions for the distribution of T. Expansions terms and conditions are expressed in terms of 'influence functions' of the statistic. These functions measure the dependence of T on up to three observations. They are usually easy to compute or estimate in applications by means of Taylor expansions. See section 2 for examples including applications to the Bootstrap. The basis of the mathematical treatment of the problem will be Hoeffding's (1948) decomposition of the statistic in terms of conditional means of influence functions (see section 4).

Denote by

$$\mathbf{E}_i T = \mathbf{E}(T|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_N), \qquad \text{for} \quad 1 \le i \le N,$$

the expectation with respect to X_i, that is the conditional expectation given all other random variables. Define the first order influence function (of X_i) as the difference

$$D_i T = T - \mathbf{E}_i T. \tag{1.1}$$

For distinct i, j, k, define the second and the third order influence functions (with respect to observations X_i, X_j and X_k) as repeated differences

$$\begin{aligned} &D_i D_j T = T - \mathbf{E}_i T - \mathbf{E}_j T + \mathbf{E}_{ij} T, \\ &D_i D_j D_k T = T - \mathbf{E}_i T - \mathbf{E}_j - \mathbf{E}_k T + \mathbf{E}_{ij} T + \mathbf{E}_{ik} T + \mathbf{E}_{jk} - \mathbf{E}_{ijk} T, \end{aligned}$$

where we write for brevity $\mathbf{E}_{ij\dots}$ instead of $\mathbf{E}_i \mathbf{E}_j \dots$. Notice that $D_i D_i T = D_i T$ and $D_i D_j T = D_i D_j T$.

Furthermore, define the random variables

(1.2)
$$T_i = \mathbf{E}(D_i T | X_i) = \mathbf{E}(T|X_i) - \mathbf{E}T,$$

(1.3)
$$T_{ij} = \mathbf{E}(D_i D_j T | X_i, X_j) = \mathbf{E}(T|X_i, X_j) - \mathbf{E}(T|X_i) - \mathbf{E}(T|X_j) + \mathbf{E}T.$$

These expectations are the well known Hoeffding's decomposition terms of first and second orders. We may write

$$T - \mathbf{E}T = \sum_{i=1}^{N} T_i + \sum_{1 \le i < j \le N} T_{ji} + R, \tag{1.4}$$

where all random variables on the right hand side of (1.4) have mean zero and are uncorrelated. The terms

$$\sum_{i=1}^{N} T_i \qquad \text{and} \qquad \sum_{1 \le i < j \le N} T_{ji} + R$$