EMPIRICAL EDGEWORTH EXPANSIONS

FOR SYMMETRIC STATISTICS[1]

BY HEIN PUTTER AND WILLEM R. VAN ZWET

University of Leiden and University of North Carolina, Chapel Hill

In this paper the validity of a one-term Edgeworth expansion for Studentized symmetric statistics is proved. We propose jackknife estimates for the unknown constants appearing in the expansion and prove their consistency. As a result we obtain the second-order correctness of the empirical Edgeworth expansion for a very general class of statistics, including U-statistics, L-statistics and smooth functions of the sample mean. We illustrate the application of the bootstrap in the case of a U-statistic of degree two.

1. Introduction

Let $(\mathcal{X}, \mathcal{A})$ be a measurable space and P a probability measure on $(\mathcal{X}, \mathcal{A})$. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, taking values in $\mathcal{X}$ with unknown common distribution P. Let $t_N : \mathcal{X}^N \times \mathcal{P} \to \mathbb{R}$ be symmetric as a function on $\mathcal{X}^N$, i.e. for every $x_1, \ldots, x_N \in \mathcal{X}$ and every permutation $\{\alpha_1, \ldots, \alpha_N\}$ of $\{1, \ldots, N\}$, we have

$$t_N(x_1, \ldots, x_N; P) = t_N(x_{\alpha_1}, \ldots, x_{\alpha_N}; P) \ .$$

It will be assumed throughout this paper that

$$T_N = t_N(X_1, \ldots, X_N; P) \ , \tag{1.1}$$

is a random variable with expectation

$$ET_N = 0 \qquad \text{for all } N \tag{1.2}$$

[1] The research for this paper was supported by the Netherlands Organization for Scientific Research (NWO) and the Sonderforschungsbereich 343 "Diskrete Strukturen in der Mathematik" at the University of Bielefeld, Germany.

AMS 1991 subject classifications. Primary 62 E 20, 62 G 09.

Key words and phrases. Edgeworth expansion, Hoeffding's decomposition, jackknife, bootstrap.

and variance $\sigma^2(T_N) = \sigma_N^2$, satisfying

$$0 < c \leq \sigma_N^2 \leq C < \infty \qquad \text{for all } N \tag{1.3}$$

for finite positive constants c and C. Suppose that T_N/σ_N converges in distribution to a standard normal distribution. Typically the accuracy of the normal approximation is of the order $N^{-\frac{1}{2}}$ as N tends to infinity. In this paper we shall focus on second-order approximations, i.e. on approximations with error $o(N^{-\frac{1}{2}})$ to the distribution functions of T_N/σ_N and also of the Studentized version T_N/S_N, where S_N^2 is an estimator of σ_N^2. We shall accomplish this by first proving the validity of a (one-term) Edgeworth expansion with remainder $o(N^{-\frac{1}{2}})$ and then estimating the unknown constants in the expansion. The procedure is therefore called an *empirical Edgeworth expansion.*

As an estimate S_N^2 of σ_N^2 we shall use the jackknife estimator of variance, introduced by Quenouille (1949, 1956) and Tukey (1958). Let us suppose that we have one additional observation X_{N+1} at our disposal, and define for $i = 1, \ldots, N$,

$$T_N^{(i)} = t_N(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{N+1}; P) \ , \qquad T_N^{(N+1)} = T_N \ , \tag{1.4}$$

and

$$\overline{T}_N = \frac{1}{N+1} \sum_{i=1}^{N+1} T_N^{(i)} \ . \tag{1.5}$$

The jackknife estimator of variance S_N^2 is then defined as

$$S_N^2 = \sum_{i=1}^{N+1} \left(\overline{T}_N - T_N^{(i)} \right)^2 \ . \tag{1.6}$$

At first sight it may seem somewhat awkward to try to estimate the variance of T_N with $N+1$ observations from P, since for practical purposes we would then wish to approximate the distribution of T_{N+1} instead of T_N, and the variance of T_{N+1} would be the object of interest. With σ_N^2 satisfying (1.3), however, σ_N^2 will typically converge to a constant and the difference between σ_N^2 and σ_{N+1}^2 will vanish. In practically all cases of interest it will not influence the one-term Edgeworth expansions either. We shall therefore not be concerned with this difference.

We shall make extensive use of the properties of Hoeffding's decomposition, introduced by Hoeffding (1948). For $k = 1, 2, \ldots$, let Ω_k denote the set of integers $\{1, \ldots, k\}$ and define for a set $A \subset \Omega_N$,

$$E(T_N|A) = E(\, T_N \mid X_i, i \in A\,) \tag{1.7}$$

Next, for $D \subset \Omega_N$, define

$$T_{N,D} = \sum_{A \subset D} (-1)^{|D|-|A|} \ E(T_N|A) \ . \tag{1.8}$$