stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SW 54/77 NOVEMBER
(DEPARTMENT OF MATHEMATICAL STATISTICS)

R. HELMERS

A BERRY-ESSEEN THEOREM FOR LINEAR COMBINATIONS OF ORDER STATISTICS

Preprint

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS(MOS) subject classification scheme (1970): Primary 62G30, Secondary 62E20

A Berry-Esseen theorem for linear combinations of order statistics *)

Abbreviated title:
Linear combinations of order statistics

by

R. Helmers

ABSTRACT

A Berry-Esseen bound of order $n^{-\frac{1}{2}}$ is established for linear combinations of order statistics with smooth weight functions. The underlying distribution function must possess a finite absolute third moment. This improves an earlier result of the author.

KEY WORDS & PHRASES: *linear combinations of order statistics, Berry-Esseen bounds.*

---

*) This report will be submitted for publication elsewhere.

1. INTRODUCTION

Linear combinations of order statistics received much attention during the last ten years. Much is known about them including their asymptotic normality under quite general conditions. Berry-Esseen type bounds for the normal approximation of linear combinations of order statistics were established by BJERVE (1977) and HELMERS (1977) (see VAN ZWET (1977) for a review of these results). For the related problem of establishing Edgeworth expansions for linear combinations of order statistics we refer to the papers of HELMERS (1976) and VAN ZWET (1977).

BJERVE (1977) obtained the order bound $\mathcal{O}(n^{-\frac{1}{2}})$ (n being the sample size) for trimmed linear combinations of order statistics. His result admits quite general weights between the αth and βth sample percentile ($0<\alpha<\beta<1$) but he does not allow weights to be put on the remaining observations. In HELMERS (1977) the order bound $\mathcal{O}(n^{-\frac{1}{2}})$ was established for linear combinations of order statistics with weights of the form $c_{in} = J(\frac{i}{n+1})$, $i = 1,2,\ldots,n$, for a smooth function J on (0,1). The underlying distribution function F must possess a finite absolute third moment. Though we avoid the assumption that there are no weights in the tails the use of a technique of BICKEL (1974) in the second part of the proof given in HELMERS (1977) leads to the assumption $\int_0^1 |J'(s)|\, dF^{-1}(s) < \infty$ (J' being the derivative of J). In this note we shall show that this assumption is superfluous. The present theorem will be proved by modifying the proof given in HELMERS (1977); we shall argue as in CHAN & WIERMAN (1977) and CALLAERT & JANSSEN (1977), rather than using Bickel's idea (BICKEL (1974)).

2. THE THEOREM

Let, for each $n \geq 1$, $T_n = n^{-1} \sum_{i=1}^n J(\frac{i}{n+1}) X_{in}$ where X_{in}, $i = 1,2,\ldots,n$ denotes the i^{th} order statistic of a random sample $X_1,\ldots,X_n$ of size n from a distribution with distribution function (d.f.) F and J is a bounded measurable function on (0,1). The inverse of a d.f. will always be the left-continuous one. Let $F_n^*(x) = P(T_n^* \leq x)$ for $-\infty < x < \infty$, where

$$T_n^* = (T_n - \mathcal{E}(T_n))/\sigma(T_n).$$

Let Φ denote the standard normal distribution function. We prove the following theorem.

THEOREM 1. *Suppose*

(1) J *is bounded and continuous on* (0,1). *The derivative* J' *exists, except possibly at a finite number of points;* J' *satisfies a Lipschitz condition of order* $> \frac{1}{2}$ *on the open intervals where it exist. The inverse* F^{-1} *satisfies a Lipschitz condition of order* $> \frac{1}{2}$ *on neighborhoods of the points where* J' *does not exist.*

(2) $E|X_1|^3 < \infty$.

Then $\sigma^2(J,F) > 0$, *where*

$$\sigma^2(J,F) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(F(x)J(F(y))(F(\min(x,y)) - F(x)F(y))dxdy$$

implies that there exists a constant C, *depending on* J *and* F *but not on* n, *such that for all* $n \geq 1$

$$\sup_x |F_n^*(x) - \Phi(x)| \leq Cn^{-\frac{1}{2}}.$$

As indicated in our introduction the present theorem was obtained in HELMERS (1977) under the additional requirement $\int_0^1 |J'(s)|dF^{-1}(s) < \infty$.

3. THE PROOF

Let, for each $n \geq 1$, $U_1, U_2, \ldots, U_n$ be independent uniform (0,1) random variables (r.v.'s). For any r.v. X with $0 < \sigma(X) < \infty$ we denote by X^* the r.v. $(X-E(X))/\sigma(X)$. Let χ_E denote the indicator of a set E. Define, for each $n \geq 1$, the r.v. S_n by

$$S_n = I_{1n} + I_{2n} \tag{3.1}$$

where

$$I_{1n} = -n^{-1} \sum_{i=1}^n \int_0^1 J(s)(\chi_{(0,s]}(U_i)-s)dF^{-1}(s) \tag{3.2}$$

and

$$(3.3) \qquad I_{2n} = -n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{i-1} \int_0^1 J'(s)(\chi_{(0,s]}(U_i)-s)(\chi_{(0,s]}(U_j)-s)dF^{-1}(s).$$

In section 2 of HELMERS (1977) it is proved that under the assumption (1) of the present theorem and the assumption that $E|X_1|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$ $T_n^* - S_n^*$ is of negligible order for our purposes; i.e.

$$(3.4) \qquad P(|T_n^*-S_n^*| \geq n^{-\frac{1}{2}}) = O(n^{-\frac{1}{2}}), \quad \text{as} \quad n \to \infty.$$

Hence it suffices to prove that the assumptions of the present theorem ensure that the order of the normal approximation for S_n^* is $n^{-\frac{1}{2}}$.

The r.v. S_n^* is given by $S_n^* = J_{1n} + J_{2n}$, where $J_{mn} = I_{mn}/\sigma(S_n)$ for $m = 1,2$ and all $n \geq 1$. For convenience we shall write

$$(3.5) \qquad g(U_i) = -\int_0^1 J(s)(\chi_{(0,s]}(U_i)-s)dF^{-1}(s) \quad \text{for} \quad 1 \leq i \leq n$$

and

$$(3.6) \qquad h(U_i,U_j) = -\int_0^1 J'(s)(\chi_{(0,s]}(U_i)-s)(\chi_{(0,s]}(U_j)-s)dF^{-1}(s) \quad \text{for} \quad 1 \leq i < j \leq n$$

and also $\sigma_n = \sigma(S_n)$.

Using the idea of CHAN & WIERMAN (1977) we introduce r.v.'s J'_{2n} and J''_{2n} as follows

$$(3.7) \qquad J'_{2n} = (n^2\sigma_n)^{-1} \sum_{i=1}^{I} \sum_{j=1}^{i-1} h(U_i,U_j)$$

$$(3.8) \qquad J''_{2n} = (n^2\sigma_n)^{-1} \sum_{i=I+1}^{n} \sum_{j=1}^{i-1} h(U_i,U_j)$$

where $I = [n-3\, n^{\frac{3}{2}}\, \sigma_n^2\, \sigma^{-2} \log n \wedge n]$. Note that $J_{2n} = J'_{2n} + J''_{2n}$. Since our proof will depend on characteristic functions (c.f) arguments let us denote by ρ_{1n} and ρ'_n the c.f. of J_{1n} and $J_{1n} + J'_{2n}$. The c.f. of a summand of $n\sigma_n J_{1n}$, that is of $g(U_1)$, will be denoted by ρ. Clearly we have $\rho_{1n}(t) = \rho^n(\frac{t}{n\sigma_n})$ for all t and $n \geq 1$.

Following now the pattern of the Chan-Wierman proof we shall first show that there exist $\varepsilon_1 > 0$ such that

(3.9) $$\int_{|t|\leq\varepsilon_1 n^{\frac{1}{2}}} |\rho_{1n}(t) - e^{-\frac{t^2}{2}}||t|^{-1}dt = O(n^{-\frac{1}{2}}), \text{ as } n \to \infty.$$

Next we show that there exist $\varepsilon_2 > 0$ such that

(3.10) $$\int_{|t|\leq\varepsilon_2 n^{\frac{1}{2}}} |\rho_n'(t) - \rho_{1n}(t)||t|^{-1}dt = O(n^{-\frac{1}{2}}), \text{ as } n \to \infty.$$

Finally we prove that J''_{2n} is of negligible order for our purposes; i.e.

(3.11) $$P(|J''_{2n}| \geq n^{-\frac{1}{2}}) = O(n^{-\frac{1}{2}}), \text{ as } n \to \infty.$$

The Berry-Esseen bound of order $n^{-\frac{1}{2}}$ for S_n^* then follows from (3.9),(3.10), (3.11) and the usual argument based on Esseen's smoothing lemma (see e.g. FELLER (1966)).

LEMMA 3.1. *Suppose that the conditions* (1) *and* (2) *of theorem* 1 *are satisfied. Then* $\sigma^2(J,F) > 0$ *implies* (3.9).

PROOF. See lemma 3.1 of HELMERS (1977). The proof is essentially that of the classical Berry-Esseen theorem for a properly standarized sum of independent identically distributed random variables with finite absolute third moment. The main difficulty is that we have standarized I_{1n} by multiplying with σ_n^{-1} rather than with $n^{\frac{1}{2}}\sigma^{-1}(J,F)$. □

Next we shall be concerned with the problem of showing that (3.10) holds under appropriate conditions.

LEMMA 3.2. *Let* $E|X_1|^{2+\varepsilon} < \infty$ *for some* $\varepsilon > 0$ *and suppose that condition* (1) *of theorem* 1 *is satisfied. Then* $\sigma^2(J,F) > 0$ *implies* (3.10).

PROOF. The proof is essentially the Chan & Wierman proof. See CHAN & WIERMAN (1977), p.137. First note that $0 < \sigma^2(g(U_1)) = \sigma^2(J,F) < \infty$ because $\sigma^2(J,F) > 0$, J is bounded and $EX_1^2 < \infty$. Hence there exists $\varepsilon_2 > 0$ such that $|\rho(t)| \le 1 - \frac{t^2\sigma^2(J,F)}{3} \le \exp(-\frac{t^2\sigma^2(J,F)}{3})$ for $|t| \le \varepsilon_2/\sigma(J,F)$. Next we remark that

$$E(J_{2n}')^2 = O(n^{-1}) \quad \text{as } n \to \infty. \tag{3.12}$$

To prove this we note first that it follows directly from the proof of lemma 3.1 (see HELMERS (1977)) that $0 < \lim_{n\to\infty} n\sigma_n^2 = \sigma^2(J,F) < \infty$ and $EI_{2n}^2 = O(n^{-2})$ as $n \to \infty$ holds under the assumptions of the present lemma. This implies that $E(J_{2n})^2 = O(n^{-1})$ as $n \to \infty$ from which (3.12) follows immediately.

To proceed we remark first that

$$\int_{|t|\le\varepsilon_2 n^{\frac12}} |\rho_n'(t) - \rho_{1n}(t)||t|^{-1}dt =$$

$$= \int_{|t|\le n^{\frac14}\wedge\varepsilon_2 n^{\frac12}} |\rho_n'(t) - \rho_{1n}(t)||t|^{-1}dt +$$

$$+ \int_{n^{\frac14}\wedge\varepsilon_2 n^{\frac12}<|t|\le\varepsilon_2 n^{\frac12}} |\rho_n'(t) - \rho_{1n}(t)||t|^{-1}dt.$$

Next we note that

$$\int_{|t|\le n^{\frac14}\wedge\varepsilon_2 n^{\frac12}} |\rho_n'(t) - \rho_{1n}(t)||t|^{-1}dt =$$

$$= \int_{|t|\le n^{\frac14}\wedge\varepsilon_2 n^{\frac12}} |Ee^{itJ_{1n}}(e^{itJ_{2n}'}-1)||t|^{-1}dt \le$$

$$\le \int_{|t|\le n^{\frac14}\wedge\varepsilon_2 n^{\frac12}} (|Ee^{itJ_{1n}} J'_{2n}| + |t| E(J'_{2n})^2)dt \le$$

$$\le (n^2\sigma_n)^{-1} \sum_{i=1}^{I} \sum_{j=1}^{i-1} \int_{|t|\le n^{\frac14}\wedge\varepsilon_2 n^{\frac12}} |Ee^{itJ_{1n}} h(U_i,U_j)|dt +$$

$$+ n^{\frac12} E(J'_{2n})^2 \le$$

$$\le \sigma_n^{-1} \int_{|t|\le n^{\frac14}\wedge\varepsilon_2 n^{\frac12}} |\rho^{n-2}(\tfrac{t}{n\sigma_n})|.|Ee^{\frac{it}{n\sigma_n}(g(U_1)+g(U_2))} h(U_1,U_2)|dt +$$

$$+ O(n^{-\frac12}) \le$$

$$\le n^{-2}\sigma_n^{-3} \int_{|t|\le n^{\frac14}\wedge\varepsilon_2 n^{\frac12}} t^2 \exp(-(n-2)\frac{t^2\sigma^2(J,F)}{3n^2\sigma_n^2}).$$

$$. E|(g(U_1) + g(U_2))^2 h(U_1,U_2)|dt$$

$$+ O(n^{-\frac12}) =$$

$$= O(n^{-\frac12}), \quad \text{as} \quad n \to \infty.$$

In the fourth inequality we have applied (3.12), whereas in the fifth inequality we use that $g(U_1) + g(U_2)$ and $h(U_1,U_2)$ are uncorrelated. Note also that $E|(g(U_1)+g(U_2))^2h(U_1,U_2)| < \infty$ under the assumptions of the lemma.

Finally we remark that

$$\int_{n^{\frac14}\wedge\varepsilon_2 n^{\frac12}<|t|\le\varepsilon_2 n^{\frac12}} |\rho'_n(t) - \rho_{1n}(t)||t|^{-1}dt \le$$

$$\le 2 \int_{n^{\frac14}\wedge\varepsilon_2 n^{\frac12}<|t|\le\varepsilon_2 n^{\frac12}} |\rho^{n-I}(\tfrac{t}{n\sigma_n})||t|^{-1} dt \le$$

$$\le 2 \int_{n^{\frac14}\wedge\varepsilon_2 n^{\frac12}<|t|\le\varepsilon_2 n^{\frac12}} |t|^{-1} \exp(-\frac{t^2(n-I)\sigma^2}{3n^2\sigma_n^2})dt.$$

But for n sufficiently large this integral is equal to

$$2\int_{n^{\frac{1}{4}}\wedge\varepsilon_2 n^{\frac{1}{2}}<|t|\le\varepsilon_2 n^{\frac{1}{2}}} |t|^{-1}n^{-1}dt = O(n^{-1}\log n),\text{ as } n\to\infty.$$

Combining all the results we find that (3.10) is proved. □

Finally we prove (3.11). The basic idea of the proof is the same as that of a lemma of CALLAERT & JANSSEN (1977).

LEMMA 3.3. *Suppose that the conditions* (1) *and* (2) *of theorem* 1 *are satisfied. Then* $\sigma^2(J,F) > 0$ *implies* (3.11).

PROOF. For $i = 1,2,\ldots$ define $\xi_i = \sum_{j=1}^{i-1} h(U_i,U_j)$. We have $\xi_1 = 0$ and $E(\xi_{i+1}|\xi_1,\ldots,\xi_i) = 0$ a.s. for $i = 1,2,\ldots$. Hence

$$\sum_{i=I+1}^{I+k} \xi_i,\quad k = 1,2,\ldots$$

is a martingale. Note also that, for fixed i, $\sum_{j=1}^{m} h(U_i,U_j)$, $m = 1,2,\ldots,i-1$ is also a martingale. Hence the martingale inequality of DHARMADHIKARI, FABIAN and JOGDEO (1968) can be applied twice. We obtain, after some elementary computations and making again use of the fact that $0 < \lim_{n\to\infty} n\sigma_n^2 < \infty$ (see the proof of lemma 3.2), that there exists a number $B > 0$ such that for all $n \ge 1$

$$E|J''_{2n}|^3 \le Bn^{-9/4}(\log n)^{\frac{3}{2}}\, E|h(U_1,U_2)|^3.$$

To proceed we shall show that the assumptions of this lemma imply that $E|h(U_1,U_2)|^3 < \infty$. We start with an application of the c_r-inequality:

$$E|h(U_1,U_2)|^3 \le 3^2[E(\int_0^{U_1\wedge U_2} |J'(s)|s^2dF^{-1}(s))^3 +$$

$$+ E(\int_{U_1\wedge U_2}^{U_1\vee U_2} |J'(s)|s(1-s)dF^{-1}(s))^3 + E(\int_{U_1\vee U_2}^{1} |J'(s)|(1-s)^2dF^{-1}(s))^3].$$

Since J' is bounded on its domain, F^{-1} is continuous at the points where J' is not defined, and $E|X_1| < \infty$, we can easily check that

$$E\Big(\int_{U_1\wedge U_2}^{U_1\vee U_2} |J'(s)|s(1-s)dF^{-1}(s)\Big)^3 \le \Big(\int_0^1 |J'(s)|s(1-s)dF^{-1}(s)\Big)^3 < \infty.$$

Also note that

$$E\Big(\int_0^{U_1\wedge U_2} |J'(s)|s^2dF^{-1}(s)\Big)^3 =$$

$$= 2\int_0^1 (1-t)\Big(\int_0^t |J'(s)|s^2dF^{-1}(s)\Big)^3dt.$$

It is not difficult to verify that the integral on the right is finite under the assumptions of this lemma. Similarly we can show that also

$$E\Big(\int_{U_1\vee U_2}^1 |J'(s)|(1-s)^2dF^{-1}(s)\Big)^2 < \infty.$$

Combining all these results we find that $E|h(U_1,U_2)|^3 < \infty$, so that we can conclude that

$$E|J''_{2n}|^3 = O(n^{-9/4}(\log n)^{\frac{3}{2}}), \quad \text{as} \quad n \to \infty.$$

An application of Markov's inequality yields now

$$P(|J''_{2n}| \le n^{-\frac{1}{2}}) \le n^{\frac{3}{2}} E|J''_{2n}|^3 = O(n^{-\frac{3}{4}}(\log n)^{\frac{3}{2}}),$$

as $n \to \infty$, which completes our proof of the lemma. □

ACKNOWLEDGMENT

The author wishes to thank W.R. Van Zwet for pointing out the relevance of the references [4] and [5] for our problem.

REFERENCES

[1] BICKEL, P.J. (1974), *Edgeworth expansions in non-parametric statistics*, Ann. Statist., 2, 1-20.

[2] BJERVE, S. (1977), *Error bounds for linear combinations of order statistics*, Ann. Statist. 5, 2, 357-369.

[3] CALLAERT, H. and P. JANSSEN (1977), *The Berry-Esseen theorem for U-statistics*, to appear in the Annals of Statistics.

[4] CHAN, Y. and J. WIERMAN (1977), *On the Berry-Esseen theorem for U-statistics*, Ann. Prob. 5, 136-139.

[5] DHARMADHIKARI, S.W., V. FABIAN and K. JOGDEO (1968), *Bounds on the moments of martingales*. Ann. Math. Statist. 39, 1719-1723.

[6] FELLER, W. (1966), *An Introduction to Probability Theory and Its Applications*, Vol. 2, Wiley, New York.

[7] HELMERS, R. (1977), *The order of the normal approximation for linear combinations of order statistics with smooth weight functions*, Ann. Prob. 5, no. 6.

[8] HELMERS, R. (1976), *Edgeworth expansions for linear combinations of order statistics with smooth weight functions*, submitted to the Annals of statistics.

[9] ZWET, W.R. VAN, (1977), *Asymptotic expansions for the distribution functions of linear combinations of order statistics*, *Statistical Decision Theory and Related Topics* II, 421-438, Academic Press, New York.