The Annals of Probability
1981, Vol. 9, No. 2, 342–347

A BERRY-ESSEEN THEOREM FOR LINEAR COMBINATIONS OF ORDER STATISTICS[1]

By R. Helmers

Mathematisch Centrum, Amsterdam

A Berry-Esseen bound of order $n^{-1/2}$ 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.

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 [2] and the author [5]. Bjerve obtained the order bound $O(n^{-1/2})$ (n being the sample size) for trimmed linear combinations of order statistics. In [5] the order bound $O(n^{-1/2})$ was established for linear combinations of order statistics with weights of the form $c_{in} = J(i/(n+1))$, $i = 1, 2, \ldots, n$ for a smooth function J on (0, 1). The underlying distribution function F must possess a finite absolute third moment. Though the assumption that there are no weights in the tails is avoided, the use of a technique of Bickel [1] in the second part of the proof given in [5] 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 and, moreover, that the smoothness conditions needed in [5] can be relaxed. The result of this paper, as well as similar results employing a different, more practical standardization, and for a Studentized version of a linear combination of order statistics are summarized in [6]. Boos and Serfling [3] recently obtained the Berry-Esseen theorem for statistical functions. As an application they obtain a Berry-Esseen theorem for a class of linear combinations of order statistics.

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

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

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

Theorem. *Suppose*
(1) *the function J satisfies a Lipschitz condition of order* 1 *on* (0, 1);
(2) $E|X_1|^3 < \infty$.
Then $\sigma^2(J, F) > 0$ *where*

Received September 1977; revised April 30, 1979.

[1] This paper is a completely revised version of Mathematical Centre report SW54 (1977). It was written while the author was on leave at the Université de Montréal.

AMS 1970 *subject classifications.* Primary 62G30; secondary 62E20.

Key words and phrases. Linear combinations of order statistics, Berry-Esseen bounds.

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

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

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

3. Proof. Let, for each $n \geq 1$, $U_1, \ldots, U_n$ be independent uniform (0, 1) random variables (rv's). For any rv X with $0 < \sigma(X) < \infty$ we denote by X^* the rv $X^* = (X - E(X))/\sigma(X)$. Let χ_E denote the indicator of a set E. In the first lemma we approximate T_n by a rv V_n given by

$$V_n = \int_0^1 J(s)F_n^{-1}(s)\, ds = \sum_{i=1}^n \int_{(i-1)/n}^{i/n} J(s)\, ds\, X_{in} \tag{3.1}$$

where F_n denotes the empirical df based on $X_1, \ldots, X_n$.

LEMMA 3.1. *Let $EX_1^2 < \infty$ and suppose that assumption* (1) *is satisfied. Then $\sigma^2(J, F) > 0$ implies that as $n \to \infty$*

$$\sigma^2(T_n^* - V_n^*) = O(n^{-2}). \tag{3.2}$$

PROOF. The present lemma will be proved by modifying the proof of Lemma 2.2 of [5]. First note that we can follow the argument given on page 943 of [5] to check that $\lim_{n\to\infty} n\sigma^2(T_n) = \sigma^2(J, F) > 0$ and to find that it suffices then to prove that

$$\sigma^2(T_n - V_n) = O(n^{-3}) \text{ as } n \to \infty. \tag{3.3}$$

To see that (3.3) is true we simply apply the inequalities (2.8) and (2.10) of [5] and use the Lipschitz condition for J. This completes the proof of the lemma. □

Define for $0 < u < 1$ the function

$$\psi(u) = \int_u^1 J(s)\, ds - (1 - u) \int_0^1 J(s)\, ds \tag{3.4}$$

and let $c = \int_0^1 J(s)\, ds$. Then (cf. (2.18) of [5])

$$V_n = \int_0^1 \psi(\Gamma_n(s))\, dF^{-1}(s) + cn^{-1} \sum_{i=1}^n F^{-1}(U_i) \tag{3.5}$$

with probability 1. Here and elsewhere Γ_n will denote the empirical df based on $U_1, \ldots, U_n$. To proceed we note that, as J is Lipschitz of order 1 on (0, 1), we can approximate V_n from above and below by

$$\begin{aligned} W_{n+} = &\int_0^1 \{\psi(s) + (\Gamma_n(s) - s)\psi'(s)\}\, dF^{-1}(s) + cn^{-1} \sum_{i=1}^n F^{-1}(U_i) \\ &+ K \int_0^1 (\Gamma_n(s) - s)^2\, dF^{-1}(s) \end{aligned} \tag{3.6}$$

and

$$W_{n-} = \int_0^1 \{\psi(s) + (\Gamma_n(s) - s)\psi'(s)\}\, dF^{-1}(s) + cn^{-1} \textstyle\sum_{i=1}^n F^{-1}(U_i) \tag{3.7}$$
$$- K \int_0^1 (\Gamma_n(s) - s)^2\, dF^{-1}(s)$$

for some fixed $K > 0$ and all $n \geq 1$; i.e., for all $n \geq 1$

$$W_{n-} \leq V_n \leq W_{n+}. \tag{3.8}$$

It will be convenient to have

LEMMA 3.2. *Let $E|X_1|^{2+\epsilon} < \infty$ for some $\epsilon > 0$ and suppose that assumption* (1) *is satisfied. Then $\sigma^2(J, F) > 0$ implies that as $n \to \infty$*

$$\frac{\sigma(W_{n+})}{\sigma(V_n)} = 1 + O(n^{-1/2}), \frac{E(V_n - W_{n+})}{\sigma(V_n)} = O(n^{-1/2}) \tag{3.9}$$

and

$$\frac{\sigma(W_{n-})}{\sigma(V_n)} = 1 + O(n^{-1/2}), \frac{E(V_n - W_{n-})}{\sigma(V_n)} = O(n^{-1/2}). \tag{3.10}$$

PROOF. It is immediate from (3.5), (3.6) and assumption (1) that

$$|V_n - W_{n+}| = O\left(\int_0^1 (\Gamma_n(s) - s)^2\, dF^{-1}(s)\right)$$

as $n \to \infty$. A simple moment calculation, using the moment assumption of the lemma, yields that

$$E|V_n - W_{n+}| = O(n^{-1}),$$

and

$$\sigma^2(V_n - W_{n+}) \leq E(V_n - W_{n+})^2 = O(n^{-2})$$

as $n \to \infty$. As in the proof of Lemma 3.1 we also have that $\lim_{n\to\infty} n\sigma^2(V_n) = \sigma^2(J, F) > 0$ under the present assumptions. The Cauchy-Schwarz inequality implies that $|\sigma(W_{n+}) - \sigma(V_n)| \leq \sigma(W_{n+} - V_n)$ and (3.9) follows. The proof of (3.10) is similar. □

In the following lemma we relate W_{n+} and W_{n-} to appropriate U-statistics. Define for each $n \geq 1$

$$U_{n+} = \binom{n}{2}^{-1} \textstyle\sum_{i=1}^n \sum_{j=1}^{i-1} h_+(U_i, U_j) \tag{3.11}$$

and

$$U_{n-} = \binom{n}{2}^{-1} \textstyle\sum_{i=1}^n \sum_{j=1}^{i-1} h_-(U_i, U_j) \tag{3.12}$$

where the functions h_+ and h_- are given for $0 < u, v < 1$ by

$$h_+(u, v) = -\int_0^1 J(s)\{\chi_{(0,s]}(u) + \chi_{(0,s]}(v) - 2s\}\, dF^{-1}(s) \tag{3.13}$$
$$+ 2K \int_0^1 (\chi_{(0,s]}(u) - s)(\chi_{(0,s]}(v) - s)\, dF^{-1}(s)$$

and $h_-(u, v)$ similarly by replacing K by $-K$ in (3.13). The constant K is as in (3.6) and (3.7).

LEMMA 3.3. *Let $EX_1^2 < \infty$ and suppose that assumption* (1) *is satisfied. Then $\sigma^2(J, F) > 0$ implies that as $n \to \infty$*

$$\sigma^2(W_{n+}^* - U_{n+}^*) = O(n^{-2}) \tag{3.14}$$

and

$$\sigma^2(W_{n-}^* - U_{n-}^*) = O(n^{-2}). \tag{3.15}$$

PROOF. The present lemma will be proved by modifying part of the proof of Lemma 2.3 of [5]. We first prove (3.14). To start with we note that the argument leading to relation (2.26) of [5] can be repeated (replace $-2^{-1}J'(s)$ by K and W_n by W_{n+}) to find that

$$\begin{aligned} W_{n+} - E\,W_{n+} = &-n^{-1} \textstyle\sum_{i=1}^n \displaystyle\int_0^1 J(s)(\chi_{(0,s]}(U_i) - s)\, dF^{-1}(s) \\ &+ Kn^{-2} \textstyle\sum_{i=1}^n \sum_{j=1}^n \displaystyle\int_0^1 (\chi_{(0,s]}(U_i) - s)(\chi_{(0,s]}(U_j) - s)\, dF^{-1}(s) \\ &- Kn^{-1} \int_0^1 s(1 - s)\, dF^{-1}(s). \end{aligned} \tag{3.16}$$

Combining (3.16) with (3.11) and using the assumptions of the lemma we find after a little calculation that

$$\sigma^2\left(\tfrac{1}{2}\left(1 - \frac{1}{n}\right) U_{n+} - W_{n+}\right) = O(n^{-3}) \quad \text{as} \quad n \to \infty. \tag{3.17}$$

As it is easily verified that $\lim_{n\to\infty} n\sigma^2(W_{n+}) > 0$ under the present assumptions we have (cf. the proof of Lemma 3.1) proved (3.14). The proof of (3.15) is of course similar. □

In the fourth and final lemma of this section we establish Berry-Esseen bounds for U_{n+}^* and U_{n-}^*. This lemma is a direct consequence of a Berry-Esseen theorem for U-statistics due to Callaert and Janssen [4].

LEMMA 3.4. *Let $E|X_1|^3 < \infty$ and suppose that J is bounded on* (0, 1). *Then $\sigma^2(J, F) > 0$ implies that as $n \to \infty$*

$$\sup_x |P(U_{n+}^* \le x) - \Phi(x)| = O(n^{-1/2}) \tag{3.18}$$

and

$$\sup_x |P(U_{n-}^* \le x) - \Phi(x)| = O(n^{-1/2}). \tag{3.19}$$

PROOF. It is immediate from (3.13) that

$$E(h_+(U_1, U_2) \mid U_1) = -\int_0^1 J(s)(\chi_{(0,s]}(U_1) - s)\, dF^{-1}(s) \tag{3.20}$$

with probability 1. Also note that

$$\sigma^2\left(\int_0^1 J(s)(\chi_{(0,s]}(U_1) - s)\, dF^{-1}(s)\right) = \sigma^2(J, F) > 0 \tag{3.21}$$

so that we find that the conditional expectation (3.20) has a positive variance. Moreover it is immediate from (3.7) and (3.8) of [5] that

$$E\left|\int_0^1 J(s)(\chi_{(0,s]}(U_1) - s)\,dF^{-1}(s)\right|^3 < \infty. \tag{3.22}$$

Since

$$\left|\int_0^1 (\chi_{(0,s]}(U_1) - s)(\chi_{(0,s]}(U_2) - s)\,dF^{-1}(s)\right| \le \int_0^1 |\chi_{(0,s]}(U_1) - s|\,dF^{-1}(s)$$

we also have by a similar argument that

$$E\left|\int_0^1 (\chi_{(0,s]}(U_1) - s)(\chi_{(0,s]}(U_2) - s)\,dF^{-1}(s)\right|^3 < \infty \tag{3.23}$$

under the present assumptions. Hence it follows that $E\,|h_+(U_1, U_2)|^3 < \infty$. The conditions of the Berry-Esseen theorem for U-statistics ([4]) are therefore satisfied and (3.18) follows. The proof of (3.19) is, of course, similar. □

We are now in a position to prove our theorem. First we use Lemma 3.1 and Chebychev's inequality to find that

$$P(|T_n^* - V_n^*| \ge n^{-2/3}) \le n^{4/3}\sigma^2(T_n^* - V_n^*) = O(n^{-2/3}). \tag{3.24}$$

Using this we see that

$$\begin{aligned} F_n^*(x) &= P(T_n^* \le x) \\ &\le P(V_n^* \le x + n^{-2/3}) + P(|T_n^* - V_n^*| \ge n^{-2/3}) \\ &= P(V_n^* \le x + n^{-2/3}) + O(n^{-2/3}) \end{aligned} \tag{3.25}$$

uniformly in x. A similar argument yields the opposite inequality

$$F_n^*(x) \ge P(V_n^* \le x - n^{-2/3}) + O(n^{-2/3}) \tag{3.26}$$

uniformly in x. Secondly we remark that, because of inequality (3.8),

$$P(V_n^* \le x + n^{-2/3}) \le P\left(W_{n-}^* \frac{\sigma(W_{n-})}{\sigma(V_n)} + \frac{E(W_{n-} - V_n)}{\sigma(V_n)} \le x + n^{-2/3}\right) \tag{3.27}$$

and

$$P(V_n^* \le x - n^{-2/3}) \ge P\left(W_{n+}^* \frac{\sigma(W_{n+})}{\sigma(V_n)} + \frac{E(W_{n+} - V_n)}{\sigma(V_n)} \le x - n^{-2/3}\right). \tag{3.28}$$

This, together with Lemma 3.2 yields that

$$P(V_n^* \le x + n^{-2/3}) \le P(W_{n-}^* \le x_{n+}) \tag{3.29}$$

and

$$P(V_n^* \le x - n^{-2/3}) \ge P(W_{n+}^* \le x_{n-}) \tag{3.30}$$

for appropriate sequences x_{n+}, $n = 1, 2, \ldots$ and x_{n-}, $n = 1, 2, \ldots$ satisfying

$$x_{n\pm} = x(1 + O(n^{-1/2})) + O(n^{-1/2}) \tag{3.31}$$

uniformly in x. We can now simply repeat the argument leading to (3.25) and (3.26), using this time Lemma 3.3 and Chebychev's inequality, to see that

$$P(W^*_{n-} \le x_{n+}) \le P(U^*_{n-} \le x_{n+} + n^{-2/3}) + O(n^{-2/3}) \tag{3.32}$$

and

$$P(W^*_{n+} \le x_{n-}) \ge P(U^*_{n+} \le x_{n-} - n^{-2/3}) + O(n^{-2/3}) \tag{3.33}$$

as $n \to \infty$, uniformly in x. Combining all these inequalities we see that

$$P(T^*_n \le x) \le P(U^*_{n-} \le x_{n+} + n^{-2/3}) + O(n^{-2/3}) \tag{3.34}$$

and

$$P(T^*_n \le x) \ge P(U^*_{n+} \le x_{n-} - n^{-2/3}) + O(n^{-2/3}) \tag{3.35}$$

as $n \to \infty$, uniformly in x. Applying now Lemma 3.4 we see that the first terms on the right of (3.34) and (3.35) are equal to $\Phi(x_{n+} + n^{-2/3}) + O(n^{-1/2})$ and $\Phi(x_{n-} - n^{-2/3}) + O(n^{-1/2})$ respectively, uniformly in x. As these two expressions are easily seen to be equal to $\Phi(x) + O(n^{-1/2})$, as $n \to \infty$, uniformly in x, the proof of our theorem is complete.

Acknowledgment. The author wishes to thank W. R. van Zwet for pointing out the relevance of reference [4] for our problem.

REFERENCES

[1] BICKEL, P. J. (1974). Edgeworth expansions in nonparametric statistics, *Ann. Statist.* **2** 1–20.
[2] BJERVE, S. (1977). Error bounds for linear combinations of order statistics. *Ann. Statist.* **5** 357–369.
[3] BOOS, D. and SERFLING R. J. (1979). On Berry-Esseen rates for statistical functions, with application to L-estimates. Technical report, Dept. Statist., The Florida State Univ.
[4] CALLAERT, H. and JANSSEN P. (1978). The Berry-Esseen theorem for U-statistics. *Ann. Statist.* **6** 417–421.
[5] HELMERS, R. (1977). The order of the normal approximation for linear combinations of order statistics with smooth weight functions. *Ann. Probability* **5** 940–953.
[6] HELMERS, R. (1980). *Edgeworth expansions for linear combinations of order statistics.* Mathematical Centre Tract 105, Amsterdam. To appear.

DEPARTMENT OF STATISTICS
MATHEMATICAL CENTRE
KRUISLAAN 413
1098 SJ AMSTERDAM
THE NETHERLANDS