stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SW 76/81 SEPTEMBER
(DEPARTMENT OF MATHEMATICAL STATISTICS)

R.J.M.M. DOES & C.A.J. KLAASSEN

A BERRY-ESSEEN THEOREM FOR FUNCTIONS OF UNIFORM SPACINGS

PREPRINT

kruislaan 413 1098 SJ amsterdam

Printed at the Mathematical Centre, 413 Kruislaan, 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.).

1980 Mathematics subject classification: Primary 62E20
Secondary 60E10, 62G10

A Berry-Esseen theorem for functions of uniform spacings *)

by

Ronald J.M.M. Does**) & Chris A.J. Klaassen

ABSTRACT

A Berry-Esseen bound of order $\mathcal{O}(n^{-\frac{1}{2}})$ is established for suitably normalized sums of nonlinear functions of uniform spacings under a natural moment assumption and mild regularity conditions. Furthermore, it is shown that these regularity conditions are fulfilled for a wide class of functions.

KEY WORDS & PHRASES: *Berry-Esseen theorem, uniform spacings, conditional distribution, characteristic function*

---

*) This report will be submitted for publication elsewhere.

**) Now at the University of Limburg, Maastricht

1. INTRODUCTION AND NOTATION

Let $U_1, U_2, \ldots$ be a sequence of independent random variables which are uniformly distributed on the interval $(0,1)$. For $n = 1,2,\ldots$ the ordered $U_1, U_2, \ldots, U_n$ are denoted by $U_{1:n} \le U_{2:n} \le \ldots \le U_{n:n}$. Let $U_{0:n} = 0$ and $U_{n+1:n} = 1$. The uniform spacings D_{in} are defined by

$$D_{in} = U_{i:n} - U_{i-1:n}, \quad i = 1,2,\ldots,n+1. \tag{1.1}$$

Let $g_n\colon (0,\infty) \to \mathbb{R}$, $n = 1,2,\ldots$, be measurable functions and define the statistics T_n by

$$T_n = \sum_{i=1}^{n+1} g_n((n+1)D_{in}), \quad n = 1,2,\ldots . \tag{1.2}$$

Statistics of this form can be used for testing uniformity. A survey of the asymptotic theory on these statistics may be found in PYKE (1972). We also refer to KOZIOL (1980). According to PYKE (1972) a study of the rates of convergence of sums of functions of spacings to their limiting distributions is of interest. In the case of uniform spacings Pyke suggests to use the following well-known characterization, which has been applied by LE CAM (1958) in order to prove first order limit theorems. Let Y_i, $i = 1,2,\ldots$, be independent exponential random variables with expectation 1.

If for $n = 1,2,\ldots$

$$W_n = \sum_{i=1}^{n+1} g_n(Y_i) \tag{1.3}$$

and

$$S_n = \sum_{i=1}^{n+1} (Y_i - 1), \tag{1.4}$$

then (cf. Lemma 3.1)

$$\mathcal{L}(T_n) = \mathcal{L}(W_n \mid S_n = 0), \tag{1.5}$$

i.e. T_n has the same distribution as a sum of independent random variables given another sum of independent random variables. Applying (1.5) we obtain an expression for the characteristic function of T_n (cf. (3.36)). Using this expression we prove a Berry-Esseen theorem for T_n under quite general conditions, which in most cases are easy to verify (cf. Corollary 2.1).

In a forthcoming paper we shall derive, under some version of Cramér's condition, an Edgeworth expansion for T_n using the same techniques as in the present paper. In DOES & HELMERS (1980) such an Edgeworth expansion has been established by expanding the density of the vector $(n^{-\frac{1}{2}}W_n, n^{-\frac{1}{2}}S_n)$. That paper also contains, as a by-product, a Berry-Esseen bound of order $\mathcal{O}(n^{-\frac{1}{2}})$ under a condition which is rather hard to check.

To conclude this section we define for n = 1,2,...

$$(1.6)\qquad \rho_n(s,t) = E(e^{is(Y_1-1)+itg_n(Y_1)}), \quad s \in \mathbb{R}, \; t \in \mathbb{R},$$

$$(1.7)\qquad \mu_n = Eg_n(Y_1) = \int_0^\infty g_n(y)e^{-y}dy,$$

$$(1.8)\qquad \sigma_n^2 = \mathrm{var}\; g_n(Y_1) = \int_0^\infty (g_n(y) - \mu_n)^2\, e^{-y}dy,$$

$$(1.9)\qquad \tau_n = \mathrm{cov}(g_n(Y_1),Y_1) = \int_0^\infty (g_n(y) - \mu_n)(y-1)\, e^{-y}dy,$$

$$(1.10)\qquad 1_A(x) = \begin{cases} 1 & \text{for } x \in A, \\ 0 & \text{for } x \notin A, \end{cases} \quad A \subset \mathbb{R},$$

and

$$(1.11)\qquad [x]^+ = \begin{cases} x & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

and we note that (cf. Section 2.1 of PYKE (1965))

$$(1.12)\qquad ET_n = n \int_0^{n+1} g_n(y)\left(1 - \frac{y}{n+1}\right)^{n-1} dy$$

and

$$(1.13)\qquad ET_n^2 = n \int_0^{n+1} g_n^2(y)\left(1 - \frac{y}{n+1}\right)^{n-1} dy +$$

$$+ n^2 \frac{n-1}{n+1} \int_0^{n+1} \int_0^{n+1-y} g_n(x) g_n(y) (1 - \frac{x+y}{n+1})^{n-2} \, dxdy.$$

2. A BERRY-ESSEEN THEOREM

Our main result reads as follows:

THEOREM 2.1. *Let* F_n *be the distribution function of*

(2.1) $$T_n^* = (T_n - ET_n)(\text{var } T_n)^{-\frac{1}{2}}, \quad n = 1,2,\ldots,$$

and let the functions $K_m\colon (0,\infty)^2 \to [0,\infty]$, $m = 1,2,\ldots$, *be defined by*

(2.2) $$K_m(b,d) = \sup_{|t|\le b} \sup_{n\ge m} n \log(n+1) \int_{\log|s|\ge dn} |\rho_n(s,t)|^n \, ds.$$

If for some positive and finite constants c *and* C *and some integer* n_0

(2.3) $$\sup_{n\ge n_0} \int_0^\infty |g_n(y)|^3 e^{-y} dy \le C,$$

(2.4) $$\inf_{n\ge n_0} (\sigma_n^2 - \tau_n^2) \ge c,$$

if

(2.5) g_n *is Lebesgue almost everywhere continuous for* $n \ge n_0$

and if for all positive d *there exists a positive* b *with*

(2.6) $$K_{n_0}(b,d) < \infty,$$

then var T_n *and consequently* F_n *are properly defined for* $n \ge n_0$ *and if* var T_n *and* F_n *are well defined for* $n < n_0$ *too, then there exists a finite constant* A *only depending on* n_0, K_{n_0}, c *and* C, *such that*

(2.7) $$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \le An^{-\frac{1}{2}}, \quad n = 1,2,\ldots,$$

where Φ *is the standard normal distribution function.*

It should be noted that T_n and consequently F_n are insensitive to changes of g_n on the interval $(n+1,\infty)$, but that the conditions (2.3) through (2.6) do not have this property for reasons of simplicity. Nevertheless, condition (2.3) seems to be a natural moment condition. Assumption (2.4) protects against degeneration of the variance of T_n. Loosely speaking it means that g_n is nonlinear both for finite n and asymptotically. The relevance of (2.5) will be clear from Lemma 3.1. However, assumption (2.6) is rather unpleasant. In the following corollary we shall give conditions which imply (2.3) through (2.6), which are easier to verify and which are satisfied for most of the statistics T_n used in practice.

COROLLARY 2.1. *Let* H_n *be the distribution function of*

$$T'_n = n^{-\frac{1}{2}} \sum_{i=1}^{n+1} g((n+1)D_{in}), \quad n = 1,2,..., \tag{2.8}$$

where $g: (0,\infty) \to \mathbb{R}$ *is a (fixed) measurable function for which there exist nondecreasing functions* g_0 *and* g_1 *such that*

$$g = g_0 - g_1. \tag{2.9}$$

If there exists a positive number α *such that*

$$\lim_{y\downarrow 0} y^{\alpha} g_i(y) = \lim_{y\to\infty} e^{-\alpha y} g_i(y) = 0, \quad i = 0,1, \tag{2.10}$$

and if

$$\int_0^\infty g(y)e^{-y}\, dy = 0, \tag{2.11}$$

$$\int_0^\infty g^2(y)e^{-y}\, dy - \left[\int_0^\infty g(y)y\, e^{-y}\, dy\right]^2 = 1 \tag{2.12}$$

and

$$\int_0^\infty |g(y)|^3\, e^{-y}\, dy < \infty, \tag{2.13}$$

then there exists a constant A *only depending on* g *such that*

$$\sup_{x\in\mathbb{R}} |H_n(x) - \Phi(x)| \le An^{-\frac{1}{2}}, \quad n = 1,2,\ldots . \tag{2.14}$$

It is clear that by a suitable linear transformation every nonlinear g satisfying (2.13) can be standardized to a function satisfying (2.11) and (2.12). The examples given in Section 3 of PYKE (1965) are related to the functions $f_1(x) = x^r$, $r > 0$, $r \neq 1$, $f_2(x) = (x-1)^2$, $f_3(x) = |x-1|$, $f_4(x) = \log x$ and $f_5(x) = x^{-1}$. With the corollary f_1 through f_4 can be handled. The statistic corresponding to f_5 has a nonnormal limit distribution.

3. PROOF OF THEOREM 2.1.

Since the representation (1.5) is crucial to our setup, we examine it in the following lemma.

LEMMA 3.1. *Let* g_n, T_n, W_n *and* S_n *be as defined in Section* 1. *If* g_n *is Lebesgue almost everywhere continuous, then there exists a version of the conditional distribution* $L(W_n \mid S_n = x)$ *which is continuous in* $x = 0$ *(with respect to weak convergence). Moreover, for such a version (cf.* (1.5)*)*

$$L(T_n) = L(W_n \mid S_n = 0). \tag{3.1}$$

PROOF. A regular version of the conditional density of $(Y_1,\ldots,Y_n)$ given $S_n = x$ is given by

$$f_{Y_1,\ldots,Y_n|S_n}(y_1,\ldots,y_n|x) = n!(n+1+x)^{-n}, \quad 0 \le y_i, \quad i = 1,\ldots,n, \tag{3.2}$$
$$\sum_{i=1}^{n} y_i \le n+1+x, \quad -n-1 < x < \infty.$$

According to this version we define the conditional distribution $L(W_n \mid S_n = x)$ of W_n given $S_n = x$. From (3.2) and the Mann-Wald theorem (cf. Theorem 5.1 of BILLINGSLEY (1968)) we see that $L(W_n \mid S_n = x)$ weakly converges to $L(W_n \mid S_n = 0)$ as x tends to 0. Furthermore, we note that the density of $((n+1)D_{1n},\ldots,(n+1)D_{nn})$ equals (cf. Section 4 of PYKE (1965))

$$f_{(n+1)D_{1n},\ldots,(n+1)D_{nn}}(z_1,\ldots,z_n) = n!(n+1)^{-n}, \quad 0 \le z_i, \ i = 1,\ldots,n, \tag{3.3}$$
$$\sum_{i=1}^{n} z_i \le n+1.$$

Combining (3.2) and (3.3) we arrive at (3.1). □

In the notation of Section 9 of BHATTACHARYA & RAO (1976) we have the following preliminary lemma:

LEMMA 3.2. *Let* n_0 *be a positive integer and let* G_n, $n = n_0, n_0+1, \ldots,$ *be probability measures on* $\mathbb{R}^2$ *with zero mean, positive-definite covariance matrix* V_n *and finite third absolute moment. Let* B_n *be the symmetric positive-definite matrix satisfying* $B_n^2 = V_n^{-1}$ *and let*

$$(3.4) \qquad c_3 = \sup_{n \geq n_0} \int \| B_n x \|^3 \, dG_n(x).$$

There exist two positive constants c_1 *and* c_2 *such that for all integers* $n \geq n_0$, *for all* $t \in \mathbb{R}^2$ *satisfying*

$$(3.5) \qquad \| t \| \leq c_1 c_3^{-1} n^{\frac{1}{2}}$$

and for all nonnegative integral vectors α *with* $0 \leq |\alpha| \leq 3$ *the inequality*

$$(3.6) \qquad |D^{\alpha}[\hat{G}_n^n(n^{-\frac{1}{2}} B_n t) - \exp\{-\tfrac{1}{2}\| t \|^2\}]|$$
$$\leq c_2 c_3 n^{-\frac{1}{2}} [\| t \|^{3-|\alpha|} + \| t \|^{3+|\alpha|}] \exp\{-\tfrac{1}{4}\| t \|^2\}$$

holds.

PROOF. This is an immediate consequence of Theorem 9.10 of BHATTACHARYA & RAO (1976). □

For the proof of Theorem 2.1 we assume that conditions (2.3) through (2.6) are satisfied and we define

$$(3.7) \qquad \tilde{g}_n(y) = g_n(y) - \tau_n(y-1), \qquad y > 0.$$

By $\tilde{T}_n$, $\tilde{\rho}_n$, $\tilde{\tau}_n$, $\tilde{K}_m(b,d)$ etc. we denote the quantities defined in (1.2), (1.6), (1.9), (2.2) etc., with g_n replaced by $\tilde{g}_n$. It is easy to verify that

$$\tilde{T}_n = T_n, \quad \tilde{\mu}_n = \mu_n, \quad \tilde{\sigma}_n^2 = \sigma_n^2 - \tau_n^2, \quad \tilde{\tau}_n = 0,$$

(3.8)

$$\tilde{\rho}_n(s,t) = \rho_n(s - \tau_n t, t)$$

and that for $|b| \le C^{-\frac{1}{3}} e^{\frac{1}{2}dn_0}(e^{\frac{1}{2}dn_0} - 1)$

(3.9) $$\tilde{K}_{n_0}(b,d) \le K_{n_0}(b,\tfrac{1}{2}d).$$

Together with (2.3), (2.4) and (2.6) this implies that

$$\sup_{n \ge n_0} \int_0^\infty |\tilde{g}_n(y)|^3 e^{-y} dy \le 14C,$$

$$\inf_{n \ge n_0} \tilde{\sigma}_n^2 \ge c$$

and that for all positive d there exists a positive b with

$$\tilde{K}_{n_0}(b,d) < \infty.$$

We conclude that we may assume without loss of generality that (2.3) through (2.6) hold with

(3.10) $$\tau_n = 0.$$

Next we define

(3.11) $$\nu_n = (n+1)^{-1} ET_n,$$

(3.12) $$\omega_n = n^{-\frac{1}{2}}(\text{var } T_n)^{\frac{1}{2}}$$

and

(3.13) $$g_n^*(y) = \omega_n^{-1}(g_n(y) - \nu_n), \quad y > 0.$$

With the Cauchy-Schwarz inequality we obtain

$$(3.14)\quad \left|\int_0^{n+1} g_n(y)\left(1-\frac{y}{n+1}\right)^{n-1} dy - \int_0^{\infty} g_n(y)\left(1+\frac{2y-\frac{1}{2}y^2}{n+1}\right)e^{-y}\,dy\right|$$

$$\le \left[\int_0^{\infty} g_n^2(y)e^{-y}\,dy\right]^{\frac{1}{2}}\left[\int_0^{\infty}\left|\left(\left[1-\frac{y}{n+1}\right]^+\right)^{n-1} - \left(1+\frac{2y-\frac{1}{2}y^2}{n+1}\right)e^{-y}\right|^2 e^y\,dy\right]^{\frac{1}{2}}.$$

Note that for $a_n \in (1,n+1)$

$$\int_0^{\infty}\left|\left(\left[1-\frac{y}{n+1}\right]^+\right)^{n-1} - \left(1+\frac{2y-\frac{1}{2}y^2}{n+1}\right)e^{-y}\right|^2 e^y\,dy$$

$$(3.15)\quad \le \int_0^{a_n} [O(n^{-2}a_n^4)]^2\, e^{-y}\,dy + \int_{a_n}^{n+1}\left(1-\frac{y}{n+1}\right)^{2(n+1)}\left(1-\frac{a_n}{n+1}\right)^{-4} e^y\,dy$$

$$+ \int_{a_n}^{\infty}\left(1+\frac{2y-\frac{1}{2}y^2}{n+1}\right)^2 e^{-y}\,dy$$

$$= O\left(n^{-4}a_n^8 + \left(1-\frac{a_n}{n+1}\right)^{-4} e^{-a_n} + (1+n^{-2}a_n^4)e^{-a_n}\right).$$

Substituting $a_n = 4\log(n+1)$ in (3.15) we see that there exists a finite fixed constant A_0 such that (3.14) and (2.3) imply

$$(3.16)\quad \left|\int_0^{n+1} g_n(y)\left(1-\frac{y}{n+1}\right)^{n-1} dy - \int_0^{\infty} g_n(y)\left(1+\frac{2y-\frac{1}{2}y^2}{n+1}\right)e^{-y}\,dy\right|$$

$$\le A_0 C^{\frac{1}{3}} n^{-2}\log^4(n+1).$$

Furthermore,

$$(3.17)\quad |\nu_n| \le \left|\int_0^{n+1} g_n(y)\left(1-\frac{y}{n+1}\right)^{n-1} dy\right| \le \int_0^{n+1} |g_n(y)|e^{-y}\, e^{\frac{2y}{n+1}}\,dy \le e^2 C^{\frac{1}{3}}.$$

Let

$$(3.18)\quad \alpha_n = \omega_n^2 - \int_0^{\infty} g_n^2(y)e^{-y}\,dy$$

$$- \frac{n(n-1)}{n+1}\int_0^{\infty}\int_0^{\infty} g_n(x)g_n(y)\left(1+\frac{3(x+y)-\frac{1}{2}(x+y)^2}{n+1}\right)e^{-x-y}\,dxdy$$

$$+ n\left(\int_0^{\infty} g_n(y)\left(1+\frac{2y-\frac{1}{2}y^2}{n+1}\right)e^{-y}dy\right)^2.$$

With the aid of (1.12), (1.13), (3.16), (3.17), Hölder's inequality and inequalities like (3.15), we arrive at

$$
\begin{aligned}
|\alpha_n| &\le \left\{\int_0^\infty |g_n(y)|^3 e^{-y} dy\right\}^{\frac{2}{3}} \left\{\int_0^\infty \left|\left(\left[1-\frac{y}{n+1}\right]^+\right)^{n-1} - e^{-y}\right|^3 e^{2y} dy\right\}^{\frac{1}{3}} \\
&\quad + \frac{n(n-1)}{n+1}\left\{\int_0^\infty \int_0^\infty g_n^2(x) g_n^2(y) e^{-x-y} dxdy\right\}^{\frac{1}{2}} \\
&\quad \cdot \left\{\int_0^\infty z\left[\left(\left[1-\frac{z}{n+1}\right]^+\right)^{n-2} - \left(1+\frac{3z-\frac{1}{2}z^2}{n+1}\right)e^{-z}\right]^2 e^z dz\right\}^{\frac{1}{2}} \qquad (3.19) \\
&\quad + 2n\, e^2\, C^{\frac{1}{3}}\, A_0\, C^{\frac{1}{3}}\, n^{-2} \log^4(n+1) \\
&\quad + n\, C^{\frac{2}{3}}\, A_0^2\, n^{-4} \log^8(n+1) \\
&\le A_1\, C^{\frac{2}{3}}\, n^{-1} \log^4(n+1),
\end{aligned}
$$

where A_1 is a finite constant independent of n and g_n. Straightforward computation shows that

$$
\begin{aligned}
&|\omega_n^2 - \alpha_n - \sigma_n^2| \\
&= (n+1)^{-2}\left|(-13n+1)\mu_n^2 + n\int_0^\infty g_n(x)x^2 e^{-x}dx\left\{4\mu_n - \tfrac{1}{4}\int_0^\infty g_n(x)x^2e^{-x}dx\right\}\right| \qquad (3.20) \\
&\le A_2\, C^{\frac{2}{3}}\, n^{-1}
\end{aligned}
$$

for some finite constant A_2 which is independent of n and g_n. From (3.19) and (3.20) we conclude that

$$
|\omega_n^2 - \sigma_n^2| \le A_3\, C^{\frac{2}{3}}\, n^{-\frac{2}{3}}, \qquad (3.21)
$$

where, again, A_3 is independent of n and g_n. Together with assumptions (2.3) and (2.4) this implies that for $n \ge n_0^* = n_0 \vee \{(2A_3)^{\frac{3}{2}} c^{-\frac{3}{2}} C\}$

$$
\omega_n^2 \ge \sigma_n^2 - \tfrac{1}{2}c \ge \tfrac{1}{2}c, \qquad (3.22)
$$

$$
\omega_n^2 \le \sigma_n^2 + \tfrac{1}{2}c \le \tfrac{3}{2}\sigma_n^2 \le \tfrac{3}{2}C^{\frac{2}{3}}. \qquad (3.23)
$$

With (3.13), (3.17), (3.22) and (3.23) we arrive at

$$\sup_{n\geq n_0^*} \int_0^\infty |g_n^*(y)|^3 e^{-y}\, dy \leq 2^{\frac{7}{2}}(1+e^6)c^{-\frac{3}{2}} C, \tag{3.24}$$

$$\inf_{n\geq n_0^*} \{\operatorname{var} g_n^*(Y_1) - [\operatorname{cov}(g_n^*(Y_1),Y_1)]^2\} \geq \tfrac{2}{3}c\, C^{-\frac{2}{3}}. \tag{3.25}$$

Studying (1.6), (3.24) and (3.25) we see that it suffices to prove (2.7) under (2.3) through (2.6) and the supplementary assumptions (cf. (3.11) and (3.12))

$$\nu_n = 0, \tag{3.26}$$

$$\omega_n = 1 \tag{3.27}$$

and (3.10). Note that in this case (cf. (3.13) and (2.1)) $g_n^* = g_n$ and $T_n^* = n^{-\frac{1}{2}} T_n$.

We shall apply Lemma 3.2 with G_n the probability measure on $\mathbb{R}^2$ corresponding to the random vector $(Y_1-1, g_n(Y_1)-\mu_n)$. Then

$$V_n = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_n^2 \end{pmatrix}, \quad B_n = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_n^{-1} \end{pmatrix} \tag{3.28}$$

and with the aid of (2.3) and (2.4) it is not difficult to prove that there exists a finite constant c_4 only depending on c and C such that c_3 from Lemma 3.2 with our choice of G_n satisfies

$$c_3 \leq c_4. \tag{3.29}$$

Let now (cf. (3.5) and (3.41))

$$\delta = 2^{\frac{1}{2}}\, 3^{-\frac{1}{2}}\, c_1 c_4^{-1} \tag{3.30}$$

and fix $b \in (0, C^{-\frac{1}{3}}\{(1+\frac{1}{2}\delta^2)^{-\frac{1}{2}} - (1+\delta^2)^{-\frac{1}{2}}\})$ in such a way that (cf. (2.2), (2.6) and (3.37))

(3.31) $\quad K_{n_0}(b, \frac{1}{4}\log(1+\frac{1}{2}\delta^2)) < \infty.$

Let $\chi_n(t)$ be the characteristic function of $T_n^* = n^{-\frac{1}{2}}T_n$. By Esseen's smoothing lemma (see e.g. Lemma XVI.3.2 of FELLER (1971)) it suffices to prove that

$$(3.32) \qquad \int_{|t|\le bn^{\frac{1}{2}}} |t|^{-1}|\chi_n(t) - e^{-\frac{1}{2}t^2}|dt \le An^{-\frac{1}{2}}, \qquad n \ge n_0,$$

where A is a generic constant only depending on n_0, K_{n_0}, c and C. Since for all $t \in \mathbb{R}$

$$|t^{-1}(\chi_n(t) - 1)| \le E|t^{-1}(e^{itT_n^*} - 1)| \le E|T_n^*| \le 1,$$

we have

$$(3.33) \qquad \int_{|t|\le n^{-\frac{1}{2}}} |t|^{-1}|\chi_n(t) - e^{-\frac{1}{2}t^2}|dt \le 4n^{-\frac{1}{2}}.$$

We now choose a version of the conditional distribution of $n^{-\frac{1}{2}}W_n$ given $n^{-\frac{1}{2}}S_n = x$ which is continuous in $x = 0$. Lemma 3.1 shows that this is possible and moreover it implies

$$(3.34) \qquad \chi_n(t) = E\, e^{itT_n^*} = E\left(e^{itn^{-\frac{1}{2}}W_n} \mid n^{-\frac{1}{2}}S_n = 0\right).$$

Let h_n be the density of $n^{-\frac{1}{2}}S_n$ and let $\psi_n(s,t)$ be the characteristic function of $(n^{-\frac{1}{2}}S_n, n^{-\frac{1}{2}}W_n)$, i.e. (cf. (1.6))

$$(3.35) \qquad \psi_n(s,t) = [\rho_n(n^{-\frac{1}{2}}s, n^{-\frac{1}{2}}t)]^{n+1}.$$

We note that (3.31) implies that for all $|t| \le bn^{\frac{1}{2}}$ and all $n \ge n_0$

$$\int_{-\infty}^{\infty} |\psi_n(s,t)|ds < \infty$$

and that Lemma 3.1 yields the continuity of $E(\exp\{itn^{-\frac{1}{2}}W_n\} \mid n^{-\frac{1}{2}}S_n = x)$ in $x = 0$. Together with

$$\psi_n(s,t) = \int_{-\infty}^{\infty} e^{isx}\left\{E\left(e^{itn^{-\frac{1}{2}}W_n} \mid n^{-\frac{1}{2}}S_n = x\right)h_n(x)\right\}dx$$

and (3.34) these facts imply by Fourier inversion (cf. Theorem 21.49 of HEWITT & STROMBERG (1965))

$$\chi_n(t) = (2\pi h_n(0))^{-1} \int_{-\infty}^{\infty} \psi_n(s,t)ds, \quad |t| \le bn^{\frac{1}{2}}. \tag{3.36}$$

In view of (3.31), the inequality

$$|\rho_n(s,t)| \le |\rho_n(s,0)| + E\left|e^{is(Y_1-1)}\left\{e^{itg_n(Y_1)} - 1\right\}\right|$$
$$\le (1+s^2)^{-\frac{1}{2}} + C^{\frac{1}{3}}|t|$$

and the choice of b we also have

$$\int_{|s|\ge\delta n^{\frac{1}{2}}} |\psi_n(s,t)|ds \le K_{n_0}(b,\tfrac{1}{4}\log(1+\tfrac{1}{2}\delta^2))\, n^{-\frac{1}{2}}(\log(n+1))^{-1} + 2n^{\frac{1}{2}}(1+\tfrac{1}{2}\delta^2)^{-\frac{1}{4}n}, \quad |t| \le bn^{\frac{1}{2}}. \tag{3.37}$$

From the theory of Edgeworth expansions for the densities of sums of independent and identically distributed random variables (see e.g. Theorem XVI.2.2 of FELLER (1971)) it follows that

$$h_n(0) = (2\pi)^{-\frac{1}{2}}(1 - \frac{7}{12n}) + O(n^{-\frac{3}{2}}). \tag{3.38}$$

Combining (3.32), (3.33), (3.36), (3.37) and (3.38) we see that it suffices to show that

$$\int_{n^{-\frac{1}{2}}\le|t|\le bn^{\frac{1}{2}}} \left| \int_{|s|\le\delta n^{\frac{1}{2}}} t^{-1}\left[\frac{\psi_n(s,t)}{2\pi h_n(0)} - (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}(s^2+t^2)}\right]ds\right| dt \le An^{-\frac{1}{2}}. \tag{3.39}$$

The left-hand side of (3.39) is dominated by the quantity

$$\int_{n^{-\frac{1}{2}}\le|t|\le bn^{\frac{1}{2}}} (2\pi|t|)^{-1}\left|\int_{|s|\le\delta n^{\frac{1}{2}}} \left[(h_n(0))^{-1} - (2\pi)^{\frac{1}{2}}\right]e^{-\frac{1}{2}(s^2+t^2)}ds\right| dt +$$

$$+\int_{n^{-\frac12}\le|t|\le bn^{\frac12}}(2\pi h_n(0)|t|)^{-1}\left|\int_{|s|\le\delta n^{\frac12}}\left[e^{itn^{-\frac12}(n+1)\mu_n}-1\right]e^{-\frac12(s^2+t^2)}ds\right|dt$$

(3.40)

$$+\int_{n^{-\frac12}\le|t|\le bn^{\frac12}}(2\pi h_n(0)|t|)^{-1}\left|\int_{|s|\le\delta n^{\frac12}}e^{itn^{-\frac12}(n+1)\mu_n}\cdot\left[e^{-\frac12(s^2+\sigma_n^2t^2)}-e^{-\frac12(s^2+t^2)}\right]ds\right|dt$$

$$+\int_{n^{-\frac12}\le|t|\le bn^{\frac12}}(2\pi h_n(0)|t|)^{-1}\left|\int_{|s|\le\delta n^{\frac12}}\left[e^{-itn^{-\frac12}(n+1)\mu_n}\psi_n(s,t)-e^{-\frac12(s^2+\sigma_n^2t^2)}\right]ds\right|dt.$$

In view of (3.38) the first term of (3.40) equals at most $An^{-1}\log(n+1)$. Since $|\exp\{itn^{-\frac12}(n+1)\mu_n\}-1|\le|t|n^{-\frac12}(n+1)|\mu_n|$ and since in view of (3.16) and (3.26) $|\mu_n|\le An^{-1}$, the second term of (3.40) is dominated by $An^{-\frac12}$. In view of (3.21) and (3.27) the third term of (3.40) equals at most $An^{-\frac23}$. From Lemma 3.2 with $\alpha=(0,0)$, from (3.28), (3.29) and (3.30) and from the inequality

$$[\delta^2+C^{\frac23}b^2]^{\frac12}\le[\delta^2+(1+\tfrac12\delta^2)^{-1}-(1+\delta^2)^{-1}]^{\frac12}\le[\tfrac32\delta^2]^{\frac12}=c_1c_4^{-1} \tag{3.41}$$

it follows that

$$\int_{1\le|t|\le bn^{\frac12}}|t|^{-1}\left|\int_{|s|\le\delta n^{\frac12}}\left[e^{-itn^{-\frac12}(n+1)\mu_n}\psi_n(s,t)-e^{-\frac12(s^2+\sigma_n^2t^2)}\right]ds\right|dt\le An^{-\frac12}. \tag{3.42}$$

Furthermore, Lemma 3.2 with $\alpha=(0,1)$ and the mean value theorem imply that

$$\int_{n^{-\frac12}\le|t|\le1}|t|^{-1}\left|\int_{|s|\le\delta n^{\frac12}}\left\{\left[e^{-itn^{-\frac12}(n+1)\mu_n}\psi_n(s,t)-e^{-\frac12(s^2+\sigma_n^2t^2)}\right]-\left[\psi_n(s,0)-e^{-\frac12 s^2}\right]\right\}ds\right|dt\le An^{-\frac12}. \tag{3.43}$$

Combining (3.36), (3.37) and (3.38) we see that

$$(3.44)\qquad \int_{n^{-\frac{1}{2}}\le|t|\le 1} |t|^{-1}\left|\int_{|s|\le\delta n^{\frac{1}{2}}}\left[\psi_n(s,0) - e^{-\frac{1}{2}s^2}\right]ds\right|dt \le An^{-\frac{1}{2}}.$$

From (3.39), (3.40), (3.42), (3.43) and (3.44) we conclude that assertion (3.32) and hereby Theorem 2.1 have been proved. □

4. PROOF OF COROLLARY 2.1.

We shall apply Theorem 2.1 with $g_n = g$. In order to verify (2.6) we have to consider

$$(4.1)\qquad \rho(s,t) = \int_0^\infty e^{is(y-1)+itg(y)}\, e^{-y}\, dy.$$

For all $\varepsilon \in (0,\frac{1}{2})$ we have

$$(4.2)\qquad |\rho(s,t)| \le 2\varepsilon + \left|\int_\varepsilon^{-\log\varepsilon} e^{(is-1)y}\, e^{itg(y)}\, dy\right|.$$

For all integers m and all real numbers $\varepsilon = a_0 \le a_1 \le \ldots \le a_m = -\log\varepsilon$, $b_1 \le b_2 \le \ldots \le b_m$ and $c_1 \le c_2 \le \ldots \le c_m$ we define

$$(4.3)\qquad g_0(y) = \sum_{i=1}^m b_i\, 1_{(a_{i-1},a_i]}(y),\quad g_1(y) = \sum_{i=1}^m c_i\, 1_{(a_{i-1},a_i]}(y).$$

For the function $g(y) = g_0(y) - g_1(y)$ the following string of (in)equalities holds:

$$\left|\int_\varepsilon^{-\log\varepsilon} e^{(is-1)y}\, e^{itg(y)}dy\right| = \left|\sum_{j=1}^m \int_{a_{j-1}}^{a_j} e^{(is-1)y}\, e^{it(b_j-c_j)}dy\right|$$

$$(4.4)\qquad = |1-is|^{-1}\left|\sum_{j=1}^{m-1}\left(e^{it(b_{j+1}-c_{j+1})} - e^{it(b_j-c_j)}\right)e^{-(1-is)a_j} + e^{it(b_1-c_1)}e^{-(1-is)a_0} - e^{it(b_m-c_m)}e^{-(1-is)a_m}\right|$$

$$\le (1+s^2)^{-\frac{1}{2}}\left\{|t|\sum_{j=1}^{m-1}(b_{j+1}-b_j+c_{j+1}-c_j)e^{-a_j}+e^{-a_0}+e^{-a_m}\right\}.$$

From (4.2) and (4.4) it follows that

$$(4.5)\quad |\rho(s,t)| \le 3\varepsilon + (1+s^2)^{-\frac{1}{2}}\Big\{1 + |t|\Big[\int_{\varepsilon}^{-\log\varepsilon} (g_0(y)+g_1(y))e^{-y}\,dy - (g_0(\varepsilon)+g_1(\varepsilon))e^{-\varepsilon} + (g_0(-\log\varepsilon)+g_1(-\log\varepsilon))\varepsilon\Big]\Big\}.$$

By a limiting argument we see that (4.5) holds for all nondecreasing functions g_0 and g_1. From (2.10) we see that there exists a constant d_0 such that for $i = 0,1$

$$|g_i(y)| \le \begin{cases} d_0 y^{-\alpha} & \text{for } y < 1, \\ d_0 e^{\alpha y} & \text{for } y \ge 1, \end{cases}$$

which together with (4.5) yields the existence of a constant d_1 such that for all $\varepsilon > 0$

$$(4.6)\quad |\rho(s,t)| \le 3\varepsilon + (1+s^2)^{-\frac{1}{2}}\{1 + d_1|t|\varepsilon^{-\alpha}\}.$$

Choosing $\varepsilon = (|t|(1+s^2)^{-\frac{1}{2}})^{1/(\alpha+1)}$ we see that for

$$|t| \le \{(3+d_1)^{-1}[1 - (1+s^2)^{-\frac{1}{2}\alpha/(\alpha+1)}]\}^{\alpha+1}$$

the inequality

$$(4.7)\quad |\rho(s,t)| \le (1+s^2)^{-\frac{1}{2}/(\alpha+1)}$$

holds, which implies the validity of (2.6) for $n_0 > \alpha+1$. Since (2.3), (2.4) and (2.5) are trivially fulfilled for $g_n = g$, Theorem 2.1 now yields

$$(4.8)\quad \sup_{x\in\mathbb{R}} \left|H_n(x) - \Phi\left(\frac{x - ET_n'}{(\operatorname{var} T_n')^{\frac{1}{2}}}\right)\right| \le An^{-\frac{1}{2}}.$$

Combining (2.8), (1.2), (1.12), (2.11) and (3.16) we see that $ET_n' = O(n^{-\frac{1}{2}})$ and combining (3.12), (3.21), (3.10), (1.8), (1.9) and (2.12) that $\operatorname{var} T_n' = 1 + O(n^{-\frac{2}{3}})$. Together with (4.8) this yields (2.14) and hence the corollary. □

ACKNOWLEDGEMENT. *We are grateful to Dr. R. Helmers and Prof. W.R. van Zwet for stimulating discussions.*

REFERENCES

BHATTACHARYA, R.N. & R.R. RAO (1976), *Normal approximation and asymptotic expansions*, Wiley, New York.

BILLINGSLEY, P. (1968), *Convergence of probability measures*, Wiley, New York.

DOES, R.J.M.M. & R. HELMERS (1980), *Edgeworth expansions for functions of uniform spacings*, Mathematical Centre Report SW 55, Amsterdam, to appear in the Proceedings of the Colloquium on Nonparametric Statistical Inference, I. Vincze (ed.), Budapest.

FELLER, W. (1971), *An introduction to probability theory and its applications*, Vol. II, second edition, Wiley, New York.

HEWITT, E. & K. STROMBERG (1965), *Real and abstract analysis*, Springer-Verlag, Berlin.

KOZIOL, J.A. (1980), *A note on limiting distributions for spacings statistics*, Z. Wahrsch. Verw. Gebiete 51, 55-62.

LE CAM, L. (1958), *Un théorème sur la division d'un intervalle par des points pris au hasard*, Publ. Inst. Statist. Univ. Paris 7, 7-16.

PYKE, R. (1965), *Spacings*, J. Roy. Statist. Soc. Ser. B 27, 359-449.

PYKE, R. (1972), *Spacings revisited*, Proc. Sixth Berkeley Symp. Math. Statist. Prob. 1, 417-427.