COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
32. NONPARAMETRIC STATISTICAL INFERENCE,
BUDAPEST (HUNGARY), 1980.

EDGEWORTH EXPANSIONS FOR FUNCTIONS OF UNIFORM SPACINGS*

R.J.M.M. DOES — R. HELMERS

1. INTRODUCTION

Let $U_1, U_2, \ldots$ be a sequence of independent uniform $(0, 1)$ random variables (r.v.'s). For $n = 1, 2, \ldots,$ $U_{1:n} \leqslant U_{2:n} \leqslant \ldots \leqslant U_{n:n}$ denote the ordered $U_1, U_2, \ldots, U_n$. Let $U_{0:n} = 0$ and $U_{n+1:n} = 1$. Uniform spacings are defined by

$$D_{in} = U_{i:n} - U_{i-1:n} \tag{1.1}$$

for $i = 1, 2, \ldots, n + 1$. Let g be a fixed real-valued nonlinear measurable function and define statistics T_n $(n = 1, 2, \ldots)$ by

$$T_n = \sum_{i=1}^{n+1} g((n + 1)D_{in}). \tag{1.2}$$

It is well-known that suitably normalized statistics of the form (1.2) are asymptotically $\mathcal{N}(0, 1)$ distributed under quite general conditions. We refer to the papers of Darling [3], Le Cam [6], Shorack [10],

*Report SW 55/80 Mathematisch Centrum, Amsterdam.

Pyke [9] and Koziol [5] (see also the theorem contained in the discussion between Kingman and Pyke in Pyke [8]). A survey of the general area of limit theorems for spacings may be found in Pyke [9].

The purpose of this paper is to establish Edgeworth expansions for statistics of the form (1.2) with remainder $o\left(\frac{1}{n}\right)$. Using a well-known characterization of the joint distribution of uniform spacings the problem is reduced to that of deriving an asymptotic expansion for the distribution function (d.f.) of a normalized sum of independent r.v.'s conditioned on another sum of independent r.v.'s. This latter problem is then dealt with in a standard manner by applying classical results for sums of independent random vectors.

After the result of the present paper was obtained, a related paper of Michel [7] appeared. In his paper he proves, using a similar method of proof, a general theorem on asymptotic expansions for conditional distributions. However, Michel's result is, though more general in scope, less explicit than ours when applied to the conditional distributions we have to consider. Another related paper is that of Van Zwet [11]. He derives the Edgeworth expansion for linear combinations of uniform order statistics under minimal conditions.

2. THE THEOREM

For any r.v. X let us denote by X^* the r.v. $X^* = \dfrac{X - \mathsf{E}(X)}{\sigma(X)}$ whenever $0 < \sigma(X) < \infty$. Let Y be an exponential r.v. with expectation 1 and let g be a fixed real-valued nonlinear measurable function defined on R^+. Introduce, whenever well-defined, for integers i and j, quantities m_{ij} by

$$(2.1) \qquad m_{ij} = \mathsf{E}((g^*(Y))^i (Y^*)^j).$$

We shall establish an asymptotic expansion with remainder $o\left(\frac{1}{n}\right)$ for the d.f.

$$(2.2) \qquad G_n(x) = \mathsf{P}\left(\frac{1}{\sigma\tau\sqrt{n}}(T_n - (n+1)\mu) \leqslant x\right) \qquad (-\infty < x < \infty),$$

where

(2.3) $$\mu = \int_0^\infty g(y)e^{-y}\, dy, \quad \sigma^2 = \int_0^\infty (g(y) - \mu)^2 e^{-y}\, dy$$

and

(2.4) $$\tau = \sqrt{1 - m_{11}^2}\,.$$

Note that $\sigma\tau > 0$ if and only if g is not linear.

Let Φ and φ denote the d.f. and the density of the standard normal distribution and let ρ denote the characteristic function (ch. f.) of $(g^*(Y), Y^*)$; i.e.

(2.5) $$\rho(s, t) = \mathsf{E} e^{isg^*(Y) + itY^*} \quad \text{for} \quad (s, t) \in R^2.$$

Theorem 1. *Suppose that the assumptions*

(i) $\int_0^\infty g^4(y)e^{-y} dy < \infty$;

(ii) $\int_{-\infty}^\infty \int_{-\infty}^\infty |\rho(s, t)|^p\, dsdt < \infty$ *for some* $p \geqslant 1$

are satisfied. Then we have that

(2.6) $$\sup_x |G_n(x) - \widetilde{G}_n(x)| = o\left(\frac{1}{n}\right) \quad \text{as} \quad n \to \infty,$$

where

$$\widetilde{G}_n(x) = \Phi(x) - \varphi(x) \Big\{ \frac{1}{\sqrt{n}} \left(\frac{1}{6} \kappa_3(x^2 - 1) + a\right) +$$
$$+ \frac{1}{n} \left(\frac{1}{24} \kappa_4(x^3 - 3x) + \frac{1}{72} \kappa_3^2(x^5 - 10x^3 + 15x) +\right.$$
$$\left. + \frac{1}{8} (-4a\kappa_3 + b)x + \frac{1}{6} a\kappa_3 x^3\right)\Big\}$$

with

$$\kappa_3 = \frac{1}{\tau^3} (m_{30} - 3m_{21}m_{11} + 3m_{12}m_{11}^2 - 2m_{11}^3),$$

$$\kappa_4 = \frac{1}{\tau^4}(m_{40} - 3 - 3m_{21}^2 + 12m_{12}m_{21}m_{11} - 4m_{31}m_{11} +$$

$$+ 6m_{11}^2 + 6m_{22}m_{11}^2 - 12m_{12}^2m_{11}^2 - 12m_{21}m_{11}^2 +$$

$$+ 24m_{12}m_{11}^3 - 4m_{13}m_{11}^3 - 6m_{11}^4),$$

$$a = \frac{1}{2\tau}(-m_{12} + 2m_{11})$$

and

$$b = \frac{1}{\tau^2}(6 + 4m_{21} + 3m_{12}^2 - 2m_{22} - 20m_{12}m_{11} +$$

$$+ 4m_{13}m_{11} - 4m_{11}^2).$$

In Theorem 1 we establish an asymptotic expansion for the d.f. of suitably normalized sums of a function of uniform spacings with remainder $o\left(\frac{1}{n}\right)$. In fact it can be proved that under the assumptions of Theorem 1 there exists an expansion for $\mathrm{P}\left(\frac{1}{\sigma\tau\sqrt{n}}(T_n - (n+1)\mu) \in A\right)$ for every Borel set A, the remainder being $o\left(\frac{1}{n}\right)$ uniform over all Borel sets. However, for statistical applications, the theorem in its present form will be of sufficient generality. The theorem requires a natural moment condition (i). Moreover, the ch.f. ρ (cf. (2.5)) must satisfy an integrability assumption (ii) which is commonly encountered in problems of establishing expansions for conditional distributions (see, e.g., Albers [1], Ch. 2, and Michel [7]). An assumption, equivalent to (ii) and obviously implying that g is nonlinear, is that the n-th convolution of the distribution of $(g^*(Y), Y^*)$ possesses almost everywhere on R^2 a bounded density for all sufficiently large n (see Bhattacharya and Rao [2], Th. 19.1). The verification of this assumption is a problem in itself which we have not attempted to solve.

We conclude this section with some remarks. In the first place we note that it is not difficult to check that under the assumptions of Theorem 1 not only (2.6) holds but that also

$$\sup_x |F_n^*(x) - \widetilde{F}_n(x)| = o\left(\frac{1}{n}\right) \quad \text{as} \quad n \to \infty,$$

is true, where $F_n^*(x) = \mathrm{P}(T_n^* \leqslant x)$ and

$$\widetilde{F}_n(x) = \Phi(x) - \varphi(x)\left\{\frac{1}{6\sqrt{n}}\kappa_3(x^2 - 1) + \frac{1}{24n}\kappa_4(x^3 - 3x) + \right.$$
$$\left. + \frac{1}{72n}\kappa_3^2(x^5 - 10x^3 + 15x)\right\}.$$

Note that $\widetilde{F}_n$ does not depend on the quantities a and b appearing in the expansion $\widetilde{G}_n$. This is due to the exact standardization we have employed here. Secondly we remark that Theorem 1 provides a partial answer to a question posed by Pyke [9] (his problem 5) concerning the rate of convergence to normality for functions of uniform spacings of the form T_n. Typically the error committed when the normal approximation is applied is of order $\frac{1}{\sqrt{n}}$. Finally, we note that, as the error of the expansion given in (2.6) is $o\left(\frac{1}{n}\right)$, we may expect, that, at least for not too small sample sizes, the expansion will yield a better numerical approximation for the distribution function of T_n than can be provided by the normal approximation.

3. PROOF OF THE THEOREM

As indicated in the introduction we begin the proof by reducing our problem to that of deriving an asymptotic expansion for the d.f. of a sum of independent r.v.'s, conditioned on another sum of independent r.v.'s. Let $Y_1, Y_2, \ldots$ be independent exponentially distributed r.v.'s with expectation 1 and define r.v.'s X_n and S_n $(n = 1, 2, \ldots)$, by

$$(3.1) \qquad X_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n+1} g^*(Y_i), \quad S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n+1} Y_i^*.$$

A well-known characterization of the joint distribution of uniform spacings (see, e.g., Pyke [8]) implies that the distribution of

$$\frac{1}{\sigma\sqrt{n}}(T_n - (n+1)\mu) \qquad (n = 1, 2, \ldots),$$

is the same as the conditional distribution of X_n given $S_n = 0$ $(n = 1, 2, \ldots)$; i.e.

$$G_n(x) = \mathsf{P}\left(\frac{1}{\sigma\tau\sqrt{n}}(T_n - (n+1)\mu) \leq x\right) = \mathsf{P}(X_n \leq x\tau \mid S_n = 0). \tag{3.2}$$

The problem to establish an asymptotic expansion for the conditional probabilities in (3.2) will now be solved in a number of steps. In the first place we establish an expansion for the ch.f. ψ_n of (X_n, S_n); i.e. of

$$\psi_n(s, t) = \mathsf{E}\, e^{isX_n + itS_n} = \rho^{n+1}\left(\frac{s}{\sqrt{n}}, \frac{t}{\sqrt{n}}\right), \tag{3.3}$$

where ρ is defined in (2.5). We shall prove that for some $\delta > 0$ and uniformly for all $(s, t) \in R^2$ satisfying $s^2 + t^2 \leq \delta n$ we have that

$$\psi_n(s, t) = \widetilde{\psi}_n(s, t) + o\left(\frac{P(s, t)e^{-\frac{1}{4}Q(s,t)}}{n}\right), \tag{3.4}$$

where P is a polynomial of the sixth degree in s and t, without constant term and $Q(s, t) = s^2 + 2m_{11}st + t^2$. The function $\widetilde{\psi}_n$ is given by

$$\begin{aligned} \widetilde{\psi}_n(s, t) = {} & e^{-\frac{1}{2}(s^2 + 2m_{11}st + t^2)} \times \\ & \times \Big\{1 - \frac{i}{6\sqrt{n}}(m_{30}s^3 + 3m_{21}s^2t + 3m_{12}st^2 + 2t^3) + \\ & + \frac{1}{24n}((m_{40} - 3)s^4 + 4(m_{31} - 3m_{11})s^3t + \\ & + 6(m_{22} - 2m_{11}^2 - 1)s^2t^2 + 4(m_{13} - 3m_{11})st^3 + 6t^4) - \\ & - \frac{1}{72n}(m_{30}s^3 + 3m_{21}s^2t + 3m_{12}st^2 + 2t^3)^2 - \\ & - \frac{1}{2n}(s^2 + 2m_{11}st + t^2)\Big\}. \end{aligned} \tag{3.5}$$

To check this we first note that because of assumption (i) we can expand the ch.f. of $(g^*(Y_1), Y_1^*)$ in a neighbourhood of the origin. A simple computation yields that uniformly for all $(s, t) \in R^2$ satisfying $s^2 + t^2 = o(n)$

$$\rho\left(\frac{s}{\sqrt{n}}, \frac{t}{\sqrt{n}}\right) = 1 - \frac{1}{2n}(s^2 + 2m_{11}st + t^2) -$$

$$(3.6) \qquad - \frac{i}{6n^{\frac{3}{2}}}(m_{30}s^3 + 3m_{21}s^2t + 3m_{12}st^2 + 2t^3) +$$

$$+ \frac{1}{24n^2}(m_{40}s^4 + 4m_{31}s^3t + 6m_{22}s^2t^2 + 4m_{13}st^3 + 9t^4) +$$

$$+ o\left(\frac{R(s, t)}{n^2}\right),$$

where R is a polynomial of the fourth degree in s and t without constant term. Taking now the $(n+1)$-th power of the expansion in (3.6) and expanding further we arrive at (3.4) (see Ch. 2, Sec. 9 in Bhattacharya and Rao [2] for more details).

Next we convert the expansion (3.4) into an expansion for the joint density f_n of (X_n, S_n) in the point $(x, 0)$. After long but easy computations we find that for all sufficiently large n

$$(3.7) \qquad f_n(x, 0) = \tilde{f}_n(x, 0) + r_n(x),$$

where

$$\tilde{f}_n(x, 0) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-isx} \tilde{\psi}_n(s, t)\, ds dt =$$

$$= \frac{1}{2\pi\tau} e^{-\frac{1}{2}x^2\tau^{-2}} \left\{1 - \frac{\kappa_3}{6\sqrt{n}}\left(\frac{3x}{\tau} - \frac{x^3}{\tau^3}\right) + \right.$$

$$(3.8) \qquad + \frac{\kappa_4}{24n}\left(3 - \frac{6x^2}{\tau^2} + \frac{x^4}{\tau^4}\right) -$$

$$- \frac{\kappa_3^2}{72n}\left(15 - \frac{45x^2}{\tau^2} + \frac{15x^4}{\tau^4} - \frac{x^6}{\tau^6}\right) +$$

$$\left. + \frac{ax}{\tau\sqrt{n}} + \frac{a\kappa_3}{6n}\left(3 - \frac{6x^2}{\tau^2} + \frac{x^4}{\tau^4}\right) - \frac{b}{8n}\left(1 - \frac{x^2}{\tau^2}\right) - \frac{7}{12n}\right\}$$

and

$$(3.9) \qquad (1 + x^4)r_n(x) = o\left(\frac{1}{n}\right)$$

uniformly in x. Here we have used assumption (ii) to guarantee that f_n exists for all sufficiently large n (see Th. 19.1 of Bhattacharya and Rao [2]) and to validate the application of the Fourier inversion theorem in R^2 (see Feller [4], p. 524). The nonuniform estimate of the remainder $r_n(x)$ in (3.9) follows directly from Theorem 19.2 of Bhattacharya and Rao [2]. The validity of this application of Theorem 19.2 of Bhattacharya and Rao [2] can be inferred from the assumptions of Theorem 1.

We also need an expansion for the marginal density f_{n2} of S_n in the point 0. Since $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n+1} (Y_i - 1)$ (cf. (3.1)) the theory of asymptotic expansions for the densities of sums of i.i.d. r.v.'s can be applied (see e.g. Feller [4], Th. XVI. 2.2). It follows that

$$f_{n2}(0) = \frac{1}{\sqrt{2\pi}} \left(1 - \frac{7}{12n}\right) + O\left(\frac{1}{n^{\frac{3}{2}}}\right). \tag{3.10}$$

We are now in a position to prove (2.6). Remark first that using (3.2) and the fact that the conditional density of X_n, given $S_n = 0$, in the point x is obviously equal to $\frac{f_n(x, 0)}{f_{n2}(0)}$, yields that for all sufficiently large n

$$G_n(x) = \int_{-\infty}^{x\tau} \frac{f_n(y, 0)}{f_{n2}(0)} \, dy. \tag{3.11}$$

The relations (3.7)-(3.10) provide expansions for the numerator and denominator of the integrand in (3.11). To find an expansion for $G_n(x)$ from these results we note that

$$\frac{f_n(y, 0)}{f_{n2}(0)} = \sqrt{2\pi} \left(1 + \frac{7}{12n} + O\left(\frac{1}{n^{\frac{3}{2}}}\right)\right) (\tilde{f}_n(y, 0) + r_n(y)) =$$

$$= \sqrt{2\pi} \tilde{f}_n(y, 0) + \frac{7\varphi\left(\frac{y}{\tau}\right)}{12n\tau} + \frac{o\left(\frac{1}{n}\right)}{1 + y^2} + \frac{o\left(\frac{1}{n}\right)}{1 + y^4} + \tag{3.12}$$

$$+ O\left(\frac{1}{n^{\frac{3}{2}}}\right) \tilde{f}_n(y, 0)$$

uniformly in y. Note that to obtain the third term in the second line we have used the remark following Theorem 19.2 of Bhattacharya and Rao [2] to check that $\sqrt{2\pi} f_n(y, 0) = \frac{\varphi\left(\frac{y}{\tau}\right)}{\tau} + \frac{o(1)}{1 + y^2}$ uniformly in y. Combining now (3.8) and (3.12) with (3.11) we find that

$$(3.13) \qquad G_n(x) = \int_{-\infty}^{x\tau} \left\{ \sqrt{2\pi} \widetilde{f}_n(y, 0) + \frac{7\varphi\left(\frac{y}{\tau}\right)}{12n\tau} \right\} dy + o\left(\frac{1}{n}\right)$$

uniformly in x. The theorem now follows from (3.8) and (3.13) after a number of elementary integrations.∎

Acknowledgements. The authors are grateful to F.H. Ruymgaart for suggesting the problem and to W.R. Van Zwet for his valuable comments.

REFERENCES

[1] W. Albers, *Asymptotic expansions and the deficiency concept in statistics,* Mathematical Centre Tract 58, Amsterdam, 1974.

[2] R.N. Bhattacharya – R.R. Rao, *Normal approximation and asymptotic expansions.* Wiley, New York, 1976.

[3] D.A. Darling, On a class of problems related to the random division of an interval, *Ann. Math. Statist.*, 24 (1953), 239-253.

[4] W. Feller, *An introduction to probability theory and its applications* 2, 2nd ed., Wiley, New York, 1971.

[5] J.A. Koziol, A note on limiting distributions for spacings statistics, *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 51 (1980), 55-62.

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

[7] R. Michel, Asymptotic expansions for conditional distributions, *J. Multivariate Anal.*, 9 (1979), 393-400.

[8] R. Pyke, Spacings, *J. Roy. Statist. Soc., Ser. B*, 27 (1965), 359-449.

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

[10] G.R. Shorack, Convergence of quantile and spacings processes with applications, *Ann. Math. Statist.*, 43 (1972), 1400-1411.

[11] W.R. Van Zwet, The Edgeworth expansion for linear combinations of uniform order statistics, *Proc. Second Prague Symp. Asymp. Statist.*, (P. Mandl and M. Hušková (eds.)), 93-101, North Holland, Amsterdam, 1979.

R.J.M.M. Does

University of Limburg, Data Processing Department, P.O. Box 616, 6200 MD Maastricht, The Netherlands.

R. Helmers

Mathematical Centre, Dept. of Statistics, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands.