The Annals of Probability
1979, Vol. 7, No. 6, 1066–1072

ASYMPTOTIC NORMALITY OF SUM-FUNCTIONS OF SPACINGS

BY LARS HOLST

Uppsala University

Take n points at random on a circle of unit circumference and order them clockwise. Let $S_0^{(m)}, \cdots, S_{n-1}^{(m)}$ be the mth order spacings, i.e., the clockwise arc-lengths between every pair of points with $m-1$ points between. Ordinary spacings correspond to the case $m = 1$. A central limit theorem is proved for $Z_n = \Sigma_{k=0}^{n-1} h(nS_k, \cdots, nS_{k+m-1})$, where h is a given function. Using this, asymptotic distributions of central order statistics and sums of the logarithms of mth order spacings are derived.

1. Introduction. Let $U_1, U_2, \cdots, U_{n-1}$ be i.i.d. with a uniform distribution over [0, 1]. The corresponding order statistics are $0 \leqslant U_{(1)} \leqslant U_{(2)} \leqslant \cdots \leqslant U_{(n-1)} \leqslant 1$. Define $U_{(0)} = 0$, $U_{(n)} = 1$ and $U_{(k)} = 1 + U_{(k-n)}$ for $k \geqslant n$. The successive distances between the order statistics $S_k = U_{(k+1)} - U_{(k)}$ are usually called spacings. For an integer $m > 1$ generalized spacings of the form

$$S_k^{(m)} = U_{(k+m)} - U_{(k)}, \qquad k = 0, 1, \cdots, n-1,$$

are considered in the present paper. Sometimes these are called mth order spacings, or mth order gaps. A huge literature exists on usual spacings, see, e.g., the reviews by Pyke (1965), (1972).

In connection with directional data it is natural to consider the uniform distribution over a circle of unit circumference. When n points are taken at random from this distribution the successive arc-lengths (with respect to one of the n points chosen at random) have the same distribution as $S_0, S_1, \cdots, S_{n-1}$. Also the mth order spacings have obvious interpretations. The spacings can be used for testing uniformity in various ways, see, e.g., Rao (1976). For this purpose mth order spacings are useful too, see Cressie (1978).

Let h be a given real measurable function and consider the random variable

$$Z_n = \Sigma_{k=0}^{n-1} h(nS_k).$$

A complete characterization of the possible limit laws when $n \to \infty$ was given by Le Cam (1958). It is easy to prove that nS_k converges in distribution to $\Gamma(1, 1)$, where $\Gamma(a, 1)$ stands for a gamma-distribution with mean a and scale-parameter 1. However, the dependence is not asymptotically negligible, so classical limit theorems cannot be used directly. Analogously $nS_k^{(m)}$ converges to $\Gamma(m, 1)$. The $S_k^{(m)}$'s are "more" dependent because of the overlapping. Using a central limit theorem for m-dependent random variables and a generalization of Le Cam's method it is proved below that

$$Z_n^{(m)} = \Sigma_{k=0}^{n-1} h(nS_k, \cdots, nS_{k-m+1})$$

Received April 7, 1978; revised September 21, 1978.
AMS 1970 *subject classifications.* Primary 60F05; secondary 62E20.
Key words and phrases. Spacings, order statistics, uniform distribution, limit theorems.

is asymptotically normally distributed if $0 < \mathrm{Var}(h(X_0, \cdots, X_{m-1})) < \infty$ where $X_0, \cdots, X_{m-1}$ are independent and $\Gamma(1, 1)$ random variables. In particular follows asymptotic normality of $\Sigma_{k=0}^{n-1}\ln(nS_k^{(m)})$. This case has been studied by Cressie (1976) in detail.

Let $S_{(n)} \geqslant \cdots \geqslant S_{(k)} \geqslant \cdots \geqslant S_{(1)}$ be the order statistics of the ordinary spacings. From the results of Le Cam (1958) the limit distribution of the extreme-values and the central order statistics can be deduced. Below asymptotic normality is proved for the central order statistics of mth order spacings. The corresponding lower-extreme values has been studied by Cressie (1977); upper-extreme values can be studied analogously.

The problems considered in this paper have been motivated on the circle of unit circumference, but it is not difficult to see that because the order m is finite, all the asymptotic results will be identical on the unit interval.

2. Asymptotic normality. For notational convenience let $X_0, X_1, \cdots, X_{n-1}$ be i.i.d. $\Gamma(1, 1)$ rv's in the following, and set

$$X_{n+j} = X_j,$$

and

$$T_k = \Sigma_{j=0}^{m-1}X_{k+j}.$$

THEOREM. *If*

$$Eh(X_0, \cdots, X_{m-1}) = 0, \quad 0 < \mathrm{Var}(h(X_0, \cdots, X_{m-1})) < \infty,$$

then

$$\mathcal{L}\big(n^{-\frac{1}{2}}\Sigma_{k=0}^{n-1}h(nS_k, \cdots, nS_{k+m-1})\big) \rightarrow N(0, A - B^2), \quad n \rightarrow \infty,$$

where

$$A = \Sigma_{j=-m+1}^{m-1}\mathrm{Cov}\big(h(X_0, \cdots, X_{m-1}), h(X_j, \cdots, X_{j+m-1})\big)$$

and

$$B = m^{-1}\Sigma_{j=-m+1}^{m-1}\mathrm{Cov}\big(h(X_0, \cdots, X_{m-1}), T_j\big).$$

PROOF. In the lemma below it is proved that for $M \leqslant n - 3$ the characteristic function of $\Sigma_{k=0}^{M-m}h(nS_k, \cdots, nS_{k+m-1})$ can be written

$$g_M(t) = a_n^{-1}\int_{-\infty}^{\infty}E\Big(\exp\big(itn^{-\frac{1}{2}}\Sigma_{k=0}^{M-m}h(X_k, \cdots, X_{k+m-1}) + iu\Sigma_{k=0}^{M}(X_k - 1)\big)\Big)$$
$$\cdot E\big(\exp\big(iu\Sigma_{k=M+1}^{n-1}(X_k - 1)\big)\big)du,$$

where, using Stirling's formula,

$$a_n = 2\pi n^{n-1}e^{-n}/(n-1)! \sim (2\pi/n)^{\frac{1}{2}}, \quad n \rightarrow \infty.$$

Let $n, M \rightarrow \infty$ in such a way that $M/n \rightarrow \alpha$, $0 < \alpha < 1$. From the central limit theorem it follows that

$$E\big(\exp\big(iu\Sigma_{k=M+1}^{n-1}(X_k - 1)/n^{1/2}\big)\big) \rightarrow e^{-(1-\alpha)u^2/2}.$$

Either by direct calculation or using a local central limit theorem it can be shown that

$$\int_{-\infty}^{\infty}|E(\exp(iu\Sigma_{k=M+1}^{n-1}(X_k - 1)/n^{\frac{1}{2}}))|du \to \int_{-\infty}^{\infty}\exp(-(1-\alpha)u^2/2)du.$$

By a variant of the Lebesgue dominated convergence theorem (see Rao (1973), page 136) it now follows

$$g_M(t) \to g_\alpha(t) = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty} f_\alpha(t, u)e^{-(1-\alpha)u^2/2}du,$$

if

$$f_M(t, u) = E\big(\exp\big(itn^{-\frac{1}{2}}\Sigma_{k=0}^{M-m}h(X_k, \cdots, X_{k+m-1}) + iun^{-\frac{1}{2}}\Sigma_{k=0}^{M}(X_k - 1)\big)\big) \to f_\alpha(t, u).$$

Obviously, the random variable

$$tn^{-\frac{1}{2}}\Sigma_{k=0}^{M-m}h(X_k, \cdots, X_{k+m-1}) + un^{-\frac{1}{2}}\Sigma_{k=0}^{M}(X_k - 1)$$

has the same asymptotic distribution (if any) as

$$n^{-\frac{1}{2}}\Sigma_{k=0}^{M-m}(th(X_k, \cdots, X_{k+m-1}) + u(T_k - m)/m)$$

(recall that $T_k = \Sigma_{j=0}^{m-1}X_{k+j}$). By the central limit theorem for m-dependent sequences (see, e.g., Billingsley (1968), page 174)

$$\mathcal{L}\big(n^{-\frac{1}{2}}\Sigma_{k=0}^{M-m}(th(X_k, \cdots, X_{k+m-1}) + u(T_k - m)/m)\big) \to N(0, \alpha\sigma^2(t, u)),$$

where

$$\sigma^2(t, u) = \mathrm{Var}(th(X_0, \cdots, X_{m-1}) + uT_0/m) + 2\Sigma_{j=1}^{m-1}\mathrm{Cov}(th(X_0, \cdots, X_{m-1}) + uT_0/m, th(X_j, \cdots, X_{j+m-1}) + uT_j/m).$$

Hence it follows that

$$f_M(t, u) \to f_\alpha(t, u) = \exp(-\alpha\sigma^2(t, u)/2),$$

and, therefore,

$$g_M(t) \to g_\alpha(t) = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}\exp(-(\alpha\sigma^2(t, u) + (1-\alpha)u^2)/2)du.$$

It is easily seen that $g_\alpha(t) \to 1$ when $\alpha \searrow 0$. Hence, by Le Cam (1958), page 13, Lemma 5,

$$E\big(\exp\big(itn^{-\frac{1}{2}}\Sigma_{k=0}^{n-1}h(nS_k, \cdots, nS_{k+m-1})\big)\big) \to \lim_{\alpha \nearrow 1} g_\alpha(t) = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}\exp(-\sigma^2(t, u)/2)du.$$

Now an elementary calculation gives

$$\begin{aligned}\sigma^2(t, u) = t^2\big(&\mathrm{Var}(h(X_0, \cdots, X_{m-1})) \\ &+ 2\Sigma_{j=1}^{m-1}\mathrm{Cov}(h(X_0, \cdots, X_{m-1}), h(X_j, \cdots, X_{j+m-1}))\big) \\ &+ 2tum^{-1}\big(\mathrm{Cov}(h(X_0, \cdots, X_{m-1}), T_0) + \Sigma_{j=1}^{m-1}\mathrm{Cov}(h(X_0, \cdots, X_{m-1}), T_j) \\ &+ \Sigma_{j=1}^{m-1}\mathrm{Cov}(h(X_j, \cdots, X_{j+m-1}), T_0)\big) + u^2 = t^2A + 2tuB + u^2,\end{aligned}$$

with A and B as in the assertion. Hence

$$(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}\exp(-\sigma^2(t, u)/2)du = \exp(-t^2(A - B^2)/2)$$

which proves the theorem.

LEMMA. *For any real measurable function g on R^M with $M \leqslant n - 3$*

$$E(\exp(itg(nS_0, \cdots, nS_M))) = (2\pi \cdot n^{n-1} \cdot e^{-n}/(n-1)!)^{-1} \cdot \int_{-\infty}^{\infty} E(\exp(itg(X_0, \cdots, X_M) + iu\Sigma_0^{n-1}(X_k - 1)))du.$$

PROOF. Using conditional expectation

$$\begin{aligned} &E(\exp(itg(X_0, \cdots, X_M) + iu\Sigma_0^{n-1}X_k)) \\ &= \int_{-\infty}^{\infty} E(\exp(itg(X_0, \cdots, X_M) + iu\Sigma_0^{n-1}X_k)|\Sigma_0^{n-1}X_k = s) \cdot f_n(s)ds \\ &= \int_{-\infty}^{\infty} e^{ius}E(\exp(itg(X_0, \cdots, X_M))|\Sigma_0^{n-1}X_k = s) \cdot f_n(s)ds, \end{aligned} \tag{2.1}$$

where $f_n(s)$ is the density of $\Sigma_0^{n-1}X_k$, which is $\Gamma(n, 1)$-distributed. As the X's are independent

$$|E(\exp(itg(X_0, \cdots, X_M) + iu\Sigma_0^{n-1}X_k))| \leqslant |E(\exp(iu\Sigma_{M+1}^{n-1}X_k))| = |1 - iu|^{M-n+1}.$$

Thus, the left hand side of (2.1) is integrable because $M \leqslant n - 3$. Therefore, by Fourier's inversion formula it follows from (2.1) that

$$\begin{aligned} &E(\exp(itg(X_0, \cdots, X_M))|\Sigma_0^{n-1}X_k = n) \cdot f_n(n) \\ &= (2\pi)^{-1}\int_{-\infty}^{\infty} e^{-iun} \cdot E(\exp(itg(X_0, \cdots, X_M) + iu\Sigma_0^{n-1}X_k))du. \end{aligned} \tag{2.2}$$

It is a well-known fact that

$$\mathfrak{L}((nS_0, \cdots, nS_{n-1})) = \mathfrak{L}((X_0, \cdots, X_{n-1})|\Sigma_0^{n-1}X_k = n).$$

Using this, the assertion follows from (2.2).

REMARK. The idea of using partial inversion for obtaining characteristic functions of conditional distributions goes back at least to Bartlett (1938).

In the following two corollaries the function h satisfies $h(x_1, \cdots, x_m) = h(x_1 + \cdots + x_m)$. Asymptotic normality in this special case was proved by Cressie (1976); he did not, however, derive the asymptotic variance.

COROLLARY 1. *When $n \to \infty$,*

$$\mathfrak{L}(n^{-\frac{1}{2}}(\Sigma_0^{n-1}(\ln(nS_k^{(m)}) - \psi(m)))) \to N(0, \sigma^2),$$

with

$$\sigma^2 = (2m^2 - 2m + 1)\psi'(m) - 2m + 1,$$

where ψ is the di-gamma- or psi-function with

$$\psi(m) = \Sigma_{j=1}^{m-1}j^{-1} - \gamma,$$

$$\psi'(m) = -\Sigma_{j=1}^{m-1}j^{-2} + \pi^2/6,$$

and γ is Euler's constant.

PROOF. As $T_0 = \Sigma_{j=0}^{m-1} X_j$ is $\Gamma(m, 1)$-distributed

$$E \ln T_0 = \int_0^\infty \ln(x) x^{m-1} e^{-x}/\Gamma(m)dx = \Gamma'(m)/\Gamma(m) = \psi(m).$$

Therefore, consider in the theorem the function $h(x) = \ln x - \psi(m)$. As

$$\operatorname{Var}(\ln T_0) = \int_0^\infty (\ln x - \psi(m))^2 x^{m-1} e^{-x}/\Gamma(m)dx = \psi'(m) = \Sigma_{j=m}^\infty j^{-2},$$

the assumptions of the theorem are satisfied. Using the method of Cressie (1976), pages 346–347, it is possible to obtain after some lengthy calculations that

$$\Sigma_{k=1}^{m-1} \operatorname{Cov}(\ln T_0, \ln T_k) = m(m-1)\psi'(m) - (m-1).$$

Hence the constant A in the theorem can be written

$$A = \psi'(m) + 2m(m-1)\psi'(m) - 2(m-1).$$

The number B is much easier to calculate. First

$$\begin{aligned}\operatorname{Cov}(\ln T_0, T_k) &= \operatorname{Cov}\left(\ln \Sigma_{j=0}^{m-1} X_j, \Sigma_{j=0}^{m-k-1} X_{j+k}\right)\\ &= E\left(\left(\ln \Sigma_0^{m-1} X_j\right)\Sigma_{j=0}^{m-k+1} X_{j+k}\right) - \psi(m)(m-k)\\ &= (m-k)\psi(m+1) - (m-k)\psi(m) = 1 - k/m\end{aligned}$$

and, therefore,

$$B = m^{-1}\left(1 + 2\Sigma_{k=1}^{m-1}(1 - k/m)\right) = 1.$$

Thus

$$A - B^2 = (2m^2 - 2m + 1)\psi'(m) - 2m + 1,$$

proving the assertion.

Using the theorem the asymptotic distribution of the central order statistic $S_{(np)}^{(m)}$, $0 < p < 1$, can be obtained as follows.

COROLLARY 2. *For $0 < p < 1$*

$$\mathcal{L}\left(n^{\frac{1}{2}}\left(nS_{(np)}^{(m)} - \mu_p\right)\right) \to N(0, \sigma^2), \qquad n \to \infty,$$

where

$$\begin{aligned} p &= F_m(\mu_p),\\ \sigma^2 &= \left(p(1-p) + 2\Sigma_{k=1}^{m-1} c_k - \left(\mu_p f_m(\mu_p)\right)^2\right)/\left(f_m(\mu_p)\right)^2,\\ c_k &= \int_0^{\mu_p} (F_k(u))^2 \cdot f_{m-k}(\mu_p - u)du - p^2,\end{aligned}$$

with $f_r(x) = x^{r-1}e^{-x}/\Gamma(r)$, the density of the $\Gamma(r, 1)$-distribution, and $F_r(x) = \int_0^x f_r(t)dt$, the corresponding distribution function.

PROOF. Using indicator functions it follows that

$$\begin{aligned}&P\left(nS_{(np)}^{(m)} \leqslant \mu_p + x/n^{\frac{1}{2}}\right) = P\left(\Sigma_{k=0}^{n-1} I\left(nS_k^{(m)} \leqslant \mu_p + x/n^{\frac{1}{2}}\right) \geqslant np\right)\\ &= P\left(n^{-\frac{1}{2}}\Sigma_{k=0}^{n-1}\left(I\left(nS_k^{(m)} \leqslant \mu_p\right) - p\right) + n^{-\frac{1}{2}}\Sigma_{k=0}^{n-1} I\left(\mu_p \leqslant nS_k^{(m)} \leqslant \mu_p + x/n^{\frac{1}{2}}\right) \geqslant 0\right).\end{aligned}$$

After some elementary calculations one obtains

$$E\left(n^{-\frac{1}{2}}\Sigma_{k=0}^{n-1}I\left(\mu_p \leqslant nS_k^{(m)} \leqslant \mu_p + x/n^{\frac{1}{2}}\right)\right) \to xf_m(\mu_p),$$

and

$$\operatorname{Var}\left(n^{-\frac{1}{2}}\Sigma_{k=0}^{n-1}I\left(\mu_p \leqslant nS_k^{(m)} \leqslant \mu_p + x/n^{\frac{1}{2}}\right)\right) \to 0,$$

when $n \to \infty$. With $h(x) = I(x \leqslant \mu_p) - p$ in the theorem

$$\mathcal{L}\left(n^{-\frac{1}{2}}\Sigma_{k=0}^{n-1}\left(I\left(nS_k^{(m)} \leqslant \mu_p\right) - p\right)\right) \to N(0, A - B^2).$$

Combining the results above gives

$$P\left(nS_{(np)}^{(m)} \leqslant \mu_p + x/n^{\frac{1}{2}}\right) \to 1 - \Phi\left(-xf_m(\mu_p)/(A - B^2)^{\frac{1}{2}}\right)$$
$$= \Phi\left(xf_m(\mu_p)/(A - B^2)^{\frac{1}{2}}\right),$$

where Φ is the standardized normal distribution function.

The constants A and B remain to be calculated.

As T_0 is $\Gamma(m, 1)$

$$\operatorname{Var} I(T_0 \leqslant \mu_p) = p \cdot (1 - p).$$

Let the independent random variables U, V, X be $\Gamma(k, 1)$, $\Gamma(m - k, 1)$, $\Gamma(1, 1)$. By elementary calculation it follows that

$$\operatorname{Cov}\left(I(T_0 < \mu_p), T_k\right) = \operatorname{Cov}\left(I(U + V < \mu_p), V\right)$$
$$= E\left(I(U + V < \mu_p)V\right) - p \cdot (n - k) = (m - k)P(U + V + X < \mu_p)$$
$$- p \cdot (n - k).$$

In a similar way

$$\operatorname{Cov}\left(I(T_0 < \mu_p), T_0\right) = m\left(P(U + V + X < \mu_p) - p\right).$$

Hence

$$B = m^{-1} \cdot \left(m + 2\Sigma_{k=1}^{m-1}(m - k)\right)\left(P(T_0 + X < \mu_p) - p\right)$$
$$= -m \cdot \mu_p^m e^{-\mu_p}/m! = -\mu_p f_m(\mu_p).$$

With the same notation one finds

$$\operatorname{Cov}\left(I(T_0 < \mu_p), I(T_k < \mu_p)\right) = \int_0^{\mu_p}\left(P(U < \mu_p - v)\right)^2 f_V(v)dv \cdot p^2$$
$$= \int_0^{\mu_p}\left(P(U < u)\right)^2 f_V(\mu_p - u)du - p^2 = c_k.$$

Thus

$$A - B^2 = p(1 - p) + 2\Sigma_{k=1}^{m-1}c_k - \left(\mu_p f_m(\mu_p)\right)^2,$$

proving the assertion.

Remark. The author has not been able to get a simple expression for $\Sigma_{k=1}^{m-1}c_k$. But the given expression is easy to compute numerically for given values of p and m.

REFERENCES

[1] BARTLETT, M. S. (1938). The characteristic function of a conditional statistic. *J. London Math. Soc*. **13** 62–67.

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[3] CRESSIE, N. (1976). On the logarithms of high-order spacings. *Biometrika* **63** 343–355.

[4] CRESSIE, N. (1977). The minimum of higher order gaps. *Austral. J. Statist.* **19** 132–143.

[5] CRESSIE, N. (1978). Power results for tests based on high-order gaps. *Biometrika* **65** 214–218.

[6] 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.

[7] PYKE, R. (1965). Spacings (with discussion). *J. Roy. Statist. Soc. Ser. B* **7** 395–449.

[8] PYKE, R. (1972). Spacings revisited. *Proc. Sixth Berkeley Symp. Math. Statist. Probability*. **1** 417–427.

[9] RAO, C. R. (1973) *Linear Statistical Inference and its Applications, 2nd ed.* Wiley, New York.

[10] RAO, J. S. (1976) Some tests based on arc-lengths for the circle. *Sankhyā Ser. B* **33** 1–10.

DEPARTMENT OF MATHEMATICS
THUNBERGSVÄGEN 3
S-752 38 UPPSALA
SWEDEN