AUSTRIAN JOURNAL OF STATISTICS
Volume 32 (2003), Number 1&2, 29–47

Exact Laws for Sums of Logarithms of Uniform Spacings

Paul Deheuvels and Gérard Derzko
Université Pierre et Marie Curie, Paris
Sanofi-Synthelabo, Montpellier

Abstract: In this paper we provide some simple expressions for the exact law of sums of logarithms of uniform spacings. We also obtain closed form expressions for the corresponding cumulants of arbitrary order, in terms of the Riemann zeta function. Our results follow from two representations of this statistic through sums of independent random variables, the latter being of independent interest. We conclude by showing that the distribution of sums of logarithms of uniform spacings are very closely approximated by a Gamma distribution even for very small values of the sample size. This renders easy the use of this statistic for goodness-of-fit tests of the uniformity assumption.

Keywords: Order Statistics, Spacings, Weak Laws of Large Numbers, Goodness-of-Fit Tests, Normal Distribution, Gamma Distribution, Central Limit Theorem, Zeta Function, Exact Distributions of Test Statistics.

1 Introduction and Main Result

Let $X = X_1, X_2, \ldots$ be i.i.d.r.v.'s with density $f(x)$, continuous and positive on $[0,1]$, and null elsewhere. Denote for $n \geq 2$ by $0 < X_{1,n-1} < \ldots < X_{n-1,n-1} < 1$ the order statistics of $X_1, \ldots, X_{n-1}$. Setting $X_{0,n-1} = 0$ and $X_{n,n-1} = 1$ for $n \geq 1$, denote by

$$D_{i,n} = X_{i,n-1} - X_{i-1,n-1} \quad \text{for} \quad i = 1, \ldots, n, \tag{1}$$

the spacings of order $n \geq 1$. Darling (1953) introduced the statistic

$$T_n = \sum_{i=1}^{n} \{ -\log(nD_{i,n}) \} = -\log \Big(n^n \prod_{i=1}^{n} D_{i,n} \Big), \tag{2}$$

for testing the uniformity assumption $(H.0)$: $\{f(x) = 1 : x \in [0,1]\}$, against the alternative. Blumenthal (1968), established the weak asymptotic behavior of T_n by showing that, as $n \to \infty$,

$$n^{-1/2} \Big\{ T_n - n\gamma - n\mathbb{E} \log f(X) \Big\} \xrightarrow{d} N\Big(0, \zeta(2) - 1 + \mathrm{Var}(\log f(X))\Big). \tag{3}$$

Here, we denote respectively by (see, e.g., 3:3:1-3:3:5, pp. 26-27 and 3:7:1, p. 28 in Spanier and Oldham, 1987)

$$\zeta(r) = \frac{1}{\Gamma(r)} \int_0^\infty \frac{t^{r-1}dt}{e^t - 1} = \sum_{j=1}^{\infty} \frac{1}{j^r} \quad \text{for} \quad r > 1, \tag{4}$$

$$\gamma = \int_0^\infty (-\log t) e^{-t} dt = \lim_{r \downarrow 1} \Big\{ \zeta(r) - \frac{1}{r-1} \Big\} \tag{5}$$

$$= \lim_{n \to \infty} \Big\{ \sum_{j=1}^{n-1} \frac{1}{j} - \log n \Big\} = 0.577\,215\,664\ldots,$$

Riemann's zeta function and Euler's constant (see, e.g., 9.73, p. 1080 in Gradshteyn and Ryzhik, 1982). For $m \geq 1$,

$$\zeta(2) = \frac{\pi^2}{6},\ \zeta(3) \approx 1.202\,056\,903,\ \zeta(4) = \frac{\pi^4}{90}, \ldots, \zeta(2m) = \frac{2^{2m-1}\pi^{2m}|B_{2m}|}{(2m)!}, \tag{6}$$

with $B_2 = \frac{1}{6}$, $B_4 = -\frac{1}{30}$, $B_6 = \frac{1}{42}, \ldots$ denoting the Bernoulli numbers conveniently defined through the formal expansion $\frac{t}{e^t-1} = \sum_{n=0}^{\infty} \frac{t^n}{n!} B_n$.

The problem of testing uniformity via statistics based upon spacings together with the investigation of related limit laws has been discussed by a number of authors, including Aly (1990), Aly et al. (1984), Blumenthal (1966a), Blumenthal (1966b), Borovikov (1988), Cox (1955), Deheuvels (1983), del Pino (1975), Ekstrom (1991), Hall (1982), Hall (1984), Koziol (1980), Le Cam (1958), L'Ecuyer (1997), Lévy (1939), Naus (1966), Pyke (1965), Pyke (1972), Rao (1976), Sethuraman and Rao (1969), Siegel (1978), Siegel (1979), Slud (1978) and Weiss (1956), Weiss (1957), among others. Beirlant and Horváth (1984), Beirlant et al. (1991a), Beirlant et al. (1991b), Beirlant et al. (1994), Beirlant et al. (1982), Einmahl and van Zuijlen (1988), and Shorack (1972), provided an empirical process approach to the derivation of limiting laws as in (3), Cressie (1976), Cressie (1978) discussed the power of tests based on sums of functions of spacings, Holst (1979, 1981), devised a general methodology to establish the asymptotic normality of such statistics. Further related papers are those of Czekala (1993), Guttorp and Lockhart (1989), Ranneby (1984), Shao and Hahn (1995), Shao and Jiménez (1998), and Swartz (1992).

For $0 < \alpha < 1$, denote by ν_α the upper α quantile of the normal $N(0,1)$ law. A simple consequence of (3), in combination with the observation that $\mathbb{E} \log f(X) \geq 0$, with equality iff $f(x) = 1$ for all $x \in [0,1]$, shows that the test rejecting $(H.0)$ when

$$T_n \geq n\gamma + n^{1/2}\nu_\alpha\{\tfrac{\pi^2}{6} - 1\}^{1/2}, \tag{7}$$

is asymptotically consistent for $(H.0)$ with size tending to α as $n \to \infty$. A finite sample application of this test is rendered difficult by the fact that the exact distribution of T_n under $(H.0)$ has remained unknown until present. Our main purpose is to give an answer to this question through the following theorem. We denote by $\Gamma(r) = \int_0^\infty t^{r-1}e^{-t}dt$ and $\beta(r,s) = \Gamma(r)\Gamma(s)/\Gamma(r+s)$ for $r, s > 0$, the Euler Gamma and Beta functions.

Theorem 1.1 *Under $(H.0)$, for each $n \geq 1$, it holds that*

$$\mathbb{E}\Big(\exp(sT_n)\Big) = \Gamma(1-s)^n \left\{\frac{n^{-ns}\Gamma(n)}{\Gamma(n(1-s))}\right\} \quad \textit{for} \quad s < 1. \tag{8}$$

The proof of Theorem 1.1 is postponed until Section 2, where this result will be shown to follow from a representation of T_n as a sum of $n-1$ independent random variables. Some useful consequences of Theorem 1.1 are given in Corollaries 1.1 and 1.2 below.

Corollary 1.1 *Under $(H.0)$, for $n \geq 1$, the cumulants $\kappa_1, \kappa_2, \ldots$ of T_n are given by*

$$\kappa_1 = \mathbb{E}(T_n) = n\Big\{\sum_{j=1}^{n-1} \frac{1}{j} - \log n\Big\}, \tag{9}$$

$$\kappa_k = n(k-1)!\Big\{\zeta(k) - n^{k-1}\Big(\zeta(k) - \sum_{j=1}^{n-1}\frac{1}{j^k}\Big)\Big\} \quad \textit{for} \quad k \geq 2. \tag{10}$$

and fulfill

$$\log \mathbb{E}\Big(\exp(sT_n)\Big) = \sum_{k=1}^{\infty} \frac{\kappa_k}{k!} s^k \quad \textit{for} \quad |s| < 1, \tag{11}$$

Moreover, as $n \to \infty$,

$$\kappa_1 - n\gamma \to -\frac{1}{2}, \tag{12}$$

$$\kappa_k - n(k-1)!\Big\{\zeta(k) - \frac{1}{k-1}\Big\} \to -\frac{(k-1)!}{2} \quad \textit{for} \quad k \geq 2, \tag{13}$$

and $\kappa_1, \kappa_2, \ldots$ fulfill the inequalities, for $n \geq 1$,

$$-1 \leq -\frac{n+1}{2n} < \kappa_1 - n\gamma < 0, \tag{14}$$

$$-(k-1)! < \kappa_k - n(k-1)!\Big\{\zeta(k) - \frac{1}{k-1}\Big\} < 0 \quad \textit{for} \quad k \geq 2. \tag{15}$$

Proof. By (1), $D_{1,1} = 1$ and $T_1 = 0$ has k-th cumulant $\kappa_k = 0$ for any $k \geq 1$. This is in agreement with (9)–(10) and (14)–(15) via the convention $\sum_{i=1}^{0}(\cdot) = 0$. For $n \geq 2$, we recall from Abramowitz and Stegun (1970), with $\psi(z) = \frac{d}{dz}\log\Gamma(z)$ denoting the digamma function, that

$$\log\Gamma(1-s) = \gamma s + \sum_{k=2}^{\infty} \frac{\zeta(k)}{k} s^k \quad \text{for} \quad |s| < 1, \tag{16}$$

$$\log\Gamma(n(1-s)) - \log\Gamma(n) = \sum_{k=1}^{\infty} (-1)^k \frac{\psi^{(k-1)}(n)}{k!}\, n^k s^k, \tag{17}$$

this last expansion holding in a neighborhood of $s = 0$. By 6.4.3, p. 260 of Abramowitz and Stegun (1970),

$$\psi(n) = \psi^{(0)}(n) = -\gamma + \sum_{j=1}^{n-1} \frac{1}{j}, \tag{18}$$

$$\psi^{(k-1)}(n) = (-1)^{k-1}(k-1)!\Big\{ -\zeta(k) + \sum_{j=1}^{n-1} \frac{1}{j^k}\Big\} \quad \text{for} \quad k \geq 2. \tag{19}$$

By replacing $\psi^{(k-1)}(n)$ in (17) by its values given in (18) and (19), we obtain that

$$\log\Gamma(n(1-s)) - \log\Gamma(n) = \Big\{\gamma - \sum_{j=1}^{n-1}\frac{1}{j}\Big\} sn + \sum_{k=2}^{\infty}\Big\{\zeta(k) - \sum_{j=1}^{n-1}\frac{1}{j^k}\Big\}\frac{n^k s^k}{k}. \tag{20}$$

In view of (8), we obtain readily (9) and (10) by combining (16) with (20).

To establish (14), we set $u_1 = 0$ and $u_m = \sum_{j=1}^{m-1} \frac{1}{j} - \log m$ for $m \geq 2$. For all $m \geq 1$,

$$0 < u_{m+1} - u_m = \frac{1}{m} - \log\Big(1 + \frac{1}{m}\Big) < \frac{1}{2m^2}, \tag{21}$$

whence, by combining (5) with (9) and (21),

$$0 < n\gamma - \kappa_1 = n(\gamma - u_n) = n\sum_{m=n}^{\infty}(u_{m+1} - u_m) < \frac{n}{2}\sum_{m=n}^{\infty}\frac{1}{m^2}$$
$$< \frac{n}{2}\Big\{\frac{1}{n^2} + \int_n^{\infty}\frac{dt}{t^2}\Big\} = \frac{n}{2}\Big\{\frac{1}{n^2} + \frac{1}{n}\Big\},$$

which is (14). We observe likewise that, for $k \geq 2$,

$$0 < \zeta(k) - \sum_{j=1}^{n-1}\frac{1}{j^k} = \sum_{k=n}^{\infty}\frac{1}{j^k} < \frac{1}{n^k} + \int_n^{\infty}\frac{dt}{t^k} = \frac{1}{n^k} + \frac{1}{(k-1)n^{k-1}},$$

whence, by (10),

$$-(k-1)! < \kappa_k - n(k-1)!\Big\{\zeta(k) - \frac{1}{k-1}\Big\}$$
$$= n(k-1)!\Big\{\frac{1}{k-1} - n^{k-1}\Big(\zeta(k) - \sum_{j=1}^{n-1}\frac{1}{j^k}\Big)\Big\} < 0,$$

which is (15). Finally, we conclude from (14) and (15) that

$$\sum_{k=1}^{\infty}\frac{|\kappa_k|}{k!}|s|^k \leq \sum_{k=1}^{\infty}\frac{|s|^k}{k} + n\sum_{k=1}^{\infty}\frac{1}{k}\Big\{\zeta(k) - \frac{1}{k-1}\Big\}|s|^k < \infty \quad \text{for} \quad |s| < 1,$$

which shows that the series in (11) converges for $|s| < 1$. □

Remark 1.1 *Recall the definition* $(r)_k = r(r+1)\dots(r+k-1)$ *of the Pochhammer symbol. Sharp asymptotic evaluations of the cumulants* $\kappa_1, \kappa_2, \dots$ *of* T_n *may be obtained by combining* (9)–(10) *with the expansion (see, e.g., 3:3:6, p. 27 in Spanier and Oldham, 1987),*

$$n^{r-1}\Big\{\zeta(r) - \frac{1}{r-1} - \sum_{j=1}^{n-1}\frac{1}{j^r}\Big\} = \frac{1}{2n} + \frac{r}{12n^2} + \frac{r(r+1)(r+2)}{6n^2} \tag{22}$$
$$+ \dots + (1 + o(1))\frac{(r)_{k-1}2^k(2^k-1)B_k}{2k!\, n^k} \text{ as } n \to \infty.$$

In (22) *we set, in agreement with* (5), $\zeta(r) - 1/(r-1) = \gamma$ *for* $r = 1$.

Remark 1.2 *Since* $\zeta(r) - 1/(r-1)$ *increases from* $\gamma = 0.577\dots$ *to* 1, *as* r *increases from* 1 *to* ∞, (15) *implies that the* k*-th cumulant* κ_k *of* T_n *fulfills, for* $n \geq 1$ *and* $k \geq 2$,

$$0 \leq \kappa_k < n(k-1)!\Big\{\zeta(k) - \frac{1}{k-1}\Big\} < n(k-1)! \tag{23}$$

Corollary 1.2 below gives a refinement of (3) when $(H.0)$ holds.

Corollary 1.2 *Under $(H.0)$, as $n \to \infty$,*

$$n^{-1/2}\{T_n - n\gamma\} \xrightarrow{d} N(0, \tfrac{\pi^2}{6} - 1). \tag{24}$$

Moreover, for each $k \geq 1$, the k-th moment of $n^{-1/2}\{T_n - n\gamma\}$ converges as $n \to \infty$ to the k-th moment of the $N(0, \frac{\pi^2}{6} - 1)$ law.

Proof. Let $\kappa_1^*, \kappa_2^*, \ldots$ denote the cumulants of $n^{-1/2}\{T_n - n\gamma\}$. By (9), (15) and (23), for $n \geq 1$,

$$\begin{aligned} -n^{-1/2} &\leq \kappa_1^* = n^{-1/2}\{\kappa_1 - n\gamma\} < 0, \\ -n^{-1} &< \kappa_2^* - \{\zeta(2) - 1\} = n^{-1}\{\kappa_2 - n\{\zeta(2) - 1\}\} < 0, \\ |\kappa_k^*| &= n^{-k/2}|\kappa_k| < n^{1-k/2}(k-1)! \quad \text{for} \quad k \geq 3. \end{aligned}$$

Thus, for any fixed $\delta > 0$, we have, ultimately in $n \to \infty$, uniformly over $|s| \leq \delta$,

$$\begin{aligned} &\Big| \log \mathbb{E}\Big(\exp\left(sn^{-1/2}\{T_n - n\gamma\}\right) - \tfrac{1}{2}s^2\{\zeta(2) - 1\} \Big| \\ &\leq n^{-1/2}\delta + \tfrac{1}{2}(n^{-1/2}\delta)^2 + n\sum_{k=3}^{\infty} \tfrac{1}{k}(n^{-1/2}\delta)^k = O(n^{-1/2}) \to 0, \end{aligned} \tag{25}$$

which readily implies (24) and completes the proof of the corollary. $\square$

The remainder of our paper is organized as follows. In Section 2, we derive a useful representation of T_n as a sum of independent random variables, which is of interest in and of itself. This representation, stated in Theorem 2.1 below, provides a key argument for the proof of Theorem 1.1, which is given at the end of the section. In Section 3 we give an alternate representation of T_n, stated in Theorem 3.1, by showing that $T_n = Q_n - R_n$, where Q_n is a partial sum of i.i.d.r.v.'s, $R_n = O_{\mathbb{P}}(1)$, and with R_n and T_n mutually independent. The limiting behavior of R_n is described in Theorem 3.2 of the same section. By combining Theorems 3.1 and 3.2, we get an alternate proof of Theorem 1.1, as well as a simple direct proof of Blumenthal's limit law (3) under $(H.0)$. In Section 4, we apply the moment calculations of Section 1 to show that an approximation of T_n by a Gamma distribution with suitably chosen parameters is extremely sharp even for values of n as small as $n = 3$. A theoretical justification for this fact is given in Theorem 4.1.

2 First Representation of T_n

In Theorem 2.1 below, we show that T_n has a simple expression as a sum of $n - 1$ independent random variables. This representation is then used to prove Theorem 1.1.

We start by some notation and facts. Under $(H.0)$, it is convenient to denote by $U_1 = X_1, U_2 = X_2, \ldots$ i.i.d. uniform $(0, 1)$ random variables. Set $U_{0,N} = 0$ and $U_{N+1,N} = 1$ for $N \geq 0$, and let $0 < U_{1,N} < \ldots < U_{N,N} < 1$ for $n \geq 1$ be the order statistics of $U_1, \ldots, U_N$ for $N \geq 1$. The *uniform spacings* of order $N \geq 1$ are then defined by

$$D_{i,N} = \mathcal{E}_{i,N-1} = U_{i,N-1} - U_{i-1,N-1} \quad \text{for} \quad 1 \leq i \leq N. \tag{26}$$

For each $n \geq 1$, introduce the random variables

$$Y_{i,n} = \frac{U_{i,n} - U_{i-1,n}}{U_{n,n}} = \frac{\mathcal{E}_{i,n}}{U_{n,n}} = \frac{D_{i,n+1}}{U_{n,n}} \quad \text{for} \quad 1 \leq i \leq n. \tag{27}$$

The following distributional identities hold for each $n \geq 1$ (see, e.g., Fact 2 and Remark 3.1 in §3), with $D_{i,n}$ being as in (26) (or equivalently, as in (1) under $(H.0)$).

$$\{Y_{i,n} : 1 \leq i \leq n\} \stackrel{d}{=} \{D_{i,n} : 1 \leq i \leq n\}. \tag{28}$$

As follows from (26), (27) and (28), we have the distributional identity

$$\mathcal{T}_n = \sum_{i=1}^{n} \{ -\log(nY_{i,n})\} \stackrel{d}{=} T_n = \sum_{i=1}^{n} \{ -\log(nD_{i,n})\}. \tag{29}$$

which allows to use equivalently $\mathcal{T}_n$ and T_n in the statement of our theorems. For convenience, in the present section, we will work with $\mathcal{T}_n$ rather than with T_n. Set, for $n \geq 1$,

$$V_{\ell,n} = \left\{ \frac{U_{\ell,n}}{U_{\ell+1,n}} \right\}^{\ell} \quad \text{for} \quad 1 \leq \ell \leq n+1. \tag{30}$$

We first recall (see, e.g., Sukhatme, 1937; Malmquist, 1950; David, 1981, pp. 20-21):

Fact 1 *For each $n \geq 1$, the random variables $\{V_{\ell,n} : 1 \leq \ell \leq n\}$ are independent and uniformly distributed on $(0,1)$.*

The main result of this section may now be stated in Theorem 2.1 below.

Theorem 2.1 *For each $n \geq 1$, the random variable $\mathcal{T}_n$ may be decomposed into the sum of $n-1$ independent random variables as follows.*

$$\mathcal{T}_n = \sum_{\ell=1}^{n-1} \left\{ -\log\left(nV_{\ell,n}\{1 - V_{\ell,n}^{1/\ell}\}\right)\right\}. \tag{31}$$

Proof. For $n = 1$, $\mathcal{T}_n = 0$ and (31) holds via $\sum_{\ell=1}^{0}(\cdot) = 0$. For $n \geq 2$, we write

$$U_{i,n} = \prod_{\ell=i}^{n} V_{\ell,n}^{1/\ell} \quad \text{for} \quad i = 1, \ldots, n \quad \text{and} \quad U_{n,n} = V_{n,n}^{1/n},$$

whence, by (28)

$$Y_{i,n} = \left\{ \prod_{\ell=i}^{n-1} V_{\ell,n}^{1/\ell} \right\} \left\{ 1 - V_{i-1,n}^{1/(i-1)} \right\} \quad \text{for} \quad 2 \leq i \leq n-1,$$

$$Y_{1,n} = \left\{ \prod_{\ell=1}^{n-1} V_{\ell,n}^{1/\ell} \right\} \quad \text{and} \quad Y_{n,n} = \left\{ 1 - V_{n-1,n}^{1/(n-1)} \right\}.$$

We then observe that

$$Y_{1,n}Y_{2,n} = \Big\{\prod_{\ell=1}^{n-1} V_{\ell,n}^{1/\ell}\Big\}\Big\{\prod_{\ell=2}^{n-1} V_{\ell,n}^{1/\ell}\Big\}\Big\{1 - V_{i-1,n}^{1/1}\Big\} = V_{1,n}\{1 - V_{1,n}\}\Big\{\prod_{\ell=2}^{n-1} V_{\ell,n}^{2/\ell}\Big\},$$

$$Y_{1,n}Y_{2,n}Y_{3,n} = V_{1,n}\{1 - V_{1,n}\}V_{2,n}\{1 - V_{2,n}^{1/2}\}\Big\{\prod_{\ell=3}^{n-1} V_{\ell,n}^{3/\ell}\Big\},$$

and, via a straightforward induction,

$$\prod_{i=1}^{n} Y_{i,n} = \prod_{i=1}^{n-1} V_{\ell,n}\{1 - V_{\ell,n}^{1/\ell}\}.$$

After taking logarithms and recalling (29), we conclude readily (31). □

Remark 2.1 *In view of* (31)*, we may give a direct proof of* (9) *by showing that*

$$\mathbb{E}(T_n) = \mathbb{E}(\mathcal{T}_n) = n\Big\{\sum_{j=1}^{n-1}\frac{1}{j} - \log n\Big\}.$$

For this, we observe by (31) *that* $\mathbb{E}(\mathcal{T}_n) = -n\log n + \sum_{\ell=1}^{n-1} I_\ell$*, where, for* $\ell \geq 1$*,*

$$\begin{aligned} I_\ell &= -\mathbb{E}\Big(\log V_{\ell,n} + \log(1 - V_{\ell,n}^{1/\ell})\Big) = -\int_0^1 (\log z)dz - \int_0^1 \log(1 - z^{1/\ell})dz \\ &= 1 - \ell\int_0^1 (1-t)^{\ell-1}(\log t)dt = 1 - \ell\sum_{k=0}^{\ell-1}\binom{\ell-1}{k}(-1)^k\int_0^1 t^k(\log t)dt \\ &= 1 - \sum_{m=1}^{\ell}\binom{\ell}{m}\frac{(-1)^m}{m} = 1 + \int_0^1\Big\{\frac{1-(1-z)^\ell}{z}\Big\}dz = 1 + \int_0^1\Big\{\frac{1-z^\ell}{1-z}\Big\}dz \\ &= 1 + \int_0^1\Big\{1 + z + \ldots + z^{\ell-1}\Big\}dz = 1 + \Big\{1 + \frac{1}{2} + \ldots + \frac{1}{\ell}\Big\}. \end{aligned}$$

Here, we have used Fact 1, together with the change of variable $t = 1 - z^{1/\ell}$*, and the identity* $\int_0^1 t^k(\log t)dt = -1/(k+1)^2$ *for* $k = 0, 1, \ldots$*. This, in turn, shows that*

$$\begin{aligned} \mathbb{E}(\mathcal{T}_n) + n\log n &= \sum_{\ell=1}^{n-1} I_\ell = n - 1 + \sum_{\ell=1}^{n-1}\Big\{1 + \frac{1}{2} + \ldots + \frac{1}{\ell}\Big\} \\ &= n\Big\{1 + \frac{1}{2} + \ldots + \frac{1}{n-1}\Big\}, \end{aligned}$$

which is in agreement with (9).

The next lemma is oriented towards proving Theorem 1.1.

Lemma 2.1 *For each* $\ell = 1, \ldots, n$*, we have the identity*

$$\mathbb{E}\exp\Big\{ - s\log\big(V_{\ell,n}\{1 - V_{\ell,n}^{1/\ell}\}\big)\Big\} = \ell\beta\Big(\ell(1-s), 1-s\Big) \tag{32}$$

$$= \ell\Gamma(1-s)\Big\{\frac{\Gamma(\ell(1-s))}{\Gamma((\ell+1)(1-s))}\Big\} \quad \textit{for} \quad s < 1. \tag{33}$$

Proof. For each $s < 1$, we may write, via the change of variable $z = t^\ell$,

$$\mathbb{E}\,\exp\Big\{-s\log\big(V_{\ell,n}\{1-V_{\ell,n}^{1/\ell}\}\big)\Big\} = \int_0^1 \exp\Big\{-s\log\big(z\{1-z^{1/\ell}\}\big)\Big\}dz$$
$$= \int_0^1 z^{-s}(1-z^{1/\ell})^{-s}dz = \ell\int_0^1 t^{\ell(1-s)-1}(1-t)^{-s}dt = \ell\beta\Big(\ell(1-s),1-s\Big),$$

which is (32), from where (33) is straightforward (see 43:13:12, p. 419 in Spanier and Oldham, 1987). □

Proof of Theorem 1.1. By combining (31) in Theorem 2.1 with (33) in Lemma 2.1, we see that, for each $n \geq 1$ and $s < 1$,

$$\mathbb{E}\,\exp(s\mathcal{T}_n) = \exp\big(-sn\log n\big)\prod_{\ell=1}^{n-1}\mathbb{E}\,\exp\Big\{-s\log\big(V_{\ell,n}\{1-V_{\ell,n}^{1/\ell}\}\big)\Big\}$$
$$= n^{-ns}\prod_{\ell=1}^{n-1}\ell\Gamma(1-s)\Big\{\frac{\Gamma(\ell(1-s))}{\Gamma((\ell+1)(1-s))}\Big\} = \frac{n^{-ns}\Gamma(n)\Gamma(1-s)^n}{\Gamma(n(1-s))},$$

which, in view of the distributional equality $\mathcal{T} \overset{d}{=} T_n$ in (29), yields (8). □

3 Second Representation of T_n

Recalling from (29) that $\mathcal{T}_n \overset{d}{=} T_n$, below we will work with T_n rather than with $\mathcal{T}_n$. We recall the following fact (see, e.g., Pyke, 1965; Shorack and Wellner, 1986, Prop. 8.2.1).

Fact 2 *Let $\omega = \omega_1, \omega_2, \ldots$ be i.i.d. exponentially distributed random variables with $\mathbb{P}(\omega > t) = e^{-t}$ for $t \geq 0$, and set $S_k = \omega_1 + \ldots + \omega_k$ for $k \geq 1$. Then, for each $n \geq 1$, we have the distributional identity*

$$\Big\{D_{i,n} : 1 \leq i \leq n\Big\} \overset{d}{=} \Big\{\frac{\omega_i}{S_n} : 1 \leq i \leq n\Big\}. \tag{34}$$

Remark 3.1 *The following distributional identity is a straightforward consequence of Fact 2. For each specified $N \geq n \geq 1$ and $1 \leq j_1 < \ldots < j_n \leq N$, we have*

$$\Big\{D_{i,n} : 1 \leq i \leq n\Big\} \overset{d}{=} \Big\{\frac{D_{j_i,N}}{\sum_{m=1}^n D_{j_m,N}} : 1 \leq i \leq n\Big\}. \tag{35}$$

In particular, (35) reduces to (28) when $N = n+1$ and $j_m = m$ for $m = 1,\ldots,n$.

We will need a refinement of Fact 2 stated in the following lemma.

Lemma 3.1 *Let S_n^* denote a random variable, independent of $\{D_{i,n} : 1 \leq i \leq n\}$, and with a $\Gamma(n)$ distribution. Then, the random variables $\omega_i^* = D_{i,n}S_n^*$ for $i = 1,\ldots,n$, are mutually independent and exponentially distributed with unit mean.*

Proof. In view of the notation (27), set $\mathcal{E}_{i,n} = U_{i,n} - U_{i-1,n}$ for $i = 1, \ldots, n+1$. It is well-known (refer to David (1981), p. 99) that $\{\mathcal{E}_{1,n}, \ldots, \mathcal{E}_{n,n}\}$ is uniformly distributed over $A_n := \{(u_1, \ldots, u_n) \in \mathbb{R}^n : \sum_{j=1}^n u_j \leq 1, u_i \geq 0, i = 1, \ldots, n\}$. Let S_{n+1}^* be independent of $\{\mathcal{E}_{1,n}, \ldots, \mathcal{E}_{n,n}\}$, and with a $\Gamma(n+1)$ distribution. The joint density of $(\mathcal{E}_{1,n}, \ldots, \mathcal{E}_{n,n}, S_{n+1}^*)$ is then $f(u_1, \ldots, u_n, s) = s^{n+1}e^{-s}$ on $A_n \times (0, \infty)$. By a straightforward change of variables, it follows that the joint density $g(x_1, \ldots, x_{n+1})$ of $(\mathcal{E}_{1,n}S_{n+1}^*, \ldots, \mathcal{E}_{n,n}S_{n+1}^*, (1 - \sum_{j=1}^n \mathcal{E}_{j,n})S_{n+1}^*) = (\mathcal{E}_{1,n}S_{n+1}^*, \ldots, \mathcal{E}_{n+1,n+1}S_{n+1}^*)$ is given by $g(x_1, \ldots, x_{n+1}) = \exp(-\sum_{i=1}^{n+1} x_i)$ for $x_1 \geq 0, \ldots, x_{n+1} \geq 0$. We so obtain the version of the lemma obtained with the formal change of n into $n+1$. $\square$

In the sequel, we may and do assume that the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ on which sits $\{U_n : n \geq 1\}$ is enlarged by products to carry a sequence $\{\omega_n^* : n \geq 1\}$, independent of $\{U_n : n \geq 1\}$, and composed of independent and (unit mean) exponentially distributed random variables. Via Lemma 3.1, for each $n \geq 1$, we set $S_n^* = \sum_{i=1}^n \omega_i^*$, and define a new set of random variables $\{\omega_{i,n} : 1 \leq i \leq n\}$ by setting

$$\omega_{i,n} = D_{i,n}S_n^* \quad \text{for} \quad i = 1, \ldots, n. \tag{36}$$

Since (26) implies that $\sum_{i=1}^n D_{i,n} = 1$, we observe from (36) that, for each $n \geq 1$,

$$S_n := \sum_{i=1}^n \omega_{i,n} = S_n^* \sum_{i=1}^n D_{i,n} = S_n^*. \tag{37}$$

This, in turn, shows that we may obtain a version of (34) by writing

$$\{D_{i,n} : 1 \leq i \leq n\} \overset{d}{=} \{\omega_{i,n}/S_n : 1 \leq i \leq n\}, \tag{38}$$

with $S_n = \sum_{i=1}^n \omega_{i,n}$ *independent* of $\{D_{i,n} : 1 \leq i \leq n\}$. Set now, for each $n \geq 1$,

$$Q_n = \sum_{i=1}^n \Big\{\omega_{i,n} - 1 - \log \omega_{i,n}\Big\}, \tag{39}$$

$$R_n = n\Big\{\frac{S_n}{n} - 1 - \log\Big(\frac{S_n}{n}\Big)\Big\}. \tag{40}$$

Given the above preliminaries, we may now state the main result of the present section.

Theorem 3.1 *For each $n \geq 1$, T_n and R_n are independent and such that*

$$Q_n = T_n + R_n. \tag{41}$$

Moreover, the moment-generating functions of Q_n and R_n are given by

$$\mathbb{E}(\exp(sQ_n)) = \Gamma(1-s)^n \Big\{\frac{e^{-ns}}{(1-s)^{n(1-s)}}\Big\} \quad \textit{for} \quad s < 1, \tag{42}$$

$$\mathbb{E}(\exp(sR_n)) = \Big\{\frac{\Gamma(n(1-s))}{\Gamma(n)}\Big\}\Big\{\frac{n^{ns}e^{-ns}}{(1-s)^{n(1-s)}}\Big\} \quad \textit{for} \quad s < 1. \tag{43}$$

Proof. In view of (36)–(38) and Lemma 3.1, we obtain (41) by combining (29) with (39)–(40). Recalling that, for $r > 0$, $\int_0^\infty z^{r-1}e^{-\lambda z}dz = \lambda^{-r}\Gamma(r)$, to establish (43), we use the fact that S_n follows a $\Gamma(n)$ distribution to write that, for $n \geq 1$ and $s < 1$,

$$\mathbb{E}(\exp(sR_n)) = \int_0^\infty \exp\Big(ns\Big\{\frac{z}{n} - 1 - \log\Big(\frac{z}{n}\Big)\Big\}\Big)\frac{z^{n-1}}{\Gamma(n)}e^{-z}dz$$
$$= \frac{n^{ns}e^{-ns}}{\Gamma(n)}\int_0^\infty z^{n(1-s)-1}e^{-z(1-s)}dz = \Big\{\frac{\Gamma(n(1-s))}{\Gamma(n)}\Big\}\Big\{\frac{n^{ns}e^{-ns}}{(1-s)^{n(1-s)}}\Big\}.$$

Setting $n = 1$ in this last expression, we see that, whenever ω is exponentially distributed,

$$\mathbb{E}\Big(\exp(s\{\omega - 1 - \log\omega\})\Big) = \Gamma(1-s)\Big\{\frac{e^{-s}}{(1-s)^{1-s}}\Big\} \quad \text{for} \quad s < 1. \tag{44}$$

Since $\mathbb{E}\,\exp(sQ_n) = \{\mathbb{E}\,\exp(s\{\omega - 1 - \log\omega\})\}^n$, we infer readily (43) from (44).□

Remark 3.2 *It is easy to check that* (11) *and* (42)–(43) *are in agreement, since*

$$\mathbb{E}(\exp(sQ_n)) = \mathbb{E}(\exp(s\mathcal{T}_n))\mathbb{E}(\exp(sR_n)). \tag{45}$$

The next corollaries of Theorem 3.1 evaluate the cumulants of R_n and Q_n. Their proofs are very similar to that of Corollary 1.1, and are therefore omitted.

Corollary 3.1 *For each* $n \geq 1$, *the cumulants* $\kappa_1^q, \kappa_2^q, \dots$ *of* Q_n *are given by*

$$\kappa_1^q = \mathbb{E}(Q_n) = n\gamma, \tag{46}$$
$$\kappa_k^q = n(k-1)!\Big\{\zeta(k) - \frac{1}{k-1}\Big\} \quad \textit{for} \quad k \geq 2. \tag{47}$$

Corollary 3.2 *For each* $n \geq 1$, *the cumulants* $\kappa_1^r, \kappa_2^r, \dots$ *of* R_n *are given by*

$$\kappa_1^r = \mathbb{E}(R_n) = n\Big\{\gamma - \sum_{j=1}^{n-1}\frac{1}{j} + \log n\Big\}, \tag{48}$$
$$\kappa_k^r = n(k-1)!\Big\{n^{k-1}\Big(\zeta(k) - \sum_{j=1}^{n-1}\frac{1}{j^k}\Big) - \frac{1}{k-1}\Big\} \quad \textit{for} \quad k \geq 2. \tag{49}$$

Moreover, for each $k \geq 1$, *we have, as* $n \to \infty$,

$$\kappa_k^r \to \frac{(k-1)!}{2}. \tag{50}$$

In addition, $\kappa_1^r, \kappa_2^r, \dots$ *fulfill the inequalities, for* $n \geq 1$,

$$0 < \kappa_1^r \leq \frac{n+1}{2n} \leq 1, \tag{51}$$
$$0 < \kappa_k^r < (k-1)! \quad \textit{for} \quad k \geq 2. \tag{52}$$

We now show that, in the decomposition $Q_n = T_n + R_n$ in (41), $R_n = O_{\mathbb{P}}(1)$, whence

$$T_n = Q_n + O_{\mathbb{P}}(1) \quad \text{as} \quad n \to \infty. \tag{53}$$

This, together with the explicit limiting behavior of R_n is captured in the next theorem.

Theorem 3.2 *As $n \to \infty$, we have the weak convergence*

$$2R_n = 2n\Big\{\frac{S_n}{n} - 1 - \log\Big(\frac{S_n}{n}\Big)\Big\} \xrightarrow{d} \chi_1^2, \tag{54}$$

and, for each $k \geq 1$, the k-th moment of $2R_n$ converges, as $n \to \infty$, to the k-th moment of a χ_1^2 random variable.

Proof. By the central limit theorem, as $n \to \infty$, $n^{-1/2}(S_n - n) \xrightarrow{d} N(0,1)$, so that $n^{-1}S_n - 1 = O_{\mathbb{P}}(n^{-1/2})$, and hence

$$\log\Big(\frac{S_n}{n}\Big) = \frac{S_n}{n} - 1 - \frac{1}{2}\Big(\frac{S_n}{n} - 1\Big)^2 + O_{\mathbb{P}}\Big(n^{-3/2}\Big).$$

Therefore, as $n \to \infty$,

$$2R_n = 2n\Big\{\frac{S_n}{n} - 1 - \log\Big(\frac{S_n}{n}\Big)\Big\} = n\Big(\frac{S_n}{n} - 1\Big)^2 + O_{\mathbb{P}}\Big(n^{-1/2}\Big) \xrightarrow{d} \chi_1^2, \tag{55}$$

which is (54). The cumulants $\kappa_1^\chi, \kappa_2^\chi, \ldots, \kappa_k^\chi = 2^{k-1}(k-1)!, \ldots$ of a χ_1^2 random variable are readily obtained from the expansion

$$-\frac{1}{2}\log(1-2s) = \sum_{k=1}^{\infty} \frac{2^k s^k}{2k} = \sum_{k=1}^{\infty} \frac{\kappa_k^\chi s^k}{k!} \quad \text{for} \quad s < 1. \tag{56}$$

The proof of the theorem is completed by (50), which implies that the k-th cumulant $2^k\kappa_k^r$ of $2R_n$ converges to $\kappa_k^\chi = 2^{k-1}(k-1)!$, as $n \to \infty$. □

Remark 3.3 *By combining (53) with an application of the central limit theorem to Q_n, one obtains an immediate proof of the weak convergence (3) under $(H.0)$, namely*

$$n^{-1/2}\{T_n - n\gamma\} = n^{-1/2}\{Q_n - n\gamma\} + O_{\mathbb{P}}(n^{-1/2}) \xrightarrow{d} N(0, \tfrac{\pi^2}{6} - 1).$$

4 Approximation of T_n by a Gamma Distribution

In this section, we consider the problem of finding simple and sharp numerical approximations of $F_n(x) := \mathbb{P}(T_n \leq x)$ under $(H.0)$, allowing to derive non-asymptotic versions of the test procedure (7) for small values of the sample size n. The easiest case is when $n = 2$, the exact distribution of T_n being then provided by the proposition below.

Proposition 4.1 *We have*

$$\mathbb{P}(T_2 \leq x) = \sqrt{1 - 2e^{-x}} \quad \textit{for} \quad x \geq \log 2. \tag{57}$$

Proof. Recalling (31), let $V = V_{1,2}$ denote a uniform $(0,1)$ r.v. An application of Theorem 2.1 for $n = 2$ shows that

$$\mathbb{P}(T_2 \leq x) = \mathbb{P}\Big(\log\left(2V(1-V)\right) \geq -x \Big) = \mathbb{P}\Big(V(1-V) \geq \tfrac{1}{2} e^{-x} \Big). \tag{58}$$

Since, for $0 \leq z \leq \frac{1}{4}$, $v(1-v) \geq z \quad\Leftrightarrow\quad |v - \frac{1}{2}| \leq \frac{1}{2}\sqrt{1-4z}$, we infer readily (57) from the uniformity of V, in combination with (58), taken with $z = \frac{1}{2}e^{-x}$. □

From now on, in view of Proposition 4.1, we will concentrate on the study of the d.f. $F_n(x) = \mathbb{P}(T_n \leq x)$ of T_n for $n \geq 3$. As will become obvious from Table 2 in the sequel, the approximation of $F_n(x)$ by $\Phi_n(x) := \Phi((x - \mathbb{E}\,T_n)/\sqrt{\mathrm{Var}(T_n)}\,)$, (with $\Phi(t) = (2\pi)^{-1/2}\int_{-\infty}^{t} e^{-s^2/2} ds$ denoting the normal $N(0,1)$ d.f.) is too rough for small values of n, and some other method is needed. As a main result of the present section, we will show that (under $(H.0)$ and after centering) the distribution of T_n is very closely approximated by a $\Gamma(r,\lambda)$ distribution for appropriate choices of $r = r_n$ and $\lambda = \lambda_n$ specified below. In practice, this replacement turns out to generate a uniform error less than 2.7% for $n \geq 3$ and less than 1.0% for $n \geq 4$. Moreover, for most applications of interest, the approximation error can be neglected if $n \geq 5$. Since the practical computation of the distribution of a Gamma random variable is routine, this observation allows us to implement easily, for arbitrary values of n, the uniformity test (rejecting $(H.0)$ when T_n exceeds a critical level) described in §1.

Recall that a random variable Z follows a $\Gamma(r,\lambda)$ distribution iff it has density $g(\cdot)$ and moment-generating function $\psi(\cdot)$ given respectively by

$$g(z) = \frac{\lambda^r}{\Gamma(r)} z^{r-1} e^{-\lambda z} \text{ for } z > 0, \quad \text{and} \quad \psi(s) = \mathbb{E}(e^{sZ}) = (1 - s/\lambda)^{-r} \text{ for } s < 1.$$

Setting $\mu = \mathbb{E}(Z)$ and $M_k = \mathbb{E}((Z-\mu)^k)$ for $k = 1, 2, \ldots$, we denote by $\beta_1^g = M_3^2/M_2^3$ (resp. $\beta_2^g = M_4/M_2^2$) the skewness (resp. kurtosis) coefficient of Z. For $r > 0$ and $\lambda > 0$,

$$\mu = \frac{r}{\lambda}, \quad M_2 = \frac{r}{\lambda^2}, \quad \beta_1^g = \frac{4}{r} \quad \text{and} \quad \beta_2^g = 3 + \frac{6}{r} = 3 + \frac{3}{2}\beta_1^g. \tag{59}$$

We choose $r = r_n > 0$, $\lambda = \lambda_n > 0$ and c_n in such a way that the moments of order 1, 2 and 3 of $Z = Z_n$ and $T_n - c_n$ coincide. In view of (6), (12)–(13) and (59), we obtain therefore, with $\kappa_1, \kappa_2, \ldots$ denoting as in (9)–(10) the cumulants of T_n, that

$$r_n := \frac{4\kappa_2^3}{\kappa_3^2} = \frac{n\Big\{\zeta(2) - n\Big(\zeta(2) - \sum_{j=1}^{n-1}\frac{1}{j^2}\Big)\Big\}^3}{\Big\{\zeta(3) - n^2\Big(\zeta(3) - \sum_{j=1}^{n-1}\frac{1}{j^3}\Big)\Big\}^2} \tag{60}$$

$$= \frac{(1+o(1))n\Big\{\zeta(2) - 1\Big\}^3}{\Big\{\zeta(3) - \frac{1}{2}\Big\}^2} \approx (0.544\,253) \times n \quad \text{as} \quad n \to \infty,$$

$$\lambda_n := \frac{2\kappa_2}{\kappa_3} = \frac{\zeta(2) - n\Big(\zeta(2) - \sum_{j=1}^{n-1}\frac{1}{j^2}\Big)}{\zeta(3) - n^2\Big(\zeta(3) - \sum_{j=1}^{n-1}\frac{1}{j^3}\Big)} \tag{61}$$

$$
\begin{aligned}
&= \frac{(1+o(1))\Big\{\zeta(2)-1\Big\}}{\zeta(3)-\frac{1}{2}} \approx 0.918\,635 \quad \text{as} \quad n\to\infty,\\
\mu_n &:= \mathbb{E}(Z_n) = \frac{r_n}{\lambda_n} = \frac{2\kappa_2^2}{\kappa_3} = \frac{n\Big\{\zeta(2)-n\Big(\zeta(2)-\sum_{j=1}^{n-1}\frac{1}{j^2}\Big)\Big\}^2}{\zeta(3)-n^2\Big(\zeta(3)-\sum_{j=1}^{n-1}\frac{1}{j^3}\Big)} \qquad (62)\\
&= \frac{(1+o(1))n\Big\{\zeta(2)-1\Big\}^2}{\zeta(3)-\frac{1}{2}} \approx (0.592\,459)\times n \quad \text{as} \quad n\to\infty,\\
c_n &:= \mathbb{E}(T_n-Z_n) = \kappa_1 - \frac{2\kappa_2^2}{\kappa_3} = n\Big\{\sum_{j=1}^{n-1}\frac{1}{j} - \log n\Big\} \qquad (63)\\
&\quad - \frac{n\Big\{\zeta(2)-n\Big(\zeta(2)-\sum_{j=1}^{n-1}\frac{1}{j^2}\Big)\Big\}^2}{\zeta(3)-n^2\Big(\zeta(3)-\sum_{j=1}^{n-1}\frac{1}{j^3}\Big)}\\
&= (1+o(1))n\Big\{\gamma - \frac{\{\zeta(2)-1\}^2}{\zeta(3)-\frac{1}{2}}\Big\} \approx -(0.015\,243)\times n \quad \text{as} \quad n\to\infty.
\end{aligned}
$$

Below, we follow the notation (59) by letting $\beta_2^g = 3+6/r_n$ denote the kurtosis coefficient of Z_n. In the same spirit, we let $\beta_2 = 3+(\kappa_4/\kappa_2^2)$ denote the kurtosis coefficient of T_n, with κ_k being as in (10) for $k\geq 2$. We will make use later on of the fact that

$$
\begin{aligned}
\beta_2-\beta_2^g &= \kappa_4 - \frac{6}{r_n} = (1+o(1))\frac{6}{n}\left\{\frac{\zeta(4)-\frac{1}{3}}{\{\zeta(2)-1\}^2} - \frac{\{\zeta(3)-\frac{1}{2}\}^2}{\{\zeta(2)-1\}^3}\right\}\\
&= -(1+o(1))n^{-1}(0.219\,973\,574) \quad \text{as} \quad n\to\infty.
\end{aligned}
$$

The choices of r_n, λ_n and $c_n = \mathbb{E}(T_n - Z_n)$ provided by (60)–(61) and (63) ensure that the first three moments of T_n and $Z_n + c_n$ coincide. The evaluation of the asymptotic sharpness of this approximation is achieved in the next theorem. Recalling that $F_n(x) = \mathbb{P}(T_n \leq x)$, we set $F_n^g(x) = \mathbb{P}(Z_n \leq x - c_n)$ and denote as above, respectively, by β_2 and β_2^g the kurtosis coefficients of T_n and Z_n.

Theorem 4.1 *We have, as $n\to\infty$,*

$$
\begin{aligned}
\sup_{x\in\mathbb{R}}|F_n(x)-F_n^g(x)| &= \frac{1+o(1)}{24}|\beta_2-\beta_2^g|\sup_{z\in\mathbb{R}}|H_3(z)|\varphi(z) \qquad (64)\\
&= \frac{1+o(1)}{4n}\left\{\frac{\zeta(4)-\frac{1}{3}}{\{\zeta(2)-1\}^2} - \frac{\{\zeta(3)-\frac{1}{2}\}^2}{\{\zeta(2)-1\}^3}\right\}\left\{\sup_{z\in\mathbb{R}}|z^3-3z|\frac{e^{-z^2/2}}{\sqrt{2\pi}}\right\} \approx \frac{0.005\,046}{n}.
\end{aligned}
$$

Proof. Set $Z_n^r = (Z_n - \mathbb{E}(Z_n))/\sqrt{\mathrm{Var}(Z_n)}$ and $T_n^r = (T_n - \mathbb{E}(T_n))/\sqrt{\mathrm{Var}(T_n)}$, $F_n^{g,r}(z) = \mathbb{P}(Z_n^r \leq z)$ and $F_n^r(z) = \mathbb{P}(T_n^r \leq z)$. Set, as usual $\varphi(t) = (2\pi)^{-1/2}e^{-t^2/2}$ and $\Phi(x) = \int_{-\infty}^x \varphi(t)dt$ for the normal $N(0,1)$ density and d.f. Denote by $H_2(x) = x^2-1$, $H_3(x) = x^3-3x$, $H_5(x) = x^5-10x^3+3x, \ldots$ (see p. 137 in Petrov, 1975) the Chebyshev-Hermite polynomials of order $2,3,5,\ldots$. Fix any $0 < M < \infty$. Since Z_n follows a

$\Gamma(r_n, \lambda_n)$ distribution with skewness and kurtosis coefficients β_1^g, β_2^g such that

$$\sqrt{\beta_1^g} = \frac{2}{\sqrt{r_n}} \quad \text{and} \quad \beta_2^g - 3 = \frac{6}{r_n} = (1+o(1)) \frac{6\Big\{\zeta(3) - \frac{1}{2}\Big\}^2}{n\Big\{\zeta(2) - 1\Big\}^3}, \tag{65}$$

it follows from Ch. VI in Petrov (1975) that, as $n \to \infty$, uniformly over $|z| \le M$

$$\begin{aligned} F_n^{g,r}(z) - \Phi(z) = -\varphi(z)\Big\{ & \frac{\sqrt{\beta_1^g}}{6} H_2(z) + \frac{\beta_1^g}{72} H_5(z) \\ & + \frac{\beta_2^g - 3}{24} H_3(z) + O(n^{-3/2})\Big\}. \end{aligned} \tag{66}$$

Likewise, making use of Theorem 2.1, which shows that T_n may be represented as a sum of independent random variables, we may write that, uniformly over $|z| \le M$,

$$\begin{aligned} F_n^{r}(z) - \Phi(z) = -\varphi(z)\Big\{ & \frac{\sqrt{\beta_1}}{6} H_2(z) + \frac{\beta_1}{72} H_5(z) \\ & + \frac{\beta_2 - 3}{24} H_3(z) + O(n^{-3/2})\Big\}, \end{aligned} \tag{67}$$

where (recall (9)–(10) and (12)–(13)) $\beta_2 - 3$ is given by

$$\beta_2 - 3 = \frac{\kappa_4}{\kappa_2^2} = (1+o(1)) \frac{6\Big\{\zeta(4) - \frac{1}{3}\Big\}}{n\Big\{\zeta(2) - 1\Big\}^2}. \tag{68}$$

Recalling (6), it follows readily from (66)–(67) that, uniformly over $|z| \le M$,

$$\begin{aligned} F_n^{r}(z) - F_n^{g,r}(z) &= -\frac{(1+o(1))}{n}\, \varphi(z) H_3(z) \\ &\quad \times \frac{\Big\{\zeta(4) - \frac{1}{3}\Big\}\Big\{\zeta(2) - 1\Big\} - \Big\{\zeta(3) - \frac{1}{2}\Big\}^2}{4\Big\{\zeta(2) - 1\Big\}^3} \\ &\approx -\frac{(1+o(1))}{n}\, \varphi(z) H_3(z) \Big\{\frac{0.219\,973\,574}{24}\Big\}. \end{aligned} \tag{69}$$

Finally, we observe that the supremum of $|\phi(z) H_3(z)|$ over $z \in \mathbb{R}$ is reached for $z_0 = \sqrt{3 - \sqrt{6}}$ and equals $0.550\,587\,939\ldots$. The conclusion (64) is now straightforward. □

To appreciate the relevance of the limiting bound (64) for small values of n, it is natural to estimate $F_n(x) = \mathbb{P}(T_n \le x)$ by the empirical distribution function $F_{n,N}(x)$, obtained from a simulated sample of size N of T_n for some suitably large value of N. It is easy to evaluate numerically $F_n^g(x)$, and we so estimate $\Delta_n = \sup_x |F_n(x) - F_n^g(x)|$ by $\widehat{\Delta}_n = \sup_x |F_{n,N}(x) - F_n^g(x)|$. The convergence of $\widehat{\Delta}_n$ to Δ_n is achieved with rate $O(N^{-1/2})$. Making use of the rough inequality

$$\mathbb{P}\Big(\sup_{x \in \mathbb{R}} |F_{n,N}(x) - F_n(x)| \ge 3/\sqrt{N}\Big) \le 10\%,$$

we see that the (rather huge and beyond our present reach) sample size $N = 900\,000\,000$ would be required to achieve (with 90% confidence) a precision of $0.000\,1$ or less in the estimation of Δ_n by $\widehat{\Delta}_n$. Below, we provide estimations obtained with the more tractable $N = 2\,000\,000$. The figures in Table 1 are therefore given only with a guaranteed precision of $\pm\delta$, where $\delta = 3/\sqrt{2\,000\,000} \approx 0.0021$ (with 90% confidence).

Table 1: $\widehat{\Delta}_n = \sup_x |F_{n,N}(x) - F_n^g(x)|$

n	$\widehat{\Delta}_n$	$\pm\delta$	$n\widehat{\Delta}_n$	$\pm n\delta$
3	0.02463	±0.002	0.07389	±0.006
4	0.00734	±0.002	0.02936	±0.008
5	0.00360	±0.002	0.01800	±0.010
8	0.00122	±0.002	0.00976	±0.016
10	0.00102	±0.002	0.01020	±0.021
15	0.00064	±0.002	0.00960	±0.032
20	0.00042	±0.002	0.00840	±0.042

In spite of the lack of precision induced by our choice of N, we see that the estimations in Table 1 come close to the order of the asymptotic rate in (64), even for very small values of n. In particular, Table 1 gives numerical evidence to support the conjecture that the replacement of F_n by F_n^g should yield an error less than 1% (resp. 0.5%) for $n \geq 4$ (resp. $n \geq 8$). We leave the question of checking the validity of this numerical conjecture for future research. In particular, the lack of precision on $\widehat{\Delta}_n$ obtained from our simulations, performed with $N = 2\,000\,000$ only, does not allow us to observe properly from the values in Table 1 the convergence of $n\widehat{\Delta}_n$ to the constant $0.005\,046\ldots$ (following from (64)). This is in contrast with the behavior of $n^{1/2}\widetilde{\Delta}_n$ discussed later on in Remark 4.1, where the simulated sample size of $N = 2\,000\,000$ turns out to be sufficient, to give a numerical confirmation of Proposition 4.2 below.

In Table 2 below, we give similar figures as in Table 1, with F_n^g being replaced by the normal approximant $\Phi_n(x) = \Phi((x - \mathbb{E}\,T_n)/\sqrt{\mathrm{Var}(T_n)})$.

Table 2: $\widetilde{\Delta}_n = \sup_x |F_{n,N}(x) - \Phi_n(x)|$

n	$\widetilde{\Delta}_n$	$\pm\delta$	$n^{1/2}\widetilde{\Delta}_n$	$\pm n^{1/2}\delta$
3	0.15270	±0.002	0.26448	0.00346
4	0.10630	±0.002	0.21260	0.00400
5	0.09082	±0.002	0.20308	0.00447
8	0.06816	±0.002	0.19278	0.00565
10	0.06047	±0.002	0.19122	0.00632
15	0.04822	±0.002	0.18676	0.00775
20	0.04116	±0.002	0.18407	0.00894

Making use of (67), it is not too difficult to prove the proposition

Proposition 4.2 *We have, as* $n \to \infty$*,*

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi_n(x)| = \frac{1+o(1)}{6}\Big\{\frac{2}{\sqrt{r_n}}\Big\} \sup_{z\in\mathbb{R}} |H_2(z)|\varphi(z)$$
$$= \frac{1+o(1)}{3\sqrt{2\pi}\,\sqrt{n}} \left\{\frac{\zeta(3)-\frac{1}{2}}{\{\zeta(2)-1\}^{3/2}}\right\} \approx \frac{0.180\,255}{\sqrt{n}}. \tag{70}$$

Proof. The supremum $\sup_{z\in\mathbb{R}} |H_2(z)|\phi(z) = 1/\sqrt{2\pi}$ is reached for $z = 0$. The remainder of the proof, being very similar to the proof of Theorem 4.1 is omitted. □

Remark 4.1 *The limiting constant* $0.180\,255\ldots$ *in* (70) *is in agreement with the figures obtained in Table 2. As follows from Proposition 4.2 and Tables 2, a normal fit of* F_n *is very poor with respect to the Gamma fit described in Theorem 4.1 and Table 1.*

Acknowledgement

We thank the referee for a careful reading of our manuscript and insightful comments.

References

M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 9th edition, 1970.

E.-E.A.A. Aly. Normalized spacings and goodness-of-fit tests. *J. Statist. Comput. Simul.*, 36:117–127, 1990.

E.-E.A.A. Aly, J. Beirlant, and L. Horváth. Strong and weak approximations of k-spacings processes. *Z. Wahrscheinlichkeit. verw. Gebiete*, 66:461–484, 1984.

J. Beirlant, P. Deheuvels, J.H.J. Einmahl, and D.P. Mason. Bahadur-Kiefer theorems for uniform spacings processes. *Teoria Verojantost. Primen.*, 36:724–743, 1991a.

J. Beirlant, L. Győrfi, and G. Lugosi. On the asymptotic normality of the L_1-errors and L_2-errors in histogram density estimation. *Can. J. Statist.*, 22:308–318, 1994.

J. Beirlant and L. Horváth. Approximations of m-overlapping spacings processes. *Scand. J. Statist.*, 11:225–245, 1984.

J. Beirlant, P. Janssen, and N. Veraverbeke. On the asymptotic normality of functions of uniform spacings. *Can. J. Statist.*, 19:93–101, 1991b.

J. Beirlant, E.C. van der Meulen, and F.H. Ruymgaart. On functions bounding the empirical distribution of uniform spacings. *Z. Wahrscheinlichkeit. verw. Gebiete*, 61:417–430, 1982.

S. Blumenthal. Contribution to sample spacings theory, I: Limit distributions of sums of ratios of spacings. *Ann. Math. Statist.*, 37:904–924, 1966a.

S. Blumenthal. Contribution to sample spacings theory, II: Test of the parametric goodness of fit and two-sample problems. *Ann. Math. Statist.*, 37:925–939, 1966b.

S. Blumenthal. Logarithms of sample spacings. *SIAM J. Appl. Math.*, 16:1184–1191, 1968.

V.P. Borovikov. Limit theorems for statistics that are partial-sums of functions of spacings. *Theor. Probab. Appl.*, 32:86–97, 1988.

D.R. Cox. Some statistical methods connected with series of events. *J. R. Statist. Soc. B*, 17:129–164, 1955.

N. Cressie. On the logarithms of high-order spacings. *Biometrika*, 63:343–355, 1976.

N. Cressie. Power results for tests based on high-order gaps. *Biometrika*, 65:214–218, 1978.

F. Czekala. Asymptotic distribution of statistics based on logarithm of spacings. *Zastos. Mat.*, 21:511–519, 1993.

D.A. Darling. On a class of problems related to the random division of an interval. *Ann. Math. Statist.*, 24:239–253, 1953.

H.A. David. *Order Statistics*. J. Wiley, New York, 2nd edition, 1981.

P. Deheuvels. Strong bounds for multidimensional spacings. *Z. Wahrscheinlichkeit. verw. Gebiete*, 64:411–424, 1983.

G. del Pino. On the asymptotic distribution of some goodness of fit tests based on spacings (abstract). *Bull. Inst. Math. Statist.*, 4:137, 1975.

J.H.J. Einmahl and M.C.A. van Zuijlen. Strong bounds for weighted empirical distribution functions based on uniform spacings. *Ann. Probab.*, 16:108–125, 1988.

M. Ekstrom. Strong limit theorems for sums of logarithms of high order spacings. *Statistics*, 33:153–169, 1991.

I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series and Products*. Academic Press, New York, 1982.

P. Guttorp and R.A. Lockhart. On the asymptotic distributions of high-order spacing statistics. *Can. J. Statist.*, 19:419–426, 1989.

P. Hall. Limit theorems for estimators based on inverses of spacings of order statistics. *Ann. Probab.*, 10:992–1003, 1982.

P. Hall. Limit theorems for general functions of m-spacings. *Math. Proc. Cambridge Philos. Soc.*, 96:517–532, 1984.

L. Holst. Asymptotic normality of sum-functions of spacings. *Ann. Probab.*, 7:1066–1072, 1979.

L. Holst. On convergence of the coverage by random arcs on a circle and the largest spacing. *Ann. Probab.*, 9:648–655, 1981.

J. Koziol. A note on limiting distributions for spacings statistics. *Z. Wahrscheinlichkeit verw. Gebiete*, 51:55–62, 1980.

L. Le Cam. Un théorème sur la division d'un intervalle par des points pris au hasard. *Publ. Inst. Statist. Univ. Paris*, 7:7–16, 1958.

P. L'Ecuyer. Tests based on sum-functions of spacings for uniform random numbers. *J. Stat. Comput. Sim.*, 59:251–269, 1997.

P. Lévy. Sur la division d'un segment par des points choisis au hasard. *C. R. Acad. Sci. Paris*, 208:147–149, 1939.

S. Malmquist. On a property of order statistics from a rectangular distribution. *Skand. Aktuarietidskr.*, 33:214–222, 1950.

J. Naus. A power comparison of two tests of non random clustering. *Technometrics*, 8: 493–517, 1966.

V.V. Petrov. *Sums of Independent Random Variables*. Springer, New York, 1st edition, 1975.

R. Pyke. Spacings. *J. R. Statist. Soc. B*, 27:395–436 and discussion 437–449, 1965.

R. Pyke. Spacings revisited. *Proc. 6th Berkeley Symp.*, 1:417–427, 1972.

B. Ranneby. The maximum spacing method: An estimation method related to the maximum likelihood method. *Scand. J. Statist.*, 11:93–112, 1984.

C.R. Rao. Some tests based on arc-lengths for the circle. *Sankhyā Ser. B*, 33:1–10, 1976.

J. Sethuraman and J.S. Rao. Pitman efficiencies of tests based on spacings. In M.L. Puri, editor, *Nonparametric Techniques in Statistical Inference*, pages 405–415. Cambridge University Press, 1969.

Y. Shao and M.G. Hahn. Limit theorems for the logarithms of sample spacings. *Statist. Prob. Lett.*, 24:121–132, 1995.

Y. Shao and R. Jiménez. Entropy for random partitions and its applications. *J. Theor. Probab.*, 11:417–433, 1998.

G.R. Shorack. Convergence of quantile and spacings processes with applications. *Ann. Math. Statist.*, 43:1400–1411, 1972.

G.R. Shorack and J.A. Wellner. *Empirical Processes with Applications to Statistics*. J. Wiley, New York, 1st edition, 1986.

A.F. Siegel. Random arcs on the circle. *J. Appl. Probab.*, 15:774–779, 1978.

A.F. Siegel. Asymptotic coverage distributions on the circle. *Ann. Probab.*, 7:651–661, 1979.

E. Slud. Entropy and maximal spacings for random partitions. *Z. Wahrscheinlichkeit. verw. Gebiete*, 41:341–352, 1978.

J. Spanier and K.B. Oldham. *An Atlas of Functions*. Hemisphere Publ. Co, Washington, 1987.

P.V. Sukhatme. Tests of significance for samples of the χ^2 population with two degrees of freedom. *Ann. Eugen.*, 8:52–56, 1937.

T. Swartz. Goodness of fit tests using Kullback-Leibler information. *Commun. Stat. Simulat.*, 21:711–729, 1992.

L. Weiss. A certain class of tests of fit. *Ann. Math. Statist.*, 27:1165–1170, 1956.

L. Weiss. The asymptotic power of certain tests of fit based on sample spacings. *Ann. Math. Statist.*, 28:783–786, 1957.

Authors' addresses:

Paul Deheuvels
L.S.T.A.
Université Pierre et Marie Curie
7 avenue du Château
F-92340 Bourg-la-Reine
France

E-mail: pd@ccr.jussieu.fr

Gérard Derzko
Centre de Recherche
Sanofi-Synthélabo
371 rue du Professeur Joseph Blayac
F-34184 Montpellier Cedex 04
France

E-mail: Gerard.Derzko@sanofi-synthelabo.com