

# UC Santa Barbara

## UC Santa Barbara Previously Published Works

### Title

Small sample approximations for spacing statistics

### Permalink

<https://escholarship.org/uc/item/8nv569zp>

### Journal

Journal of Statistical Planning and Inference, 69(2)

### ISSN

0378-3758

### Authors

Ghosh, Kaushik  
Jammalamadaka, S Rao

### Publication Date

1998-06-01

### DOI

10.1016/s0378-3758(97)00164-x

Peer reviewed



## Small sample approximations for spacing statistics

Kaushik Ghosh\*, S. Rao Jammalamadaka

*Department of Statistics and Applied Probability, University of California, Santa Barbara,  
CA 93106, USA*

Received 21 May 1997; received in revised form 21 August 1997

---

### Abstract

Edgeworth expansions as well as saddle-point methods are used to approximate the distributions of some spacing statistics for small to moderate sample sizes. By comparing with the exact values when available, it is shown that a particular form of Edgeworth expansion produces extremely good results even for fairly small sample sizes. However, this expansion suffers from negative tail probabilities and an accurate approximation without this disadvantage, is shown to be the one based on saddle-point method. Finally, quantiles of some spacing statistics whose exact distributions are not known, are tabulated, making them available in a variety of testing contexts. © 1998 Elsevier Science B.V. All rights reserved.

*AMS classifications:* primary 62E20; 62G10; 62G20

*Keywords:* Goodness-of-fit; Uniform spacings; Small sample asymptotics; Edgeworth expansion; Saddle-point approximation; Rao's spacing statistic; Greenwood statistic; Tests for uniformity

---

### 1. Introduction

Given a random sample from some unknown distribution, it is often necessary to check whether the sample comes from a particular completely specified distribution  $F_0$ . If  $F_0$  is continuous, by the probability integral transform  $x \mapsto F_0(x)$ , this is equivalent to testing for uniformity of the transformed sample. From now on in this article, we assume that such a transformation has already been applied and our data is on the unit interval  $(0, 1)$ . Suppose  $X_1, X_2, \dots, X_n$  is a random sample from  $F$  where  $F$  is a continuous distribution on  $(0, 1)$  and we are interested in testing  $H_0: F = U(0, 1)$  where  $U(0, 1)$  is the uniform distribution on the interval  $(0, 1)$ . This is the classical goodness-of-fit problem and in this article, we will be concerned with tests based on spacings, i.e., the gaps between consecutive ordered observations.

---

\* Corresponding author. E-mail: [ghosh@pstat.ucsb.edu](mailto:ghosh@pstat.ucsb.edu).

Spacings are the natural choice for directional data problems where they are the maximal invariants with respect to change of the zero direction and the sense of rotation.

Let  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  be the corresponding order statistics for the data in hand. Define the spacings obtained from the sample by

$$D_i = X_{(i)} - X_{(i-1)}, \quad i = 1, 2, \dots, n + 1,$$

where  $X_{(0)} \equiv 0$  and  $X_{(n+1)} \equiv 1$ . Under  $H_0$ ,  $X_1, X_2, \dots, X_n$  are i.i.d.  $U(0, 1)$  and the  $D_i$ 's are called Uniform Spacings. For the special case of Uniform Spacings, we denote them by  $T_i$ .

Various statistics based on spacings have been proposed to test for uniformity. A large class of these can be written in the general form

$$G_n = \sum_{i=1}^{n+1} h((n + 1)D_i).$$

Some standard cases, which have been discussed in the literature, correspond to  $h(x) = x^2$ ,  $h(x) = |x - 1|$ ,  $h(x) = \log x$  and  $h(x) = x \log x$ . It has been shown (see, Sethuraman and Rao, 1970) that under a smooth sequence of alternatives, the Greenwood statistic has the maximum efficacy in this class  $G_n$ .

In order to use these tests based on spacings, it is necessary to know the null distributions of the corresponding statistics. It turns out that the exact small sample null distributions are not known in most of the cases. The asymptotic null distributions are known to be normal under mild conditions on  $h(\cdot)$ . However, this asymptotic normality is potentially misleading since it is generally good only for considerably large sample sizes. In this paper we discuss the use of approximations which ease calculations and which are quite accurate for small to moderate sample sizes.

Section (2) contains a brief review of what has been previously done. Section (3) discusses the use of Edgeworth expansion techniques while Section (4) considers saddle-point approximations.

## 2. Background

Let us denote the Greenwood statistic (corresponding to  $h(x) = x^2$ ), by

$$G_{1,n} = \sum_{i=1}^{n+1} T_i^2, \tag{2.1}$$

the Rao's spacing statistic (corresponding to  $h(x) = |x - 1|$ ), by

$$G_{2,n} = \sum_{i=1}^{n+1} \left| T_i - \frac{1}{n + 1} \right| \tag{2.2}$$

and the ‘log’ statistic by

$$G_{3,n} = \sum_{i=1}^{n+1} \log(T_i), \tag{2.3}$$

with the corresponding cumulative distribution functions denoted by  $F_{1,n}(\cdot)$ ,  $F_{2,n}(\cdot)$  and  $F_{3,n}(\cdot)$ , respectively. Early attempts to tabulate exact values of  $F_{1,n}(\cdot)$  were by Greenwood (1946) and Gardner (1952), who obtained exact distributions upto  $n = 3$ . Burrows (1979) and Currie (1981) tabulated selected percentage points of the Greenwood statistic up to  $n = 20$  using a recursive algorithm. The method breaks down for higher  $n$  due to complicated nature of the algorithm and hence, no exact tables are available for such cases.

Hill (1979) used Johnson and Log-normal curves to approximate  $F_{1,n}(\cdot)$  for  $n$  up to 10 and Stephens (1981) fitted Pearson curves to approximate  $F_{1,n}(\cdot)$  for  $n \geq 12$ . Although the latter is pretty close to the exact values for  $n = 20$ , its behavior for higher  $n$  is not known. Moreover, this method does not give any easy approximating formula for  $F_{1,n}(\cdot)$ .

The exact density of  $G_{2,n}/2$  is given by

$$f(x) = n! \sum_{j=1}^n \binom{n+1}{j} \frac{x^{n-j} f_j((n+1)x)}{(n-j)!(n+1)^{j-1}},$$

where

$$f_j(x) = \frac{1}{(j-1)!} \sum_{0 \leq k < x} (-1)^k \binom{j}{k} (x-k)^{j-1}, \quad 0 < x < j.$$

Here  $f_j(\cdot)$  stands for the density of the convolution of  $j$  independent  $U(0,1)$  random variables (see, e.g., Rao (1976) for circular case and also Darling (1953)). Using numerical integration, Rao (1976) and Batschelet (1981) provide exact critical values of  $G_{2,n}$  for  $\alpha = 0.01, 0.05$  and  $0.1$  and selected values of  $n = 1$  to 200. Their tables were further extended by Russell and Levitin (1995) to incorporate more values of  $n$  and  $\alpha$ . As is obvious from the formula of density, this method is very slow since it is computationally involved and hence, impractical.

For all other spacing statistics, neither the exact distributions nor tables of critical values are available for finite  $n$ .

Using the well-known fact that

$$\{(n+1)T_i\}_{i=1}^{n+1} \stackrel{d}{=} \left\{ \frac{W_i}{\bar{W}} \right\}_{i=1}^{n+1}, \tag{2.4}$$

where  $W_1, W_2, \dots, W_{n+1}$  are i.i.d. exponential with mean 1 and  $\bar{W}$  is their sample mean (see, e.g., Pyke, 1965), it can be shown that all the spacing based statistics discussed previously are asymptotically normal (see, e.g., Rao and Sethuraman, 1975).

Moran (1947) gives the first four raw moments of Greenwood statistic to be

$$\begin{aligned}
 \mu'_1 &= \frac{2}{n+2}, \\
 \mu'_2 &= \frac{2^2(n+6)}{(n+2)(n+3)(n+4)}, \\
 \mu'_3 &= \frac{2^3(n^2+17n+90)}{(n+2)\cdots(n+6)}, \\
 \mu'_4 &= \frac{2^4(n^3+33n^2+434n+2520)}{(n+2)\cdots(n+8)},
 \end{aligned}
 \tag{2.5}$$

from which, the measures of skewness and kurtosis are obtained to be

$$\begin{aligned}
 \beta_{1,n} &= \frac{\mu_3}{\mu_2^{3/2}} = \frac{(10n-4)(n+3)^{1/2}(n+4)^{1/2}}{n^{1/2}(n+5)(n+6)}, \\
 \beta_{2,n} &= \frac{\mu_4}{\mu_2^2} = \frac{(3n^3+303n^2+42n-24)(n+3)(n+4)}{n(n+5)(n+6)(n+7)(n+8)}.
 \end{aligned}
 \tag{2.6}$$

Darling (1953) gives the formula for the characteristic function of the ‘log’ statistic  $G_{3,n}$  to be

$$\varphi(t) = \frac{\Gamma(n+1)\{\Gamma(it+1)\}^{n+1}}{\Gamma((n+1)(it+1))}.
 \tag{2.7}$$

Differentiating Eq. (2.7), the first four raw moments of the ‘log’ statistic can be obtained in terms of polygamma functions. These are very complicated but can be handled using software capable of symbolic mathematics like Mathematica or Maple. Using them and Eq. (2.6), we found the skewness and kurtosis of the two statistics for various values of  $n$ . Table 1 summarizes the results.

In comparison with the skewness and kurtosis values for a normal distribution of 0 and 3, respectively, these are quite ‘non-normal’ even for  $n = 50$  or  $100$ . The skewness and kurtosis values of Rao’s spacing statistic can be found to be (0.247, 2.940) for  $n = 5$

Table 1  
Skewness and kurtosis of some spacing statistics

$n$	$G_{1,n}$		$G_{3,n}$	
	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
5	1.587	6.827	-1.202	5.116
10	1.706	8.351	-0.853	4.068
20	1.584	8.201	-0.605	3.537
50	1.218	6.378	-0.383	3.216
100	0.926	5.026	-0.271	3.108
500	0.440	3.473	-0.121	3.022
1000	0.314	3.241	-0.086	3.011
$\infty$	0.000	3.000	0.000	3.000

and (0.180, 2.971) for  $n = 10$ . This shows that in some cases like  $G_{2,n}$ , asymptotic normality kicks in rather quickly while for others, a normal approximation would be quite bad for small samples.

Hence, there is a need to look for approximations to the exact distributions; in particular, the Edgeworth expansions and saddle-point approximations.

### 3. Edgeworth expansion method

From now on, we use the symbols  $\Phi(\cdot)$  and  $\phi(\cdot)$  to denote the standard normal distribution and density functions, respectively. Edgeworth expansions for statistics have a long history and we first quote the following general result from Hall (1992, pp. 46–48):

**Proposition 3.1.** *Suppose  $S_n$  is a statistic with a limiting standard normal distribution and is a ‘smooth’ function of vector means. Then,*

$$P(S_n \leq x) = \Phi(x) + \phi(x) \left( \frac{p_1(x)}{\sqrt{n}} + \frac{p_2(x)}{n} + \dots + \frac{p_j(x)}{n^{j/2}} \right) + o(n^{-j/2}),$$

where

$$\begin{aligned} p_1(x) &= -\{k_{1,2} + \frac{1}{6}k_{3,1}(x^2 - 1)\}, \\ p_2(x) &= -x\{\frac{1}{2}(k_{2,2} + k_{1,2}^2) + \frac{1}{24}(k_{4,1} + 4k_{1,2}k_{3,1})(x^2 - 3) \\ &\quad + \frac{1}{72}k_{3,1}^2(x^4 - 10x^2 + 15)\}, \\ &\vdots \end{aligned} \tag{3.8}$$

and

$$\kappa_{j,n} = n^{-(j-2)/2}(k_{j,1} + n^{-1}k_{j,2} + n^{-2}k_{j,3} + \dots), \quad j \geq 1,$$

is the  $j$ th cumulant of  $S_n$ .

A different kind of Edgeworth expansion correct to  $o(n^{-1})$  for any asymptotic normal statistic can be obtained as follows:

**Proposition 3.2.** *Suppose  $G_n$  is a statistic with an asymptotic normal distribution and*

$$\tilde{S}_n = \frac{G_n - E(G_n)}{\sqrt{\text{Var}(G_n)}}.$$

Let  $\beta_{1,n} \stackrel{\text{def}}{=} \beta_1(G_n)$  and  $\beta_{2,n} \stackrel{\text{def}}{=} \beta_2(G_n)$ . Then, if  $\beta_{1,n} = O(n^{-1/2})$ ,  $\beta_{2,n} - 3 = O(n^{-1})$  and higher-order cumulants of  $\tilde{S}_n$  are  $O(n^{-\alpha})$ ,  $\alpha > 1$ , we have

$$P(\tilde{S}_n \leq x) = \Phi(x) - \phi(x) \left[ \frac{\beta_{1,n}(x^2 - 1)}{6} + \frac{(\beta_{2,n} - 3)(x^3 - 3x)}{24} + \frac{\beta_{1,n}^2(x^5 - 10x^3 + 15x)}{72} + o(n^{-1}) \right]. \tag{3.9}$$

**Proof.** By the given conditions, the characteristic function of  $\tilde{S}_n$  is

$$\begin{aligned} \chi(t) &= \exp \left[ 0 + 1 \frac{(it)^2}{2} + \beta_{1,n} \frac{(it)^3}{6} + (\beta_{2,n} - 3) \frac{(it)^4}{24} + o(n^{-1}) \right] \\ &= e^{-t^2/2} \left[ 1 + \beta_{1,n} \frac{(it)^3}{6} + (\beta_{2,n} - 3) \frac{(it)^4}{24} + \beta_{1,n}^2 \frac{(it)^6}{72} + o(n^{-1}) \right]. \end{aligned}$$

On inversion, this gives

$$P(\tilde{S}_n \leq x) = \Phi(x) - \phi(x) \left[ \beta_{1,n} \frac{H_2(x)}{6} + \frac{(\beta_{2,n} - 3)}{24} H_3(x) + \frac{\beta_{1,n}^2}{72} H_5(x) + o(n^{-1}) \right],$$

where  $H_j(x)$  is the  $j$ th Hermite polynomial and satisfies

$$\int_{-\infty}^{\infty} e^{-itx} (it)^j dt = -H_{j-1}(x)\phi(x), \quad j \geq 1. \quad \square$$

A similar result without complete proof can be found in Does et al. (1988) where it is specifically applied to the Greenwood statistic. We now apply these results to the Greenwood and ‘log’ statistics. Some of the regularity conditions like smoothness of  $h(\cdot)$ , do not hold for Rao’s spacing statistic and hence these results are not directly applicable.

### 3.1. The Greenwood statistic

Using the characterization of spacings given in Eq. (2.4), Proposition 3.1 can be applied to obtain an Edgeworth expansion of normalized Greenwood statistic:

$$S_n = \frac{\sqrt{n+1}}{2} \{(n+1)G_{1,n} - 2\}. \tag{3.10}$$

Kabaila (1993) describes a method for computer calculation of Edgeworth expansions of smooth function models accurate to the  $o(n^{-1})$  term. The method expands the cumulants of the statistic asymptotically and collects the relevant coefficients to form

the polynomials  $p_1(\cdot)$  and  $p_2(\cdot)$ . We modified Kabaila’s method for use on the normalized Greenwood statistic, Eq. (3.10), and implemented it in *Mathematica*. Using Eq. (2.5), asymptotic expansion of the cumulants of Eq. (3.10) gives

$$\begin{aligned} \kappa_{1,n} &= n^{1/2} \left\{ 0 - \frac{1}{n} + \frac{1}{n^2} - \frac{1}{n^3} + \frac{1}{n^4} + \dots \right\}, \\ \kappa_{2,n} &= n^0 \left\{ 1 - \frac{8}{n} + \frac{39}{n^2} - \frac{154}{n^3} + \frac{545}{n^4} + \dots \right\}, \\ \kappa_{3,n} &= n^{-1/2} \left\{ 10 - \frac{194}{n} + \frac{2152}{n^2} - \frac{18180}{n^3} + \frac{130878}{n^4} + \dots \right\}, \\ \kappa_{4,n} &= n^{-1} \left\{ 246 - \frac{8640}{n} + \frac{168204}{n^2} - \frac{2420688}{n^3} + \frac{28902666}{n^4} + \dots \right\}. \end{aligned}$$

Thus, Eq. (3.8) gives

$$\begin{aligned} p_1(x) &= \frac{8}{3} - \frac{5}{3}x^2, \\ p_2(x) &= \frac{101}{12}x + \frac{191}{36}x^3 - \frac{25}{18}x^5. \end{aligned}$$

Hence,

$$\begin{aligned} P(S_n \leq x) &= \Phi(x) + \phi(x) \left\{ \frac{1}{\sqrt{n}} \left( \frac{8}{3} - \frac{5}{3}x^2 \right) \right. \\ &\quad \left. + \frac{1}{n} \left( \frac{101}{12}x + \frac{191}{36}x^3 - \frac{25}{18}x^5 \right) \right\} + o(n^{-1}). \end{aligned} \tag{3.11}$$

We can of course apply Proposition 3.2 to obtain an Edgeworth expansion of the statistic

$$\tilde{S}_n = \left( G_{1,n} - \frac{2}{n+2} \right) \sqrt{\frac{(n+2)^2(n+3)(n+4)}{4n}}. \tag{3.12}$$

From Eq. (2.6), we have  $\beta_1(G_{1,n}) = O(n^{-1/2})$ ,  $\beta_2(G_{1,n}) = 3 + O(n^{-1})$ . Also, since  $\tilde{S}_n = O(1)S_n + O(n^{1/2})$ ,  $\kappa_r(\tilde{S}_n) = O(n^{-(r-2)/2})$  for  $r \geq 5$  by properties of cumulants.

Hence,

$$\begin{aligned} P(\tilde{S}_n \leq x) &= \Phi(x) - \phi(x) \left\{ \beta_{1,n} \frac{(x^2 - 1)}{6} + \frac{(\beta_{2,n} - 3)(x^3 - 3x)}{24} \right. \\ &\quad \left. + \frac{\beta_{1,n}^2(x^5 - 10x^3 + 15x)}{72} \right\} + o(n^{-1}) \end{aligned} \tag{3.13}$$



Table 2  
Quantiles of Greenwood statistic using Edgeworth expansion,  $n = 10$

$\alpha$	$t_\alpha$				
	Exact	Normal	Edgeworth of order		DHK
			$n^{-1/2}$	$n^{-1}$	
0.01	0.111694	0.054287	0.096327	0.130842	0.116383
0.05	0.121088	0.091647	0.104439	0.133684	0.121583
0.1	0.127248	0.111563	0.112076	0.137016	0.126723
0.2	0.136050	0.13568	0.12381	0.143213	0.135104
0.3	0.143499	0.15307	0.133414	0.149072	0.142477
0.4	0.150744	0.16793	0.142088	0.154819	0.149639
0.5	0.158375	0.181818	0.150434	0.160627	0.15711
0.6	0.166976	0.195707	0.15892	0.166661	0.165441
0.7	0.177436	0.210566	0.168109	0.17311	0.175426
0.8	0.191648	0.227956	0.179087	0.180226	0.188419
0.9	0.215717	0.252073	0.196186	0.188392	0.208281
0.95	0.240356	0.271989	0.31685	0.193063	0.262817
0.99	0.300793	0.309349	0.35937	0.197196	0.310416

Table 2 compares the results obtained when the two expansions are used in estimating the quantiles of Greenwood statistic. Note that  $t_\alpha$  satisfies  $P(G_{1,n} \leq t_\alpha) = \alpha$ . Fig. 1 summarizes the information in Table 2.  $Q_{ex}$  is the exact quantile and  $Q_{ap}$  is the approximate quantile obtained using the appropriate approximation. The  $x$ -axis gives the probabilities and the  $y$ -axis gives the corresponding absolute difference in the quantiles,  $|Q_{ex} - Q_{ap}|$  for the various approximation formulae. The exact quantiles are taken from Burrows (1979), Currie (1981) and Stephens (1981).

Visual inspection of Table 2 and Fig. 1 suggests that the approximation given by Proposition 3.2 outperforms the one obtained from Proposition 3.1. This can be explained by the fact that the expansion based on Proposition 3.2 uses the exact cumulants whereas the one based on Proposition 3.1 just collects the relevant coefficients from the series expansions of the cumulants upto a given order. Such high accuracy may also be explained by the fact that normalization is done with the exact mean and variance in Proposition 3.2 while asymptotic values are used in Proposition 3.1. Thus, the approximation of Proposition 3.2 provides an exactly centered and scaled statistic which behaves more like a  $N(0, 1)$  random variable in small sample sizes.

Adding higher-order terms to Eqs. (3.10) and (3.12) do not improve the results. This is because, the oscillations of higher-order Hermite polynomials offset any of the benefits of taking higher powers of  $1/n$ .

Even though the approximation using Eq. (3.12) was very close to the exact one, Fig. 2 shows the inherent problems with all Edgeworth expansions – they do not provide us with an exact distribution function. Hence, a density obtained by this method may turn out to be negative at some points or may not integrate to 1.

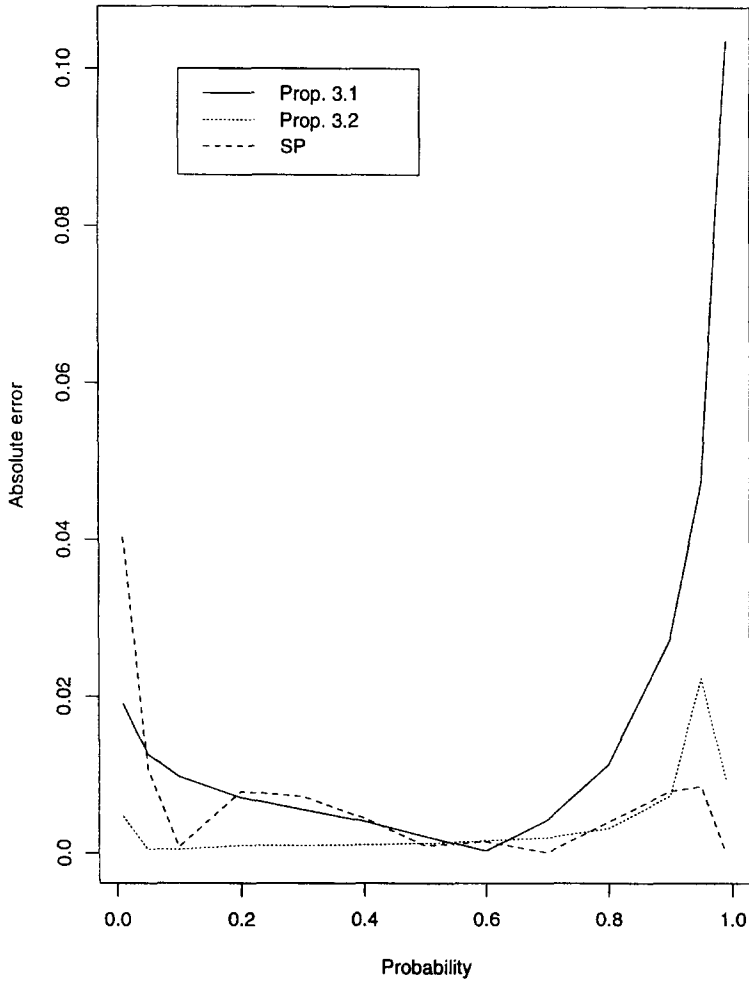


Fig. 1. Errors in quantile estimation of Greenwood statistic,  $n = 10$ .

### 3.2. The 'log' statistic

As in the case of Greenwood statistic, Propositions 3.1 and 3.2 can be applied to obtain two Edgeworth expansions of the 'log' statistic  $G_{3,n}$  discussed originally in Moran (1951). Defining

$$S_n = \left( \frac{G_{3,n}}{n+1} + \log(n+1) + \gamma \right) \div \sqrt{\frac{(\pi^2/6) - 1}{n+1}}, \quad (3.14)$$

where  $\gamma = 0.57722$  is the Euler number, it is easy to check that  $S_n$  is a smooth function of vector means with a limiting standard normal distribution. Hence, the conditions of

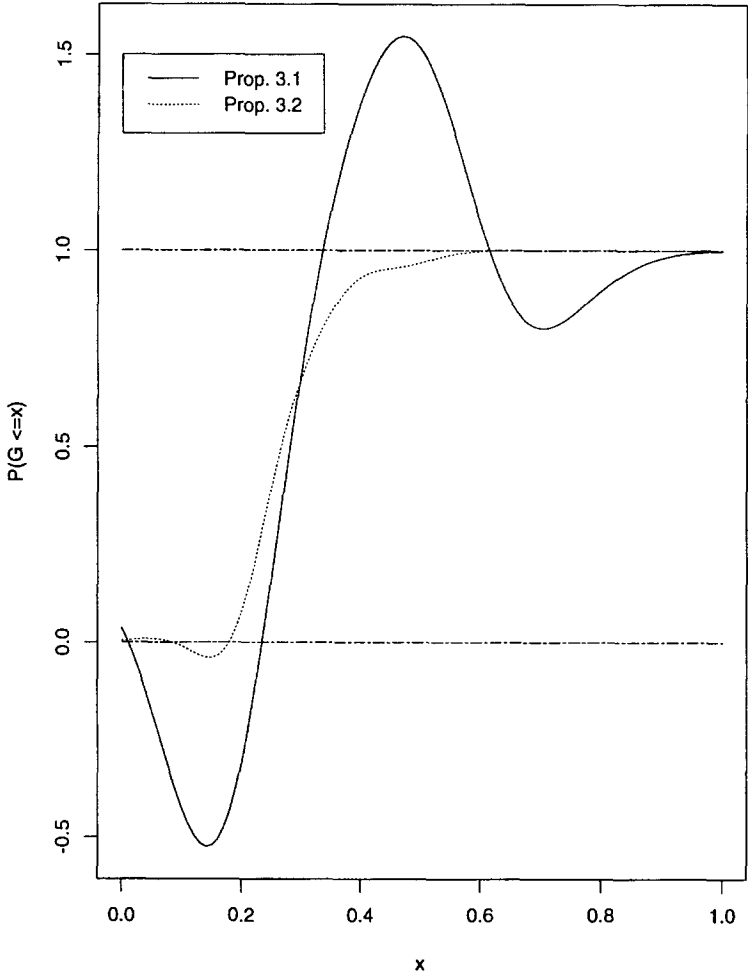


Fig. 2. Edgeworth expansions of Greenwood Statistic  $n = 5$ .

Proposition 3.1 are satisfied. Calculations yield

$$\begin{aligned}
 P(S_n \leq x) = \Phi(x) + \phi(x) \left\{ \frac{1}{\sqrt{n}}(-0.244 + 0.452x^2) \right. \\
 \left. + \frac{1}{n}(0.467x + 0.477x^3 - 0.102x^5) \right\} + o(n^{-1}).
 \end{aligned}
 \tag{3.15}$$

Similarly, defining

$$\tilde{S}_n = \frac{G_{3,n} - E(G_{3,n})}{\sqrt{\text{Var}(G_{3,n})}},
 \tag{3.16}$$

Proposition 3.2 yields another Edgeworth expansion of the ‘log’ statistic.

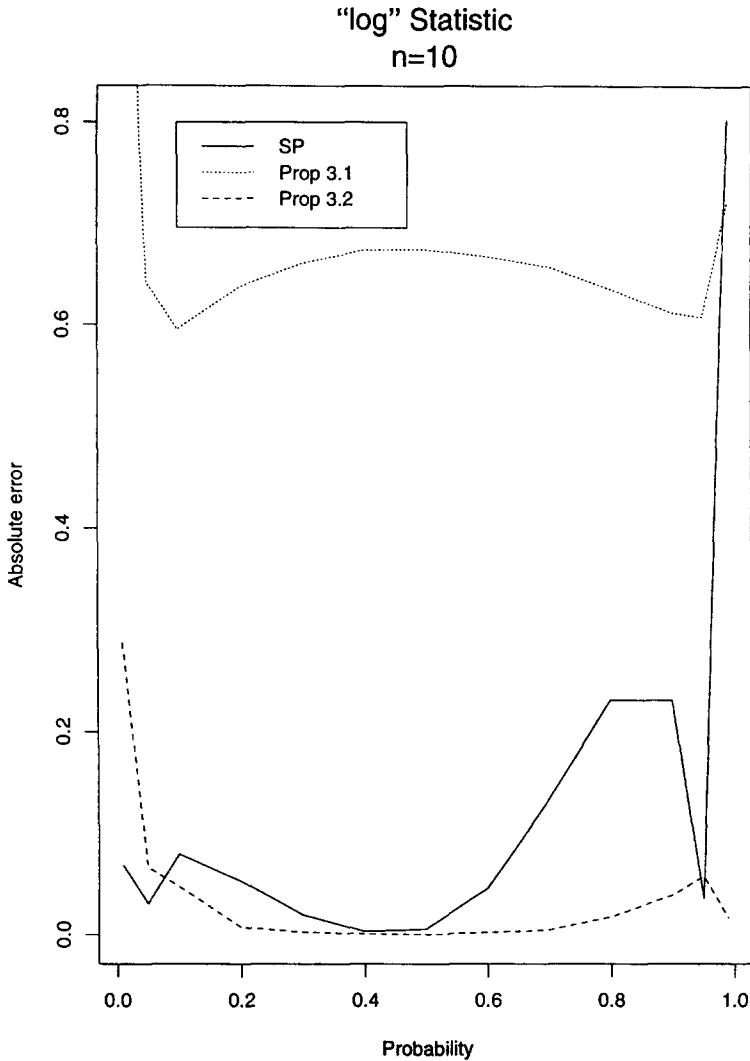


Fig. 3. Errors in quantile estimation of  $G_{3,n}$ ,  $n = 10$ .

Fig. 3 shows the errors in approximating quantiles using the two methods. We have used the results of  $10^5$  Monte-Carlo simulations for exact values.

Fig. 4 plots the two approximations. A quick glance at Fig. 3 shows that Proposition 3.2 outperforms Proposition 3.1.

We observe that of the two Edgeworth expansions discussed here, the one derived from Proposition 3.2 seems to perform better. Even though this is quite accurate for small sample sizes, it suffers from oscillations in the tails, where it can be outside  $[0, 1]$ . This is a problem since in practical situations, it is the tail areas that require accurate approximations. One method of overcoming this is through an alternative approach,

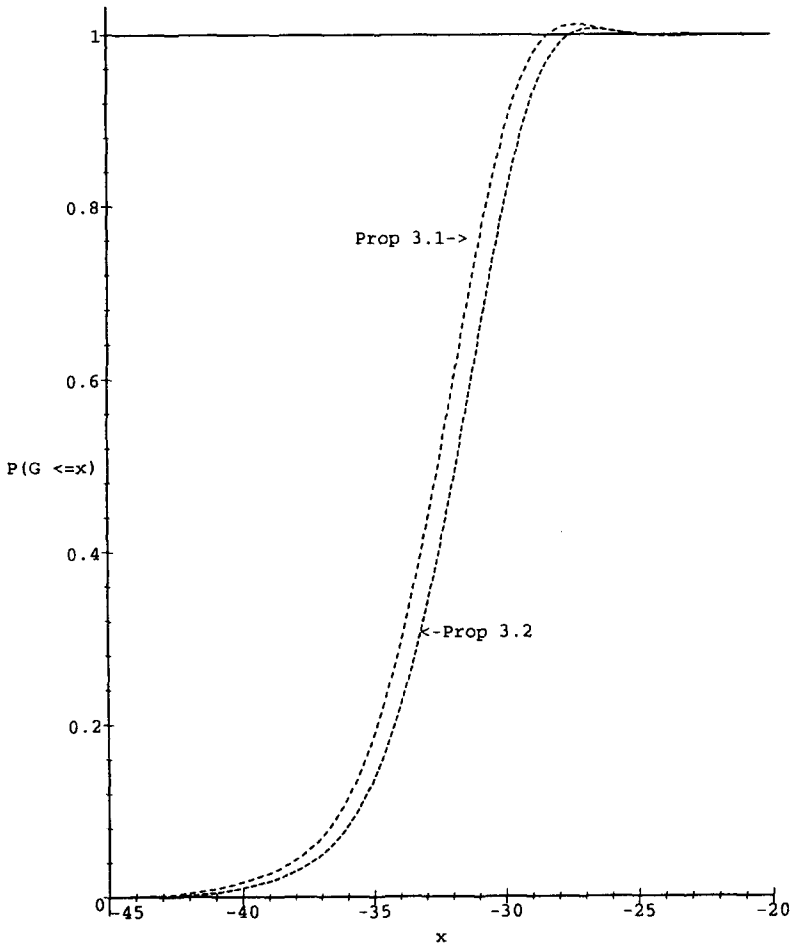


Fig. 4. Edgeworth expansions of  $G_{3,n}, n = 10$ .

namely saddle-point approximations which always provide a positive density. This is the topic of the next section.

#### 4. Saddle-point approximation

##### 4.1. Theory and notations

Daniels (1954) in a pioneering paper, proposed saddle-point methods to approximate distributions. Two good review papers on the topic are Daniels (1987) and Reid (1988). Let  $T_n$  be a real valued statistic and  $K_n(t)$  be its cumulant generating function. Let  $R_n(t) = K_n(nt)/n$ . Then the saddle-point approximation of the density of  $T_n$  with

uniform error of order  $n^{-1}$  is given by

$$g_n(x) = \sqrt{\frac{n}{2\pi R_n''(t_0)}} \exp[n\{R_n(t_0) - t_0x\}], \tag{4.17}$$

where  $t_0$  is the saddle-point, determined as the root of the equation

$$R_n'(t_0) = x. \tag{4.18}$$

The saddle-point approximation of the distribution of  $T_n$  is given by

$$G_n(x) = \Phi(y) + \phi(y)(y^{-1} - z^{-1}), \tag{4.19}$$

where  $\Phi$  and  $\phi$  are the standard normal distribution and density, respectively,

$$y = \sqrt{2n(t_0x - R_n(t_0))} \operatorname{sgn}(t_0),$$

$$z = t_0 \sqrt{nR_n''(t_0)}$$

and where  $R_n', R_n''$  denote the first two derivatives of  $R_n$  (see, Easton and Ronchetti, 1986).

Often, the exact  $R_n$  is not available and an approximate form is used. Easton and Ronchetti (1986) suggest approximating the cumulant generating function by a fourth-degree polynomial using only the first four cumulants of  $T_n$ . Hence, the new method replaces  $R_n$  by  $\tilde{R}_n$  defined by

$$\tilde{R}_n(t) = \kappa_{1,n}t + \kappa_{2,n}nt^2/2 + \kappa_{3,n}n^2t^3/6 + \kappa_{4,n}n^3t^4/24, \tag{4.20}$$

where  $\kappa_{i,n}; i = 1, 2, 3, 4$  are the first four cumulants of  $T_n$ .

One drawback with Eq. (4.20) is that since  $\tilde{R}_n'(t)$  is not always strictly increasing, it results in multiple roots to the saddle-point equation, Eq. (4.18). In such cases, Wang (1992) suggests replacing  $\tilde{R}_n$  by the following:

$$\hat{R}_n(t; b) = \kappa_{1,n}t + \kappa_{2,n}nt^2/2 + (\kappa_{3,n}n^2t^3/6 + \kappa_{4,n}n^3t^4/24)e^{-(\kappa_{2,n}nb^2t^2/2)}, \tag{4.21}$$

where  $b$  is chosen so that  $\hat{R}_n''(t; b) > 0 \forall t$ . This ensures a unique solution to Eq. (4.18).

#### 4.2. The Greenwood statistic

Since the exact mgf of  $G_{1,n}$  is not known, we use the approximate version, Eq. (4.20), of  $R_n$ . Using the first four cumulants of this statistic from Moran (1947), it is easily checked that  $\tilde{R}_n''(t) > 0 \forall t$  for  $n > 1$ . Hence, modifications of Eq. (4.21) are not necessary. The results are summarized in Table 3. Fig. 1 compares the accuracy of the results with those obtained from Edgeworth expansions. Proposition 3.2 seems to perform better overall.

#### 4.3. Rao's spacing statistic

Since in this case, the exact pdf as well as tables of percentage points are available, we used this as a test case of how well saddle-point methods based on the four cumulant approximation work.

Table 3  
Quantiles of Greenwood statistic using saddle-point approximation

$\alpha$	$t_\alpha$					
	$n = 5$		$n = 10$		$n = 20$	
	SP	Exact	SP	Exact	SP	Exact
0.01	0.1159	0.1839	0.0714	0.1117	0.0482121	0.064864
0.05	0.1831	0.1994	0.1103	0.1211	0.065479	0.069786
0.10	0.2131	0.2101	0.1280	0.1272	0.0733643	0.073
0.20	0.2399	0.2260	0.1440	0.1360	0.0806165	0.077
0.30	0.2523	0.2399	0.1508	0.1435	0.0838482	0.081
0.40	0.2610	0.2537	0.1552	0.1507	0.0859238	0.084
0.50	0.2702	0.2684	0.1594	0.1584	0.0879011	0.088
0.60	0.2846	0.2853	0.1655	0.1670	0.0907076	0.091
0.70	0.3089	0.3060	0.1773	0.1774	0.0958185	0.096
0.80	0.3428	0.3344	0.1956	0.1916	0.103712	0.102
0.90	0.3946	0.3830	0.2237	0.2157	0.115993	0.113
0.95	0.4410	0.4320	0.2490	0.2404	0.127029	0.123
0.99	0.5360	0.5475	0.3009	0.3008	0.149747	0.149

Table 4  
Distribution function of Rao’s spacing statistic using saddle-point approximation

$t$	$P(G_{2,n} \leq t)$			
	$n = 5$		$n = 10$	
	SP	Exact	SP	Exact
0.278	0.012	0.013	0	0
0.333	0.030	0.032	0.002	0.003
0.389	0.065	0.067	0.010	0.011
0.444	0.120	0.121	0.032	0.032
0.500	0.196	0.196	0.078	0.079
0.555	0.289	0.289	0.158	0.159
0.611	0.397	0.399	0.275	0.276
0.667	0.511	0.512	0.419	0.418
0.722	0.619	0.619	0.569	0.569
0.778	0.718	0.717	0.708	0.708
0.833	0.801	0.801	0.819	0.819
0.889	0.867	0.868	0.899	0.899
0.944	0.915	0.916	0.948	0.948
1.000	0.949	0.948	0.976	0.976
1.056	0.971	0.970	0.990	0.99
1.111	0.984	0.984	0.996	0.996
1.167	0.992	0.992	0.999	0.999
1.222	0.996	0.996	1	1

We used the correction from Eq. (4.21) to alleviate the problem of non-unique solutions and the results obtained are in Table 4. In every case  $b$  turned out to be  $\frac{1}{2}$ . The results are very accurate, even for very small  $n$ . Fig. 5 illustrates its accuracy in estimating quantiles. The exact values are from Russell and Levitin (1995).

### Rao's Spacing Statistic

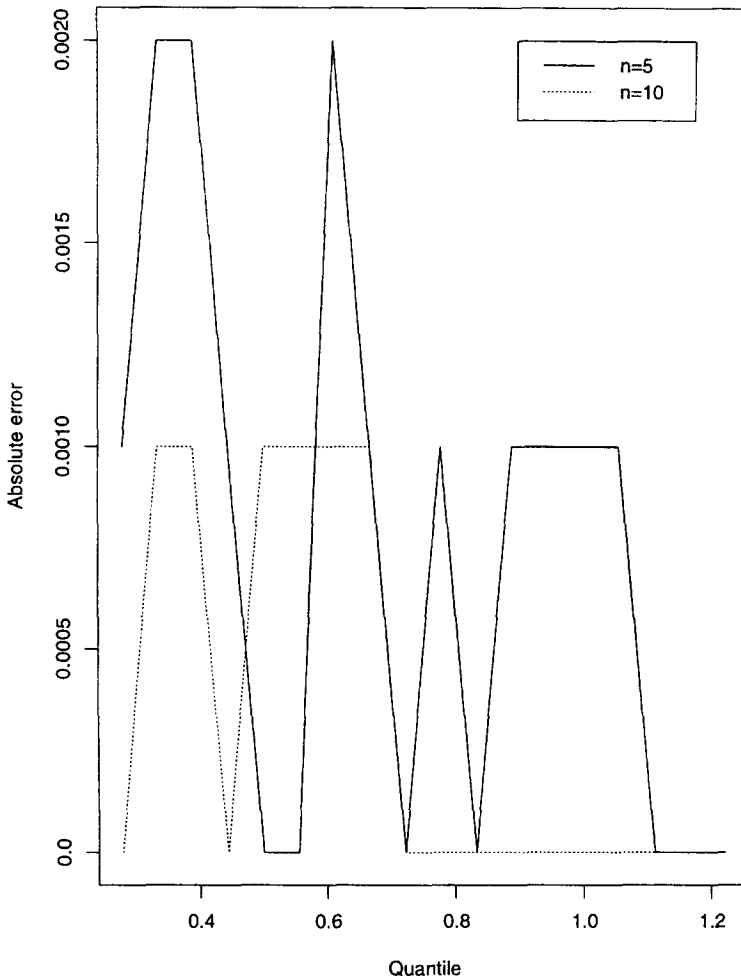


Fig. 5. Errors in quantile estimation of Rao's spacing statistic (using saddle-point approximation).

Instead of using just the first four cumulants, we could have used the whole cumulant generating function as was done recently by Gatto and Jammalamadaka (1996). The method is considerably slower and for all practical purposes the gain is not much. In fact, for Rao's spacing statistic with  $n = 10$ , our maximum observed error is 0.001 while theirs is 0.003.

#### 4.4. The 'log' statistic

Since no previous results are available for the statistic with  $h(x) = \log(x)$ , we cannot determine the accuracy of the results as before but instead we compare them



Table 5  
Quantiles of  $G_{3,n}$  using saddle-point approximation

$\alpha$	$t_\alpha$		$n = 10$		$n = 20$	
	$n = 5$		SP	Simulated	SP	Simulated
	SP	Simulated				
0.01	-19.3834	-19.3710	-39.6712	-39.7399	-85.4782	-85.4932
0.05	-17.295	-17.1897	-37.0352	-37.0047	-82.0942	-82.0425
0.1	-16.2716	-16.1406	-35.7328	-35.6531	-80.4045	-80.3903
0.2	-15.1251	-15.0253	-34.2619	-34.2092	-78.4743	-78.4547
0.3	-14.3717	-14.3353	-33.2827	-33.2632	-77.1683	-77.1785
0.4	-13.7983	-13.7968	-32.5119	-32.5090	-76.1158	-76.1363
0.5	-13.3632	-13.3350	-31.8626	-31.8574	-75.1915	-75.2115
0.6	-13.0438	-12.9234	-31.3000	-31.2545	-74.332	-74.3334
0.7	-12.7707	-12.5350	-30.7907	-30.6570	-73.4895	-73.452
0.8	-12.4482	-12.1536	-30.2562	-30.0247	-72.5932	-72.4578
0.9	-11.8735	-11.7159	-29.4792	-29.2480	-71.4253	-71.2035
0.95	-11.2607	-11.4378	-28.735	-28.6995	-70.4365	-70.2815
0.99	-9.87405	-11.0778	-27.0856	-27.8886	-68.3877	-68.78

with Monte-Carlo results based on  $10^5$  simulations. In this case, we did not have to modify the saddle-point approximation formula since the cumulant generating function was well behaved. The results are summarized in Table 5. Fig. 3 compares the accuracy of this method with Edgeworth expansions. Proposition 3.2 seems to outperform the saddle-point approximation.

### 5. Conclusions

Overall, we see that the Edgeworth expansion obtained using Proposition 3.2 outperforms the other approximation methods that we have studied, although the saddle-point approximations are equally accurate in the tails. Saddle-point methods also have the advantage that they always yield non-negative probabilities. This is especially important in the tails where Edgeworth expansions can give rise to negative probabilities.

Gatto and Jammalamadaka (1996) discuss a general conditional saddle-point approach and apply it, in particular, to obtain distributions of spacings statistics. Our method based on only the first four moments is simpler and often equally accurate for spacings statistics.

### References

Batschelet, E., 1981. Circular Statistics In Biology. Academic Press, London.  
 Burrows, P.M., 1979. Selected percentage points of Greenwood’s statistic. J. Roy. Statist. Soc. Ser. A 142(2), 256–258.  
 Currie, I.D., 1981. Further percentage points of Greenwood’s statistic. J. Roy. Statist. Soc. Ser. A 144(3), 360–363.  
 Daniels, H.E., 1954. Saddlepoint approximations in statistics. Ann. Math. Statist. 25, 631–650.

- Daniels, H.E., 1987. Tail probability approximations. *Internat. Statist. Rev.* 55(1), 37–48.
- Darling, D.A., 1953. On a class of problems related to the random division of an interval. *Ann. Math. Statist.* 24, 239–253.
- Does, R.J.M.M. et al., 1988. Approximating the distribution of Greenwood's statistic. *Statist. Neerlandica* 42(3), 153–161.
- Easton, G.S., Ronchetti, E., 1986. General saddlepoint approximations with applications to  $L$  statistics. *J. Amer. Statist. Assoc.* 81(394), 420–430.
- Gardner, A., 1952. Greenwood's 'Problem of intervals': an exact solution for  $n = 3$ . *J. Roy. Statist. Soc. Ser. B* 14, 135–139.
- Gatto, R., Jammalamadaka, S.R., 1996. A conditional saddlepoint approximation for testing problems. Preprint.
- Greenwood, M., 1946. The statistical study of infectious diseases. *J. Roy. Statist. Soc. Ser. A* 109, 85–110.
- Hall, P., 1992. *The Bootstrap And Edgeworth Expansion*. Springer, New York.
- Hill, I.D., 1979. Approximating The Distribution Of Greenwood's Statistic With Johnson Distributions. *J. Roy. Statist. Soc. Ser. A* 142, 378–380 (Corrigendum (1981), *J. Roy. Statist. Soc. Ser. A* 144, 388).
- Kabaila, P., 1993. A method for the computer calculation of Edgeworth expansions of smooth function models. *J. Comput. Graphical Statist.* 2(2), 199–207.
- Moran, P.A.P., 1947. The random division of an interval. *J. Roy. Statist. Soc. Ser. B* 9, 92–98 (Corrigendum (1981), *J. Roy. Statist. Soc. Ser. A* 144, 388).
- Moran, P.A.P., 1951. The random division of an interval II. *J. Roy. Statist. Soc. Ser. B* 13, 147–150.
- Pyke, R., 1965. Spacings (With Discussion). *J. Roy. Statist. Soc. Ser. B* 27(3), 395–449.
- Rao, J.S., 1976. Some tests based on arc-lengths for the circle. *Sankhyā: Indian J. Statist. Ser. B* 38, 329–338.
- Rao, J.S., Sethuraman, J., 1975. Weak Convergence Empirical Distribution Functions of Random Variables Subject to Perturbations and Scale Factors, *Annals of Statistics* 3(2), 299–313.
- Reid, N., 1988. Saddlepoint methods and statistical inference. *Statist. Sci.* 3(2), 213–238.
- Russell, G.S., Levitin, D.J., 1995. An expanded table of probability values for Rao's spacing test. *Commun. Statist. Simulation Comput.* 24(4), 879–888.
- Sethuraman, J., Rao, J.S., 1970. Pitman efficiencies of tests based on spacings. In: Puri, M.L. (Ed.), *Nonparametric Techniques in Statistical Inference*. Cambridge University Press, Cambridge, pp. 405–415.
- Stephens, M.A., 1981. Further percentage points for Greenwood's statistic. *J. Roy. Statist. Soc. Ser. A* 144(3), 364–366.
- Wang, S., 1992. General saddlepoint approximations in the bootstrap. *Statist. Probab. Lett.* 13, 61–66.