



HAL
open science

Saddlepoint approximations for the probability mass functions of some nonparametric test statistics

Cyrille Joutard

► **To cite this version:**

Cyrille Joutard. Saddlepoint approximations for the probability mass functions of some nonparametric test statistics. *Statistics and Probability Letters*, 2022, 188, pp.109544. 10.1016/j.spl.2022.109544 . hal-04529013

HAL Id: hal-04529013

<https://hal.science/hal-04529013>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro



Saddlepoint approximations for the probability mass functions of some nonparametric test statistics

Cyrille Joutard

Univ Paul-Valéry Montpellier 3 and IMAG, Univ Montpellier, CNRS, Montpellier, France



ARTICLE INFO

Article history:

Received 29 September 2021
Received in revised form 16 February 2022
Accepted 14 May 2022
Available online 23 May 2022

Keywords:

Saddlepoint approximation
Edgeworth expansion
Wilcoxon signed-rank statistic
Kendall's tau coefficient

ABSTRACT

We give classical saddlepoint approximations for the probability mass functions of two nonparametric test statistics, the Kendall's tau coefficient and the Wilcoxon signed-rank statistic. Then, we provide numerical comparisons, by comparing the exact probabilities with the saddlepoint approximation, a large deviation local limit theorem, an Edgeworth expansion and a normal approximation.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

In this paper, we give asymptotic approximations of saddlepoint type for the probability mass function of the statistics of two nonparametric tests, namely the Kendall's rank coefficient test ([Kendall 1938](#)) and the Wilcoxon signed-rank test ([Wilcoxon 1945](#)).

Several works have been dealing specifically with asymptotic approximations for the Wilcoxon signed-rank statistic (like, for example, [Fellingham and Stoker 1964](#), [Klotz 1965](#), [Sievers 1969](#), [Chaganty and Sethuraman 1985](#), [Quine 1994](#) and [Joutard 2020](#)) and the Kendall's tau coefficient (e.g., [David et al. 1951](#), [Sievers 1969](#), [Gatto 2009](#) and [Joutard 2020](#)). However, most of these works mainly deal with approximations and numerical comparisons of tail probabilities. [David et al. \(1951\)](#) and [Quine \(1994\)](#) gave some brief formulas of Edgeworth type for the probability mass functions of the Kendall's tau coefficient and of the Wilcoxon signed-rank statistic respectively, but without illustrating them with numerical studies.

Here, we also carry out numerical comparisons for the probabilities $\mathbb{P}(Z_n = a)$, where Z_n is the test statistic. For different values of a , these probabilities are compared with the saddlepoint approximation, a large deviation local limit theorem of [Chaganty and Sethuraman \(1985\)](#), an Edgeworth expansion (see [David et al. 1951](#) for the Kendall's tau coefficient and [Quine 1994](#) for the Wilcoxon statistic) and a normal approximation.

Besides the theoretical interest, note that these asymptotic approximations (especially of the saddlepoint and Edgeworth types) can also be of statistical interest, in particular for the probability mass function of the Kendall's tau coefficient, for which most statistical software can compute exact values only for small values of n (for example, for $n < 13$ in the case of the software R).

Saddlepoint approximations of densities were first studied by [Daniels \(1954\)](#), see also [Barndorff-Nielsen and Cox \(1979\)](#) regarding saddlepoint results for some multivariate and conditional densities, and [Blackwell and Hodges \(1959\)](#) for the particular case of the sample mean with discrete random variables.

E-mail address: cyrille.joutard@univ-montp3.fr.

<https://doi.org/10.1016/j.spl.2022.109544>

0167-7152/© 2022 Elsevier B.V. All rights reserved.

© 2022 published by Elsevier. This manuscript is made available under the CC BY NC user license

<https://creativecommons.org/licenses/by-nc/4.0/>

General information about asymptotic approximations (including saddlepoint results) for tail probabilities as well as for densities, can be found, e.g. in [Bhattacharya and Rao \(1976\)](#), [Barndorff-Nielsen and Cox \(1989\)](#), [Butler \(2007\)](#), [Jensen \(1995\)](#), [Kolassa \(1994\)](#) and [Petrov \(1995\)](#).

The paper is organized as follows. In Sections 2 and 3, we give the saddlepoint results and also present the other asymptotic approximations for the Kendall's tau coefficient and the Wilcoxon signed-rank statistic, respectively. Finally, Section 4 is devoted to the numerical comparisons for both test statistics.

2. The Kendall's tau coefficient

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a sequence of i.i.d. continuous random couples having distribution function $F(x, y)$ and let F_X and F_Y be the marginal distributions. The Kendall's tau coefficient K_n can be defined by

$$K_n = 2 \sum_{1 \leq i < j \leq n} \frac{(\mathbb{I}_{\{X_i \geq X_j\}} - \mathbb{I}_{\{X_i \leq X_j\}})(\mathbb{I}_{\{Y_i \geq Y_j\}} - \mathbb{I}_{\{Y_i \leq Y_j\}})}{n(n-1)}. \tag{1}$$

It was first used by [Kendall \(1938\)](#) to test the null hypothesis $H_0 : F(x, y) = F_X(x)F_Y(y)$ for all x, y .

Let $T_n = nK_n$. The m.g.f. of T_n under the null hypothesis is given below (see, e.g., [Kendall and Stuart 1979](#) or [Van de Wiel et al. 1999](#)),

$$\phi_K(t) = E(e^{tK_n}) = \frac{e^{nt}}{n!} \prod_{k=1}^n \frac{e^{-\frac{4kt}{n-1}} - 1}{e^{-\frac{4t}{n-1}} - 1}.$$

Then, the normalized c.g.f. of T_n is, for $t > 0$,

$$\begin{aligned} \kappa_K(t) &= n^{-1} \log \phi_K(t) \\ &= t - n^{-1} \log(n!) + n^{-1} \sum_{k=1}^n \log \left(1 - e^{-\frac{4kt}{n-1}} \right) - \log \left(1 - e^{-\frac{4t}{n-1}} \right). \end{aligned}$$

Denoting $u_{n,i}(t) = e^{-\frac{4t}{n-1}} / (1 - e^{-\frac{4t}{n-1}})^i$, the first and second derivatives of κ_K are

$$\begin{aligned} \kappa'_K(t) &= 1 + \frac{4}{n(n-1)} \sum_{k=1}^n k u_{n,1}(kt) - \frac{4}{n-1} u_{n,1}(t), \\ \kappa''_K(t) &= -\frac{16}{n(n-1)^2} \sum_{k=1}^n k^2 u_{n,2}(kt) + \frac{16}{(n-1)^2} u_{n,2}(t). \end{aligned}$$

The expressions for the third and the fourth derivatives can be found below:

$$\begin{aligned} \kappa^{(3)}_K(t) &= -\frac{64}{n(n-1)^3} \sum_{k=1}^n k^3 (u_{n,2}(kt) - 2u_{n,3}(kt)) + \frac{64}{(n-1)^3} (u_{n,2}(t) - 2u_{n,3}(t)), \\ \kappa^{(4)}_K(t) &= -\frac{256}{n(n-1)^4} \sum_{k=1}^n k^4 (u_{n,2}(kt) + 6e^{-\frac{4kt}{n(n-1)}} u_{n,4}(kt)) + \frac{256}{(n-1)^4} (u_{n,2}(t) \\ &\quad + 6e^{-\frac{4t}{n(n-1)}} u_{n,4}(t)). \end{aligned}$$

Using these derivatives, we can state the saddlepoint approximation in the next section.

2.1. Saddlepoint approximation

In the second application example of their main theorem for lattice-valued random variables, [Chaganty and Sethuraman \(1985\)](#) gave a large deviation local limit approximation for the mass function of the Kendall's tau coefficient (this is actually the same approximation as the one below without the terms of order n^{-1}). Under H_0 , a "full" saddlepoint approximation (including the terms of order n^{-1}) for the probability mass function of K_n is given below.

Theorem 1. *Let K_n be defined as in (1). Then for n large enough,*

$$\mathbb{P}(K_n = a) = \frac{4}{n-1} \frac{\exp(-n\kappa_K^*(a))}{\sqrt{2\pi n\kappa''_K(\tau_a)}} \left(1 + \frac{P_{2K}(\tau_a)}{n} + O(n^{-3/2}) \right) \tag{2}$$

where the saddle point τ_a is such that $\kappa'_K(\tau_a) = a$. Furthermore, $\kappa_K^*(a) = \tau_a a - \kappa_K(\tau_a)$ and

$$P_{2K}(x) = \frac{\kappa_K^{(4)}(x)}{8(\kappa''_K(x))^2} - \frac{5(\kappa_K^{(3)}(x))^2}{24(\kappa''_K(x))^3}.$$

Proof. In order to prove (2), we will use arguments from proofs of Chaganty and Sethuraman (1985, Theorem 2.2) and Jin and Robinson (2003, Theorem 1). First, let us define the tilted variable $T_n^* = nK_n^*$ by

$$\mathbb{P}(T_n^* = na) = e^{-n\kappa_K(\tau_a) + \tau_a na} \mathbb{P}(T_n = na). \tag{3}$$

Recall that the moment generating function of T_n^* is

$$\frac{\phi_K(\tau_a + it)}{\phi_K(\tau_a)} = e^{-n\kappa_K(\tau_a)} \phi_K(\tau_a + it). \tag{4}$$

Therefore, by applying the inversion formula (see, e.g., Feller 1966) to the lattice valued random variable T_n^* with span $4/(n - 1)$, and using (3) and (4), we have

$$\mathbb{P}(K_n = a) = \mathbb{P}(T_n = na) = \frac{4}{(n - 1)2\pi} \int_{-\frac{\pi(n-1)}{4}}^{\frac{\pi(n-1)}{4}} \phi_K(\tau_a + it) e^{-(\tau_a + it)na} dt$$

Then,

$$\sqrt{n} \left(\frac{n - 1}{4} \right) \mathbb{P}(K_n = a) = \left(\frac{1}{2\pi \kappa_K''(\tau_a)} \right)^{1/2} e^{-n\kappa_K^*(a)} I_n \tag{5}$$

where $\kappa_K^*(a) = \tau_a a - \kappa_K(\tau_a)$ and

$$I_n = \left(\frac{n\kappa_K''(\tau_a)}{2\pi} \right)^{1/2} \int_{-\frac{\pi(n-1)}{4}}^{\frac{\pi(n-1)}{4}} e^{-nG_n(t)} dt,$$

with $G_n(t) = -\kappa_K(\tau_a + it) + \kappa_K(\tau_a) + ita$. Choosing a positive constant λ such that $\lambda < 1/2$, let

$$\begin{aligned} I_n &= \left(\frac{n\kappa_K''(\tau_a)}{2\pi} \right)^{1/2} \left[\int_{|t| < n^{-\lambda}} e^{-nG_n(t)} dt + \int_{n^{-\lambda} < |t| < \frac{\pi(n-1)}{4}} e^{-nG_n(t)} dt \right] \\ &= I_{n1} + I_{n2}. \end{aligned} \tag{6}$$

Using the same arguments as Chaganty and Sethuraman (1985), we can show that there exists a constant $\alpha > 0$ such that

$$|I_{n2}| \leq O(n) \exp(-\alpha n^{1-2\lambda}/4), \tag{7}$$

which goes exponentially fast to zero since $\lambda < 1/2$. In fact, the assumptions (A), (B), (C'), (D') of Chaganty and Sethuraman (1985, Theorem 2.2), needed to prove (7), hold for the Kendall's tau coefficient (see Chaganty and Sethuraman 1985, Example 3.2). In particular we use the fact that there exists $\alpha > 0$ such that $\inf_{n^{-\lambda} < |t| < \pi(n-1)/4} \text{Re}(G_n(t)) \geq \alpha n^{-2\lambda}/4$.

For the first term I_{n1} , by an expansion of G_n (see, e.g., the proof of Jin and Robinson 2003, Theorem 1) and using the change of variables $v = t\sqrt{n\kappa_K''(\tau_a)}$, we have

$$\begin{aligned} I_{n1} &= \frac{1}{\sqrt{2\pi}} \int_{|v| < n^{1/2-\lambda} \sqrt{\kappa_K''(\tau_a)}} e^{-\frac{v^2}{2}} \left[1 - \frac{1}{\sqrt{n}} \frac{iv^3 \kappa_K^{(3)}(\tau_a)}{3!(\kappa_K''(\tau_a))^{3/2}} + \frac{1}{\sqrt{n}} \left(\frac{v^4 \kappa_K^{(4)}(\tau_a)}{4!(\kappa_K''(\tau_a))^2} \right. \right. \\ &\quad \left. \left. - \frac{v^6 (\kappa_K^{(3)}(\tau_a))^2}{2 \times 6^2 (\kappa_K''(\tau_a))^3} \right) + |v|^5 e^{\frac{v^2}{4}} O(n^{-3/2}) \right] \\ &= 1 + \frac{1}{n} \left(\frac{3\kappa_K^{(4)}(\tau_a)}{24(\kappa_K''(\tau_a))^2} - \frac{15(\kappa_K^{(3)}(\tau_a))^2}{72(\kappa_K''(\tau_a))^3} \right) + O(n^{-3/2}). \end{aligned} \tag{8}$$

Combining (5), (6), (7) and (8), we eventually obtain (2). □

2.2. Normal approximation

Let ϕ_G be the standard normal density function, and Φ_G the corresponding cumulative distribution function. Under H_0 , the expectation and the variance of K_n are

$$E(K_n) = 0 \quad \text{and} \quad V(K_n) = \frac{2(2n + 5)}{9n(n - 1)}.$$

Besides, the span of the lattice-valued random variable K_n is equal to $4/n(n - 1)$. Then, the normal approximation (including a continuity correction) for the mass function of the Kendall's tau coefficient is given below:

$$P(K_n = a) \sim \Phi_G \left(\frac{a + 2/n(n - 1)}{\sqrt{(2(2n + 5))/(9n(n - 1))}} \right) - \Phi_G \left(\frac{a - 2/n(n - 1)}{\sqrt{(2(2n + 5))/(9n(n - 1))}} \right), \tag{9}$$

where \sim denotes the asymptotic equivalence.

2.3. Edgeworth expansion

Let $S_n = \sum_{1 \leq i < j \leq n} (\mathbb{I}_{\{X_i \geq X_j\}} - \mathbb{I}_{\{X_i \leq X_j\}})(\mathbb{I}_{\{Y_i \geq Y_j\}} - \mathbb{I}_{\{Y_i \leq Y_j\}})$. David et al. (1951) gave an Edgeworth expansion with seven terms for the mass function of S_n . First, denoting by κ_i the i th cumulant (under H_0) of S_n , we have

$$\begin{aligned} \kappa_2 &= \frac{n(n-1)(2n+5)}{18}, \\ \kappa_4 &= -\frac{n(6n^4 + 15n^3 + 10n^2 - 31)}{225}, \\ \kappa_6 &= \frac{8n(6n^6 + 21n^5 + 21n^4 - 7n^2 - 41)}{1323}, \\ \kappa_8 &= -\frac{8n(10n^8 + 45n^7 + 60n^6 - 42n^4 + 20n^2 - 93)}{675}. \end{aligned}$$

Note that the cumulants of odd order are null. Next, let us define the standardized cumulants (under H_0) $\tilde{\kappa}_i$ of S_n by

$$\tilde{\kappa}_i = \frac{\kappa_i}{\kappa_2^{i/2}}. \tag{10}$$

Using the fact that the span of the lattice valued random variable $S_n = n(n-1)K_n/2$ is 2, the Edgeworth approximation of David et al. (1951) can therefore be expressed as follows:

$$\begin{aligned} P(S_n = k) &\sim 2\kappa_2^{-1/2} \phi_C(x_k) \left(1 + \frac{\tilde{\kappa}_4}{24} H_4(x_k) + \frac{\tilde{\kappa}_6}{720} H_6(x_k) + \frac{\tilde{\kappa}_4^2}{1152} H_8(x_k) \right. \\ &\quad \left. + \frac{\tilde{\kappa}_8}{40320} H_8(x_k) + \frac{\tilde{\kappa}_6 \tilde{\kappa}_4}{17280} H_{10}(x_k) + \frac{\tilde{\kappa}_4^3}{82944} H_{12}(x_k) \right) \end{aligned} \tag{11}$$

where $x_k = k/\sqrt{\kappa_2}$, the standardized cumulants $\tilde{\kappa}_i$ are defined in (10) and the Hermite polynomials are defined as follows,

$$\begin{aligned} H_4(x) &= x^4 - 6x^2 + 3, \\ H_6(x) &= x^6 - 15x^4 + 45x^2 - 15, \\ H_8(x) &= x^8 - 28x^6 + 210x^4 - 420x^2 + 105, \\ H_{10}(x) &= x^{10} - 45x^8 + 630x^6 - 3150x^4 + 4725x^2 - 945 \\ &\text{and} \\ H_{12}(x) &= 2^{-6} \tilde{H}_{12}(x/\sqrt{2}) \end{aligned}$$

where

$$\tilde{H}_{12}(x) = 4096x^{12} - 135168x^{10} + 1520640x^8 - 7096320x^6 + 13305600x^4 - 7983360x^2 + 665280.$$

3. The Wilcoxon signed-rank statistic

Let X_1, \dots, X_n be a sequence of i.i.d. absolutely continuous random variables having distribution function F and let R_i be the rank of $|X_i|$ among $|X_1|, \dots, |X_n|$. In other words, if one arranges $|X_1|, |X_2|, \dots, |X_n|$ in increasing order of magnitude, R_i denotes the rank of $|X_i|$. Assume that the random variables X_i are symmetric about their median m . The Wilcoxon signed-rank statistic W_n is defined as the sum of the quantities R_i corresponding to the positive X_i 's, that is,

$$W_n = \sum_{i=1}^n \mathbb{I}_{\{X_i > 0\}} R_i. \tag{12}$$

The statistic W_n was introduced by Wilcoxon (1945) and is used to test the null hypothesis $H_0 : m = 0$.

Let $T_n = W_n/n$. Then, under the null hypothesis H_0 , the moment generating function (m.g.f.) of T_n is given below (see, e.g., Van de Wiel et al. 1999),

$$\phi_W(t) = E(e^{tW_n/n}) = \prod_{k=1}^n \frac{\exp(tk/n) + 1}{2}, \quad t \in \mathbb{R}.$$

Next, the normalized cumulant generating function (c.g.f.) of W_n/n is

$$\begin{aligned} \kappa_W(t) &= n^{-1} \log \phi_W(t) \\ &= n^{-1} \sum_{k=1}^n \log \left(\frac{\exp(tk/n) + 1}{2} \right). \end{aligned}$$

The first and second derivatives of κ_W are

$$\kappa'_W(t) = n^{-2} \sum_{k=1}^n \frac{k \exp(tk/n)}{\exp(tk/n) + 1} \quad \text{and} \quad \kappa''_W(t) = n^{-3} \sum_{k=1}^n \frac{k^2 \exp(tk/n)}{(\exp(tk/n) + 1)^2}.$$

We will also need the third and the fourth derivatives which can be written as follows after simplification:

$$\kappa^{(3)}_W(t) = n^{-4} \sum_{k=1}^n \frac{k^3 \exp(tk/n) (1 - \exp(tk/n))}{(\exp(tk/n) + 1)^3}$$

and

$$\kappa^{(4)}_W(t) = n^{-5} \sum_{k=1}^n \frac{k^4 \exp(tk/n)}{(\exp(tk/n) + 1)^2} \left(1 - \frac{6 \exp(tk/n)}{(\exp(tk/n) + 1)^2} \right).$$

Using these derivatives, we can present the saddlepoint approximation in the next section.

3.1. Saddlepoint approximation

As for the Kendall's tau coefficient, [Chaganty and Sethuraman \(1985\)](#) gave a large deviation approximation (which is also the same approximation as the one below without the terms of order n^{-1}) for the mass function of the Wilcoxon statistic in the first application example of their main theorem.

Under H_0 , we have the following saddlepoint approximation for the mass function of $T_n = W_n/n$.

Theorem 2. Let $T_n = W_n/n$ where W_n is defined in (12). Then for n large enough,

$$\mathbb{P}(T_n = nb) = \frac{\exp(-n\kappa_W^*(b))}{n\sqrt{2\pi n\kappa''_W(\tau_b)}} \left(1 + \frac{P_{2W}(\tau_b)}{n} + O(n^{-3/2}) \right) \tag{13}$$

where the saddle point τ_b is such that $\kappa'_W(\tau_b) = b$. Furthermore, $\kappa_W^*(b) = \tau_b b - \kappa_W(\tau_b)$ and

$$P_{2W}(x) = \frac{\kappa^{(4)}_W(x)}{8(\kappa''_W(x))^2} - \frac{5(\kappa^{(3)}_W(x))^2}{24(\kappa''_W(x))^3}.$$

Proof. As for the Kendall's tau coefficient, we first have to define the tilted variable $T_n^* = W_n^*/n$ by

$$\mathbb{P}(T_n^* = nb) = e^{-n\kappa_W(\tau_b) + \tau_b nb} \mathbb{P}(T_n = nb).$$

By applying the inversion formula to the lattice valued random variable T_n^* with span n^{-1} , we have

$$\mathbb{P}(W_n = n^2b) = \mathbb{P}(T_n = nb) = \frac{1}{2\pi n} \int_{-\pi n}^{\pi n} \phi_W(\tau_b + it) e^{-(\tau_b + it)nb} dt.$$

The rest of the proof parallels the one in Section 2.1. In particular, the assumptions (A), (B), (C'), (D') of [Chaganty and Sethuraman \(1985, Theorem 2.2\)](#), also hold for the Wilcoxon statistic (see [Chaganty and Sethuraman 1985, Example 3.1](#)). \square

3.2. Normal approximation

Under H_0 , the expectation and the variance of W_n are

$$E(W_n) = \frac{n(n+1)}{4} \quad \text{and} \quad V(W_n) = \frac{n(n+1)(2n+1)}{24}. \tag{14}$$

Then, the normal approximation (including a continuity correction) for the Wilcoxon signed-rank statistic is given below:

$$P(W_n = a) \sim \Phi_G \left(\frac{a + 1/2 - \mu_W}{\sqrt{\sigma_W^2}} \right) - \Phi_G \left(\frac{a - 1/2 - \mu_W}{\sqrt{\sigma_W^2}} \right), \tag{15}$$

where $\mu_W = E(W_n)$ and $\sigma_W^2 = V(W_n)$ are given in (14).

Table 1
Approximations of $\mathbb{P}(K_n = a)$ for the Kendall's tau coefficient ($n = 10$).

k	1	3	5	7	11	15	17	19
a	0.0222	0.0667	0.1111	0.1556	0.2444	0.3333	0.3778	0.4222
<i>Exact</i>	0.06910	0.06716	0.06342	0.05817	0.04474	0.03032	0.02371	0.01788
Pr_{NORM}	0.07099	0.06876	0.06450	0.05862	0.04398	0.02905	0.02250	0.01688
Pr_{EDG}	0.06910	0.06716	0.06342	0.05817	0.04474	0.03031	0.02371	0.01788
Pr_{LD}	0.07111	0.06911	0.06526	0.05985	0.04601	0.03116	0.02436	0.01836
Pr_{SAD}	0.06919	0.06724	0.06350	0.05824	0.04479	0.03035	0.02374	0.01790

k	21	23	25	27	31	35	39	43
a	0.4667	0.5111	0.5556	0.6000	0.6889	0.7778	0.8667	0.9556
<i>Exact</i>	0.01296	0.00901	0.00597	0.00376	0.00124	0.00029	4.27e−05	2.48e−06
Pr_{NORM}	0.01227	0.00864	0.00589	0.00389	0.00154	0.00054	1.65e−04	4.46e−05
Pr_{EDG}	0.01296	0.00900	0.00597	0.00376	0.00124	0.00030	3.98e−05	−4.80e−07
Pr_{LD}	0.01330	0.00924	0.00612	0.00385	0.00126	0.00030	4.40e−05	2.69e−06
Pr_{SAD}	0.01297	0.00901	0.00597	0.00376	0.00124	0.00029	4.27e−05	2.46e−06

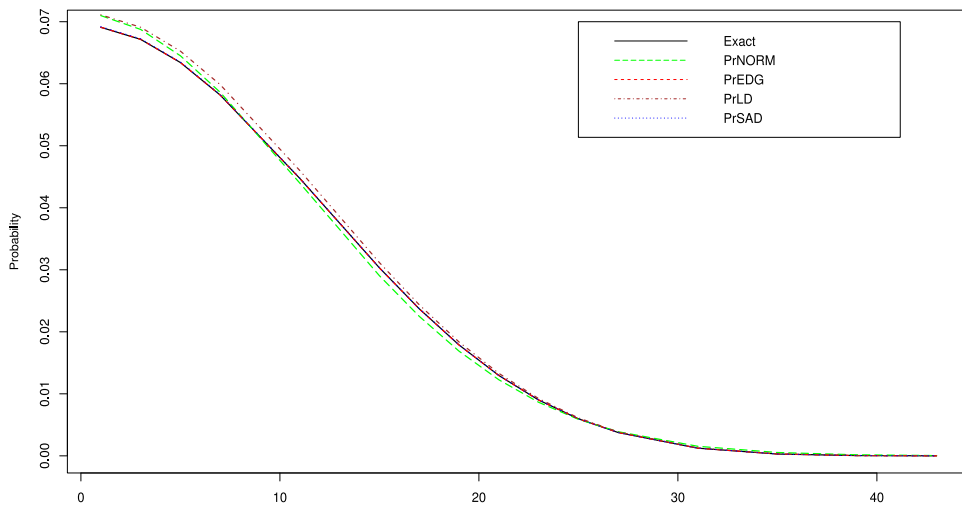


Fig. 1. Approximations of $\mathbb{P}(K_n = a)$ for the Kendall's tau coefficient ($n = 10$).

3.3. Edgeworth expansion

Quine (1994) gave an Edgeworth expansion (including a remainder term of order $n^{-7/2}$) for the mass function of W_n :

$$P(W_n = a) = \phi_G(x_a) \left(1 - \frac{H_4(x_a)(6n^5 + 15n^4 + 10n^3 - n)}{5760\sigma_W^2} \right) + O\left(\frac{1}{n^{7/2}}\right), \tag{16}$$

where $x_a = (a - \mu_W)/\sqrt{\sigma_W^2}$, $\mu_W = E(W_n)$ and $\sigma_W^2 = V(W_n)$ are given in (14), and the Hermite polynomial H_4 is defined as $H_4(x) = x^4 - 6x^2 + 3$.

4. Numerical comparisons

4.1. For the Kendall's tau coefficient

The numerical comparisons are given in Table 1 and illustrated in Fig. 1 (note that for this last one, we chose a curve graph for a better visualization of the different approximations). We consider the case $n = 10$ and compare the exact probabilities $\mathbb{P}(K_n = a)$, where $a = 2k/n(n - 1)$, for different values of k ranging from 1 to 43, with the normal approximation (9), the Edgeworth expansion (11) of David et al. (1951), the large deviation local limit result of Chaganty and Sethuraman (1985) which is obtained from Eq. (2) by excluding the $O(1/n)$ terms, and the “full” saddlepoint approximation (2).

Note that for the numerical computations, we used R and the exact tail probabilities were obtained by using the function *dKendall* of the R package *SuppDists*.

Table 2
Approximations of $\mathbb{P}(W_n = a)$ for the Wilcoxon signed-rank statistic ($n = 10$).

a	1	5	10	15	20	23	26	28
<i>Exact</i>	0.00098	0.00293	0.00977	0.01953	0.03027	0.03516	0.03809	0.03906
<i>Pr_{NORM}</i>	0.00106	0.00294	0.00829	0.01806	0.03035	0.03659	0.04017	0.04059
<i>Pr_{EDG}</i>	0.00087	0.00297	0.00899	0.01911	0.03043	0.03567	0.03855	0.03888
<i>Pr_{LD}</i>	0.00115	0.00337	0.00981	0.02024	0.03182	0.03725	0.04027	0.04062
<i>Pr_{SAD}</i>	0.00090	0.00303	0.00918	0.01920	0.03037	0.03563	0.03855	0.03888

a	31	34	38	41	43	46	49	53
<i>Exact</i>	0.03711	0.03223	0.02344	0.01660	0.01270	0.00781	0.00391	0.00098
<i>Pr_{NORM}</i>	0.03814	0.03264	0.02294	0.01578	0.01168	0.00688	0.00369	0.00139
<i>Pr_{EDG}</i>	0.03693	0.03239	0.02377	0.01685	0.01263	0.00743	0.00383	0.00123
<i>Pr_{LD}</i>	0.03857	0.03384	0.02499	0.01793	0.01359	0.00817	0.00431	0.00147
<i>Pr_{SAD}</i>	0.03690	0.03233	0.02378	0.01697	0.01280	0.00761	0.00392	0.00123

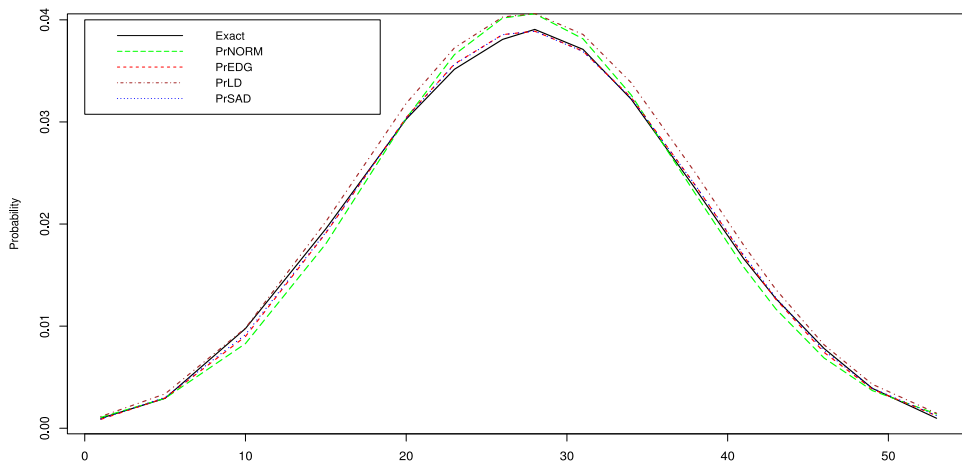


Fig. 2. Approximations of $\mathbb{P}(W_n = a)$ for the Wilcoxon signed-rank statistic ($n = 10$).

As we have already noticed for the approximation of tail probabilities of K_n in Joutard (2020), the Edgeworth expansion of David et al. (1951), including all the seven terms, is very accurate here and only fails for the last value ($k=43$) of our example (with a negative probability). It is actually the best approximation for $k = 1, \dots, 25$ while the saddlepoint approximation is slightly better for $k = 27, 31, 35$. As expected, this last one is accurate everywhere and beats all other approximations for $k = 27, \dots, 43$.

Finally, the normal approximation and the large deviation local limit theorem generally give correct results. Pr_{NORM} is a little better than Pr_{LD} for values near the center of the distribution ($k = 1, 3, 5, 7, 11$) and for $k = 25$ while Pr_{LD} beats the normal approximation for all other values (in particular for extreme ones).

4.2. For the Wilcoxon statistic

The numerical comparisons are given in Table 2 and illustrated in Fig. 2. As in the previous section, we consider the case $n = 10$ and compare the exact probabilities $\mathbb{P}(W_n = a)$, for different values of a ranging from 1 to 53, with the normal approximation (15), the Edgeworth expansion (16) of Quine (1994), the large deviation result of Chaganty and Sethuraman (1985) which is obtained from Eq. (13) by excluding the $O(1/n)$ terms, and the “full” saddlepoint approximation (13) (with $b = a/n^2$).

Here, the exact tail probabilities were obtained with the R function *dsignrank*.

Once again, the saddlepoint approximation and the Edgeworth expansion give the best results for most values, but this time with a slight advantage for the saddlepoint expression. In fact, the most accurate approximation is Pr_{SAD} for $a = 1, 15, 23, 26, 34, 46, 49$ while Pr_{EDG} beats all other approximations for $a = 28, 31, 38, 41, 43, 53$. The large deviation result is the sharpest approximation only for $a = 10$, while the normal approximation does a pretty good job here and outperforms even all other approximations for $a = 5, 20$.

References

Barndoff-Nielsen, O.E., Cox, D.R., 1979. Edgeworth and saddlepoint approximations with statistical applications. *J. R. Stat. Soc. Ser. B* 41, 279–312.
 Barndoff-Nielsen, O.E., Cox, D.R., 1989. *Asymptotic Techniques for Uses in Statistics*. Chapman and Hall.

- Bhattacharya, R.N., Rao, R.R., 1976. Normal Approximation and Asymptotic Expansions. John Wiley, New York.
- Blackwell, D., Hodges, J.L., 1959. The probability in the extreme tail of a convolution. *Ann. Math. Stat.* 30, 1113–1120.
- Butler, R.W., 2007. Saddlepoint Approximations with Applications. Cambridge University Press.
- Chaganty, N.R., Sethuraman, J., 1985. Large deviation local limit theorems for arbitrary sequences of random variables. *Ann. Probab.* 13, 97–114.
- Daniels, H.E., 1954. Saddlepoint approximations in statistics. *Ann. Math. Stat.* 25, 631–650.
- David, S.T., Kendall, M.G., Stuart, A., 1951. Some questions of distribution in the theory of rank correlation. *Biometrika* 38, 131–140.
- Feller, W., 1966. An Introduction to Probability Theory and Its Applications. II. John Wiley & Sons.
- Fellingham, S.A., Stoker, D.J., 1964. An approximation for the exact distribution of the wilcoxon test for symmetry. *J. Amer. Statist. Assoc.* 59, 899–905.
- Gatto, R., 2009. Asymptotic approximations to the distribution of Kendall's sample τ . *J. Stat. Comput. Simul.* 79, 671–679.
- Jensen, J.L., 1995. Saddlepoint Approximations. Oxford University Press.
- Jin, R., Robinson, J., 2003. Saddlepoint approximations of the two-sample wilcoxon statistic. *Lecture Notes-Monograph Series* 43, 149–158.
- Joutard, C., 2020. Asymptotic tail approximations for some nonparametric test statistics. *J. Nonparametr. Stat.* 32, 756–768.
- Kendall, M.G., 1938. A new measure of rank correlation. *Biometrika* 30, 81–89.
- Kendall, M.G., Stuart, A., 1979. The Advanced Theory of Statistics, II. MacMillan, New-York.
- Klotz, J., 1965. Alternative efficiencies for signed rank tests. *Ann. Math. Stat.* 36, 1759–1766.
- Kolassa, J., 1994. Series Approximation Methods in Statistics. In: *Lecture Notes in Statistics*, vol. 88, Springer-Verlag, New York.
- Petrov, V.V., 1995. Limit Theorems of Probability Theory: Sequences of Independent Random Variables. Clarendon Press, Oxford.
- Quine, M.P., 1994. Probability approximations for divisible discrete distributions. *Austr. J. Stat.* 36, 339–349.
- Sievers, G.L., 1969. On the probability of large deviations. *Ann. Math. Stat.* 40, 1908–1921.
- Van de Wiel, M.A., Bucchianico, A.Di., Laan, P.Van.der., 1999. Symbolic computation and exact distributions of nonparametric test statistics. *J. R. Stat. Soc. Ser. D* 48, 507–516.
- Wilcoxon, F., 1945. Individual comparison by ranking methods. *Biometrics* 1, 80–83.