

Article

# Discrete Parameter-Free Zone Distribution and Its Application in Normality Testing

Atif Avdović \*  and Vesna Jevremović

Department of Science and Mathematics, State University of Novi Pazar, Vuka Karadžića 9, 36300 Novi Pazar, Serbia

\* Correspondence: aavdovic@np.ac.rs

**Abstract:** In recent research endeavors, discrete models have gained considerable attention, even in cases where the observed variables are continuous. These variables can often be effectively approximated by a normal distribution. Given the prevalence of processes requiring robust quality control, models associated with the normal distribution have found widespread applicability; nevertheless, there remains a persistent need for enhanced accuracy in normality analysis, prompting the exploration of novel and improved solutions. This paper introduces a discrete parameter-free distribution linked to the normal distribution, derived from a quality control methodology rooted in the renowned ‘3-sigma’ rule. The development of a novel normality test, based on this distribution, is presented. A comprehensive examination encompasses mathematical derivation, distribution tables generated through Monte Carlo simulation studies, properties, power analysis, and comparative analysis, all with key features illustrated graphically. Notably, the proposed normality test surpasses conventional methods in performance. Termed the ‘Zone distribution’, this newly introduced distribution, along with its accompanying ‘Zone test’, demonstrates superior efficacy through illustrative examples. This research contributes a valuable tool to the field of normality analysis, offering a robust alternative for applications requiring precise and reliable assessments.

**Keywords:** zone; 3-sigma rule; discrete distribution; quality control; normal distribution; power of the test

**MSC:** 62E10; 62E17; 62G10; 62G30; 62Q05



**Citation:** Avdović, A.; Jevremović, V. Discrete Parameter-Free Zone Distribution and Its Application in Normality Testing. *Axioms* **2023**, *12*, 1087. <https://doi.org/10.3390/axioms12121087>

Academic Editor: Tomas Ruzgas

Received: 30 October 2023

Revised: 21 November 2023

Accepted: 22 November 2023

Published: 28 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Existing and well-established solutions concerning discrete distributions, such as the Binomial distribution, Geometric distribution, and Poisson distribution, are still extensively in use [1]; however, new approaches are frequently being developed as solutions to theoretical as well as empirical problems [2–8]. In the past two years, most attention has been directed toward describing and forecasting COVID-19 time series data [3,7,8]. However, other areas of application, such as medicine and agriculture [2], quality control [4], genetics and biology [7], life index, reading accuracy and intelligence [9], the failure times of devices and electronic components [2,3,9], age dependency ratio modelling [10], etc., have not been neglected.

In the authors’ last paper, a new discrete distribution was developed and its application in the new Quantile-Zone normality test was elaborated on. While that distribution has proven exceptionally useful in normality testing, its application in other contexts is challenging. This difficulty arises from its connection to the empirical distribution function (EDF), making it impractical or extremely complicated to determine functional and numerical characteristics [4]. The idea for such an approach was derived from research in quality control that the authors have also been working on [11,12]. Essentially, the concept of the X-bar chart has been modified to give new approaches in control charts, Quantile-Zone distribution, and normality testing. These findings have been crucial to this paper as the

newly introduced Zone distribution relies on a similar zoning concept inspired by the ‘3-sigma’ ( $3\sigma$ ) rule and is applied for normality testing. The research suggests that both the Quantile-Zone distribution and Zone distribution hold significant potential for applications in quality control. This recognition aligns with the growing importance of quality control as a notable research topic, both in theory and practical application [11–17].

Another significant subject pertains to distributions associated with the normal distribution [4,9,10], as it is one of the most commonly observed distributions in empirical variables, such as those found in nature, medicine, engineering, and other fields. Additionally, the reliability of parametric statistical analysis methods depends on the normality of the referent variables. The normal distribution and its properties, graphical and other methods of preliminary normality analysis, as well as normality tests, have all been analyzed, elaborated on, and widely used [18]; nevertheless, given its significance, it appears that research on the normal distribution, particularly regarding normality tests, is far from reaching conclusion. This paper comprehensively addresses these issues. Specifically, a novel discrete distribution, linked to the normal distribution through  $3\sigma$ -rule zoning, derived from a modified Shewhart-type control chart approach, has been formulated. This development yields noteworthy outcomes in the domain of quality control. Lastly, the resulting distribution is employed in a novel normality test, adding to the advancements in this field.

When it comes to normality tests, the most used, as well as other known tests, have been a topic of discussion [4,18,19], in which their properties, advantages, disadvantages, and power analyses are elaborated on in detail. Extensive power and comparative analysis have been provided as well [4,19,20]. Power analysis, in cases of less usual alternative distributions, has also been discussed [21–23]. Using parameter estimates, which better reflect actual values than assumed values, and employing the cumulative distribution function (CDF) of the test statistics in the null hypothesis, will enhance the power of normality tests. Such an approach has shown to be very efficient with the Quantile-Zone test [4] and Lilliefors test [19]. New approaches are constantly suggested, implemented, improved on, and analyzed in order to determine and emphasize their advantages over existing ones [4,19,22–26]. Some insights into the proficient mathematical development of goodness-of-fit tests, as well as their properties, are also available [22,27–29].

This study focuses on leveraging the Zone distribution to describe potential models arising in various practical applications. The consequential Zone test is employed to enhance normality analysis due to its simplicity and superior performance compared to many conventional tests. The outcomes of this research have broader implications for improving quality control achieved by control charts. The relevance of these innovations is underscored by the findings reported in recent publications [2–17,19,21,26,29].

In this section, a brief display of the previous and current results on topics of importance is offered. In Section 2, the definition and mathematical development of the functional characteristics of the new Zone distribution, with graphical illustrations, as well as some of its most important numerical characteristics, are given. In Section 3, the use of the Zone distribution application in normality testing with extensive power and comparative analysis, with highly illustrative graphics, is elaborated on. In Section 4, a brief insight into the application in normality testing via the Zone test, is given via examples. Finally, Section 5 concludes the paper with some important remarks on the results obtained.

## 2. Zone Distribution

### 2.1. Motivation and Definition

Let

$$X_1, X_2, \dots, X_n; n \in \mathbb{N}$$

be iid variables, where  $X_j; j = 1, \dots, n$  has the normal  $N(\mu, \sigma^2)$  distribution.

The  $3\sigma$  rule can be interpreted in two variants: standard and adjusted, as illustrated in Figure 1.

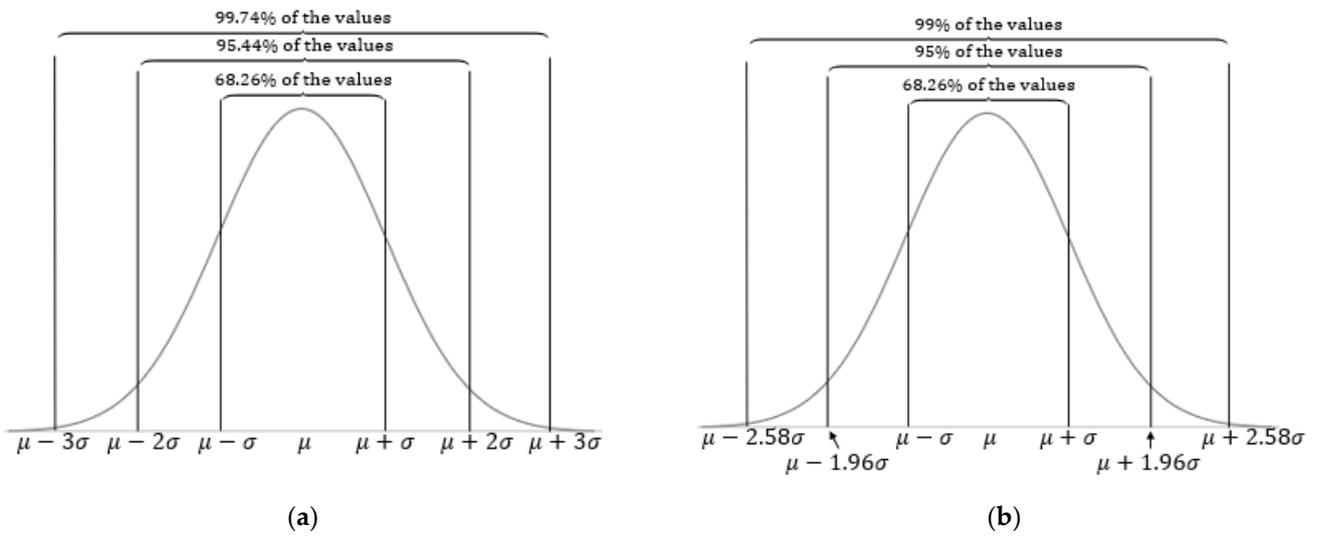


Figure 1. The 3σ rule. (a) Standard variant. (b) Adjusted variant.

Using the 3σ rule in the adjusted variant, one can obtain the following definition.

**Definition 1.** The zone function is given with the equation

$$zone(x) = \begin{cases} 1; & \mu - \sigma \leq x < \mu + \sigma \\ 1.96; & \mu - 1.96\sigma \leq x < \mu - \sigma \vee \mu + \sigma < x \leq \mu + 1.96\sigma \\ 2.58; & \mu - 2.58\sigma \leq x < \mu - 1.96\sigma \vee \mu + 1.96\sigma < x \leq \mu + 2.58\sigma \\ 2.81; & x < \mu - 2.58\sigma \vee x > \mu + 2.58\sigma \end{cases}, \quad x \in \mathbb{R}.$$

The zone function is graphically illustrated in Figure 2.

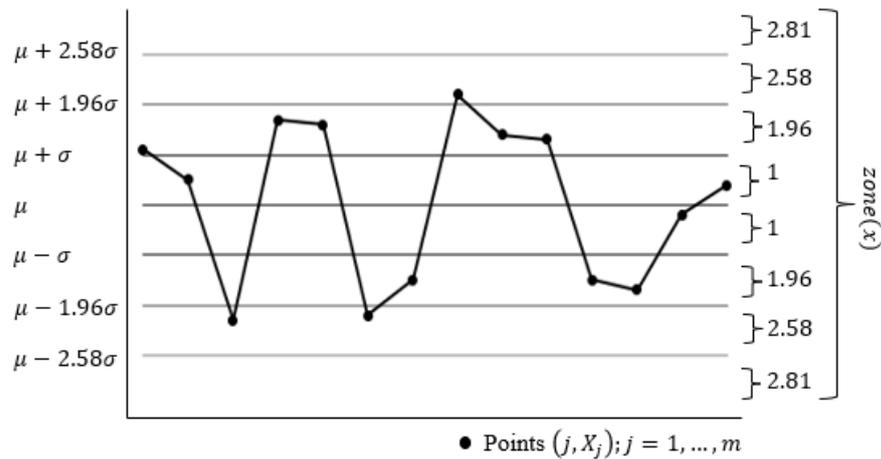


Figure 2. Graphical illustration of the zone function.

**Remark 1.** The choice of the value 2.81 is explained by the expressions

$$F^{-1} \left( \frac{F(2.58) + \lim_{x \rightarrow +\infty} F(x)}{2} \right) = 2.81$$

and

$$F^{-1} \left( \frac{\lim_{x \rightarrow -\infty} F(x) + F(-2.58)}{2} \right) = -2.81$$

where  $F$  is the CDF of the normal distribution with mean 0 and variance 1. Though any other value can be taken as well, this way, the analogy with the  $3\sigma$  rule in the standard form is ensured.

In Figure 2, one can see that the idea of the zone function definition has been developed by modification of Shewhart’s approach to quality control via  $\bar{X}$ -chart. The idea is to apply the zone function on all sample elements, and to quantify the discrepancy of each sample element from the central line  $x = \mu$ .

Using the zone function, we obtain the sample

$$zone(X_1), zone(X_2), \dots, zone(X_n) \tag{1}$$

(see Figure 2). The adjusted variant of the  $3\sigma$  rule implies the following proposition.

**Proposition 1.** *The distribution of the variable  $zone(X_j); j = 1, \dots, n$  (rounded to two decimal places) is given by*

$$zone(X_j) : \begin{pmatrix} 1 & 1.96 & 2.58 & 2.81 \\ 0.68 & 0.27 & 0.04 & 0.01 \end{pmatrix}.$$

Note that the following equations hold.

$$(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6826;$$

$$P(\mu - 1.96\sigma \leq X < \mu - \sigma) + P(\mu + \sigma < X \leq \mu + 1.96\sigma) = 0.2673;$$

$$P(\mu - 58\sigma \leq X < \mu - 1.96\sigma) + P(\mu + 1.96\sigma < X \leq \mu + 2.58\sigma) = 0.0401;$$

$$P(X > \mu + 2.58) + P(X < \mu - 2.58) = 0.0099.$$

**Definition 2.** *The discrete random variable defined as*

$$A = \frac{1}{n} \sum_{i=1}^n zone(X_i) \tag{2}$$

has the **Zone distribution with  $n$  degrees of freedom**, which will be denoted by  $A \sim zone_n$ .

**Theorem 1.** *Let*

$$C(n_1, n_{1.96}, n_{2.58}, n_{2.81}) = \left\{ \underbrace{1, 1, \dots, 1}_{n_1}, \underbrace{1.96, 1.96, \dots, 1.96}_{n_{1.96}}, \underbrace{2.58, 2.58, \dots, 2.58}_{n_{2.58}}, \underbrace{2.81, 2.81, \dots, 2.81}_{n_{2.81}} \right\},$$

$\sum C(n_1^{\{k\}}, n_{1.96}^{\{k\}}, n_{2.58}^{\{k\}}, n_{2.81}^{\{k\}})$  be the sum of elements in this set and number  $k$  be the order of  $\sum C(n_1^{\{k\}}, n_{1.96}^{\{k\}}, n_{2.58}^{\{k\}}, n_{2.81}^{\{k\}})$  when all such sums are sorted in ascending order.

The probability mass function (PMF) of statistic  $A$  (Zone distribution) is given with

$$P\left(A = \frac{1}{n} \sum C(n_1^{\{k\}}, n_{1.96}^{\{k\}}, n_{2.58}^{\{k\}}, n_{2.81}^{\{k\}})\right) = \frac{n!}{n_1^{\{k\}}! \cdot n_{1.96}^{\{k\}}! \cdot n_{2.58}^{\{k\}}! \cdot n_{2.81}^{\{k\}}!} \cdot 0.68^{n_1^{\{k\}}} \cdot 0.27^{n_{1.96}^{\{k\}}} \cdot 0.04^{n_{2.58}^{\{k\}}} \cdot 0.01^{n_{2.81}^{\{k\}}},$$

$$k = 1, \dots, \binom{n+3}{n}.$$

**Proof of Theorem 1.** Appendix A.  $\square$

**Remark 2.** Though the standard variant of the  $3\sigma$  rule would not make the essential difference in the distribution, some modalities, such as  $a_i$  and  $a_k$  ( $i \neq k$ ), would be equal due to  $1 + 3 = 2 + 2$ . The adjusted variant of the  $3\sigma$  rule overcomes this issue.

**Corollary 1.** The CDF of variable  $A$  is

$$F(x) = P(A \leq x) = \begin{cases} 0; & x < a_1 \\ \sum_{i=1}^k \frac{n!}{n_1^{\{k\}}! \cdot n_{1.96}^{\{k\}}! \cdot n_{2.58}^{\{k\}}! \cdot n_{2.81}^{\{k\}}!} \cdot 0.68^{n_1^{\{k\}}} \cdot 0.27^{n_{1.96}^{\{k\}}} \cdot 0.04^{n_{2.58}^{\{k\}}} \cdot 0.01^{n_{2.81}^{\{k\}}}, & a_k \leq x < a_{k+1}, k = 1, \dots, \binom{n+3}{n} - 1 \\ 1; & x \geq a_{\binom{n+3}{n}} \end{cases} \quad (3)$$

where  $a_j = \frac{1}{n} \sum C(n_1^{\{j\}}, n_{1.96}^{\{j\}}, n_{2.58}^{\{j\}}, n_{2.81}^{\{j\}}), j = 1, \dots, \binom{n+3}{n}$ .

Calculating  $F(x); x \in \mathbb{R}$  or  $F^{-1}(p); p \in [0, 1]$  is rather complicated, even to program. To overcome this issue, we obtain its quantiles in the table of distribution. The  $zone_n$  distribution is a discrete one; hence,  $F^{-1}(p)$  does not exist for every  $p \in [0, 1]$ . Additionally, for higher  $n$ , the number of modalities of the random variable  $A$  also becomes higher; hence, it is impossible to make the distribution table by calculating  $F(x)$  using (3). That is why a Monte Carlo simulation study with 100,000 simulations for each  $n$  is conducted; thus, Table 1 is obtained. The simulation study has been conducted using MATLAB.

**Table 1.**  $F(x) = P(A \leq x); A \sim zone_m$ .

$n$	$p$												
	0.01	0.025	0.05	0.1	0.15	0.2	0.5	0.8	0.85	0.9	0.95	0.975	0.99
5	*	*	*	*	1.0011	1.0339	1.3840	1.5540	1.5651	1.6382	1.7421	1.8198	1.9201
10	*	1.0038	1.0315	1.0870	1.1840	1.1897	1.3393	1.4749	1.5080	1.5420	1.6270	1.6890	1.7518
20	1.0925	1.1391	1.1440	1.1914	1.2199	1.2385	1.3334	1.4303	1.4612	1.4940	1.5365	1.5730	1.6206
30	1.1280	1.1600	1.1918	1.2220	1.2447	1.2557	1.3370	1.4148	1.4359	1.4573	1.4970	1.5290	1.5647
50	1.1728	1.2016	1.2233	1.2428	1.2620	1.2744	1.3366	1.3990	1.4134	1.4325	1.4596	1.4846	1.5149
100	1.2228	1.2394	1.2552	1.2716	1.2840	1.2936	1.3366	1.3811	1.3912	1.4038	1.4235	1.4405	1.4602
200	1.2552	1.2680	1.2789	1.2916	1.3004	1.3072	1.3376	1.3684	1.3756	1.3849	1.3984	1.4108	1.4249
300	1.2703	1.2806	1.2897	1.2999	1.3070	1.3127	1.3375	1.3627	1.3686	1.3759	1.3872	1.3968	1.4084
500	1.2853	1.2933	1.3003	1.3084	1.3140	1.3184	1.3378	1.3572	1.3671	1.3675	1.3760	1.3835	1.3925
1000	1.3002	1.3061	1.3111	1.3170	1.3210	1.3241	1.3378	1.3515	1.3547	1.3588	1.3647	1.3699	1.3762
1500	1.3077	1.3121	1.3162	1.3209	1.3242	1.3267	1.3379	1.3490	1.3516	1.3548	1.3597	1.3641	1.3690
2000	1.3112	1.3154	1.3190	1.3231	1.3259	1.3281	1.3378	1.3476	1.3499	1.3527	1.3569	1.3605	1.3650

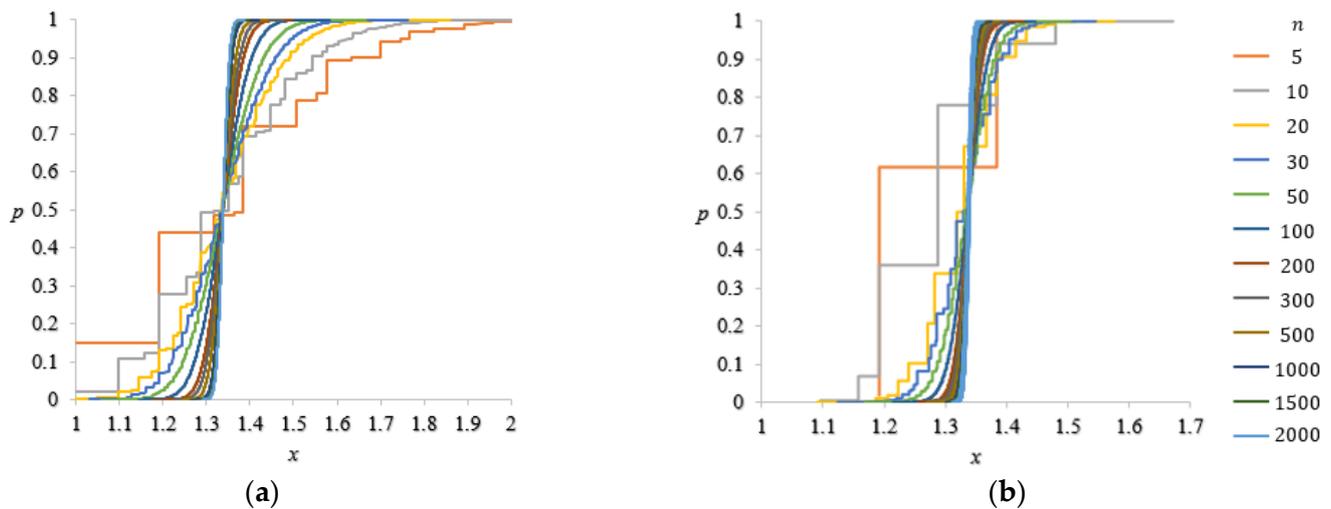
When the distribution parameters are unknown, they can be estimated by the maximum likelihood method. Since  $X \sim N(\mu, \sigma^2)$ , the maximum likelihood method estimates of parameters  $\mu$  and  $\sigma^2$  are  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  and  $\bar{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ , respectively; however, the estimate for  $\sigma^2$  used in this paper is  $\tilde{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ , because it is an unbiased estimate for  $\sigma^2$ , and  $\bar{S}_n^2$  is not.

In this case, the distribution of statistic  $A$  given by (2) is determinable only via Monte Carlo simulations. Another table of the distribution of statistic  $A$  (Table 2) obtained via 100,000 Monte Carlo simulations (runs), performed by MATLAB, is given.

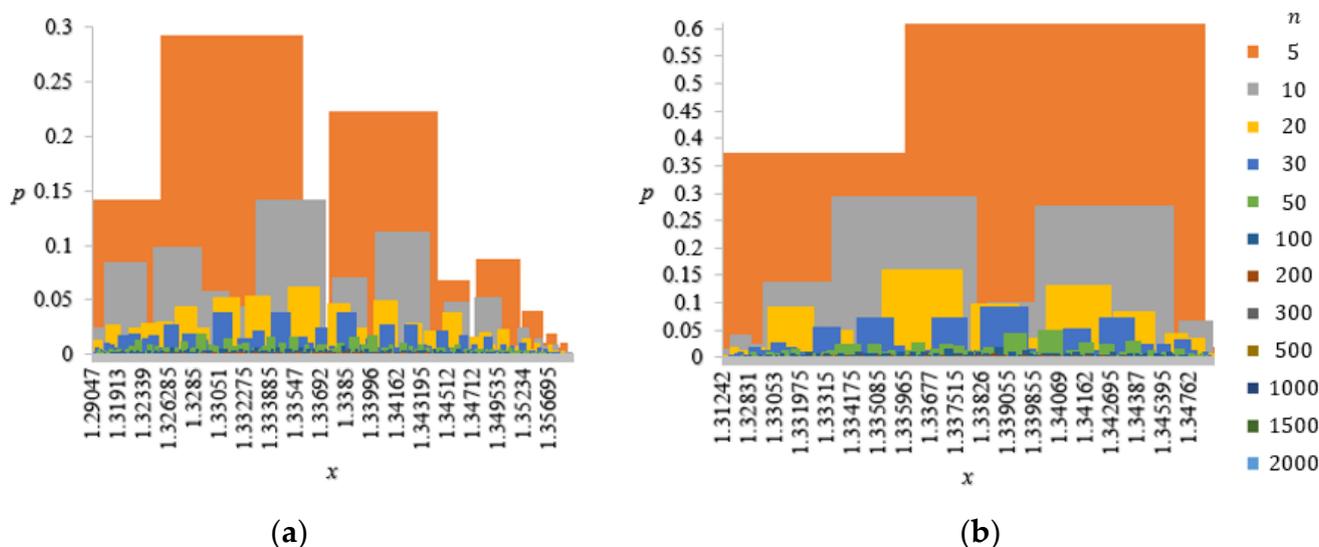
**Table 2.**  $F(x) = P(A \leq x); A \sim zone_m$  when parameters  $\mu$  and  $\sigma^2$  are estimated with  $\bar{X}_n$  and  $\frac{S_n^2}{n}$ , respectively.

$n$	$p$												
	0.01	0.025	0.05	0.1	0.15	0.2	0.5	0.8	0.85	0.9	0.95	0.975	0.99
5	*	1.1920	1.1920	1.1920	1.1920	1.1920	1.1920	1.3840	1.3840	1.3840	1.3840	1.3840	1.3840
10	1.1580	1.1580	1.1580	1.1920	1.1920	1.1920	1.2880	1.3840	1.3840	1.3840	1.4800	1.4800	1.4800
20	1.1865	1.2230	1.2230	1.2400	1.2710	1.2710	1.3190	1.3785	1.3840	1.3840	1.4320	1.4320	1.4630
30	1.2203	1.2410	1.2540	1.2730	1.2843	1.2860	1.3293	1.3727	1.3820	1.3933	1.4047	1.4253	1.4480
50	1.2508	1.2608	1.2722	1.2868	1.2970	1.3050	1.3320	1.3636	1.3704	1.3828	1.3964	1.4088	1.4212
100	1.2750	1.2846	1.2925	1.3021	1.3089	1.3145	1.3366	1.3586	1.3636	1.3704	1.3800	1.3868	1.3987
200	1.2936	1.3008	1.3067	1.3134	1.3180	1.3216	1.3371	1.3530	1.3564	1.3611	1.3676	1.3735	1.3812
300	1.3019	1.3076	1.3125	1.3179	1.3217	1.3247	1.3375	1.3503	1.3533	1.3570	1.3623	1.3674	1.3731
500	1.3100	1.3144	1.3182	1.3225	1.3254	1.3277	1.3377	1.3477	1.3501	1.3528	1.3572	1.3608	1.3651
1000	1.3184	1.3213	1.3240	1.3270	1.3290	1.3307	1.3377	1.3448	1.3464	1.3484	1.3514	1.3541	1.3572
1500	1.3219	1.3244	1.3266	1.3291	1.3307	1.3320	1.3378	1.3436	1.3449	1.3466	1.3491	1.3512	1.3538
2000	1.3241	1.3263	1.3281	1.3302	1.3317	1.3328	1.3378	1.3428	1.3440	1.3455	1.3476	1.3495	1.3516

In the following figures, graphical illustrations of the CDF for the Zone distribution, for different sample sizes (Figure 3) and histograms of PMFs (Figure 4), are given. This Figure illustrates data based on which Tables 1 and 2 were formed.



**Figure 3.** The graph of the CDF  $F(x) = p$  for each  $n$  (approximated values). (a) Known parameters. (b) Estimated parameters.



**Figure 4.** The histogram of the PMF  $P(A = x) = p$  for each  $n$  (approximated values). (a) Known parameters. (b) Estimated parameters.

As can be seen in Figures 3 and 4, when parameters are unknown and estimated, the distribution changes significantly; however, for higher  $n$ , the departure is lower. Both Figures 3 and 4 are obtained using Monte Carlo simulations.

A crucial aspect is that the Zone distribution is not constrained by the parameters, leading to the application of this discrete distribution yielding more precise models.

2.2. Basic Properties and Numerical Characteristics

The following proposition is a direct consequence of Proposition 1.

**Proposition 2.** Expectation and variance of the variable zone( $X_k$ ) are

$$E(\text{zone}(X_k)) = 1.3405$$

and

$$\text{Var}(\text{zone}(X_k)) = 0.2655$$

for  $k = 1, \dots, n$ .

**Proposition 3.** Expectation and variance of variable  $A$  are  $E(A) = 1.3405$  and  $\text{Var}(A) = \frac{0.2655}{n}$ .

**Proof of Proposition 3.** Appendix A. □

**Remark 3.** Since

$$\text{Var}(A) = E(A^2) - (E(A))^2$$

we obtain

$$E(A^2) = \frac{0.2655}{n} + 1.7969. \tag{4}$$

**Theorem 2.** Skewness and kurtosis of variable  $A$  are

$$\text{Skew}(A) = \frac{35.2148n^2 - 0.0015n + 1.0636}{\sqrt{n}}$$

and

$$\text{Kurt}(A) = \frac{-0.0071n^3 + 375.8936n^2 + 543.6709n - 779.4681}{n}.$$

**Proof of Theorem 2.** Appendix A.  $\square$

### 3. Zone Distribution in Normality Testing

#### 3.1. The Testing Procedure

Let

$$X_1, X_2, \dots, X_n; n \in \mathbb{N}$$

be iid variables that are distributed identically as random variable  $X$ . In this section, statistic  $A$  given with (2) is used to test the null hypothesis that variable  $X$  is normally distributed, i.e.,  $H_0(X \sim N(\mu, \sigma^2))$  against the alternative  $H_1(X \text{ is not normally distributed})$ .

The critical region is two-sided interval  $W = [1, c_1] \cup [c_2, 2.81]$ , determined by the critical condition given with

$$P(A \leq c_1|H_0) = P(A \geq c_2|H_0) = \frac{\alpha}{2},$$

where  $\alpha$  is the level of significance.

If the empirical value  $a$  of test-statistic  $A$  is inside the critical region  $W$ , the null hypothesis  $H_0(X \sim N(\mu, \sigma^2))$  is rejected.

The  $p$ -value for this test is calculated as usual by

$$p = \min\{P(A \leq a|H_0), P(A \geq a|H_0)\}$$

and then could be compared to the level of significance  $\alpha$ . If  $p \leq \alpha$ , the null hypothesis is rejected.

All of the advantages of the Quantile-Zone test discussed in [4] hold for the Zone test due to the similar definition of the zone function. For instance, the method used in constructing the test provides tools for assessing the frequency of sample elements within a selected interval. On some level, it is even sensitive to outliers, because more of them will increase the likelihood of hypothesis rejection, and few of them, especially for large-sized samples, will not affect hypothesis acceptance. That is important since, in theory, normal variates are not bounded.

An additional advantage is that sample (1) consists of iid random variables, which make the distribution of the test statistic available in both formal and simulated variants. That also means that the values of the order statistic do not affect the value of test statistic  $A$ , which is the case for many other test statistics, especially ones based on the EDF; hence, test statistic  $A$  is not sensitive to repetitive values, i.e., it does not result in false negative decisions, which can occur very frequently in the case of such an event. Specifically, theoretical consideration allows for the existence of repetitive values (given a reasonable number of repetitions) within our normality hypothesis; however, empirically, these repetitions do not hinder the overall alignment of the sample structure with the fundamental characteristics of the normality hypothesis. The presence of such invariance properties proves advantageous in the context of normality tests

The testing procedure for the Quantile-Zone test is not very complicated to perform or implement in a technical solution, but the Zone test is even less complicated and gives the possibility of fast performance and implementation. This is an important observation because, despite new, more powerful tests being developed [19,26], these tests are not yet widely accepted as an alternative, probably due to the complexity of their application, or the fact that they can only produce high power on certain occasions, etc.

#### 3.2. Power Analysis

For power analysis, we use 10,000 runs of statistic  $A$  Monte Carlo simulations. Statistic  $A$  samples are modelled for various alternative distributions, divided into symmetric and asymmetric ones. Monte Carlo simulations, as well as the alternative distribution modelling, have been performed via MATLAB codes. Some modelling algorithms available

for MATLAB functions, such as ‘normrnd’, ‘chi2rnd’, and ‘trnd’, etc., are used [30]. For the alternative distributions with no available MATLAB sampling function, we coded algorithms using the inverse function (CDF) method [31]. The numbers of simulations, 10,000 and 100,000, have both proven to be satisfactory [32]. Additionally, in our simulation study, we used both for null distribution, and the results are asymptotically identical.

The null distribution is  $N(0, 1)$ . We have used the level of significance  $\alpha = 0.05$ . Various sample sizes have been discussed and are available in the following tables. The selection of parameters for alternative distributions has been done in a way that maximizes alignment with the null distribution, within reasonable bounds.

The power of the Zone test is calculated by the following algorithm:

1. Modelling the sample  $x_1, x_2, \dots, x_n$  of the chosen alternative distribution for the observed sample size  $n$ ;
2. Calculating  $zone(x_1), zone(x_2), \dots, zone(x_n)$  and the empirical value of the test statistic

$$a = \frac{1}{n} \sum_{i=1}^n zone(x_i);$$

3. Repeating the first two steps  $m = 10,000$  times and thus obtaining the sample  $a_1, a_2, \dots, a_m$ ;
4. Determining the EDF

$$F_m^*(x) = \frac{1}{m} \sum_{i=1}^m I(a_i \leq x)$$

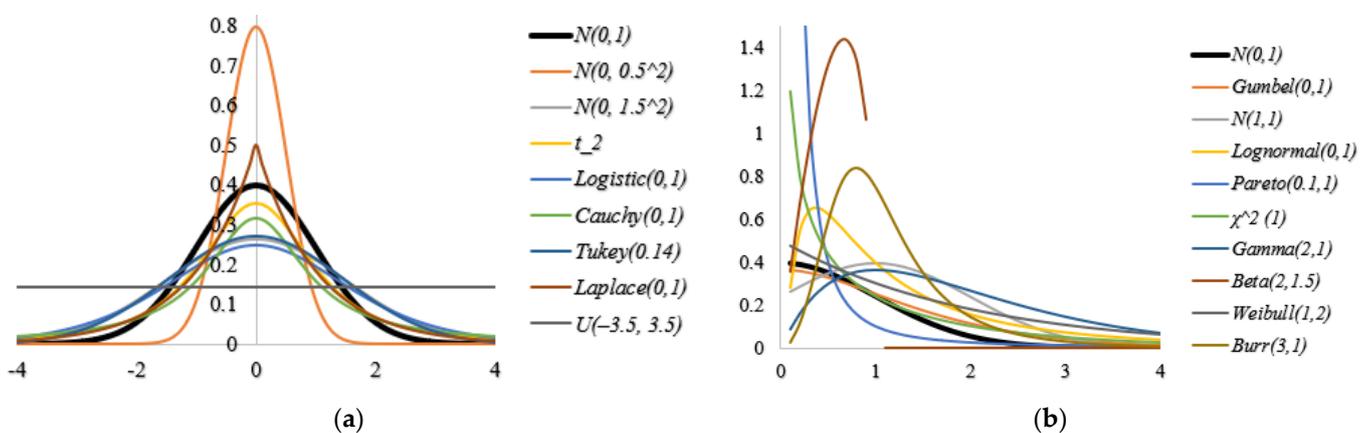
of the sample obtained in the third step ( $I$  is the event indicator; it equals 1 if the event had occurred and 0 if it had not);

5. Calculating the power  $1 - \beta$  of the test by

$$1 - \beta = F_m^*(c_1) + (1 - F_m^*(c_2))$$

where  $c_1$  and  $c_2$  are the critical values of statistic  $A$  for  $\alpha = 0.05$ .

In several other papers [19–22], power calculations encompass multiple distributions; however, for many, such analyses might be considered unnecessary, as preliminary methods, such as a histogram, suffice for normality assessment. Furthermore, this situation inflates the average power value, concealing alternative distributions where the test exhibits low power. In essence, the power value is artificially boosted by large but irrelevant data. To address this, alternative distributions are thoughtfully selected in this study to provide a more accurate representation of the test’s applicability. The following figure (Figure 5) presents a comprehensive list of all alternative distributions for which power calculations are conducted.



**Figure 5.** Alternative distributions compared to the null distribution. (a) Symmetric alternatives. (b) Asymmetric alternatives.

Further on, Tables 3 and 4 of the power values are available.

When examining symmetric alternative distributions, it becomes evident that the Zone test exhibits excellent performance, especially when the parameters of the null hypothesis distribution are known. The test achieves the highest performance with the uniform distribution and the lowest with the Laplace (0, 1) distribution. The test is powerful even for small sample sizes, though there are some exceptions. For instance, when the alternative distribution is Laplace (0, 1), it is the best to use the Zone test for  $n \approx 150$  or higher; when the alternative distribution is it is the best to use the Zone test for  $n > 50$  (see Table 3). Observing Figure 5a, we can see that the discrepancy of the chosen symmetric alternative distributions from the null distribution is not as large as it usually is in the power analysis studies [19,20]. Given this observation, and noting that the Zone test performs better for alternative distributions that deviate more from the null hypothesis distribution, we can confidently regard our test as a suitable choice for normality testing against symmetric alternative distributions.

**Table 3.** Empirical power of the Zone test for various sample sizes with the level of significance  $\alpha = 0.05$  when the parameters are known—symmetric alternative distributions.

Distribution	<i>n</i>					
	10	20	30	50	100	200
Laplace (0, 1)	0.1441	0.2187	0.2692	0.3818	0.6030	0.8508
$t_2$	0.2679	0.4392	0.5574	0.7541	0.9489	0.9990
Tukey (0.14)	0.3331	0.5776	0.7361	0.9029	0.9943	1
$N(0, 1.5^2)$	0.3640	0.6172	0.7734	0.9283	0.9966	1
Logistic (0, 1)	0.5106	0.7924	0.9068	0.9873	1	1
Cauchy (0, 1)	0.5434	0.8027	0.9147	0.9865	1	1
$N(0, 0.5^2)$	0.6321	0.9446	0.9980	1	1	1
$U(-3.5, 3.5)$	0.9216	0.9974	0.9998	1	1	1
<b>Average</b>	<b>0.4646</b>	<b>0.6737</b>	<b>0.7694</b>	<b>0.8676</b>	<b>0.9428</b>	<b>0.9812</b>

**Table 4.** Empirical power of the Zone test for various sample sizes with the level of significance  $\alpha = 0.05$  when the parameters are known—asymmetric alternative distributions.

Distribution	<i>n</i>					
	10	20	30	50	100	200
$\chi_1^2$	0.1476	0.1944	0.2139	0.2861	0.4226	0.6630
Gumbel (0, 1)	0.1478	0.2087	0.2636	0.3784	0.6002	0.8517
Burr (3, 1)	0.2574	0.4649	0.6216	0.8290	0.9816	1
Pareto (0.1, 1)	0.3641	0.6064	0.6670	0.8504	0.9882	1
$N(1, 1)$	0.3850	0.6215	0.7977	0.9378	0.9974	1
Lognormal (0, 1)	0.4822	0.7294	0.8721	0.9701	0.9998	1
Weibull (1, 2)	0.7867	0.9628	0.9932	0.9998	1	1
Gamma (2, 1)	0.9382	0.9976	1	1	1	1
Beta (2, 1.5)	1	1	1	1	1	1
<b>Average</b>	<b>0.5010</b>	<b>0.6429</b>	<b>0.7143</b>	<b>0.8057</b>	<b>0.8878</b>	<b>0.9461</b>

The scenario remains largely consistent in the context of asymmetric alternative distributions. Specifically, when examining all the alternative distributions presented in Figure 5b, in this case, the same conclusions as those drawn in the previous paragraph hold true. There is a slight difference considering the performance of the test for small sample sizes where the alternative distributions are the  $\chi_1^2$  distribution and the Gumbel (0, 1) distribution; however, the difference noted does not have a significant impact on the final conclusion (see Table 4).

In the following tables, power analysis for the same levels of significance, sample sizes, and alternative distributions is given; however, it is the case of the estimated parameters that is considered here. Now, the distribution of statistic  $A$  is given in Table 2.

Zone test performance (and that of any other test) gets better when parameters are estimated because the empirical mean  $\bar{X}_n$  and standard deviation  $\tilde{S}_n$ , practically, are better representatives of average value and average deviation than the assumed values  $\mu$  and  $\sigma$ . If the sample is truly drawn from the  $N(\mu, \sigma^2)$  distribution,  $\bar{X}_n \approx \mu$  and  $\tilde{S}_n \approx \sigma^2$  hold; otherwise, the Zone test will register a larger discrepancy. The performance of the test has improved for all of the symmetric alternative distributions. The biggest improvement can be noticed in the Laplace (0, 1) distribution. An interesting observation is that, through parameter estimation, the Zone test demonstrates improved performance for the Logistic (0, 1) distribution compared to the Cauchy (0, 1) distribution. This outcome differs from the one with known parameters.

Figure 6 provides a visual comparison of the Zone test’s performance for symmetric alternative distributions, considering both specified and estimated parameters. This graphical representation corresponds to the information presented in Tables 3 and 5.

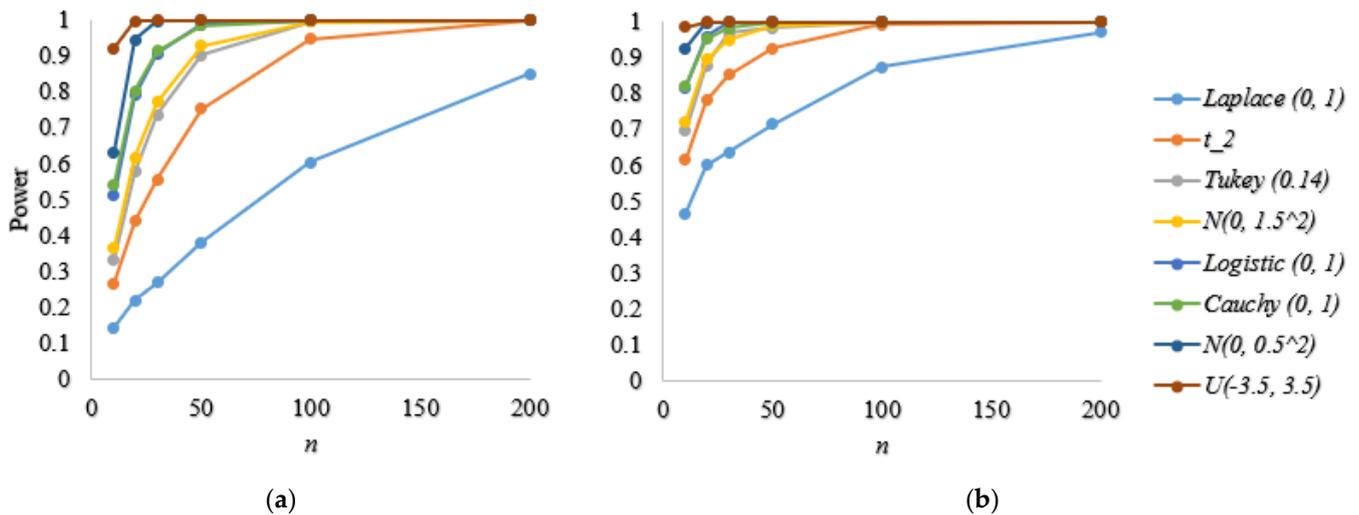


Figure 6. Empirical power of the Zone test for various sample sizes with the level of significance  $\alpha = 0.05$ —symmetric alternative distributions. (a) Known parameters. (b) Estimated parameters.

Table 5. Empirical power of the Zone test for various sample sizes with the level of significance  $\alpha = 0.05$  when the parameters are estimated—symmetric alternative distributions.

Distribution	$n$					
	10	20	30	50	100	200
Laplace (0, 1)	0.4668	0.6013	0.6365	0.7150	0.8739	0.9716
$t_2$	0.6163	0.7812	0.8532	0.9269	0.9921	1
Tukey (0.14)	0.6971	0.8795	0.9682	0.9827	0.9995	1
$N(0, 1.5^2)$	0.7232	0.8946	0.9509	0.9905	1	1
Logistic (0, 1)	0.8171	0.9606	0.9972	1	1	1
Cauchy (0, 1)	0.8215	0.9522	0.9826	0.9986	1	1
$N(0, 0.5^2)$	0.9271	0.9973	0.9998	1	1	1
$U(-3.5, 3.5)$	0.9866	0.9997	1	1	1	1
Average	0.7570	0.8833	0.9236	0.9517	0.9832	0.9965

Figure 7 provides a visual comparison of the Zone test’s performance for symmetric alternative distributions, considering both specified and estimated parameters. This graphical representation corresponds to the information presented in Tables 4 and 6.

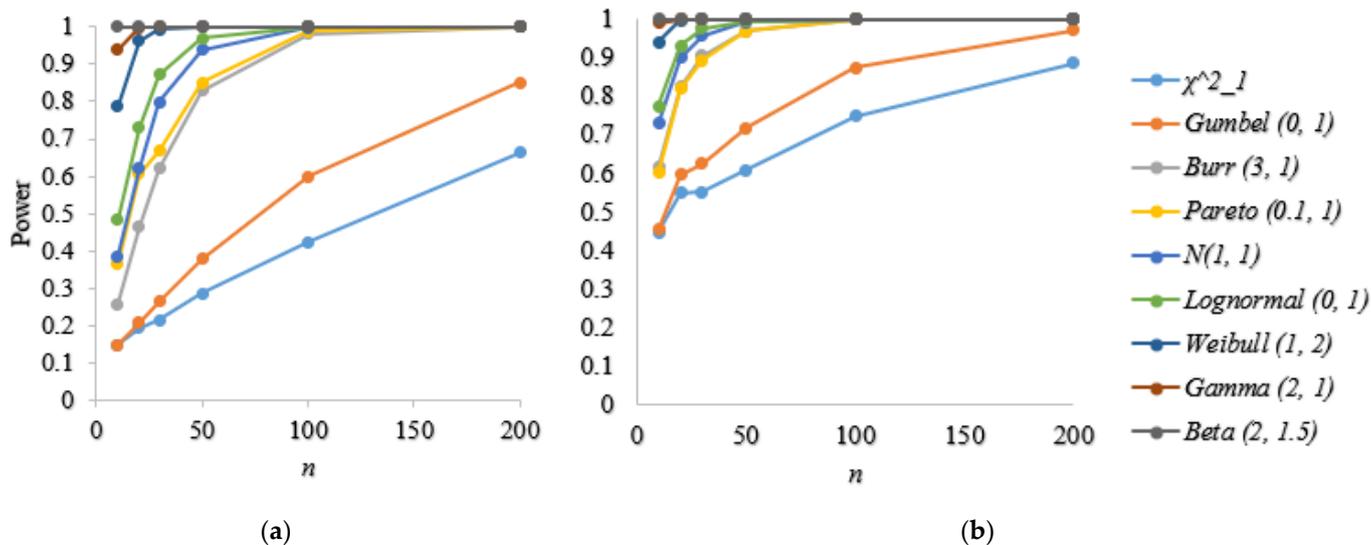


Figure 7. Empirical power of the Zone test for various sample sizes with the level of significance  $\alpha = 0.05$ —asymmetric alternative distributions. (a) Known parameters. (b) Estimated parameters.

Table 6. Empirical power of the Zone test for various sample sizes with the level of significance  $\alpha = 0.05$  when the parameters are estimated—asymmetric alternative distributions.

Distribution	n					
	10	20	30	50	100	200
$\chi^2_1$	0.4479	0.5499	0.5524	0.6078	0.7485	0.8865
Gumbel (0, 1)	0.4548	0.5963	0.6246	0.7169	0.8745	0.9710
Burr (3, 1)	0.6164	0.8248	0.9044	0.9691	0.9990	1
Pareto (0.1, 1)	0.6026	0.8190	0.8941	0.9686	0.9991	1
N (1, 1)	0.7305	0.8995	0.9581	0.9939	1	1
Lognormal (0, 1)	0.7744	0.9312	0.9767	0.9950	1	1
Weibull (1, 2)	0.9411	0.9967	0.9997	1	1	1
Gamma (2, 1)	0.9900	0.9999	1	1	1	1
Beta (2, 1.5)	1	1	1	1	1	1
Average	0.7286	0.8464	0.8789	0.9168	0.9579	0.9842

This conclusion holds for asymmetric alternative distributions as well. The most noticeable improvements in the power of the test are identified for the  $\chi^2_1$  distribution and Gumbel (0, 1) distribution. When the parameters are specified, the Zone test shows better performance for Pareto (0.1, 1) compared to the Burr (3, 1) distribution, but for estimated parameters the opposite conclusion holds. The test performs better for symmetric alternative distributions; however, it has been shown to be very powerful in both variants.

The following section delves into a more detailed discussion of consistency properties, addressing both the advantages and disadvantages of the Zone test in comparison to other commonly employed normality tests.

### 3.3. Comparative Analysis

Here, the obtained power values are used for calculating the average ones, which are then compared to the average power values for other well-known and widely used normality tests. The tests that the Zone test is compared to are the Kolmogorov–Smirnov test [33] with its variant for estimated parameters (Lilliefors test) [34], Chi-square test [28], Shapiro–Wilk test [35], and Anderson–Darling test [36]. For most of these tests, the same data obtained in [4] are used.

Table 7 reveals that, with known parameters, the Zone test outperforms other tests in the case of symmetric alternative distributions. This conclusion holds even for small

sample sizes. The power of the Zone test is the largest for all sample sizes. Even for large sample sizes, such as  $n \approx 200$ , the Zone test is still noticeably more powerful than other analyzed tests.

**Table 7.** Average empirical power values of the Zone test and some other normality tests for various sample sizes with the level of significance  $\alpha = 0.05$ —symmetric alternative distributions.

Test	<i>n</i>					
	10	20	30	50	100	200
Zone (EP <sup>1</sup> )	<b>0.7570</b>	<b>0.8833</b>	<b>0.9236</b>	<b>0.9517</b>	<b>0.9832</b>	<b>0.9965</b>
Zone (KP <sup>2</sup> )	<b>0.4646</b>	<b>0.6737</b>	<b>0.7694</b>	<b>0.8676</b>	<b>0.9428</b>	<b>0.9812</b>
Shapiro–Wilk	0.2328	0.4548	0.6768	0.7712	0.8469	0.9005
Anderson–Darling	0.2316	0.4522	0.6730	0.7644	0.8368	0.8901
$\chi^2$	0.2158	0.4208	0.6257	0.7307	0.8160	0.8672
Lilliefors	0.2177	0.4245	0.6313	0.7222	0.7996	0.8605
Kolmogorov–Smirnov	0.1917	0.3726	0.5534	0.6723	0.7628	0.8175

<sup>1</sup> Estimated Parameters. <sup>2</sup> Known Parameters.

In Table 8, results indicate a different outcome; namely, that the Zone test has not performed as well as the competitor tests when the parameters of the normal distribution are known and in cases involving asymmetric alternative distributions. There is an exception when  $n = 200$ , where this test has a higher power value than the Kolmogorov–Smirnov test. The power of this test is still very high, and the differences are not an argument against using the Zone test for known parameters of the normal null distribution.

**Table 8.** Average empirical power values of the Zone test and some other normality tests for various sample sizes with the level of significance  $\alpha = 0.05$ —asymmetric alternative distributions.

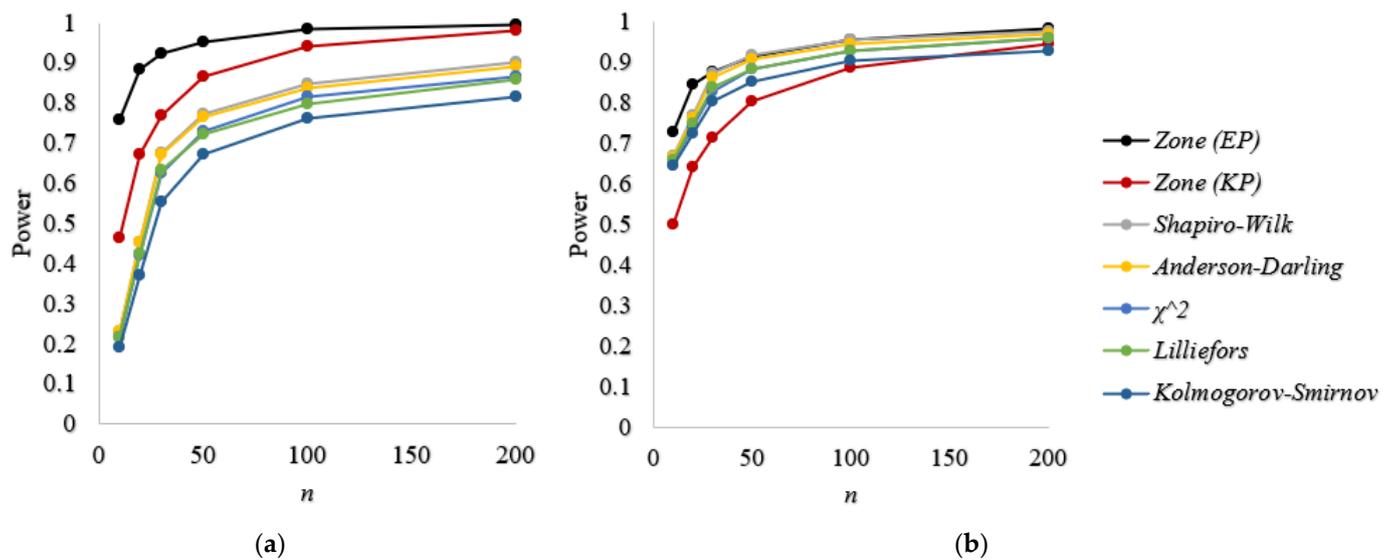
Test	<i>n</i>					
	10	20	30	50	100	200
Zone (EP <sup>1</sup> )	<b>0.7286</b>	<b>0.8464</b>	<b>0.8789</b>	<b>0.9168</b>	<b>0.9579</b>	<b>0.9842</b>
Zone (KP <sup>2</sup> )	<b>0.5010</b>	<b>0.6429</b>	<b>0.7143</b>	<b>0.8057</b>	<b>0.8878</b>	<b>0.9461</b>
Shapiro–Wilk	0.6698	0.7714	0.8730	0.9191	0.9552	0.9759
Anderson–Darling	0.6666	0.7649	0.8633	0.9087	0.9465	0.9702
$\chi^2$	0.6552	0.7423	0.8293	0.8841	0.9293	0.9615
Lilliefors	0.6587	0.7493	0.8398	0.8859	0.9285	0.9586
Kolmogorov–Smirnov	0.6467	0.7253	0.8038	0.8543	0.9040	0.9308

<sup>1</sup> Estimated Parameters. <sup>2</sup> Known Parameters.

When the parameters of the null distribution are estimated, the results show that the Zone test is more powerful than all considered competitor tests. This result holds for all of the sample sizes observed. The variant of the test for estimated parameters is more powerful because a better fit with the Zone distribution in Table 2 is obtained.

The following figure (Figure 8) provides a graphical interpretation of the previous results, offering a comparative analysis of power function graphs for both symmetric and asymmetric alternative distributions.

The results highlight the Zone test as a good choice for normality testing. In the cases of both the known and estimated parameters of the null hypothesis distribution, the Zone test exhibits similar or even better performance compared to the tests included in this comparative analysis. The alternative distributions utilized are likely among the most representative of commonly used distributions [19,21,22]. Furthermore, upon reviewing power analysis results for tests not covered in this comparative analysis, an additional advantage of the Zone test becomes apparent. These observations can therefore be generalized when comparing the Zone test to the majority of well-known normality tests [19–26].



**Figure 8.** Average empirical power values of the Zone test and some other normality tests for various sample sizes with the level of significance  $\alpha = 0.05$ . (a) Symmetric alternative distributions. (b) Asymmetric alternative distributions.

Finally, Zone test and Quantile-Zone test performances are compared since, in the authors' previous paper, it was shown that, in a similar analysis, the Quantile-Zone test yielded the best results.

Table 9 illustrates that, for symmetric alternative distributions, the Quantile-Zone test demonstrates significantly better performance in the case of known parameters; however, in the case of estimated parameters, the Zone test is the one with better performance. For large sample sizes, the differences are of no significance.

**Table 9.** Average empirical power values of the Zone test and Quantile-Zone test for various sample sizes with the level of significance  $\alpha = 0.05$ —symmetric alternative distributions.

Test	<i>n</i>					
	10	20	30	50	100	200
Zone (EP <sup>1</sup> )	0.7570	0.8833	0.9236	0.9517	0.9832	0.9965
Quantile-Zone (EP)	0.7309	0.8365	0.8945	0.9466	0.9855	0.9962
Zone (KP <sup>2</sup> )	0.4646	0.6737	0.7694	0.8676	0.9428	0.9812
Quantile-Zone (KP)	0.4921	0.7294	0.8462	0.9295	0.9832	0.9962

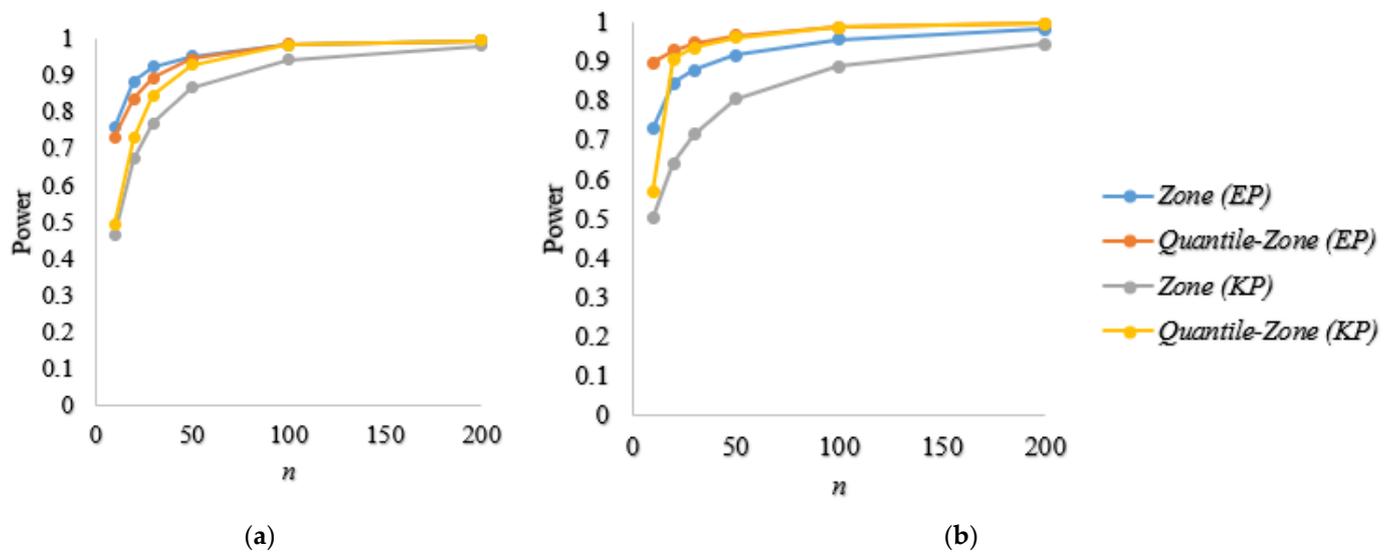
<sup>1</sup> Estimated Parameters. <sup>2</sup> Known Parameters.

Table 10 indicates that, for asymmetric alternative distributions, the Quantile-Zone test outperforms the Zone test, establishing it as the most powerful normality test based on our results. In Figure 9, the results of this comparison are graphically illustrated.

**Table 10.** Average empirical power values of the Zone test and Quantile-Zone test for various sample sizes with the level of significance  $\alpha = 0.05$ —asymmetric alternative distributions.

Test	<i>n</i>					
	10	20	30	50	100	200
Zone (EP <sup>1</sup> )	0.7286	0.8464	0.8789	0.9168	0.9579	0.9842
Quantile-Zone (EP)	0.8984	0.9294	0.9470	0.9664	0.9881	0.9978
Zone (KP <sup>2</sup> )	0.5010	0.6429	0.7143	0.8057	0.8878	0.9461
Quantile-Zone (KP)	0.5673	0.9066	0.9367	0.9623	0.9878	0.9977

<sup>1</sup> Estimated Parameters. <sup>2</sup> Known Parameters.



**Figure 9.** Average empirical power values of the Zone test and Quantile-Zone test for various sample sizes with the level of significance  $\alpha = 0.05$ . (a) Symmetric alternative distributions. (b) Asymmetric alternative distributions.

Despite the best performance of the Quantile-Zone test, there are advantages to the Zone test that might make it a better choice, on some occasions, for the following reasons:

- It is still more powerful than the other normality tests usually used;
- It is very simple to apply and program;
- It performs faster than the Quantile-Zone test, which is significant for big data;
- The elements of the sample  $zone(X_j); j = 1, \dots, n$  are not mutually dependent, which is not the case for zones in Quantile-Zone distribution. That makes the Zone distribution and many of its characteristics determinable theoretically (3);
- The invariance for outliers, to some extent, is the same as with the Quantile-Zone test, etc.

#### 4. Examples

The following examples will illustrate how to apply the Zone distribution in normality testing; in other words, the data in the example will be tested for normality via the Zone test. This example also indicates the connection this discrete distribution makes between quality control and normality testing.

To control the quantity of protein in milk, 48 packages with 100 g of milk are taken from the production line. Measurements have yielded the following results (in %): 3.04, 3.12, 3.12, 3.22, 3.09, 3.13, 3.21, 3.18, 3.10, 3.18, 3.21, 3.18, 3.04, 3.11, 3.17, 3.06, 3.13, 3.12, 3.11, 3.07, 3.15, 3.05, 3.14, 3.18, 3.11, 3.21, 3.22, 3.13, 3.06, 3.07, 3.17, 3.22, 3.05, 3.19, 3.18, 3.20, 3.08, 3.20, 3.21, 3.09, 3.05, 3.14, 3.22, 3.08, 3.19, 3.18, 3.21, 3.06. The concentration of protein in milk is usually between three and four percent. Two examples are being considered.

##### 4.1. Known Parameters Case

Assume that the milk packages meet the standard if the protein concentration is distributed by the normal  $N(3.15, 0.08^2)$  distribution. This is tested using the Zone test. The empirical value of test statistic  $a = 1.16$  is inside the critical region  $W = [1, 1.1974] \cup [1.5846, 2.81]$ . That indicates that the distribution of the protein intake in the milk packages is not normally distributed with the distribution  $N(3.15, 0.08^2)$ . The significance is  $p \approx 0.01 < 0.05$ .

In this example, as depicted in Figure 10, despite the sample elements being within the band  $[\mu - \sigma, \mu + \sigma]$ , the test statistic detects an excessive concentration within this range, resulting in a smaller realized value of  $A$ . Additionally, it is illustrated that several repetitions of some sample elements did not cause extreme variations in the  $A$  value.

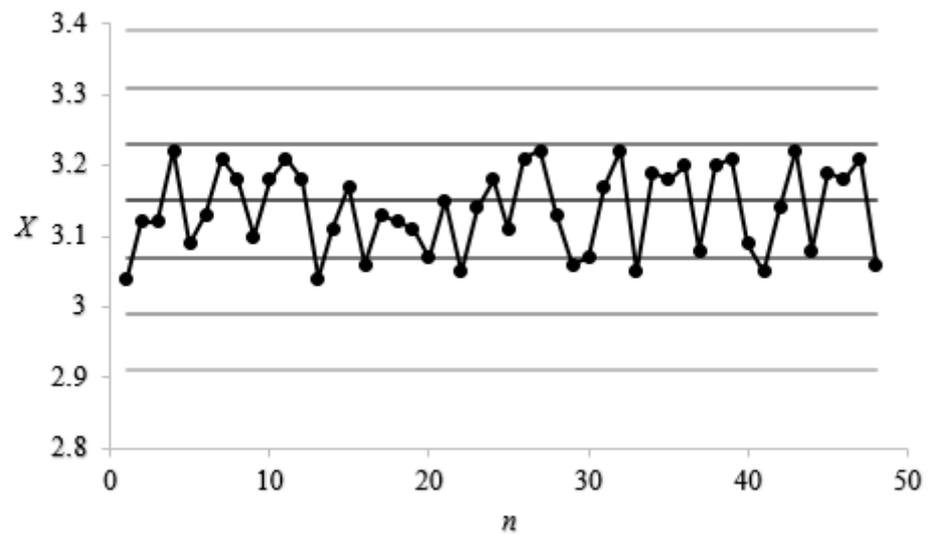


Figure 10. Distribution of the sample in relation to zone lines—known parameters.

4.2. Estimated Parameters Case

Assume that the milk packages meet the standard if the protein concentration is distributed by the normal  $N(\bar{x}_n, \tilde{s}_n^2) \sim N(3.14, 0.06^2)$  distribution. This is tested using the Zone test. The empirical value of test statistic  $a = 1.42$  is inside the critical region  $W = [1, 1.2588] \cup [1.4105, 2.81]$ . That indicates that the distribution of the protein intake in the milk packages is not normally distributed with the distribution  $N(3.14, 0.06^2)$ . The significance is  $p \approx 0.035 < 0.05$ .

Figure 11 illustrates that estimating the parameters has positioned the zone lines in a more adequate way in terms of normality; however, the structure of the sample elements still appears to deviate from the normal one due to too many of the sample elements being detected in the zones between one and two standard deviations from the mean.

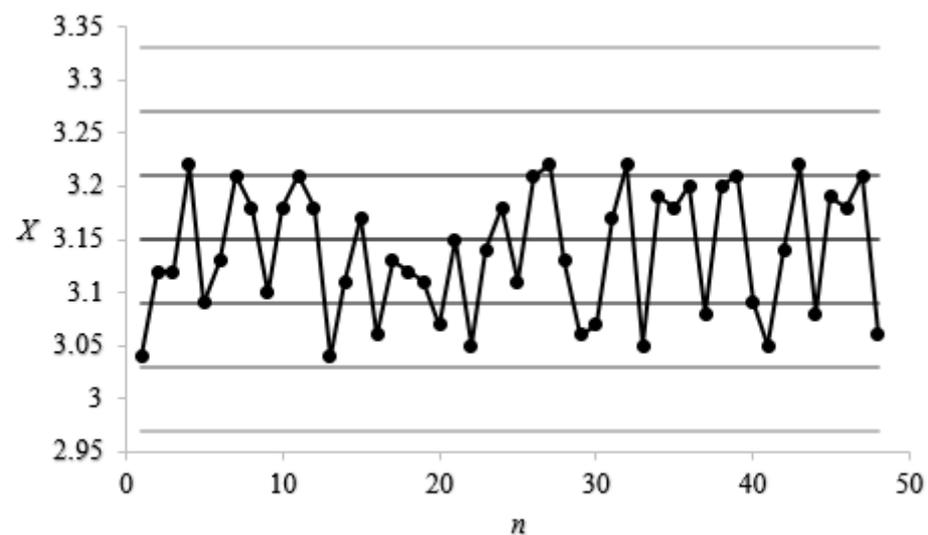


Figure 11. Distribution of the sample in relation to zone lines—estimated parameters.

5. Conclusions

This paper accomplishes the following:

- Development of a novel discrete distribution named the Zone distribution, associated with normal distributions, including the presentation of functional characteristics, PMF, CDF, and corresponding graphical illustrations;

- Provision of quantile tables for the Zone distribution in the cases of both known and estimated parameters of related normal distributions;
- Computation of key numerical characteristics for the Zone distribution;
- Illustration of the application of the Zone distribution in normality testing, along with an exploration of its advantages and properties;
- Calculation of empirical power for the new Zone test, conducted separately for symmetric and asymmetric alternative distributions, providing power analysis results for both known and estimated parameters;
- Presentation of a highly illustrative graphical interpretation of the power analysis;
- Comparative power analysis of both variants of the Zone test (specified and estimated parameters) against other commonly used normality tests, accompanied by detailed results and graphical representations.

Future work will involve further exploration of the Zone and Quantile-Zone approaches in defining discrete distributions and their application in goodness-of-fit testing. Additionally, there are plans to investigate new possibilities for applying these discrete models, conduct additional power and efficiency analyses, extend these approaches to multivariate normality testing, explore adjustments to the continuous variant, and pursue other possibilities for enhancing the methodology.

**Author Contributions:** Conceptualization, A.A.; methodology, A.A. and V.J.; software, A.A.; validation, A.A. and V.J.; formal analysis, A.A. and V.J.; investigation, A.A.; data curation, A.A.; writing—original draft preparation, A.A.; writing—review and editing, A.A. and V.J.; visualization, A.A.; supervision, A.A. and V.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Simulated general data sets were used in this study.

**Acknowledgments:** The authors express their gratitude for the valuable revision suggestions provided by the reviewers, which have significantly enhanced the quality of the paper. Additionally, they extend their thanks to the editors for their prompt handling of the manuscript preparation process.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A

**Proof of Theorem 1.** To adequately describe the  $zone_n$  distribution, combinations with repetitions of the set  $\{1, 1.96, 2.58, 2.81\}$  are used. Namely, in (2), calculating  $A$  requires the sample given in (1). The order of the sample elements does not affect the value of  $A$ ; hence, the modalities of  $A$  are equivalent to the class  $n$  combination of numbers 1, 1.96, 2.58, 2.81 with repetition. That means that  $A$  has

$$\binom{4 + n - 1}{n} = \binom{n + 3}{n} = \frac{(n + 3)!}{n! \cdot 3!}$$

modalities, where  $n! = n \cdot (n - 1) \cdot \dots \cdot 2 \cdot 1$ .

The above-mentioned combinations are

$$C(n_1, n_{1.96}, n_{2.58}, n_{2.81}) = \left\{ \underbrace{1, 1, \dots, 1}_{n_1}, \underbrace{1.96, 1.96, \dots, 1.96}_{n_{1.96}}, \underbrace{2.58, 2.58, \dots, 2.58}_{n_{2.58}}, \underbrace{2.81, 2.81, \dots, 2.81}_{n_{2.81}} \right\},$$

then the modality equivalent to this combination is

$$a_k = \frac{1}{n} \sum C\left(n_1^{\{k\}}, n_{1.96}^{\{k\}}, n_{2.58}^{\{k\}}, n_{2.81}^{\{k\}}\right), k = 1, \dots, \binom{n + 3}{n}.$$

The order of the elements in sample (1) does not affect the value of  $A$ , but it does affect the probability of that value being realized. That is why the number of permutations of each combination, equivalent to a specific modality, is observed. Since the combinations are with repetition, the referred set of permutations are with repetition as well. The number of permutations with repetition of the set  $C(n_1^{\{k\}}, n_{1.96}^{\{k\}}, n_{2.58}^{\{k\}}, n_{2.81}^{\{k\}})$  elements is

$$\frac{n!}{n_1^{\{k\}}! \cdot n_{1.96}^{\{k\}}! \cdot n_{2.58}^{\{k\}}! \cdot n_{2.81}^{\{k\}}!}$$

For independent random variables  $X_1, X_2, \dots, X_n$ , the equation

$$P(\text{zone}(X_1) = s_1; \text{zone}(X_2) = s_2; \dots; \text{zone}(X_n) = s_n) = P(\text{zone}(X_1) = s_1) \cdot P(\text{zone}(X_2) = s_2) \cdot \dots \cdot P(\text{zone}(X_n) = s_n)$$

holds for  $s_k \in \{1, 1.96, 2.58, 2.81\}$ ,  $k = 1, \dots, n$ , thus

$$\begin{aligned} P(nA = \sum C(n_1^{\{k\}}, n_{1.96}^{\{k\}}, n_{2.58}^{\{k\}}, n_{2.81}^{\{k\}})) &= P\left(\sum_{i=1}^n \text{zone}(X_i) = \sum C(n_1^{\{k\}}, n_{1.96}^{\{k\}}, n_{2.58}^{\{k\}}, n_{2.81}^{\{k\}})\right) \\ &= \frac{n!}{n_1^{\{k\}}! \cdot n_{1.96}^{\{k\}}! \cdot n_{2.58}^{\{k\}}! \cdot n_{2.81}^{\{k\}}!} P\left(\bigcap_{i=1}^n \text{zone}(X_i)\right) \\ &= C(n_1^{\{k\}}, n_{1.96}^{\{k\}}, n_{2.58}^{\{k\}}, n_{2.81}^{\{k\}})(i) \\ &= \frac{n!}{n_1^{\{k\}}! \cdot n_{1.96}^{\{k\}}! \cdot n_{2.58}^{\{k\}}! \cdot n_{2.81}^{\{k\}}!} P\left(\bigcap_{i=1}^{n_1^{\{k\}}} \text{zone}(X_i) = 1 \cap \bigcap_{i=n_1^{\{k\}}+1}^{n_{1.96}^{\{k\}}} \text{zone}(X_i) = 1.96 \cap \bigcap_{i=n_{1.96}^{\{k\}}+1}^{n_{2.58}^{\{k\}}} \text{zone}(X_i) = 2.58 \cap \bigcap_{i=n_{2.58}^{\{k\}}+1}^{n_{2.81}^{\{k\}}} \text{zone}(X_i) = 2.81\right) \\ &= \frac{n!}{n_1^{\{k\}}! \cdot n_{1.96}^{\{k\}}! \cdot n_{2.58}^{\{k\}}! \cdot n_{2.81}^{\{k\}}!} \prod_{i=1}^{n_1^{\{k\}}} P(\text{zone}(X_i) = 1) \prod_{i=n_{1.96}^{\{k\}}}^{n_{2.58}^{\{k\}}} P(\text{zone}(X_i) = 1.96) \\ &= 1.96) \prod_{i=n_{1.96}^{\{k\}}}^{n_{2.58}^{\{k\}}} P(\text{zone}(X_i) = 2.58) \prod_{i=n_{2.58}^{\{k\}}}^{n_{2.81}^{\{k\}}} P(\text{zone}(X_i) = 2.81) \\ &= \frac{n!}{n_1^{\{k\}}! \cdot n_{1.96}^{\{k\}}! \cdot n_{2.58}^{\{k\}}! \cdot n_{2.81}^{\{k\}}!} \cdot 0.68^{n_1^{\{k\}}} \cdot 0.27^{n_{1.96}^{\{k\}}} \cdot 0.04^{n_{2.58}^{\{k\}}} \cdot 0.01^{n_{2.81}^{\{k\}}} \end{aligned}$$

due to Proposition 1. The equation

$$P(nA = \sum C(n_1^{\{k\}}, n_{1.96}^{\{k\}}, n_{2.58}^{\{k\}}, n_{2.81}^{\{k\}})) = P\left(A = \frac{1}{n} \sum C(n_1^{\{k\}}, n_{1.96}^{\{k\}}, n_{2.58}^{\{k\}}, n_{2.81}^{\{k\}})\right)$$

implies the claim of the theorem. □

**Proof of Proposition 3.** Using basic properties of expectation and variance, the following equality sequences hold.

$$E(A) = E\left(\frac{1}{n} \sum_{i=1}^n \text{zone}(X_i)\right) = \frac{1}{n} \sum_{i=1}^n E(\text{zone}(X_i)) = \frac{1}{n} \cdot n \cdot 1.3405 = 1.3405 \tag{A1}$$

and

$$Var(A) = Var\left(\frac{1}{n} \sum_{i=1}^n zone(X_i)\right) = \frac{1}{n^2} \sum_{i=1}^n Var(zone(X_i)) = \frac{1}{n^2} n \cdot 0.2655 = \frac{0.2655}{n}. \tag{A2}$$

□

**Proof of Theorem 2.** Based on Proposition 2.1, the equality

$$E\left((zone(X_k))^3\right) = 3.6218, k = 1, \dots, n$$

holds. Note that there is  $n(n - 1)$  of the variations  $(i, j), i, j \in \{1, \dots, n\}, i \neq j$  and  $\binom{n}{3} = \frac{n!}{6(n - 3)!}$  of combinations  $\{i, j, k\} \subset \{1, \dots, n\}$ . Hence,

$$\begin{aligned} E(A^3) &= E\left(\left(\frac{1}{n} \sum_{i=1}^n zone(X_i)\right)^3\right) \\ &= \frac{1}{n^3} \left( \sum_{i=1}^n E\left((zone(X_i))^3\right) + 3 \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n E\left((zone(X_i))^2\right)E(zone(X_j)) \right. \\ &\quad \left. + 6 \sum_{\{i,j,k\} \subset \{1,\dots,n\}} E(zone(X_i))E(zone(X_j))E(zone(X_k)) \right) \\ &= \frac{1}{n^3} \left( \sum_{i=1}^n 3.6218 + 3 \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n 2.0624 \cdot 1.3405 + 6 \sum_{\{i,j,k\} \subset \{1,\dots,n\}} 1.3405^3 \right) \tag{A3} \\ &= \frac{1}{n^3} \left( 3.6218n + 3n(n - 1)2.0624 \cdot 1.3405 + 6 \binom{n}{3} 1.3405^3 \right) \\ &= \frac{2.4088n^2 + 1.0675n + 0.1455}{n^2}. \end{aligned}$$

Substituting (4), (A1), (A2) and (A3) in the following formula, skewness is calculated.

$$\begin{aligned} Skew(A) &= E\left(\frac{A - E(A)}{\sqrt{Var(A)}}\right)^3 \\ &= \frac{E(A^3) - E(A)(3E(A^2) - 2(E(A))^2)}{\frac{3}{(Var(A))^{\frac{3}{2}}}} \\ &= \frac{\frac{2.4088n^2 + 1.0675n + 0.1455}{n^2} - 1.3405 \left( 3 \frac{0.2655}{n} + 1.7969 - 2 \cdot 1.7969 \right)}{\frac{3}{\left(\frac{0.2655}{n}\right)^{\frac{3}{2}}}} \\ &= \frac{35.2148n^2 - 0.0015n + 1.0636}{\sqrt{n}}. \end{aligned}$$

Analogously,

$$E\left((zone(X_k))^4\right) = 7.0604, k = 1, \dots, n$$

and

$$\begin{aligned}
 E(A^4) &= E\left(\left(\frac{1}{n}\sum_{i=1}^n \text{zone}(X_i)\right)^4\right) \\
 &= \frac{1}{n^4}\left(\sum_{i=1}^n E\left((\text{zone}(X_i))^4\right) + 4\sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n E\left((\text{zone}(X_i))^3\right)E(\text{zone}(X_j))\right. \\
 &\quad + 6\sum_{\{i,j\} \subset \{1,\dots,n\}} E\left((\text{zone}(X_i))^2\right)E\left((\text{zone}(X_j))^2\right) \\
 &\quad \left. + 12\sum_{i=1}^n \sum_{\substack{\{j,k\} \subset \{1,\dots,n\} \\ j,k \neq i}} \left((\text{zone}(X_i))^3\right)E(\text{zone}(X_j))E(\text{zone}(X_k))\right) \tag{A4} \\
 &\quad + 24\sum_{\{i,j,k,l\} \subset \{1,\dots,n\}} E(\text{zone}(X_i))E(\text{zone}(X_j))E(\text{zone}(X_k))E(\text{zone}(X_l))\Big) \\
 &= \frac{1}{n^4}\left(n \cdot 7.0604 + 4n(n-1)3.6218 \cdot 2.0624 + 6\binom{n}{2}2.0624^2 + 12n\binom{n-1}{2}3.6218 \cdot 1.3405^2\right. \\
 &\quad \left.+ 24\binom{n}{4}1.3405^4\right) \\
 &= \frac{3.2290n^3 + 29.3619n^2 + 39.1090n - 54.9525}{n^3}
 \end{aligned}$$

are calculated. Finally, substituting (4), (A1)–(A4) in the following formula, kurtosis is calculated.

$$\begin{aligned}
 \text{Kurt}(A) &= E\left(\frac{A - E(A)}{\sqrt{\text{Var}(A)}}\right)^4 \\
 &= \frac{E(A^4) - 4E(A^3)E(A) + 6E(A^2)(E(A))^2 - 3(E(A))^4}{(\text{Var}(A))^2} \\
 &= \left(\frac{3.2290n^3 + 29.3619n^2 + 39.1090n - 54.9525}{n^3} - 4\frac{2.4088n^2 + 1.0675n + 0.1455}{n^2}1.3405\right. \\
 &\quad \left.+ 6\left(\frac{0.2655}{n} + 1.7969\right)1.3405^2 - 3 \cdot 1.3405^4\right)\left(\frac{n^2}{0.2655^2}\right) \\
 &= \frac{-0.0071n^3 + 375.8936n^2 + 543.6709n - 779.4681}{n}.
 \end{aligned}$$

□

### Appendix B

MATLAB codes for the (approximate) CDF of the Zone distribution (with minor changes to the PMF also programmed), realized value of the Zone test statistic along with its approximate *p*-value, and the Monte Carlo simulations we have performed are available at the following link: <https://drive.google.com/drive/folders/1TLi1atLd7LOpnrtD3xB3FFtY8FdFfomt?usp=sharing> (accessed on 21 November 2023).

### References

1. Hogg, R.V.; McKean, J.W.; Craig, A.T. *Introduction to Mathematical Statistics*, 8th ed.; Pearson Education, Inc.: Boston, MA, USA, 2019.
2. Aboraya, M.; Yousof, H.M.; Hamedani, G.G.; Ibrahim, M. A New Family of Discrete Distributions with Mathematical Properties, Characterizations, Bayesian and Non-Bayesian Estimation Methods. *Mathematics* **2020**, *8*, 1648. [CrossRef]
3. Gillariose, J.; Balogun, O.S.; Almetwally, E.M.; Sherwani, R.A.K.; Jamal, F.; Joseph, J. On the Discrete Weibull Marshall–Olkin Family of Distributions: Properties, Characterizations, and Applications. *Axioms* **2021**, *10*, 287. [CrossRef]

4. Avdović, A.; Jevremović, V. Quantile-Zone Based Approach to Normality Testing. *Mathematics* **2022**, *10*, 1828. [[CrossRef](#)]
5. Fabiano, N.; Gardašević-Filipović, M.; Mirkov, N.; Todorčević, V.; Radenović, S. On the Distribution of Kurepa's Function. *Axioms* **2022**, *11*, 388. [[CrossRef](#)]
6. Hassan, A.; Shalbaf, G.A.; Bilal, S.; Rashid, A. A New Flexible Discrete Distribution with Applications to Count Data. *J. Stat. Theory Appl.* **2020**, *19*, 102–108. [[CrossRef](#)]
7. Vaz, S.; Torres, D.F.M. A Discrete-Time Compartmental Epidemiological Model for COVID-19 with a Case Study for Portugal. *Axioms* **2021**, *10*, 314. [[CrossRef](#)]
8. Almetwally, E.M.; Abdo, D.A.; Hafez, E.H.; Jawa, T.M.; Sayed-Ahmed, N.; Almongy, H.M. The new discrete distribution with application to COVID-19 Data. *Results Phys.* **2021**, *32*, 10498. [[CrossRef](#)] [[PubMed](#)]
9. Korkmaz, M.Ç.; Chesneau, C.; Korkmaz, Z.S. On the Arcsecant Hyperbolic Normal Distribution. Properties, Quantile Regression Modeling and Applications. *Symmetry* **2021**, *13*, 117. [[CrossRef](#)]
10. Nafidi, A.; Bahij, M.; Gutiérrez-Sánchez, R.; Achchab, B. Two-Parameter Stochastic Weibull Diffusion Model: Statistical Inference and Application to Real Modeling Example. *Mathematics* **2020**, *8*, 160. [[CrossRef](#)]
11. Jevremović, V.; Avdović, A. Control Charts Based on Quantiles—New Approaches. *Sci. Publ. State Univ. Novi Pazar Ser. A Appl. Math. Inform. Mech.* **2020**, *12*, 99–104. [[CrossRef](#)]
12. Jevremović, V.; Avdović, A. Empirical Distribution Function as a Tool in Quality Control. *Sci. Publ. State Univ. Novi Pazar Ser. A Appl. Math. Inform. Mech.* **2020**, *12*, 37–46. [[CrossRef](#)]
13. Ajadi, J.O.; Zwetsloot, I.M.; Tsui, K.-L. A New Robust Multivariate EWMA Dispersion Control Chart for Individual Observations. *Mathematics* **2021**, *9*, 1038. [[CrossRef](#)]
14. Hu, X.; Sun, G.; Xie, F.; Tang, A. Monitoring the Ratio of Two Normal Variables Based on Triple Exponentially Weighted Moving Average Control Charts with Fixed and Variable Sampling Intervals. *Symmetry* **2022**, *14*, 1236. [[CrossRef](#)]
15. Owens, A.O.; Rioborue, A.B. Control Chart and Its Application in Modelling Body Mass Index (BMI) of Students in Delta State Polytechnic, Oghara. *Am. J. Theor. Appl. Stat.* **2022**, *11*, 19–26. [[CrossRef](#)]
16. Kotb, K.A.M.; El-Ashkar, H.A. Quality Control for Feedback M/M/1/N Queue with Balking and Retention of Reneged Customers. *Filomat* **2020**, *34*, 167–174. [[CrossRef](#)]
17. Veljkovic, K.; Elfaghihe, H.; Jevremovic, V. Economic Statistical Design of X Bar Control Chart for Non-Normal Symmetric Distribution of Quality Characteristic. *Filomat* **2015**, *29*, 2325–2338. [[CrossRef](#)]
18. Thode, H.C., Jr. *Testing for Normality*; Marcel Dekker AG: Basel, Switzerland, 2002.
19. Arnastauskaitė, J.; Ruzgas, T.; Bražėnas, M. An Exhaustive Power Comparison of Normality Tests. *Mathematics* **2021**, *9*, 788. [[CrossRef](#)]
20. Thadewald, T.; Büning, H. Jarque–Bera Test and its Competitors for Testing Normality—A Power Comparison. *J. Appl. Stat.* **2007**, *34*, 87–105. [[CrossRef](#)]
21. Noughabi, H.A. A Comprehensive Study on Power of Tests for Normality. *J. Stat. Theory Appl.* **2018**, *17*, 647–660. [[CrossRef](#)]
22. Sürücü, B. A power comparison and simulation study of goodness-of-fit tests. *Comput. Math. Appl.* **2008**, *56*, 1617–1625. [[CrossRef](#)]
23. Gel, Y.R.; Miao, W.; Gastwirth, J.L. Robust directed tests of normality against heavy-tailed alternatives. *Comput. Stat. Data Anal.* **2007**, *51*, 2734–2746. [[CrossRef](#)]
24. Bakshaev, A. Goodness of fit and homogeneity tests on the basis of N-distances. *J. Stat. Plan. Inference* **2009**, *139*, 3750–3758. [[CrossRef](#)]
25. Coin, D. A goodness-of-fit test for normality based on polynomial regression. *Comput. Stat. Data Anal.* **2008**, *52*, 2185–2198. [[CrossRef](#)]
26. Desgagné, A.; Lafaye de Micheaux, P. A powerful and interpretable alternative to the Jarque–Bera test of normality based on 2nd-power skewness and kurtosis, using the Rao's score test on the APD family. *J. Appl. Stat.* **2017**, *45*, 2307–2327. [[CrossRef](#)]
27. Anderson, T.W.; Darling, D.A. Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. *Ann. Math. Stat.* **1952**, *23*, 193–212. [[CrossRef](#)]
28. Cochran, W.G. The  $\chi^2$  Test of Goodness of Fit. *Ann. Math. Stat.* **1952**, *23*, 315–345. [[CrossRef](#)]
29. Obradović, M.; Jovanović, M.; Milošević, B. Goodness-of-fit tests for Pareto distribution based on a characterization and their asymptotics. *Statistics* **2014**, *49*, 5–1026. [[CrossRef](#)]
30. MATLAB Help Center. Creating and Controlling a Random Number Stream. Available online: <https://www.mathworks.com/help/matlab/math/creating-and-controlling-a-random-number-stream.html> (accessed on 21 November 2023).
31. Gentle, J.E. *Random Numbers Generation and Monte Carlo Methods*, 2nd ed.; George Mason University: Fairfax, VA, USA, 2002.
32. Ritter, F.E.; Schoelles, M.J.; Quigley, K.S.; Klein, L.C. Determining the Number of Simulation Runs: Treating Simulations as Theories by Not Sampling Their Behavior. In *Human-in-the-Loop Simulations*; Springer: London, UK, 2011.
33. Massey, F.J., Jr. The Kolmogorov–Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68–78. [[CrossRef](#)]
34. Lilliefors, H.W. On the Kolmogorov–Smirnov Test for Normality with Mean and Variance Unknown. *J. Am. Stat. Assoc.* **1967**, *62*, 399–402. [[CrossRef](#)]

35. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611. [[CrossRef](#)]
36. Anderson, T.W.; Darling, D.A. A Test of Goodness of Fit. *J. Am. Stat. Assoc.* **1954**, *49*, 765–769. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.