International Journal of Mathematics and
Computer Science, **17**(2022), no. 3, 1327–1339

Assessment of Anderson-Darling and their Modified Tests for right skewed distribution

Jutaporn Neamvonk, Bumrungsak Phuenaree

Department of Mathematics
Faculty of Science
Burapha University
Chonburi, Thailand

email:jutaporn@buu.ac.th, bumrungsak@buu.ac.th

(Received March 13, 2022, Accepted April 25, 2022)

Abstract

The objective of this study is to assess performance of goodness of fit of Anderson-Darling test and modified Anderson-Darling tests. Ahmad, Sinclair, and Spurr [1], Zhang [4], and Seathow & Neamvonk [3] modified Anderson-Darling test for testing right skewed distribution, including Lognormal Gamma and Weibull distribution. Critical values of the 4 tests are estimated through a simulation study. These values are applied to study type I error probability and power of the tests with sample size of 10, 20, 30, 50, 100, and 200, and significant level of 0.01 and 0.05. The results show that all tests can control type I error probability, close to the significant level. In testing whether data are Lognormal distribution, the Zhang modified Anderson-Darling test produces the highest power value in all sets of parameters, sample sizes and significant levels. In testing Gamma distribution, the Zhang modified Anderson-Darling test has the most powerful in all set of parameters, significant levels and sample sizes. However, the Seathow & Neamvonk modified Anderson-Darling test provides higher power than the Zhang one when the sample size is 10 and alternative hypothesis is Lognormal distribution. In testing Weibull distribution, the Seathow & Neamvonk modified Anderson-Darling test is more powerful than

Key words and phrases: Goodness of fit test, Anderson-Darling and Modification, Power of test, Right skewed distribution.
AMS (MOS) Subject Classifications: 6208, 62E17.
ISSN 1814-0432, 2022, http://ijmcs.future-in-tech.net

the others when the sample size is 10. The Zhang modified Anderson-Darling test has the most power when the sample sizes are 20 30 50 100 and 200.

1 Introduction

Let X be a continuous random variable with cumulative distribution function $F(x)$ and let $x_1, x_2, \cdots, x_n$ be a random sample of X with order statistics $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$. The test hypotheses whether the data are from a defined distribution is stated as

$$H_0 : F(x) = F_0(x),\ \forall x$$

$$H_a : F(x) \neq F_0(x),\ \exists x$$

where $F_0(x)$ is the defined cumulative distribution function and $x \in (-\infty, \infty)$.

Zhang [4] presented the goodness of fit test statistic as

$$Z = \int_{-\infty}^{\infty} Z_x dw(x) \tag{1.1}$$

or

$$Z_{\texttt{max}} = \sup_{x \in (-\infty,\infty)} \left\{ Z_x w(x) \right\}. \tag{1.2}$$

The $w(x)$ is weight function and Z_x is replaced by Chi-square statistic (χ^2_x) or likelihood ratio statistic (G^2_x) as shown below

$$\chi^2_x = \frac{nF_n(x) - F_0(x)^2}{F_0(x)1 - F_0(x)} \tag{1.3}$$

and

$$G^2_x = 2n\left[F_n(x)\texttt{log}\left\{\frac{F_n(x)}{F_0(x)}\right\} + \{1 - F_n(x)\}\,\texttt{log}\left\{\frac{1 - F_n(x)}{1 - F_0(x)}\right\}\right] \tag{1.4}$$

where $F_0(x)$ and $F_n(x)$ are hypothesized and empirical distribution function respectively.

Goodness of fit tests have been around for long time with Anderson-Darling [1] standing out in the literature. Consequently, there are a few modified Anderson-Darling tests including those by Ahmad et al.[1], Zhang [4], and Saethow & Neamvonk [3]. Evolution of the four tests are shown below:

1. Anderson-Darling test (A^2) [1] is developed by replacing Z_x by χ_x^2 in 1.1 and $dw(x) = dF_0(x)$. We have

$$A^2 = \int_{-\infty}^{\infty} \chi_x^2 dw(x)$$
$$= \int_{-\infty}^{\infty} \frac{nF_n(x) - F_0(x)^2}{F_0(x)1 - F_0(x)} dF_0(x).$$

Therefore,

$$A^2 = -\frac{2}{n}\sum_{i=1}^{n}\left[\left(i-\frac{1}{2}\right)\texttt{log}\left\{F_0(x_{(i)})\right\} + \left(n-i+\frac{1}{2}\right)\texttt{log}\left\{1-F_0\left(X_{(i)}\right)\right\}\right] - n$$

2. Ahmad et al.[2] modified the Anderson-Darlng test through emphasis at the upper and lower tail of distribution. If we consider a left skewed distribution, then the test statistic is developed by replace Z_x by χ_x^2 in 1.1 and the weight function is defined as $dw(x) = \{1 - F_0(x)\} dF_0(x)$. Then, the Ahmad et al. test statistic (AL^2) is shown as

$$AL^2 = \int_{-\infty}^{\infty} \chi_x^2 dw(x)$$
$$= \int_{-\infty}^{\infty} \frac{nF_n(x) - F_0(x)^2}{F_0(x)1 - F_0(x)} \{1 - F_0(x)\} dF_0(x).$$

Therefore,

$$AL^2 = -\frac{3n}{2} + 2\sum_{i=1}^{n} F_0\left(X_{(i)}\right) - \sum_{i=1}^{n}\left[\left(\frac{2i-1}{n}\right)\texttt{log}\left\{F_0\left(X_{(i)}\right)\right\}\right]$$

If we consider a right skewed distribution, then the weight function is replaced by $dw(x) = F_0(x)dF_0(x)$. Then, the test statistic (AU^2) is shown as

$$AU^2 = \int_{-\infty}^{\infty} \chi_x^2 dw(x)$$
$$= \int_{-\infty}^{\infty} \frac{nF_n(x) - F_0(x)^2}{F_0(x)1 - F_0(x)} \{F_0(x)\} dF_0(x).$$

Therefore,

$$AU^2 = \frac{n}{2} - 2\sum_{i=1}^{n} F_0\left(X_{(i)}\right) - \sum_{i=1}^{n}\left[\left\{2 - \left(\frac{2i-1}{n}\right)\right\}\texttt{log}\left\{1 - F_0\left(X_{(i)}\right)\right\}\right]$$

Note that the sum of these two statistics is the original A^2 test statistic, that is $A^2 = AU^2 + AL^2$.

3. Zhang [4] developed modified Anderson-Darling test by using likelihood ratio statistic and a new adjusted weight function. The test statistics is replaced Z_x by G_x^2 in 1.1 and weighted function is $dw(x) = \frac{1}{F_n(x)\left\{1 - F_n(x)\right\}} dF_n(x)$. Hence, the Zhang test statistic (Z_A) is defined as

$$\begin{aligned} Z_A &= \int_{-\infty}^{\infty} G_x^2 dw(x) \\ &= \int_{-\infty}^{\infty} G_x^2 \frac{1}{F_n(x)\left\{1 - F_n(x)\right\}} dF_n(x) \end{aligned}$$

Then, Zhang test statistics is shown below

$$Z_A = 2\sum_{i=1}^{n} \left[\frac{n}{n - i + \frac{1}{2}} \texttt{log} \left\{ \frac{i - \frac{1}{2}}{nF_0(X_{(i)}} \right\} + \frac{n}{i - \frac{1}{2}} \texttt{log} \left\{ \frac{n - i + \frac{1}{2}}{n\left\{1 - F_0(X_{(i)}\right\}} \right\} \right]$$

4. Saethow & Neamvonk [3] proposed modified Anderson-Darling test by adapting Ahmad et al.(1998) and Zhang (2002). The test statistic is replaced by G_X^2 and also consider the tail of distribution. When the considered distribution is left skewed distribution, the weight function is $\frac{1}{F_n(x)} dF_n(X)$. The Saethow & Neamvonk test statistic is

$$\begin{aligned} Z_{AL} &= \int_{-\infty}^{\infty} G_x^2 dw(x) \\ &= \int_{-\infty}^{\infty} G_x^2 \frac{1}{F_n(x)} dF_n(x) \end{aligned}$$

Then, Saethow & Neamvonk test statistic for testing left skewed distribution is shown below

$$Z_{AL} = 2\sum_{i=1}^{n} \left[\texttt{log} \left\{ \frac{F_n(X_{(i)})}{F_0(X_{(i)})} \right\} + \frac{1 - F_n(X_{(i)}}{F_n(X_{(i)}} \texttt{log} \left\{ \frac{1 - F_n(X_{(i)})}{1 - F_0(X_{(i)})} \right\} \right]$$

and $F_n(X_{(i)}) = \dfrac{i - \frac{1}{2}}{n}$.

If we consider a right skewed distribution, then the weight function is $\dfrac{1}{1-F_n(x)} dF_n(X)$. The test statistic is as follows

$$\begin{aligned} Z_{AU} &= \int_{-\infty}^{\infty} G_x^2 dw(x) \\ &= \int_{-\infty}^{\infty} G_x^2 \frac{1}{1-F_n(x)} dF_n(x) \end{aligned}$$

Then, Saethow & Neamvonk test statistics for testing right skewed distribution is shown below

$$Z_{AU} = 2\sum_{i=1}^{n} \left[\frac{F_n(X_{(i)})}{1-F_n(X_{(i)})} \log\left\{ \frac{F_n(X_{(i)})}{F_0(X_{(i)})} \right\} + \log\left\{ \frac{1-F_n(X_{(i)})}{1-F_0(X_{(i)})} \right\} \right]$$

In this paper, we will assess the four goodness of fit tests for testing right skewed, Lognormal Gamma and Weibull distributions. The critical values, type I error Probability and the power of the test are estimated through Monte Carlo simulation.

2 Main Results

In estimating the probability distribution of statistic tests using Monte Carlo simulation, we generate the Lognormal, Gamma, and Weibull distributed random numbers and calculate the A^2, AU^2, Z_A, Z_{AU} statistics. This process is repeated 100,000 times. For each statistic, the 100,000 values of test statistics are ranked in ascending order. The 99^{th} and 95^{th} percentile of ordered statistics are the 0.01 and 0.05 significant level respectively of critical values presented in Tables 1-3 .

In the next step, we apply the critical values to the four tests in order to examine the type I error probability of the tests; A^2, AU^2, Z_A, Z_{AU}. We generate random numbers from the assumed distribution; i.e., Lognormal, Gamma, and Weibull distribution. Without loss of generality, parameters of each distribution are varied at least four sets according to the skewness of the distribution. The test statistics for each set of parameters are calculated with 10,000 replication. The numbers of rejecting the assumed distribution

Table 1: Critical values for testing Lognormal distribution

n	α =0.01				α =0.05			
	A^2	AU^2	Z_A	Z_{AU}	A^2	AU^2	Z_A	Z_{AU}
10	0.9740	0.5085	8.3652	4.6421	0.7237	0.3750	5.6179	3.0248
20	1.0060	0.5299	11.3942	6.5690	0.7384	0.3840	7.8146	4.3190
30	1.0130	0.5344	13.2150	7.8211	0.7424	0.3860	9.1554	5.1592
50	1.0244	0.5402	15.5738	9.4978	0.7461	0.3887	10.9549	6.2729
100	1.0277	0.5432	19.0076	11.9964	0.7488	0.3898	13.5278	7.8431
200	1.0278	0.5454	22.5588	14.5394	0.7490	0.3900	16.1101	9.4052

Table 2: Critical values for testing Gamma distribution

n	α =0.01				α =0.05			
	A^2	AU^2	Z_A	Z_{AU}	A^2	AU^2	Z_A	Z_{AU}
10	1.0112	0.5236	8.2686	4.4977	0.7445	0.3831	5.5504	2.9232
20	1.0509	0.5495	11.2669	6.3735	0.7652	0.3953	7.7279	4.1727
30	1.0641	0.5577	13.0172	7.5678	0.7706	0.3987	9.0581	4.9871
50	1.0732	0.5635	15.3655	9.2200	0.7755	0.4011	10.8309	6.0736
100	1.0782	0.5667	18.7300	11.6002	0.7776	0.4025	13.3301	7.6054
200	1.0834	0.5691	22.3343	14.0995	0.7803	0.4046	15.9549	9.1881

are counted and divided by 10,000. We average the type I error probabilities of various set parameter with the same sample size. The results are shown in Tables 4-6 representing that the type I error probabilities of the four tests across different sample sizes are consistently close to the related significance level and also in the range of Cocharn criteria [5]; i.e., (0.007, 0.015) for $\alpha = 0.01$ and (0.04, 0.06) for $\alpha = 0.05$. This signifies that the critical values are reliable to the tested distributions.

To investigate the power of the tests, we generate random numbers from an alternative hypothesized distribution. The test statistics are again calculated with 10,000 replications. The number of rejecting the null distribution are counted and divided by 10,000. The power of the tests for testing Lognormal distribution is presented in Figure 1. We can see that the Z_A test is the most powerful for all sample sizes and significance levels. The Z_{AU} is the second highest one when the sample size is 50 or more. The A^2 and AU^2

Table 3: Critical values for testing Weibull distribution

	α =0.01				α =0.05			
n	A^2	AU^2	Z_A	Z_{AU}	A^2	AU^2	Z_A	Z_{AU}
10	0.9807	0.4698	8.6625	3.7066	0.7262	0.3496	5.8030	2.4822
20	1.0107	0.4950	11.7403	5.3731	0.7416	0.3612	8.0264	3.5743
30	1.0247	0.5037	13.6234	6.4539	0.7471	0.3651	9.4029	4.2983
50	1.0283	0.5080	16.0765	7.9344	0.7501	0.3672	11.1933	5.2830
100	1.0392	0.5135	19.5367	10.1391	0.7536	0.3698	13.7321	6.7370
200	1.0394	0.5149	23.1673	12.5546	0.7538	0.3708	16.2986	8.2221

Table 4: Type I error probability of the tests for testing Lognormal distribution

	α =0.01				α =0.05			
n	A^2	AU^2	Z_A	Z_{AU}	A^2	AU^2	Z_A	Z_{AU}
10	0.0092	0.0097	0.0091	0.0102	0.0493	0.0498	0.0493	0.0503
20	0.0098	0.0099	0.0095	0.0092	0.0508	0.0506	0.0507	0.0515
30	0.0101	0.0106	0.0102	0.0108	0.0504	0.0505	0.0505	0.0495
50	0.0100	0.0102	0.0105	0.0099	0.0506	0.0502	0.0506	0.0505
100	0.0102	0.0101	0.0101	0.0097	0.0509	0.0503	0.0504	0.0515
200	0.0102	0.0104	0.0102	0.0101	0.0509	0.0502	0.0502	0.0505

Table 5: Type I error probability of the tests for testing Gamma distribution

	α =0.01				α =0.05			
n	A^2	AU^2	Z_A	Z_{AU}	A^2	AU^2	Z_A	Z_{AU}
10	0.0080	0.0086	0.0095	0.0104	0.0445	0.0459	0.0482	0.0530
20	0.0083	0.0082	0.0101	0.0112	0.0459	0.0466	0.0520	0.0541
30	0.0087	0.0088	0.0102	0.0120	0.0457	0.0449	0.0497	0.0524
50	0.0077	0.0083	0.0095	0.0103	0.0470	0.0484	0.0536	0.0542
100	0.0088	0.0090	0.0105	0.0111	0.0456	0.0463	0.0529	0.0532
200	0.0080	0.0089	0.0097	0.0118	0.0472	0.0468	0.0515	0.0546

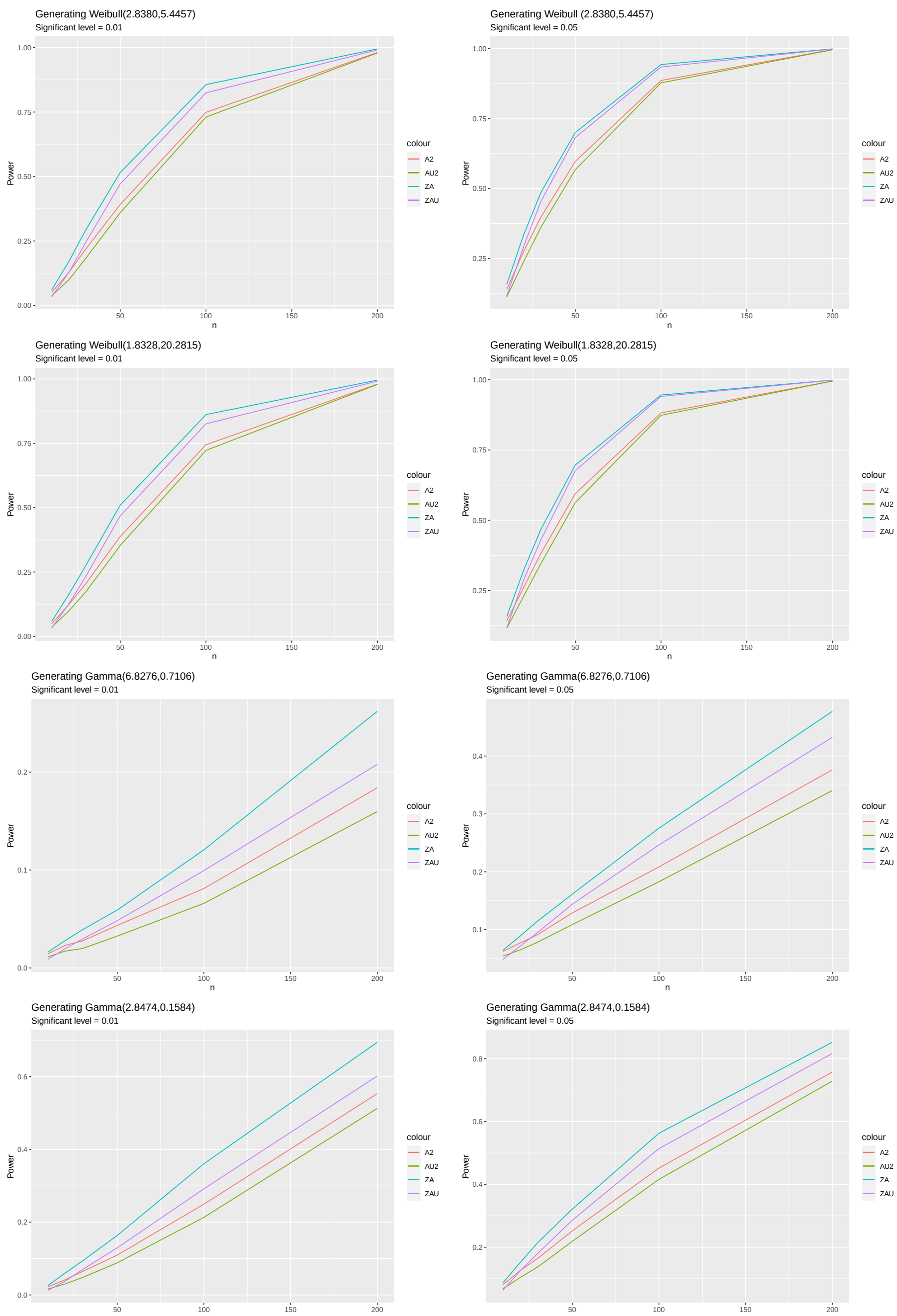

Figure 1: Power of the tests for Lognormal distribution at significant level of 0.01 and 0.05

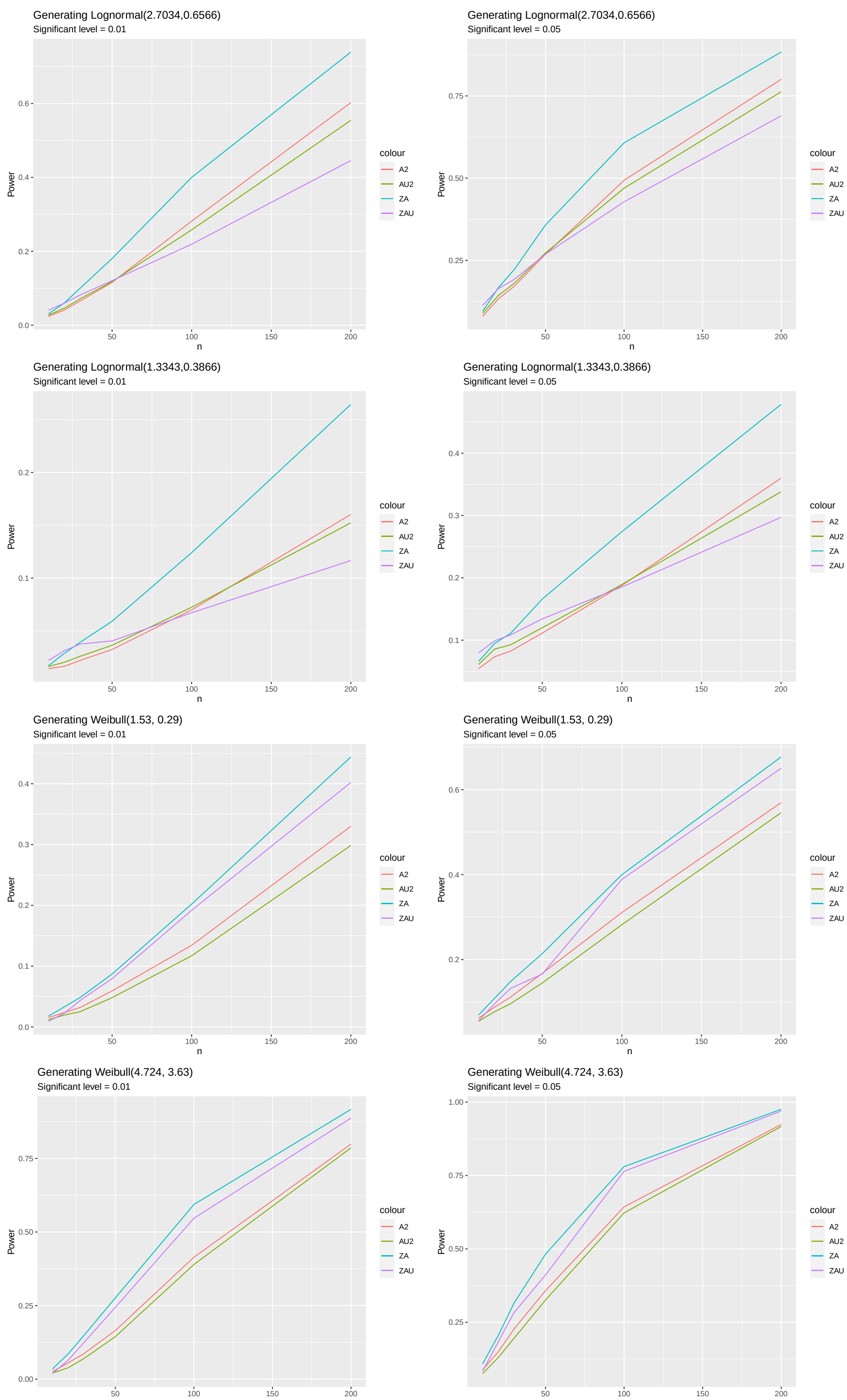

Figure 2: Power of the tests for Gamma distribution at significant level of 0.01 and 0.05

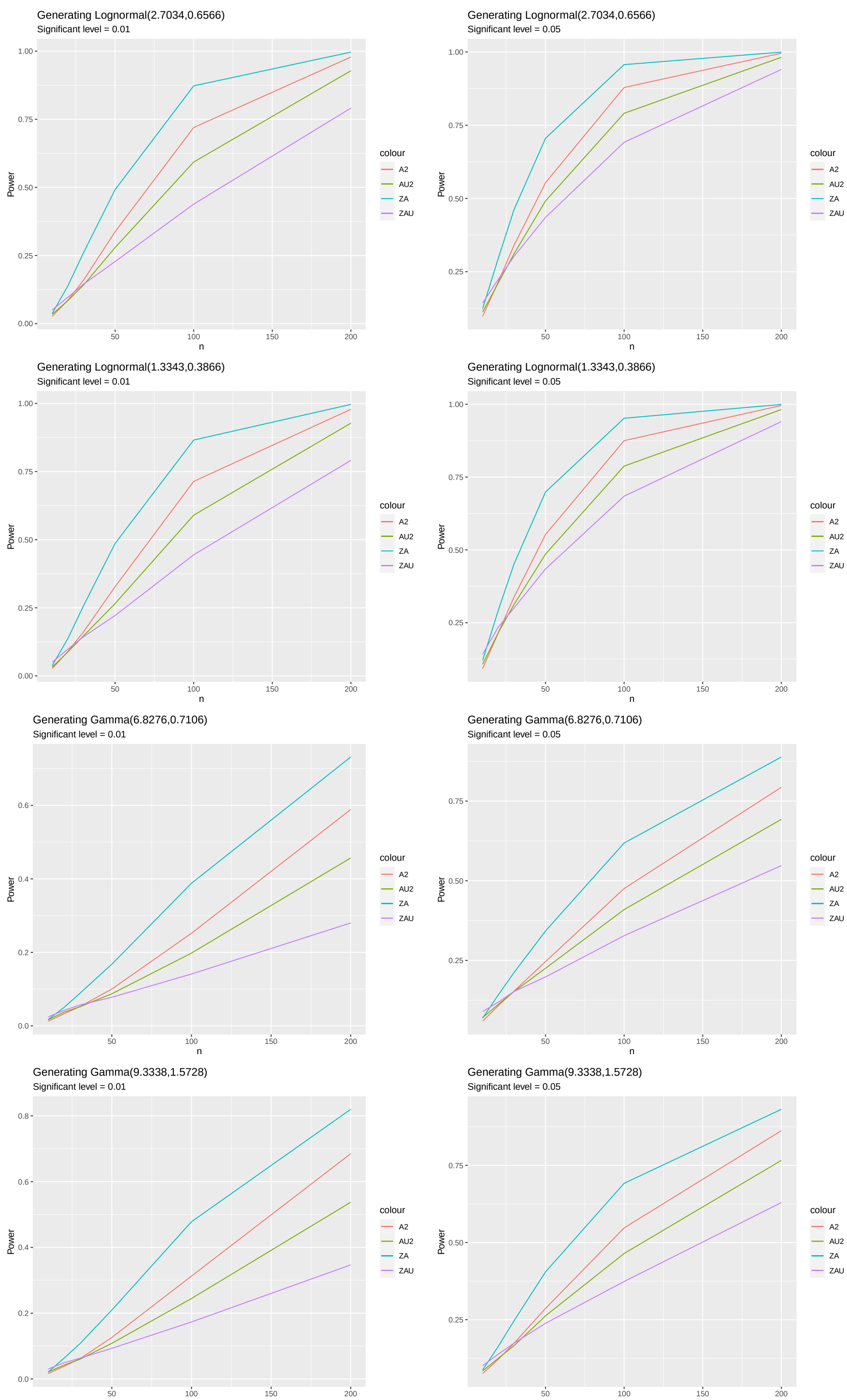

Figure 3: Power of the tests for Weibull distribution at significant level of 0.01 and 0.05

Table 6: Type I error probability of the tests for testing Weibull distribution

	α =0.01				α =0.05			
n	A^2	AU^2	Z_A	Z_{AU}	A^2	AU^2	Z_A	Z_{AU}
10	0.0097	0.0093	0.0097	0.0101	0.0514	0.0517	0.0507	0.0525
20	0.0112	0.0100	0.0106	0.0095	0.0483	0.0508	0.0486	0.0507
30	0.0103	0.0099	0.0097	0.0100	0.0491	0.0499	0.0481	0.0507
50	0.0099	0.0094	0.0107	0.0103	0.0513	0.0527	0.0518	0.0508
100	0.0093	0.0091	0.0093	0.0102	0.0493	0.0487	0.0502	0.0512
200	0.0099	0.0090	0.0100	0.0094	0.0492	0.0510	0.0483	0.0483

have similar power when the alternative hypothesis is Weibull distribution; however, the power of A^2 is higher than AU^2 when the hypothesis is Gamma distribution.

For testing Gamma and Weibull distribution in Figures 2 and 3, the power of the Z_A tests is, generally, the most powerful than others. The Z_{AU} is superior than others when sample size is small as 10.

3 Conclusion

In this paper, we assessed the performance of the Anderson-Darling and their modification tests for a right skewed distribution. The critical values for the four tests were evaluated through simulations. The results showed that the Z_A test were generally the most powerful among the original test and other modified one. The Z_{AU} was superior than others only when the sample size was 10.

In order to apply these tests more conveniently, we find equations presenting the relationship between the critical values and sample sizes as shown in Tables 7-8. The sample size well explains the critical values as the r^2s are higher than 0.99.

Table 7: Critical value functions of Z_A and Z_{AU} tests for $\alpha = 0.01$

Distribution	Z_A test	Z_{AU} test
Lognormal	$y = -2.775 + 4.742\texttt{log}(n)$ $r^2 = 0.9988$	$y = -3.311 + 3.328\texttt{log}(n)$ $r^2 = 0.9967$
Gamma	$y = -2.7672 + 4.689\texttt{log}(n)$ $r^2 = 0.9984$	$y = -3.2062 + 3.2243\texttt{log}(n)$ $r^2 = 0.9966$
Weibull	$y = -2.7277 + 4.849\texttt{log}(n)$ $r^2 = 0.9989$	$y = -3.4306 + 2.965\texttt{log}(n)$ $r^2 = 0.9947$

Table 8: Critical value functions of Z_A and Z_{AU} tests for $\alpha = 0.05$

Distribution	Z_A	Z_{AU}
Lognormal	$y = -2.6783 + 3.5199\texttt{log}(n)$ $r^2 = 0.9987$	$y = -2.0565 + 2.148\texttt{log}(n)$ $r^2 = 0.9985$
Gamma	$y = -2.6541 + 3.4811\texttt{log}(n)$ $r^2 = 0.9986$	$y = -2.0766 + 2.1057\texttt{log}(n)$ $r^2 = 0.9979$
Weibull	$y = -2.4582 + 3.5179\texttt{log}(n)$ $r^2 = 0.9991$	$y = -2.1542 + 1.933\texttt{log}(n)$ $r^2 = 0.9961$

Acknowledgment. This work was financially supported by the Research Grant of Burapha University through National Research Council of Thailand (Grant no. 154/2560) and Faculty of Science, Burapha University.

References

[1] M. I. Ahmad, C. D. Sinclair, B. D. Suprr, Assessment of flood frequency models using empirical distribution function statistics, Water resource research, **24**, no. 8, (1988), 1323–1328.

[2] T. W. Anderson, D. A. Darling, Asymptotic theory of certain "goodness of fit" criteria based on stochastic process, Ann. Math. Statist. Soc. B, **23**, (1952), 193–212.

[3] P. Saethow, J. Neamvonk, Modified Anderson-Darling test based on the likelihood ratio for skewed distributions, International Conference on Applied Statistics (Thailand), May 21–24, 2014, 27–34.

[4] J. Zhang, Powerful goodness of fit test based on likelihood ratio, J. R. Statist. Soc. B, **64**, no. 2, (2002), 281–294.

[5] W. G. Cocharn, Some methods of strengthening the common chi-square tests, Biometrics, **10**, (1954), 417–451.