Thai Journal of Mathematics
Volume 22 Number 4 (2024)
Pages 595–604

http://thaijmath.in.cmu.ac.th

ISSN 1686-0209

Objective Judgement on Q-Q Plot for Testing Laplace Distribution

Chaichana Jaiboon[1] **and Wanpen Chantarangsi**[2,*]

[1] *Faculty of Science and Technology, Rajamangala University of Technology Rattanakosin, Nakhon Pathom, 73170, Thailand*
email : chaichana.jai@rmutr.ac.th

[2] *Faculty of Science and Technology, Nakhon Pathom Rajabhat University, Nakhon Pathom, 73000, Thailand*
e-mail : wanpen@webmail.npru.ac.th

Abstract Q-Q plot is one of the graphical tools for evaluating the validity of an underlying distribution which compares quantiles of a sample against the quantiles of the distribution. In this paper, we focus on how a Q-Q plot can be augmented by intervals for all the points so that, if the population distribution is Laplace then all the points should fall into the corresponding intervals simultaneously with probability $1-\alpha$. Therefore, the major concern of this paper is to construct the objective mean based graphical tests to assess whether the plotted points fall close to the straight line. Simulations are used to compare powers among graphical tests and some popular non-graphical Anderson-Darling and Cramer-von-Mises tests. Based on this power study, the tests which have higher powers are recommended to apply real world problems and should be used in what circumstances. Finally, the example is provided to illustrate the methods.

MSC: 62J15; 62E20; 62A09

Keywords: exponential distribution; graphical methods; hypothesis testing; Laplace distribution; power; Q-Q plot; simultaneous intervals; statistical simulation

Submission date: 12.06.2021 / Acceptance date: 01.02.2022

1. Introduction

When a simple random sample $Y_1, \cdots, Y_n$ is drawn from a population, one important question is whether the population has a distribution of the form $F_0\left((y-\mu)/\sigma\right)$, where $F_0(\cdot)$ is a given cumulative distribution function (cdf), and $-\infty < \mu < \infty$ and $\sigma > 0$ are two unknown parameters. Note that μ is not necessarily the mean and σ is not necessarily the standard deviation of Y. One widely used graphical technique for dealing with this question is the Q-Q plot.

A Q-Q plot consists of the n points $(q_k, Y_{[k]})$, $k = 1, \cdots, n$, where $Y_{[1]} \leq \cdots \leq Y_{[n]}$ are the ordered Y_k's and $q_1 < \ldots < q_n$ are a set of n reference values which represent the ordered values of a typical sample of size n from the distribution $F_0(y)$. There are several ways to choose the reference values $q_k = F_0^{-1}(p_k)$ where $F_0^{-1}(\cdot)$ is the inverse function of $F_0(\cdot)$. Various slightly different forms of p_k have been suggested in the statistical

*Corresponding author.

Published by The Mathematical Association of Thailand.

literature. See, e.g., [1–3]. Throughout this paper, we use $p_k = (k-0.5)/n$ $(k = 1, \ldots, n)$, which are firstly given in [4] and used in the software packages R (when $n > 10$) and Matlab. Note that the choices of the p_k's do not affect the tests discussed in this paper.

If $Y_1, \cdots, Y_n$ have the distribution $F_0\left((y-\mu)/\sigma\right)$, then the n points $(q_k, Y_{[k]})$ should fall close to a straight line. In order to provide an objective judgement on whether the points $(q_k, Y_{[k]})$ fall close to a straight line, one can augment the Q-Q plot by providing an interval for each $Y_{[k]}$ $(k = 1, \cdots, n)$ so that, if the population follows the distribution $F_0\left((y-\mu)/\sigma\right)$, then all the $Y_{[k]}$ $(k = 1, \cdots, n)$ will fall inside the corresponding intervals simultaneously with probability $1-\alpha$. Each of these n intervals can be depicted in the Q-Q plot as a vertical interval at the corresponding q_k. Therefore, if at least one point $(q_k, Y_{[k]})$ $(1 \le k \le n)$ does not fall within the corresponding interval then one can claim, with $1-\alpha$ confidence, that the population does not follow the distribution $F_0\left((y-\mu)/\sigma\right)$. This is in effect a size α test for the null hypothesis H_0: the population distribution is $F_0\left((y-\mu)/\sigma\right)$ for some $-\infty < \mu < \infty$ and $\sigma > 0$ against the alternative hypothesis H_a: H_0 is not true, but with a clear graphical interpretation on the Q-Q plot.

In order to provide an objective judgement on whether the points $(z_k, Y_{[k]})$ fall close to a straight line and building on the work of [5]. In [6], it considers augmenting the normal probability plot by providing an interval for each $Y_{[k]}$ $(k = 1, \ldots, n)$ so that, if the population is normally distributed then all the $Y_{[k]}$ $(k = 1, \cdots, n)$ will fall into the corresponding intervals simultaneously with probability $1-\alpha$. Next, [7] construct the simultaneous intervals on Q-Q plot plots to judge whether a sample is drawn from the Weibull or exponential distributions. For testing Weibull, the authors uses log-transformation to obtain the smallest extreme value distribution which is a member of location-scale family. The main purpose of this paper is to apply the graphical tests on Q-Q plot and to compare powers of these graphical tests in order to identify the one having larger overall power.

The layout of the paper is as follows. Section 2 presents the methods of parameter estimation for Laplace distribution. Section 3 then constructs graphical tests for testing Laplace distribution based on the tests proposed in [6]. The powers of these graphical and two non-graphical tests are then compared in a simulation study in order to identify the tests that have overall good power in Section 4. Constructing simultaneous probability intervals and illustrative example are presented in Section 5.

2. Distribution Function and Parameter Estimation

A random variable X is said to have the Laplace distribution, $L(a, b)$, if its probability density function (pdf) is given by

$$f(x|a,b) = \frac{1}{2b}\exp\Big(-\frac{|x-a|}{b}\Big), \quad -\infty < x < \infty, -\infty < a < \infty, b > 0$$

where a is called the location parameter and b the scale parameter. According to [8], the maximum likelihood estimator (MLE) of a is the sample median, $\hat{a}$, of $X_1, \ldots, X_n$ and the MLE of b is given by $\hat{b} = \frac{1}{n}\Sigma_{k=1}^{n}|X_k - \hat{a}|$. Since the Laplace distribution $L(a, b)$ is one of location-scale family, we can use the method of transformation in order to obtain the standard exponential distribution $Exp(0, 1)$, which is also a member of a location-scale family. We will then be able to use prior knowledge about the graphical tests proposed by [7] to construct the corresponding simultaneous intervals as follows.

If $Y = \frac{X-a}{b}$, by the method of transformation,

$$g(y) = \frac{1}{2b} \exp\big(-|y|\big), \quad -\infty < y < \infty,$$

and cumulative distribution function (cdf)

$$G(y) = \left\{ \begin{array}{ll} \frac{1}{2}\exp(-|y|), & y \le 0 \\ 1 - \frac{1}{2}\exp(-|y|), & y > 0 \end{array} \right\}$$

which is the so-called double exponential distribution. It is obvious that if $Z = |Y|$ then $T(z) = 2G(z) - 1$ and $t(z) = 2g(z)$ for $z > 0$ are the cdf and pdf of Z from the standard exponential distribution $Exp(0,1)$, respectively. Specifically, the p^{th} quantile of a random variable Z is given by $F^{-1}(p) = -\ln(1-p)$. The original null hypothesis $H_0 : X_1, \ldots, X_n$ come from $L(a,b)$ is therefore the same as $H_0 : Z_1 = \frac{|X_1-a|}{b}, \ldots, Z_n = \frac{|X_n-a|}{b}$ are from $Exp(0,1)$. Among the three estimators of the exponential distribution: the maximum likelihood estimatros (MLE), the best linear unbiased estimators (BLUE), the best linear invariant estimators (BLIE), [7] recommended that BLIE often gives the best power, even though the power differences between BLIE and BLUE are often small. Therefore, in this paper we employ BLIE as an estimator to test exponential.

The best linear invariant estimators (BLIE) are given in [9] by

$$\hat{\mu} = (1 + \frac{1}{n})Y_{[1]} - \frac{\bar{Y}}{n}, \tag{2.1}$$

$$\hat{\sigma} = \bar{Y} - Y_{[1]}. \tag{2.2}$$

Let $Z_{[1]} \le \ldots \le Z_{[n]}$ be the ordered sample from $Exp(0,1)$. Then we have [10] as

$$\mu_k = \mathrm{E}(Z_{[k]}) = \sum_{i=1}^{k} \frac{1}{n-i+1}, \ k = 1, \ldots, n$$

$$\sigma_k^2 = \mathrm{Var}(Z_{[k]}) = \sum_{i=1}^{k} \frac{1}{(n-i+1)^2}, \ k = 1, \ldots, n$$

$$\sigma_{rk}^2 = \mathrm{Cov}(Z_{[r]}, Z_{[k]}) = \sum_{i=1}^{k} \frac{1}{(n-i+1)^2}, \ 1 \le r \le k \le n.$$

where $\sigma_{rk}^2 = \sigma_{kr}^2 = \mathrm{Cov}(Z_{[r]}, Z_{[k]})$.

3. The Tests

For testing the Exponential distribution, the five graphical and two non-graphical tests given above are easily modified by simply assuming that $T(\cdot)$ is the cdf of $Exp(0,1)$ and that $Z_1, \ldots, Z_n$ are a simple random sample from $Exp(0,1)$ to give the five graphical and two non-graphical tests also denoted as D, D_m, D_e, D_{be}, D_{bi}, AD and CvM.

Our focus is on the five graphical tests D, D_m, D_e, D_{bi} and D_{be}, each providing a set of simultaneous $1-\alpha$ probability intervals for the $Z_{[k]}$'s. These intervals can be used in a Q-Q plot to objectively judge whether the n points $(q_k, Z_{[k]})$ fall close to a straight line. We also want to compare the powers of the five graphical and the two non-graphical tests.

3.1. The D Test

One way to construct the intervals is to use the Kolmogorov-Smirnov (1983) statistic

$$D = \max_{1\le k\le n} \left| F\left((X_{[k]} - \hat{a})/\hat{b}\right) - (k-0.5)/n \right|$$

where $\hat{a}$ and $\hat{b}$ are the estimates of a and b, respectively. Note that F is the cdf of the Laplace distribution $F(\cdot)$. As shown in Section 2, if X is a random variable following $L(a,b)$, then $Y = \frac{X-a}{b}$ follows a double exponential distribution and finally $Z = |Y|$ having the standard exponential distribution with cdf $T(\cdot)$. Therefore, the Kolmogorov-Smirnov statistic is

$$D = \max_{1\le k\le n} \left| T\left((Z_{[k]} - \hat{\mu})/\hat{\sigma}\right) - (k-0.5)/n \right|$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the BLIE shown in (2.1) and (2.2) based on the exponential, respectively. Note that D is sometimes also referred to as Lilliefors' statistic in [11]. Let c_D be a critical constant so that $P\{D \le c_D\} = 1-\alpha$ under H_0. This probability statement can be rewritten as

$$P\left\{ Z_{[k]} \in \hat{\mu} + \hat{\sigma}T^{-1}\left((k-0.5)/n \pm c_D\right), \quad k = 1, \cdots, n \right\} = 1-\alpha.$$

Hence, under H_0, each $Z_{[k]}$ should fall in the corresponding interval $\hat{\mu} + \hat{\sigma}T^{-1}\big((k-0.5)/n \pm c_D\big)$ simultaneously for $k = 1, \cdots, n$ with probability $1-\alpha$.

3.2. The D_m Test

The second set of intervals is due to [5] and based on the statistic

$$D_m = \max_{1\le k\le n} \left| (2/\pi)\arcsin\sqrt{T\left((Z_{[k]} - \hat{\mu})/\hat{\sigma}\right)} - (2/\pi)\arcsin\sqrt{(k-0.5)/n} \right|.$$

Let c_{D_m} be a critical constant so that $P\{D_m \le c_{D_m}\} = 1-\alpha$ under H_0. This probability statement can be rewritten as

$$P\left\{ Z_{[k]} \in \hat{\mu} + \hat{\sigma}T^{-1}\left(\sin^2[\arcsin\sqrt{(k-0.5)/n} \pm \frac{\pi}{2}c_{D_m}]\right) \text{ for } k = 1, \ldots, n \right\} = 1-\alpha.$$

3.3. The D_e Test

Recall that $Z_1, ..., Z_n$ denote a simple random sample drawn from $Exp(0,1)$ and $Z_{[1]} \le ... \le Z_{[n]}$ be the ordered values. The expected values and variances of $Z_{[k]}$ for $k = 1, \ldots, n$ are given by

$$\begin{aligned} \mu_k &= \mathrm{E}(Z_{[k]}), \\ \sigma_k^2 &= \mathrm{Var}(Z_{[k]}) = \mathrm{E}(Z_{[k]}^2) - \mu_k^2 \end{aligned}$$

where $t_k(z)$ is the probability density function of $Z_{[k]}$ and is defined by

$$t_k(z) = \frac{n!}{(k-1)!(n-k)!}\left(T_Z(z)\right)^{k-1}\left(1 - T_Z(z)\right)^{n-k} t_Z(z), \ z > 0.$$

The test D_e uses the test statistic

$$D_e = \max_{1\le k\le n} \left| \frac{Z_{[k]} - (\hat{\mu} + \hat{\sigma}\mu_k)}{\hat{\sigma}\sigma_k} \right|, \tag{3.1}$$

where $(\hat{\mu}, \hat{\sigma})$ is the BLIE as in (2.1) and (2.2).

It is clear from expression (3.1) that the distribution of D_e does not depend on the unknown parameters μ and σ^2. The critical constant c_e, which satisfies $P\{D_e \le c_e\} = 1-\alpha$ under H_0, can easily be computed accurately by using a large number of simulations, as in [6]. See [12] and [13] for ways to assess the accuracy of this approach. It is noteworthy that simulation methods are also used to compute the critical constants of the D and D_m tests; see, e.g., [5] and [14]. The probability statement $P\{D_e \le C_e\} = 1-\alpha$ produces the following simultaneously probability intervals for $Z_{[1]}, ..., Z_{[n]}$:

$$P\left\{\ Z_{[k]} \in [\hat{\mu} + \hat{\sigma}\mu_k \pm c_e\hat{\sigma}\sigma_k] \ \text{ for } \ k = 1, \cdots, n \ \right\} = 1 - \alpha.$$

3.4. The D_{be} Test

The D_{be} test is constructed in the following steps. Let $T(\cdot)$ denote the cdf of $Exp(0,1)$. Note that, under H_0, $U_k = T\Big(\dfrac{Z_k - \mu}{\sigma}\Big)$, $k = 1, \ldots, n$ has a uniform distribution on the interval $(0,1)$ and the order statistics $U_{[k]} = T\Big(\dfrac{Z_{[k]} - \mu}{\sigma}\Big)$ has the beta distribution with parameters k and $n-k+1$.

- **Step 1**. Construct p^* level highest-density probability interval $[L(p^*, k, n), U(p^*, k, n)]$ for $U_{[k]}$, which is the shortest probability interval for $U_{[k]}$ among all the p^* level probability intervals for $U_{[k]}$.
- **Step 2**. Find p^* so that

$$K(p^*) \equiv P\Big\{T^{-1}(L(p^*, k, n)) \le \frac{Z_{[k]} - \hat{\mu}}{\hat{\sigma}} \le T^{-1}(U(p^*, k, n)) \ \text{ for } k = 1, ..., n\Big\} \ = \ 1 - \alpha.$$

Such a p^* can be found by simulation and a standard numerical searching algorithm in a similar way as computing such p^* for normality.

- **Step 3**. Under H_0, the simultaneous $1-\alpha$ probability intervals for $Y_{[1]} \le ... \le Y_{[n]}$ are therefore given by

$$\hat{\mu} + \hat{\sigma}T^{-1}(L(p^*, k, n)) \le Y_{[k]} \le \hat{\mu} + \hat{\sigma}T^{-1}(U(p^*, k, n)), \quad k = 1, \ldots, n.$$

Hence test D_{be} rejects H_0 if and only if at least one $Z_{[k]}$ is not included in its corresponding interval $[\hat{\mu} + \hat{\sigma}T^{-1}(L(p^*, k, n)), \hat{\mu} + \hat{\sigma}T^{-1}(U(p^*, k, n))]$.

3.5. The D_{bi} Test

The D_{bi} test uses statistic

$$D_{bi} \ = \ \max_{1 \le k \le n} \frac{\big|T\big((Z_{[k]} - \hat{\mu})/\hat{\sigma}\big) - (k - 0.5)/n\big|}{\sqrt{(k - 0.5)(n - k + 0.5)/n^3}}.$$

Let c_{bi} be a critical constant so that $P\Big\{D_{bi} < c_{bi}\Big\} = 1 - \alpha$, under H_0, which can be determined by using simulation as before. The simultaneous $1-\alpha$ probability intervals for $Y_{[1]} \le ... \le Y_{[n]}$ are therefore given by

$$Z_{[k]} \in \hat{\mu} + \hat{\sigma}T^{-1}\Big(\frac{k - 0.5}{n} \pm c_{bi}\sqrt{\frac{(k - 0.5)(n - k + 0.5)}{n^3}}\Big) \ \text{ for } k = 1, \ldots, n.$$

The non-graphical Anderson-Darling (AD) test rejects H_0 if and only if AD $> c$ where

$$\text{AD} = -\sum_{k=1}^{n}\Big[\frac{(2k-1)\{\ln(T(Z_{[k]})) + \ln(1-T(Z_{[n+1-k]}\}}{n}\Big] - n. \tag{3.2}$$

The critical constant c, which satisfies $P\{\text{AD} < c\} = 1 - \alpha$ under H_0, can be determined by simulation as before.

The non-graphical Cramér-von Mises (CvM) test rejects H_0 if and only if CvM $> c$ where

$$\text{CvM} = \sum_{k=1}^{n}\Big[T(Y_{[k]}) - \frac{2k-1}{2n}\Big]^2 + \frac{1}{12n}. \tag{3.3}$$

The critical constant c, which satisfies $P\{\text{CvM} < c\} = 1 - \alpha$ under H_0, can again be determined by simulation.

4. Power Comparison

The power of a test is evaluated by simulation as the proportion of times the null hypothesis H_0 is rejected by the test for a given alternative distribution. In our simulation study, each critical value is based on 10,000 simulations as shown in Table 1 and each power value is also based on 10,000 simulations. The powers of the seven tests are computed for all possible combinations of $\alpha = 0.05$ and sample size n from a set of values, and the alternative distribution from a set of distributions in Table 2. The set of alternative distributions consists of various distributions : Normal(0,1), logistic(0,1), Cauchy(0,1), Uniform(0,1), Tukey(0,5), Beta(2,2) and EV(0,1) (Extreme value distribution) used in [15].

The AD and CvM tests are probably the best choices recommended by [7] and [16] for comparison to the graphical tests here. This is the reason why AD and CvM are included in this paper. Overall, the non-graphical AD and CvM tests perform well for almost all situations but they cannot be depicted on Q-Q plot by using them. Therefore, the practitioners will not see the deviation which may be presented on the plot. The graphical test based on D_e does better than the numerical tests based on AD and CvM for long tailed symmetric distribution: see Cauchy(0,1). Nonetheless, the D_e has the worst powers among all tests on most occasions.

Notice that for symmetric alternatives the D_{bi} are more powerful than the other tests; see, for example, Normal(0,1), Uniform(0,1), Beta(2,2) when $n \geq 50$. However, if $n \leq 20$, the D turns out to have better power than the D_{bi}. It can also be observed from the simulation study that the powers of D_{be} are similar to those of the D_m against all alternative distributions which this fact is found by [7] for testing Weibull and exponential distributions.

Table 1. Critical values of the D, D_m, D_e, D_{be}, D_{bi} tests for Laplace distribution at significance level $\alpha = 0.05$ and n =20, 50, 100

n	D	D_m	D_e	D_{bi}	AD	CvM
20	0.2086	0.1439	2.4283	2.4733	1.1783	0.2107
50	0.1396	0.1031	2.6425	2.9968	1.2462	0.216
100	0.1033	0.0786	2.7784	3.3784	1.2864	. 0.2244

TABLE 2. Powers (in %) of the D, D_m, D_e, D_{be}, D_{bi} for testing Laplace distribution at $\alpha = 0.05$ against various alternative distributions with $n = 20$, 50, 100

Alternatives	n	D	D_m	D_e	D_{be}	D_{bi}	AD	CvM
Normal(0,1)	20	12.76	10.89	9.09	11.65	7.39	9.59	13.38
	50	35.37	31.06	24.16	29.43	33.38	34.53	41.61
	100	65.15	65.1	48.87	61.29	66.78	72.23	75.82
Logistic(0,1)	20	8.72	7.55	5.94	8.1	4.64	6.13	8.4
	50	19.58	16.56	13.26	16.58	15.08	16.76	21.8
	100	36.31	32.59	26.29	31.37	26.14	37.53	42.05
Cauchy(0,1)	20	56.03	55.9	63.76	56.12	44.57	62.85	60.39
	50	89.78	87.37	92.71	87.78	74.93	94.14	91.71
	100	99.18	98.74	99.54	98.85	94.25	99.51	99.49
Uniform(0,1)	20	36.64	32.82	28.85	32.9	34.45	37.19	43.47
	50	88.34	96.59	77.32	94.54	99.07	93.93	94.78
	100	99.7	100	98.66	100	100	99.95	99.97
Tukey(0,1)	20	25.75	21.95	18.77	22.76	19.38	23.51	29.66
	50	71.86	81.17	56.36	75.68	92.25	80.11	83.04
	100	96.78	99.92	89.46	99.74	99.98	99.5	99.39
Beta(2,2)	20	24.36	20.46	16.81	21.3	16.9	21.36	27.47
	50	67.37	75.85	50.78	69.57	88.42	75.36	78.88
	100	95.24	99.78	86.31	99.56	100	99.13	99.09
EV(0,1)	20	10.43	9.35	8.66	9.97	5.89	7.72	10.47
	50	23.31	19.37	17.97	19.8	13.75	20.61	25.81
	100	41.52	35.38	33.8	35.68	19.24	43.37	47.72

5. Constructing Simultaneous Probability Intervals and the Illustrative Example

The null (H_0) : a random sample $X_1, \ldots, X_n$ is from the Laplace distribution $L(a, b)$, without loss of generality, Define $Z_k = |\frac{X_k - a}{b}|$, $k = 1, \ldots, n$. Then each Z_k has standard exponential distribution, $Exp(0, 1)$, as discussed above. Hence a test of H_0 is equivalent to a test of $H_0 : Z_1, \ldots, Z_n \overset{i.i.d.}{\sim} Exp(0, 1)$. Any of the five graphical tests can then be used to test this H_0 in the following way:

(1) Choose the significance level α and choose the graphical test.
(2) Compute the BLIE of the parameters μ and σ of $Exp(\mu, \sigma)$. Moreover, calculate the critical value which depends on α, sample size n and the number of simulations.
(3) Sort the Z_k's in ascending order $Z_{[1]} \leq \ldots \leq Z_{[n]}$ and plot the points $(-\ln(1 - p_k), Z_{[k]})$ for $k = 1, \ldots, n$ where $p_k = \frac{k-0.5}{n}$. This constructs the Q-Q plot.
(4) For each k, plot a vertical interval based on a graphical test considered.
(5) Join the upper limits of the k vertical intervals and the lower limits of the k vertical intervals step to obtain a band.
(6) Reject the null hypothesis H_0 at level α if at least one point $(-\ln(1-p_k), Z_{[k]})$ falls outside its corresponding vertical interval.

TABLE 3. The ordered differences in flood heights for two stations on the Fox River Wisconsin for 33 years and their corresponding probability intervals at $\alpha = 0.05$

				D		D_m		D_e		D_{be}		D_{bi}	
k	$Y_{[k]}$	$Z_{[k]}$	quantiles	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
1	1.96	0	0.0153	NaN*	0.7737	-0.1053	0.3315	-0.3665	0.3665	-0.1311	0.3861	NaN	0.1981
2	1.96	0.1416	0.0465	NaN	0.9384	-0.1303	0.6385	-0.3912	0.6617	-0.1285	0.6526	NaN	0.5444
3	3.6	0.1545	0.0788	NaN	1.1096	-0.1047	0.9044	-0.3804	0.9301	-0.103	0.9041	NaN	0.8432
4	3.8	0.3863	0.1121	NaN	1.2879	-0.0566	1.1581	-0.3503	1.1884	-0.0556	1.1515	NaN	1.1283
5	4.79	1.2232	0.1466	NaN	1.4739	0.006	1.4081	-0.3068	1.4434	0.0067	1.3981	NaN	1.4098
6	5.66	1.6996	0.1823	NaN	1.6682	0.0798	1.6586	-0.253	1.6987	0.0805	1.6466	NaN	1.6928
7	5.76	1.7382	0.2194	-0.0285	1.8717	0.1629	1.9125	-0.1907	1.9569	0.1638	1.8989	-0.0972	1.9805
8	5.78	2.0858	0.2578	0.1079	2.0852	0.2544	2.1715	-0.1207	2.2197	0.2557	2.1569	-0.0085	2.2756
9	6.27	2.0858	0.2978	0.2488	2.3097	0.3538	2.4375	-0.0436	2.4889	0.3556	2.422	0.0889	2.5802
10	6.3	2.1631	0.3395	0.3944	2.5466	0.4609	2.712	0.0401	2.7658	0.4632	2.6957	0.1947	2.8964
11	6.76	2.7554	0.383	0.545	2.7972	0.5756	2.9963	0.1303	3.0518	0.5786	2.9795	0.3087	3.2265
12	7.65	2.8455	0.4285	0.7011	3.0632	0.6981	3.2922	0.227	3.3486	0.7018	3.2749	0.4312	3.5728
13	7.84	2.9485	0.4761	0.863	3.3466	0.8287	3.6013	0.3302	3.6575	0.8331	3.5837	0.5623	3.9381
14	7.99	3.1931	0.5261	1.0313	3.6499	0.9677	3.9255	0.4401	3.9803	0.973	3.9076	0.7026	4.3254
15	8.51	3.412	0.5787	1.2063	3.9761	1.1157	4.2669	0.5569	4.3191	1.1218	4.2488	0.8525	4.7386
16	9.18	3.7725	0.6343	1.3887	4.3288	1.2733	4.6281	0.6809	4.6759	1.2804	4.6097	1.0129	5.1822
17	10.13	4.0687	0.6931	1.5792	4.7129	1.4414	5.0118	0.8125	5.0533	1.4495	4.9934	1.1846	5.6619
18	10.24	4.3391	0.7557	1.7784	5.1345	1.6209	5.4217	0.9523	5.4545	1.6301	5.4033	1.3688	6.1853
19	10.25	4.9313	0.8224	1.9872	5.6016	1.8131	5.8621	1.1008	5.883	1.8234	5.8438	1.5668	6.762
20	10.43	4.9571	0.8938	2.2066	6.1252	2.0194	6.3381	1.2586	6.3434	2.031	6.3201	1.7803	7.4054
21	11.45	4.97	0.9708	2.4377	6.7211	2.2417	6.8567	1.4266	6.8411	2.2547	6.8389	2.0113	8.1345
22	11.48	5.2146	1.0542	2.6819	7.4123	2.4823	7.4264	1.6057	7.3832	2.4967	7.4091	2.2627	8.9777
23	11.75	5.4979	1.1451	2.9407	8.2354	2.744	8.0587	1.797	7.9788	2.76	8.042	2.5376	9.9805
24	11.81	5.6009	1.2452	3.216	9.2526	3.0306	8.7693	2.0015	8.6397	3.0482	8.7534	2.8405	11.2231
25	12.34	5.6266	1.3564	3.5099	10.5849	3.3469	9.5802	2.2206	9.3822	3.3661	9.5653	3.1771	12.8687
26	12.78	5.7554	1.4816	3.8253	12.5198	3.6993	10.524	2.4555	10.2292	3.7203	10.5102	3.5554	15.3493
27	13.06	6.8755	1.6247	4.1655	16.1101	4.0971	11.6515	2.707	11.2141	4.1197	11.6383	3.9867	20.8684
28	13.29	7.8412	1.7918	4.5348	NaN	4.5534	13.0479	2.9748	12.3887	4.5771	13.0339	4.4876	NaN
29	13.98	8.1502	1.9924	4.9385	NaN	5.0889	14.8731	3.2551	13.8393	5.1122	14.8521	5.0848	NaN
30	14.18	8.4077	2.2437	5.3838	NaN	5.7376	17.4808	3.5342	15.7239	5.7565	17.4257	5.8245	NaN
31	14.4	8.9227	2.5802	5.8803	NaN	6.5644	21.936	3.764	18.379	6.5658	21.6717	6.7987	NaN
32	16.22	10.5193	3.091	6.4411	NaN	7.7204	36.6957	3.7502	22.7202	7.6488	31.8613	8.2392	NaN
33	17.06	10.5193	4.1897	7.0857	NaN	9.7732	22.0392	2.1927	32.9323	9.3171	∞	11.1799	NaN

: NaN stands for Not a Number representing an undefined value.

[8] presented a Q-Q plot of the difference in flood heights data for two stations on the Fox River in Wisconsin for 33 years being 1.96, 1.96, 3.6, 3.8, 4.79, 5.66, 5.76, 5.78, 6.27, 6.3, 6.76, 7.65, 7.84, 7.99, 8.51, 9.18, 10.13, 10.24, 10.25, 10.43, 11.45, 11.48, 11.75, 11.81, 12.34, 12.78, 13.06, 13.29, 13.98, 14.18, 14.4, 16.22, 17.06. He also concludes that the data set follows the Laplace distribution although there exists obvious curvature in the Q-Q plot as shown in Figure 1. [15] mentioned that [16] and [17] doubt whether the Laplace model fits these data. The Q-Q plot and their corresponding intervals for the $Z_{[k]}$'s of D, D_m, D_e, D_{be} and D_{bi} are given in Figure 1. Obviously, there is one point, $Z_{[33]}$, of the plot D_{bi} in Figure 1 lies outside the corresponding interval; therefore, these data do not follow the Laplace distribution based on the D_{bi} test. However, there may be other points which are outside their corresponding intervals, but it is rather subjective to judge that these points fall outside such intervals. Hence, the numerical comparison must be used to

assess whenever points cannot be distinguished by eyes. Table 3 illustrates the values of 33 ordered observations with their corresponding probability intervals. The table is very helpful when any points cannot be diagnosed by eyes; see for example $Z_{[6]}$ falling outside the corresponding intervals based on all tests and $Z_{[8]}$ based on the D test.

For the non-graphical tests, the test statistics AD and CvM tests computed by (3.2) and (3.3) are 1.6153 and 0.3320, respectively. Also, the corresponding critical values of AD and CvM statistics at $\alpha = 0.05$ are 1.2206 and 0.2274, respectively. Hence, the null hypothesis H_0 is rejected by AD and CvM.

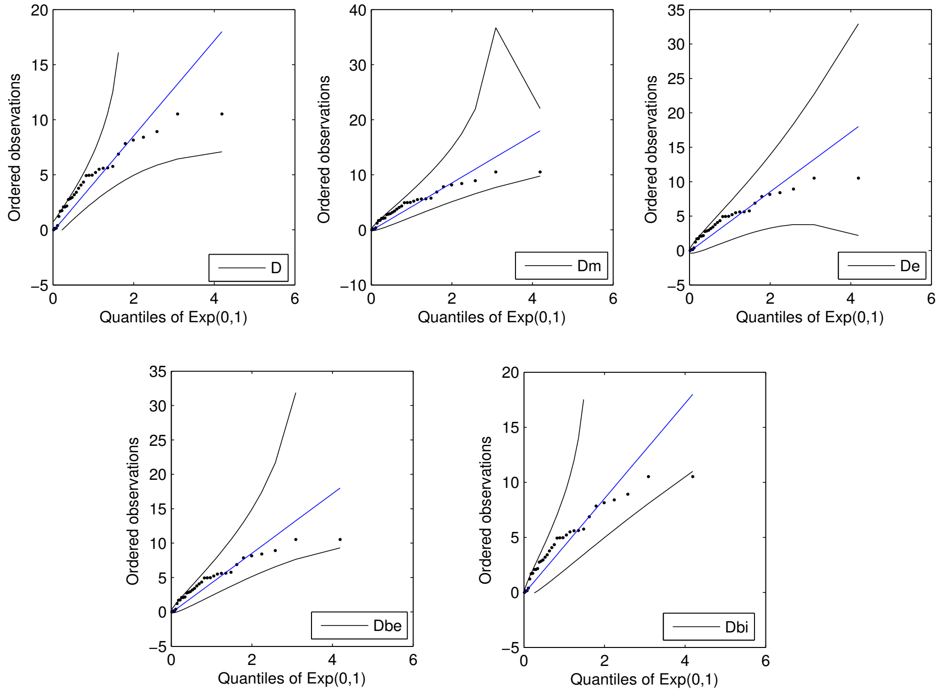

FIGURE 1. The Q-Q plot and simultaneous probability intervals of differences in flood heights for two stations for testing Laplace distribution using transformation to exponential distribution based on the graphical D, D_m, D_e, D_{be}, D_{bi} at $\alpha = 0.05$.

6. Conclusions

Although not completely dominated, we recognize that no test of fit will be optimal for all alternatives. Due to the fact that different people can make different interpretations on a plot, graphical tests and their simultaneous probability intervals can deal with this pitfall. Specifically, we obtain the simultaneous $1 - \alpha$ probability intervals suitable for Q-Q plots on testing the Laplace distributions. They become the useful objective judgement on Q-Q plots for practitioners who want to use the graphical tests.

Acknowledgements

The authors sincerely thank all the referees, associate editor and editor for constructive comments. This research was supported by The Thailand Research Fund (TRF) and Office of the Higher Education Commission (OHEC), Ministry of Education, Thailand, under the grant MRG6180258.

References

[1] W. Weibull, The Phenomenon of Rupture in Solids, Ingeniors Vetenskaps Akademien Handlingar Report, 153, Stockholm, 1939.

[2] G. Blom, Statistical Estimates and Transformed Beta Variables, Wiley, New York, 1958.

[3] J.J. Filliben, The probability plot correlation coefficient test for normality, Technometrics 17 (1975) 111–117.

[4] A. Hazen, Storage to be provided in the impounding reservoirs for municipal water supply, Trans. Am. Soc. Civil Eng. 77 (1914) 1547–1550.

[5] J.R. Michael, The stabilized probability plot, Biometrika 70 (1983) 11–17.

[6] W. Chantarangsi, W. Liu, F. Bretz, S. Kiatsupaibul, A.J. Hayter, W. Fang, Normal probability plots with confidence, Biom. J. 57 (1) (2015) 52–63.

[7] W. Chantarangsi, W. Liu, F. Bretz, S. Kiatsupaibul, A.J. Hayter, W. Fang, Q-Q Plots with confidence for testing Weibull and exponential distributions, Hacettepe J. Math. Stat. 43 (3) (2016) 887–904.

[8] K. Krishnamoorthy, Handbook of Statistical Distributions with Applications, Chapman & Hall CRC, New York, 2006.

[9] N.R. Mann, Optimum estimators for linear functions of location and scale parameters, Ann. Math. Stat. 40 (1969) 2149–2155.

[10] M. Ahsanullah, G.G. Hamedani, Exponential Distribution: Theory and Methods, Nova Science Publishers, Inc., New York, 2010.

[11] H.W. Lilliefors, On the Kolmogorov-Smirnov test for normality with mean and variance unknown, J. Am. Stat. Assoc. 62 (318) (1967) 399–402.

[12] D. Edwards, J.J. Berry, The efficiency of simulation-based multiple comparisons, Biometrics 43 (1987) 913–928.

[13] W. Liu, M. Jamshidian, Y. Zhang, J. Donnelly, Simulation-based simultaneous confidence bands in multiple linear regression with predictor variables constrained in intervals, J. Comput. Graph. Stat. 14 (2005) 459–484.

[14] W.F. Scott, B. Stewart, Tables for the Lilliefors and modified Cramer-von Mises tests of normality, Commun. Stat. Theor. M. 40 (4) (2011) 726–730.

[15] D.J. Best, J.C.W. Rayner, O. Thas, Comparison of some tests of fit for the Laplace distributon, Comput. Statist. Data Anal. 52 (2008) 5338–5343.

[16] P. Puig, M.A. Stephens, Tests of fit for the Laplace distribution, with applications, Technometrics 42 (2000) 417–424.

[17] M.A. Meintanis, A class of omnibus tests for the Laplace distribution based on the empirical characteristic function, Commun. Stat. Theor. M. 70 (2004) 11–17.