American Journal of Applied Mathematics
2016; 4(2): 110-113
http://www.sciencepublishinggroup.com/j/ajam
doi: 10.11648/j.ajam.20160402.17
ISSN: 2330-0043 (Print); ISSN: 2330-006X (Online)

Quantile-Quantile Plot Compared with Stablized Probability Plot in Figure on the Distribution of the Test Research

Hu Yue, Yu Jin-yang

Department of Mathematics and Information Science, Zhejiang University of Science and Technology, Hangzhou, P. R. China

Email address:
huyuexxxx@sina.com (Hu Yue), m13588294489_1@163.com (Yu Jin-yang)

To cite this article:
Hu Yue, Yu Jin-yang. Quantile-Quantile Plot Compared with Stablized Probability Plot in Figure on the Distribution of the Test Research. *American Journal of Applied Mathematics*. Vol. 4, No. 2, 2016, pp. 110-113. doi: 10.11648/j.ajam.20160402.17

Received: March 31, 2016; **Accepted**: April 12, 2016; **Published**: April 19, 2016

Abstract: Introduces the structure and function of PP (Stablized Probability Plot) and QQ (Quantile-Quantile Plot), and uses MATLAB to produce two sets of exponential distribution and normal distribution. The sample number is 20 and 45 respectively. The fitting of QQ plot and PP plot are respectively used to obtain the superiority of PP in the exponential distribution. In normal distribution, the QQ plot is more advantageous.

Keywords: QQ Plot, PP Plot, Goodness-of-Fit Test, Life Distribution

1. Introduction

At present, people used to test the distributions method has many kinds, mainly divided into parameter test and non parameter test [1]. Parameter test including t-test, F-testt, non parameter test including chi square, K-S test and so on. In addition, there are used as a means to test, such as the QQ plot and PP plot. The method of drawing test is more intuitive. Wu Dong Tang Yincai (2004) in "the life distribution of the PP plot"this paper compared the QQ plot and PP plot on the life distribution of three kinds of distribution: normal distribution, logarrithmic and exponential distribution, exponential distribution, concluded PP plot is more stable [7]. Du He (2014) used the same type of relationship to explain the quantile regression [11]; FAN Lijun ang XIONG Zhe using quantile regression on the relationship between weather and rainfall forecast, quantile of the application is very broad, QQ plot that the use of quantile mapping, and its advantages are self-evident. PP plot is a reliability plot, which can be used for reliability annlysis based on PP plot [12]. In this paper, the comparative analysis of the superiority of QQ plot and PP plot in the life distribution is analyzed. Paper comparsion test of superiority of QQ plot and PP plot from the life distribution and the sampledata of different sizes of two aspects.

2. QQ Plot and PP Plot

QQ Plot, it is mainly to show the difference between the observed value and the predicted value. Or to detect whether a set of data from a distribution, or verify whether the two sets of data from the same distribution.

PP plot is based on the cumulative probability of the variable corresponding to the specified theoretical distribution of cumulative probability plot, for the intuitive detection of sample data is consistent with a probability distribution. If the tested data is in accordance with the specified distribution, the points on the representative sample data should be represented on the diagonal of the theoretical distribution.

2.1. QQ Plot

Constructing QQ plot [2-4] does not need to ensure the distribution of the mean, calculated from the standard distribution theory in the specified class. Linear model showed that the class of the specified reasonable representation of data distribution (mean) location and scale parameters (standard deviation). Visual intercept and slope of the linear model can be estimated.

If $x_1, x_2, \cdots, x_n$ are a set of sample observations from a continuous population X of n, and X distribution function is $F(x)$, these observations $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ are sorted by size, $x_{(1)}$ is the smallest measured value, $x_{(n)}$ s the largest measured value. Because X is a continuous random variable, so the probability of $x_{(i)}$ equal probability is zero, we might as well put these sort after the observation value is expressed

as $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$. Because the probability of occurrence of events $\{X \le x_{(i)}\}$ and $\{X < x_{(i)}\}$ respectively $\frac{i}{n}$ and $\frac{i-1}{n}$, and the probability of the two events of the same, so the estimation $F(x_{(i)}) = P(X \le x_{(i)})$ for the middle of the middle of the two frequency value $\frac{i-1}{n}$ is more appropriate [5]. Make the estimated [6]

$$F(x_{(i)}) \approx \frac{i-0.375}{n+0.25} \tag{1}$$

Then use (1) to structure normal distribution of QQ plot. A normal distribution X is established $X \sim N(\mu, \sigma^2)$, which μ, σ are an unknown parameter. Combined (1)

$$\phi(\frac{x_{(i)} - \mu}{\sigma}) \approx q_{(i)} \tag{2}$$

Which $q_{(i)} = \phi^{-1}(\frac{i-0.375}{n+0.25})$, conbined (2)

$$x_{(i)} \approx \mu + \sigma q_{(i)} \tag{3}$$

Use ($q_{(i)}, x_{(i)}$)(i=1,2,$\cdots$,n) a scatter plot on the coordinate plane, which is called the normal distribution of QQ. The masimum likelihood estimation method is used to estimate the unknow μ, σ .

2.2. PP Plot

If random variables Y to obey the position-scale parameters, the distribution function is $F(\frac{y-\mu}{\sigma})$, and changed $U = F(\frac{y-\mu}{\sigma})$, the transformation was:

$$Y = \mu + \sigma F^{-1}(U) \tag{4}$$

Transform (4):

$$S = \frac{2}{\pi}\arcsin(U^{1/2}) \tag{5}$$

And its density function is

$$p(s) = \frac{1}{\pi}\sin(\pi s), 0 \le s \le 1 \tag{6}$$

If set: $s_i = \frac{2}{\pi}\arcsin(F^{1/2}(\frac{y_i - \mu}{\sigma})), i = 1,2,\cdots,n,$ (7)

$$r_i = \frac{2}{\pi}\arcsin((\frac{i-\frac{1}{2}}{n})^{\frac{1}{2}}), i = 1,2,\cdots,n \tag{8}$$

The scatter plot (r_i, s_i) which is composed of points is called the PP plot [7].

3. Life Distribution of Commonly Is Uesd

From plants to aniamls, each one has its own life, and is not a unique biological properties, and those who have no life, which is often said that the life of the product. To better monitor the merits of the product, for different products have different life distribution. The life distribution of a common exponential distribution, Weibull distribution [8], normal distribution, lognormal distribution, extreme value distribution and so on.

In this paper, the comparison between QQ and PP is simulated and compared with the exponential distribution and normal distribution [9] [10].

A sample with a capacity of 20 is generated from the overall compliance of EXP(0.4).

20 Samples of exponential distribution

1.6880	1.1097	1.3923	0.7770
13.3291	1.1875	13.7890	1.4564
0.6542	3.7437	5.1361	4.5942
0.2731	1.6896	12.3016	0.2042
1.5501	7.0600	9.3267	13.4734

According to the definition, make the corresponding QQ plot and PP plot:

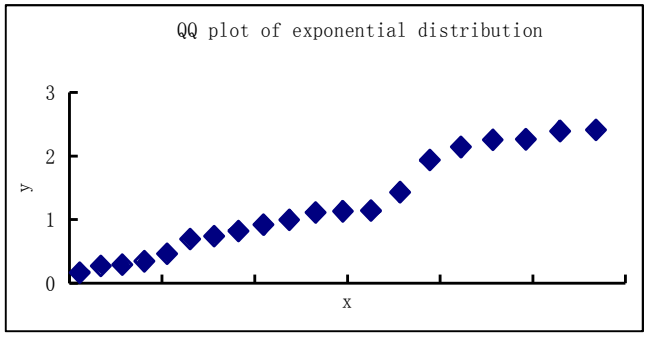

Fig. 1. *Exponential distribution of QQ plot.*

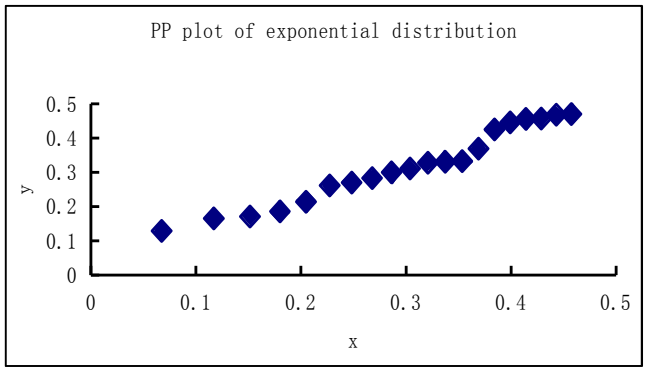

Fig. 2. *Exponential distribution of PP plot.*

A sample with a capacity of 45 is generated from the total of EXP (0,4) which is subject to exponential distribution.

45 Samples of exponential distribution

1.1440	0.1662	0.6940	4.1995	4.1804
5.0093	10.3158	8.1638	7.2761	5.4645
2.4117	5.4772	6.5064	0.7413	1.1125
11.6788	2.2567	0.9213	2.7252	7.9039
0.8218	5.5279	2.1428	1.1309	2.5335
3.0253	4.6691	1.4320	7.6074	5.6501
1.9377	2.3934	3.8656	0.9981	17.7242
2.5506	0.4621	5.4268	0.2935	2.9922
0.3457	2.2638	0.2731	4.3492	7.1915

Make the corresponding graph:

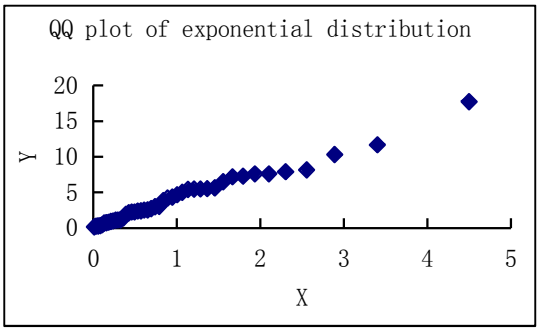

Fig. 3. *Exponential distribution of QQ plot.*

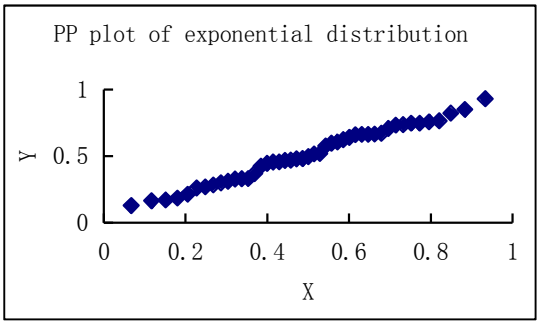

Fig. 4. *Exponential distribution of PP plot.*

Samples from the total sample size of N(0,1) obeying normal distribution is 20.

20 samples off normal distribution

-0.4397	-0.0273	-1.2272	-1.3770	0.6352
-1.4238	-1.2276	0.7009	-0.6681	0.8512
0.3449	-1.9276	0.7326	0.8782	1.2362
1.6791	-0.4843	-0.1107	-0.7836	-0.9889

Make the corresponding graph:

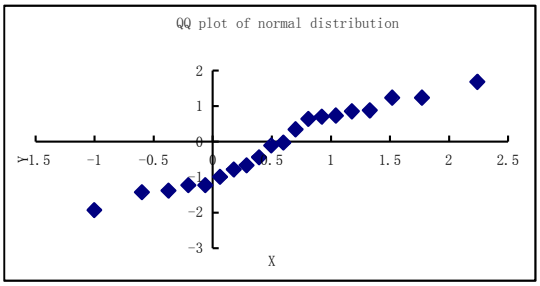

Fig. 5. *Normal distribution of QQ plo.*

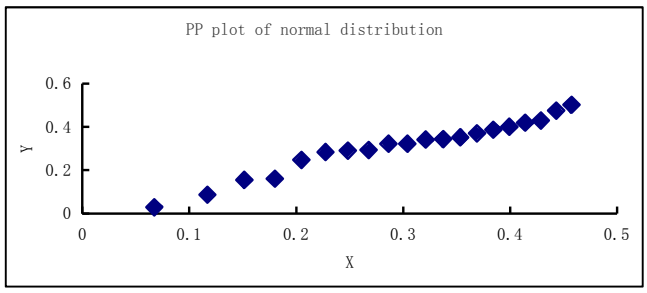

Fig. 6. *Normal distribution of PP plot.*

Samples from the total sample size of 45 to N(0,1).

45 samples off normal distribution

0.2062		0.7761	1.5478	-0.5968	0.8921
2.6757	-1.5362	0.0338	1.8783	-1.0657	
-2.8704	-1.5697	-2.0880	-0.8934	-0.6400	
1.2217	-0.0979	1.5562	-0.2803	0.6623	
-0.3197	-0.7215	0.3962	1.5764	-0.7227	
0.1884		0.7524	-0.5206	0.7740	-0.6310
1.0691	-0.8505	0.9716	0.4542	-0.3946	
0.1697		0.0071	-0.4512	0.8692	0.8758
0.2143	-0.8637	0.7387	0.4341	0.3109	

Make the corresponding graph:

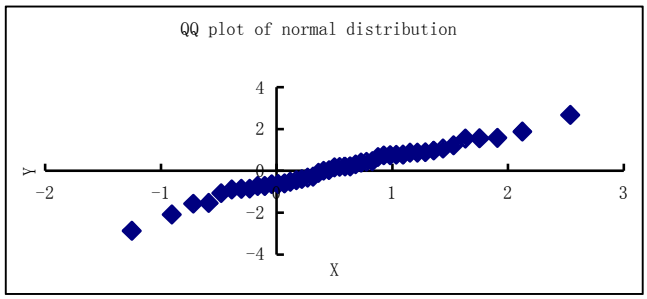

Fig. 7. *Normal distribution of QQ plot.*

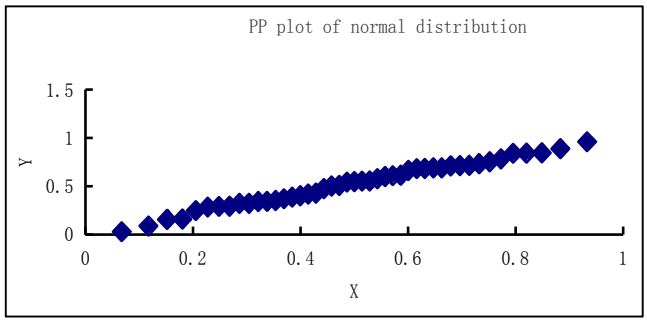

Fig. 8. *Normal distribution of PP plot.*

From above a few views can be obtained, when the sample data is small, the sample data point distribution is uniform; when more data and index distribution of the QQ plots of the data points at both ends of the more intensive, and PP plot distribution is uniform, so in the index distribution of the test PP plot has advantages. The test of normal distribution, when data quantity is large, the QQ plot of the data points concentrated in the middle part, which is in line with the characteristics of normal distribution; and PP plot data points

are uniformly distributed, cannot directly through the observations made by the graph to determine the test of normal distribution or exponential distribution. Therefore, in the normality test, the QQ chart is more superior, but in the test of the exponential distribution, the effect of PP plot is better.

4. Conclusion

QQ and the PP plot in the analysis of lifetime has a prominent role, it can be very good to test the lifetime distribution is consistent with a distribution, compared to in the traditional method for data analysis, the use of PP and QQ diagram more concise. In different life distribution, QQ diagram and PP diagram have different characteristics, in the exponential distribution of PP figure has better explanation, and in the normal distribution of QQ figure is more clear.

The conclusion of this paper and the conclusion of Wu Dong Tang Yincai is different, this paper cansider the test between different life distribution, and cansiders whether the data sample size will affect the test of life distribution. It is this a little improvenment, But there are many defects in this paper, the life distribution of less test, the other distribution test, which method should be used for inspection, further research is needed.

Acknowledgements

This work was financially supported by the science and Technology Department of Zhejiang Province of China (2015C33088), Educational Commission of Zhejiang Province of China (Y201327793) and Foundation of Education Research Zhejiang University of Science and Technology (2011039).

References

[1] Guo Lihong, The goodness of fit test of multivariate distribution and its application [D], Master's degree thesis, North China Electric Power University2014.

[2] Liang J, Pan Wolliam S Y, Yang Z H. Characterization-baesd Q-Q Plots for Testing Multinormality [J]. Statistics and Probablity Letters. 2004, 70; 183-1190.

[3] Yan Su. Smooth tset for elliptical symmetry [C]. Proceedings of the 2011 International Conference on Machine Learing and Cybernetics. 2012.

[4] Yang Zhenhai, Cheng Weihu, Zhang Junjian, Goodness of fit test [M]. Science Press, 2011: 95-122.

[5] Wang Xueming, <Applied probability statistics> [M], Beijing: High Education Press, 2005.

[6] Zhang J. Powerful goodness of fit tests baesd on the likelihood ratio [J]. J Roy Soc, 2002, 64: 281-294.

[7] Wu Dong, TANG Yin-cai, PP Plot of Lifetime Distributions [J], Data statistics and management 2004, 23(5), 33-40 .

[8] Wang Shan, A comparative study on the life distribution [D], Master's degree thesis, Zhejiang Gongshang University, 2012.

[9] Liu Lianhua, Luo Wenqiang, Comparison of goodness of fit [J], Journal of Yangtze University: Natural Science Edition, 2013, 10(4), 14-16.

[10] Surucu B. A power comparison and simulation study of goodness of fit tests [J]. Computersand Mathematics with Application, 2008, 56: 1617-1625

[11] Du He, Quantile regression [D], Master's degree thesis, Huazhong Normal University, 2014.

[12] FAN Lijun and XIONG Zhe, Using Quantile Regression to Detect Relationships between Large-scale Predictors and Local Precipitation over Northern China [J], Advances in atmospheric science, 2015, 32, 541-552.