

Goodness-of-fit-test for Exponential Power Distribution

A. A. Olosunde^{1,*}, A. M. Adegoke^{2,*}

¹Department of Mathematics, Obafemi Awolowo University, Ile-Ife, Nigeria

²Department of Physics, Obafemi Awolowo University, Ile-Ife, Nigeria

*Corresponding author: akinolosunde@gmail.com, adegoke00@gmail.com

Abstract Given a set of data, one of the statistical issues is to see how well the data fit into postulated model. This technique necessitates the corresponding table of the probability distribution for the proposed model. In this paper, we examined, to what limit of p can normal approximate this sample without falling into type I error (i.e. a random variable x having normal distribution when indeed it has exponential power distribution with estimated parameter p). We also present the goodness-of-fit test for exponential power distribution using the conventional testing methods which are discussed, one is Pearson's χ^2 test and the other one is kolmogorov-Smirnov test. An example in poultry feeds data and a simulation example are included, comparison with the fitting of the normal distribution is also examined for further illustration.

Keywords: shape parameter, short tails, cumulative distribution function, kolmogorov-Smirnov test, pearson's χ^2 test, poultry feeds data data

Cite This Article: A. A. Olosunde, and A. M. Adegoke, "Goodness-of-fit-test for Exponential Power Distribution." *American Journal of Applied Mathematics and Statistics*, vol. 4, no. 1 (2015): 1-8. doi: 10.12691/ajams-4-1-1.

1. Introduction

For $p > 0$, consider the random variable Z with density function

$$f(x) = \frac{1}{2\sigma_p p^{1/p} \Gamma\left(1 + \frac{1}{p}\right)} \exp\left\{-\frac{|x-\mu|^2}{p\sigma_p^p}\right\} \quad (1)$$

where $-\infty < x < \infty, -\infty < \mu < \infty, p > 0$ and $\sigma_p > 0$. (1) is called exponential power distribution with shape parameter p which regulates the tail region. Several properties of this distribution have been studied by many authors both as univariate and its multivariate extension. Among them are Olosunde [7], Agro [1,2], Mineo and Ruggieri [6] and others. Many authors have also found this distribution useful as substitute to normal distribution in applications; Lindsey [4] have applied it in repeated measurements, also Olosunde [7] in fitting poultry feeds data, just to mention few. The defined random variable family exponential power distribution retains many statistical properties of the normal and Laplace distribution, that is, what differentiate exponential power ($EP(p)$) from the normal and Laplace is the shape factor, which makes the tail becomes thicker as $p \rightarrow \infty$. The distribution can also be regarded as generalized normal(or Laplace) because at $p = 2$ (or 1), we have the normal(or Laplace) distribution with parameters μ and σ . The distribution (1) because of its shape parameter performed better when compare to its normal distribution subclass

especially in an experiment where small sample size is feasible.

The moment can be obtained from the maximization of the log-likelihood function [6]

$$\begin{aligned} \ell(x; \mu, \sigma_p, p) &= \log L(x; \mu, \sigma_p, p) \\ &= -n \log \left[2p^{1/p} \sigma_p \Gamma\left(1 + \frac{1}{p}\right) \right] \\ &\quad - \frac{\sum_{i=1}^n |x_i - \mu|^p}{p\sigma_p^p} \end{aligned} \quad (2)$$

the derivative of (1.2) with respect to μ, σ and p and equating to zero gives the following equations:

$$\frac{d\ell(x)}{d\mu} = -\frac{1}{\sigma_p^p} \sum_{i=1}^n |x_i - \mu|^{p-1} \text{sign}(x_i - \mu) = 0 \quad (3)$$

$$\frac{d\ell(x)}{d\sigma p} = -\frac{n}{\sigma p} + \frac{1}{\sigma_p^{p+1}} \sum |x_i - \mu|^p = 0 \quad (4)$$

$$\begin{aligned} \frac{d\ell(x)}{dp} &= -\frac{n}{p^2} [\ln p + \psi(1 + 1/p) - 1] \\ &\quad + \frac{\sum_{i=1}^n |x_i - \mu|^p}{p\sigma_p^p} \left[\frac{1}{p} + \ln \sigma p - \ln |x_i - \mu| \right] = 0 \end{aligned} \quad (5)$$

where ψ is the *di-gamma* function. The equation 1.3 and 1.5 can only be resolved using numerical approach, while the explicit solution to 1.4 is:

$$\hat{\sigma}_p = \left(\frac{\sum_{i=1}^n |x_i - \mu|^p}{n} \right)^{1/p}. \quad (6)$$

Therefore, the location, scale and shape parameters can be estimated from the sample by maximizing the log-likelihood function, using numerical approach because, the maximizing expressions for location and shape parameters are not in close form. Although statistical properties about exponential power and its generalization have been discussed extensively in the literature, but given a data set X_1, \dots, X_n , there is a problem about how well the underlying distribution can be represented by an exponential power distribution. The Pearson χ^2 and Kolmogorov-Smirnov's test and so many others are important tools commonly used in statistical practice. However, to carry this out these test effectively, we require the associated probability distribution table. Just like the normal distribution, the cumulative distribution table of exponential power random variable, to the best of our knowledge are not available in any literature. This deficiency may be one of the hindrances to wide use of exponential power distribution when compare to normal distribution as exponential power distribution generalized the normal distribution and its cdf is not in close form. In what follows, we establish distribution table for the random variable having (1) with different values of p the shape parameter. In section 2 we describe how the table was developed. Section 3 discuss the two procedures for the goodness-of-fit test (Pearson and Kolmogorov-Smirnov). Section 4 presents a simulation data set of exponential power distribution, this is also included to illustrate the use of the table. Importantly, the normal distribution has been well known to be the limiting distribution for many density functions. In this paper, we examined, to what limit of p can normal approximate this sample without falling into type I error (i.e. a random variable x having normal distribution when indeed it has exponential power distribution with estimated parameter p) and finally an example of potential application to poultry feeds data from Olosunde, [7] is also examined.

2. Exponential Power Distribution Table

The cumulative distribution function (cdf) for a standardized random variable X having (1) with real p can be expressed as

$$P(X \leq x) = \int_{-\infty}^x \frac{1}{2p^{1/p}\Gamma\left(1 + \frac{1}{p}\right)} \exp\left\{-\frac{|x|^p}{p}\right\} dx \quad (7)$$

Thus, for each specified p , we can calculate the corresponding probability for each value of t . In the table we present the corresponding probabilities for t ranging from 0.00 until $P(X \leq x) \approx 1$ to 3 decimal places, with each increase in length by 0.01. We employed Simpson rule in Numerical Computation in conjunction with R program developed by Ihaka and Gentleman [3]. We prefer Simpson's method compare to other methods because its guarantees the accuracy level of the table. The

table is arranged as follows, if we wish to compute, say $x = 0.15$, the table in the appendix can be used in the this way:

$$P(Y \leq 0.15) = 0.5586, \text{ when } p=2.6$$

$$P(Y \leq 0.15) = 0.5585, \text{ when } p=3.4$$

from the table we can see that the probability distribution of exponential power distribution depends on the shape parameter, p , and as p increases the cdf changed. For example the, $P(Y \leq 3.0) = 0.9998$ remain the same at the accuracy of 10^{-4} for p ranging from 2.60 to 5.60, values that normal gave a good approximations are left out. Therefore, the tables were truncated at some points, when the resulting values of $P(X \leq x)$ repeat the previous values for increase in shape parameter p . To check the accuracy of the table in the appendix, from our program we allowed $p = 1$ which of course gave the values for the cdf of Laplace distribution otherwise known as double exponential (not reproduce here). Also, when $p = 2$ we have the values for the cdf of a random variable having a standard normal probability distribution function (not reproduce here, but available in many Statistical texts), we carefully select our p for some values for illustration of its usefulness and applications purposes in order to save pages. The algorithm on R to further developed extensive table can be made available upon request.

3. Goodness-of-Fit Tests for The Exponential Power Distribution

In this section, we present the two commonly used procedures for goodness-of-fit test but now with exponential power distribution as the underlining distribution of interest. One is *Pearson's χ^2 test* and the other one is *Kolmogorov - Smirnov test*. These are two well known tests in the literature to examine how well a set of data fits into a postulated model provided that the probability distribution of the postulated random variable is available.

3.1. χ^2 Procedure for Exponential Power Distribution

Given a set of data X_1, \dots, X_n , carrying out Pearson's χ^2 test to ascertain if the data is well fit into exponential power distribution $EP(p_0)$, the procedures are well known in most statistical text.

The χ^2 test statistic with degree of freedoms $K - 1$ is then defined as

$$\chi^2 = \sum_{i=1}^K \frac{(N_i - E_i)^2}{E_i} \quad (8)$$

where N_i is the number of outcomes that fall in the i th interval and E_i is the expected number in the i th interval. The selection of K follows the general rule in the application of Pearson's χ^2 test.

To illustrate this a simulation of 1000 samples from exponential power with $p = 4.4$ and comparatio was done with the normal distribution.

Example 2: (Simulation from exponential power distribution)

Table 4.1. Pearson's χ^2 Test

| Intervals | n | p_i | EP(4.4) | normal |
|--------------------|-----|--------|---------|--------|
| $(-\infty, -1.75]$ | 3 | 0.0033 | 3.3 | 40.1 |
| $(-1.75, -1.25]$ | 55 | 0.0540 | 54.0 | 65.5 |
| $(-1.25, -0.75]$ | 149 | 0.1524 | 152.4 | 121.0 |
| $(-0.75, -0.25]$ | 193 | 0.1924 | 192.4 | 174.7 |
| $(-0.25, 0.25]$ | 197 | 0.1958 | 195.8 | 197.4 |
| $(0.25, 0.75]$ | 196 | 0.1924 | 192.4 | 174.7 |
| $(0.75, 1.25]$ | 151 | 0.1524 | 152.4 | 121.0 |
| $(1.25, 1.75]$ | 52 | 0.0540 | 54.0 | 65.5 |
| $(1.75, \infty]$ | 4 | 0.0033 | 3.3 | 40.1 |

The Table 4.1 shows a simulation of 1000 samples from exponential power distribution with $p = 4.4$, where n is the observed frequency in the i th interval. Np_i and normal are the expected frequency in the i th interval for EPD(4.4) and normal distribution respectively. We obtained χ^2 value of 0.4207 for EP(4.40) with degree of freedom 9, thus EP(4.40) is accepted as expected. However, the goodness-of-fit for $N(0,1)$ gives an observed χ^2 value of 89.72, which results in the rejection of $N(0,1)$ model for the same data set. See Table 4.2 for detail report.

3.2. Kolmogorov Test Procedure on the Exponential Power Distribution

If we have a random sample X_1, \dots, X_n from a population with distribution function $F(x)$, we desire to see if a postulated exponential power distribution (with specified p_0) can be used to fit the underlying population of the data. The null hypothesis can be stated as follows

$$H_0 : F(x) = G_0(x) \text{ for all } x$$

against the alternative

$$H_1 : F(x) \neq G_0(x) \text{ for at least one } x.$$

where $G_0(x)$ denotes the cdf of $EP(p_0)$

$$D(F_n(x), G_p(x)) = \sup_z |F_n(x) - G_p(x)| \tag{9}$$

where

$$F_n(x) = \begin{cases} 0, & x < X_{(1)} \\ \frac{i}{n}, & X_{(i)} \leq x \leq X_{(i+1)}, k = 1, \dots, n-1; \\ 1, & x \geq X_{(n)} \end{cases}$$

where $X_{(1)}, \dots, X_{(n)}$ in the expression of $F_n(x)$ are the ordered statistics of X_1, \dots, X_n . $G_p(x)$ at each sample points of X_i can be found from the exponential power distribution table. In this case, the Kolmogorov-Smirnov test statistic $D(.,.)$ is the maximum distance between empirical distribution function and postulated distribution function at the sample points. At significant level of α , the test endpoint d_α for test statistic D can be found from Miller [5]. The rule is that if the calculated D is larger than d_α the postulated exponential power distribution function is too far away from the observed distribution function. Thus H_0 is rejected at α level of significance, otherwise, H_0 is accepted at the same significance level. To carry out this test, it is critical to find the $F_n(x)$'s for the postulated exponential power distribution. The table provide in this paper makes it possible for the implementation of this test.

4. Applications

Example 1: (Approximation of the Exponential Power Distribution by The Normal Distribution)

Normal distribution has been well known to be the limiting distribution for so many distribution in the literature. In this section with explore to what value of the parameter p will normal give an acceptable approximation to data having exponential power distribution with parameter p_i . This will also examine the closeness between exponential power and normal distributions, using the Kolmogorov-Smirnov test of normality distance. Let $X \sim N(0,1)$ and $F(x)$ be the cdf, also let $Y \sim EP(p)$ and $G_p(y)$ be the cdf. The Kolmogorov distance between $F(x)$ and $G_p(y)$ is defined as

$$D(F, G_p) = \sup_z |F - G_p| \tag{10}$$

The values of $D(F, G_p)$ can be obtained from the Tables in the appendix. The values of $D(F, G_p)$ from some selected $p = 1.6 - 4.4$. These are shown in the table below

Table 4.2. Kolmogorov distance between F and G_p

| p | 1.6 | 2.2 | 2.6 | 2.8 | 3.0 | 3.4 | 3.8 | 4.0 | 4.2 | 4.4 |
|-------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| $D(F, G_p)$ | 0.0146 | 0.0065 | 0.0197 | 0.0226 | 0.0268 | 0.0348 | 0.0415 | 0.0447 | 0.0478 | 0.0504 |

we observed from Table 4.2, that as p increases $D(F, G_p)$ also increases, this implies that approximation by normal distribution becomes poorer with large estimated p from experimental samples. Large $D(F, G_p)$

is noticeable in all p 's when $t = 1.6$. Therefore, normal assumption in such case of large p value may lead to error in conclusion. It should be noted that the significance of $D(F, G_p)$ also depends on the sample size.

Example 3: (Applications to Poultry feeds data) The data was obtained from Olosunde [7], where parameters have been estimated using maximum likelihood approach numerically. The cholesterol level x_i of 48 eggs of chicken fed with organic copper-salt are measured in mg/egg , where 5.20 is the estimated p value for exponential power distribution and 131.457 and 37.232 are the population mean and standard deviation respectively. Also for Normal we have 59.10 and 1.822 as the estimated mean and standard deviation respectively. The ordered data set x_i are given in Table 4.3, z_i and t_i is the standardized values for x_i for $EP(5.20)$ and normal respectively.

$$Z_i = P(EP(5.20 \leq z_i)) \text{ and } T_i = (P(N(0,1)) \leq t_i)$$

normal counterpart. We define D_{EP} as the $\max(|Z_i - i/n|, |Z_i - (i-1)/n|)$ for $EP(5.20)$ and D_N as $\max(|T_i - i/n|, |T_i - (i-1)/n|)$ for normal distribution. From Table 4.3, using Kolmogorov-Smirnov test, we find the corresponding $|D| = 0.061833$ for $EP(5.20)$ and $|D| = 0.0779$ for normal distribution. One can easily see that the fit of exponential power cdf is uniformly better than that of the standard normal cdf in this example. All these have been made possible using the Table in the appendix. Details are provided in Table 4.3.

Table 4.3. Kolomogorov Goodness-of-Fit Test

| x_i | z_i | t_i | Z_i | $ D_{EP} $ | T_i | $ D_N $ |
|--------|--------------|--------------|--------|------------|---------|----------|
| 60.73 | -1.489196365 | -1.899629351 | 0.0115 | 0.0115 | 0.5294 | 0.5294 |
| 66.03 | -1.374792238 | -1.757278685 | 0.0254 | 0.01627 | 0.5392 | 0.518367 |
| 71.33 | -1.260388111 | -1.614928019 | 0.0452 | 0.0173 | 0.5537 | 0.512033 |
| 76.63 | -1.145983983 | -1.472577353 | 0.0713 | 0.01203 | 0.5708 | 0.5083 |
| 81.86 | -1.033090854 | -1.33210679 | 0.1065 | 0.023167 | 0.5918 | 0.508467 |
| 81.93 | -1.031579856 | -1.330226687 | 0.1065 | 0.0185 | 0.5918 | 0.487633 |
| 81.93 | -1.031579856 | -1.330226687 | 0.1065 | 0.03933 | 0.5918 | 0.4668 |
| 87.16 | -0.918686727 | -1.189756124 | 0.1429 | 0.02377 | 0.617 | 0.471167 |
| 92.46 | -0.8042826 | -1.047405458 | 0.1829 | 0.016233 | 0.6469 | 0.480233 |
| 92.52 | -0.802987459 | -1.045793941 | 0.1864 | 0.02193 | 0.96492 | 0.77742 |
| 97.76 | -0.689878473 | -0.905054792 | 0.2284 | 0.020067 | 0.6814 | 0.473067 |
| 97.82 | -0.688583332 | -0.903443275 | 0.2284 | 0.0216 | 0.6841 | 0.454933 |
| 103.06 | -0.575474345 | -0.762704125 | 0.2707 | 0.0207 | 0.7236 | 0.4736 |
| 103.11 | -0.574395061 | -0.761361195 | 0.2747 | 0.01697 | 0.7236 | 0.452767 |
| 108.36 | -0.461070218 | -0.620353459 | 0.3182 | 0.026533 | 0.7676 | 0.475933 |
| 108.41 | -0.459990934 | -0.619010529 | 0.3182 | 0.01513 | 0.7709 | 0.4584 |
| 113.66 | -0.346666091 | -0.478002793 | 0.3613 | 0.027967 | 0.8156 | 0.482267 |
| 113.7 | -0.345802664 | -0.476928449 | 0.3613 | 0.0137 | 0.8156 | 0.461433 |
| 118.96 | -0.232261964 | -0.335652127 | 0.4088 | 0.0338 | 0.8669 | 0.4919 |
| 119 | -0.231398536 | -0.334577783 | 0.4088 | 0.012967 | 0.8707 | 0.474867 |
| 124.26 | -0.117857837 | -0.193301461 | 0.4563 | 0.039633 | 0.9247 | 0.508033 |
| 124.3 | -0.116994409 | -0.192227116 | 0.4524 | 0.0149 | 0.9247 | 0.4872 |
| 129.56 | -0.003453709 | -0.050950795 | 0.4998 | 0.041467 | 0.9801 | 0.521767 |
| 129.6 | -0.002590282 | -0.04987645 | 0.4996 | 0.020433 | 0.9801 | 0.500933 |
| 134.86 | 0.110950418 | 0.091399871 | 0.5437 | 0.0437 | 0.0359 | 0.48493 |
| 134.89 | 0.111597988 | 0.09220563 | 0.5437 | 0.022867 | 0.0359 | 0.50577 |
| 140.16 | 0.225354545 | 0.233750537 | 0.5912 | 0.049533 | 0.091 | 0.4715 |
| 140.19 | 0.226002115 | 0.234556296 | 0.5912 | 0.0287 | 0.091 | 0.49233 |
| 145.46 | 0.339758672 | 0.376101203 | 0.6347 | 0.051367 | 0.148 | 0.45617 |
| 145.48 | 0.340190386 | 0.376638376 | 0.6347 | 0.030533 | 0.148 | 0.477 |
| 150.76 | 0.454162799 | 0.518451869 | 0.6778 | 0.0528 | 0.1985 | 0.44733 |
| 150.78 | 0.454594513 | 0.518989042 | 0.6778 | 0.031967 | 0.1985 | 0.46817 |
| 156.06 | 0.568566926 | 0.660802535 | 0.7253 | 0.058633 | 0.2454 | 0.4421 |
| 156.08 | 0.56899864 | 0.661339708 | 0.7253 | 0.0378 | 0.2454 | 0.46293 |
| 161.36 | 0.682971054 | 0.803153202 | 0.7681 | 0.059767 | 0.2881 | 0.44107 |
| 161.37 | 0.68318691 | 0.803421788 | 0.7681 | 0.038933 | 0.2881 | 0.4619 |
| 166.66 | 0.797375181 | 0.945503868 | 0.8136 | 0.0636 | 0.3289 | 0.44193 |
| 166.67 | 0.797591038 | 0.945772454 | 0.8136 | 0.042767 | 0.3289 | 0.46277 |
| 171.96 | 0.911779308 | 1.087854534 | 0.8535 | 0.061833 | 0.3621 | 0.4504 |
| 171.97 | 0.911995165 | 1.08812312 | 0.8535 | 0.041 | 0.3621 | 0.47123 |
| 177.26 | 1.026183435 | 1.2302052 | 0.8935 | 0.060167 | 0.3907 | 0.46347 |
| 177.26 | 1.026183435 | 1.2302052 | 0.8935 | 0.039333 | 0.3907 | 0.4843 |
| 182.56 | 1.140587562 | 1.372555866 | 0.9259 | 0.0509 | 0.4147 | 0.48113 |
| 182.56 | 1.140587562 | 1.372555866 | 0.9259 | 0.030067 | 0.4147 | 0.50197 |
| 182.56 | 1.140587562 | 1.372555866 | 0.9259 | 0.0116 | 0.4147 | 0.5228 |
| 187.86 | 1.25499169 | 1.514906532 | 0.9528 | 0.0153 | 0.4345 | 0.52383 |
| 187.86 | 1.25499169 | 1.514906532 | 0.9528 | 0.02637 | 0.4345 | 0.54467 |
| 193.16 | 1.369395817 | 1.657257198 | 0.9746 | 0.0254 | 0.4515 | 0.5485 |

References

[1] Agro, G. (1992), Maximum Likelihood and L_p -norm Estimators. Statistica Applicata. 4(2), 171-182.

[2] Agro, G. (1995), Maximum Likelihood Estimation for the Exponential Power Distribution. Communications in Statistics (Simulation and Computation), 24(2), 523-536.

[3] Ihaka, R and Gentleman, R. (1996), R. A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics. Vol. 4(3), pp. 299-314.

