IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

Lindley Regression with Application

Adnan Mostafa Al-Sinjary[1] and Hayder Sabah Tuama[2]

[1,3] Statistics & Information Techniques, Technical college of Management, Northern Technical University, Iraq
[2] Department of Economics, College of Administration and Economics, University of Misan, Iraq

ARTICLE INFO

Article history:
Received 07 October 2025
Revised 07 October 2025,
Accepted 13 January 2026,
Available online 16 January 2026

Keywords:
Lindley regression model
Maximum Likelihood Method
Bayesian Information Criterion
Cumulative distribution function
Quantile function

ABSTRACT

In this work, Lindley regression model is studied standing up Lindley distribution which is one of the generalized linear models. It is like with distributions for example Weibull, Exponential, Gamma and other distributions which are positive skew. This is Lindly Distribution. It aims to dealing with variables have positively skewed problems. The model has been composed and then reaching to a required link function .Parameters are estimated by Maximum Likelihood Method by using Newton-Raphson because it is difficult to obtain closed parameters. In application, real data is used to compare a model depended on exponential distribution and other depended on Lindley`s distribution by using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) after showing by Kolmogorrov-Smirnov Test that the dependent variable follows both distributions The outcomes show that Lindley`s distribution provides accurate results comparing with results of the exponential distribution.

1. Introduction

Lindley`s distribution is a continuous mixed that has an essential importance in the statistical analysis especially with positively skewed variables. . It is an expanded case from exponential distributions and has properties of exponential and Gamma distributions. Lindley`s distribution is flexible and has ability to represent data accurately in applied fields such as medical and social studies. A progression model is built by using Lindley`s distribution in the context of generalised linear models which leads to finding a link function and estimating its parameters depending on Maximum Likelihood Method. Also, the model`s efficiency is tested in comparing with other models depend on the exponential distribution by using different link functions like linear, logarithm and inverse functions. The applied side includes real data showing influence many explanatory variables on period of staying patients where the data is analyzed by using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to evaluate these models. The results appear efficiency of Lindley`s distribution in representing and explaining data when it compared with the other models. This supports that Lindley`s distribution is a flexible and accurate statistic when it deals with positively skewed data.

2. Methodology (Lindly`s Distribution)

It is called and suggested according to the researcher Dennis Victor Lindley in 1958. This distribution comes from two random variables one of them follows the exponential

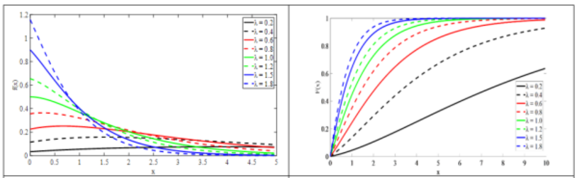
Corresponding author E-mail address: mudher68irri@yahoo.com
https://doi.org/10.62933/7k7j2657

distribution with λ , while the other follows gamma distribution with (2,λ). The probability density function and cumulative distribution function are defined as below : (Ghitany et al., 2008)

$$f(y,\lambda)=\frac{\lambda^2}{\lambda+1}(1+y)e^{-\lambda y} \quad (1)$$

$$F(y,\lambda)=1-[\frac{\lambda+1+\lambda y}{\lambda+1}]e^{-\lambda y}\ ; y>0, \lambda>0 \quad (2)$$

The graphs of these functions for different values of λ are shown below :



a. Probability Density Function b. Cumulative Distribution Function

Figure 1. Probability Density Function and Cumulative Distribution Function for Lindley Distribution

Figure 1. Probability Density Function and Cumulative Distribution Function for Lindley Distribution

The mean and variance of Lindley`s distribution are written as follows :

$$\mu=\frac{(\lambda+2)}{(\lambda+1)\lambda}, \qquad \sigma^2=\frac{\lambda^2+4\lambda+2}{(\lambda+1)^2\lambda^2}$$

In 2010, Adrian Jodra has found out the quantile function for this distribution as seen :

$$y=-1-\frac{1}{\lambda}-\frac{1}{\lambda}W_{-1}[\frac{(\lambda+1)}{e^{(\lambda+1)}}(\mu-1), \quad 0<\mu<1 \quad (3)$$

where W_{-1} denotes the negative branch of the Lambert W function.

The generalized linear models

The generalized linear model is an extension for the linear regression, and it includes all probability distributions belonging to the exponential family. This means each probability distribution can be written as below: (Hardin & Hilbe, 2018)

$$f(y;\theta,\varphi)=e^{[\frac{y\theta-b(\theta)}{a(\varphi)}+c(y,\varphi)]} \quad (4)$$

where y is the random variable which a member of exponential family, θ is the location parameter or the canonical parameter and φ is the dispersion parameter.

The generalized linear model have the following components (Dobson & Barnett, 2018)

- Random component represented by a variable has probability distribution belongs to exponential family.
- Linear systematic component which includes views of explanatory variables X and parameter vector β with dimension k×1, where k is number of explanatory variables. The systematic part is $\eta_i=x_i^T\beta$ called linear predictor.
- Differentiable link function g(,) that relate random part to systematic part where the inverse function for this function is equal to expectation conditional by β and x_i^T as follows:

$$\mu_i=g^{-1}(\eta_i)=E(y_i|x_i^T,\beta), \quad (5)$$

The link function of this distribution of exponential family with canonical parameter θ conjugating for this variable can be obtained as below :

$$g(\mu_i)=\theta_i=\eta_i=x_i^T\beta$$

Also, there are not requirements about link function to be derived from probability density function. (Brown and Prescott, 2015)

Lindley`s Regression Model

The general form for Lindley`s regression can be obtained by knowing if Lindley`s distribution whether it belong to exponential family. The probability density function for this distribution could written as below.

$$f(y,\lambda)=e^{-\left[y\lambda-\ln\left[\frac{\lambda^2}{\lambda+1}\right]+\ln[1+y]\right]} \quad (6)$$

where $a = -1$, $b = \ln[\frac{\lambda^2}{\lambda+1}]$ and $c = \ln[1 + y]$ with unit link function, that is:

$$g(\lambda) = \lambda$$

As long as we have arbitrary sample taken from Lindley`s regression with probability density function which defined in the equation (1), so it will be :

$$L(\boldsymbol{\beta}) = exp\left\{\sum_{i=1}^{n}\left[ln\left[\frac{(\boldsymbol{x}_i^T\boldsymbol{\beta})^2}{\boldsymbol{x}_i^T\boldsymbol{\beta}+1}\right] - y_i\boldsymbol{x}_i^T\boldsymbol{\beta} + ln[1+y_i]\right]\right\}, \qquad (7)$$

By taking natural logarithm in (7), we obtain:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} ln\left[\frac{(\boldsymbol{x}_i^T\boldsymbol{\beta})^2}{\boldsymbol{x}_i^T\boldsymbol{\beta}+1}\right] - \sum_{i=1}^{n} y_i\boldsymbol{x}_i^T\boldsymbol{\beta} + \sum_{i=1}^{n} ln[1+y_i], \qquad (8)$$

The idea of method of Maximum Likelihood is maximizing of natural logarithm of likelihood function (Hussain et al., 2022) which defined in equ. (8) with respect to parameters vector of regression β. So it is getting estimation of these parameters, where the first derivative for natural logarithm function is Likelihood function with respect to vector β, as below:

$$\frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\frac{(\boldsymbol{x}_i^T\boldsymbol{\beta}+1)}{(\boldsymbol{x}_i^T\boldsymbol{\beta})^2}.\frac{2(\boldsymbol{x}_i^T\boldsymbol{\beta})\boldsymbol{x}_i(\boldsymbol{x}_i^T\boldsymbol{\beta}+1) - (\boldsymbol{x}_i^T\boldsymbol{\beta})^2\boldsymbol{x}_i}{(\boldsymbol{x}_i^T\boldsymbol{\beta}+1)^2} - \sum_{i=1}^{n} y_i\boldsymbol{x}_i,$$

$$\frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\frac{\boldsymbol{x}_i\left(2(\boldsymbol{x}_i^T\boldsymbol{\beta}) + 2 - (\boldsymbol{x}_i^T\boldsymbol{\beta})\right)}{(\boldsymbol{x}_i^T\boldsymbol{\beta})(\boldsymbol{x}_i^T\boldsymbol{\beta}+1)} - \sum_{i=1}^{n} y_i\boldsymbol{x}_i,$$

$$\frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\frac{\boldsymbol{x}_i(\boldsymbol{x}_i^T\boldsymbol{\beta}+2)}{(\boldsymbol{x}_i^T\boldsymbol{\beta})(\boldsymbol{x}_i^T\boldsymbol{\beta}+1)} - \sum_{i=1}^{n} y_i\boldsymbol{x}_i, \qquad (9)$$

By making the first derivative equal to zero, then

$$\frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\frac{\boldsymbol{x}_i(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}+2)}{(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}+1)} - \sum_{i=1}^{n} y_i\boldsymbol{x}_i = \boldsymbol{0}. \qquad (10)$$

We can not find a closed form foe estimators, so we use Newton- Raphson method. Then we find solution vector for equations system $g(\hat{\beta})$ = 0, where $g\ (\hat{\beta})= \frac{\partial l(\beta)}{\partial\beta}$ by using the following equation:

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \widehat{\boldsymbol{\beta}}^{(k)} - \left[J(\widehat{\boldsymbol{\beta}}^{(k)})\right]^{-1} g(\widehat{\boldsymbol{\beta}}^{(k)}), \qquad (11)$$

$\hat{\beta}^{(k+1)}$ represents parameters estimations which we want to find at the present frequency (k+1) and $\hat{\beta}^{(k)}$ represents the vector in the preceding frequency (k), while $g(\hat{\beta}^{(k)}$ means the equation in (10) after substituting parameters estimations vector with preceding values. $J(\hat{\beta}^{(k)})$ represents partial derivatives matrix for (10) after substituting the preceding values, as follows:

$$J(\widehat{\boldsymbol{\beta}}^{(k)}) = \frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\boldsymbol{T}}} = \sum_{i=1}^{n}\frac{(\boldsymbol{x}_i\boldsymbol{x}_i^T)\left[-(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}) - 4(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}) - 2\right]}{(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})^2(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}+1)^2},$$

$$J\left(\widehat{\boldsymbol{\beta}}^{(k)}\right) = -\sum_{i=1}^{n} \frac{(\boldsymbol{x}_i\boldsymbol{x}_i^T)\left[(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}})\{\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} + \boldsymbol{4}\} + 2\right]}{\left(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}\right)^2\left(\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} + 1\right)^2}. \quad (12)$$

Equation (11) does to find the optimal solution where here the preceding values converging to new values, satisfying the following condition (Epperson, 2021) :

$$\left\|\widehat{\boldsymbol{\beta}}^{(k+1)} - \widehat{\boldsymbol{\beta}}^{(k)}\right\| < \delta$$

where |(|.|)| means the Euclidean norm when $0<\delta<1$.

3. Results and discussion

The collected data by (Salman, 2024) is depended which show influence many variables in period of 200- patient remain. Some statistic properties for study variables are shown below:

Table 1. Descriptive Statistics for Dependent and Independent Variables Kolmogorov-Smirnov Test is done to ensure that the independent variable follows Lindley and exponential distributions. The results are seen below:

Table 1. Kolmogorov-Smirnov results

Table 1. Descriptive Statistics for Dependent and Independent Variables

Variables	Min	Median	Max	Mean	Std Dev	Skewness	Kurtosis
Duration of stay in months (Y)	0.2667	1.7833	11.9000	2.7478	2.6328	1.5165	4.7309
Age in months (X1)	188.1000	788.9500	1244.8333	777.9243	197.3062	-0.3586	2.7387
Gender (X2)	1.0000	1.0000	2.0000	1.4600	0.4996	0.1605	1.0258
Job (X3)	1.0000	5.0000	9.0000	4.0100	2.7049	0.1055	1.5532
Area (X4)	1.0000	6.0000	19.0000	6.8450	5.6605	0.5592	2.0975
Tumor grade (X5)	2.0000	9.0000	9.0000	8.7600	1.1830	-4.7495	23.7438
Chemotherapy (X6)	1.0000	9.0000	9.0000	5.6300	3.7501	-0.2252	1.0826
Radiotherapy (X7)	1.0000	9.0000	9.0000	6.0200	3.6725	-0.4309	1.2165
Tumor behavior (X8)	0.0000	3.0000	3.0000	2.3050	1.1127	-1.2144	2.8164
Tumor extent (X9)	1.0000	9.0000	9.0000	7.3100	2.6926	-1.4187	3.3879

The table 2 suggests that data probe both of distributions because p-value is greater than significance level $\alpha=0.05$. The following diagram states the histogram with fit for both distributions:

Table 2. Kolmogorov-Smirnov results

Distribution	Statistic	p-value
Lindley	0.09249	0.06533
Exponential	0.09361	0.06007

The table 2 suggests that data probe both of distributions because p-value is greater than significance level $= 0.05$. The following diagram states the histogram with fit for both distributions:

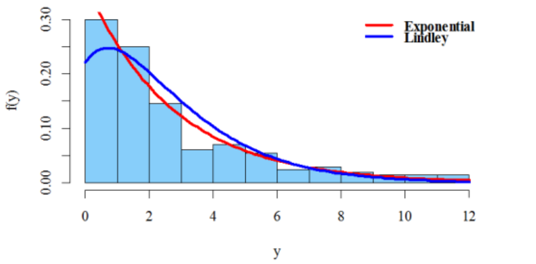


Figure 2. The histogram and the fit of Exponential and Lindley distributions

Parameters estimates of regression model depended on the exponential distribution with three different link functions like, Logarithm, inverse and identity. Also are estimated with regression depend on Lindley distribution with Identity link function. It is a map obtained from differentiation with Lindley regression. The following table represents results of parameters estimates for each model with values AIC and BIC for each one:

Table 3. Parameter Estimates and Model Selection Criteria

Table 3. Parameter Estimates and Model Selection Criteria

Model	EXP-Identity	EXP-Log	EXP-Inv	Lind-Identity
β0	4.848452	1.061931	1.155518	1.443182
β1	0.000195	0.000244	-0.000122	-0.000172
β2	-0.719585	-0.331961	0.093492	0.136879
β3	-0.066265	-0.024925	0.006650	0.009785
β4	-0.098968	-0.052963	0.023895	0.033609
β5	0.118781	0.151534	-0.137292	-0.166029
β6	-0.093011	-0.031038	0.009613	0.013993
β7	0.017900	0.000236	0.001681	0.002220
β8	-0.249630	-0.096140	0.026699	0.039128
β9	-0.049340	-0.043558	0.018520	0.025990
AIC	791.537226	787.432231	787.754757	776.973304
BIC	824.520399	820.415405	820.737930	809.956478

The above table shows a comparison among four regression models which are EXP-Identity, EXP. Log, EXP- Inv and Lind-Identity. Each of these explains shows a different method to estimate parameters depending on an certain distribution and link function. The calculated values in this table parameters estimates (β_0-β_9) for each model with BIC and AIC criterions used to evaluate efficiency of models.

Depending on parameters estimates, it is seen that there are noticeable differences between models. For example, in Lind-Identity model, parameters estimates seem more stable if it is compared with models that depend on exponential distribution. This refers that Lindley`s distribution is may be more compatible with these data especially if data has a certain properties like positively skewed.

Each of AIC and BIC shows that Lind-Identity model satisfies least values proving it is the perfect in representing data. For example, value of AIC in Lind-Identity model is 776.973304, which is less than other values for the rest models. This shows that the model is well-balanced and BIC in the same model supports this result.

It is important to refer that the other three models which depend on the exponential distribution does not state much improvement when changing the link function among linear, logarithm and inverse functions, This means that there is an influence of link function on model quality is limited in exponential distribution.

According to these results, it is seen that using Lindley`s distribution with linear link function is the best choice to analyse these data. This distribution suggests high efficiency in statistic explanation for data and shows ability of Lindley`s distribution to introduce accurate results comparing with other models that depend on exponential distribution.

4. Conclusions

Results of this work show efficiency of regression model that depends on Lindley`s distribution compared with other models which depends on exponential distribution by using AIC and BIC that sure this fact. This model has stability of parameters estimation showing the relation between independent variables and dependent variable. Also it shows that usage linear link function for Lind- Identity has efficiency if it compared with exponential functions like logarithmic and exponential functions. In addition, Kolmogorov-Smirnov test states data follows Lindley and exponential distributions despite of Lindley shows accurate data which sure its validity. In applied part, the real data for 200 patients shows Lindley`s ability to deal with effectively appearing it can be used with data has positive skewed or inhomogeneous distributions. Further tests like reminder analysis and variance tests are advised to ensure appropriateness of this model. In future, my view is to integrate more variables or finding out new link functions to improve Lindley`s efficiency explaining other subjects.

References

[1] Brown, H., & Prescott, R. (2015). Applied mixed models in medicine. John Wiley & Sons.

[2] Dobson, A. J., & Barnett, A. G. (2018). An introduction to generalized linear models. CRC press.

[3] Epperson, J. F. (2021). An introduction to numerical methods and analysis. 3rd Edition. John Wiley & Sons.

[4] Ghitany, M. E., Atieh, B., & Nadarajah, S. (2008). Lindley distribution and its application. Mathematics and computers in simulation, 78(4), 493-506.

[5] Hardin, J. W., Hardin, J. W., Hilbe, J. M., & Hilbe, J. (2007). Generalized linear models and extensions. 4th Edition. Stata press.

[6] Hussain, E. A., Al-Shallawi, A. N., & Saied, H. A. (2022). Using maximum likelihood method to estimate parameters of the linear regression t truncated model. NTU Journal of Pure Sciences, 1(4), 26-34.

[7] Jodrá, P. (2010). Computer generation of random variables with Lindley or Poisson–Lindley distribution via the Lambert W function. Mathematics and Computers in Simulation, 81(4), 851-859.

[8] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. Journal of the Royal Statistical Society. Series B (Methodological), 102-107.

[9] Salman, A. S. M. (2024). Using the Bayesian method to estimate some survival regression models with the application. Master Thesis, Mustansiriyah University. Iraq.