Research Article

Estimation of the Parameters of Mixed Frechet Distribution and Its Employment in Simple Linear Regression

Ahmed Dheyab Ahmed[1], Baydaa Ismael Abdulwahhab[2], Ebtisam Karim Abdulah[3]

[1]College of Administration and Economics, University of Baghdad, Statistics Department
[2]College of Administration and Economics, University of Baghdad, Statistics Department
[3]College of Administration and Economics, University of Baghdad, Statistics Department
[1]ahmedthieb19@gmail.com,[2] baidaa29@yahoo.com,[3]ebtisamsa@yahoo.com

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: Probability distribution is important to those who are interested in the subject of mathematical statistics, so in this research a probability function of the mixed distribution Frechet-Frechet was formed and its parameters were estimated using maximum likelihood and moment methods. These methods are applied on real data which represents the number of cancer cases in Iraq for the years (1991-2016).

Keywords: Mixed Frechet distribution, Maximum likelihood method, Moment method, Central moment, Cancer diseases.

1. Introduction

A noticeable increase of cancer disease in Iraq puts people at risk of death. The main type of cancers seen are lung cancer, stomach cancers, liver cancers, breast cancers and brain tumors. The main reason of these cases being widely seen is the past wars that have affected the environment and atmosphere that the Iraqi people live in, where they experienced nuclear weapons and radiation. These are highly poisonous and can be lethal to humanity or can cause future diseases such as the one noted in this research.

The estimation of probability distribution is important to researchers whom are majoring in mathematical statistic. This is due to the development of estimation method which will play a role in getting best estimator for the parameters. The Frechet distribution is one of the most widely and differently used distributions in many fields and different data. This data could be homogenous or nonhomogenouty and have different probability or same distribution but with different parameters. This will reflect on the population and divide to subpopulations. For each subpopulation, there is a probability density function for each population. However by merging this subpopulation it will give us mixed distribution. There are many researchers who studied this mixture distribution among them; Bader[6], Jaheen[9], Nassar[11], Abbas[1], Adeoye[2]and others.
By assuming the random variable X has Frechet distribution[8] with the following probability density function:

$$f(x,\theta) = \theta\, x^{-2}\, e^{-\left(\frac{x}{\theta}\right)^{-1}} \qquad , \qquad x \geq 0\ ,\ \ \theta \succ 0 \qquad .... (1)$$

Where θ is a shape parameters, and the cumulative density function is:

$$F(x,\theta) = e^{-\left(\frac{x}{\theta}\right)^{-1}} \qquad ...... (2)$$

The importance of this research is through forming probability density for mixed distribution by merging two Frechet distributions. The estimated parameters of this mixed distribution are done byusing maximum likelihood and moment methods.

2. Mixed distribution

This is a distribution that results from mixing two or more distributions based on a determined proportion from each population; however in this research just two distributions are assumed. The probability density function for mixed distribution can be found from the following forms [7]:

$$f(x) = \zeta\ f_1(x) + (1-\zeta)\, f_2(x) \qquad ...... (3)$$

Where ζ represents mixing proportion parameter and isaproportion of subpopulation from the origin population, such that $\sum_{i=1}^{n} \zeta_i = 1$ and $0 \prec \zeta \prec 1$,so the probability density function for mixed distribution will be as follows:

$$L(x_1, x_2, ..., x_n, \alpha, \theta_1, \theta_2) = \prod_{i=1}^{n} f(x_1, x_2, ..., x_n, \alpha, \theta_1, \theta_2) \qquad ....(4)$$

The cumulative distribution functionfor the mixed distribution can be found *by* the following form:

$$F(x) = \zeta\, F_1(x) + (1-\zeta) F_2(x) \qquad ...... (5)$$

So, the cumulative distribution function for the mixed distribution as the following forms:

$$F(x, \zeta, \theta_1, \theta_2) = \zeta e^{-\left(\frac{x}{\theta_1}\right)^{-1}} + (1-\zeta) e^{-\left(\frac{x}{\theta_2}\right)^{-1}} \qquad .... (6)$$

3. The central moment about origin point

$$E(x^r) = \int_0^\infty x^r f_1(x) d_1 x + \int_0^\infty x^r f_2(x) d_2 x = m_r \qquad ...... (7)$$

$$E(x^r) = \int_0^\infty x^r \zeta \theta_1\, x^{-2}\, e^{-\left(\frac{x}{\theta_1}\right)^{-1}} d_1 x + \int_0^\infty x^r (1-\zeta) \theta_2\, x^{-2}\, e^{-\left(\frac{x}{\theta_2}\right)^{-1}} d_2 x$$ Let

$$y = \left(\frac{x}{\theta}\right)^{-1} \Rightarrow x = \frac{\theta}{y} \Rightarrow dx = -\frac{\theta}{y^2} dy$$

$$m_r = \zeta \int_0^\infty \left(\frac{\theta_1}{y}\right)^r \theta_1 \left(\frac{\theta_1}{y}\right)^{-2} e^{-y^{-1}} \left(-\frac{\theta_1}{y^2}\right) d_1 y + (1-\zeta) \int_0^\infty \left(\frac{\theta_2}{y}\right)^r \theta_2 \left(\frac{\theta_2}{y}\right)^{-2} e^{-y^{-1}} \left(-\frac{\theta_2}{y^2}\right) d_2 y$$

$$m_r = -\left(\zeta \int_0^\infty \left(\frac{\theta_1}{y}\right)^r e^{-y^{-1}} d_1 y + (1-\zeta) \int_0^\infty \left(\frac{\theta_2}{y}\right)^r e^{-y^{-1}} d_2 y \right)$$

$$m_r = -\zeta \theta_1^r \Gamma(r+1) - (1-\zeta) \theta_2^r \Gamma(r+1) \qquad ...... (8)$$

when r = 1

$$m_1 = -\zeta \theta_1 - (1-\zeta) \theta_2 \qquad ...... (9)$$

when r = 2

$$m_2 = -2\zeta \theta_1^2 - 2(1-\zeta) \theta_2^2 \qquad ...... (10)$$

when r = 3

$$m_3 = -6\zeta \theta_1^3 - 6(1-\zeta) \theta_2^3 \qquad ...... (11)$$

Maximum Likelihood Method (MLM):

This method is one of the well-known methods for estimating the parameters of probability distributions, these estimators have consistentand stability properties. By assuming $(x_1, x_2, ..., x_n)$ is a random sample ~ mixed Frechet distribution, the maximum likelihood is as follows

$$L(x_1, x_2, ..., x_n, \zeta, \theta_1, \theta_2) = \prod_{i=1}^{n} f(x_1, x_2, ..., x_n, \zeta, \theta_1, \theta_2) \qquad ...... (12)$$

$$L(x_1, x_2, ..., x_n, \zeta, \theta_1, \theta_2) = \prod_{i=1}^{n} \zeta \theta_1\, x^{-2}\, e^{-\left(\frac{x}{\theta_1}\right)^{-1}} + (1-\zeta) \theta_2\, x^{-2}\, e^{-\left(\frac{x}{\theta_2}\right)^{-1}}$$

$$x \geq 0 \quad , \theta_1 \succ 0, \theta_2 \succ 0$$

$$L(x_1,x_2,...,x_n,\zeta,\theta_1,\theta_2)=\prod_{i=1}^{n}\left(\zeta\, x^{-2}\,\theta_1 e^{-\left(\frac{x}{\theta_1}\right)^{-1}}+(1-\zeta)x^{-2}\,\theta_2 e^{-\left(\frac{x}{\theta_2}\right)^{-1}}\right),$$

$$x\geq 0\quad,\theta_1\succ 0,\theta_2\succ 0$$

By taking natural logarithm of both sides:

$$LnL(x_1,x_2,...,x_n,\zeta,\theta_1,\theta_2)=\sum_{i=1}^{n}Ln\left(\zeta\, x^{-2}\,\theta_1 e^{-\left(\frac{x}{\theta_1}\right)^{-1}}+(1-\zeta)x^{-2}\,\theta_2 e^{-\left(\frac{x}{\theta_2}\right)^{-1}}\right)\quad ......(13)$$

After taking partial derivatives for distribution parameters and setting it equal to zero:

$$\frac{\partial LnL(x_1,x_2,...,x_n,\zeta,\theta_1,\theta_2)}{\partial\theta_1}=\sum_{i=1}^{n}\left(\frac{\left(-\zeta\, x^{-3}\,\theta_1\, e^{-\left(\frac{x}{\theta_1}\right)^{-1}}+\zeta x^{-2}\, e^{-\left(\frac{x}{\theta_1}\right)^{-1}}\right)}{\left(\zeta\, x^{-2}\,\theta_1 e^{-\left(\frac{x}{\theta_1}\right)^{-1}}+(1-\zeta)x^{-2}\,\theta_2 e^{-\left(\frac{x}{\theta_2}\right)^{-1}}\right)}\right)=0\quad .... (14)$$

$$\frac{\partial LnL(x_1,x_2,...,x_n,\zeta,\theta_1,\theta_2)}{\partial\theta_2}=\sum_{i=1}^{n}\left(\frac{\left(-(1-\zeta)x^{-3}\,\theta_2\, e^{-\left(\frac{x}{\theta_2}\right)^{-1}}+(1-\zeta)x^{-2}\, e^{-\left(\frac{x}{\theta_2}\right)^{-1}}\right)}{\left(\zeta\, x^{-2}\,\theta_1 e^{-\left(\frac{x}{\theta_1}\right)^{-1}}+(1-\zeta)x^{-2}\,\theta_2 e^{-\left(\frac{x}{\theta_2}\right)^{-1}}\right)}\right)=0\quad ...... (15)$$

$$\frac{\partial LnL(x_1,x_2,...,x_n,\zeta,\theta_1,\theta_2)}{\partial\zeta}=\sum_{i=1}^{n}\left(\frac{\left(x^{-2}\theta_1\, e^{-\left(\frac{x}{\theta_1}\right)^{-1}}\right)-\left(x^{-2}\theta_2\, e^{-\left(\frac{x}{\theta_2}\right)^{-1}}\right)}{\left(\zeta\, x^{-2}\,\theta_1 e^{-\left(\frac{x}{\theta_1}\right)^{-1}}+(1-\zeta)x^{-2}\theta_2\, e^{-\left(\frac{x}{\theta_2}\right)^{-1}}\right)}\right)=0\quad .... \quad (16)$$

the equations (13), (14), and (15) will be solved by using (fsolve) function which exist in MATLAB software in order to obtain the parameters estimators $\hat{\theta}_{1MLM}$, $\hat{\theta}_{2MLM}$, ζ_{MLM} respectively.

Moments Method (MOM):

This method depends on the equality of the sample size with population size to obtain the equations for the parameters to be estimated as follows:

$$m_1=\frac{\sum_{i=1}^{n}x_i}{n}=-\zeta\theta_1-(1-\zeta)\theta_2\quad ...... (17)$$

$$m_2=\frac{\sum_{i=1}^{n}x_i^2}{n}=-2\zeta\theta_1^2-2(1-\zeta)\theta_2^2\quad ...... (18)$$

$$m_3 = \frac{\sum_{i=1}^{n} x_i^3}{n} = -6\zeta\theta_1^3 - 6(1-\zeta)\theta_2^3 \qquad ......(19)$$

the equations (16), (17), and (18) will be solved by using (fsolve) function which exist in MATLAB software in order to obtain the parameters estimators $\hat{\theta}_{1MOM}$, $\hat{\theta}_{2MOM}$, ζ_{MOM} respectively.

4. Simple linear regression model

Suppose the simple linear regression model as follows:

$$y = b_0 + b_1 w + u \qquad ...... (20)$$

Where y is a dependent variable distributed with mixed Frechet distribution, w is an explanatory variable, b_0, b_1 are regression parameters, so the mixed Frechet distribution can be employed in linear regression model. This can be seenin many researches when the errors are not normally distributed like Ahmed [3], Ebtisam[8],Ahmed [5], and others.

5. Applied experiment

Table (1) shows the number of cases of cancer diseases in Iraq for the years (1991-2016).Kolmogorov test is used to test the data and seems that it is distributed with Frechet distribution where the value of the test equal (D = 0.2624) less than the tabulated value (0.2667) at significant level $\alpha = 0.05$.

T*able*1 : Number of Cases of Cancer Diseases in Iraq (1991-2016) *(1)*

years	Cancer cases	years	Cancer cases
1991	5720	2004	14520
1992	8526	2005	15172
1993	8471	2006	15226
1994	7785	2007	14213
1995	7948	2008	14180
1996	8360	2009	15251
1997	8592	2010	18482
1998	9033	2011	20278
1999	8963	2012	21101
2000	10888	2013	23308
2001	13332	2014	25598
2002	13985	2015	25269
2003	11248	2016	25556

The parameters of mixed Frechet distribution are estimated by (MLM) and (MOM) methods as shown in Table 2.

Table 2:The Estimated Parameters of the Mixed FrechetDistribution

method	ζ	θ_1	θ_2
MLM	1	1.0149	1.0150
MOM	1.1835	0.6753	12.1332

The values of estimated parameters of simple linear regression model can be employed as initial values for the random error distribution.

6. Conclusion

Through mixing the two Frechet distributions, a new distribution is obtained, which is mixed Frechet distribution. The parameters are estimated through the estimated formulas that have been derived from the theory sections. Maximum likelihood and moments methods are used for this estimation and for the data which representsthe number of cancer cases in Iraq.

References

A. Abbas, S. and others (2020). Exponentiated Exponential – Exponentiatd Weibull Linear Mixed Distribution. PakistanJournal Statistical OperatingRresearch, 16(3), 517-527.

B. Adeoye, A. and others (2019). On The Exponential-Gamma-X Mixed Distribution. International Journal of Research and Innovation in Applied Science, 10(3), 139-141.

C. Ahmed, D., Ebtisam, A., Baydaa, A. and Noura, O. (2020). Solving multicollinearity problem of gross domestic product using ridge regression method. Periodicals of Engineering and Natural Sciences, 8(2), 668-672.

D. Ahmed, D., Ebtisam, A. andBaydaa, A. (2020).Estimating the Coefficients of Polynomial Regression Model when the Error Distributed Generalized Logistic with Three Parameters.International Journal of Psychosocial Rehabilitation, 24(8), 4487-4496.

E. Ahmed, D.,Baydaa, A. and Ebtisam, A. (2020).A comparison among Different Methods for Estimating Regression Parameters with Autocorrelation Problem under Exponentially Distributed Error.Baghdad Science Journal, 17(3), 980-987.

F. Badr, M. and Shawky, A.I. (2014). Mixture of ExponentiatedFrechet Distribution. Life Science Journal, 11(3), 392-404.

G. Badr, M. (2015). Mixture of exponentiatedFrechet distribution based on upper record values. Journal of American Science , 11(7), 56-64.

H. Ebtisam, A., Ahmed, A. and Baydaa, A. (2019). Comparison of Estimate Methods of Multiple Linear Regression Model with Auto-Correlated Errors when The Error distributed with General Logistic. Journal of Engineering and Applied Sciences, 14(19), 7072-7076.

I. Jaheen, Z.F. (2005). On record statistics from amixture of two exponential distributions.Journal of Statistical omputation and Simulation, 75(1), 1-11.

J. Kamran, A. and Tang, Y. (2012). Comparison of Estimation Methods for Frechet Distribution with Known Shape. Caspian Journal of Applied Sciences Research, 1(10), 58-64.

K. Nassar, M.M. (1988). Two properties of mixtures of exponential distributions. IEEE Transactions on Reliability, 37(4), 383-385.