Journal of Advances in Mathematics and Computer Science

36(9): 30-36, 2021; Article no.JAMCS.75856

ISSN: 2456-9968

(Past name: British Journal of Mathematics & Computer Science, Past ISSN: 2231-0851)

Statistical Modelling with Birnbaum-Saunders Distribution

Nuri Celik[1*]

[1] *Department of Mathematics, Gebze Technical University, 41400, Kocaeli, Turkey.*

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/JAMCS/2021/v36i930399
Editor(s):
(1) Dr. Metin Basarir, Sakarya University, Turkey.
Reviewers:
(1) B. Roopashri Tantri, India.
(2) V. Nirmala, Sri Krishna Arts and Science College, India.
Complete Peer review History: https://www.sdiarticle4.com/review-history/75856

Received: 17 August 2021
Accepted: 18 October 2021
Published: 27 October 2021

Original Research Article

Abstract

In this article, it is assumed that the distribution of the error terms is the Birnbaum-Saunders distribution in the process of one-way ANOVA. The Birnbaum-Saunders distribution has been widely used in reliability analysis especially in fatigue-life models. In reliability analysis, nonnormal distribution is much more common than the normal distribution. We obtain the estimation of the parameters og interest by maximum likelihood method. We also propose new test statistics based on these estimators . The efficiencies of the maximum likelihood estimators and the Type I errors obtained by using the proposed estimators are compared with normal theory via Monte Carlo simulation study. At the end of the study, the real life example is given just for the illustration of the method.

Keywords: Statistical modelling; Birnbaum-Saunders distribution; ANOVA; maximum likelihood.

2010 Mathematics Subject Classification: 62E10, 62K25.

1 Introduction

The Birnbaum-Saunders (BS) distribution was proposed in order to solve the problems of vibration in commercial aircrafts causing fatigue in materials. It is also called as the fatigue-life distribution.

**Corresponding author: E-mail: nuricelik@gtu.edu.tr;*

However, BS distribution has been widely used for the situations where the accumulation of a certain factor forces a quantifiable characteristic to exceed a critical threshold, such as migration of metallic flaw, generation of action potentials, late human mortality and earthquakes and tsunamis. The distribution was firstly given by [1] and it was originally obtained by [2] according to paper written by [3]. Additionally, in reliablity and life testing it was proposed by [4]. However, as a clear explanation the BS distribution was originally derived by [5].

Any random variable with BS distribution can be obtained by the transformation of standard normal random variable Z as

$$Z = \frac{1}{\lambda}\Big(\sqrt{\frac{T}{\sigma}} - \sqrt{\frac{\sigma}{T}}\Big) \tag{1.1}$$

where T is the random variable whose disrtibution is the BS with shape parameter $\lambda > 0$ and scale parameter $\sigma > 0$, then it is notated as $T \sim BS(\lambda, \sigma)$. The probability density function (pdf) of T is given by

$$f(t, \lambda, \sigma) = \frac{1}{\sqrt{2\pi}} exp\Big(-\frac{1}{2\lambda^2}\Big(\frac{t}{\sigma} + \frac{\sigma}{t} - 2\Big)\Big)\frac{1}{2\lambda\sigma}\Big(\sqrt{\frac{t}{\sigma}} - \sqrt{\Big(\frac{\sigma}{t}\Big)^3}\Big) \tag{1.2}$$

The cumulative distribution function (cdf) of BS distribution can be obtained by using the transformation given in equation (1.1) as,

$$F(t, \lambda, \sigma) = \Phi\Big(\frac{1}{\lambda}\Big(\sqrt{\frac{t}{\sigma}} - \sqrt{\frac{\sigma}{t}}\Big)\Big) \tag{1.3}$$

The expected value and the variance of the BS distribution are

$$E(T) = \sigma\Big(1 + \frac{\lambda^2}{2}\Big) \tag{1.4}$$

and

$$V(T) = \sigma^2\lambda^2\Big(1 + \frac{5}{4}\lambda^2\Big) \tag{1.5}$$

respectively. It can be easily noticed that the distribution is positively skewed and unimodal. It can be also added that when λ converges to 0, the distribution becomes symmetric. The detailed information about the distribution characteristics, features, derivation and the application area of BS distribution can be found in [6].

It is generally assumed that the error terms have normal distribution with mean 0 and variance σ^2 in the context of statistical modeling. Based on normality assumption, the parameter of interest are estimated by using the least square (LS) estimation method. Additionally, the model significance is tested by using F statistics by using the corresponding LS estimators. It is also known that when the normality assumption is satisfied, the LS estimates are optimal and the test statistics based on them are most powerful. However, in literature, there are plenty of studies ([7], [8], [9] and [10]) showing that the non-normal distributions are more prevalent than the normal distribution, and there are so many works ([11], [12] and [13]) using these nonnormal distributions as the error terms of the statistical modelling. As a result of these papers it can be concluded that Type I error of the test statistic based on normal distribution is not much different from the nonnormal distribution by the help of central limit theorem. Additionally, power value of the test statistics based on noemal distribution is lower than the test statistics based on nonnormal distributions because of the inefficiency of the sample mean. Therefore, it is recommended to obtain robust test statistics according to these properties.

In this study, we assume that the error distribution of one-way ANOVA model as BS distribution. The reason for assuming BS distribution as an error distribution is that it is widely used in reliability analysis especially for fatigue data. Furthermore, there is no work in the statistics literature assuming BS distribution as an error distribution.

The rest of the paper is organized as follows. In Section 2, one-way ANOVA model is studied by using BS distribution as an error distribution. The model parameters and the test statistics based on them are estimated by using maximum likelihood (ML) method. In Section 3, for some representative parameter (α) values, these estimators are compared by using relative efficiency (RE) criteria and the test statistics are compared by using the powers . A real data sets is analysed in order to show the performances of these models in Section 4. Conclusion is given at the end of this paper.

2 One-way ANOVA

Consider the following one-way ANOVA model,

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad i = 1, 2, ..., a; \; j = 1, 2, ..., n \tag{2.1}$$

where y_{ij} is the response of the jth observation in the ith treatment, μ is the grand mean, α_i is the effect of ith treatment and ϵ_{ij} are the independently and identically distributed (iid) error terms. It is generally assumed that the error distribution is normal with mean zero and constant variance σ^2. Consider the model (1.2) and assume the distribution of ϵ_{ij} is $BS(\lambda, \sigma)$.

$$f(\epsilon, \lambda, \sigma) = \frac{1}{\sqrt{2\pi}} exp\Big(- \frac{1}{2\lambda^2}\Big(\frac{\epsilon}{\sigma} + \frac{\sigma}{\epsilon} - 2\Big)\Big) \frac{1}{2\lambda\sigma}\Big(\sqrt{\frac{\epsilon}{\sigma}} - \sqrt{\Big(\frac{\sigma}{\epsilon}\Big)^3}\Big) \tag{2.2}$$

To obtain the ML estimators of the parameters of interest in model (2.1), the loglikelihood function

$$lnL = c_1 + \frac{N}{\sigma^2} - \frac{1}{2\sigma^2}\sum_{i=1}^{a}\sum_{j=1}^{n}\Big(\frac{z_{ij}}{\lambda} + \frac{\lambda}{z_{ij}}\Big) - \frac{3}{2}\sum_{i=1}^{a}\sum_{j=1}^{n} ln(z_{ij}) - Nln(\sigma) - \frac{N}{2}ln(\lambda) + \sum_{i=1}^{a}\sum_{j=1}^{n} ln(z_{ij} + \lambda) \tag{2.3}$$

is maximized with respect to the parameters. Here $z_{ij} = \frac{y_{ij} - \mu - \alpha_i}{\sigma}$ and $N = an$. By differentiating the log-likelihood function with respect to the parameters of interest and setting them to zero we obtain the following likelihood equations

$$\begin{aligned}
\frac{\partial \ln L}{\partial \mu} &= \frac{1}{2\sigma^2}\sum_{i=1}^{a}\sum_{j=1}^{n}\Big(\frac{1}{\lambda} - \frac{\lambda}{z_{ij}^2}\Big) + \frac{3}{2}\sum_{i=1}^{a}\sum_{j=1}^{n}\frac{1}{z_{ij}} - \sum_{i=1}^{a}\sum_{j=1}^{n}\Big(\frac{1}{z_{ij} + \lambda}\Big) = 0 \\
\frac{\partial \ln L}{\partial \alpha_i} &= \frac{1}{2\sigma^2}\sum_{j=1}^{n}\Big(\frac{1}{\lambda} - \frac{\lambda}{z_{ij}^2}\Big) + \frac{3}{2}\sum_{j=1}^{n}\frac{1}{z_{ij}} - \sum_{j=1}^{n}\Big(\frac{1}{z_{ij} + \lambda}\Big) = 0 \\
\frac{\partial \ln L}{\partial \lambda} &= \frac{1}{2\sigma^2}\sum_{i=1}^{a}\sum_{j=1}^{n}\Big(\frac{z_{ij}}{\lambda^2} - \frac{1}{z_{ij}}\Big) - \frac{N}{2\lambda} + \sum_{i=1}^{a}\sum_{j=1}^{n}\Big(\frac{1}{z_{ij} + \lambda}\Big) = 0
\end{aligned} \tag{2.4}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{2N}{\sigma^3} - \frac{1}{\sigma^3}\sum_{i=1}^{a}\sum_{j=1}^{n}\left(\frac{z_{ij}}{\lambda} + \frac{\lambda}{z_{ij}}\right) - \frac{N}{\sigma} = 0.$$

Solutions of these equations are called the ML estimators. As can be seen from these equations, there is no explicit solutions have no; therefore we must use iterative methods.

In one-way ANOVA, our aim is to compare the equality of treatment effects, in other words, to test the following null hypothesis

$$H_0 = \alpha_1 = \alpha_2 = ... = \alpha_a = 0 \tag{2.5}$$

For testing the hypothesis, we traditionally use the following test statistics based on the LS estimators

$$F_{LS} = \frac{n\sum_{i=1}^{a}\tilde{\alpha}_{i,LS}^2}{(a-1)\tilde{\sigma}_{LS}^2} \tag{2.6}$$

It is known that powers of this test statistics are very sensitive to the data anomalies. Therefore, in this paper, we propose the following test statistics based on the ML estimators.

$$F_{ML} = \frac{n\sum_{i=1}^{a}\hat{\alpha}_{i,ML}^2}{(a-1)\hat{\sigma}_{ML}^2} \tag{2.7}$$

3 Simulation Study

In this part, a Monte Carlo simulation study is performed for different shape parameters(λ) of BS distribution and for different sample sizes (n). In all situation μ is taken 0, σ is taken 1 and $a = 3$ for the sake of brevity. The process is replicated by 100.000/n times for obtaining RE values for the parameters $\mu_i = \mu + \alpha_i$ and σ. As can be seen from Table 1, the performances of ML estimates increase when the shape parameter λ increases for the parameter μ_i. The same interpretation can be made for the parameter σ by using Table 2.

Table 1. RE values for $\mu_i, a = 3$

		$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1.0$	$\alpha = 2.0$	$\alpha = 5.0$
	5	97	85	80	73	66
n	10	94	83	78	**69**	62
	15	91	81	74	65	59

Table 2. RE values for $\sigma, a = 3$

		$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 1.0$	$\alpha = 2.0$	$\alpha = 5.0$
	5	95	88	87	79	73
n	10	94	85	85	**78**	71
	15	90	80	82	75	70

For large n values, the distribution of the test statistics based on ML estimators is cenrtal F with $\nu_1 = a - 1, \nu_2 = N - a$ degrees of freedoms. On the other hand, we resort Monte Carlo simulation study in order to approximate the null distribution of F_{ML}. Type I errors of the test statistics based on ML estimators are simulated by using the following probability.

$$P\{F_{ML} > F_\alpha(a-1, N-a)|H_0\}$$

It can be seen from Table 3, Type I errors of F_{ML} remains stable around the true value 0.05, on the other hand Type I errors of F_{LS} increase, when the shape parameter increases.

Table 3. Simulated Type I errors of F_{LS} and F_{ML} for the treatments; $a = 3$ and $\alpha = 0.050$

	α =0.3		α =0.5		α =1.0		α =2.0		α =5.0	
n	F_{LS}	F_{ML}	F_{LS}	F_{ML}	F_{LS}	F_{ML}	F_{LS}	F_{ML}	F_{LS}	F_{ML}
5	0.049	0.055	0.057	0.048	0.065	0.048	0.069	0.048	0.075	0.045
10	0.052	0.052	0.058	0.047	0.067	0.046	0.069	0.047	0.079	0.044
15	0.055	0.054	0.060	0.052	0.067	0.045	0.071	0.048	0.081	0.045

Fig. 1 shows the power values of traditional F_{LS} statistic and the proposed test statistic F_{ML} with $\alpha = 0.3, 0.5, 1.0, 2.0$ and 5.0 respectively. As can be shown from the figure the power values of the proposed test statistics F_{ML} is much more higher than the corresponding F_{LS} test statistics.

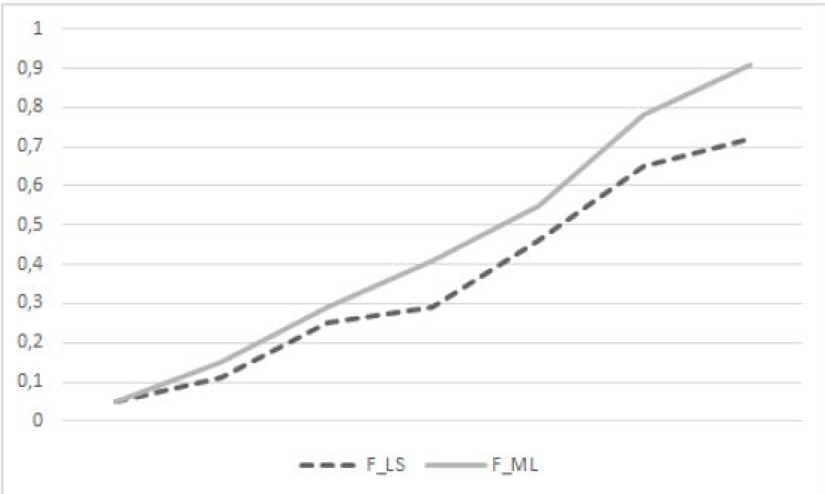

Fig. 1. Power values of F_{LS} and F_{ML}.

4 Numerical Example

As a numerical example, we consider the data given by [14]. These data are about the initial remission times of leukemia patients (in days) induced by three treatments. There are 66 patients and three treatments (25, 19 and 22 respectively).

To clarify the distribution of the error terms, we fit some selected statistical distributions to the data then calculate lnL and the Akaike information criterion (AIC) values for these distribution.

Table 4. Initial Remission Times of Leukemia Patients

1	2	3
4, 5, 9, 10, 12, 13, 10,	8, 10, 10, 12, 14,	8, 10, 11, 23, 25, 25,
23, 28, 28, 28, 29,	20, 48, 70, 75, 99, 103,	28, 28, 31, 31, 40,
31, 32, 37, 41, 41,	162, 169, 195, 220,	48, 89, 124, 143,
57, 62, 74, 100, 139,	161, 199, 217,	12, 159, 190, 196,
20, 258, 269	245	197, 205, 219

Estimates of the parameters of interest are given in Table 5. Estimations are obtained by traditional LS method and ML method based on BS distribution of error terms. Table 6 gives the test statistics and p-values obtained by LS and ML method. We consider also Kruskal Wallis test as a nonparametric method. It is clear from Table 6 that the test statistics based on LS and Kruskal Wallis test do not reject the null hypotheses corresponding to the treatments, However, the test statistics based on ML method reject the null hypothesis.

Table 5. Estimation values of parameters with LS and ML method.

	LS	ML
μ	81.902	77.335
α_1	-27.133	-29.265
α_1	25.308	26.954
α_1	1.824	2.311
σ	76.631	69.231

Table 6. The calculted test statistics and the p-values of the related tests.

$Method$	$TestStatistics$	$p-value$
LS	2.629	0.080
$Kruskal\ Wallis$	3.476	0.176
ML	11.101	0.001

5 Conclusion

In the context of ANOVA, LS methodology is used to estimate the parameters of interest. LS estimators are the best when the error distribution is normal, however when the normality assumption is not satisfied, the LS estimators loose their efficiency. In this article, we assume that the distribution of the error terms in one-way ANOVA as BS distribution. BS distribution is called as fatigue-life distribution and is widely used especially in reliability analysis. We obtain the estimation of the unknown parameters by using ML methodology. We also prpose a new test statistics based on these new estimators. The simulation results show that the ML estimators are much more efficient than the LS estimators. Simulation studies show that Type I errors of the test statistics based on ML estimators are much more accurate than the test statistics based on LS estimators. Additionally, the power values of the proposed test statistics are higher than the traditional methodology.

Competing Interests

Author has declared that no competing interests exist.

References

[1] Fletcher H. A verification of the theory of Brownian movements and a direct determination of the value of NE for gaseous ionization. Phys Rev. 1911;33:81-110.

[2] Konstantinowsky D. Elektrische ladungen und Brownsche bewegung schrkleiner metallteilchen gasen. Sitzungsberichte der KaiserlichenAkademie der Wissenschaften. 1914;123:1697-1752.

[3] Shredinger E. Zur Theorie der fall-und steigversuche an teilchenn mit Brownscher Bewegung. Physikalische Zeitschrift. 1915;16:289-295.

[4] Freudenthal AM, Shinozuka M. Structural safety under conditions of ultimate load failure and fatigue. [Technical Report].WADD-TR-61-77. Wright Air Development Division; 1961.

[5] Birnbaum ZW, Saunders SC. A new family of life distribution. J Appl Probab. 1969;6(2):319-327.

[6] Leiva V. The BirnbaumSaunders Distribution, Elsevier, London; 2016.

[7] Pearson ES. The analysis of variance in cases of nonnormal variation. Biometrika. 1932;23:114-133.

[8] Geary RC. Testing for normality. Biometrika. 1974;34:209-242.

[9] Huber PJ. Robust Statistics, Jonh Wiley, New York; 1981.

[10] Tan WY, Tiku ML. Sampling distributions in terms of Laguerre Polynomials with applications. New Age International (formerly, Wiley Eastern), New Delhi; 1999.

[11] Senoglu B, Avcoglu MD. Analysis of Covariance with Non-normal error terms. International Statistical Review. 2009;77-3:366–377.

[12] Lin TI. Robust mixture modeling using multivariate skew t distribution, Stat Comput. 2010;20:343-356.

[13] Celik N, Senoglu B. Two-way ANOVA when the distribution of the error terms is skew t, Communication in Statistics-Simulation and Computation. 2017;48(1):287-301.

[14] Lee ET, Wang JW. Statistical Methods for survival data analysis. Jonh Wiley, New York; 2003.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
https://www.sdiarticle4.com/review-history/75856